Optimum Full Information, Unlimited Complexity, Invariant and Minimax Clock Skew and Offset Estimators for IEEE 1588

Anantha K. Karthik, *Student Member, IEEE* and Rick S. Blum, *IEEE Fellow*

***Abstract*—This paper addresses the problem of clock skew and offset estimation for the IEEE 1588 precision time protocol. Built on the classical two-way message exchange scheme, IEEE 1588 is a prominent synchronization protocol for packet switched networks. Due to the presence of random queuing delays in a packet switched network, the joint recovery of clock skew and offset from the received packet timestamps can be viewed as a statistical estimation problem. Recently, assuming perfect clock skew information, minimax optimum clock offset estimators were developed for IEEE 1588. Building on this work, we first develop joint optimum invariant clock skew and offset estimators for IEEE 1588 for known queuing delay statistics and unlimited computational complexity. We then show the developed estimators are minimax optimum, i.e., these estimators minimize the maximum skew normalized mean square estimation error over all possible values of the unknown parameters. Minimax optimum estimators that utilize information from past timestamps to improve accuracy are also introduced. The developed optimum estimators provide useful fundamental limits for evaluating the performance of clock skew and offset estimation schemes. These performance limits can aid system designers to develop algorithms with the desired computational complexity that achieve performance close to the performance of the optimum estimators. If a designer finds an approach with a complexity they find acceptable and which provides performance close to the optimum performance, they can use it and know they have near optimum performance. This is precisely the approach used in communications when comparing to capacity.**

***Index Terms*—Time synchronization, IEEE 1588 Precision Time Protocol, Optimum Invariant Estimation, Minimax Estimation, Cellular networks.**

I. Introduction

Precise synchronization of events is essential to ensure the proper functioning of a distributed network as it ensures a common time frame for all the nodes in the network. The IEEE 1588 Precision Time Protocol (PTP) [1] is a popular time synchronization protocol for synchronizing slave clocks to a master clock. It is cost effective and offers accuracy comparable to Global Positioning System (GPS)-based timing. PTP is utilized in various applications including electrical grid networks [2], cellular base station synchronization in 4G Long Term Evaluation (LTE) [3], substation communication networks [4] and industrial control [5]. In this paper, we will develop clock synchronization algorithms for PTP in a packet switched network.

The clock time at the slave node can be modeled mathematically, as a function $c(t)$ of the master node's clock time t. When the clocks at the slave and master node are synchronized, then $c(t) = t$. However, in practice these clocks are not synchronized, implying a synchronization error $e(t) = |c(t) - t|$, that tends to grow over large time scales. In general, the clock time of the slave node is modeled as $c(t) = \phi t + \delta$ [6]–[10], where ϕ and δ denote the relative clock skew and offset of the slave's clock time with respect to the master's clock time respectively.

A number of time synchronization protocols including PTP, Timing Protocol for Sensor Networks (TPSN) [11], and Lightweight Time Synchronization (LTS) [12] are built on the classical two-way message exchange scheme. In these protocols, the slave node exchanges a series of synchronization packets with the master node and uses the packet timestamps to estimate ϕ and δ. The messages traveling between the master and slave nodes can encounter several intermediate switches and routers, accumulating delays at each node. The main factors contributing to the overall delay are: (1) the fixed propagation and processing delays at the intermediate nodes along the network path between the master and slave nodes and (2) the random queuing delays at each such node. This randomness in the overall network traversal time is referred to as Packet Delay Variation (PDV) [10], and the problem of estimating ϕ and δ, while combating the noise in the observations that occurs due to PDV is called the "Clock Skew and Offset Estimation" (CSOE) problem. Maximum-likelihood (ML)-based CSOE schemes have been proposed in [6]–[8].

Previously, assuming complete knowledge of the clock skew and a known affine relationship between the fixed path delays, members of our research team studied estimators for clock offset. In particular, Guruswamy *et al.* [10] developed optimum invariant clock offset estimators for PTP under the squared error loss function. Further, in [13], we developed robust clock offset estimation schemes for PTP in the presence of unknown path asymmetries. In [14], Guruswamy *et al.* developed an approximate approach for estimating the clock skew and offset. However, the approximate approach is not optimum. In this paper, assuming complete knowledge of the statistical information describing the PDV and unlimited computational complexity, we develop the joint optimum invariant clock skew and offset estimators for PTP.

To study the CSOE problem, we consider three observation

This work was supported by the Department of Energy under Award DE-OE0000779, and by the National Science Foundation under Grants ECCS 1744129 and CNS-1702555.

Anantha K. Karthik and Rick S. Blum are with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: akk314@lehigh.edu; rblum@lehigh.edu).

models, namely the known fixed delays model (*K-model*), the standard model (*S-model*) and the multi-block model (*M-model*), to describe the observations available to the slave node in our work. Under the *K-model*, we assume that the fixed delays in both the forward and reverse directions are known to the slave node, while under the *S-model*, we assume that the fixed path delays are unknown, but there is a prior relationship between the fixed path delays. Further, under the *M-model*, we assume a prior known relationship between the fixed path delays, as well additional timestamps that contain the same clock skew, but different clock offsets. For all the considered observation models, the problem of estimating the clock skew and offset in the presence of PDV falls under a variant of the location-scale parameter estimation problem [15], with the unknown clock skew as the scale parameter and the unknown clock offset as the location parameter. Fixing the loss function as the skew-normalized squared error loss and assuming complete knowledge of the statistical information describing the PDV along with unlimited computational complexity, we use invariant decision theory (see chapter 6 of [15]) to design the optimum invariant CSOE scheme for the considered observation models. Then, using results from [15]–[17], we show that the developed optimum invariant CSOE schemes are minimax optimum for the skew-normalized squared error loss, i.e., these estimators minimize the maximum skew normalized mean square estimation error over all possible values of the unknown parameters.

In this paper, we focus our numerical results on the LTE backhaul network scenario. In this scenario, PTP is used to synchronize the cellular base stations using the mobile backhaul networks. The optimum approaches are general, so they can be applied to other applications, for example smart grids. In the cellular base station application, the backhaul networks are leased from commercial Internet Service Providers (ISPs), and the network is shared with other commercial and non-commercial users. The background traffic generated by these users often results in PDV for the synchronization packets. Based on an extensive study [10] employing a detailed simulation package we built based on the recommendations by standard committees focused on IEEE 1588, the popular models for the probability density functions (pdfs) of the random variables describing the PDV that were considered in the literature (Gaussian, exponential, Weibull, and log-normal [9]) do not always provide a close match to the queuing delay pdfs [10]. In this paper, we use the pdfs obtained from the simulation package to evaluate the performance of the considered clock skew and offset estimators. Our key new contributions in this paper are as follows:

1) *Optimum invariant clock skew and offset estimators:* Given the joint pdf of the random variables describing the PDV and without complexity limitations, we develop the optimum invariant clock skew and offset estimators for PTP under the considered observation models.
2) *Minimax optimum clock skew and offset estimators:* The developed optimum invariant estimators are shown to be minimax optimum.

The developed optimum estimators are very useful to understand the possible performance when we have the complete statistical information on the queuing delays and unlimited computational complexity. As the previously proposed approaches to solve the IEEE 1588 timing synchronization problem are all invariant, the optimum estimators can provide useful performance benchmarks for evaluating the performance of these CSOE schemes. The performance comparison between the realistic schemes and the optimum estimators are performed off-line, where complexity is not a stringent issue.

To demonstrate the utility of the results given in this paper, we use the developed optimum estimators in a robust approach [18] that does not require complete information on the pdf of the queuing delays but requires unlimited computational complexity. Simulation results indicate that there is no significant loss in performance for the robust estimator when compared to the optimum scheme. These results illustrate how the developed optimum estimator can help us understand the performance loss due to incomplete knowledge of the queuing delay pdfs. The results can also help in evaluating limited complexity approaches. If a designer finds an approach with an acceptable computational complexity that exhibits performance close to the optimum estimators, they can use it and know they have near optimum performance. This is precisely the approach used in communications when comparing to capacity.

Notations: We use bold upper case, bold lower case, and italic lettering to denote matrices, column vectors and scalars respectively. The notations $(.)^T$ and $\otimes$ denote the transpose and Kronecker product, respectively. $\boldsymbol{I}_N$ stands for a N-dimensional identity matrix and $\mathbf{1}_N$ denotes a column vector of length N with all the elements equal to 1. Further, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}^+$ denotes the set of positive real numbers, $\mathbb{R}_0^+$ denotes the set of non-negative real numbers and $\mathcal{I}_A(x)$ denotes the indicator function having the value 1 when $x \in A$ and 0 when $x \notin A$.

II. Signal Model and Problem Statement

Consider a scenario where the slave clock has a clock offset δ and a clock skew ϕ with respect to its master clock. To help the slave determine δ and ϕ, PTP allows a two-way message exchange between the master and slave node, which we now describe. The master node initiates a two-way message exchange by sending a *sync* packet to the slave at time t_1. The value of t_1 is later communicated to the slave via a *follow_up* message. The slave node records the time of reception of the *sync* message as t_2. The slave node sends a *delay_req* message to the master node while recording the time of transmission as t_3. The master records the time of arrival of the *delay_req* packet at time t_4 and this value is later communicated to the slave using a *delay_resp* packet. The relationship between the timestamps is given by

$$t_2 = (t_1 + d_{ms} + w_1)\phi + \delta, \quad (1)$$

$$t_3 = (t_4 - d_{sm} - w_2)\phi + \delta, \quad (2)$$

where d_{ms} and d_{sm} denote the fixed propagation delays, while w_1 and w_2 model the random queuing delays. Assuming the values of δ, ϕ, d_{ms} and d_{sm} remain constant over the duration of P two-way message exchanges, we can collect the timestamps from multiple two-way message exchanges to estimate δ and ϕ [6]–[9]. We denote these timestamps as

$$t_{2i} = (t_{1i} + d_{ms} + w_{1i})\phi + \delta, \quad (3)$$
$$t_{3i} = (t_{4i} - d_{sm} - w_{2i})\phi + \delta \quad (4)$$

for $i = 1, 2, \cdots, P$. Define $\boldsymbol{w}_k = [w_{k1}, w_{k2}, \cdots, w_{kP}]$ for $k = 1, 2$ and $\boldsymbol{t}_k = [t_{k1}, t_{k2}, \cdots, t_{kP}]$ for $k = 1, 2, 3, 4$. The joint pdf of $\boldsymbol{w}_k$ is defined as $f_{\boldsymbol{w}_k}(\boldsymbol{w}_k) = f_k(w_{k1}, w_{k2}, \cdots, w_{kP})$ for $k = 1, 2$. We next consider three observation models based on the amount of information available regarding the fixed path delays.

1) Known fixed delay model (K-Model): In this model, we assume complete knowledge of the fixed-path delays d_{ms} and d_{sm}. The received timestamps shown in (3) and (4) can be arranged in vector form as follows

$$\boldsymbol{y} = \boldsymbol{u}\phi + \delta \mathbf{1}_{2P}, \quad (5)$$

where we have $\boldsymbol{y} = [\boldsymbol{t}_2, \boldsymbol{t}_3]^T$, and $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2]^T$, $\boldsymbol{u}_1 = (\boldsymbol{t}_1 + \boldsymbol{w}_1 + d_{ms}\mathbf{1}_P^T)$ and $\boldsymbol{u}_2 = (\boldsymbol{t}_4 - \boldsymbol{w}_2 - d_{sm}\mathbf{1}_P^T)$. Further, we have $f_{\boldsymbol{u}}(\boldsymbol{u}) = f_{\boldsymbol{u}_1}(\boldsymbol{u}_1) f_{\boldsymbol{u}_2}(\boldsymbol{u}_2)$ with $f_{\boldsymbol{u}_1}(\boldsymbol{u}_1) = f_{\boldsymbol{w}_1}(\boldsymbol{u}_1 - \boldsymbol{t}_1 - d_{ms}\mathbf{1}_P^T)$ and $f_{\boldsymbol{u}_2}(\boldsymbol{u}_2) = f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{u}_2 - d_{sm}\mathbf{1}_P^T)$. The unknown parameters in this model are ϕ and δ.

2) Standard model (S-Model): Freris *et al.* [19] provided some necessary conditions for obtaining a unique solution for the system of equations given in (3) and (4), when the complete information regarding the fixed delays is not available. We need to know either one of the fixed path delays (either d_{ms} or d_{sm}), or have a prior known affine relationship between the fixed delays (see Theorem 4 in [19]). Hence, in this model, we assume a prior known affine relationship between the fixed path delays. Let $d_{ms} = d$ and $d_{sm} = a_0 d_{ms} + c_0$, where the parameter d is unknown, but the constants a_0 and c_0 are known. The received time stamps shown in (3) and (4) can be arranged in vector form as

$$\boldsymbol{y} = (\boldsymbol{h}d + \boldsymbol{v})\phi + \delta \mathbf{1}_{2P}, \quad (6)$$

where $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2]^T$, $\boldsymbol{v}_1 = (\boldsymbol{t}_1 + \boldsymbol{w}_1)$, $\boldsymbol{v}_2 = (\boldsymbol{t}_4 - c_0\mathbf{1}_P^T - \boldsymbol{w}_2)$, $\boldsymbol{h} = [\mathbf{1}_P^T, -a_0\mathbf{1}_P^T]^T$, and $\boldsymbol{y} = [\boldsymbol{t}_2, \boldsymbol{t}_3]^T$. Further, we have $f_{\boldsymbol{v}}(\boldsymbol{v}) = f_{\boldsymbol{v}_1}(\boldsymbol{v}_1) f_{\boldsymbol{v}_2}(\boldsymbol{v}_2)$ with $f_{\boldsymbol{v}_1}(\boldsymbol{v}_1) = f_{\boldsymbol{w}_1}(\boldsymbol{v}_1 - \boldsymbol{t}_1)$ and $f_{\boldsymbol{v}_2}(\boldsymbol{v}_2) = f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{v}_2 - c_0\mathbf{1}_P^T)$. The unknown parameters in this model are ϕ, d and δ.

3) Multiblock model (M-Model): Here we assume, as in the *S-model*, that there is a prior known affine relationship between the fixed path delays, i.e., $d_{ms} = d$ and $d_{sm} = a_0 d_{ms} + c_0$, where the parameter d is unknown, but the constants a_0 and c_0 are known. Suppose we refer to the set of timestamps obtained from P two-way message exchanges as a block. In this model, we further assume that in addition to the current block, we have additional timestamps from B previous blocks. The clock offset δ is modeled as being constant within each block, but varying between different blocks. However, the parameters d and ϕ are modeled as constant across all $B + 1$ blocks. This model is representative of scenarios where changes in the clock skew ϕ, occur over longer time scales than changes in the clock offset δ. We denote the timestamps in the past blocks using the notation

$$t_{2ij} = (t_{1ij} + d + w_{1ij})\phi + \delta_j, \quad (7)$$
$$t_{3ij} = (t_{4ij} - a_0 d - c_0 - w_{2ij})\phi + \delta_j \quad (8)$$

for $i = 1, 2, \cdots, P$ and $j = 1, 2, \cdots, B$, and the timestamps in the current block as

$$t_{2i} = (t_{1i} + d + w_{1i})\phi + \delta, \quad (9)$$
$$t_{3i} = (t_{4i} - a_0 d - c_0 - w_{2i})\phi + \delta \quad (10)$$

for $i = 1, 2, \cdots, P$. In (10), δ denotes the clock offset of the current block which we want to estimate along with the clock skew ϕ. In (8), δ_j denotes the clock offset corresponding to the j^{th} previous block i.e., δ_B corresponds to the clock offset of the 'oldest' block. For notational convenience, we define $\boldsymbol{t}_{kj} = [t_{k1j}, t_{k2j}, \cdots, t_{kPj}]$ for $k = 1, 2, 3, 4$ and $j = 1, 2, \cdots, B$ and $\boldsymbol{w}_{kj} = [w_{k1j}, w_{k2j}, \cdots, w_{kPj}]$ for $k = 1, 2$ and $j = 1, 2, \cdots, B$. The complete set of timestamps from the $(B + 1)$ blocks can be arranged in vector form as

$$\boldsymbol{y} = (\boldsymbol{h}_M d + \boldsymbol{z})\phi + (\boldsymbol{\delta} \otimes \mathbf{1}_{2P}), \quad (11)$$

where $\boldsymbol{\delta} = [\delta, \delta_1, \delta_2, \cdots, \delta_B]$; $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2]^T$ with $\boldsymbol{y}_1 = [\boldsymbol{t}_2, \boldsymbol{t}_{21}, \cdots, \boldsymbol{t}_{2B}]$ and $\boldsymbol{y}_2 = [\boldsymbol{t}_3, \boldsymbol{t}_{31}, \cdots, \boldsymbol{t}_{3B}]$; $\boldsymbol{h}_M = [\mathbf{1}_{P(B+1)}^T, -a_0\mathbf{1}_{P(B+1)}^T]^T$; and $\boldsymbol{z} = [\boldsymbol{z}_1, \boldsymbol{z}_{11}, \cdots, \boldsymbol{z}_{1B}, \boldsymbol{z}_2, \boldsymbol{z}_{21}, \cdots, \boldsymbol{z}_{2B}]^T$ with $\boldsymbol{z}_1 = (\boldsymbol{t}_1 + \boldsymbol{w}_1)$, $\boldsymbol{z}_{1j} = (\boldsymbol{t}_{1j} + \boldsymbol{w}_{1j})$ for $j = 1, 2, \cdots, B$, $\boldsymbol{z}_2 = (\boldsymbol{t}_4 - \boldsymbol{w}_2 - c_0\mathbf{1}_P^T)$ and $\boldsymbol{z}_{2j} = (\boldsymbol{t}_{4j} - \boldsymbol{w}_{2j} - c_0\mathbf{1}_P^T)$ for $j = 1, 2, \cdots, B$. Further assuming the timestamps across different blocks are independent and the time stamps have identical pdfs over all blocks for both the forward and reverse path, we have $f_{\boldsymbol{z}}(\boldsymbol{z}) = f_{\boldsymbol{z}_1}(\boldsymbol{z}_1) f_{\boldsymbol{z}_2}(\boldsymbol{z}_2)$ with $f_{\boldsymbol{z}_1}(\boldsymbol{z}_1) = f_{\boldsymbol{w}_1}(\boldsymbol{z}_1 - \boldsymbol{t}_1) \prod_{j=1}^{B} f_{\boldsymbol{w}_1}(\boldsymbol{z}_{1j} - \boldsymbol{t}_{1j})$ and $f_{\boldsymbol{z}_2}(\boldsymbol{z}_2) = f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{z}_2 - c_0\mathbf{1}_P^T) \prod_{j=1}^{B} f_{\boldsymbol{w}_2}(\boldsymbol{t}_{4j} - \boldsymbol{z}_{2j} - c_0\mathbf{1}_P^T)$. The unknown parameters in this model are ϕ, d and $\boldsymbol{\delta}$.

Given any of the observation models, the CSOE problem is to estimate ϕ and δ from the received time stamps. We now state all the assumptions made in our work.

Assumption 1: *All the queuing delays are strictly positive random variables and have finite support.*

Assumption 2: *The queuing delays in the forward and reverse path are independent, the pdfs of the random variables describing the queuing delays are assumed to be completely known and we assume unlimited computational complexity.*

Assumption 3: *For the K-model and S-model, the parameters d, ϕ, and δ are assumed to be constant over P two-way message exchanges.*

Assumption 4: *In the case of the M-model, we assume the queuing delays across different blocks are independent from blocks to block and have identical pdfs for each of the $(B + 1)$ blocks in both the forward and reverse path. Also, the parameters ϕ and d are assumed to be constant over all $(B+1)$ blocks, while the clock offset is assumed to be constant for a block, but is varying from block to block. The value of B can be chosen according to the time interval across which the clock skew ϕ, can be assumed to be constant.*

III. Statistical Preliminaries

In this section, we present some important definitions for characterizing the performance of estimators along with some useful results regarding invariant estimators. It is assumed throughout this section that the observed data $\boldsymbol{x} \in \mathbb{R}^N$ is characterized by the pdf $f(\boldsymbol{x}|\boldsymbol{\theta})$, which depends upon the vector of unknown parameters $\boldsymbol{\theta}$ with the corresponding parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^M$. Suppose we are interested in estimating a scalar $\boldsymbol{c}^T\boldsymbol{\theta}$, where $\boldsymbol{c} \in \mathbb{R}^M$ is a constant vector. Let ψ denote an estimator of $\boldsymbol{c}^T\boldsymbol{\theta}$, $\psi(\boldsymbol{x})$ denote the estimate of $\boldsymbol{c}^T\boldsymbol{\theta}$ obtained using the estimator ψ on $\boldsymbol{x}$, and $L(\psi(\boldsymbol{x}),\boldsymbol{\theta})$ denote the considered loss function. The performance of the estimator ψ can be characterized by [16]:

1) The conditional risk of an estimator
$$\mathcal{R}(\psi,\boldsymbol{\theta}) \quad = \quad \int_{\mathbb{R}^N} L(\psi(\boldsymbol{x}),\boldsymbol{\theta})f(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}, \tag{12}$$
2) The maximum risk of an estimator
$$\mathcal{M}(\psi) \quad = \quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{R}(\psi,\boldsymbol{\theta}), \tag{13}$$
3) The average risk of an estimator
$$\mathcal{B}(\psi,p) \quad = \quad \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{R}(\psi,\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\Theta}, \tag{14}$$
where $p(\boldsymbol{\theta})$ is a prior distribution defined over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

See Chapter 6 of [15] for definitions of a group, invariant loss function and an invariant estimator. We now present an important theorem regarding the conditional risk of invariant estimators.

Theorem 1 (Section 6.2.3, [15])**.** *For an invariant loss function, the conditional risk of an invariant estimator ψ of $\boldsymbol{c}^T\boldsymbol{\theta}$, is constant for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.*

Remark. If ψ is an invariant estimator of $\boldsymbol{c}^T\boldsymbol{\theta}$, we have
$$\mathcal{R}(\psi,\boldsymbol{\theta}) = \mathcal{M}(\psi) = \mathcal{B}(\psi,p), \tag{15}$$
for any $p(\boldsymbol{\theta})$ defined over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

For an invariant loss function, we can construct the optimum (or minimum conditional risk) invariant estimator using the theory from [15]. An attractive property of optimum invariant estimators is that they frequently turn out to be minimax optimum [15]. We now present the definition of a minimax estimator from [16].

Definition 1 (Minimax estimators)**.** *An estimator $\boldsymbol{\psi}_{MinMax}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$ is said to be a minimax estimator of $\boldsymbol{c}^T\boldsymbol{\theta}$ for the considered loss function, if*
$$\mathcal{M}(\psi_{MinMax}) \quad = \quad \inf_{\psi} \mathcal{M}(\psi) = \inf_{\psi}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{R}(\psi,\boldsymbol{\theta}). \tag{16}$$

In this paper, assuming complete knowledge of the joint queuing delay pdfs and unlimited computational complexity, we use the concepts of invariant estimation theory to design the optimum invariant CSOE schemes under the considered observation models. As we are primarily interested in estimating δ and ϕ, we consider the skew normalized squared error loss functions defined by
$$L_1(a_\delta,\boldsymbol{\theta}) \quad = \quad \frac{(a_\delta-\delta)^2}{\phi^2}, \tag{17}$$
and
$$L_2(a_\phi,\boldsymbol{\theta}) \quad = \quad \frac{(a_\phi-\phi)^2}{\phi^2} \tag{18}$$
for δ and ϕ, respectively[1]. In (17) and (18), a_δ and a_ϕ denote estimates of δ and ϕ, respectively, $\boldsymbol{\theta} = [\phi,\delta]$ in the case of the *K-model*, $\boldsymbol{\theta} = [\phi,d,\delta]$ for the *S-model* and $\boldsymbol{\theta} = [\phi,d,\delta,\delta_1,\cdots,\delta_B]$ for the *M-model*. We then use results from [15]–[17] to show the derived optimum invariant estimators of δ and ϕ are minimax optimum.

IV. Optimum Invariant CSOE scheme under K-model

In this section, we apply invariant decision theory to derive an optimum invariant estimator of ϕ and δ for the *K-model* assuming complete knowledge of the joint queuing delay pdfs and unlimited computational complexity. Recall from (5), the observations under the *K-model* can be represented as
$$\boldsymbol{y} \quad = \quad \boldsymbol{u}\phi + \delta\mathbf{1}_{2P}, \tag{19}$$
where $\boldsymbol{y} \in \mathbb{R}^{2P}$, $\boldsymbol{u} \in \mathbb{R}^{2P}$, $\phi \in \mathbb{R}^+$ and $\delta \in \mathbb{R}$. Let $\boldsymbol{\theta} = [\phi,\delta]$ denote the vector of unknown parameters. The parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}$, is given by
$$\boldsymbol{\Theta} \quad = \quad \{(\phi,\delta) : \phi \in \mathbb{R}^+, \delta \in \mathbb{R}\}. \tag{20}$$
From (5), we have $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2-\delta\mathbf{1}_P^T}{\phi} - d_{ms}\mathbf{1}_P^T - \boldsymbol{t}_1\right) f_{\boldsymbol{w}_2}\left(\frac{\delta\mathbf{1}_P^T-\boldsymbol{t}_3}{\phi} - d_{sm}\mathbf{1}_P^T + \boldsymbol{t}_4\right)$, where the factor $\frac{1}{\phi^{2P}}$ comes from the Jacobian of the transformation of the random variable.

Let $\mathcal{F}_{KModel}$ denote the class of all pdfs $f(\boldsymbol{y}|\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The class of such pdfs is invariant under the group of location-scale transformations (see Example 5, Section 6.2.1, [15]) $\mathcal{G}_{KModel}$, on $\mathbb{R}^{2P}$, defined as
$$\begin{aligned}\mathcal{G}_{KModel} \quad = \quad & \{g_{a,b}(\boldsymbol{m}) : g_{a,b}(\boldsymbol{m}) = a\boldsymbol{m} + b\mathbf{1}_{2P}, \\ & \forall(a,b) \in \mathbb{R}^+ \times \mathbb{R}\},\end{aligned} \tag{21}$$
where $\boldsymbol{m} \in \mathbb{R}^{2P}$, since $\boldsymbol{y}_g = g_{a,b}(\boldsymbol{y})$ has the pdf $\frac{1}{(a\phi)^{2P}} f_{\boldsymbol{u}}\left(\frac{\boldsymbol{y}_g-(a\delta+b)\mathbf{1}_{2P}}{a\phi}\right)$ which has the scale and shift parameters $(a\phi, a\delta+b)$ as opposed to the parameters (ϕ,δ) for $f(\boldsymbol{y}|\boldsymbol{\theta})$. This shows that the group, $\bar{\mathcal{G}}_{KModel}$, of induced transformations is given by
$$\begin{aligned}\bar{\mathcal{G}}_{KModel} \quad = \quad & \{\bar{g}_{a,b}((\phi,\delta)) : \bar{g}_{a,b}((\phi,\delta)) = (a\phi,(a\delta+b)), \\ & \forall(a,b) \in \mathbb{R}^+ \times \mathbb{R}\},\end{aligned} \tag{22}$$
where $\phi \in \mathbb{R}^+$ and $\delta \in \mathbb{R}$.

Let $\hat{\delta}_I$ and $\hat{\phi}_I$ denote estimators of δ and ϕ, respectively and let $\hat{\delta}_I(\boldsymbol{y})$ and $\hat{\phi}_I(\boldsymbol{y})$ denote the estimates obtained from the received data $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{u}}\left(\frac{\boldsymbol{y}-\delta\mathbf{1}_{2P}}{\phi}\right)$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_{KModel}$ from (21) if for all $(a,b) \in \mathbb{R}^+ \times \mathbb{R}$
$$\hat{\delta}_I(g_{a,b}(\boldsymbol{y})) = \hat{\delta}_I(a\boldsymbol{y} + b\mathbf{1}_{2P}) \quad = \quad a\hat{\delta}_I(\boldsymbol{y}) + b, \tag{23}$$
$$\hat{\phi}_I(g_{a,b}(\boldsymbol{y})) = \hat{\phi}_I(a\boldsymbol{y} + b\mathbf{1}_{2P}) \quad = \quad a\hat{\phi}_I(\boldsymbol{y}). \tag{24}$$

[1]As seen in equations (5), (6) and (11), the unknown clock skew ϕ is multiplied with the random queuing delays. Hence, we fix our loss function as the skew normalized squared error loss.

Thus the scaling and shifting factors a and b scale and shift the estimators as one might expect. Further, the loss functions defined in (17) and (18) for δ and ϕ, respectively, are invariant under $\mathcal{G}_{KModel}$ from (21), since

$$\frac{(\hat{\delta}_I(\boldsymbol{y})-\delta)^2}{\phi^2}=\frac{\left(\hat{\delta}_I(g_{a,b}(\boldsymbol{y}))-(a\delta+b)\right)^2}{a^2\phi^2}, \tag{25}$$

and

$$\frac{(\hat{\phi}_I(\boldsymbol{y})-\phi)^2}{\phi^2}=\frac{\left(\hat{\phi}_I(g_{a,b}(\boldsymbol{y}))-a\phi\right)^2}{a^2\phi^2} \tag{26}$$

for all $g_{a,b}\in\mathcal{G}_{KModel}$. We now present the minimax optimum estimators of δ and ϕ under the *K-model*.

Proposition 1. *The optimum (or minimum conditional risk) invariant estimators of δ and ϕ, denoted by $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, respectively, under $\mathcal{G}_{KModel}$ defined in (21), for the skew-normalized squared error loss functions defined in (17) and (18), respectively, are given by*

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}}\frac{\delta}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}}\frac{1}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}, \tag{27}$$

and

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}}\frac{1}{\phi^2}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}}\frac{1}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}, \tag{28}$$

respectively, where $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2-\delta\boldsymbol{1}_P^T}{\phi}-d_{ms}\boldsymbol{1}_P^T-\boldsymbol{t}_1\right)f_{\boldsymbol{w}_2}\left(\frac{\delta\boldsymbol{1}_P^T-\boldsymbol{t}_3}{\phi}-d_{sm}\boldsymbol{1}_P^T+\boldsymbol{t}_4\right)$. *Further, the derived optimum invariant estimators are minimax for the skew-normalized squared error loss (see Appendix A for proof).*

We now present an important result with regards to the mean square estimation error performance of the minimax optimum estimators when compared to ML estimators. Let $\hat{\delta}$ and $\hat{\phi}$ denote estimators of δ and ϕ, respectively. The Mean Square estimation Errors (MSEs) of $\hat{\delta}$ and $\hat{\phi}$, denoted by MSE($\hat{\delta}$) and MSE($\hat{\phi}$), respectively, are defined as

$$\text{MSE}(\hat{\delta}) = E\left\{(\hat{\delta}-\delta)^2|\boldsymbol{\theta}\right\}, \tag{29}$$

and

$$\text{MSE}(\hat{\phi}) = E\left\{(\hat{\phi}-\phi)^2|\boldsymbol{\theta}\right\}, \tag{30}$$

where $E\{.\}$ denotes the expectation operator and $\boldsymbol{\theta}$ is the vector of unknown parameters.

Proposition 2. *Let $\hat{\delta}_{MLE}$ and $\hat{\phi}_{MLE}$ denote the ML estimators of δ and ϕ, respectively. Under the K-model, the MSE of $\hat{\delta}_{MLE}$ is always greater than or equal to the MSE of $\hat{\delta}_{MinRisk}$. Also, under the K-model, the MSE of $\hat{\phi}_{MLE}$ is always greater than or equal to the MSE of $\hat{\phi}_{MinRisk}$.*

Proof. In the *K-model*, we have $\boldsymbol{\theta}=[\phi,\delta]$. Let $\hat{\phi}_{MLE}(\boldsymbol{y})$ and $\hat{\delta}_{MLE}(\boldsymbol{y})$ denote the ML estimates obtained from $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y}|\boldsymbol{\theta})=\frac{1}{\phi^{2P}}f_{\boldsymbol{u}}(\frac{\boldsymbol{y}-\delta\boldsymbol{1}_{2P}}{\phi})$. We have

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{MLE}(\boldsymbol{y}) &= [\hat{\phi}_{MLE}(\boldsymbol{y}),\hat{\delta}_{MLE}(\boldsymbol{y})]\\ &= \underbrace{\arg\max}_{\boldsymbol{\theta}}\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}),\end{aligned} \tag{31}$$

where $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y})$ is the likelihood function and is equal to $f(\boldsymbol{y}|\boldsymbol{\theta})$. Let $g_{a,b}\in\mathcal{G}_{KModel}$ from (21) and define $\boldsymbol{y}_g=g_{a,b}(\boldsymbol{y})$. From (22), the corresponding transformation of the parameter vector $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta}_g=\bar{g}_{a,b}(\boldsymbol{\theta})=(a\phi,(a\delta+b))$. From the functional invariance of ML estimators [20] (see Chapter 7, Theorem 7.2.10), we have $\hat{\boldsymbol{\theta}}_{MLE}(\boldsymbol{y}_g)=\bar{g}_{a,b}(\hat{\boldsymbol{\theta}}_{MLE}(\boldsymbol{y}))$. So, we have the following relationship

$$\hat{\delta}_{MLE}(\boldsymbol{y}_g) = a\hat{\delta}_{MLE}(\boldsymbol{y})+b, \tag{32}$$

and

$$\hat{\phi}_{MLE}(\boldsymbol{y}_g) = a\hat{\phi}_{MLE}(\boldsymbol{y}). \tag{33}$$

As this holds true for all $g_{a,b}\in\mathcal{G}_{KModel}$ from (21), the ML estimators of δ and ϕ are invariant under $\mathcal{G}_{KModel}$ as they satisfy (23) and (24). Hence, for the skew-normalized loss function defined in (17), we have

$$\mathcal{R}(\hat{\delta}_{MinRisk},\boldsymbol{\theta}) \leq \mathcal{R}(\hat{\delta}_{MLE},\boldsymbol{\theta}), \tag{34}$$

since $\hat{\delta}_{MinRisk}$ is the optimum invariant estimator under $\mathcal{G}_{KModel}$ in (21) and achieves the minimum conditional risk among all estimators that are invariant under $\mathcal{G}_{KModel}$ (see Proposition 1). From (34), we have

$$\begin{aligned}&\int_{\mathbb{R}^{2P}}\frac{(\hat{\delta}_{MinRisk}(\boldsymbol{y})-\delta)^2}{\phi^2}f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{y}\leq\\ &\qquad\int_{\mathbb{R}^{2P}}\frac{(\hat{\delta}_{MLE}(\boldsymbol{y})-\delta)^2}{\phi^2}f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{y},\end{aligned} \tag{35}$$

$$\begin{aligned}\implies&\int_{\mathbb{R}^{2P}}(\hat{\delta}_{MinRisk}(\boldsymbol{y})-\delta)^2f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{y}\leq\\ &\qquad\int_{\mathbb{R}^{2P}}(\hat{\delta}_{MLE}(\boldsymbol{y})-\delta)^2f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{y},\end{aligned} \tag{36}$$

implies

$$\text{MSE}(\hat{\delta}_{MinRisk}) \leq \text{MSE}(\hat{\delta}_{MLE}). \tag{37}$$

Following similar steps, we can show that MSE($\hat{\phi}_{MinRisk}$) $\leq$ MSE($\hat{\phi}_{MLE}$). ☐

V. Optimum Invariant CSOE scheme under S-model

In this section, we apply invariant decision theory to derive an optimum invariant estimator of ϕ and δ under the *S-model* assuming complete knowledge of the joint queuing delay pdfs and unlimited computational complexity. Recall from (6), the observations under the *S-model* can be represented as

$$\boldsymbol{y} = (\boldsymbol{h}d+\boldsymbol{v})\phi+\delta\boldsymbol{1}_{2P}, \tag{38}$$

where $\boldsymbol{y}\in\mathbb{R}^{2P}$, $\boldsymbol{v}\in\mathbb{R}^{2P}$, $\phi\in\mathbb{R}^+$ and $\delta\in\mathbb{R}$. As the unknown fixed delay d is always non-negative, we have $d\in\mathbb{R}_0^+$. However, it is not possible to design invariant estimators under this constraint, as we cannot construct a group of transformations for which the class of pdfs in the *S-model* is invariant under the constructed group of transformations. Hence, we assume $d\in\mathbb{R}$, but later we see this is not a problem as we derive the minimax optimum estimator in Proposition 3. Let $\boldsymbol{\theta}=[\phi,d,\delta]$ denote the vector of unknown parameters. The unrestricted parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}$, is given by

$$\boldsymbol{\Theta} = \{(\phi,d,\delta):\phi\in\mathbb{R}^+,d\in\mathbb{R},\delta\in\mathbb{R}\}, \tag{39}$$

and the restricted parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}^*$, is given by

$$\boldsymbol{\Theta}^* = \{(\phi, d, \delta) : \phi \in \mathbb{R}^+, d \in \mathbb{R}_0^+, \delta \in \mathbb{R}\}. \quad (40)$$

From (6), we have $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \delta \mathbf{1}_P^T}{\phi} - d\mathbf{1}_P^T - \boldsymbol{t}_1\right) f_{\boldsymbol{w}_2}\left(\frac{\delta \mathbf{1}_P^T - \boldsymbol{t}_3}{\phi} + \boldsymbol{t}_4 - (a_0 d + c_0)\mathbf{1}_P^T\right)$.

Let $\mathcal{F}_{SModel}$ denote the class of all pdfs $f(\boldsymbol{y}|\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The class of such pdfs is invariant under the group of transformations $\mathcal{G}_{SModel}$, on $\mathbb{R}^{2P}$, defined as

$$\begin{aligned} \mathcal{G}_{SModel} = \{& g_{a,b,c}(\boldsymbol{m}) : g_{a,b,c}(\boldsymbol{m}) = a(\boldsymbol{m} + \boldsymbol{h}b) + c\mathbf{1}_{2P}, \\ & \forall (a, b, c) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}\}, \end{aligned} \quad (41)$$

where $\boldsymbol{m} \in \mathbb{R}^{2P}$, since $\boldsymbol{y}_g = g_{a,b,c}(\boldsymbol{y})$ has the pdf $\frac{1}{(a\phi)^{2P}} f_{\boldsymbol{v}}\left(\frac{\boldsymbol{y}_g - (a\delta + c)\mathbf{1}_{2P}}{a\phi} - \boldsymbol{h}\left(d + \frac{b}{\phi}\right)\right)$ which has the parameters $(a\phi, (d + b/\phi), a\delta + c)$ as opposed to the parameters (ϕ, d, δ) for $f(\boldsymbol{y}|\boldsymbol{\theta})$. This shows that the group, $\bar{\mathcal{G}}_{SModel}$, of induced transformations on $\boldsymbol{\Theta}$ is given by

$$\begin{aligned} \bar{\mathcal{G}}_{SModel} = \{& \bar{g}_{a,b,c}((\phi, d, \delta)) : \bar{g}_{a,b,c}((\phi, d, \delta)) \\ & = (a\phi, (d + b/\phi), (a\delta + c)) \\ & \forall (a, b, c) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}\}, \end{aligned} \quad (42)$$

where $\phi \in \mathbb{R}^+, d \in \mathbb{R}$ and $\delta \in \mathbb{R}$. Thus the transformations modify the three parameters but the pdf can still be represented in the same general class of pdfs which have some values for these parameters.

Let $\hat{\delta}_I$ and $\hat{\phi}_I$ denote estimators of δ and ϕ, respectively and let $\hat{\delta}_I(\boldsymbol{y})$ and $\hat{\phi}_I(\boldsymbol{y})$ denote the estimates obtained from the received data $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{v}}\left(\frac{\boldsymbol{y} - \delta \mathbf{1}_{2P}}{\phi} - \boldsymbol{h}d\right)$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_{SModel}$ from (41), if for all $(a, b, c) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}$,

$$\hat{\delta}_I(g_{a,b,c}(\boldsymbol{y})) = \hat{\delta}_I(a(\boldsymbol{y} + \boldsymbol{h}b) + c\mathbf{1}_{2P}) = (a\hat{\delta}_I(\boldsymbol{y}) + c), \quad (43)$$

$$\hat{\phi}_I(g_{a,b,c}(\boldsymbol{y})) = \hat{\phi}_I(a(\boldsymbol{y} + \boldsymbol{h}b) + c\mathbf{1}_{2P}) = a\hat{\phi}_I(\boldsymbol{y}). \quad (44)$$

Note that, by design, the estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant to the parameter d (since the changes in d in (43) and (44) do not affect $\hat{\delta}_I$ and $\hat{\phi}_I$), i.e., the estimates, as well as the performance of the estimators, are not affected by the value of d. Further, the skew-normalized loss functions defined in (17) and (18) for δ and ϕ, respectively, are invariant under $\mathcal{G}_{SModel}$ from (41), since

$$\frac{(\hat{\delta}_I(\boldsymbol{y}) - \delta)^2}{\phi^2} = \frac{\left(\hat{\delta}_I(g_{a,b,c}(\boldsymbol{y})) - (a\delta + c)\right)^2}{a^2\phi^2}, \quad (45)$$

and

$$\frac{(\hat{\phi}_I(\boldsymbol{y}) - \phi)^2}{\phi^2} = \frac{\left(\hat{\phi}_I(g_{a,b,c}(\boldsymbol{y})) - a\phi\right)^2}{a^2\phi^2} \quad (46)$$

for all $g_{a,b,c} \in \mathcal{G}_{SModel}$. We now present the minimax optimum estimators of δ and ϕ under the *S-model*.

Proposition 3. *The optimum (or minimum conditional risk) invariant estimators of δ and ϕ, denoted by $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, respectively, under $\mathcal{G}_{SModel}$ defined in (41), for the skew normalized squared error loss functions defined in (17) and (18) respectively, are given by*

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \frac{\delta}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}{\int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}, \quad (47)$$

and

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \frac{1}{\phi} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}{\int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}, \quad (48)$$

respectively, where $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \delta \mathbf{1}_P^T}{\phi} - d\mathbf{1}_P^T - \boldsymbol{t}_1\right) f_{\boldsymbol{w}_2}\left(\frac{\delta \mathbf{1}_P^T - \boldsymbol{t}_3}{\phi} + \boldsymbol{t}_4 - (a_0 d + c_0)\mathbf{1}_P^T\right)$. Further, the derived optimum invariant estimators are minimax for the skew-normalized squared error loss in the restricted parameter space $\boldsymbol{\Theta}^*$ (see Appendix B for proof). Also, the optimum invariant estimators are optimum in terms of acheiving the lowest MSE among all estimators invariant under $\mathcal{G}_{SModel}$ defined in (41).

A desirable property of any estimator of δ and ϕ is for it to be asymptotically consistent. We now present an important result regarding the invariance of asymptotically consistent estimators.

Proposition 4. *Any consistent estimator of ϕ and δ obtained from solving (38) is asymptotically invariant under $\mathcal{G}_{SModel}$ defined in (41).*

Proof. For any fixed value of d, a scale or shift transformation on the observations would lead to corresponding estimates of δ and ϕ obtained from a consistent clock skew and offset estimator to be scaled or shifted for asymptotically large sample sizes, since any consistent estimator always converges towards the real value of the parameter. Hence, any consistent estimator of ϕ and δ obtained from solving (38) is asymptotically invariant under $\mathcal{G}_{SModel}$ defined in (41) as they satisfy (43) and (44). □

Remark. For a fixed value of d, the ML-estimator of ϕ and δ under the *S-model* is invariant under $\mathcal{G}_{SModel}$ defined in (41) as any shift or scale transformation of the observations results in the corresponding transformation of the ML estimate of ϕ and δ (due to the functional invariance property of ML estimators). Following steps similar to those given in Proposition 2, we can show that the ML-estimators under the *S-model* have a MSE greater than or equal to the optimum invariant estimators under the *S-model*.

A. Imprecise Knowledge of Queuing Delay pdfs

We now consider a case where the queuing delay pdfs are not known perfectly. To this end, we assume that there K possible pdfs, defined by the set $\left\{\left(f_{\boldsymbol{w}_1}^{(1)}(.), f_{\boldsymbol{w}_2}^{(1)}(.)\right), \left(f_{\boldsymbol{w}_1}^{(2)}(.), f_{\boldsymbol{w}_2}^{(2)}(.)\right), \ldots \left(f_{\boldsymbol{w}_1}^{(K)}(.), f_{\boldsymbol{w}_2}^{(K)}(.)\right)\right\}$. We use the idea discussed in [18] along with the proposed optimum estimator to construct a robust estimator. The robust clock skew and offset estimator of ϕ and δ obtained from the observations $\boldsymbol{y}$ are given by

$$\hat{\delta}_{robust}(\boldsymbol{y}) = \frac{\sum_{k=1}^{K} c_k(\boldsymbol{y})\hat{\delta}^{(k)}(\boldsymbol{y})}{\sum_{k=1}^{K} c_k(\boldsymbol{y})}, \quad (49)$$

and

$$\hat{\phi}_{robust}(\boldsymbol{y}) = \frac{\sum_{k=1}^{K} c_k(\boldsymbol{y})\hat{\phi}^{(k)}(\boldsymbol{y})}{\sum_{k=1}^{K} c_k(\boldsymbol{y})}, \quad (50)$$

where $\hat{\delta}^{(k)}(\boldsymbol{y})$ and $\hat{\phi}^{(k)}(\boldsymbol{y})$ denote the optimum invariant estimates of δ and ϕ under the k^{th} scenario, obtained using Proposition 3 and the queuing delay pdfs $\left(f^{(k)}_{\boldsymbol{w}_1}(.), f^{(k)}_{\boldsymbol{w}_2}(.)\right)$. The weights $c_k(\boldsymbol{y})$ are based on the likelihood function $f(\boldsymbol{y}|\boldsymbol{\theta})$, and are defined as

$$\begin{aligned} c_k(\boldsymbol{y}) &= \frac{1}{(\hat{\phi}^{(k)}(\boldsymbol{y}))^{2P}} f^{(k)}_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \hat{\delta}^{(k)}(\boldsymbol{y})\boldsymbol{1}_P^T}{\hat{\phi}^{(k)}(\boldsymbol{y})} - \hat{d}^{(k)}_{est}\boldsymbol{1}_P^T - \boldsymbol{t}_1\right) \\ &\quad f^{(k)}_{\boldsymbol{w}_2}\left(\frac{\hat{\delta}^{(k)}(\boldsymbol{y})\boldsymbol{1}_P^T - \boldsymbol{t}_3}{\hat{\phi}^{(k)}(\boldsymbol{y})} + \boldsymbol{t}_4 - (a_0\hat{d}^{(k)}_{est} + c_0)\boldsymbol{1}_P^T\right), \quad (51) \end{aligned}$$

where $\hat{d}^{(k)}_{est}$ is an estimate of d corresponding to the k^{th} scenario and the constants a_0 and c_0 were defined in Section II. To obtain $\hat{d}^{(k)}_{est}$, we use $\hat{\delta}^{(k)}$ and $\hat{\phi}^{(k)}$ to calculate $\boldsymbol{z}_1 = \left(\frac{(\boldsymbol{t}_2 - \hat{\delta}^{(k)}\boldsymbol{1}_P^T)}{\hat{\phi}^{(k)}} - \boldsymbol{t}_1\right)$ and $\boldsymbol{z}_2 = \left(\boldsymbol{t}_4 - \frac{(\boldsymbol{t}_3 - \hat{\delta}^{(k)}\boldsymbol{1}_P^T)}{\hat{\phi}^{(k)}}\right)$. We now have

$$\boldsymbol{z}_1 \approx d\boldsymbol{1}_P + \boldsymbol{w}_1, \quad (52)$$

and

$$\boldsymbol{z}_2 \approx (a_0 d + c_0)\boldsymbol{1}_P + \boldsymbol{w}_2. \quad (53)$$

The problem of estimating d from $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ falls under the class of location parameter estimation problems. Hence, we use the optimum invariant location parameter estimator proposed in [10] to calculate $\hat{d}^{(k)}_{est}$, given by

$$\hat{d}^{(k)}_{est} = \frac{\int_{\mathbb{R}} d f^{(k)}_{\boldsymbol{w}_1}(\boldsymbol{z}_1 - d\boldsymbol{1}_P^T) f^{(k)}_{\boldsymbol{w}_2}(\boldsymbol{z}_2 - (a_0 d + c_0)\boldsymbol{1}_P^T) d(d)}{\int_{\mathbb{R}} f^{(k)}_{\boldsymbol{w}_1}(\boldsymbol{z}_1 - d\boldsymbol{1}_P^T) f^{(k)}_{\boldsymbol{w}_2}(\boldsymbol{z}_2 - (a_0 d + c_0)\boldsymbol{1}_P^T) d(d)}. \quad (54)$$

The weights $c_k(\boldsymbol{y})$ are chosen as a function of likelihood in order to give more weight to the more plausible models, i.e., we assign higher weight when the assumed pdf of the queuing delays is judged to be closer to the actual queuing delay pdf. As the estimators $\hat{\delta}_{robust}$ and $\hat{\phi}_{robust}$ are linear combination of estimates that are invariant under $\mathcal{G}_{SModel}$ defined in (41), the robust clock skew and offset estimators presented in (49) and (50) are invariant under $\mathcal{G}_{SModel}$.

VI. Optimum Invariant CSOE scheme under M-model

In this section, we apply invariant decision theory to derive an optimum invariant estimator of ϕ and δ under the *M-model* assuming complete knowledge of the joint queuing delay pdfs and unlimited computational complexity. Recall from (11), the observations under the *M-model* can be represented as

$$\boldsymbol{y} = (\boldsymbol{h}_M d + \boldsymbol{z})\phi + (\boldsymbol{\delta} \otimes \boldsymbol{1}_{2P}), \quad (55)$$

where $\boldsymbol{y} \in \mathbb{R}^{2P(B+1)}$, $\boldsymbol{z} \in \mathbb{R}^{2P(B+1)}$, $\phi \in \mathbb{R}^+$, $\boldsymbol{\delta} = [\delta, \delta_1, \delta_2, \cdots, \delta_B] \in \mathbb{R}^{B+1}$ and $\boldsymbol{h}_M$ is a known vector defined in Section II and is given by $\boldsymbol{h}_M = [\boldsymbol{1}^T_{P(B+1)}, -a_0\boldsymbol{1}^T_{P(B+1)}]^T$. Let $\boldsymbol{\theta} = [\phi, d, \delta, \delta_1, \cdots, \delta_B]$ denote the vector of unknown parameters. The unrestricted parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}$, is given by

$$\boldsymbol{\Theta} = \{(\phi, d, \boldsymbol{\delta}) : \phi \in \mathbb{R}^+, d \in \mathbb{R}, \boldsymbol{\delta} \in \mathbb{R}^{B+1}\}, \quad (56)$$

and the restricted parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}^*$, is given by,

$$\boldsymbol{\Theta}^* = \{(\phi, d, \boldsymbol{\delta}) : \phi \in \mathbb{R}^+, d \in \mathbb{R}_0^+, \boldsymbol{\delta} \in \mathbb{R}^{B+1}\} \quad (57)$$

From (11), the conditional pdf of $\boldsymbol{y}$ is given by

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\theta}) &= \frac{1}{\phi^{2P(B+1)}} f_{\boldsymbol{z}}\left(\frac{\boldsymbol{y} - (\boldsymbol{\delta} \otimes \boldsymbol{1}_{2P})}{\phi} - \boldsymbol{h}_M d\right), \quad (58) \\ &= \frac{1}{\phi^{2P(B+1)}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \delta\boldsymbol{1}_P^T}{\phi} - d\boldsymbol{1}_P^T - \boldsymbol{t}_1\right) \\ &\quad f_{\boldsymbol{w}_2}\left(\frac{\delta\boldsymbol{1}_P^T - \boldsymbol{t}_3}{\phi} + \boldsymbol{t}_4 - (a_0 d + c_0)\boldsymbol{1}_P^T\right) \\ &\quad \prod_{j=1}^{B}\left[f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_{2j} - \delta_j\boldsymbol{1}_P^T}{\phi} - d\boldsymbol{1}_P^T - \boldsymbol{t}_{1i}\right)\right. \\ &\quad \left. f_{\boldsymbol{w}_2}\left(\frac{\delta_j\boldsymbol{1}_P^T - \boldsymbol{t}_{3j}}{\phi} + \boldsymbol{t}_{4j} - (a_0 d + c_0)\boldsymbol{1}_P^T\right)\right], \quad (59) \end{aligned}$$

where a_0 and c_0 are known constants defined in Section II.

Let $\mathcal{F}_{MModel}$ denote the class of all pdfs $f(\boldsymbol{y}|\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The class of such pdfs is invariant under the group of transformations $\mathcal{G}_{MModel}$, on $\mathbb{R}^{2P(B+1)}$, defined as

$$\begin{aligned} \mathcal{G}_{MModel} = \{&g_{a,b,\boldsymbol{c}}(\boldsymbol{m}) : g_{a,b,\boldsymbol{c}}(\boldsymbol{m}) = \\ &(\boldsymbol{m} + \boldsymbol{h}_M b)a + \boldsymbol{c} \otimes \boldsymbol{1}_{2P}, \\ &\forall (a, b, \boldsymbol{c}) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^{B+1}\}, \quad (60) \end{aligned}$$

where $\boldsymbol{m} \in \mathbb{R}^{2P(B+1)}$, since $\boldsymbol{y}_g = g_{a,b,\boldsymbol{c}}(\boldsymbol{y})$ has a pdf given by $\frac{1}{(a\phi)^{2P(B+1)}} f_{\boldsymbol{z}}\left(\frac{\boldsymbol{y}_g - ((a\boldsymbol{\delta} + \boldsymbol{c}) \otimes \boldsymbol{1}_{2P})}{a\phi} - \boldsymbol{h}_M\left(d + \frac{b}{\phi}\right)\right)$. The corresponding group of induced transformations on $\boldsymbol{\Theta}$, denoted by $\bar{\mathcal{G}}_{MModel}$, is given by

$$\begin{aligned} \bar{\mathcal{G}}_{MModel} = \{&\bar{g}_{a,b,\boldsymbol{c}}((\phi, d, \boldsymbol{\delta})) : \bar{g}_{a,b,\boldsymbol{c}}((\phi, d, \boldsymbol{\delta})) = \\ &(a\phi, (d + b/\phi), (a\boldsymbol{\delta} + \boldsymbol{c})), \\ &\forall (a, b, \boldsymbol{c}) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^{B+1}\}, \quad (61) \end{aligned}$$

where $\phi \in \mathbb{R}^+, d \in \mathbb{R}$ and $\boldsymbol{\delta} \in \mathbb{R}^{B+1}$. Note the similarity to (42).

Let $\hat{\delta}_I$ and $\hat{\phi}_I$ denote estimators of δ and ϕ, respectively and let $\hat{\delta}_I(\boldsymbol{y})$ and $\hat{\phi}_I(\boldsymbol{y})$ denote the estimates obtained from the received data $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\phi^{2P(B+1)}} f_{\boldsymbol{z}}\left(\frac{\boldsymbol{y} - (\boldsymbol{\delta} \otimes \boldsymbol{1}_{2P})}{\phi} - \boldsymbol{h}_M d\right)$. Let $\boldsymbol{c} = [c_1, c_2, \cdots, c_{B+1}] \in \mathbb{R}^{B+1}$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_{MModel}$ from (60), if for all $(a, b, \boldsymbol{c}) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^{B+1}$,

$$\begin{aligned} \hat{\delta}_I(g_{a,b,\boldsymbol{c}}(\boldsymbol{y})) &= \hat{\delta}_I(a(\boldsymbol{y} + \boldsymbol{h}_M b) + \boldsymbol{c} \otimes \boldsymbol{1}_{2P}) \\ &= a\hat{\delta}_I(\boldsymbol{y}) + c_1, \quad (62) \\ \hat{\phi}_I(g_{a,b,\boldsymbol{c}}(\boldsymbol{y})) &= \hat{\phi}_I(a(\boldsymbol{y} + \boldsymbol{h}_M b) + \boldsymbol{c} \otimes \boldsymbol{1}_{2P}) \\ &= a\hat{\phi}_I(\boldsymbol{y}). \quad (63) \end{aligned}$$

Note that the estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant to the parameter d, i.e., the estimates, as well as the performance of the estimators, are not affected by the value of d. Further, the skew-normalized loss functions defined in (17) and (18) for δ and ϕ respectively, are invariant under $\mathcal{G}_{MModel}$ from (60), since

$$\frac{(\hat{\delta}_I(\boldsymbol{y})-\delta)^2}{\phi^2} = \frac{\left(\hat{\delta}_I(g_{a,b,\boldsymbol{c}}(\boldsymbol{y}))-(a\delta+c_1)\right)^2}{a^2\phi^2}, \quad (64)$$

and

$$\frac{(\hat{\phi}_I(\boldsymbol{y})-\phi)^2}{\phi^2} = \frac{\left(\hat{\phi}_I(g_{a,b,\boldsymbol{c}}(\boldsymbol{y}))-a\phi\right)^2}{a^2\phi^2} \quad (65)$$

for all $g_{a,b,\boldsymbol{c}} \in \mathcal{G}_{MModel}$. We now present the minimax optimum estimators of δ and ϕ under the *M-model*.

Proposition 5. *The optimum (or minimum conditional risk) invariant estimators of δ and ϕ, denoted by $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, respectively, under $\mathcal{G}_{MModel}$ defined in (60), for the skew normalized squared error loss functions defined in (17) and (18) are given by*

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{B+2}} \frac{\delta}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{B+2}} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}, \quad (66)$$

and

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{B+2}} \frac{1}{\phi} f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{B+2}} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}, \quad (67)$$

respectively, where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is defined in (59). Further, the derived optimum invariant estimators are minimax for the skew-normalized squared error loss in the restricted parameter space $\boldsymbol{\Theta}^*$.

Outline of the Proof. Following steps similar to those in Appendix B, we can show that the right invariant prior, denoted by $\pi^r(\boldsymbol{\theta})$, for $\mathcal{G}_{MModel}$ defined in (61) is given by $\pi^r(\boldsymbol{\theta}) = \mathcal{I}_{\mathbb{R}^+}(\phi)\mathcal{I}_{\mathbb{R}}(d)\mathcal{I}_{\mathbb{R}^{B+1}}(\boldsymbol{\delta})$. Using the obtained $\pi^r(\boldsymbol{\theta})$ and following similar steps to those in Appendix B, we can obtain the optimum invariant estimators and show that they are minimax optimum.

VII. Simulation Results

In this section, we illustrate the performance of the optimum estimators via numerical simulations in the LTE backhaul network scenario described in Section I. PTP is sometimes used in conjunction with Synchronous Ethernet (SyncE) for cellular base station synchronization in 4G LTE networks. Although the SyncE standards are now mature, much of the deployed base of Ethernet equipment does not support it [21]. PTP is the primary option for synchronization to operators with packet backhaul networks that do not support SyncE [21], [22]. For simplicity, we assume symmetric network conditions in the forward and reverse paths, i.e., $f_{\boldsymbol{w}_1}(.) = f_{\boldsymbol{w}_2}(.) = f_{\boldsymbol{w}}(.)$. Also, we assume the queuing delay samples are independent and identically distributed.

We follow the approach given in [10] for generating the random queuing delays in LTE backhaul networks. Specifically, we assume a Gigabit Ethernet network consisting of a cascade of 10 switches between the master and slave nodes. A two-class non-preemptive priority queue is used to model the traffic at each switch. The network traffic at the switch comprises of the lower priority background traffic and the higher priority synchronization messages. We assume cross-traffic flows, where new background traffic is injected at each switch and this traffic exits at the subsequent switch. The arrival times and size of background traffic packets injected at each switch are assumed to be statistically independent. We use Traffic Model 1 (TM-1) and Traffic Model 2 (TM-2) from the ITU-T specification G.8261 [23], described in Table I, for generating the background traffic at each switch. The interarrival times between packets in background traffic are assumed to follow an exponential distribution, and we set the rate parameter of each exponential distribution accordingly to obtain the desired load factor, i.e., the percentage of the total capacity consumed by background traffic [10]. The empirical pdf of the PDV in the backhaul networks was obtained in [10] for different load factors. The timestamps t_{1i} and t_{3i} are set to $40i\ \mu s$ and $40i\ \mu s{+}20\mu s$, respectively, for $i = 0, 1, \cdots, P-1$. For a given value of parameters $\{\phi, d, \delta\}$, the timestamps t_{2i} and t_{4i} are then generated using the appropriate equations, for example (3) and (4), assuming $d_{ms} = d$ and $d_{sm} = a_0 d + c_0$, where a_0 and c_0 are known constants.

Traffic Model	Packet Sizes (in Bytes)	% of total traffic
TM-1	{64, 576, 1518}	{80%, 5%, 15%}
TM-2	{64, 576, 1518}	{30%, 10%, 60%}

TABLE I: Composition of background packets in the considered traffic models.

A. Considered CSOE schemes

We now briefly describe the various CSOE schemes available in the literature for which we also evaluate performance:

1) Least Squares Estimate (LSE): We assume the *K-model* for this CSOE scheme. In this scheme, we assume prior information of the mean and variance of $f_{\boldsymbol{w}}(.)$. We use the least squares estimator to get an estimate of ϕ and δ from (5). It can be shown that the least squares CSOE scheme is invariant under $\mathcal{G}_{KModel}$ defined in (21).

2) Local Maximum Likelihood Estimate (LMLE): We assume the *K-model* for this CSOE scheme. As discussed in Proposition 2, the ML estimate under the *K-model* is obtaining by finding the value of parameter vector that maximizes the likelihood function (see (31)). However, for small values of P, the likelihood function need not always be concave. The likelihood function is shown in Figure 1 for a TM-1 network scenario under 40% load for $\phi = 1$ and $\delta = 0$ for different values of P. We see that for small values of P, the likelihood function is not necessarily concave and sometimes it has many local maxima. In our simulations, we use the solution obtained from the least squares estimate as the initial point in the search for the ML estimate. The obtained solution is called the *Local Maximum Likelihood Estimate* since we cannot guarantee a global maximum.

Remark. We should mention here that we have used the *K-model* for the least squares and ML-based CSOE schemes. We conjecture that this provides a lower bound on the performance of these CSOE schemes under the *S-model*, as the presence of additional unknown nuisance parameters would generally degrade the performance of an estimation scheme.

B. Performance metric

We now describe the metric used for illustrating the performance of the considered CSOE schemes. Let $\hat{\delta}$ and $\hat{\phi}$ denote estimators of δ and ϕ, respectively. The skew Normalized Root Mean Square estimation Error (NRMSE) of $\hat{\delta}$ and $\hat{\phi}$, denoted by NRMSE($\hat{\delta}$) and NRMSE($\hat{\phi}$), respectively, are defined as $\frac{\sqrt{\text{MSE}(\hat{\delta})}}{\phi}$ and $\frac{\sqrt{\text{MSE}(\hat{\phi})}}{\phi}$, respectively, where MSE($\hat{\delta}$) and MSE($\hat{\phi}$) are defined in (29). In our results, we use the NRMSE($\hat{\delta}$) and NRMSE($\hat{\phi}$) metrics to evaluate performance. Also, note that the risk of the estimators $\hat{\delta}$ and $\hat{\phi}$ under the skew normalized squared error loss are given by $\mathcal{R}(\hat{\delta}, \boldsymbol{\theta}) = (\text{NRMSE}(\hat{\delta}))^2$ and $\mathcal{R}(\hat{\phi}, \boldsymbol{\theta}) = (\text{NRMSE}(\hat{\phi}))^2$, respectively.

Remark. In scenarios where analytical expressions for the queuing delay pdfs $f_{\boldsymbol{w}_1}(.)$ and $f_{\boldsymbol{w}_2}(.)$ are known, it might be possible to further simplify the integrals in Proposition 1, 3 and 5. However, in the general case of arbitrary queuing delay pdfs $f_{\boldsymbol{w}_1}(.)$ and $f_{\boldsymbol{w}_2}(.)$, these integrals are computed by approximating them with Riemann summations. In such cases, the computational complexity associated with the optimum estimators will depend on the number of bins used in the Riemann summations. Typically, this computational complexity is significantly higher than that of conventional ML-based estimators. The performance comparison between the realistic schemes and the optimum estimators can be performed offline, where complexity is not a stringent issue.

In this paper, we approximate the integral over $\mathbb{R}^+$ using Riemann sums by setting the width of the Riemann summation bins to 0.01 and the limits of the integral to $[0.5, 2]$. Also, we approximate the integral over $\mathbb{R}$ using Riemann sums by setting the width of the Riemann summation bins to $0.01\ \mu s$ and the limits of the integral to $[-20\ \mu s, 20\ \mu s]$.

C. Numerical results

Figures 2 and 3 shows the NRMSE performance for the considered CSOE schemes for $\{\phi, d, \delta\} = \{1.01, 1\ \mu s, 1.25\ \mu s\}$ with $d_{ms} = d$ and $d_{sm} = d + 1\ \mu s$ for TM-1 and TM-2 network scenarios. We see that the performance of all the considered CSOE schemes improves with an increase in the number of two-way message exchanges. Some key observations are as follows:

1) *Performance of minimax optimum estimators:* Figure 2 compares the performance of the optimum estimator under the *K-* and *S-models*, namely the the minimax optimum estimator under the *K-model* (*Minimax-K*) and the minimax optimum estimator under the *S-model* (*Minimax-S*), to the performance of other CSOE schemes available in the literature. Interestingly, we do not observe a significant loss in performance of *Minimax-S* due to the unknown nuisance parameter d. Further, we observe that the robust estimator described in Section V (*Robust-CSOE*) exhibits a performance close to the optimum estimators[2] indicating that the robust CSOE scheme is relatively robust to network uncertainties. Figure 3 shows us the performance of the minimax optimum estimator under the *M-model* (*Minimax-M*) for different values of B for different network scenarios. We observe a noticeable gain in performance when estimating ϕ by using information from the past blocks, since the additional timestamps contain information regarding the clock skew ϕ. Also, we observe a slight gain in performance when estimating the clock offset δ. Although the previous blocks do not provide us information regarding the current block's clock offset δ, the additional timestamps help in improving the estimate of ϕ, which in turn provides a performance gain when estimating the current block's clock offset δ.
2) *Performance of the minimax optimum estimators for different values of ϕ and δ:* Figure 4 shows us the performance of the minimax optimum estimator under the *K-model* for different values of ϕ and δ. As observed from the results, the performance of the optimum invariant estimator is independent of the parameter values $\{\phi, d, \delta\}$ since the conditional risk of an invariant estimator is constant (see Theorem 1). Similarly, using Theorem 1, we can infer that the performance of the optimum invariant estimators under the *S-model* and *M-model* are also independent of the parameter values.
3) *Effect of width of Riemann summation bins:* In our simulation results, we set the width of the Riemann summation bins to small values to ensure that the additional error introduced due to the Riemann sum approximation is small relative to the estimation error. However, the computational complexity associated with the developed optimum estimators depends on the total number of bins (or the width of the Riemann summation bins) used in the Riemann sum approximation. For example, consider the optimum estimator under the *K-model*. It is easy to see that when Riemann sums are used, $\mathcal{O}(P^2 N_{b_1} N_{b_2})$ multiplications and $\mathcal{O}(N_{b_1} N_{b_2})$ additions are required per estimate, where N_{b_1} and N_{b_2} denote the total number of Riemann sum bins utilized in approximating the integral over $\mathbb{R}^+$ and the integral over $\mathbb{R}$, respectively. In Figure 5, we compare the performance of the minimax optimum estimator under the *K-model* for different values of the width of the Riemann summation bins, when approximating the integral over $\mathbb{R}$. From Figure 5, we do not observe a noticeable loss in performance when the width of the Riemann summation bins is increased from $0.01\ \mu s$ to $0.1\ \mu s$. However, there is a significant degradation in the performance of the optimum estimator when the width of the Riemann summation bins is increased to $0.5\ \mu s$ or higher. For a given bin width, there is always some large

[2] A total of 14 possible pdfs are assumed for the PDV pdf consisting of pdfs corresponding to TM-1 and TM-2 at $\{20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%\}$ load were assumed available for the robust CSOE scheme.

value of P above which the performance of the minimax optimum estimator does not improve with an increasing number of two-way message exchanges. Apparently, the error in approximating the integral is much larger than the estimation error with the exact integral at such P.

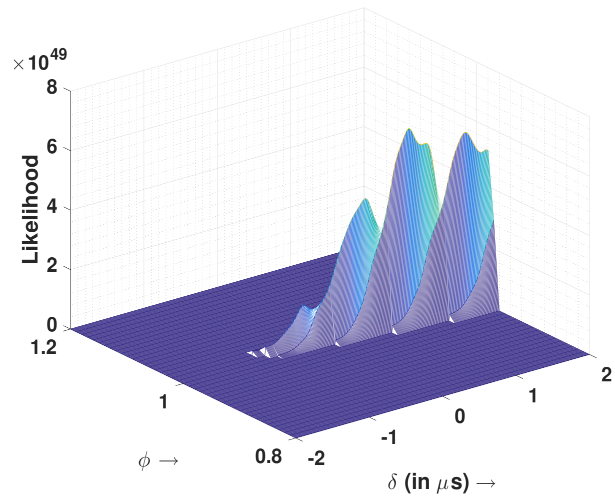
4) *Effect of unknown path asymmetries:* When designing the optimum estimators, we assumed a prior known relationship between the fixed path delays, d_{ms} and d_{sm}. We now study the possible performance loss that could occur due to the presence of unknown path asymmetries. Figure 6 compares the performance of the minimax optimum estimator under the *K-model* in the presence of such unknown path asymmetries, namely when the estimator assumes that $d_{ms} = d_{sm} = d$, when the actual relationship is given by $d_{ms} = d$ and $d_{sm} = d + 1\ \mu s$. From Figure 6, we observe a significant degradation in the performance of the clock offset estimator as well as a noticeable degradation in the performance of the clock skew estimator. The loss in performance follows from [24], [25], where it was shown that the presence of unknown path asymmetries in PTP can result in significant degradation of the performance of a clock skew and offset estimation scheme. We note that the loss can be very different in other cases.

VIII. Conclusion and Future Work

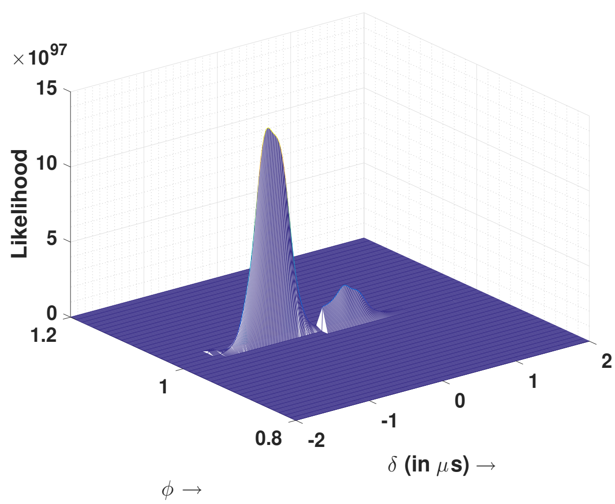
In this paper, we have developed optimum invariant estimators for the joint estimation of clock skew and offset in the IEEE 1588 PTP for different observation models assuming knowledge of queuing delay pdfs and unlimited computational complexity. The performance benchmarks obtained from the optimum estimators can aid system designers in searching for algorithms with the desired computational complexity that have near optimum performance. This is a topic of great interest. Further, using the optimum estimator and assuming unlimited computational complexity, we construct robust clock skew and offset estimators for the *S-model* under scenarios where the queuing delay pdfs are not entirely known. While these estimators show some potential, much more study is needed to fully understand their performance. Throughout this paper, we assumed either the complete knowledge of the fixed delays or a prior known affine relationship between the fixed path delays. The presence of an unknown asymmetry between the fixed path delays could significantly degrade the performance of the developed CSOE schemes. Future work can look into developing low complexity robust clock skew and offset estimation schemes when there is an unknown asymmetry between the fixed path delays.

Appendix A
Proof of Proposition 1

Proof. The optimum invariant estimator of δ under $\mathcal{G}_{KModel}$ in (21), denoted by $\hat{\delta}_{MinRisk}$, can be obtained by solving (See



(a) $P = 5$



(b) $P = 10$

Fig. 1: Likelihood function for various values of the parameter for TM-1 under 40% load for $\phi = 1$, $\delta = 0$ for different values of P.

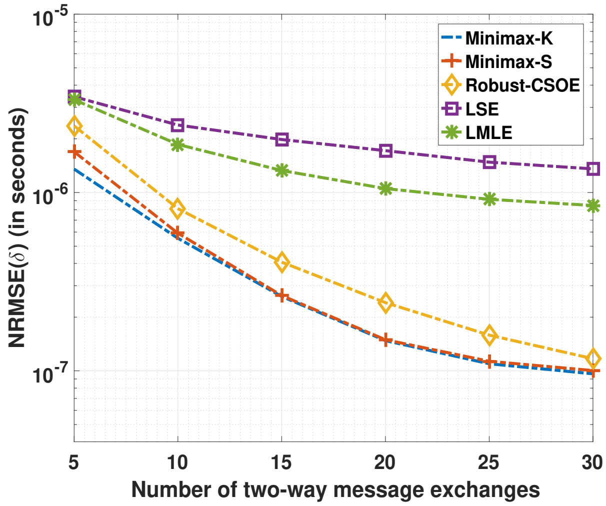
Result 3 in Section 6.6.2 of [15])

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} L_1(\hat{\delta}(\boldsymbol{y}), \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}$$

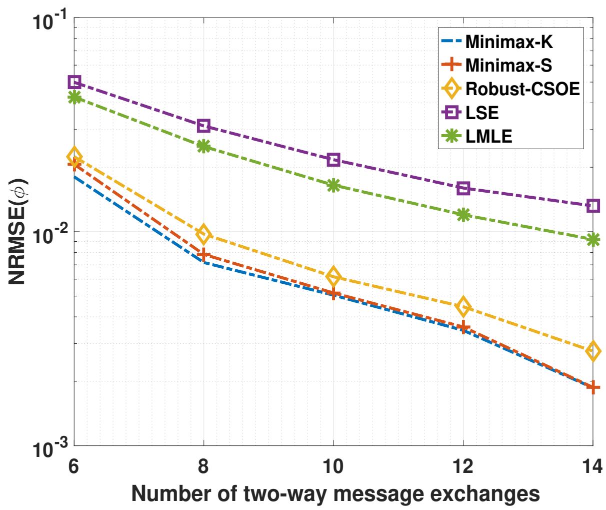
$$= \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\delta}(\boldsymbol{y}) - \delta)^2}{\phi^2} \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}, \quad (68)$$

where $\pi^r(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi^r(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y}|\boldsymbol{\theta})\pi^r(\boldsymbol{\theta})d\boldsymbol{\theta}}$ is the posterior density of $\boldsymbol{\theta}$ based on the right invariant prior π^r on $\boldsymbol{\Theta}$ (see Section 6.6.1, [15])[3]. The right invariant prior for the location-scale group was derived in [15] (see Section 6.6). As $\mathcal{G}_{KModel}$ from (21) is a location-scale group, the right invariant prior density for $\mathcal{G}_{KModel}$ is given by $\pi^r(\boldsymbol{\theta}) = \frac{1}{\phi}\mathcal{I}_{\mathbb{R}^+}(\phi)\mathcal{I}_{\mathbb{R}}(\delta)$. To

[3]The right invariant prior density need not be an actual density [15] (See section 6.6, page 409).



(a) NRMSE of clock offset for TM-1 under 40% load



(b) NRMSE of clock skew for TM-1 under 40% load

Fig. 2: NRMSE of clock offset and clock skew for various estimation schemes for $\{\phi, d, \delta\} = \{1.01, 1.0\ \mu s, 1.25\ \mu s\}$.

find $\hat{\delta}_{MinRisk}$, we differentiate the objective function in (68) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and solve for $\hat{\delta}_{MinRisk}$. We obtain

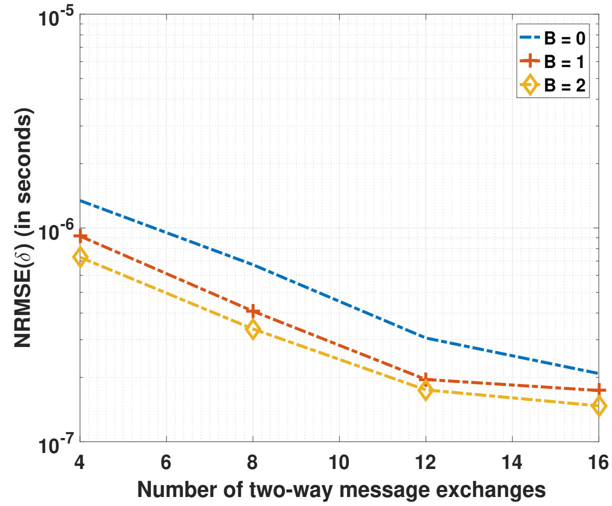
$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\delta}{\phi^2}\pi^r(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^2}\pi^r(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}} = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\delta}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{69}$$

Similarly, the optimum invariant estimator of ϕ under $\mathcal{G}_{KModel}$ in (21), denoted by $\hat{\phi}_{MinRisk}$, can be obtained by

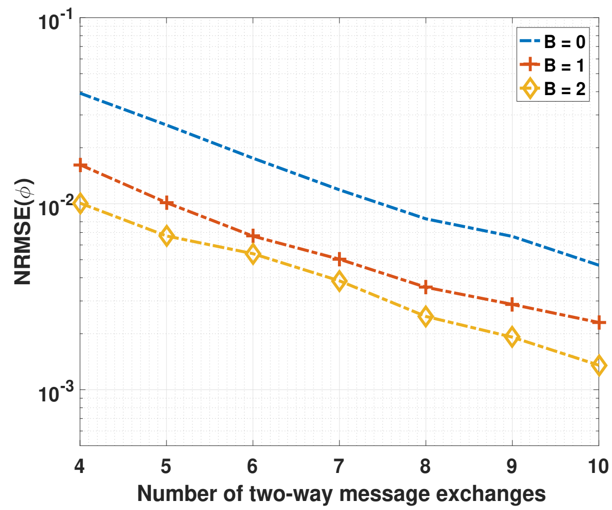
$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\phi}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\phi}(\boldsymbol{y}) - \phi)^2}{\phi^2}\pi^r(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}. \tag{70}$$

Using the same derivative-based approach, we obtain

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^2}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3}f(\boldsymbol{y}|\boldsymbol{\theta})d\delta d\phi}. \tag{71}$$



(a) NRMSE of clock offset for TM-1 under 40% load

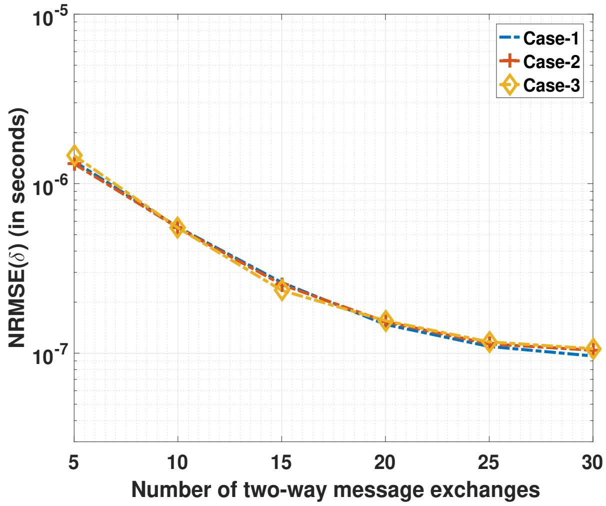


(b) NRMSE of clock skew for TM-1 under 40% load

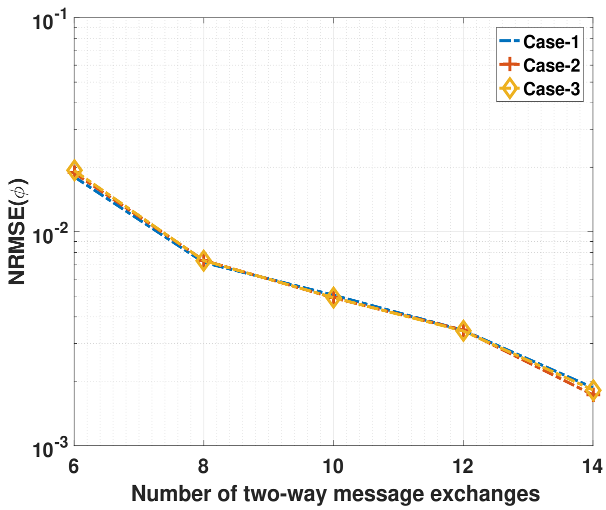
Fig. 3: NRMSE of clock offset and clock skew of minimax optimum estimator under the *M-model* for $\{\phi, d, \delta\} = \{1.01, 1.0\ \mu s, 1.25\ \mu s\}$ for different values of B past observation windows.

When the class of densities is invariant under the location-scale group, it was shown in [15] that the optimum invariant estimator of a parameter for an invariant loss function is also a minimax estimator of the parameter for the considered loss function. As the class of densities $\mathcal{F}_{KModel}$ is invariant under $\mathcal{G}_{KModel}$ in (21) (a location-scale group), and the scale invariant loss function is invariant under $\mathcal{G}_{KModel}$, the optimum invariant estimators $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, are minimax optimum estimators of δ and ϕ, respectively, for the skew-normalized squared error loss functions given in (17) and (18). □

Appendix B
Proof of Proposition 3

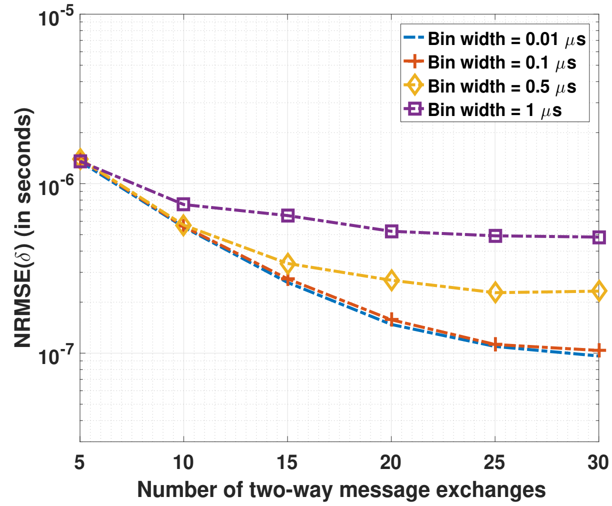


(a) NRMSE of clock offset for TM-1 under 40% load

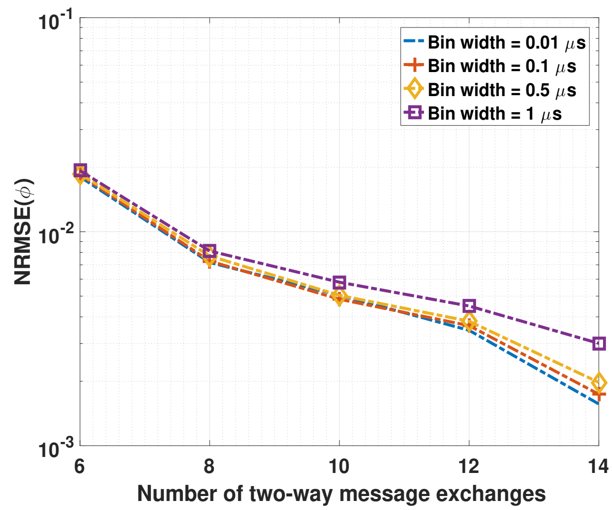


(b) NRMSE of clock skew for TM-1 under 40% load

Fig. 4: NRMSE performance of minimax optimum estimator under *K-model* for different parameter values. We have for case 1, $\{\phi, d, \delta\} = \{1.01, 1.0\mu s, 1.25\ \mu s\}$, for case 2.1, $\{\phi, d, \delta\} = \{1.05, 1.0\mu s, 1.25\ \mu s\}$ and for case 3, $\{\phi, d, \delta\} = \{0.95, 1.0\mu s, -1.25\ \mu s\}$.



(a) NRMSE of clock offset for TM-1 under 40% load



(b) NRMSE of clock skew for TM-1 under 40% load

Fig. 5: NRMSE performance of minimax optimum estimator under *K-model* for different widths of Riemann summation bins for $\{\phi, d, \delta\} = \{1.01, 1.0\mu s, 1.25\ \mu s\}$.

Proof. We first calculate the right invariant prior for $\mathcal{G}_{SModel}$, defined in (41), as it is necessary for deriving the optimum invariant estimator under $\mathcal{G}_{SModel}$. Let $\mathcal{A} \subseteq \boldsymbol{\Theta}$ and $\boldsymbol{\theta}_0 = (\phi_0, d_0, \delta_0) \in \boldsymbol{\Theta}$, with $\boldsymbol{\Theta}$ defined in (39). The right group transformation of $\mathcal{A}$ by $\boldsymbol{\theta}_0$ is given by [16]

$$\mathcal{A}_{r_0} =\{\boldsymbol{\theta}_{r_0} = (\phi_{r_0}, d_{r_0}, \delta_{r_0}) : \boldsymbol{\theta}_{r_0} = \bar{g}_{\phi,d,\delta}(\boldsymbol{\theta}_0), (\phi, d, \delta) \in \mathcal{A}\}, \quad (72)$$

$$=\{\boldsymbol{\theta}_{r_0} = (\phi\phi_0, d_0 + d/\phi_0, \phi\delta_0 + \delta) : (\phi, d, \delta) \in \mathcal{A}\}, \quad (73)$$

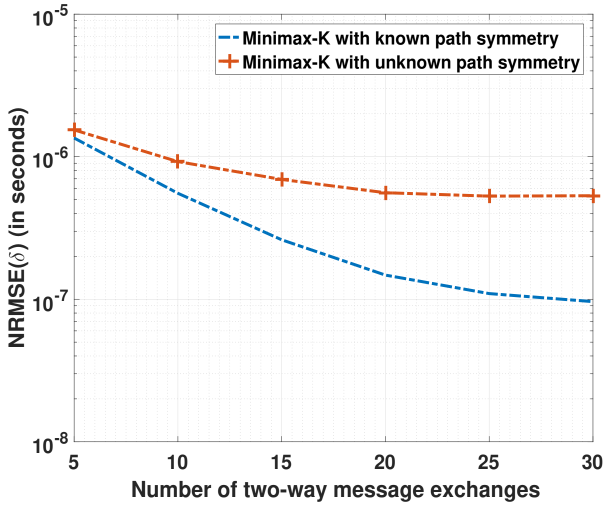
with $\bar{g}_{\phi,d,\delta} \in \bar{\mathcal{G}}_{SModel}$ from (42). The right invariant prior, π^r, on $\mathcal{G}_{SModel}$ from (41) is obtained by finding the function that satisfies[4]

$$\int_{\mathcal{A}} \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad = \quad \int_{\mathcal{A}_{r_0}} \pi^r(\boldsymbol{\theta}_{r_0}) d\boldsymbol{\theta}_{r_0}, \quad (74)$$

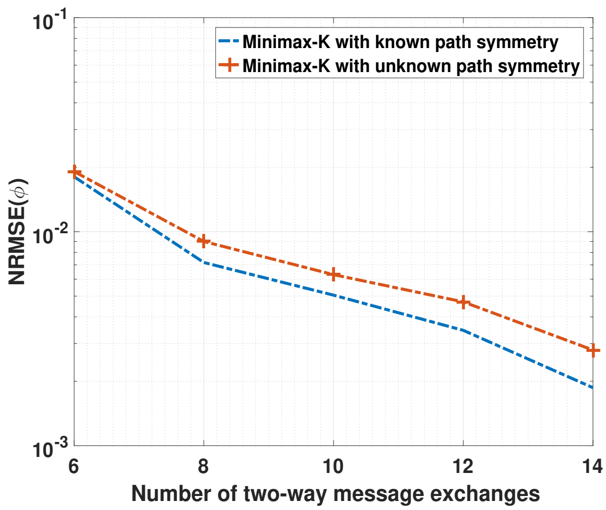
for all $\mathcal{A} \subseteq \boldsymbol{\Theta}$, for all $\bar{g}_{\phi,d,\delta} \in \bar{\mathcal{G}}_{SModel}$ and for all $\boldsymbol{\theta}_0 = (\phi_0, d_0, \delta_0) \in \boldsymbol{\Theta}$. The right invariant prior for $\bar{\mathcal{G}}_{SModel}$ is given by $\pi^r(\boldsymbol{\theta}) = \mathcal{I}_{\mathbb{R}^+}(\phi)\mathcal{I}_{\mathbb{R}}(d)\mathcal{I}_{\mathbb{R}}(\delta)$. To see this, note that

$$\int_{\mathcal{A}} 1 d\boldsymbol{\theta} = \int_{\mathcal{A}_{r_0}} \frac{d\boldsymbol{\theta}}{d\boldsymbol{\theta}_{r_0}} d\boldsymbol{\theta}_{r_0} = \int_{\mathcal{A}_{r_0}} 1 d\boldsymbol{\theta}_{r_0}, \quad (75)$$

[4]The right invariant prior is invariant to the right transformation of the parameters in the parameter space. Similarly, the left invariant prior can also be constructed. However, we are interested only in the right invariant prior as it is used in deriving the optimum invariant estimator.



(a) NRMSE of clock offset for TM-1 under 40% load



(b) NRMSE of clock skew for TM-1 under 40% load

Fig. 6: NRMSE performance of minimax optimum estimator under *K-model* in the presence of unknown path asymmetries for $\{\phi, d, \delta\} = \{1.01, 1.0\mu s, 1.25\ \mu s\}$.

since the Jacobian of the transformation in (73) is given by

$$\frac{d\boldsymbol{\theta}_{r_0}}{d\boldsymbol{\theta}} = \det\left(\begin{bmatrix} \frac{\partial \phi_{r_0}}{\partial \phi} & \frac{\partial \phi_{r_0}}{\partial d} & \frac{\partial \phi_{r_0}}{\partial \delta} \\ \frac{\partial d_{r_0}}{\partial \phi} & \frac{\partial d_{r_0}}{\partial d} & \frac{\partial d_{r_0}}{\partial \delta} \\ \frac{\partial \delta_{r_0}}{\partial \phi} & \frac{\partial \delta_{r_0}}{\partial d} & \frac{\partial \delta_{r_0}}{\partial \delta} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \phi_0 & 0 & 0 \\ 0 & 1/\phi_0 & 0 \\ \delta_0 & 0 & 1 \end{bmatrix}\right) = 1. \quad (76)$$

The optimum invariant estimators of δ under $\mathcal{G}_{SModel}$ from (41), denoted by $\hat{\delta}_{MinRisk}$, can now be obtained by solving

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\delta}(\boldsymbol{y}) - \delta)^2}{\phi^2} \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}, \quad (77)$$

where $\pi^r(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi^r(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y}|\boldsymbol{\theta})\pi^r(\boldsymbol{\theta})d\boldsymbol{\theta}}$ and $\pi^r(\boldsymbol{\theta})$ is the right invariant prior corresponding to $\bar{\mathcal{G}}_{SModel}$. To find $\hat{\delta}_{MinRisk}$, we differentiate the objective function in (77) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and solve for $\hat{\delta}(\boldsymbol{y})$. We have

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{\delta}{\phi^2} \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{1}{\phi^2} \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}} = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{\delta}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (78)$$

Similarly, the optimum invariant estimator of ϕ under $\mathcal{G}_{SModel}$ from (41), denoted by $\hat{\phi}_{MinRisk}$, can be obtained by solving

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\phi}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\phi}(\boldsymbol{y}) - \phi)^2}{\phi^2} \pi^r(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}. \quad (79)$$

Solving, we obtain

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{1}{\phi} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d(d) d\delta d\phi}. \quad (80)$$

Minimaxity of optimum invariant estimators in $\boldsymbol{\Theta}$: We now show the derived optimum invariant estimators are minimax in $\boldsymbol{\Theta}$ for the considered loss function. Consider a sequence of prior distributions, π_k for $\boldsymbol{\theta}$, defined on $\boldsymbol{\Theta}$ as follows

$$\pi_k(\boldsymbol{\theta}) = \frac{\mathcal{I}_{(0,k)}(\phi)\mathcal{I}_{(-k,k)}(d)\mathcal{I}_{(-k,k)}(\delta)}{N_k}, \quad (81)$$

for $k = 1, 2, \cdots,$ and $N_k = \int_{\boldsymbol{\Theta}} \mathcal{I}_{(0,k)}(\phi)\mathcal{I}_{(-k,k)}(d)\mathcal{I}_{(-k,k)}(\delta)d\boldsymbol{\theta}$. The support of π_k is given by

$$\boldsymbol{\Theta}_k = \{(\phi, d, \delta) : \phi \in (0, k), d \in (-k, k), \delta \in (-k, k)\}. \quad (82)$$

The optimal Bayes estimator of δ, denoted by $\hat{\delta}_{\pi_k}$, for $\pi_k(\boldsymbol{\theta})$ and the loss function given in (17) is obtained by

$$\begin{aligned} \hat{\delta}_{\pi_k} &= \underbrace{\arg\min}_{\hat{\delta}} \mathcal{B}(\hat{\delta}, \pi_k) \\ &= \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\delta}(\boldsymbol{y}) - \delta)^2}{\phi^2} \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi_k(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y}|\boldsymbol{\theta})\pi_k(\boldsymbol{\theta})d\boldsymbol{\theta}} \end{aligned} \quad (83)$$

Solving (83), we obtain

$$\hat{\delta}_{\pi_k}(\boldsymbol{y}) = \frac{\int_{\boldsymbol{\Theta}_k} \frac{\delta}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}_k} \frac{1}{\phi^2} f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (84)$$

As $k \to \infty$, we see that $\boldsymbol{\Theta}_k \to \boldsymbol{\Theta}$, $\hat{\delta}_{\pi_k} \to \hat{\delta}_{MinRisk}$, and

$$\mathcal{B}(\hat{\delta}_{\pi_k}, \pi_k) \to \mathcal{B}(\hat{\delta}_{MinRisk}, \pi_k) = \mathcal{M}(\hat{\delta}_{MinRisk}), \quad (85)$$

since $\hat{\delta}_{MinRisk}$ is an invariant estimator of δ (see (15) in Section III). Let $\hat{\delta}_r$ denote an estimator of δ. For the loss function given in (17), we have

$$\mathcal{M}(\hat{\delta}_r) \geq \mathcal{B}(\hat{\delta}_r, \pi_k) \geq \mathcal{B}(\hat{\delta}_{\pi_k}, \pi_k), \quad (86)$$

since the optimal Bayes estimator for a prior $\pi_k(\boldsymbol{\theta})$ achieves the lowest average risk. Let $k \to \infty$, we have $\mathcal{M}(\hat{\delta}_r) \geq \lim_{k\to\infty} \mathcal{B}(\hat{\delta}_{\pi_k}, \pi_k) = \mathcal{M}(\hat{\delta}_{MinRisk})$. Hence, the maximum risk of any estimator of δ is greater than or equal to the maximum risk of $\hat{\delta}_{MinRisk}$. Hence, $\hat{\delta}_{MinRisk}$ is a minimax estimator of δ for the skew-normalized loss function defined in (17). Similarly, we can show that $\hat{\phi}_{MinRisk}$ is a minimax estimator of ϕ for the skew-normalized loss function defined in (18).

Minimaxity of optimum invariant estimators in $\mathbf{\Theta}^*$: Marchand and Strawderman [17] gave conditions on $\bar{\mathcal{G}}_{SModel}$ defined in (42), under which the optimum invariant estimator remains minimax in the restricted parameter space, $\mathbf{\Theta}^*$ defined in (40). If there exists a sequence of functions $\{\bar{g}_{a_k,b_k,c_k}\}_{k=1}^{\infty} \in \bar{\mathcal{G}}_{SModel}$ from (42), such that

$$\bar{g}_{a_k,b_k,c_k}(\mathbf{\Theta}^*) \subseteq \bar{g}_{a_{k+1},b_{k+1},c_{k+1}}(\mathbf{\Theta}^*), \tag{87}$$

$$\bigcup_{k=1}^{\infty} \bar{g}_{a_k,b_k,c_k}(\mathbf{\Theta}^*) = \mathbf{\Theta}, \tag{88}$$

where $\bar{g}_{a_k,b_k,c_k}(\mathbf{\Theta}^*) = \{\bar{g}_{a_k,b_k,c_k}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbf{\Theta}^*\}$, then $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$ remains minimax in $\mathbf{\Theta}^*$ for the considered loss functions (See Theorem 1 of [17]). Consider the sequence of transformations from $\bar{\mathcal{G}}_{SModel}$ from (42), defined as $\bar{g}_{a_k,b_k,c_k} = \bar{g}_{1,-k,0}$ for $k = 1, 2, \cdots$. We have

$$\bar{g}_{a_k,b_k,c_k}(\mathbf{\Theta}^*) = \{(\phi, d, \delta) : \phi \in \mathbb{R}^+, d \geq (-k/\phi), \delta \in \mathbb{R}\}, \tag{89}$$

$$\bar{g}_{a_{k+1},b_{k+1},c_{k+1}}(\mathbf{\Theta}^*) = \{(\phi, d, \delta) : \phi \in \mathbb{R}^+, d \geq (-(k+1)/\phi), \delta \in \mathbb{R}\}. \tag{90}$$

For this sequence of transformations, (87) and (88) are satisfied. Hence, the optimum invariant estimators $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$ remain minimax in $\mathbf{\Theta}^*$ for the skew-normalized squared error loss functions defined in (17) and (18). □

Acknowledgment

The authors would like to thank the anonymous reviewers and the editor for their valuable comments and suggestions to improve the quality of the paper.

References

[1] "IEEE standard for a precision clock synchronization protocol for networked measurement and control systems," *IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002)*, pp. 1–300, July 2008.

[2] G. Gaderer, T. Sauter, and G. Bumiller, "Clock synchronization in powerline networks," in *International Symposium on Power Line Communications and Its Applications, 2005.*, April 2005, pp. 71–75.

[3] I. Hadvzic, D. R. Morgan, and Z. Sayeed, "A synchronization algorithm for packet mans," *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1142–1153, April 2011.

[4] "IEEE standard profile for use of IEEE 1588 precision time protocol in power system applications," *IEEE Std. C37.238-2011*, July 2011.

[5] K. Harris, "An application of IEEE 1588 to industrial automation," in *2008 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*, Sept 2008, pp. 71–76.

[6] K. L. Noh, Q. M. Chaudhari, E. Serpedin, and B. W. Suter, "Novel clock phase offset and skew estimation using two-way timing message exchanges for wireless sensor networks," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 766–777, April 2007.

[7] Q. M. Chaudhari, E. Serpedin, and K. Qaraqe, "On maximum likelihood estimation of clock offset and skew in networks with exponential delays," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1685–1697, April 2008.

[8] M. Leng and Y. C. Wu, "On clock synchronization algorithms for wireless sensor networks under unknown delay," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 182–190, Jan 2010.

[9] Y. C. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, Jan 2011.

[10] A. Guruswamy, R. S. Blum, S. Kishore, and M. Bordogna, "Minimax optimum estimators for phase synchronization in IEEE 1588," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3350–3362, Sept 2015.

[11] S. Ganeriwal, R. Kumar, and M. B. Srivastava, "Timing-sync protocol for sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*, ser. SenSys '03. New York, NY, USA: ACM, 2003, pp. 138–149.

[12] J. Van Greunen and J. Rabaey, "Lightweight time synchronization for sensor networks," in *Proceedings of the 2Nd ACM International Conference on Wireless Sensor Networks and Applications*, ser. WSNA '03. New York, NY, USA: ACM, 2003, pp. 11–19.

[13] A. K. Karthik and R. S. Blum, "Estimation theory-based robust phase offset determination in presence of possible path asymmetries," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1624–1635, April 2018.

[14] A. Guruswamy, R. S. Blum, S. Kishore, and M. Bordogna, "Minimax optimum joint frequency and phase recovery in IEEE 1588," in *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, July 2015, pp. 721–725.

[15] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, 1985.

[16] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. Springer Texts in Statistics, 1998.

[17] E. Marchand and W. E. Strawderman, "A unified minimax result for restricted parameter spaces," *Bernoulli*, vol. 18, no. 2, pp. 635–643, 05 2012. [Online]. Available: https://doi.org/10.3150/10-BEJ336

[18] N. Fournier, "Compromise pitman estimators," *Journal of Statistical Planning and Inference*, vol. 141, no. 8, pp. 2656 – 2669, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378375811000759

[19] N. M. Freris, S. R. Graham, and P. R. Kumar, "Fundamental limits on synchronizing clocks over networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1352–1364, June 2011.

[20] G. Casella and J. Berger, *Statistical inference*, 2nd ed. Pacific Grove, California: Duxbury/Thomson Learning, 2002.

[21] Tellabs, "Lte stretches synchronization to new limits," 2014.

[22] Huawei, "Ieee 1588v2 technology white paper," 2016.

[23] "Timing and synchronization aspects in packet networks," *Telecommunication Standardization Sector.*, pp. 1–300, April 2008.

[24] M. Ullmann and M. Vgeler, "Delay attacks implication on ntp and ptp time synchronization," in *2009 International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*, Oct 2009, pp. 1–6.

[25] T. Mizrahi, "A game theoretic analysis of delay attacks against time synchronization protocols," in *2012 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication Proceedings*, Sept 2012, pp. 1–6.