Recent Advances in Clock Synchronization for Packet-Switched Networks

Anantha K. Karthik[1] and Rick S. Blum[2]

[1] *Qualcomm Technologies Inc., USA; anankart@qti.qualcomm.com*
[2] *Lehigh University, USA; rblum@lehigh.edu*

ABSTRACT

Clock synchronization is a mechanism for providing a standard time to various devices across a network. This monograph provides a comprehensive overview of recent developments for clock synchronization protocols built on two-way message exchanges. Several clock synchronization protocols are available in the literature for distributing time from high-cost, high-stability clocks (termed masters) to low-cost, low-stability clocks (termed slaves) via an interconnecting network. A number of clock synchronization protocols are built on two-way message exchanges. These include the timing protocol for sensor networks (TPSN), lightweight time synchronization (LTS) protocol, tiny-sync and mini-sync, network time protocol (NTP) and the IEEE 1588 precision time protocol (PTP). 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 the fixed propagation and processing delays at the intermediate nodes along the network path between the master and slave, as well as the stochastic queuing delays at

Anantha K. Karthik and Rick S. Blum (2020), "Recent Advances in Clock Synchronization for Packet-Switched Networks", Foundations and Trends® in Signal Processing: Vol. 13, No. 4, pp 360–443. DOI: 10.1561/2000000108.

each such node. Popular probability density function (pdf) models for modeling the stochastic delays include Gaussian, exponential, gamma, Weibull, and log-normal. Although these pdf models for the stochastic queuing delays apply to several scenarios, they might not be suitable in specific scenarios such as cellular base station synchronization using mobile backhaul networks and IEEE 1588 in 4G Long Term Evolution (LTE) networks. Further, there could be possibly unknown asymmetries between the fixed path delays in the forward master-to-slave path and the reverse slave-to-master path. These unknown asymmetries could arise from various sources, including delay attacks or incorrect modeling. In this monograph, we present recent developments for clock synchronization protocols built on the two-way message exchange. After an introduction to the basic concepts and mathematical models, the optimum estimators are presented for estimating the clock skew and offset that are applicable for any pdf model of the stochastic delays. Robust algorithms that can also handle unknown path asymmetries are presented next. The focus is on techniques that consider practical, relevant measurement models in order to guide the reader from physical observations to the actual synchronization of the clocks at the slave and master.

1

Introduction

The proper functioning of a distributed network is critically dependent on the availability of a standard reference time for the various devices in the network. These devices, typically at different geospatial locations, usually perform timekeeping locally using clock hardware that exploits the periodicity of certain physical phenomena, such as the mechanical resonance of vibrating crystals (in low-cost quartz crystal oscillators), or electromagnetic transitions within cesium or rubidium atoms (available in expensive atomic clocks). However, such timekeeping techniques are subject to errors that can accumulate over large time scales. Further, the cost, size, and complexity of timekeeping hardware are typically proportional to clock stability. As a result, there are often scenarios where it is impractical to locally maintain the clock hardware required to achieve the desired level of stability due to space or budget constraints.

Clock synchronization is a mechanism for providing a standard reference time to various devices across a distributed network. It is critical in modern computer networks because every aspect of managing, securing, planning, and debugging a network involves determining when particular events happen. Time provides the standard frame of reference

for the various devices on the network. A few key areas where time-synchronized information can directly affect network operations are:

1. *Network Fault Diagnosis and Recovery*: Information regarding key network events are usually stored in switches, routers, and other dedicated devices. In case of a network fault or crash, the proper sequence of events can be established, and the root cause can be quickly identified, only if the timestamps associated with the recorded events are synchronized.

2. *File Timestamps*: In a distributed file-sharing system, a master file is maintained by a Network File Sharing (NFS) server for use by remote clients. NFS is network time-dependent. Thus, when presented with duplicate versions of the file, it saves the latest copy. However, if a client is not synchronized to the network and produces a timestamp for a remotely accessed file with a time earlier than the file maintained on the server, the client file, along with any changes, are discarded [65].

3. *Services:* Several user services, including billing and financial services, require highly accurate timestamps.

4. *Miscellaneous:* Many localization, security, and tracking protocols in distributed networks also demand the devices to timestamp their messages and events [74].

One of the most popular mechanisms for achieving standard time across a network is to use the Global Positioning System (GPS) [51, 63]. Each GPS satellite contains multiple atomic clocks that contribute accurate time data to the GPS signals. GPS receivers decode these signals, effectively synchronizing each receiver to the atomic clocks. Although GPS-based timing is very accurate, it may not be feasible to equip every device in a network with a GPS receiver. Further, GPS-based time synchronization requires line-of-sight between the network device and the GPS satellite, a condition that might not be possible for some devices in the network. GPS spoofing is also a serious concern [36, 45, 60, 61].

A popular alternative to GPS-based timing is network time distribution. Here the time from a high-cost, high-stability clock (termed master)

is distributed to low-cost, low-stability clocks (termed slaves) via an interconnecting network. Several clock synchronization protocols based on network time distribution are available in the literature. For instance, the network time protocol (NTP) [52] and the IEEE 1588 precision time protocol (PTP) [32] are widely used in IP networks, while protocols such as the timing protocol for sensor networks (TPSN) [20], lightweight time synchronization protocol (LTS) [70], tiny-sync and mini-sync [62], and reference broadcast time synchronization (RBS) protocol [10] are used in wireless sensor networks. Network time distribution is often more cost-effective than GPS-based timing, as it does not require any dedicated hardware and can often make use of the existing network resources for synchronizing devices across the network.

Though the time synchronization protocols for network time distribution differ from each other in many aspects, a fundamental mechanism common to a number of clock synchronization protocols including TPSN [20], LTS [70], tiny-sync and mini-sync [62], and PTP [32], is the two-way message exchange. This refers to the exchange of messages between a pair of nodes to achieve clock synchronization. During a two-way message exchange, a slave node exchanges a series of synchronization packets with a master node over an interconnecting network and collects timestamps corresponding to the departure and arrival times of these packets. The slave then attempts to utilize these timestamps to correct its clock. However, as with any packet-switched network, the exchanged packets experience difficult to predict (stochastic) delays as they traverse the network. These stochastic delays experienced by packets can significantly degrade the performance of various clock synchronization protocols. In this monograph, we present some recent developments to combat the degrading effects of stochastic delays for clock synchronization protocols based on two-way message exchange.

While the techniques presented in the monograph apply to many applications and any clock synchronization protocol based on two-way message exchanges, we mainly discuss the applications of our results in the context of IEEE 1588 PTP applied to telecommunication networks. IEEE 1588 PTP [32] is a popular time synchronization protocol used in a number of scenarios, including electrical grid networks [18], cellular base station synchronization in 4G Long Term Evolution (LTE) [24, 25],

substation communication networks [33] and industrial control [29]. It is cost-effective and offers accuracy comparable to GPS-based timing. Emerging technologies such as fog computing and industrial Internet of Things (IIoT) networks have identified the IEEE 802.1Q amendment for Time-Sensitive Networking (TSN) as the standard for time-predictable networking [46]. TSN employs PTP to provide a global notion of time over the local area network.

Packet-based time synchronization techniques based on PTP [32] are being increasingly considered as a viable alternative to GPS-based time synchronization as a means to provide sub microsecond-level synchronization between the cellular base stations in 4G LTE mobile networks [24, 25, 28, 55, 56, 73]. Further, PTP has been explored as a possible cost effective solution for synchronizing base stations in 5G new radio (NR) cellular networks [26, 27]. Such a high degree of synchronization accuracy between the cellular base stations ($<$1.5 μs) is necessary for 4G LTE/5G NR cellular networks to enable seamless handovers between cell towers, reduce inter-cell interference, and enable the use of MIMO techniques to improve capacity [2, 26, 27].

Packet-based synchronization based on PTP is often more cost-effective than GPS-based time synchronization as it can utilize the existing mobile backhaul network infrastructure that is used to interconnect cell towers. However, since backhaul networks are typically leased from commercial internet service providers (ISPs), mobile network operators must share their use with other commercial and residential users. Background traffic generated by these users often results in sizeable random network delays that hinder packet-based synchronization. Overcoming this problem is key to the adoption of packet-based synchronization schemes in mobile backhaul networks, especially given that the synchronization accuracy requirements are only expected to grow more stringent in the future.

2

Basic Models

In this section, we first describe the popular affine model for modeling the clock of a slave node in a network. We then describe the two-way message exchange employed in the IEEE 1588 precision time protocol (PTP), Timing protocol for sensor networks (TPSN), and Lightweight time synchronization protocol (LTS). Finally, we describe the popular probability density function (pdf) models available in the literature for modeling the statistical distribution of the stochastic queuing delays.

2.1 Clocks in Networks

Every individual device in a network has a clock. Ideally, the clock of the device, $c(t)$, should be configured such that $c(t) = t$, where t is the standard time reference of the network or the clock time of the master node. A simple approximation of the clock time at the slave node, $c_{slave,offset}(t)$, as a function of the clock time of the master node, is the clock offset model. In this model, we have

$$c_{slave,offset}(t) = t + \delta, \tag{2.1}$$

where δ denotes the clock offset between the clocks of the master and slave node. Many publications [8, 21, 22, 34] have considered this model

which is sometimes suitable, for example with synchronous Ethernet in LTE backhaul networks [24, 25, 30] when frequency synchronization is already maintained. Although the clock-offset model is simple, in cases where frequency synchronization is not already maintained it only coarsely reflects the actual measured clock behavior. Hence, the accuracy of clock synchronization algorithms relying on this model is generally poor when frequency synchronization is not already maintained [14].

The affine clock model is a popular alternative to the clock offset model for modeling the clock time of the slave node, $c_{slave,aff}(t)$, in terms of the clock time of the master node, t. In this model, we have

$$c_{slave,aff}(t) = \phi t + \delta, \tag{2.2}$$

where ϕ denotes the relative clock skew (the frequency offset is $(\phi - 1)$) and δ denotes the clock offset of the slave's clock time with respect to the master's clock time [7, 18, 21–23, 25, 48, 57]. Etzlinger *et al.* [12, 13] showed that the affine clock model is significantly more accurate than the simple clock offset model, where only the presence of a clock offset is assumed between the master and slave node for cases where frequency synchronization is not already maintained.

If the clocks at the slave and master node are synchronized, then $c_{slave}(t) = t$, where $c_{slave}(t) = c_{slave,offset}(t)$ for the clock offset model and $c_{slave}(t) = c_{slave,aff}(t)$ for the affine clock model. However, in practice, these clocks are not synchronized, implying a synchronization error $e(t) = |c_{slave}(t) - t|$ that can grow over large timescales. Also, over large time scales, the clock parameters slowly change due to various reasons, including temperature, atmospheric pressure, voltage changes, and aging hardware [71, 74]. Hence, even if the clock of the slave node is perfectly tuned initially, the clock can drift away from the ideal time due to the inherent imperfections in a clock oscillator [74], which in turn means the relative clock offset and skew might change with time. Hence, it is necessary for devices in the network to perform periodic clock synchronization to adjust the clock parameters to ensure that the timing error does not become significantly large.

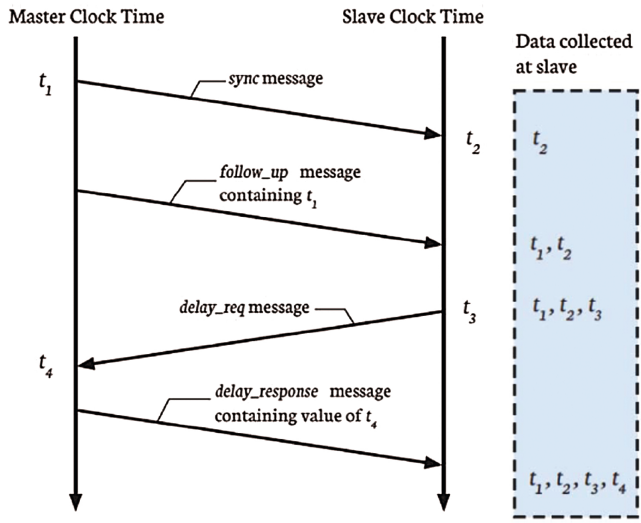


Figure 2.1: Two-way synchronization between a master and slave.

2.2 Synchronization via Two-Way Message Exchanges

Two-way message exchange is a classical mechanism for synchronizing two devices in a packet-switched network. Examples of clock synchronization protocols that employ this approach include TPSN [20], LTS [70], tiny-sync and mini-sync [62], and PTP [32]. Let us denote the relative clock skew and offset of the slave node relative to its master clock by ϕ and δ, respectively. We now describe the underlying two-way message exchange mechanism in PTP. During the two-way message exchange in PTP (see Figure 2.1), the following series of packet exchanges are performed between the master and slave to determine ϕ and δ:

1. The master initiates the 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.

2. The slave records the time of reception of the *SYNC* message as t_2. The relationship between t_1 and t_2 is given by

$$t_2 = \phi(t_1 + d_{ms}) + \delta, \tag{2.3}$$

where d_{ms} is the end-to-end (ETE) network delay between the master and the slave.

3. The slave sends a *DELAY_REQ* message to the master, recording the time of transmission as t_3.

4. The master records the time of arrival of the *DELAY_REQ* packet as t_4. The relationship between t_3 and t_4 is given by

$$t_3 = \phi(t_4 - d_{sm}) + \delta, \tag{2.4}$$

where d_{sm} is the ETE delay between the slave and the master. The value of t_4 is sent to the slave using a *DELAY_RESP* packet.

Thus, four timestamps $\{t_1, t_2, t_3, t_4\}$ are available to the slave at the end of each two-way message exchange.

2.3 Packet Delay Models

The packets traveling between the master and the slave hop across several intermediate nodes (switches or routers), that are typically part of a larger network that is shared among multiple users (see Figure 2.2 for an example).

Hence each intermediate switch concurrently services background traffic generated by other network users, in addition to synchronization traffic. The ETE delay for a timing packet is the sum of a fixed deterministic delay component and stochastic delay component. The fixed

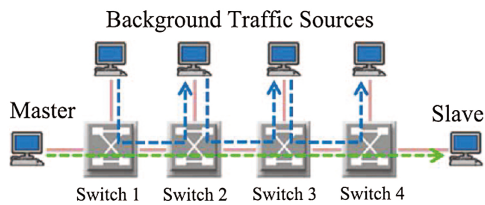


Figure 2.2: Example of a four switch network with cross and traffic flows. Red lines indicate network links, blue lines indicate the direction of background traffic flows, and green line represents the direction of synchronization traffic flows from the master to slave.

delay component corresponds to constant propagation and processing delays, while the stochastic delays occur due to contention for service with background traffic. Hence, the ETE delays can be modeled as

$$d_{ms} = d_1 + w_1; \tag{2.5}$$

$$d_{sm} = d_2 + w_2, \tag{2.6}$$

where d_1 and d_2 denote the deterministic delays in the forward master-to-slave path and the reverse slave-to-master path, respectively, while w_1 and w_2 represent the stochastic queuing delays. The statistical distribution of the stochastic delays, w_1 and w_2, depend on various network parameters (e.g., network status and traffic); therefore, no single delay model can be found to fit for every case [74]. Some popular pdf models for modeling the stochastic queuing delays are described next.

2.3.1 Gaussian Delay Model for Stochastic Delays

The Gaussian delay model assumes that the stochastic queuing delays follow a Gaussian distribution with zero mean, i.e.,

$$f_{w_k}(w) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{w^2}{2\sigma_k^2}\right), \tag{2.7}$$

where σ_k^2 denotes the variance of the Gaussian random variable for $k = 1, 2$. If the exact distribution of the stochastic queuing delays is unknown, the Gaussian pdf delay model is sometimes justified by the Central Limit Theorem (CLT), which states that the pdf of the sum of a large number of independent and identically distributed (i.i.d) random variables asymptotically approaches that of a Gaussian random variable [74]. The Gaussian pdf model for the stochastic queuing delays has been considered in [10, 48, 57].

2.3.2 Exponential Pdf Delay Model for Stochastic Delays

Another popular pdf model for stochastic delays is the exponential delay model. This model assumes that the stochastic queuing delays follow

an exponential distribution, i.e.,

$$f_{w_k}(w) = \frac{1}{\lambda_k} \exp\left(-\frac{w}{\lambda_k}\right), \tag{2.8}$$

for $k = 1, 2$. In (2.8), λ_k denotes the rate parameter of the exponential distribution. The exponential random variable is suitable for modeling the delay for point-to-point connections in an M/M/1 queue [1]. The exponential pdf delay model for the stochastic queuing delays has been considered in [7, 8, 11, 34, 49].

2.3.3 General Pdf Delay Model for Stochastic Delays

Although the exponential and Gaussian pdf delay models apply to several scenarios, they might not be suitable in certain specific scenarios. For instance, consider the case of the application of PTP to telecommunication networks. Here, as opposed to the Gaussian and exponential random variables, the stochastic delays w_1 and w_2 are non-negative random variables with a finite maximum value. The finite maximum value assumption is reasonable since synchronization packet is typically assigned higher priority than packets of background traffic, hence the worst-case queuing delay is bounded for a finite number of switches between the master and the slave [21, 23]. The empirical pdf (and the corresponding cumulative density function (cdf)) of the stochastic delays for a reasonable model of this scenario were obtained in [21–23]. To obtain these pdfs, a gigabit Ethernet network consisting of a cascade of 10 switches between the master and slave nodes was considered. Each switch is assumed to be a store-and-forward switch that implements strict priority queuing. With regard to the distribution of packet sizes in background traffic, Traffic Model 1 (TM1) and Traffic Model 2 (TM2) from the ITU-T recommendation $G.8261$ [67], as specified in Table 2.1, were employed.

The inter-arrival times between packets in all background traffic flow are assumed to follow an exponential distribution. The percentage of the link capacity consumed by background traffic is referred to as the load. The rate parameter of each exponential distribution can be set accordingly in order to achieve a particular load (see [21–23]). The empirical pdf (and the corresponding cdf) for TM1 and TM2 under

Table 2.1: Models for composition of background traffic packets

Traffic Model	Packet Sizes (in Bytes)	% of Total Load
TM1	{64, 576, 1518}	{80%, 5%, 15%}
TM2	{64, 576, 1518}	{30%, 10%, 60%}

various loads are shown in Figures 2.3 and 2.4, respectively. As seen from Figures 2.3 and 2.4, the Gaussian or the exponential delay models do not seem suitable for all network scenarios (especially at low or

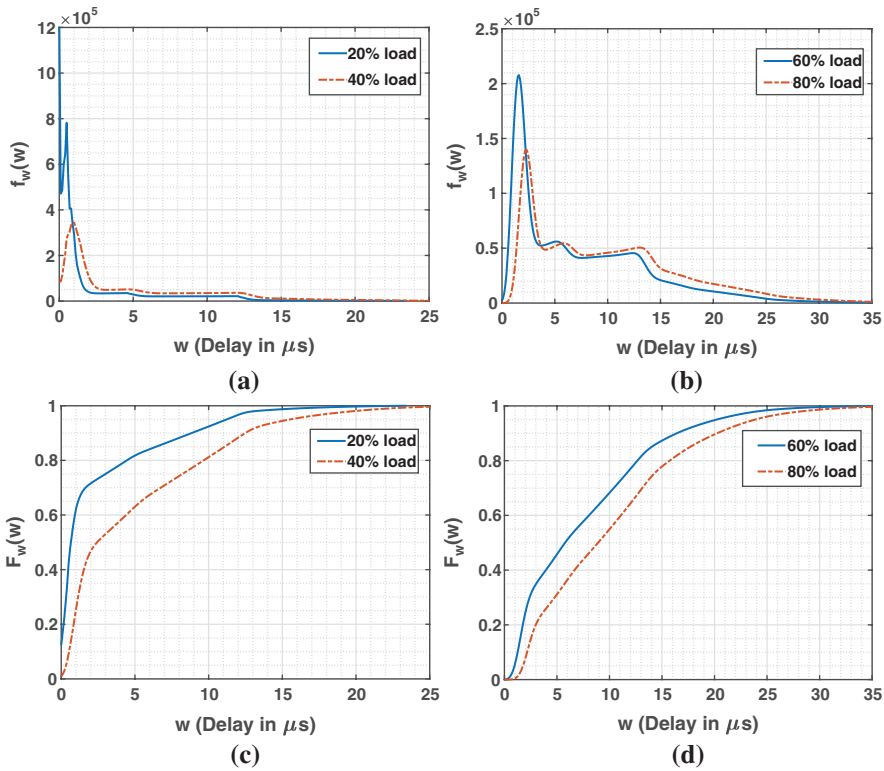


Figure 2.3: Empirical pdf and cdf of queuing delays under various loads for TM1 assuming 10 switches between the master and slave node. (a) pdf under loads 20% and 40%. (b) pdf under loads 60% and 80%. (c) cdf under loads 20% and 40%. (d) cdf under loads 60% and 80%.

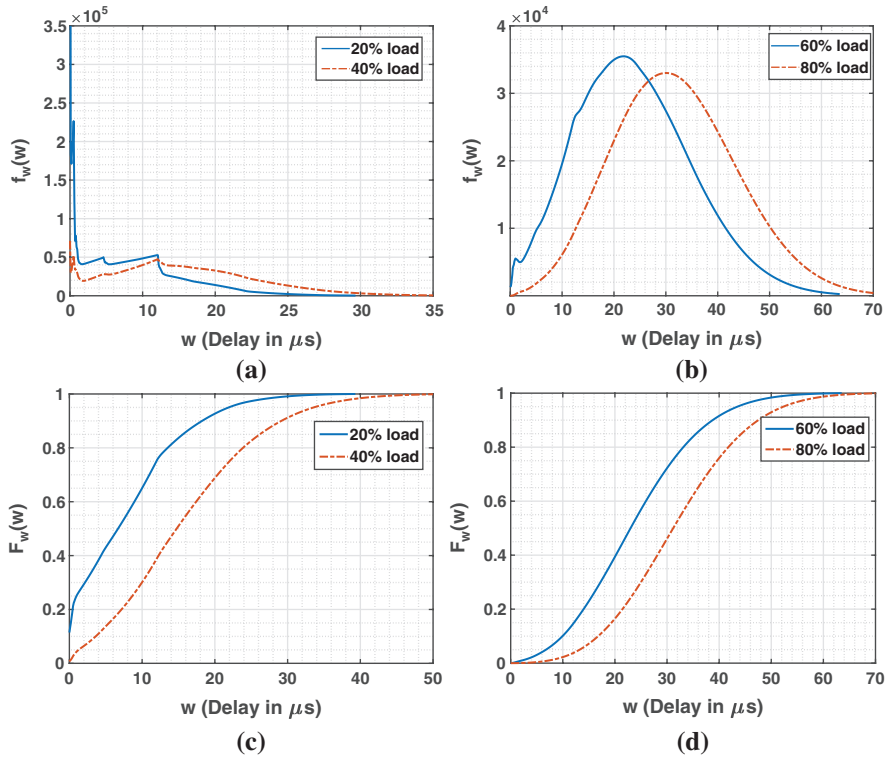


Figure 2.4: Empirical pdf and cdf of queuing delays under various loads for TM2 assuming 10 switches between the master and slave node. (a) pdf under loads 20% and 40%. (b) pdf under loads 60% and 80%. (c) cdf under loads 20% and 40%. (d) cdf under loads 60% and 80%.

medium loads). In the following sections, we present recently developed optimum clock skew and offset estimators for IEEE 1588 PTP [32].[1]

[1]The developed optimum estimators are optimum for any clock synchronization protocol built on the two-way message exchange scheme such as the TPSN [20] or LTS [70].

3

Statistical Preliminaries

The purpose of this section is to formalize the concept of invariance by defining groups of transformations over parameter and observation spaces. To this end, we repeat several essential definitions from [4, 47] to establish some concepts of invariant estimation theory. It is assumed throughout this section that the observed data $\boldsymbol{x} \in \mathbb{R}^N$ is characterized by the pdf $f(\boldsymbol{x} \mid \boldsymbol{\theta})$, which depends upon the vector of unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^M$ with the corresponding parameter space $\boldsymbol{\Theta}$.

3.1 Performance of an Estimator

Let us suppose that we are interested in estimating an unknown scalar parameter $\theta = \boldsymbol{c}^T\boldsymbol{\theta}$, where $\boldsymbol{c} \in \mathbb{R}^M$ is a constant vector. Let ψ denote an estimator of θ, $\psi(\boldsymbol{x})$ denote the estimate of θ 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 the following [47]:

1. The conditional risk of an estimator

$$\mathcal{R}(\psi, \boldsymbol{\theta}) = \int_{\mathbb{R}^N} L(\psi(\boldsymbol{x}), \boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta}) d\boldsymbol{x}. \tag{3.1}$$

2. The maximum risk of an estimator

$$\mathcal{M}(\psi) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{R}(\psi, \boldsymbol{\theta}). \tag{3.2}$$

3. The average risk of an estimator

$$\mathcal{B}(\psi, p) = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{R}(\psi, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{3.3}$$

where $p(\boldsymbol{\theta})$ is a prior distribution defined over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

3.2 Groups of Transformations

We now present some important definitions from [4] with regards to invariant estimation theory. A measurable function f: $\mathbb{R}^N \to \mathbb{R}^N$ is called a transformation on $\mathbb{R}^N$. If g_1 and g_2 are two transformations on $\mathbb{R}^N$, the composition of g_1 and g_2, denoted by $g_2 g_1$, is defined as $g_2 g_1(\boldsymbol{m}) = g_2(g_1(\boldsymbol{m}))$ for $\boldsymbol{m} \in \mathbb{R}^N$. We are now ready to define a group of transformations.

Definition 1 (Section 6.2.1, [4]). A group of transformations on $\mathbb{R}^N$, denoted by $\mathcal{G}$, is a set of bijective transformations which satisfy the following conditions:

- If $g_1 \in \mathcal{G}$ and $g_2 \in \mathcal{G}$, then $g_2 g_1 \in \mathcal{G}$.
- If $g \in \mathcal{G}$, then g^{-1}, the inverse transformation defined by the relation $g^{-1}(g(\boldsymbol{x})) = \boldsymbol{x}$, is in $\mathcal{G}$.
- The identity transformation e, defined by $e(\boldsymbol{x}) = \boldsymbol{x}$, is in $\mathcal{G}$.

Example 3.1. Let $x \in \mathbb{R}$ and consider the set of transformations $\mathcal{G} = \{g_c\colon c > 0\}$, where $g_c(x) = cx$. It is easy to see that the functions g_c are bijective. Also, note that

- $g_{c_2} g_{c_1}(x) = c_2 c_1 x = g_{c_2 c_1} x$. So $g_{c_2} g_{c_1} \in \mathcal{G}$.
- The identity element is $e = g_1$.
- For every $g_c \in \mathcal{G}$, the inverse of the element is $g_{c^{-1}} \in \mathcal{G}$.

So $\mathcal{G}$ is indeed a group of transformations.

3.3 Invariant Decision Problems

Let $\mathcal{F}$ denote the family of all pdfs $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\mathcal{G}$ denote a group of transformations on $\mathbb{R}^N$.

Definition 2 (Section 6.2.2, [4])**.** The family of pdfs $\mathcal{F}$ is said to be invariant under $\mathcal{G}$, if for every $g \in \mathcal{G}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, there exists a unique $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ such that $\boldsymbol{x}_g = g(\boldsymbol{x})$ has a pdf $f(\boldsymbol{x}_g \mid \boldsymbol{\theta}^*)$. For a given transformation g on the observations, the parameters are transformed from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$. We denote $\boldsymbol{\theta}^*$ as $\bar{g}(\boldsymbol{\theta})$. For a given g, the transformation $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ is a transformation of $\boldsymbol{\Theta}$ into itself.

Remark 3.1. If the family of pdfs $\mathcal{F}$ is invariant under $\mathcal{G}$, then

$$\bar{\mathcal{G}} = \{\bar{g}\colon\, g \in \mathcal{G}\} \tag{3.4}$$

is a group of transformations on $\boldsymbol{\Theta}$ [4], since the group of transformations $\bar{\mathcal{G}}$ defined in (3.4) satisfies the conditions listed in Definition 1. We now present a simple example to illustrate these ideas.

Example 3.2. Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$ and $h(\cdot)$ be a known pdf. Consider the family of pdfs of the form

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \frac{1}{\sigma^N} h\left(\frac{\boldsymbol{x} - \mu \mathbf{1}_N^T}{\sigma}\right), \tag{3.5}$$

where $\mu \in \mathbb{R}$ (location parameter) and $\sigma \in \mathbb{R}^+$ (scale parameter) are both unknown. The family of such pdfs is invariant under the group of location-scale transformations $\mathcal{G}_{LS}$, on $\mathbb{R}^N$, defined as

$$\mathcal{G}_{LS} = \{g_{a,b}(\boldsymbol{m})\colon\, g_{a,b}(\boldsymbol{m}) = a\boldsymbol{m} + b\mathbf{1}_N\} \tag{3.6}$$

where $a \in \mathbb{R}^+, b \in \mathbb{R}$ and $\boldsymbol{m} \in \mathbb{R}^N$, since $\boldsymbol{x}_g = g_{a,b}(\boldsymbol{x}) = [x_{g1}, x_{g2}, \ldots, x_{gN}]$ has the pdf $(a\sigma)^{-N} h(\frac{\boldsymbol{x}_g - (a\mu+b)\mathbf{1}_N^T}{a\sigma})$.

The corresponding group, $\bar{\mathcal{G}}_{LS}$, of induced transformations on $\boldsymbol{\Theta} = \{(\mu, \sigma)\colon\, \mu \in \mathbb{R} \text{ and } \sigma \in \mathbb{R}^+\}$, is given by

$$\bar{\mathcal{G}}_{LS} = \{\bar{g}_{a,b}(\mu, \sigma)\colon\, \bar{g}_{a,b}(\mu, \sigma) = (a\mu + b, a\sigma)\}. \tag{3.7}$$

For an invariant estimation problem, we need to have a loss function that is invariant to the group of transformations. We now present the definition of an invariant loss function.

Definition 3 (Section 6.2.2, [4]). Let the family of pdfs $\mathcal{F}$ be invariant under the group $\mathcal{G}$. Let ψ denote an estimator of $\theta \in \boldsymbol{\theta}$, and $\psi(\boldsymbol{x})$ denote the estimate of θ obtained from the received observations $\boldsymbol{x}$ characterized by the pdf $f(\boldsymbol{x} \mid \boldsymbol{\theta})$. A loss function $L(\psi(\boldsymbol{x}), \boldsymbol{\theta})$ is said to be invariant under $\mathcal{G}$, if for every $g \in \mathcal{G}$, there exists a transformation $\tilde{g}$ on $\psi(\boldsymbol{x})$ such that $L(\psi(\boldsymbol{x}), \boldsymbol{\theta}) = L(\tilde{g}(\psi(\boldsymbol{x})), \bar{g}(\boldsymbol{\theta}))$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The formal structures of the statistical distributions of $\boldsymbol{x}$ and $g(\boldsymbol{x})$ must be identical in an invariant estimation problem. Hence the invariance principle states that the estimates obtained from $\boldsymbol{x}$ and $g(\boldsymbol{x})$, using an estimator must be related [4]. We are now ready to present the definition of an invariant estimator.

Definition 4 (Section 6.2.3, [4]). Let ψ denote an estimator of $\theta \in \boldsymbol{\theta}$, and $\psi(\boldsymbol{x})$ denote the estimate of θ obtained from the received observations $\boldsymbol{x}$ characterized by the pdf $f(\boldsymbol{x} \mid \boldsymbol{\theta})$. We say ψ is invariant under the group $\mathcal{G}$ if for all $\boldsymbol{x} \in \mathbb{R}^N$ and $g \in \mathcal{G}$,

$$\psi(g(\boldsymbol{x})) = \tilde{g}(\psi(\boldsymbol{x})), \tag{3.8}$$

where $\tilde{g} \in \bar{\mathcal{G}}$ defined in (3.4).

Example (Example 1 Continued). Let $\hat{\mu}$ and $\hat{\sigma}$ denote estimators of μ and σ, respectively and let $\hat{\mu}(\boldsymbol{x})$ and $\hat{\sigma}(\boldsymbol{x})$ denote the estimates obtained from $\boldsymbol{x}$. The estimators $\hat{\mu}$ and $\hat{\sigma}$ are invariant under $\mathcal{G}_{LS}$ defined in (3.6), if

$$\hat{\mu}(g_{a,b}(\boldsymbol{x})) = a\hat{\mu}(\boldsymbol{x}) + b, \quad \text{and} \quad \hat{\sigma}(g_{a,b}(\boldsymbol{x})) = a\hat{\sigma}(\boldsymbol{x}) \tag{3.9}$$

for all $g_{a,b} \in \mathcal{G}_{LS}$. Suppose the loss functions for μ and σ are defined as

$$L_\mu(\hat{\mu}(\boldsymbol{x}), [\mu, \sigma]) = \frac{(\mu - \hat{\mu}(\boldsymbol{x}))^2}{\sigma^2}, \tag{3.10}$$

and

$$L_\sigma(\hat{\sigma}(\boldsymbol{x}), [\mu, \sigma]) = \frac{(\sigma - \hat{\sigma}(\boldsymbol{x}))^2}{\sigma^2}, \tag{3.11}$$

respectively. The loss functions are invariant under $\mathcal{G}_{LS}$ from (3.6), since

$$\frac{(\mu - \hat{\mu}(\boldsymbol{x}))^2}{\sigma^2} = \frac{((a\mu + b) - \hat{\mu}(g_{a,b}(\boldsymbol{x})))^2}{a^2\sigma^2} \tag{3.12}$$

and

$$\frac{(\sigma - \hat{\sigma}(\boldsymbol{x}))^2}{\sigma^2} = \frac{(a\sigma - \hat{\sigma}(g_{a,b}(\boldsymbol{x})))^2}{a^2\sigma^2}, \tag{3.13}$$

for all $g_{a,b} \in \mathcal{G}_{LS}$ from (3.6). The loss functions given in (3.10) are called the scale-normalized squared error loss. Also note that, any loss functions for μ and σ, of the form

$$L_\mu(\hat{\mu}(\boldsymbol{x}), [\mu, \sigma]) = \frac{|\mu - \hat{\mu}(\boldsymbol{x})|^k}{\sigma^k}, \tag{3.14}$$

and

$$L_\sigma(\hat{\sigma}(\boldsymbol{x}), [\mu, \sigma]) = \frac{|\sigma - \hat{\sigma}(\boldsymbol{x})|^k}{\sigma^k}, \tag{3.15}$$

respectively, are invariant under $\mathcal{G}_{LS}$ from (3.6), for $k = 1, 2, \ldots$.

We now present an important definition regarding the transitivity of the group of transformations on $\boldsymbol{\Theta}$ and the conditional risk of invariant estimators.

Definition 5 (Section 6.2.3, [4]). A group $\bar{\mathcal{G}}$ of transformations of $\boldsymbol{\Theta}$ is said to be transitive if for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, there exists some $\bar{g} \in \bar{\mathcal{G}}$ for which $\boldsymbol{\theta}_2 = \bar{g}(\boldsymbol{\theta}_1)$.

Example (Example 1 Continued). Let $\boldsymbol{\theta}_1 = (\mu_1, \sigma_1)$ and $\boldsymbol{\theta}_2 = (\mu_2, \sigma_2)$. Note that $\bar{g}_{\frac{\sigma_2}{\sigma_1}, \left(\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1\right)}(\mu_1, \sigma_1) = (\mu_2, \sigma_2)$, where $\bar{g}_{\frac{\sigma_2}{\sigma_1}, \left(\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1\right)} \in \bar{\mathcal{G}}_{LS}$ defined in (3.7). So, $\bar{\mathcal{G}}_{LS}$ is transitive.

While the conditional, maximum and average risk can be different for a estimator, for an invariant estimator and a transitive group, they are always equal, as stated in the following theorem.

Theorem 3.1 (Section 6.2.3, [4]). When $\bar{\mathcal{G}}$ is transitive and the loss function is invariant under $\mathcal{G}$, the conditional risk of an invariant estimator ψ of $\theta \in \boldsymbol{\theta}$, is constant for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Further, we have

$$\mathcal{R}(\psi, \boldsymbol{\theta}) = \mathcal{M}(\psi) = \mathcal{B}(\psi, p), \tag{3.16}$$

for any $p(\boldsymbol{\theta})$ defined over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Remark 3.2. When $\bar{\mathcal{G}}$ is transitive, we can construct the optimum (or minimum conditional risk) invariant estimator under $\mathcal{G}$, when the loss function is invariant under $\mathcal{G}$ using the theory from [4, 47].

3.4 Invariance and Minimax Optimality

In our work, we are also interested in developing estimators that are minimax optimum, that is estimators that minimize the maximum risk over all possible estimators of the parameter of interest. We first present the definition of a minimax estimator from [4, 47].

Definition 6 (Minimax Estimators)**.** An estimator $\boldsymbol{\psi}_{MinMax}$ of $\theta \in \boldsymbol{\theta}$ is said to be a minimax estimator of θ for the considered loss function, if

$$\mathcal{M}(\psi_{MinMax}) = \inf_{\psi} \mathcal{M}(\psi) = \inf_{\psi} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{R}(\psi, \boldsymbol{\theta}). \tag{3.17}$$

Remark 3.3. The class of invariant decision rules is often much smaller than the class of all possible decision rules. However, it frequently turns out that the optimum invariant decision rule is minimax optimum [4]. We use the standard approach given in [4, 47, 69] to design a minimax estimator of θ. We first construct the optimum invariant estimator of θ for a considered (invariant) loss function and then show the optimum invariant estimator is a minimax estimator of θ for the considered loss function.

4

Optimum Signal Processing Techniques for Clock Offset Estimation

Packet-based clock synchronization using PTP is increasingly being considered as an alternative to GPS-based timing to provide microsecond-level synchronization between base stations in 4G LTE mobile networks [24, 25, 55, 56, 73]. This section discusses the problem of clock offset estimation in PTP under the assumption that the clock skew is known. A practical scenario where such an assumption can be made occurs when Synchronous Ethernet (SyncE) is used in conjunction with PTP for clock synchronization. SyncE [67] is a physical layer based technology to deliver frequency synchronization (or clock skew synchronization) to packet-based Ethernet networks. SyncE is independent of the network load and supports multi-hop frequency transfer provided that all the intermediate switches and routers support SyncE. In the considered scenario of synchronization between base stations in 4G LTE mobile networks, the slave node obtains the clock skew information from the physical layer signals of SyncE, while the PTP messages are used for clock offset synchronization [21, 23].

In PTP (or any clock synchronization based on the two-way message exchange), the timing messages traveling between the master and the

slave can encounter several intermediate switches and routers, accumulating stochastic queuing delays at each such node. This randomness in the overall network traversal times is referred to as the packet delay variation (PDV) [21–23] and the problem of estimating the clock offset of the slave clock while combating the randomness in the observations that occurs due to PDV is referred to as the clock offset estimation (COE) problem.

Given the nature of the observations, the COE problem falls under a class of estimation problems referred to as the location parameter estimation problems [4, 47]. In these problems, the unknown parameters influence the observations by translating the pdf of the stochastic queuing delays, without affecting the shape of the pdf. Location parameter estimation problems occur in a wide range of practical applications; some examples include regression analysis [9] and the estimation of user position in global positioning system (GPS) receivers [5].

In this section, we describe recently developed optimum estimators from [21] for the COE problem. These estimators are optimum in terms of minimizing the mean square estimation error among the class of invariant clock offset estimators for PTP. Further, these estimators are minimax optimum, that is, they minimize the maximum mean squared error over all possible values of the unknown parameters. Additionally, we describe recently proposed L-estimators for the COE problem [22]. In the L-estimators, the estimate of the clock offset is obtained as a linear combination of the order statistics of the observations.[1] The linear L-estimators proposed in [22] exhibit performance close to the optimum estimators for many network scenarios.

4.1 Related Work

Several algorithms that address the COE problem are available in the scientific literature [7, 8, 34, 35, 42–44, 48, 49, 57]. These algorithms apply to a large number of pdf-models of PDV and address the COE problem for various scenarios. Additionally, these clock offset estimation algorithms are optimum for certain pdf-delay models of the PDV such as

[1] The order statistics of a set of observations refers simply to the same set of observations rearranged in non-decreasing order.

for the Gaussian pdf-model [57] and the exponential pdf-delay model [8]. While these estimators [7, 8, 34, 48, 49, 57] are quite useful, they may not be optimum in certain niche cases where the pdf-model of the PDV may not follow a Gaussian or exponential distribution. One such case is the LTE backhaul network scenario (see Figures 2.3 and 2.4 for the empirical pdf-model of the PDV for this scenario). This issue was addressed in [21], where the authors develop the optimum clock offset estimators for the COE problem that are optimum for any pdf-model of the PDV.

We now briefly describe some of the popular algorithms available in the literature for the COE problem [7, 8, 34, 35, 42–44, 48, 49, 57]. Depending on whether the fixed path delays are known or unknown, the maximum likelihood (ML) estimator of the clock skew and offset under the Gaussian PDV pdf delay model was derived in [57] and [48], respectively. Under the exponential PDV pdf-delay model, the ML estimate of the clock offset was derived in [34] and the joint ML-estimate of the clock skew and offset depending on whether the mean of the exponential distribution is known or unknown was derived in [7] and [49], respectively. Further, the minimum variance unbiased estimate of the clock offset under the exponential PDV pdf-delay model for the COE problem was derived in [8]. Although these algorithms [7, 8, 34, 48, 49, 57] exhibit good performance under the assumed PDV pdf-delay model, the performance of these approaches could significantly degrade if the actual pdf of PDV does not fit the assumed families very well.

In some scenarios, prior information regarding the pdf-model of the PDV might not be readily available. Several publications in the literature have addressed the COE problem for this particular scenario [35, 42–44]. In [35], the bootstrap estimate of the clock offset for the COE problem was derived, while in [42, 43], the authors approximated the unknown pdf of the PDV by a Gaussian mixture model and proposed clock offset estimators that are robust against network PDV pdf-delay model mismatches. In [44], the authors adopted a Bayesian framework and proposes a novel clock synchronization algorithm that combines the Gaussian mixture Kalman particle filter discussed in [42] with an iterative noise density estimation procedure to achieve robust performance in the presence of unknown network delay distributions.

Additionally, survey papers [72, 74] are available in the literature that discuss the current approaches available in the literature for the COE problem.

The PTP standard [32] and related literature prescribe the use of simple non-parametric COE schemes such as the sample mean, minimum, and maximum filtering schemes. Several recent papers [2, 24, 25, 55, 56, 73] have studied methods to improve the performance of the sample mean, minimum, and maximum filtering schemes, especially in the presence of substantial queuing delays due to high network loads. However, it was not known as to how close these COE schemes come to achieving the best possible estimation performance, measured in terms of the mean squared estimation error. Guruswamy *et al.* [21] addressed this issue by developing optimum clock offset estimators for an arbitrary pdf-model of the PDV.

4.2 Signal Model and Problem Statement

Recall from Section 2 that the slave node obtains four timestamps $\{t_1, t_2, t_3, t_4\}$ after a round of two-way message exchange. The relationship between the timestamps is given by

$$t_2 = \phi(t_1 + d_1 + w_1) + \delta, \tag{4.1}$$

$$t_3 = \phi(t_4 - d_2 - w_2) + \delta, \tag{4.2}$$

where ϕ and δ denote the clock skew and offset of the slave node relative to the master node, respectively. The variables d_1 and d_2 denote the deterministic fixed delays in the forward master-to-slave path and the reverse slave-to-master path, respectively.

In this section, we consider the case where the parameter ϕ is known. Hence, the clock offset δ is the only unknown parameter of interest. It is clearly sufficient to retain the pair of timestamp differences to estimate δ. We have

$$y_1 = t_2 - t_1 = d_1 + \delta + w_1; \tag{4.3}$$

$$y_2 = t_4 - t_3 = d_2 - \delta + w_2. \tag{4.4}$$

Typically, a series of timing message transmissions is required to estimate the relative clock skew and offset between the master and slave

nodes [74]. Assuming the parameters δ, d_1 and d_2 remain constant over the duration of P two-way message exchanges, we can collect multiple observation pairs (y_1, y_2) to help estimate δ. These observations can be denoted as

$$y_{1i}^* = d_1 + \delta + w_{1i}; \tag{4.5}$$
$$y_{2i}^* = d_2 - \delta + w_{2i}, \tag{4.6}$$

for $i = 1, 2, \ldots, P$. The accuracy with which we can estimate δ from the observations in (4.5) and (4.6) depends on the amount of knowledge we have about d_1 and d_2. To this end, two observation models with varying degrees of information available about d_1 and d_2 are considered as follows.

1. *Known fixed delay model (K-model):* In this model, the fixed delays d_1 and d_2 are assumed to be known at the slave. Hence, setting

$$y_{ik} = y_{ik}^* - d_k \tag{4.7}$$

for $k = 1, 2$ and $i = 1, 2, \ldots, P$, the delay compensated observations can be written as

$$y_{1i} = \delta + w_{1i}; \tag{4.8}$$
$$y_{2i} = -\delta + w_{2i}, \tag{4.9}$$

for $i = 1, 2, \ldots, P$. These observations can be collected in vector form as

$$\boldsymbol{y} = \delta \boldsymbol{e} + \boldsymbol{w}, \tag{4.10}$$

where

$$\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2]^T, \quad \boldsymbol{y}_k = [y_{k1}, y_{k2}, \ldots, y_{kP}] \text{ for } k = 1, 2, \tag{4.11}$$
$$\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2]^T, \quad \boldsymbol{w}_k = [w_{k1}, w_{k2}, \ldots, w_{kP}] \text{ for } k = 1, 2, \tag{4.12}$$
$$\boldsymbol{e} = [\mathbf{1}_P^T, -\mathbf{1}_P^T]^T, \tag{4.13}$$

and $\mathbf{1}_P$ is a $P \times 1$ vector with all elements equal to 1.

2. *Standard delay model (S-model):* Freris *et al.* [16] provided the necessary conditions for obtaining a unique solution to the clock

offset for synchronization protocols based on the two-way message exchange scheme. We need to know either one of the fixed delays (either the forward- or reverse fixed path delay) or have a prior known affine relationship between the fixed delays. For simplicity, the fixed delays in the forward and reverse path are assumed to be equal but unknown, i.e., $d_1 = d_2 = d$, where d is an unknown parameter. The observations can be written as

$$y_{1i} = \delta + d + w_{1i}; \tag{4.14}$$

$$y_{2i} = d - \delta + w_{2i}, \tag{4.15}$$

for $i = 1, 2, \ldots, P$. These observations can be collected in vector form as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{w}, \tag{4.16}$$

where $\boldsymbol{y}$ and $\boldsymbol{w}$ are defined in (4.11) and (4.12), respectively, $\boldsymbol{\theta} = [\delta, d]^T$, and

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1}_P & \mathbf{1}_P \\ -\mathbf{1}_P & \mathbf{1}_P \end{bmatrix}. \tag{4.17}$$

Given either observation models, the COE problem is to estimate δ from $\boldsymbol{y}$. Next, we state the assumptions used in developing the optimum COE scheme [23].

COE Assumptions:

1. All the queuing delays are strictly positive random variables that are mutually independent.
2. All forward queuing delays share a common pdf $f_1(\cdot)$. Similarly the reverse queuing delays share a common pdf $f_2(\cdot)$. Also, when developing the optimum estimators under the *K-model* and *S-model*, we assume $f_1(\cdot)$ and $f_2(\cdot)$ are assumed to be completely known.
3. The maximum possible value for a forward or reverse queuing delay is finite.
4. The unknown parameters are assumed to be constant over P two-way message exchanges.

4.3 Optimum Clock Offset Estimators

The COE problem under the considered observation models falls under the class of the vector location parameter estimation problems [21]. The optimum invariant clock offset estimators for the considered observation models were derived in [21] and are presented in the following subsections.

4.3.1 Optimum Invariant Estimator Under K-Model

Recall from (4.10), the observations under the *K-model* can be represented as

$$\boldsymbol{y} = \delta \boldsymbol{e} + \boldsymbol{w}, \tag{4.18}$$

where $\delta \in \mathbb{R}$ is the unknown parameter. From the COE assumptions, we now have

$$f(\boldsymbol{y} \mid \delta) = f_{\boldsymbol{w}}(\boldsymbol{y} - \delta \boldsymbol{e}), \tag{4.19}$$

$$= \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - (-1)^k \delta). \tag{4.20}$$

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

$$\mathcal{G}_{KModel} = \{g_a(\boldsymbol{m})\colon\, g_a(\boldsymbol{m}) = \boldsymbol{m} + a\boldsymbol{e}, \forall a \in \mathbb{R}\}, \tag{4.21}$$

where $\boldsymbol{m} \in \mathbb{R}^{2P}$, since $\boldsymbol{y}_g = g_a(\boldsymbol{y})$ has the pdf $f_{\boldsymbol{w}}(\boldsymbol{y}_g - (a+\delta)\boldsymbol{e})$ which has the location parameter $(a+\delta)$ as opposed to the parameters δ for $f(\boldsymbol{y} \mid \delta)$. The corresponding group of induced transformations, $\bar{\mathcal{G}}_{KModel}$, is given by

$$\bar{\mathcal{G}}_{KModel} = \{\bar{g}_a(\delta)\colon\, \bar{g}_a(\delta) = (a+\delta), \forall a \in \mathbb{R}\}, \tag{4.22}$$

where $\delta \in \mathbb{R}$.

Let $\hat{\delta}_I$ denote an estimator of δ, and let $\hat{\delta}_I(\boldsymbol{y})$ denote the estimates obtained from the received data $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y} \mid \delta) = f_{\boldsymbol{w}}(\boldsymbol{y} - \delta \boldsymbol{e})$. The estimator $\hat{\delta}_I$ is invariant under $\mathcal{G}_{KModel}$ from

(4.21) if for all $a \in \mathbb{R}$

$$\hat{\delta}_I(g_a(\boldsymbol{y})) = \hat{\delta}_I(\boldsymbol{y} + a\boldsymbol{e}) = \hat{\delta}_I(\boldsymbol{y}) + a. \tag{4.23}$$

When developing the optimum clock offset estimators, the loss function for δ is chosen as the squared error loss function [21]. The main reason for choosing the squared error loss is that the corresponding conditional risk is the mean square estimation error (MSE), which is a popular metric for evaluating the performance of an estimator. The squared error loss function is invariant under $\mathcal{G}_{KModel}$ from (4.21), since

$$(\hat{\delta}_I(\boldsymbol{y}) - \delta)^2 = (\hat{\delta}_I(g_a(\boldsymbol{y})) - (a + \delta))^2, \tag{4.24}$$

for all $g_a \in \mathcal{G}_{KModel}$.

In Theorem 4.1, we develop the optimum invariant clock offset estimator for the *K-model* so we can calculate the best performance for cases when $f_1(\cdot)$ and $f_2(\cdot)$ are known.[2] Later, we use these results to evaluate the losses from using clock offset estimation schemes designed without knowledge of $f_1(\cdot)$ and $f_2(\cdot)$.

Theorem 4.1. Assuming complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$, the optimum (or minimum conditional risk) invariant estimators of δ, denoted by $\hat{\delta}_{opt}$, under $\mathcal{G}_{KModel}$ defined in (4.21), for the squared error loss function is given by

$$\hat{\delta}_{opt}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}} \delta \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - (-1)^k \delta) d\delta}{\int_{\mathbb{R}} \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - (-1)^k \delta) d\delta}. \tag{4.25}$$

Further, the derived optimum invariant estimators are minimax optimum under the squared error loss.

Proof. The optimum invariant estimator of δ under $\mathcal{G}_{KModel}$ from (4.21), denoted by $\hat{\delta}_{opt}$, can be obtained by solving (See [4], Sec. 6.6.2, Result 3)

$$\hat{\delta}_{opt}(\boldsymbol{y}) = \arg\min_{\hat{\delta}} \int_{\mathbb{R}} (\hat{\delta}(\boldsymbol{y}) - \delta)^2 \pi^r(\delta \mid \boldsymbol{y}) d\delta, \tag{4.26}$$

where we have

$$\pi^r(\delta \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \delta) \pi^r(\delta)}{\int_{\mathbb{R}} f(\boldsymbol{y} \mid \delta) \pi^r(\delta) d\delta} \tag{4.27}$$

[2] The estimators are optimum when $f_1(\cdot)$ and $f_2(\cdot)$ have finite or infinite support.

and $\pi^r(\delta)$ is the right invariant prior defined over $\mathbb{R}$ and corresponds to $\bar{\mathcal{G}}_{KModel}$.[3] The conditional pdf of $\boldsymbol{y}$, $f(\boldsymbol{y} \mid \delta)$, is defined in (4.20). Since $\mathcal{G}_{KModel}$ corresponds to a location parameter estimation problem, the right invariant prior $\pi^r(\delta)$ for $\bar{\mathcal{G}}_{KModel}$ is given by $\pi^r(\delta) = 1$ (see Chapter 6 of [4]).

To find $\hat{\delta}_{opt}$, we differentiate the objective function in (4.26) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and solve for $\hat{\delta}(\boldsymbol{y})$. We have

$$\begin{aligned}\hat{\delta}_{opt}(\boldsymbol{y}) &= \frac{\int_{\mathbb{R}} \delta \pi^r(\delta \mid \boldsymbol{y}) d\delta}{\int_{\mathbb{R}} \pi^r(\delta \mid \boldsymbol{y}) d\delta}, \\ &= \frac{\int_{\mathbb{R}} \delta f(\boldsymbol{y} \mid \delta) \pi^r(\delta) d\delta}{\int_{\mathbb{R}} f(\boldsymbol{y} \mid \delta) \pi^r(\delta) d\delta} = \frac{\int_{\mathbb{R}} \delta f_{\boldsymbol{w}}(\boldsymbol{y} - \delta \boldsymbol{e}) \pi^r(\delta) d\delta}{\int_{\mathbb{R}} f_{\boldsymbol{w}}(\boldsymbol{y} - \delta \boldsymbol{e}) \pi^r(\delta) d\delta}, \\ &= \frac{\int_{\mathbb{R}} \delta \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - (-1)^k \delta) d\delta}{\int_{\mathbb{R}} \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - (-1)^k \delta) d\delta}. \end{aligned} \tag{4.28}$$

Further, the derived estimator $\hat{\delta}_{opt}$ can be shown to be minimax optimum using Theorem 1 from [21]. □

Remark 4.1. Theorem 4.1 provides us with the mathematical expression for the optimum invariant clock offset estimator for the *K-model* under the assumption that we have complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$. The optimum estimator described in Theorem 4.1 is applicable to any pdf-model of the PDV and achieves the lowest MSE among the class of invariant clock offset estimators for the *K-model*. Additionally, the optimum estimator presented in Theorem 4.1 is minimax optimum, i.e., it minimizes the maximum MSE among all clock offset estimators for the *K-model*. The MSE performance of the optimum invariant estimator is extremely useful as it provides a lower bound on the MSE against which we can compare the performance of any proposed estimator designed with limited information, (for example, $f_1(\cdot)$ and $f_2(\cdot)$ are unknown). If a proposed clock offset estimator exhibits performance close to the optimum estimator, the compared estimator is well suited for practical applications.

[3] We should mention here that right invariant prior, $\pi^r(\cdot)$ need not be an actual probability density function [4].

4.3.2 Optimum Invariant Estimator Under S-Model

Recall from (4.16), the observations under the *S-model* can be represented as

$$\boldsymbol{y} = \boldsymbol{A\theta} + \boldsymbol{w}, \tag{4.29}$$

where $\boldsymbol{\theta} = [\delta, d]^T$ denotes the vector of unknown parameters. From (4.16), we have

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = f_{\boldsymbol{w}}(\boldsymbol{y} - \boldsymbol{A\theta}), \tag{4.30}$$

$$= \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - d - (-1)^k \delta). \tag{4.31}$$

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

$$\mathcal{G}_{SModel} = \{g_{\boldsymbol{b}}(\boldsymbol{m})\colon\, g_{\boldsymbol{b}}(\boldsymbol{m}) = \boldsymbol{m} + \boldsymbol{Ab}, \forall \boldsymbol{b} \in \mathbb{R}^2\}, \tag{4.32}$$

where $\boldsymbol{m} \in \mathbb{R}^{2P}$, since $\boldsymbol{y}_g = g_{\boldsymbol{b}}(\boldsymbol{y})$ has the pdf $f_{\boldsymbol{w}}(\boldsymbol{y}_g - \boldsymbol{A}(\boldsymbol{\theta} + \boldsymbol{b}))$ which has the vector location parameter $(\boldsymbol{\theta} + \boldsymbol{b})$ as opposed to the parameters $\boldsymbol{\theta}$ for $f(\boldsymbol{y} \mid \boldsymbol{\theta})$. The corresponding group of induced transformations, $\bar{\mathcal{G}}_{SModel}$, is given by

$$\bar{\mathcal{G}}_{SModel} = \{\bar{g}_{\boldsymbol{b}}(\boldsymbol{\theta})\colon\, \bar{g}_{\boldsymbol{b}}(\boldsymbol{\theta}) = (\boldsymbol{\theta} + \boldsymbol{b}), \forall \boldsymbol{b} \in \mathbb{R}^2\}, \tag{4.33}$$

where $\boldsymbol{\theta} \in \mathbb{R}^2$.

Let $\hat{\delta}_I$ denote an estimator of δ, and let $\hat{\delta}_I(\boldsymbol{y})$ denote the estimates obtained from the received data $\boldsymbol{y}$ characterized by the pdf $f(\boldsymbol{y} \mid \boldsymbol{\theta}) = f_{\boldsymbol{w}}(\boldsymbol{y} - \boldsymbol{A\theta})$. The estimator $\hat{\delta}_I$ is invariant under $\mathcal{G}_{SModel}$ from (4.32) if for all $\boldsymbol{b} \in \mathbb{R}^2$

$$\hat{\delta}_I(g_{\boldsymbol{b}}(\boldsymbol{y})) = \hat{\delta}_I(\boldsymbol{y} + \boldsymbol{Ab}) = \hat{\delta}_I(\boldsymbol{y}) + \boldsymbol{c}^T\boldsymbol{b}, \tag{4.34}$$

where $\boldsymbol{c} = [1, 0]^T$ is a constant vector.

When developing the optimum clock offset estimators under the *S-model*, the loss function for δ is chosen as the squared error loss function [21]. Similar to the case of the *K-model*, it can be shown that the squared error loss function is invariant under $\mathcal{G}_{SModel}$ from (4.32).

The minimax optimum estimator of δ under the *S-model* for the squared error loss function is given in the following theorem [21].

In Theorem 4.2, we develop the optimum invariant clock offset estimator for the *S-model* and use it to calculate the best performance for cases when $f_1(\cdot)$ and $f_2(\cdot)$ are known.[4] Later, we use these results to evaluate the losses from using clock offset estimation schemes designed without knowledge of $f_1(\cdot)$ and $f_2(\cdot)$.

Theorem 4.2. Assuming complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$, the optimum (or minimum conditional risk) invariant estimators of δ, denoted by $\hat{\delta}_{opt}$, under $\mathcal{G}_{KModel}$ defined in (4.21), for the squared error loss functions is given by

$$\hat{\delta}_{opt}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^2} \delta \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - \lambda - (-1)^k \delta) d\delta d\lambda}{\int_{\mathbb{R}^2} \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - \lambda - (-1)^k \delta) d\delta d\lambda}. \tag{4.35}$$

Also, the derived optimum invariant estimators are minimax optimum under the squared error loss (see [21] for complete proof).

Proof. Let $\boldsymbol{\theta} = [\delta, d]^T$ denote the vector of unknown parameters. The optimum invariant estimator of δ under $\mathcal{G}_{SModel}$ from (4.32), denoted by $\hat{\delta}_{opt}$, can be obtained by solving (See [4], Sec. 6.6.2, Result 3)

$$\hat{\delta}_{opt}(\boldsymbol{y}) = \arg\min_{\hat{\delta}} \int_{\mathbb{R}^2} (\hat{\delta}(\boldsymbol{y}) - \delta)^2 \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}, \tag{4.36}$$

where we have

$$\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta})}{\int_{\mathbb{R}} f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{4.37}$$

and $\pi^r(\boldsymbol{\theta})$ is the right invariant prior defined over $\mathbb{R}^2$ and corresponds to $\bar{\mathcal{G}}_{SModel}$. The conditional pdf of $\boldsymbol{y}$, $f(\boldsymbol{y} \mid \boldsymbol{\theta})$, is defined in (4.31). Since $\mathcal{G}_{KModel}$ corresponds to a vector location parameter estimation problem, the right invariant prior $\pi^r(\boldsymbol{\theta})$ for $\bar{\mathcal{G}}_{SModel}$ is given by $\pi^r(\boldsymbol{\theta}) = 1$ (see Section 6 of [4]).

[4] The estimators are optimum when $f_1(\cdot)$ and $f_2(\cdot)$ have finite or infinite support.

To find $\hat{\delta}_{opt}$, we differentiate the objective function in (4.36) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and solve for $\hat{\delta}(\boldsymbol{y})$. We have

$$\begin{aligned}\hat{\delta}_{opt}(\boldsymbol{y}) &= \frac{\int_{\mathbb{R}^2} \delta \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}}{\int_{\mathbb{R}} \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}}, \\ &= \frac{\int_{\mathbb{R}^2} \delta f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^2} f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{\int_{\mathbb{R}^2} \delta f_{\boldsymbol{w}}(\boldsymbol{y} - \boldsymbol{A\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^2} f_{\boldsymbol{w}}(\boldsymbol{y} - \boldsymbol{A\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \\ &= \frac{\int_{\mathbb{R}^2} \delta \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - \lambda - (-1)^k \delta) d\delta d\lambda}{\int_{\mathbb{R}^2} \prod_{k=1}^{2} \prod_{i=1}^{N} f_k(y_{ki} - \lambda - (-1)^k \delta) d\delta d\lambda}. \end{aligned} \tag{4.38}$$

In (4.38), we have replaced the variable d with λ to avoid confusion in notation. Further, the derived estimator $\hat{\delta}_{opt}$ can be shown to be minimax optimum using Theorem 1 from [21]. □

Remark 4.2. Theorem 4.2 provides us with the mathematical expression for the optimum invariant clock offset estimator for the *S-model* under the assumption that we have complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$. The optimum estimator described in Theorem 4.2 is applicable to any pdf-model of the PDV and achieves the lowest MSE among the class of invariant clock offset estimators for the *S-model*. Additionally, the optimum estimator is minimax optimum, i.e., it minimizes the maximum MSE among all clock offset estimators for the *S-model*. The MSE performance of the optimum estimator is extremely useful as it provides a lower bound on the MSE against which we can compare the performance of any proposed estimator with limited information, (for example, $f_1(\cdot)$ and $f_2(\cdot)$ are unknown).

4.4 Optimal Linear L-Estimators for Clock Offset Estimation

A popular class of estimators in the context of location parameter estimation problems are the L-estimators [19, 31]. In these estimators, the unknown location parameter is estimated as a linear combination of the order statistics of a set of observations. Due to the linear nature of the L-estimators, they have significantly lower computational complexity than the minimax optimum estimators of [21]. Further, the conventional COE schemes for IEEE 1588, including the sample minimum, mean,

median, or maximum filtering, can all be expressed as L-estimators with fixed linear combining weights.

With regards to COE, the L-estimators with optimized weights have been studied in [34] and [8]. For the exponential pdf delay model, both the maximum likelihood estimator (MLE) and the minimum variance unbiased estimator (MVUE) of clock offset are linear combinations of order statistics. In this section, we present some results from [22] in the context of designing optimum weights for the L-estimators for the *K-model* and *S-model* that apply for any pdf delay model of the stochastic queuing delays.

Given a vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$, the ith order-statistic of this vector is defined as the ith smallest element of the vector. The vector containing all the order statistics of $\boldsymbol{x}$, ordered from smallest to largest, is referred to as the order statistic vector of $\boldsymbol{x}$ and is denoted by $\uparrow \boldsymbol{x} \uparrow$. Guruswamy *et al.* [22] proposed a linear COE scheme based on the order statistic vectors of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ as

$$\hat{\delta}(\boldsymbol{y}) = \boldsymbol{c}_1^T \uparrow \boldsymbol{y}_1 \uparrow - \boldsymbol{c}_2^T \uparrow \boldsymbol{y}_2 \uparrow + \eta, \tag{4.39}$$

where $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ are weight vectors and η is a scalar constant. Several popular conventional COE schemes can be expressed in the form given in (4.39):

1. *Sample minimum estimator*: $\boldsymbol{c}_1 = \boldsymbol{c}_2 = [0.5, \boldsymbol{0}_{P-1}^T]^T$ and $\eta = 0$.
2. *Sample mean estimator*: $\boldsymbol{c}_1 = \boldsymbol{c}_2 = P^{-1}\boldsymbol{1}_P^T$ and $\eta = 0$.
3. *Sample maximum estimator*: $\boldsymbol{c}_1 = \boldsymbol{c}_2 = [\boldsymbol{0}_{P-1}^T, 0.5]^T$ and $\eta = 0$.

Define

$$\boldsymbol{\mu}_k = E\{\uparrow \boldsymbol{w}_k \uparrow\}, \tag{4.40}$$

$$\boldsymbol{S}_k = \mathrm{cov}\{\uparrow \boldsymbol{w}_k \uparrow\}, \tag{4.41}$$

for $k = 1, 2$, where $\boldsymbol{w}_k$ is defined in (4.12). Let

$$\boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}; \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0}_{P\times P} \\ \boldsymbol{0}_{P\times P} & \boldsymbol{S}_2 \end{bmatrix}. \tag{4.42}$$

Assuming knowledge of $\boldsymbol{\mu}_k$ for $k = 1, 2$ and $\boldsymbol{S}$, Guruswamy *et al.* [22] studied the problem of designing $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ and η to minimize the mean squared error (MSE) under the constraint of constant bias. The following theorem presents the resultant optimum L-estimators from [22]:

Theorem 4.3. Under the *K-model*, the values of $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ and η that minimize the MSE of an estimator of δ of the form given in (4.39) under the constraint of constant bias is given by

$$\boldsymbol{c}^* = \frac{\boldsymbol{S}^{-1}\mathbf{1}_{2P}}{\mathbf{1}_{2P}^T\boldsymbol{S}^{-1}\mathbf{1}_{2P}}, \tag{4.43}$$

$$\eta = \boldsymbol{\mu}_2^T\boldsymbol{c}_2^* - \boldsymbol{\mu}_1^T\boldsymbol{c}_1^*. \tag{4.44}$$

The resultant optimum estimator has an MSE given by

$$\mathrm{MSE}(\hat{\delta}) = (\mathbf{1}_{2P}\boldsymbol{S}^{-1}\mathbf{1}_{2P})^{-1}. \tag{4.45}$$

For proof, see Corollary 1 in [22].

Theorem 4.4. Under the *S-model*, the values of $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ and η that minimize the MSE of an estimator of δ of the form given in (4.39) under the constraint of constant bias is given by

$$\boldsymbol{c}^* = \boldsymbol{S}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{S}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{\gamma}, \tag{4.46}$$

$$\eta = \boldsymbol{\mu}_2^T\boldsymbol{c}_1^* - \boldsymbol{\mu}_2^T\boldsymbol{c}_1^*, \tag{4.47}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1}_P^T & \mathbf{1}_P^T \\ \mathbf{1}_P^T & -\mathbf{1}_P^T \end{bmatrix} \text{ and } \quad \boldsymbol{\gamma} = [1, 0]^T. \tag{4.48}$$

The resultant optimum estimator has an MSE given by

$$\mathrm{MSE}(\hat{\delta}) = \boldsymbol{\gamma}^T(\boldsymbol{A}\boldsymbol{S}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{\gamma}. \tag{4.49}$$

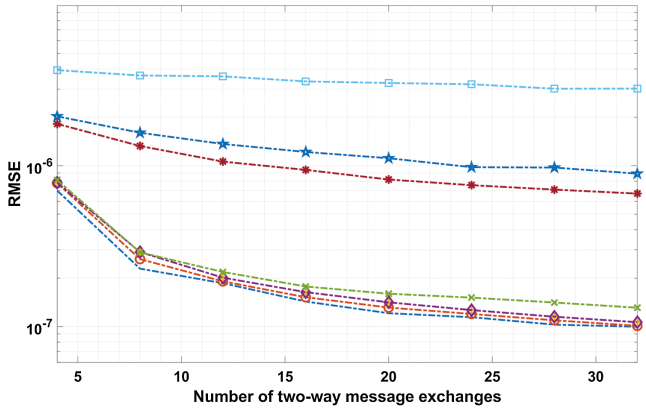
For proof, see Corollary 2 in [22].

Remark 4.3. In Theorems 4.3 and 4.4, we do not assume $f_1(\cdot)$ and $f_2(\cdot)$ are known, but we do assume the second order statistics of $f_k(\cdot)$, namely $\boldsymbol{\mu}_k$ and $\boldsymbol{S}_k$, are known for $k = 1, 2$.

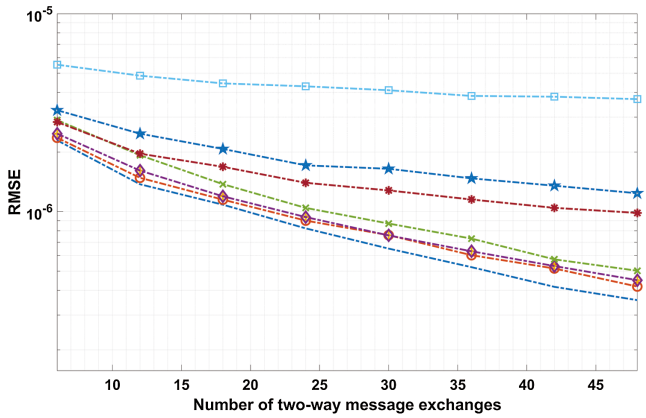
4.5 Numerical Results

We now compare the root mean square estimation error (RMSE) of the L-estimators from [22] and the existing COE schemes to the RMSE of the optimum estimators from [21] under the *K-model* and *S-model*. We focus on the LTE backhaul network scenario described in Section 2. To obtain the pdf of the stochastic queuing delays for this scenario, please refer to Subsection 2.3.3.

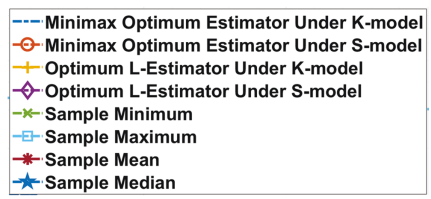
The results are presented in Figure 4.1 for Traffic Model-1 (TM1) and Traffic Model-2 (TM2) under 40% load. From the results, we see that the L-estimator-based COE schemes proposed in [22] exhibits an RMSE very close to the optimum estimators from [21] under both the *K-* and *S-model* for both considered network scenarios. The loss in performance, when restricted to only using L-estimators, is negligible while the conventional clock offset estimators such as the sample mean or the sample median exhibit an RMSE significantly larger than of the optimized L-estimators. On the other hand, the sample minimum estimator performs better than the conventional clock offset estimators but worse than the optimized L-estimators developed in [22]. Additional results highlighting the performance of the L-estimators for various other network scenarios are available in [22].



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



(b) RMSE of clock skew for TM-2 under 40% load.



(c) Legend.

Figure 4.1: RMSE of clock offset for various COE schemes under the network traffic models TM1 and TM2 from [67].

5

Optimum Signal Processing Techniques for Joint Estimation of Clock Skew and Offset

In Section 4, we described optimum clock offset estimators for PTP from [21] for the case when the clock skew is known. Typically, PTP is 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 [30, 66]. PTP is the primary option for synchronization to operators with packet backhaul networks that do not support SyncE [30, 66]. This section addresses the problem of joint estimation of clock skew and offset for PTP (or any clock synchronization protocol based on two-way message exchanges) in the presence of PDV.

The problem of jointly estimating the clock skew and offset of the slave clock while combating the randomness in the observations that occurs due to PDV is referred to as the clock skew and offset estimation (CSOE) problem [40]. Several maximum-likelihood (ML)-based joint clock skew and offset estimators have been proposed in [7, 8, 48, 49, 57] for specific pdf-models of PDV (Gaussian and exponential). Additionally, CSOE schemes based on the least squares estimation approach were proposed in [58]. While the estimators in [7, 8, 48, 49, 57]

are quite useful, they might not be optimum for an arbitrary pdf-model of PDV.

The CSOE problem falls under a variant of the location-scale parameter estimation problem [4, 47], with the unknown clock skew as the scale parameter and the unknown clock offset as the location parameter. In this section, we restrict ourselves to invariant CSOE schemes. This assumption is reasonable as all the popular estimation approaches, including the ML-based joint clock skew and offset estimator and the least-squares joint clock skew and offset estimator, are invariant for this particular problem [40]. Fixing the loss function as the skew-normalized squared error loss and assuming complete knowledge of the statistical information describing the PDV, we present recently developed optimum estimators from [40] for jointly estimating the clock skew and offset. These estimators minimize the skew-normalized mean square estimation error (NMSE) among all invariant CSOE schemes for PTP. Further, using results from [4, 47, 69], the optimum estimators are shown to be minimax optimum for the skew-normalized squared error loss, that is, these estimators minimize the maximum NMSE over all possible values of the unknown parameters.

5.1 Signal Model and Problem Statement

Recall from Section 2 that the slave node obtains four timestamps $\{t_1, t_2, t_3, t_4\}$ after a round of two-way message exchange. The relationship between the timestamps is given by

$$t_2 = \phi(t_1 + d_1 + w_1) + \delta, \tag{5.1}$$

$$t_3 = \phi(t_4 - d_2 - w_2) + \delta, \tag{5.2}$$

where ϕ and δ denote the clock skew and offset of the slave node relative to the master node, respectively. The variables d_1 and d_2 denote the deterministic fixed delays in the forward master-to-slave path and the reverse slave-to-master path, respectively. Following [7, 8, 74], we assume multiple rounds of two-way message exchanges between the master and slave node. Increasing the number of rounds of two-way message exchanges helps in reducing the estimation error when estimating the unknown parameters. It should also be mentioned that a minimum of two

rounds of two-way message exchanges to uniquely estimate ϕ and δ for the case when d_1 and d_2 are unknown, but there is a known relationship between d_1 and d_2 [16], as the number of unknown parameters is greater than the number of available equations for a single round of two-way message exchange. Assuming the parameters ϕ, δ, d_1, and d_2 remain constant over the duration of P two-way message exchanges, we can collect multiple timestamps $\{t_{1i}, t_{2i}, t_{3i}, t_{4i}\}_{i=1}^{P}$ to help estimate ϕ and δ. Similar to (5.1)–(5.2)

$$t_{2i} = (t_{1i} + d_1 + w_{1i})\phi + \delta, \tag{5.3}$$

$$t_{3i} = (t_{4i} - d_2 - w_{2i})\phi + \delta \tag{5.4}$$

for $i = 1, 2, \ldots, P$. Define $\boldsymbol{w}_k = [w_{k1}, w_{k2}, \ldots, w_{kP}]$ for $k = 1, 2$ and $\boldsymbol{t}_k = [t_{k1}, t_{k2}, \ldots, t_{kP}]$ for $k = 1, 2, 3, 4$. The joint pdf of $\boldsymbol{w}_k$ is defined as

$$\begin{aligned} f_{\boldsymbol{w}_k}(\boldsymbol{w}_k) &= f_k(w_{k1}, w_{k2}, \ldots, w_{kP}), \\ &= \prod_{i=1}^{P} f_k(w_{ki}), \end{aligned} \tag{5.5}$$

for $k = 1, 2$. Similar to Section 4, two observation models, with varying degrees of information available about d_1 and d_2 are considered. We have

1. *Known fixed delay model (K-model)*: In this model, we assume complete knowledge of the fixed-path delays d_1 and d_2. The received timestamps shown in (5.3) and (5.4) can be arranged in vector form as follows

   $$\boldsymbol{y} = \boldsymbol{u}\phi + \delta \mathbf{1}_{2P}, \tag{5.6}$$

   where we have $\boldsymbol{y} = [\boldsymbol{t}_2, \boldsymbol{t}_3]^T$, and $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2]^T$, $\boldsymbol{u}_1 = (\boldsymbol{t}_1 + \boldsymbol{w}_1 + d_1 \mathbf{1}_P^T)$ and $\boldsymbol{u}_2 = (\boldsymbol{t}_4 - \boldsymbol{w}_2 - d_2 \mathbf{1}_P^T)$. Also, note that

   $$f_{\boldsymbol{u}}(\boldsymbol{u}) = f_{\boldsymbol{u}_1}(\boldsymbol{u}_1) f_{\boldsymbol{u}_2}(\boldsymbol{u}_2), \tag{5.7}$$

   with

   $$f_{\boldsymbol{u}_1}(\boldsymbol{u}_1) = f_{\boldsymbol{w}_1}(\boldsymbol{u}_1 - \boldsymbol{t}_1 - d_1 \mathbf{1}_P^T); \tag{5.8}$$

   $$f_{\boldsymbol{u}_2}(\boldsymbol{u}_2) = f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{u}_2 - d_2 \mathbf{1}_P^T). \tag{5.9}$$

   The unknown parameters in this model are ϕ and δ.

2. *Standard delay model (S-model)*: In this model, a prior known affine relationship between the fixed path delays is assumed. Let $d_1 = d$ and $d_2 = d$, where the parameter d is unknown. The received time stamps shown in (5.3) and (5.4) can be arranged in vector form as

$$\boldsymbol{y} = (\boldsymbol{h}d + \boldsymbol{v})\phi + \delta \mathbf{1}_{2P}, \tag{5.10}$$

where $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2]^T$, $\boldsymbol{v}_1 = (\boldsymbol{t}_1 + \boldsymbol{w}_1)$, $\boldsymbol{v}_2 = (\boldsymbol{t}_4 - \boldsymbol{w}_2)$, $\boldsymbol{h} = [\mathbf{1}_P^T, \mathbf{1}_P^T]^T$, and $\boldsymbol{y} = [\boldsymbol{t}_2, \boldsymbol{t}_3]^T$. Also, note that

$$f_{\boldsymbol{v}}(\boldsymbol{v}) = f_{\boldsymbol{v}_1}(\boldsymbol{v}_1) f_{\boldsymbol{v}_2}(\boldsymbol{v}_2), \tag{5.11}$$

with

$$f_{\boldsymbol{v}_1}(\boldsymbol{v}_1) = f_{\boldsymbol{w}_1}(\boldsymbol{v}_1 - \boldsymbol{t}_1) \tag{5.12}$$

and

$$f_{\boldsymbol{v}_2}(\boldsymbol{v}_2) = f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{v}_2). \tag{5.13}$$

The unknown parameters in this model are ϕ, d and δ.

Given either observation model, the CSOE problem is to estimate ϕ and δ from the received time stamps. We now state the assumptions used in developing the optimum COE scheme [23].

CSOE Assumptions:

1. All the queuing delays are strictly positive random variables that are mutually independent.
2. All forward queuing delays share a common pdf $f_1(\cdot)$. Similarly the reverse queuing delays share a common pdf $f_2(\cdot)$. When developing the optimum estimators, we assume complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$.
3. The maximum possible value for a forward or reverse queuing delay is finite.
4. The unknown parameters are assumed to be constant over P two-way message exchanges.

5.2 Optimum Joint Clock Skew and Offset Estimators

The CSOE problem under the considered observation models falls under a variant of the location-scale parameter estimation problem [4, 47], with the unknown clock skew as the scale parameter and the unknown clock offset as the location parameter. The optimum invariant joint clock skew and offset estimators for the considered observation models were derived in [40] and are presented in the following subsections.

5.2.1 Optimum Invariant Estimator Under K-Model

Recall from (5.6), the observations under the *K-model* can be represented as

$$\boldsymbol{y} = \boldsymbol{u}\phi + \delta \mathbf{1}_{2P}, \tag{5.14}$$

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

$$\boldsymbol{\Theta} = \{(\phi, \delta)\colon \phi \in \mathbb{R}^{+}, \delta \in \mathbb{R}\}. \tag{5.15}$$

From (5.6), we have

$$\begin{aligned} f(\boldsymbol{y} \mid \boldsymbol{\theta}) &= \frac{1}{\phi^{2P}} f_{\boldsymbol{u}}\left(\frac{\boldsymbol{y} - \delta \mathbf{1}_{2P}}{\phi}\right), \\ &= \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \delta \mathbf{1}_P^T}{\phi} - d_1 \mathbf{1}_P^T - \boldsymbol{t}_1\right) \\ &\quad \times f_{\boldsymbol{w}_2}\left(\frac{\delta \mathbf{1}_P^T - \boldsymbol{t}_3}{\phi} - d_2 \mathbf{1}_P^T + \boldsymbol{t}_4\right), \end{aligned} \tag{5.16}$$

where $f_{\boldsymbol{u}}(\cdot)$ is defined in (5.7).

Let $\mathcal{F}_{KModel}$ denote the class of all pdfs $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Following steps similar to the ones described in Section 4, the class of such pdfs is invariant under the group of location-scale transformations $\mathcal{G}_{KModel}$, on $\mathbb{R}^{2P}$, defined as

$$\mathcal{G}_{KModel} = \{g_{a,b}(\boldsymbol{m})\colon g_{a,b}(\boldsymbol{m}) = a\boldsymbol{m} + b\mathbf{1}_{2P}, \forall (a,b) \in \mathbb{R}^{+} \times \mathbb{R}\}, \tag{5.17}$$

where $\boldsymbol{m} \in \mathbb{R}^{2P}$, since $\boldsymbol{y}_g = g_{a,b}(\boldsymbol{y})$ has the pdf $\frac{1}{(a\phi)^{2P}} f_{\boldsymbol{u}}(\frac{\boldsymbol{y}_g - (a\delta+b)\mathbf{1}_{2P}}{a\phi})$ (where $f_{\boldsymbol{u}}(\cdot)$ was defined in (5.7)) which has the scale and shift parameters $(a\phi, a\delta + b)$ as opposed to the parameters (ϕ, δ) for $f(\boldsymbol{y} \mid \boldsymbol{\theta})$. The corresponding group of induced transformations, $\bar{\mathcal{G}}_{KModel}$, is given by

$$\bar{\mathcal{G}}_{KModel} = \{\bar{g}_{a,b}((\phi,\delta))\colon \bar{g}_{a,b}((\phi,\delta)) = (a\phi, (a\delta+b)), \\ \forall (a,b) \in \mathbb{R}^+ \times \mathbb{R}\}, \tag{5.18}$$

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 $\boldsymbol{y}$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_{KModel}$ from (5.17) if for all $(a,b) \in \mathbb{R}^+ \times \mathbb{R}$

$$\hat{\delta}_I(g_{a,b}(\boldsymbol{y})) = \hat{\delta}_I(a\boldsymbol{y} + b\mathbf{1}_{2P}) = a\hat{\delta}_I(\boldsymbol{y}) + b; \tag{5.19}$$

$$\hat{\phi}_I(g_{a,b}(\boldsymbol{y})) = \hat{\phi}_I(a\boldsymbol{y} + b\mathbf{1}_{2P}) = a\hat{\phi}_I(\boldsymbol{y}). \tag{5.20}$$

Thus the scaling and shifting factors a and b scale and shift the estimators as one might expect. To construct the optimum invariant joint clock skew and offset estimators for the CSOE problem under the *K-model*, we need to have loss functions that are invariant under $\mathcal{G}_{KModel}$ (see Definition 3 in Section 3). To this end, we consider the skew-normalized squared error loss for δ and ϕ. The skew normalized squared error loss functions are invariant under $\mathcal{G}_{KModel}$ from (5.17) since

$$\frac{(\hat{\delta}_I(\boldsymbol{y}) - \delta)^2}{\phi^2} = \frac{(\hat{\delta}_I(g_{a,b}(\boldsymbol{y})) - (a\delta+b))^2}{a^2\phi^2}, \tag{5.21}$$

and

$$\frac{(\hat{\phi}_I(\boldsymbol{y}) - \phi)^2}{\phi^2} = \frac{(\hat{\phi}_I(g_{a,b}(\boldsymbol{y})) - a\phi)^2}{a^2\phi^2} \tag{5.22}$$

for all $g_{a,b} \in \mathcal{G}_{KModel}$. Fixing the loss function to the skew normalized squared error loss, the optimum invariant estimators of δ and ϕ under the *K-model* were derived in [40] and are presented in the following theorem.

In Theorem 5.1, we develop the optimum invariant joint clock skew and offset estimator for the *K-model*. We can use the optimum estimator

to calculate the best performance for cases where $f_1(\cdot)$ and $f_2(\cdot)$ are known.[1] Later, we use these results to evaluate the losses from using joint clock skew and offset estimation schemes designed without knowledge of $f_1(\cdot)$ and $f_2(\cdot)$. If the losses are small, the compared estimators are well suited for practical applications.

Theorem 5.1. Assuming complete information regarding $f_1(\cdot)$ and $f_2(\cdot)$, the optimum (or minimum conditional risk) invariant estimators of δ and ϕ, $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, under $\mathcal{G}_{KModel}$ defined in (5.17), for the skew normalized squared error loss functions are given by

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\delta}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta}) d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta}) d\delta d\phi}, \tag{5.23}$$

and

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^2} f(\boldsymbol{y} \mid \boldsymbol{\theta}) d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta}) d\delta d\phi}, \tag{5.24}$$

respectively, where $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ was defined in (5.16). Also, the optimum estimators are minimax for the skew-normalized squared error loss (see [40] for full proof).

Proof. The optimum invariant estimator of δ under $\mathcal{G}_{KModel}$ in (5.17), denoted by $\hat{\delta}_{MinRisk}$, can be obtained by solving (see Section 6.6.2 in [4])

$$\begin{aligned}\hat{\delta}_{MinRisk}(\boldsymbol{y}) &= \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} L_1(\hat{\delta}(\boldsymbol{y}), \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta} \\ &= \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\delta}(\boldsymbol{y}) - \delta)^2}{\phi^2} \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}, \end{aligned} \tag{5.25}$$

where $\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}}$ is the posterior density of $\boldsymbol{\theta}$ based on the right invariant prior π^r on $\boldsymbol{\Theta}$ (see Section 6.6.1, [4]).[2] The right invariant prior for the location-scale group was derived in [4] (see Section 6.6). As $\mathcal{G}_{KModel}$ from (5.17) is a location-scale group, the right invariant prior density for $\mathcal{G}_{KModel}$ is given by $\pi^r(\boldsymbol{\theta}) = \frac{1}{\phi} \mathcal{I}_{\mathbb{R}^+}(\phi) \mathcal{I}_{\mathbb{R}}(\delta)$.

[1] The estimators are optimum when $f_1(\cdot)$ and $f_2(\cdot)$ have finite or infinite support.

[2] The right invariant prior density need not be an actual density [4] (see Section 6.6, page 409).

To find $\hat{\delta}_{MinRisk}$, we differentiate the objective function in (5.25) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and solve for $\hat{\delta}_{MinRisk}$. We obtain

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\delta}{\phi^2}\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^2}\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}} = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{\delta}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{5.26}$$

Similarly, the optimum invariant estimator of ϕ under $\mathcal{G}_{KModel}$ in (5.17), denoted by $\hat{\phi}_{MinRisk}$, can be obtained by

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\phi}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\phi}(\boldsymbol{y}) - \phi)^2}{\phi^2}\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}. \tag{5.27}$$

Using the same derivative-based approach, we obtain

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^2} f(\boldsymbol{y} \mid \boldsymbol{\theta})d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{1}{\phi^3} f(\boldsymbol{y} \mid \boldsymbol{\theta})d\delta d\phi}. \tag{5.28}$$

□

Remark 5.1. Theorem 5.1 provides us with mathematical expressions for the optimum invariant joint clock skew and offset estimator for the *K-model* under the assumption that we have complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$. The optimum estimator described in Theorem 5.1 is applicable to any pdf-model of the PDV and achieves the lowest MSE among the class of invariant joint clock skew and offset offset estimators for the *K-model*. The MSE performance of the optimum estimator is extremely useful as it provides a lower bound on the MSE against which we can compare the performance of any proposed estimator with limited information, (for example, $f_1(\cdot)$ and $f_2(\cdot)$ are unknown).

5.2.2 Optimum Invariant Estimator Under S-Model

Recall from (5.10), the observations under the *S-model* can be represented as

$$\boldsymbol{y} = (\boldsymbol{h}d + \boldsymbol{v})\phi + \delta\mathbf{1}_{2P}, \tag{5.29}$$

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 $d \in \mathbb{R}_0^+$. When $d \in \mathbb{R}_0^+$, it is not

possible to construct a group for which every element has an inverse due to the non-negativity of d (unlike the group $\mathcal{G}_{SModel}$ defined in (5.34), where every element of $\mathcal{G}_{SModel}$ has a unique inverse). The existence of an inverse is necessary for constructing a group of transformations (see Definition 1 from Section 3). As it is not possible to construct a group of transformations for which the class of pdfs in the *S-model* with the constraint $d \in \mathbb{R}_0^+$, we assume that $d \in \mathbb{R}$.

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

$$\boldsymbol{\Theta} = \{(\phi, d, \delta)\colon \phi \in \mathbb{R}^+, d \in \mathbb{R}, \delta \in \mathbb{R}\}, \tag{5.30}$$

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

$$\boldsymbol{\Theta}^* = \{(\phi, d, \delta)\colon \phi \in \mathbb{R}^+, d \in \mathbb{R}_0^+, \delta \in \mathbb{R}\}. \tag{5.31}$$

From (5.10), we have

$$\begin{aligned} f(\boldsymbol{y} \mid \boldsymbol{\theta}) &= \frac{1}{\phi^{2P}} f_{\boldsymbol{v}}\left(\frac{\boldsymbol{y} - \delta \mathbf{1}_{2P}}{\phi} - \boldsymbol{h}d\right), \\ &= \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2 - \delta \mathbf{1}_P^T}{\phi} - d\mathbf{1}_P^T - \boldsymbol{t}_1\right) \\ &\quad \times f_{\boldsymbol{w}_2}\left(\frac{\delta \mathbf{1}_P^T - \boldsymbol{t}_3}{\phi} + \boldsymbol{t}_4 - \mathbf{1}_P^T\right), \end{aligned} \tag{5.32}$$

where we have

$$f_{\boldsymbol{v}}(\boldsymbol{v}) = f_{\boldsymbol{w}_1}(\boldsymbol{v}_1 - \boldsymbol{t}_1) f_{\boldsymbol{w}_2}(\boldsymbol{t}_4 - \boldsymbol{v}_2). \tag{5.33}$$

Let $\mathcal{F}_{SModel}$ denote the class of all pdfs $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ 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

$$\begin{aligned} \mathcal{G}_{SModel} = \{g_{a,b,c}(\boldsymbol{m})\colon\ & g_{a,b,c}(\boldsymbol{m}) = a(\boldsymbol{m} + \boldsymbol{h}b) + c\mathbf{1}_{2P}, \\ & \forall (a, b, c) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}\}, \end{aligned} \tag{5.34}$$

where $\boldsymbol{m} \in \mathbb{R}^{2P}$. Since $\boldsymbol{y}_g = g_{a,b,c}(\boldsymbol{y})$ has the conditional pdf $\frac{1}{(a\phi)^{2P}} \cdot f_{\boldsymbol{v}}(\frac{\boldsymbol{y}_g - (a\delta + c)\mathbf{1}_{2P}}{a\phi} - \boldsymbol{h}(d + \frac{b}{\phi}))$, which has the parameters $(a\phi, (d + b/\phi), a\delta + c)$ as opposed to the parameters (ϕ, d, δ) for $f(\boldsymbol{y} \mid \boldsymbol{\theta})$. This shows

that the group, $\bar{\mathcal{G}}_{SModel}$, of induced transformations on $\boldsymbol{\Theta}$ is given by

$$\bar{\mathcal{G}}_{SModel} = \{\bar{g}_{a,b,c}((\phi, d, \delta)): \bar{g}_{a,b,c}((\phi, d, \delta)) = (a\phi, (d + b/\phi), (a\delta + c))\tag{5.35}$$

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 valid 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 $f(\boldsymbol{y} \mid \boldsymbol{\theta})$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_{SModel}$ from (5.34), if for all $(a, b, c) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}$,

$$\hat{\delta}_I(g_{a,b,c}(\boldsymbol{y})) = \hat{\delta}_I(a(\boldsymbol{y} + \boldsymbol{h}b) + c\mathbf{1}_{2P}) = (a\hat{\delta}_I(\boldsymbol{y}) + c);\tag{5.36}$$

$$\hat{\phi}_I(g_{a,b,c}(\boldsymbol{y})) = \hat{\phi}_I(a(\boldsymbol{y} + \boldsymbol{h}b) + c\mathbf{1}_{2P}) = a\hat{\phi}_I(\boldsymbol{y}).\tag{5.37}$$

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 (5.36) and (5.37) 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 are invariant under $\mathcal{G}_{SModel}$ from (5.34), since

$$\frac{(\hat{\delta}_I(\boldsymbol{y}) - \delta)^2}{\phi^2} = \frac{(\hat{\delta}_I(g_{a,b,c}(\boldsymbol{y})) - (a\delta + c))^2}{a^2\phi^2},\tag{5.38}$$

and

$$\frac{(\hat{\phi}_I(\boldsymbol{y}) - \phi)^2}{\phi^2} = \frac{(\hat{\phi}_I(g_{a,b,c}(\boldsymbol{y})) - a\phi)^2}{a^2\phi^2}\tag{5.39}$$

for all $g_{a,b,c} \in \mathcal{G}_{SModel}$.

In Theorem 5.2, we develop the optimum invariant joint clock skew and offset estimator for the *S-model*. We can use the optimum estimator to calculate the best performance for cases when $f_1(\cdot)$ and $f_2(\cdot)$ are known.[3] Later, we use these results to evaluate the losses from using joint clock skew and offset estimation schemes designed without knowledge

[3]The estimators are optimum when $f_1(\cdot)$ and $f_2(\cdot)$ have finite or infinite support.

of $f_1(\cdot)$ and $f_2(\cdot)$. The minimax optimum estimators of δ and ϕ under the *S-model* was derived in [40] and are given by

Theorem 5.2. Assuming complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$, the optimum (or minimum conditional risk) invariant estimators of δ and ϕ, $\hat{\delta}_{MinRisk}$ and $\hat{\phi}_{MinRisk}$, under $\mathcal{G}_{SModel}$ defined in (5.34), for the skew normalized squared error loss functions, are given by

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{\delta}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}, \tag{5.40}$$

and

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}, \tag{5.41}$$

respectively, where $f(\boldsymbol{y}\mid\delta,\phi,\lambda)$ is defined as

$$\begin{aligned} f(\boldsymbol{y}\mid\delta,\phi,\lambda) =& \frac{1}{\phi^{2P}} f_{\boldsymbol{w}_1}\left(\frac{\boldsymbol{t}_2-\delta\mathbf{1}_P^T}{\phi}-\lambda\mathbf{1}_P^T-\boldsymbol{t}_1\right) \\ & f_{\boldsymbol{w}_2}\left(\frac{\delta\mathbf{1}_P^T-\boldsymbol{t}_3}{\phi}+\boldsymbol{t}_4-\lambda\mathbf{1}_P^T\right). \end{aligned} \tag{5.42}$$

Further, the derived optimum invariant estimators are minimax for the skew-normalized squared error loss in the restricted parameter space $\boldsymbol{\Theta}^*$ (see [40] for proof).

Proof. Let $\boldsymbol{\theta} = [\phi, d, \delta]$. We first calculate the right invariant prior for $\mathcal{G}_{SModel}$, defined in (5.34), 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 (5.30). The right group transformation of $\mathcal{A}$ by $\boldsymbol{\theta}_0$ is given by [47]

$$\mathcal{A}_{r_0} = \{\boldsymbol{\theta}_{r_0} = (\phi_{r_0}, d_{r_0}, \delta_{r_0})\colon \boldsymbol{\theta}_{r_0} = \bar{g}_{\phi,d,\delta}(\boldsymbol{\theta}_0), (\phi,d,\delta)\in\mathcal{A}\}, \tag{5.43}$$

$$= \{\boldsymbol{\theta}_{r_0} = (\phi\phi_0, d_0 + d/\phi_0, \phi\delta_0+\delta)\colon (\phi,d,\delta)\in\mathcal{A}\}, \tag{5.44}$$

with $\bar{g}_{\phi,d,\delta} \in \bar{\mathcal{G}}_{SModel}$ from (5.35). The right invariant prior, π^r, on $\mathcal{G}_{SModel}$ from (5.34) is obtained by finding the function that satisfies[4]

$$\int_{\mathcal{A}} \pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad = \quad \int_{\mathcal{A}_{r_0}} \pi^r(\boldsymbol{\theta}_{r_0}) d\boldsymbol{\theta}_{r_0}, \tag{5.45}$$

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 $\pi^r(\boldsymbol{\theta}) = \mathcal{I}_{\mathbb{R}^+}(\phi)\mathcal{I}_{\mathbb{R}}(d) \cdot \mathcal{I}_{\mathbb{R}}(\delta)$. To see this, note that

$$\int_{\mathcal{A}} 1 d\boldsymbol{\theta} = \int_{\mathcal{A}_{r_0}} \frac{d\boldsymbol{\theta}}{d\boldsymbol{\theta}_{r_0}} d\boldsymbol{\theta}_{r_0} = \int_{\mathcal{A}_{r_0}} 1 d\boldsymbol{\theta}_{r_0}, \tag{5.46}$$

since the Jacobian of the transformation in (5.44) is given by

$$\begin{aligned} \frac{d\boldsymbol{\theta}_{r_0}}{d\boldsymbol{\theta}} &= \det \left(\begin{bmatrix} \frac{\partial \phi_{r_0}}{\partial \phi} & \frac{\partial \phi_{r_0}}{\partial d} & \frac{\partial \phi_{r_0}}{\partial \delta} \\ \frac{\partial d_{r_0}}{\partial \phi} & \frac{\partial d_{r_0}}{\partial d} & \frac{\partial d_{r_0}}{\partial \delta} \\ \frac{\partial \delta_{r_0}}{\partial \phi} & \frac{\partial \delta_{r_0}}{\partial d} & \frac{\partial \delta_{r_0}}{\partial \delta} \end{bmatrix} \right) \\ &= \det \left(\begin{bmatrix} \phi_0 & 0 & 0 \\ 0 & 1/\phi_0 & 0 \\ \delta_0 & 0 & 1 \end{bmatrix} \right) = 1. \end{aligned} \tag{5.47}$$

The optimum invariant estimators of δ under $\mathcal{G}_{SModel}$ from (5.34), denoted by $\hat{\delta}_{MinRisk}$, can now be obtained by solving

$$\hat{\delta}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\delta}} \int_{\boldsymbol{\Theta}} \frac{(\hat{\delta}(\boldsymbol{y}) - \delta)^2}{\phi^2} \pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}, \tag{5.48}$$

where $\pi^r(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})\pi^r(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{y} \mid \boldsymbol{\theta})\pi^r(\boldsymbol{\theta}) d\boldsymbol{\theta}}$ 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 (5.48) with respect to $\hat{\delta}(\boldsymbol{y})$, set the result equal to zero and

[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.

solve for $\hat{\delta}(\boldsymbol{y})$. We have

$$\begin{aligned}\hat{\delta}_{MinRisk}(\boldsymbol{y}) &= \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{\delta}{\phi^2}\pi^r(\boldsymbol{\theta}\mid\boldsymbol{y})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}\pi^r(\boldsymbol{\theta}\mid\boldsymbol{y})d\boldsymbol{\theta}},\\ &= \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{\delta}{\phi^2}f(\boldsymbol{y}\mid\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}f(\boldsymbol{y}\mid\boldsymbol{\theta})d\boldsymbol{\theta}},\\ &= \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{\delta}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}.\end{aligned} \tag{5.49}$$

We have replaced d with λ in Theorem 5.2 to make the notation clear to the reader.

Similarly, the optimum invariant estimator of ϕ under $\mathcal{G}_{SModel}$ from (5.34), denoted by $\hat{\phi}_{MinRisk}$, can be obtained by solving

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \underbrace{\arg\min}_{\hat{\phi}}\int_{\boldsymbol{\Theta}}\frac{(\hat{\phi}(\boldsymbol{y})-\phi)^2}{\phi^2}\pi^r(\boldsymbol{\theta}\mid\boldsymbol{y})d\boldsymbol{\theta}. \tag{5.50}$$

Solving, we obtain

$$\hat{\phi}_{MinRisk}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^2}\frac{1}{\phi^2}f(\boldsymbol{y}\mid\delta,\phi,\lambda)d\lambda d\delta d\phi}. \tag{5.51}$$

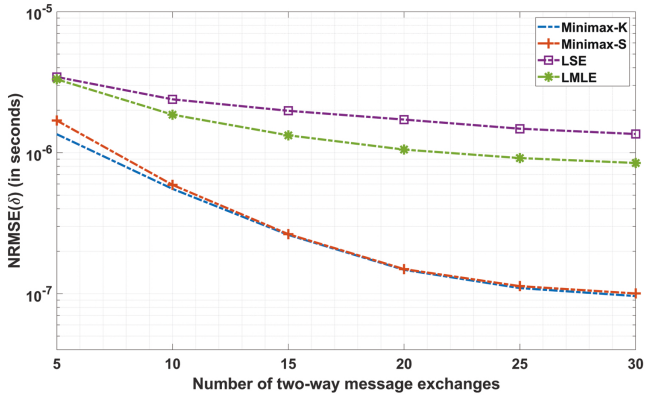
□

Remark 5.2. Theorem 5.2 provides us with mathematical expressions for the optimum invariant joint clock skew and offset estimator for the *S-model* under the assumption that we have complete knowledge of $f_1(\cdot)$ and $f_2(\cdot)$. The optimum estimator described in Theorem 5.2 is applicable to any pdf-model of the PDV and achieves the lowest MSE among the class of invariant joint clock skew and offset estimators for the *S-model*. The MSE performance of the optimum estimator is extremely useful as it provides a lower bound on the MSE against which we can compare the performance of any proposed estimator with limited information, (for example, $f_1(\cdot)$ and $f_2(\cdot)$ are unknown). If close, the compared estimators are well suited for practical applications.

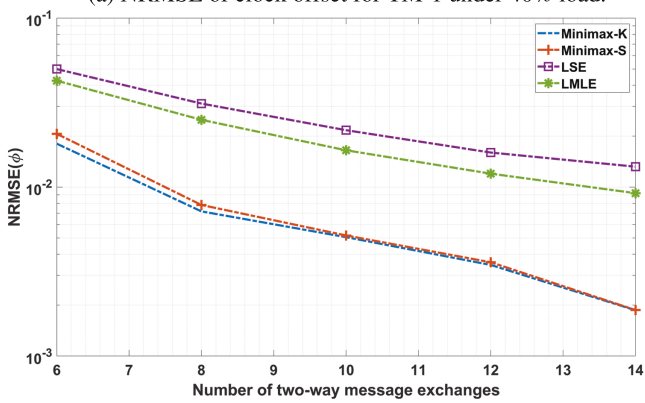
5.3 Numerical Results

We now compare the skew normalized root mean square estimation error (NRMSE) of the optimum estimators under the *K-model* (Minimax-K) and *S-model* (Minimax-S) against the conventional maximum likelihood (LMLE) and the least-squares joint clock skew and offset estimator (LSE) under the *K-model* in the LTE backhaul network scenario described in Section 2. To obtain the pdf of the stochastic queuing delays for this scenario, please refer to Subsection 2.3.3.

The results are presented in Figures 5.1 and 5.2. From the results, we see that the optimum estimators significantly outperform the conventional CSOE schemes for the considered network scenario. Minimax-K exhibits the lowest NRMSE among all the considered schemes as it has prior knowledge regarding d. However, interestingly, there is no significant loss in the performance of Minimax-S due to the unknown nuisance parameter d. 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. This is also visible in Figure 5.2.

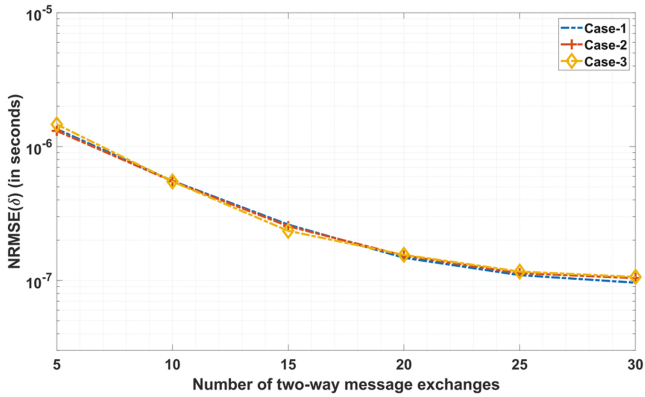


(a) NRMSE of clock offset for TM-1 under 40% load.

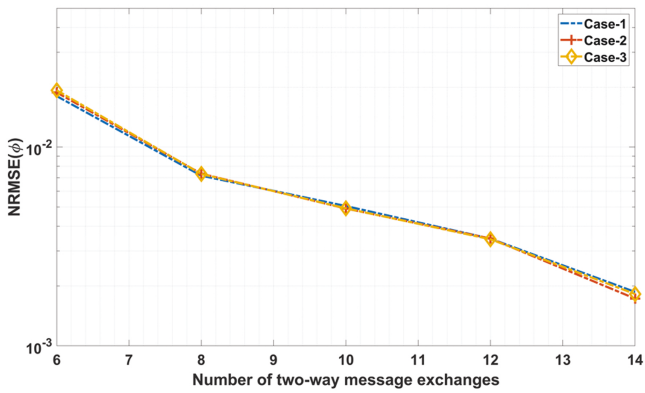


(b) NRMSE of clock skew for TM-1 under 40% load.

Figure 5.1: NRMSE of clock skew and offset for various CSOE schemes under Traffic Model-1.



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



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

Figure 5.2: NRMSE performance of minimax optimum estimator under *K-model* for different parameter values. We have for case 1, $\{\phi, d, \delta\} = \{1.01, 1\ \mu\text{s}, 1.25\ \mu\text{s}\}$, for case 2, $\{\phi, d, \delta\} = \{1.02, 1\ \mu\text{s}, 1.25\ \mu\text{s}\}$ and for case 3, $\{\phi, d, \delta\} = \{0.98, 1\ \mu\text{s}, -1.25\ \mu\text{s}\}$.

6

Joint Clock Skew and Offsets Estimators Robust Against Unknown Path Asymmetries

In Section 4, we described optimum estimation schemes for clock offset estimation, and in Section 5, we described optimum joint clock skew and offset estimation schemes for PTP. The optimum estimators described in Sections 4 and 5, assumed a prior known affine relationship between the unknown fixed path delays. However, the performance of joint clock skew and offset estimation schemes can significantly degrade in the presence of unknown path asymmetries [68]. This unknown asymmetry between the fixed path delays can arise from several sources, including delay attacks [68] and routing asymmetry [3]. This section addresses the problem of developing joint clock skew and offset estimators for PTP that are robust against unknown path asymmetries.

Several clock offset estimation schemes that are robust against unknown path asymmetries which assume complete knowledge of the clock skew, are available in the literature [17, 38, 39, 53, 54, 64]. However, to the best of our knowledge, there are no joint clock skew and offset estimation schemes available in the literature whose performance does not degrade in the presence of unknown deterministic path asymmetries. The IEEE 1588 PTP standard [32] proposes the use of timing information from multiple masters rather than a single master to help protect

against unknown path asymmetries. Following the PTP standard [32], we assume the availability of multiple master-slave communication paths in our work and further assume that fewer than half of the master-slave communication paths have an unknown asymmetry between the fixed path delays.[1]

We first present bounds from [41] on the best possible performance for invariant joint clock skew and offset estimation schemes in the presence of possible unknown path asymmetries. When developing the performance lower bounds, prior knowledge on whether a master-slave communication path has an unknown asymmetry between the fixed path delays as well as the complete knowledge of the pdf describing the PDV in the master-slave communication path is assumed. We then briefly describe recently developed robust joint clock skew and offset estimators from [41] that can handle unknown path asymmetries. Numerical results comparing the performance of the robust CSOE scheme against the performance lower bounds are also presented.

6.1 Signal Model and Problem Statement

Assume the availability of N master-slave communication paths and perfect synchronization between the clocks of the N masters. Recall that the relative clock skew and offset of the slave node with respect to a master node are denoted by $\phi \in \mathbb{R}^+$ and $\delta \in \mathbb{R}$, respectively. A total of P rounds of two-way message exchanges is assumed at each master-slave communication path. The relationship between the received timestamps obtained from the various master-slave communication paths are given by [7, 48, 57]

$$t_{2ij} = (t_{1ij} + d_i^{ms} + w_{1ij})\phi + \delta, \tag{6.1}$$

$$t_{3ij} = (t_{4ij} - d_i^{sm} + w_{2ij})\phi + \delta, \tag{6.2}$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, P$. In (6.1) and (6.2), d_i^{ms} and d_i^{sm} denote the unknown fixed propagation delays in the forward and reverse path, respectively, at the ith master-slave communication path.

[1] Alternatively, it could be assumed that one particular path must be symmetric as opposed to assuming more than half of the paths are symmetric.

The variables w_{1ij} and w_{2ij} denote the random queuing delays in the forward and reverse path, respectively, during the jth round of message exchanges for the ith master-slave communication path. The pdf of $\{w_{kij}\}_{j=1}^{P}$ is denoted by $f_{ki}(\cdot)$ for $k = 1, 2$ and $i = 1, 2, \ldots, N$.

Clock synchronization protocols including PTP [32], NTP [52], TPSN [20] used in real networks generally assume that the fixed path delays in the forward and reverse paths are equal. Freris *et al.* [16] gave the necessary conditions in order to obtain a unique solution to the clock skew and offset for protocols based on a two-way message exchange scheme. We need to know either one of the fixed path delays, or we need to have a prior known relationship between the fixed path delays in the forward and reverse paths. In this section, we classify a master-slave communication path as being symmetric or asymmetric depending on the relationship between the fixed path delays. A *symmetric master-slave communication path* denotes a path in which the fixed path delays in the forward and reverse paths are equal, i.e., $d_i^{ms} = d_i^{sm} = d_i$, where d_i denotes the unknown fixed path delay over the ith master-slave communication path. Similarly, an *asymmetric master-slave communication path* denotes a path having an unknown asymmetry between the forward and reverse fixed path delays, i.e., $d_i^{ms} = d_i + \tau_i$ and $d_i^{sm} = d_i$. The parameter τ_i denotes the unknown constant (for all j) unexpected asymmetry between the fixed path delays.

Define $\boldsymbol{w}_{ki} = [w_{ki1}, w_{ki2}, \ldots, w_{kiP}]$ for $k = 1, 2$, $i = 1, 2, \ldots, N$ and $\boldsymbol{t}_{ki} = [t_{ki1}, t_{ki2}, \ldots, t_{kiP}]$ for $k = 1, 2, 3, 4$ and $i = 1, 2, \ldots, N$. We now introduce a new binary state vector variable $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_N]$, which indicates whether a master-slave communication path is symmetric or asymmetric. The ith element of $\boldsymbol{\eta}$ is 1 when the ith master-slave communication path has asymmetric fixed path delays, else it has a value of 0. The parameter $\boldsymbol{\eta}$ is usually not known and has to estimated from the observations. If $\eta_i = 0$, the received timestamps from the ith master-slave communication path can be arranged in vector form as

$$\boldsymbol{y}_i = (\boldsymbol{h}d_i + \boldsymbol{v}_i)\phi + \delta \mathbf{1}_{2P}, \tag{6.3}$$

where $\boldsymbol{y}_i = [\boldsymbol{t}_{2i}, \boldsymbol{t}_{3i}]^T$, $\boldsymbol{h} = [\mathbf{1}_P^T, -\mathbf{1}_P^T]^T$ and $\boldsymbol{v}_i = [\boldsymbol{v}_{1i}, \boldsymbol{v}_{2i}]^T$ with $\boldsymbol{v}_{1i} = (\boldsymbol{t}_{1i} + \boldsymbol{w}_{1i})$ and $\boldsymbol{v}_{2i} = (\boldsymbol{t}_{4i} - \boldsymbol{w}_{2i})$. Similarly, when $\eta_i = 1$, the received timestamps from the ith master-slave communication path can be

arranged in vector form as

$$\boldsymbol{y}_i = (\boldsymbol{h}d_i + \boldsymbol{g}\tau_i + \boldsymbol{v}_i)\phi + \delta\mathbf{1}_{2P}, \tag{6.4}$$

where $\boldsymbol{g} = [\mathbf{1}_P^T, \mathbf{0}_P^T]^T$. The complete set of received timestamps is denoted by $\boldsymbol{t} = [\boldsymbol{t}_{11}, \boldsymbol{t}_{12}, \ldots, \boldsymbol{t}_{1N}, \boldsymbol{t}_{21}, \boldsymbol{t}_{22}, \ldots, \boldsymbol{t}_{2N}, \boldsymbol{t}_{31}, \boldsymbol{t}_{32}, \ldots, \boldsymbol{t}_{3N}, \boldsymbol{t}_{41}, \boldsymbol{t}_{42}, \ldots, \boldsymbol{t}_{4N}]$. We now state the assumptions made in [41] when developing the performance bounds as well as the robust joint clock skew and offset estimators:

CSOE with Asymmetry Assumptions

1. We assume the availability of N master-slave communication paths. Further, we assume that fewer than half of the N master-slave communication paths have an unknown asymmetry between the fixed path delays, i.e., $||\boldsymbol{\eta}||_1 < N/2$.

2. All the queuing delays are strictly positive random variables and have finite support. Also, the random queuing delays $\{w_{kij}\}_{j=1}^{P}$ are assumed to be independent and identically distributed. The pdf of the random variables are denoted by $f_{ki}(\cdot)$ for $i = 1, 2, \ldots, N$, $k = 1, 2$.

3. The unknown fixed delays $\{d_i\}_{i=1}^{N}$, unknown biases $\{\tau_i\}_{i=1}^{N}$, clock skew ϕ and the clock offset δ are assumed to be constant over P two-way message exchanges for each master-slave communication path.

4. As very small τ_i will have little impact, we officially define a master-slave communication path as having an unknown asymmetry ($\eta_i = 1$) when $|\tau_i| \geq d_\tau$, where d_τ can be chosen such that $|\tau_i| < d_\tau$ causes little impact.

6.2 Performance Lower Bounds Based on Optimum Estimators

We first present useful performance lower bounds from [41] that help in evaluating the performance of the proposed joint clock skew and offset estimation schemes that are robust to unknown path asymmetries.

We assume $\boldsymbol{\eta}$ is known and further assume complete knowledge of the forward- and reverse queuing delay pdfs $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for all N master-slave communication paths for $i = 1, 2, \ldots, N$.

Fixing the loss function to the skew normalized squared error loss, invariant decision theory was employed in [41] to develop the optimum approach for fusing information from the N master-slave communication paths. As prior information regarding the asymmetric paths is assumed, the skew normalized mean square estimation error (NMSE) performance of the optimum approach provides us a lower bound on the NMSE for an invariant joint clock skew and offset estimation scheme in the presence of possible unknown path asymmetries. We now describe the performance lower bounds developed in [41].

For ease of notation, assume the first $K(<N/2)$ master-slave communication paths have an unknown symmetry and the remaining $(N-K)$ master-slave communication paths have a known asymmetry. Under these assumptions with (6.3) and (6.4), we have

$$\boldsymbol{y}_i = (\boldsymbol{h}d_i + \boldsymbol{g}\tau_i + \boldsymbol{v}_i)\phi + \delta\mathbf{1}_{2P}, \tag{6.5}$$

for $i = 1, 2, \ldots, K$ and

$$\boldsymbol{y}_i = (\boldsymbol{h}d_i + \boldsymbol{v}_i)\phi + \delta\mathbf{1}_{2P}, \tag{6.6}$$

for $i = K+1, \ldots, N$. The complete set of observations from the N master-slave communication paths can be represented in vector form as

$$\boldsymbol{y} = (\boldsymbol{H}\boldsymbol{\gamma} + \boldsymbol{v})\phi + \delta\mathbf{1}_{2NP}, \tag{6.7}$$

where $\boldsymbol{H} = [\boldsymbol{h} \otimes \boldsymbol{I}_N, \boldsymbol{g} \otimes \boldsymbol{I}_K]$ with $\otimes$ denoting the Kronecker product operation, $\boldsymbol{y} = [\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_N^T]$, $\boldsymbol{v} = [\boldsymbol{v}_1^T, \boldsymbol{v}_2^T, \ldots, \boldsymbol{v}_N^T]$ and $\boldsymbol{\gamma} = [\boldsymbol{d}, \boldsymbol{\tau}]$ with $\boldsymbol{d} = [d_1, d_2, \ldots, d_N]$ and $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_K]$.

Let $\boldsymbol{\theta} = [\phi, \delta, d_1, d_2, \ldots, d_N, \tau_1, \ldots, \tau_K]$ denote the vector of unknown parameters. The parameter space of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\Theta}$, is given by $\boldsymbol{\Theta} = \{(\phi, \delta, \boldsymbol{d}, \boldsymbol{\tau})\colon \phi \in \mathbb{R}^+, \delta \in \mathbb{R}, \boldsymbol{d} \in \mathbb{R}^N, \boldsymbol{\tau} \in \mathbb{R}^K\}$. From (6.7), the

conditional pdf of $\boldsymbol{y}$ is given by

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \frac{1}{\phi^{2NP}} f_{\boldsymbol{v}} \left(\frac{\boldsymbol{y} - \delta \mathbf{1}_{2NP}}{\phi} - \boldsymbol{H}\boldsymbol{\gamma} \right), \tag{6.8}$$

$$= \frac{1}{\phi^{2NP}} \prod_{i=1}^{K} f_{\boldsymbol{v}_i} \left(\frac{\boldsymbol{y}_i - \delta \mathbf{1}_{2P}}{\phi} - d_i \boldsymbol{h} - \tau_i \boldsymbol{g} \right) \prod_{i=K+1}^{N} f_{\boldsymbol{v}_i} \left(\frac{\boldsymbol{y}_i - \delta \mathbf{1}_{2P}}{\phi} - d_i \boldsymbol{h} \right), \tag{6.9}$$

where $\boldsymbol{y}_i = [\boldsymbol{t}_{2i}, \boldsymbol{t}_{3i}]^T$ and $f_{\boldsymbol{v}_i}(\boldsymbol{v}_i) = \prod_{j=1}^{P} f_{1i}(v_{1ij} - t_{1ij}) f_{2i}(t_{4ij} - v_{2ij})$ for $i = 1, 2, \ldots, N$, where $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ denote the pdf of the forward and reverse path stochastic queuing delays, respectively.

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

$$\mathcal{G}_M = \{ g_{a,\boldsymbol{b},c}(\boldsymbol{y}) \colon g_{a,\boldsymbol{b},c}(\boldsymbol{y}) = (\boldsymbol{y} + \boldsymbol{H}\boldsymbol{b})a + c\mathbf{1}_{2NP}, \quad \forall (a, \boldsymbol{b}, c) \in \mathbb{R}^+ \times \mathbb{R}^{N+K} \times \mathbb{R} \}, \tag{6.10}$$

where $\boldsymbol{y} \in \mathbb{R}^{2NP}$ since $\boldsymbol{y}_g = g_{a,\boldsymbol{b},c}(\boldsymbol{y})$ has a pdf given by $\frac{1}{(a\phi)^{2NP}} f_{\boldsymbol{v}}(\frac{\boldsymbol{y}_g - ((a\delta + c)\mathbf{1}_{2NP})}{a\phi} - \boldsymbol{H}(\boldsymbol{\gamma} + \frac{\boldsymbol{b}}{\phi}))$. The corresponding group of induced transformations on $\boldsymbol{\Theta}$, denoted by $\bar{\mathcal{G}}_M$, is given by

$$\bar{\mathcal{G}}_M = \{ \bar{g}_{a,\boldsymbol{b},c}((\phi, \boldsymbol{\gamma}, \delta)) \colon \bar{g}_{a,\boldsymbol{b},c}((\phi, \boldsymbol{\gamma}, \delta)) = (a\phi, (\boldsymbol{\gamma} + \boldsymbol{b}/\phi), (a\delta + c)), \quad \forall (a, \boldsymbol{b}, c) \in \mathbb{R}^+ \times \mathbb{R}^{N+K} \times \mathbb{R} \}, \tag{6.11}$$

where $\phi \in \mathbb{R}^+, \boldsymbol{\gamma} \in \mathbb{R}^{N+K}$ 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 $f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \frac{1}{\phi^{2NP}} f_{\boldsymbol{v}}(\frac{\boldsymbol{y} - (\delta \mathbf{1}_{2NP})}{\phi} - \boldsymbol{H}\boldsymbol{\gamma})$. The estimators $\hat{\phi}_I(\boldsymbol{y})$ and $\hat{\delta}_I(\boldsymbol{y})$ are invariant under $\mathcal{G}_M$ from (6.10), if for all $(a, \boldsymbol{b}, c) \in \mathbb{R}^+ \times \mathbb{R}^{N+K} \times \mathbb{R}$,

$$\hat{\delta}_I(g_{a,\boldsymbol{b},c}(\boldsymbol{y})) = \hat{\delta}_I(a(\boldsymbol{y} + \boldsymbol{H}\boldsymbol{b}) + c\mathbf{1}_{2NP}) = a\hat{\delta}_I(\boldsymbol{y}) + c; \tag{6.12}$$

$$\hat{\phi}_I(g_{a,\boldsymbol{b},c}(\boldsymbol{y})) = \hat{\phi}_I(a(\boldsymbol{y} + \boldsymbol{H}\boldsymbol{b}) + c\mathbf{1}_{2NP}) = a\hat{\phi}_I(\boldsymbol{y}). \tag{6.13}$$

Following similar steps to the derivation in Section 4 and fixing the loss function to the skew normalized squared error loss, the optimum

invariant estimators of δ and ϕ for $\mathcal{G}_M$ from (6.10) was derived in [41] for known $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$ and prior knowledge of master-slave communication paths having unknown deterministic path asymmetries. As these optimum estimators achieve the smallest skew normalized mean square estimation error (NMSE) among the class of invariant estimators, the MSE performance of these estimators give us useful fundamental lower bounds on the MSE for a joint clock skew and offset estimation scheme for PTP. Later, we use these results to evaluate the losses from using joint clock skew and offset estimation schemes which were designed for cases where $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$ and the asymmetric paths are not known.

Theorem 6.1. Assuming knowledge of the paths having an unknown asymmetry and complete knowledge of $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$, the optimum invariant estimators for δ and ϕ, denoted by $\hat{\delta}_{opt}$ and $\hat{\phi}_{opt}$, respectively, are given by

$$\hat{\delta}_{opt}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{N+K+1}} \frac{\delta\Gamma_1(\phi,\delta,\boldsymbol{d},\boldsymbol{\tau},\boldsymbol{y})\Gamma_0(\phi,\delta,\boldsymbol{d},\boldsymbol{y})}{\phi^{2NP-N-K+3}} d\boldsymbol{\tau} d(\boldsymbol{d}) d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{N+K+1}} \frac{\Gamma_1(\phi,\delta,\boldsymbol{d},\boldsymbol{\tau},\boldsymbol{y})\Gamma_0(\phi,\delta,\boldsymbol{d},\boldsymbol{y})}{\phi^{2NP-N-K+3}} d\boldsymbol{\tau} d(\boldsymbol{d}) d\delta d\phi}, \tag{6.14}$$

and

$$\hat{\phi}_{opt}(\boldsymbol{y}) = \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{N+K+1}} \frac{\Gamma_1(\phi,\delta,\boldsymbol{d},\boldsymbol{\tau},\boldsymbol{y})\Gamma_0(\phi,\delta,\boldsymbol{d},\boldsymbol{y})}{\phi^{2NP-N-K+2}} d\boldsymbol{\tau} d(\boldsymbol{d}) d\delta d\phi}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^{N+K+1}} \frac{\Gamma_1(\phi,\delta,\boldsymbol{d},\boldsymbol{\tau},\boldsymbol{y})\Gamma_0(\phi,\delta,\boldsymbol{d},\boldsymbol{y})}{\phi^{2NP-N-K+3}} d\boldsymbol{\tau} d(\boldsymbol{d}) d\delta d\phi}, \tag{6.15}$$

respectively, where we have $\Gamma_1(\phi, \delta, \boldsymbol{d}, \boldsymbol{\tau}, \boldsymbol{y}) = \prod_{i=1}^{K} f_{\boldsymbol{v}_i}(\frac{\boldsymbol{y}_i - \delta \mathbf{1}_{2P}}{\phi} - d_i \boldsymbol{h} - \tau_i \boldsymbol{g})$ and $\Gamma_0(\phi, \delta, \boldsymbol{d}, \boldsymbol{y}) = \prod_{i=K+1}^{N} f_{\boldsymbol{v}_i}(\frac{\boldsymbol{y}_i - \delta \mathbf{1}_{2P}}{\phi} - d_i \boldsymbol{h})$. The NMSE performance of $\hat{\delta}_{opt}(\boldsymbol{y})$ and $\hat{\phi}_{opt}(\boldsymbol{y})$ give us lower bounds on the NMSE for clock skew and offset estimators in the presence of possible unknown path asymmetries.

Theorem 6.1 provides us with mathematical expressions for the optimum joint clock skew and offset estimators for IEEE 1588 under the assumption that we have complete knowledge of $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$ and prior knowledge of the master-slave communication paths having unknown deterministic path asymmetries. The NMSE performance of the optimum estimators described in Theorem 6.1 cannot generally be achieved unless we have information that

is usually not available (prior information regarding the master-slave communication paths having unknown deterministic path delay asymmetries). Hence, the NMSE performance of the optimum joint clock skew and offset estimator described in Theorem 6.1 can be viewed as a performance lower bound. In summary, the NMSE performance of the optimum joint clock skew and offset estimator described in Theorem 6.1 can be viewed as a performance lower bound on the NMSE for invariant joint clock skew and offset estimation schemes which were designed without knowledge of the asymmetric paths and $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$. We can compare the performance of any proposed estimator with limited information to this best possible performance. If close, the compared estimators are well suited for practical applications.

6.3 Robust Joint Clock Skew and Offset Estimation for IEEE 1588

In practice, when developing joint clock skew and offset estimators that are robust to unknown path asymmetries, prior information regarding the paths with unknown path asymmetries is not available. Hence, it is useful to identify the asymmetric paths when developing a joint clock skew and offset estimation scheme that is robust against unknown path asymmetries. Further, in specific scenarios, the complete information regarding the pdf of the PDV, $f_{1i}(\cdot)$ and $f_{2i}(\cdot)$ for $i = 1, 2, \ldots, N$, might not be readily available. To address this issue, the pdf of the PDV can be approximated by a finite mixture of Gaussian distributions [41]. The Gaussian mixture distribution is a prominent model for approximating a pdf as it is a universal approximator in a certain sense [6, 37, 59]. We have

$$f_{1i}(w) = \sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik},\sigma_{1ik}}(w); \tag{6.16}$$

$$f_{2i}(w) = \sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2il},\sigma_{2il}}(w), \tag{6.17}$$

for $i = 1, 2, \ldots, N$. In (6.16) and (6.17), $\{\alpha_{ik}\}_{k=1}^{M_i}$ and $\{\beta_{il}\}_{l=1}^{L_i}$ denote the unknown mixing coefficients in the forward and reverse path at

the ith master-slave communication path with M_i and L_i denoting the number of assumed mixture components in the forward and reverse path, respectively. Also, we have $\alpha_{ik} \in [0,1]$ and $\beta_{il} \in [0,1]$ with the constraints $\sum_{k=1}^{M_i} \alpha_{ik} = 1$ and $\sum_{l=1}^{L_i} \beta_{il} = 1$. Further, $\mathcal{P}_{\mu,\sigma}(\cdot)$ denotes a normal distribution with mean μ and standard deviation σ. The variables $\{\mu_{1ik}, \sigma_{1ik}\}$ denote the mean and standard deviation of the kth component in the mixture models in the ith forward path and the variables $\{\mu_{2il}, \sigma_{2il}\}$ denote the mean and standard deviation of the lth component in the mixture models in the ith reverse path. Further, we assume a set of samples $\tilde{\boldsymbol{w}}_k = [\tilde{\boldsymbol{w}}_{k1}, \tilde{\boldsymbol{w}}_{k2}, \ldots, \tilde{\boldsymbol{w}}_{kN}]$ for $k = 1, 2$, where $\tilde{\boldsymbol{w}}_{ki} = [\tilde{w}_{ki1}, \tilde{w}_{ki2}, \ldots, \tilde{w}_{kiP_t}]$ for $i = 1, 2, \ldots, N$ to be available [41]. These samples have similar statistical properties to $\boldsymbol{w}_k$ for $k = 1, 2$.

Let $\boldsymbol{\Omega}$ denote the vector of unknown parameters defined as $\boldsymbol{\Omega} = [\boldsymbol{\Psi}, \boldsymbol{\eta}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, \boldsymbol{\mu}_{11}, \ldots, \boldsymbol{\mu}_{1N}, \boldsymbol{\sigma}_{11}, \ldots, \boldsymbol{\sigma}_{1N}, \boldsymbol{\mu}_{21}, \ldots, \boldsymbol{\mu}_{2N}, \boldsymbol{\sigma}_{21}, \ldots, \boldsymbol{\sigma}_{2N}]$ where we have $\boldsymbol{\Psi} = [\phi, \delta, d_1, \ldots, d_N, \tau_1, \ldots, \tau_N]$, $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_N]$, $\boldsymbol{\alpha}_i = [\alpha_{i1}, \ldots, \alpha_{iM_i}]$ for $i = 1, 2, \ldots, N$, $\boldsymbol{\beta}_i = [\beta_{i1}, \ldots, \beta_{iL_i}]$ for $i = 1, 2, \ldots, N$, $\boldsymbol{\mu}_{1i} = [\mu_{11}, \ldots, \mu_{1M_i}]$ for $i = 1, 2, \ldots, N$, $\boldsymbol{\mu}_{2i} = [\mu_{21}, \ldots, \mu_{2L_i}]$ for $i = 1, 2, \ldots, N$, $\boldsymbol{\sigma}_{1i} = [\sigma_{1i1}, \ldots, \sigma_{1iM_i}]$ for $i = 1, 2, \ldots, N$ and $\boldsymbol{\sigma}_{2i} = [\sigma_{2i1}, \ldots, \sigma_{2iL_i}]$ for $i = 1, \ldots, N$.

Given $\boldsymbol{\Omega}$, the log-likelihood function of the observed data $\boldsymbol{t}$, $\tilde{\boldsymbol{w}}_1$ and $\tilde{\boldsymbol{w}}_2$, denoted by $\mathcal{L}(\boldsymbol{\Omega} \mid \boldsymbol{t}, \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2)$, is defined as

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{\Omega} \mid \boldsymbol{t}, \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2) \\
&\quad = \sum_{i=1}^{N} \sum_{j=1}^{P} \Bigg[\ln \Bigg(\eta_i \Bigg(\sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik}, \sigma_{1ik}} \left(\frac{t_{2ij} - \delta}{\phi} - d_i - \tau_i - t_{1ij} \right) \Bigg) \\
&\qquad \times \Bigg(\sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2il}, \sigma_{2il}} \left(t_{4ij} - d_i + \frac{\delta - t_{3ij}}{\phi} \right) \Bigg) \Bigg) \\
&\qquad + (1 - \eta_i) \Bigg(\sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik}, \sigma_{1ik}} \left(\frac{t_{2ij} - \delta}{\phi} - d_i - t_{1ij} \right) \Bigg) \\
&\qquad \times \Bigg(\sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2il}, \sigma_{2il}} \left(t_{4ij} - d_i + \frac{\delta - t_{3ij}}{\phi} \right) \Bigg) \Bigg) \Bigg]
\end{aligned}
$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{P_t}\ln\left(\sum_{k=1}^{M_i}\alpha_{ik}\mathcal{P}_{\mu_{1ik},\sigma_{1ik}}(\tilde{w}_{1ij})\right)$$
$$+\sum_{i=1}^{N}\sum_{j=1}^{P_t}\ln\left(\sum_{l=1}^{L_i}\beta_{il}\mathcal{P}_{\mu_{2il},\sigma_{2il}}(\tilde{w}_{2ij})\right)-2NM\ln\phi. \tag{6.18}$$

The maximum likelihood method is widely used and has many attractive features including consistency and asymptotic unbiasedness. The maximum likelihood estimate (MLE) of $\boldsymbol{\Omega}$, denoted by $\hat{\boldsymbol{\Omega}}_{mle}$, is obtained by solving the following constrained optimization problem

$$\hat{\boldsymbol{\Omega}}_{mle}=\arg\max_{\boldsymbol{\Omega}}\mathcal{L}(\boldsymbol{\Omega}\mid\boldsymbol{t},\tilde{\boldsymbol{w}}_1,\tilde{\boldsymbol{w}}_2) \tag{6.19}$$

$$\text{such that } \eta_i\in\{0,1\}\quad \text{for } i=1,2,\ldots,N, \tag{6.19a}$$
$$\alpha_{ik}\in[0,1]\quad \text{for } i=1,2,\ldots,N$$
$$\text{and } k=1,2,\ldots,M \text{ with } \sum_{k=1}^{M_i}\alpha_{ik}=1, \tag{6.19b}$$
$$\beta_{il}\in[0,1] \text{ for } i=1,2,\ldots,N$$
$$\text{and } l=1,2,\ldots,M \text{ with } \sum_{l=1}^{L_i}\beta_{il}=1, \tag{6.19c}$$
$$|\tau_i|\geq d_\tau \text{ when } \eta_i=1, \tag{6.19d}$$
$$\sum_{i=1}^{N}\eta_i\leq N/2. \tag{6.19e}$$

The mixed integer nonlinear programming problem presented in (6.19) is computationally intensive to solve for large values of N as we would have to generally search across 2^N possibilities of $\boldsymbol{\eta}$. To address this issue, we use the idea discussed in [75] to formulate a relaxed version of the mixed integer non-linear programming problem.

6.3.1 Binary Variable Relaxation and EM Algorithm

As the constraints in (6.19a) correspond to binary variables, we relax the problem and introduce real variables with constraints defined as $\pi_i=\Pr(\eta_i=1)\in(0,1)$ for $i=1,2,\ldots,N$. Let $\boldsymbol{\Omega}_\pi=[\boldsymbol{\Psi},\boldsymbol{\pi},\boldsymbol{\alpha}_1,\ldots,$

$\boldsymbol{\alpha}_N, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, \boldsymbol{\mu}_{11}, \ldots, \boldsymbol{\mu}_{1N}, \boldsymbol{\sigma}_{11}, \ldots, \boldsymbol{\sigma}_{1N}, \boldsymbol{\mu}_{21}, \ldots, \boldsymbol{\mu}_{2N}, \boldsymbol{\sigma}_{21}, \ldots, \boldsymbol{\sigma}_{2N}]$, where we have $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_N]$. Replacing the binary variables with the corresponding real variables and dropping the constraints in (6.19d) and (6.19e), we can rewrite the optimization problem in (6.19) as

$$\hat{\boldsymbol{\Omega}}_{\pi,mle} = \arg\max_{\boldsymbol{\Omega}_\pi} \mathcal{L}_{EM}(\boldsymbol{\Omega}_\pi \mid \boldsymbol{t}, \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2) \tag{6.20}$$

$$\text{such that } \pi_i \in (0,1) \text{ for } i = 1, 2, \ldots, N, \tag{6.20a}$$

$$\alpha_{ik} \in (0,1) \text{ with } \sum_{k=1}^{M_i} \alpha_{ik} = 1 \quad \text{for } i = 1, 2, \ldots, N, \tag{6.20b}$$

$$\beta_{il} \in (0,1) \text{ with } \sum_{l=1}^{L_i} \beta_{il} = 1 \quad \text{for } i = 1, 2, \ldots, N, \tag{6.20c}$$

where $\hat{\boldsymbol{\Omega}}_{\pi,mle}$ denotes the MLE of $\boldsymbol{\Omega}_\pi$ and $\mathcal{L}_{EM}(\boldsymbol{\Omega}_\pi \mid \boldsymbol{t}, \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2)$, referred to as the incomplete log-likelihood is defined as

$$\begin{aligned}
&\mathcal{L}_{EM}(\boldsymbol{\Omega}_\pi \mid \boldsymbol{t}, \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{P} \ln \Bigg[\pi_i \Bigg(\Bigg(\sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik}, \sigma_{1ik}} \Big(\frac{t_{2ij} - \delta}{\phi} - d_i - \tau_i - t_{1ij} \Big) \Bigg) \\
&\quad \times \Bigg(\sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2il}, \sigma_{2il}} \Big(t_{4ij} - d_i + \frac{\delta - t_{3ij}}{\phi} \Big) \Bigg) \Bigg) \\
&\quad + (1 - \pi_i) \Bigg(\Bigg(\sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik}, \sigma_{1ik}} \Big(\frac{t_{2ij} - \delta}{\phi} - d_i - t_{1ij} \Big) \Bigg) \\
&\quad \times \Bigg(\sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2il}, \sigma_{2il}} \Big(t_{4ij} - d_i + \frac{\delta - t_{3ij}}{\phi} \Big) \Bigg) \Bigg) \Bigg] \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{P_t} \ln \Bigg[\sum_{k=1}^{M_i} \alpha_{ik} \mathcal{P}_{\mu_{1ik}, \sigma_{1ik}} (\tilde{w}_{1ij}) \Bigg] \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{P_t} \ln \Bigg[\sum_{l=1}^{L_i} \beta_{il} \mathcal{P}_{\mu_{2ik}, \sigma_{2ik}} (\tilde{w}_{2ij}) \Bigg] - 2NM \ln \phi.
\end{aligned} \tag{6.21}$$

The iterative algorithm for solving (6.20) is next enumerated in steps (1)–(16). The SAGE algorithm proposed in [15] is used to derive steps (1)–(16), and the details are discussed in [41]. The algorithm begins with the current estimates $\hat{\mathbf{\Omega}}'_\pi$ of $\mathbf{\Omega}_\pi$ and produces updated estimates of $\mathbf{\Omega}_\pi$ as follows:

1. In this step, we calculate the variables D_{ij}, $\chi^{(1)}_{ijkl}$ and $\chi^{(0)}_{ijkl}$ based on the current parameter estimates, $\hat{\mathbf{\Omega}}'_\pi$, and the observed timestamps. These variables are necessary for calculating the updated estimates of the parameters in $\mathbf{\Omega}_\pi$. Define D_{ij} as

$$\begin{aligned} D_{ij} = \sum_{k_c=1}^{M_i}\sum_{l_c=1}^{L_i}\Bigg[& \hat{\pi}'_i\hat{\alpha}'_{ik_c}\hat{\beta}'_{il_c}\mathcal{P}_{\mu'_{2il_c},\sigma'_{2il_c}}\left(t_{4ij}-\hat{d}'_i+\frac{\hat{\delta}'-t_{3ij}}{\hat{\phi}'}\right) \\ & \times\mathcal{P}_{\mu'_{1ik_c},\sigma'_{1ik_c}}\left(\frac{t_{2ij}-\hat{\delta}'}{\hat{\phi}'}-\hat{d}'_i-\hat{\tau}'_i-t_{1ij}\right) \\ & +(1-\hat{\pi}'_i)\hat{\alpha}'_{ik_c}\hat{\beta}'_{il_c}\mathcal{P}_{\mu'_{2il_c},\sigma'_{2il_c}}\left(t_{4ij}-\hat{d}'_i+\frac{\hat{\delta}'-t_{3ij}}{\hat{\phi}'}\right) \\ & \times\mathcal{P}_{\mu'_{1ik_c},\sigma'_{1ik_c}}\left(\frac{t_{2ij}-\hat{\delta}'}{\hat{\phi}'}-\hat{d}'_i-t_{1ij}\right)\Bigg] \end{aligned} \tag{6.22}$$

for $i=1,2,\ldots,N$ and $j=1,2,\ldots,P$. Then, compute

$$\begin{aligned} \chi^{(1)}_{ijkl} = D^{-1}_{ij}\hat{\pi}'_i\hat{\alpha}'_{ik}\hat{\beta}'_{il}\mathcal{P}_{\mu'_{2il},\sigma'_{2il}}\left(t_{4ij}-\hat{d}'_i+\frac{\hat{\delta}'-t_{3ij}}{\hat{\phi}'}\right) \\ \times\mathcal{P}_{\mu'_{1ik},\sigma'_{1ik}}\left(\frac{t_{2ij}-\hat{\delta}'}{\hat{\phi}'}-\hat{d}'_i-\hat{\tau}'_i-t_{1ij}\right) \end{aligned} \tag{6.23}$$

and

$$\begin{aligned} \chi^{(0)}_{ijkl} = D^{-1}_{ij}(1-\hat{\pi}'_i)\hat{\alpha}'_{ik}\hat{\beta}'_{il}\mathcal{P}_{\mu'_{2il},\sigma'_{2il}}\left(t_{4ij}-\hat{d}'_i+\frac{\hat{\delta}'-t_{3ij}}{\hat{\phi}'}\right) \\ \times\mathcal{P}_{\mu'_{1ik},\sigma'_{1ik}}\left(\frac{t_{2ij}-\hat{\delta}'}{\hat{\phi}'}-\hat{d}'_i-t_{1ij}\right) \end{aligned} \tag{6.24}$$

for $i=1,2,\ldots,N$, $j=1,2,\ldots,P$, $k=1,2,\ldots,M_i$ and $l=1,2,\ldots,L_i$.

2. Similar to the first step, we calculate the variables $\tilde{D}_{ij}$ and $\tilde{a}_{ijkl}$ based on the current parameter estimates, $\hat{\mathbf{\Omega}}'_{\pi}$, and the observed timestamps. These variables are necessary for calculating the updated estimates of parameters in $\mathbf{\Omega}_{\pi}$. First, we calculate $\tilde{D}_{ij} = \sum_{k_c=1}^{M_i}\sum_{l_c=1}^{L_i}\hat{\alpha}'_{ik_c}\hat{\beta}'_{il_c}\mathcal{P}_{\mu'_{2ik_c},\sigma'_{2ik_c}}(\tilde{w}_{1ij})\mathcal{P}_{\mu'_{2il_c},\sigma'_{2il_c}}(\tilde{w}_{2ij})$ for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, P_t$. We then compute

$$\tilde{a}_{ijkl} = \tilde{D}_{ij}^{-1}\hat{\alpha}'_{ik}\hat{\beta}'_{il}\mathcal{P}_{\mu'_{2ik},\sigma'_{2ik}}(\tilde{w}_{1ij})\mathcal{P}_{\mu'_{2il},\sigma'_{2il}}(\tilde{w}_{2ij}) \tag{6.25}$$

for $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, P_t$, $k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$.

3. In this step, we calculate the updated estimate of π_i, α_{ik} and β_{il}, denoted by π'_i, α'_{ik} and β'_{il}, respectively, for $i = 1, 2, \ldots, N, k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$. We have

$$\hat{\pi}_i = \frac{1}{P}\sum_{j=1}^{P}\sum_{k=1}^{M_i}\sum_{l=1}^{L_i}\chi^{(1)}_{ijkl}, \tag{6.26}$$

$$\hat{\alpha}_{ik} = \frac{1}{(P+P_t)}\left[\sum_{j=1}^{P}\sum_{l=1}^{L_i}\left(\chi^{(1)}_{ijkl}+\chi^{(0)}_{ijkl}\right)+\sum_{j=1}^{P_t}\sum_{l=1}^{L_i}\tilde{a}_{ikjl}\right], \tag{6.27}$$

$$\hat{\beta}_{il} = \frac{1}{(P+P_t)}\left[\sum_{j=1}^{P}\sum_{k=1}^{M_i}\left(\chi^{(1)}_{ijkl}+\chi^{(0)}_{ijkl}\right)+\sum_{j=1}^{P_t}\sum_{k=1}^{M_i}\tilde{a}_{ikjl}\right]. \tag{6.28}$$

4. In this step, we update the current estimates of π_i, α_{ik} and β_{il} with the estimates obtained from step (3) and we recompute the variables in steps (1) and (2). Set $\hat{\pi}'_i = \hat{\pi}_i$, $\hat{\alpha}'_{ik} = \hat{\alpha}_{ik}$ and $\hat{\beta}'_{il} = \hat{\beta}_{il}$ for $i = 1, 2, \ldots, N, k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$ and recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi^{(1)}_{ijkl}$, $\chi^{(0)}_{ijkl}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

5. In this step, we calculate the updated estimates of μ_{1ik} and μ_{2il}, denoted by μ'_{1ik} and μ'_{2il}, respectively, for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, \ldots, L_i$. Define

$$D_{\mu_1,ik} = \sum_{j=1}^{P}\sum_{l=1}^{L_i}\left(\chi^{(1)}_{ijkl}+\chi^{(0)}_{ijkl}\right)+\sum_{j=1}^{P_t}\sum_{l=1}^{L_i}\tilde{a}_{ijkl}, \tag{6.29}$$

and

$$D_{\mu_2,il} = \sum_{j=1}^{P}\sum_{k=1}^{M_i}\left(\chi_{ijkl}^{(1)} + \chi_{ijkl}^{(0)}\right) + \sum_{j=1}^{P_t}\sum_{k=1}^{M_i}\tilde{a}_{ijkl}, \tag{6.30}$$

for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, \ldots, L_i$. Then compute

$$\hat{\mu}_{1ik} = D_{\mu_1,ik}^{-1}\left[\sum_{j=1}^{P}\sum_{l=1}^{L_i}\left(\left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - \hat{d}_i' - t_{1ij}\right)\right.\right.$$
$$\left.\left.\times\left(\chi_{ijkl}^{(1)} + \chi_{ijkl}^{(0)}\right) - \chi_{ijkl}^{(1)}\hat{\tau}_i'\right) + \sum_{j=1}^{P_t}\sum_{l=1}^{L_i}\tilde{a}_{ijkl}\tilde{w}_{1ij}\right], \tag{6.31}$$

and

$$\hat{\mu}_{2il} = D_{\mu_2,il}^{-1}\left[\sum_{j=1}^{P}\sum_{k=1}^{M_i}\left(\chi_{ijkl}^{(1)} + \chi_{ijkl}^{(0)}\right)\right.$$
$$\left.\times\left(t_{4ij} - \hat{d}_i' + \frac{\hat{\delta}' - t_{3ij}}{\hat{\phi}'}\right) + \sum_{j=1}^{P_t}\sum_{k=1}^{M_i}\tilde{a}_{ijkl}\tilde{w}_{2ij}\right] \tag{6.32}$$

for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, \ldots, L_i$.

6. In this step, we update the current estimates of μ_{1ik} and μ_{2il} with the estimates obtained from step (5) and we recompute the variables in steps (1) and (2). Set $\hat{\mu}_{1ik}' = \hat{\mu}_{1ik}$ and $\hat{\mu}_{2il}' = \hat{\mu}_{2il}$ for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$. Recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi_{ijkl}^{(1)}$, $\chi_{ijkl}^{(0)}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

7. In this step, we calculate the updated estimates of σ_{1ik}^2 and σ_{2il}^2, denoted by $\hat{\sigma}_{1ik}^2$ and $\hat{\sigma}_{2il}^2$, respectively, for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$. Define

$$D_{\mu_1,ik} = \sum_{j=1}^{P}\sum_{l=1}^{L_i}\left(\chi_{ijkl}^{(1)} + \chi_{ijkl}^{(0)}\right) + \sum_{j=1}^{P_t}\sum_{l=1}^{L_i}\tilde{a}_{ijkl}, \tag{6.33}$$

and

$$D_{\mu_2,il} = \sum_{j=1}^{P}\sum_{k=1}^{M_i}\left(\chi_{ijkl}^{(1)} + \chi_{ijkl}^{(0)}\right) + \sum_{j=1}^{P_t}\sum_{k=1}^{M_i}\tilde{a}_{ijkl} \tag{6.34}$$

for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, \ldots, L_i$. Then compute

$$\hat{\sigma}^2_{1ik} = D^{-1}_{\mu_1,ik} \left[\sum_{j=1}^{P_t} \sum_{l=1}^{L_i} \tilde{a}_{ijkl} (\tilde{w}_{1ij} - \hat{\mu}'_{1ik})^2 + \sum_{j=1}^{P} \sum_{l=1}^{L_i} \chi^{(0)}_{ijkl} \left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - \hat{d}'_i - t_{1ij} - \hat{\mu}'_{1ik} \right)^2 + \chi^{(1)}_{ijkl} \left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - \hat{d}'_i - \hat{\tau}'_i - t_{1ij} - \hat{\mu}'_{1ik} \right)^2 \right], \quad (6.35)$$

and

$$\hat{\sigma}^2_{2il} = D^{-1}_{\mu_2,il} \left[\sum_{j=1}^{P_t} \sum_{k=1}^{M_i} \tilde{a}_{ijkl} (\tilde{w}_{2ij} - \hat{\mu}'_{2il})^2 + \sum_{j=1}^{P} \sum_{k=1}^{M_i} \left(\chi^{(1)}_{ijkl} + \chi^{(0)}_{ijkl} \right) \times \left(t_{4ij} - \hat{d}'_i + \frac{\hat{\delta}' - t_{3ij}}{\hat{\phi}'} - \hat{\mu}'_{2il} \right)^2 \right] \quad (6.36)$$

for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, \ldots, L_i$.

8. In this step, we update the current estimates of σ^2_{1ik} and σ^2_{2il} with the estimates obtained from step (7) and we recompute the variables in steps (1) and (2). Set $\hat{\sigma}'^2_{1ik} = \hat{\sigma}^2_{1ik}$ and $\hat{\sigma}'^2_{2il} = \hat{\sigma}^2_{2il}$ for $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, M_i$ and $l = 1, 2, \ldots, L_i$. Recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi^{(1)}_{ijkl}$, $\chi^{(0)}_{ijkl}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

9. In this step, we calculate the updated estimates of d_i, denoted by $\hat{d}_i$, for the various master-slave communication paths. Define $D_{d,i} = \sum_{j=1}^{P} \sum_{k=1}^{M_i} \sum_{l=1}^{L_i} (\chi^{(1)}_{ijkl} + \chi^{(0)}_{ijkl}) (\frac{1}{\sigma'^2_{1ik}} + \frac{1}{\sigma'^2_{2il}})$ for

$i = 1, 2, \ldots, N$. Then compute

$$\hat{d}_i = D_{d,i}^{-1} \left[\sum_{j=1}^{P} \sum_{k=1}^{M_i} \sum_{l=1}^{L_i} \left(\chi_{ijkl}^{(1)} \left[\frac{\left(\frac{\hat{\delta}' - t_{3ij}}{\hat{\phi}'} + t_{4ij} - \hat{\mu}'_{2il} \right)}{\sigma'^2_{2il}} + \frac{\left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - \hat{\tau}'_i - t_{1ij} - \hat{\mu}'_{1ik} \right)}{\sigma'^2_{1ik}} \right] + \chi_{ijkl}^{(0)} \left[\frac{\left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - t_{1ij} - \hat{\mu}'_{1ik} \right)}{\sigma'^2_{1ik}} + \frac{\left(\frac{\hat{\delta}' - t_{3ij}}{\hat{\phi}'} + t_{4ij} - \hat{\mu}'_{2il} \right)}{\sigma'^2_{2il}} \right] \right) \right] \tag{6.37}$$

for $i = 1, 2, \ldots, N$.

10. In this step, we update the current estimates of d_i using the estimates obtained from step (9) and we recompute the variables in steps (1) and (2). Set $\hat{d}'_i = \hat{d}_i$ for $i = 1, 2, \ldots, N$. Recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi_{ijkl}^{(1)}$, $\chi_{ijkl}^{(0)}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

11. In this step, we calculate the updated estimates of τ_i, denoted by $\hat{\tau}_i$, for the various master-slave communication paths. Define $D_{\tau,i} = \sum_{j=1}^{P} \sum_{k=1}^{M_i} \sum_{l=1}^{L_i} \frac{\chi_{ijkl}^{(1)}}{\sigma'^2_{1ik}}$ for $i = 1, 2, \ldots, N$. Then compute

$$\hat{\tau}_i = D_{\tau,i}^{-1} \left[\sum_{j=1}^{P} \sum_{k=1}^{M_i} \sum_{l=1}^{L_i} \chi_{ijkl}^{(1)} \left(\frac{\left(\frac{t_{2ij} - \hat{\delta}'}{\hat{\phi}'} - \hat{d}'_i - t_{1ij} - \hat{\mu}'_{1ik} \right)}{\sigma'^2_{1ik}} \right) \right] \tag{6.38}$$

for $i = 1, 2, \ldots, N$.

12. In this step, we update the current estimates of τ_i using the estimates obtained from step (11) and we recompute the variables in steps (1) and (2). Set $\hat{\tau}'_i = \hat{\tau}_i$ for $i = 1, 2, \ldots, N$. Recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi_{ijkl}^{(1)}$, $\chi_{ijkl}^{(0)}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

13. In this step, we calculate the updated estimate of the clock offset δ, denoted by $\hat{\delta}$. Define $D_\delta = \sum_{i=1}^{N}\sum_{j=1}^{P}\sum_{k=1}^{M_i}\sum_{l=1}^{L_i}(\chi_{ijkl}^{(1)}+\chi_{ijkl}^{(0)})(\frac{1}{\sigma_{1ik}'^2}+\frac{1}{\sigma_{2il}'^2})$ and compute

$$\hat{\delta} = \hat{\phi}' D_\delta^{-1}\left[\sum_{i=1}^{N}\sum_{j=1}^{P}\sum_{k=1}^{M_i}\sum_{l=1}^{L_i}\left(\chi_{ijkl}^{(1)}\left[\frac{\left(\frac{t_{2ij}}{\hat{\phi}'}-\hat{d}_i'-\hat{\tau}_i'-t_{1ij}-\hat{\mu}_{1ik}'\right)}{\sigma_{1ik}'^2} - \frac{\left(t_{4ij}-\frac{t_{3ij}}{\hat{\phi}'}-\hat{d}_i'-\hat{\mu}_{2il}'\right)}{\sigma_{2il}'^2}\right] + \chi_{ijkl}^{(0)}\left[\frac{\left(\frac{t_{2ij}}{\hat{\phi}'}-\hat{d}_i'-t_{1ij}-\hat{\mu}_{1ik}'\right)}{\sigma_{1ik}'^2} - \frac{\left(t_{4ij}-\frac{t_{3ij}}{\hat{\phi}'}-\hat{d}_i'-\hat{\mu}_{2il}'\right)}{\sigma_{2il}'^2}\right]\right)\right]. \tag{6.39}$$

14. In this step, we update the current estimates of δ using $\hat{\delta}$ obtained from step (13) and we recompute the variables in steps (1) and (2). Set $\hat{\delta}' = \hat{\delta}$. Recompute D_{ij} and $\tilde{D}_{ij}$. Then recompute $\chi_{ijkl}^{(1)}$, $\chi_{ijkl}^{(0)}$ and $\tilde{a}_{ijkl}$ using (6.23), (6.24) and (6.25), respectively.

15. In this step, we calculate the updated estimate of the clock skew ϕ, denoted by $\hat{\phi}$. Define $c_\phi = \sum_{i=1}^{N}\sum_{j=1}^{P}\sum_{k=1}^{M_i}\sum_{l=1}^{L_i}(\chi_{ijkl}^{(1)}+\chi_{ijkl}^{(0)})(\frac{(t_{2ij}-\hat{\delta}')^2}{\sigma_{1ik}'^2}+\frac{(\hat{\delta}'-t_{3ij})^2}{\sigma_{2il}'^2})$, $a_\phi = 2NP$ and b_ϕ as

$$\begin{aligned} b_\phi = &\sum_{i=1}^{N}\sum_{j=1}^{P}\sum_{k=1}^{M_i}\sum_{l=1}^{L_i}\chi_{ijkl}^{(1)}\left[\frac{(\hat{d}_i'+\hat{\tau}_i'+t_{1ij}+\hat{\mu}_{1ik}')(t_{2ij}-\hat{\delta}')}{\sigma_{1ik}'^2} - \frac{(t_{4ij}-\hat{d}_i'-\hat{\mu}_{2il}')(\hat{\delta}'-t_{3ij})}{\sigma_{2il}'^2}\right] \\ &+\chi_{ijkl}^{(0)}\left[\frac{(\hat{d}_i'+t_{1ij}+\hat{\mu}_{1ik}')(t_{2ij}-\hat{\delta}')}{\sigma_{1ik}'^2} - \frac{(t_{4ij}-\hat{d}_i'-\hat{\mu}_{2il}')(\hat{\delta}'-t_{3ij})}{\sigma_{2il}'^2}\right]. \end{aligned} \tag{6.40}$$

Then compute $\hat{\phi} = \frac{\sqrt{b_\phi^2-4a_\phi c_\phi}-b_\phi}{2a_\phi}$.

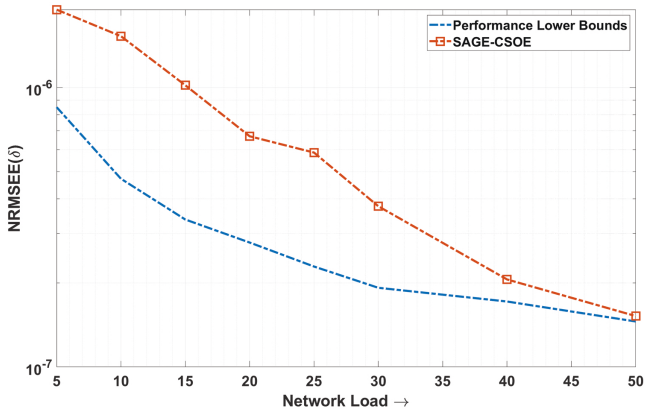
16. In this step, we update the current estimates of ϕ using $\hat{\phi}$ obtained from step (15) and we recompute the variables in steps (1) and (2). Set $\hat{\phi}' = \hat{\phi}$, and repeat steps (1)–(16).

Since the update equations in steps (1)–(16) employ the SAGE algorithm, they inherit the desirable property that the likelihood is non-decreasing at each iteration [15]. When the algorithm converges, we obtain the estimate of the clock skew and offset from $\mathbf{\Omega}'_{\pi}$. Initial values for the parameters are required to begin the SAGE algorithm. A simple ad-hoc scheme to obtain the initial values of the various parameters in $\mathbf{\Omega}_{\pi}$ is presented in [41].

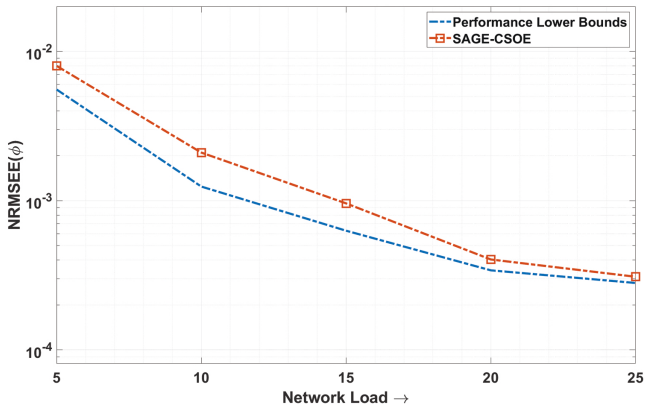
6.4 Numerical Results

We now compare the skew normalized root mean square estimation error (NRMSE) of the robust CSOE scheme (SAGE-CSOE) proposed in [41] against the performance lower bounds obtained using the NRMSE of the optimum estimators in the LTE backhaul network scenario described in Section 2. To obtain the pdf of the stochastic queuing delays for this scenario, please refer to Subsection 2.3.3.

The results are presented in Figures 6.1 and 6.2. From Figure 6.1, we see that the SAGE-CSOE scheme developed in [41] exhibits an NRMSE close to the performance bounds for the considered TM1 network scenario. Also, from Figure 6.2, we observe that for a sufficiently large number of two-way message exchanges, the SAGE-CSOE exhibits an NRMSE close to the lower bounds for several network scenarios indicating the relative robustness of the developed approach for different network scenarios.

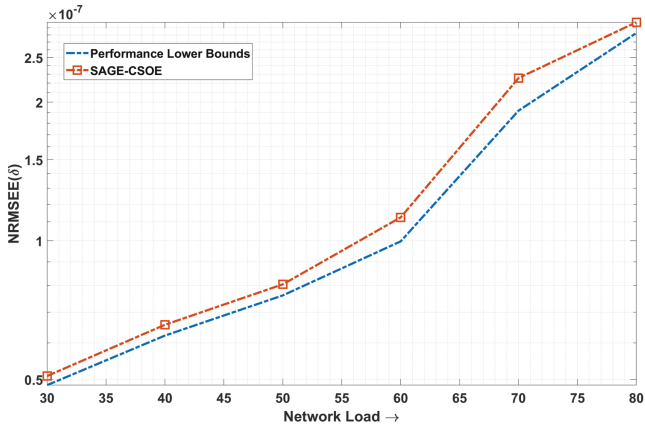


(a) NRMSE of clock offset for TM-1 under 60% load.

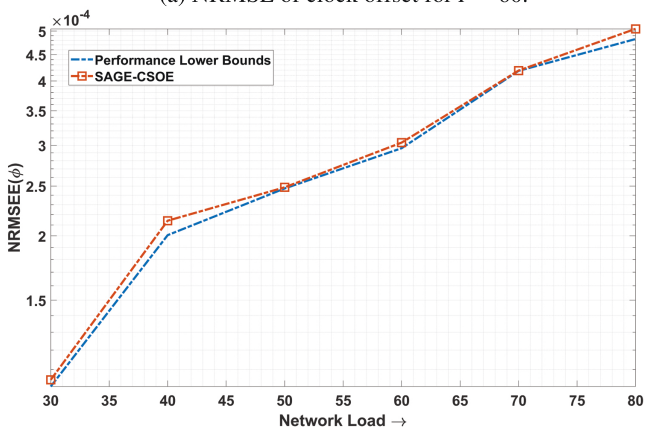


(b) NRMSE of clock skew for TM-1 under 60% load.

Figure 6.1: NRMSE of clock skew and offset for the considered CSOE schemes under traffic model-1.



(a) NRMSE of clock offset for $P = 60$.



(b) NRMSE of clock offset for $P = 25$.

Figure 6.2: NRMSE of clock offset and skew for the considered CSOE schemes under traffic model-1 under different loads.

7

Conclusions

In this monograph, we have presented an overview of recent developments in the field of clock synchronization techniques for packet-switched networks. Here we briefly present some concluding remarks on the material presented in Sections 2 through to 6. In Section 2, we presented the popular models to model the clock time at the slave node in terms of clock time at the master node. We also described the two-way message exchange in the context of PTP and presented the popular pdf-models for modeling the stochastic queuing delays in packet-switched networks. From Sections 4 through to 6, we have used invariant decision theory to develop the optimum invariant estimators for PTP. To this end, in Section 3, we formalized the concept of invariance by defining groups of transformations over parameter and observation spaces.

In Section 4, we presented the recently developed optimum clock offset estimators under various observation models for PTP from [21]. The optimum estimators achieve the lowest mean square estimation error among the class of invariant clock offset estimators and are also minimax-optimum. These estimators are guaranteed to provide the best possible performance under any network scenario for jointly estimating the clock skew and offset. Hence, these estimators are suitable for

obtaining lower bounds on achievable estimation performance. Also, in Section 4, we presented the optimum L-estimators for estimating the clock offset in PTP. The L-estimators have significantly lower computational complexity than the optimum estimators while exhibiting a mean square estimation error very close to the lower bounds. In Section 5, we addressed the problem of joint clock skew and offset estimation for PTP in the presence of stochastic queuing delays. We presented the optimum clock skew and offset estimators under various observation models for PTP from [40]. Similar to the optimum clock offset estimators, the optimum clock skew and offset estimators are guaranteed to provide the best possible performance under any network scenario.

When developing the optimum estimators in Sections 4 and 5, we assumed either the complete knowledge of the fixed path delays or a prior known affine relationship between the delays. However, the presence of an unknown asymmetry between the fixed path delays can significantly degrade the performance of clock skew and offset estimation schemes [68]. This unknown asymmetry between the fixed path delays can arise from several sources, including delay attacks [68] and routing asymmetry [3]. Section 6 addresses the problem of developing clock skew and offset estimators for PTP that are robust against unknown path asymmetries. We present useful performance bounds on the mean square estimation error of a clock skew and offset estimator in the presence of unknown path asymmetries. We also describe a robust clock skew and offset estimator that exhibits a mean square estimation error close to the bounds for a sufficiently large number of two-way message exchanges. The presented approaches from Sections 3 through to 6 are very general as they do not assume any particular pdf-model of the stochastic queuing delays. Also, the presented approaches apply to any clock synchronization based on the two-way message exchange, including TPSN [20], tiny-sync, and mini-sync [62], and LTS [70]. Further, these estimators can be easily modified for clock synchronization protocols based on message exchanges including RBS [10] and the flooding time synchronization protocol [50].

Acknowledgments

The work was supported by the Department of Energy under Grant Number DE-OE0000779, the U. S. Army Research Laboratory and the U. S. Army Research Office under Grant Number W911NF-17-1-0331, the National Science Foundation under Grant ECCS-1744129, and a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA).

References

[1] Abdel-Ghaffar, H. S. (2002). "Analysis of synchronization algorithms with time-out control over networks with exponentially symmetric delays". *IEEE Transactions on Communications*. 50(10): 1652–1661.

[2] Anyaegbu, M., C. Wang, and W. Berrie (2012). "A sample-mode packet delay variation filter for IEEE 1588 synchronization". In: *12th International Conference on ITS Telecommunications*. 1–6. DOI: 10.1109/ITST.2012.6425166.

[3] Balakrishnan, H., V. Padmanabhan, G. Fairhurst, and M. Sooriyabandara (2002). "TCP performance implications of network path asymmetry". *Tech. rep.* No. RFC 3449.

[4] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd edition. Springer.

[5] Blewitt, G. (1997). "Basics of the GPS technique: Observation equations". In: *Geodetic Applications of GPS*. Ed. by B. Johnson. Stockholm: Nordic Geodetic Commission. 10–54.

[6] Blum, R. S., R. J. Kozick, and B. M. Sadler (1999). "An adaptive spatial diversity receiver for non-Gaussian interference and noise". *IEEE Transactions on Signal Processing*. 47(8): 2100–2111.

[7] Chaudhari, Q. M., E. Serpedin, and K. Qaraqe (2008). "On maximum likelihood estimation of clock offset and skew in networks with exponential delays". *IEEE Transactions on Signal Processing*. 56(4): 1685–1697.

[8] Chaudhari, Q. M., E. Serpedin, and K. Qaraqe (2010). "On minimum variance unbiased estimation of clock offset in a two-way message exchange mechanism". *IEEE Transactions on Information Theory*. 56(6): 2893–2904.

[9] Draper, N. and H. Smith (1988). *Applied Regression Analysis*. 2nd edition. John Wiley & Sons, Inc.

[10] Elson, J., L. Girod, and D. Estrin (2002). "Fine-grained network time synchronization using reference broadcasts". *SIGOPS Oper. Syst. Rev.* 36(SI): 147–163.

[11] Elson, J., R. M. Karp, C. H. Papadimitriou, and S. Shenker (2004). "Global synchronization in sensornets". In: *LATIN 2004: Theoretical Informatics*. Ed. by M. Farach-Colton. Berlin, Heidelberg: Springer. 609–624.

[12] Etzlinger, B., B. F. Meyer, F. Hlawatsch, A. Springer, and H. Wymeersch (2016). "Cooperative simultaneous localization and synchronization in mobile agent networks". *IEEE Transactions in Signal Processing*. 65(14): 3587–3602.

[13] Etzlinger, B., B. F. Meyer, H. Wymeersch, F. Hlawatsch, and G. Muller (2014). "Cooperative simultaneous localization and synchronization: Towards a low-cost hardware implementation". In: *Proc. IEEE Sensor Array, Multichan. Signal Processing (SAM) Workshop*. Vol. b. 33–36.

[14] Etzlinger, B. and H. Wymeersch (2018). "Synchronization and localization in wireless networks". *Foundations and Trends in Signal Processing*. 12(1): 1–106.

[15] Fessler, J. A. and A. O. Hero (1994). "Space-alternating generalized expectation-maximization algorithm". *IEEE Transactions on Signal Processing*. 42(10): 2664–2677.

[16] Freris, N. M., S. R. Graham, and P. R. Kumar (2011). "Fundamental limits on synchronizing clocks over networks". *IEEE Transactions on Automatic Control*. 56(6): 1352–1364.

[17] Gaderer, G., P. Loschmidt, and T. Sauter (2010). "Improving fault tolerance in high-precision clock synchronization". *IEEE Transactions on Industrial Informatics*. 6(2): 206–215.

[18] Gaderer, G., T. Sauter, and G. Bumiller (2005). "Clock synchronization in powerline networks". In: *International Symposium on Power Line Communications and Its Applications, 2005.* 71–75. DOI: 10.1109/ISPLC.2005.1430468.

[19] Gandhi, P. P. and S. A. Kassam (1991). "Design and performance of combination filters for signal restoration". *IEEE Trans. on Signal Processing*. 39(7): 1524–1540.

[20] Ganeriwal, S., R. Kumar, and M. B. Srivastava (2003). "Timing-sync protocol for sensor networks". In: *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. New York, NY, USA: ACM. 138–149.

[21] Guruswamy, A., R. S. Blum, S. Kishore, and M. Bordogna (2015a). "Minimax optimum estimators for phase synchronization in IEEE 1588". *IEEE Transactions on Communications*. 63(9): 3350–3362.

[22] Guruswamy, A., R. S. Blum, S. Kishore, and M. Bordogna (2015b). "On the optimum design of L-estimators for phase offset estimation in IEEE 1588". *IEEE Transactions on Communications*. 63(12): 5101–5115.

[23] Guruswamy, A., R. S. Blum, S. Kishore, and M. Bordogna (2015c). "Performance lower bounds for phase offset estimation in IEEE 1588 synchronization". *IEEE Transactions on Communications*. 63(1): 243–253.

[24] Hadvzic, I. and D. R. Morgan (2010). "Adaptive packet selection for clock recovery". In: *IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*. 42–47. DOI: 10.1109/ISPCS.2010.5609775.

[25] Hadvzic, I., D. R. Morgan, and Z. Sayeed (2011). "A synchronization algorithm for packet MANs". *IEEE Transactions on Communications*. 59(4): 1142–1153.

[26] Han, L. (2017). "Synchronization requirements of 5G and corresponding solutions - China Mobile". *Tech. rep.*

[27] Han, L., L. Han, R. Duan, and G. Garner (2017). "Analysis of the synchronization requirements of 5G and corresponding solutions". *IEEE Communications Standards Magazine*. 1(1): 52–58.

[28] Han, L., H. Li, L. Wang, N. Hua, C. Hu, J. Wang, S. Liu, L. He, Z. Chen, and Y. Xu (2014). "First national high-precision time synchronization network with sub-microsecond accuracy over commercial optical networks for wireless applications". In: *Asia Communications and Photonics Conference*. AF4B.6.

[29] Harris, K. (2008). "An application of IEEE 1588 to industrial automation". In: *2008 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*. 71–76. DOI: 10.1109/ISPCS.2008.4659216.

[30] Huawei (2016). "IEEE 1588v2 technology white paper". *Tech. rep.*

[31] Huber, P. J. (1996). *Robust Statistical Procedures*. Philadelphia, PA, USA: SIAM.

[32] IEEE (2008). "IEEE standard for a precision clock synchronization protocol for networked measurement and control systems". *IEEE Std 1588–2008 (Revision of IEEE Std 1588-2002)*. July: 1–300. DOI: 10.1109/IEEESTD.2008.4579760.

[33] "IEEE Standard Profile for Use of IEEE 1588 Precision Time Protocol in Power System Applications" (2011). *IEEE Std. C37.238-2011*. July.

[34] Jeske, D. R. (2005). "On maximum-likelihood estimation of clock offset". *IEEE Transactions on Communications*. 53(1): 53–54.

[35] Jeske, D. R. and A. Sampath (2003). "Estimation of clock offset using bootstrap bias-correction techniques". *Technometrics*. 45: 256–261.

[36] Jiang, X., J. Zhang, B. J. Harding, J. J. Makela, and A. D. Dominguez-Garcia (2013). "Spoofing GPS receiver clock offset of phasor measurement units". *IEEE Transactions on Power Systems*. 28(3): 3253–3262.

[37] Kariya, T. and B. K. Sinha (1989). *Robustness of Statistical Tests*. New York, NY, USA: Academic.

[38] Karthik, A. K. and R. S. Blum (2018a). "Estimation theory-based robust phase offset determination in presence of possible path asymmetries". *IEEE Transactions on Communications*. 66(4): 1624–1635.

[39] Karthik, A. K. and R. S. Blum (2018b). "Robust phase offset estimation for IEEE 1588 PTP in electrical grid networks". In: *2018 IEEE Power Energy Society General Meeting (PESGM)*. 1–5. DOI: 10.1109/PESGM.2018.8586488.

[40] Karthik, A. K. and R. S. Blum (2019). "Optimum full information, unlimited complexity, invariant, and minimax clock skew and offset estimators for IEEE 1588". *IEEE Transactions on Communications*. 67(5): 3624–3637.

[41] Karthik, A. K. and R. S. Blum (n.d.). "Robust clock skew and offset estimation for IEEE 1588 in the presence of unexpected deterministic path delay asymmetries". *IEEE Transactions on Communications*. DOI: 10.1109/TCOMM.2020.2991212.

[42] Kim, J., J. Lee, E. Serpedin, and K. Qaraqe (2009). "A robust estimation scheme for clock phase offsets in wireless sensor networks in the presence of non-Gaussian random delays". *Eurasip Journal on Advances in Signal Processing*. 89: 1155–1161.

[43] Kim, J., J. Lee, E. Serpedin, and K. Qaraqe (2010). "A robust approach for clock offset estimation in wireless sensor networks". *Eurasip Journal on Advances in Signal Processing*. 2010: 1–8.

[44] Kim, J., J. Lee, E. Serpedin, and K. Qaraqe (2011). "Robust clock synchronization in wireless sensor networks through noise density estimation". *IEEE Transactions on Signal Processing*. 59(7): 3035–3047.

[45] Kim, T., C. S. Sin, and S. Lee (2012). "Analysis of effect of spoofing signal in GPS receiver". In: *2012 12th International Conference on Control, Automation and Systems*. 2083–2087.

[46] Kyriakakis, E., J. Sparsø, and M. Schoeberl (2018). "Hardware assisted clock synchronization with the IEEE 1588–2008 precision time protocol". In: *Proceedings of the 26th International Conference on Real-Time Networks and Systems. RTNS '18*. Chasseneuil-du-Poitou, France: ACM. 51–60. DOI: 10.1145/3273905.3273920.

[47] Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation*. 2nd edition. Springer Texts in Statistics.

[48] Leng, M. and Y. Wu (2010). "On clock synchronization algorithms for wireless sensor networks under unknown delay". *IEEE Transactions on Vehicular Technology*. 59(1): 182–190.

[49] Li, J. and D. R. Jeske (2009). "Maximum likelihood estimators of clock offset and skew under exponential delays". *Applied Stochastic Models in Business and Industry*. 25(4): 445–459.

[50] Maróti, M., B. Kusy, G. Simon, and Á. Lédeczi (2004). "The flooding time synchronization protocol". In: *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems. SenSys '04*. Baltimore, MD, USA: ACM. 39–49. DOI: 10.1145/1031495.1031501.

[51] MeinBerg (n.d.). "Time synchronization in electrical systems". https://www.meinbergglobal.com/english/info/time-synchronization-electrical-systems.htm.

[52] Mills, D. L. (1991). "Internet time synchronization: the network time protocol". *IEEE Transactions on Communications*. 39(10): 1482–1493.

[53] Mizrahi, T. (2012a). "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*. 1–6. DOI: 10.1109/ISPCS.2012.6336612.

[54] Mizrahi, T. (2012b). "Slave diversity: Using multiple paths to improve the accuracy of clock synchronization protocols". In: *2012 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication Proceedings*. 1–6. DOI: 10.1109/ISPCS.2012.6336621.

[55] Mochizuki, B. and I. Hadzic (2010). "Improving IEEE 1588v2 clock performance through controlled packet departures". *IEEE Communications Letters*. 14(5): 459–461.

[56] Murakami, T., Y. Horiuchi, and K. Nishimura (2011). "A packet filtering mechanism with a packet delay distribution estimation function for IEEE 1588 time synchronization in a congested network". In: *2011 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*. 114–119. DOI: 10.1109/ISPCS.2011.6070159.

[57] Noh, K., Q. M. Chaudhari, E. Serpedin, and B. W. Suter (2007). "Novel clock phase offset and skew estimation using two-way timing message exchanges for wireless sensor networks". *IEEE Transactions on Communications*. 55(4): 766–777.

[58] Pettyjohn, J. S., D. R. Jeske, and J. Li (2010). "Least squares-based estimation of relative clock offset and frequency in sensor networks with high latency". *IEEE Transactions on Communications*. 58(12): 3613–3620.

[59] Plataniotis, K. and D. Hatzinakos (2000). "Gaussian mixtures and their applications to signal processing". In: *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar, and Medical Imaging Real Time Systems*. Boca Raton, FL, USA: CRC.

[60] Psiaki, M. L. and T. E. Humphreys (2016). "GNSS spoofing and detection". *Proceedings of the IEEE*. 104(6): 1258–1270.

[61] Risbud, P., N. Gatsis, and A. Taha (2019). "Vulnerability analysis of smart grids to GPS spoofing". *IEEE Transactions on Smart Grid*. 10(4): 3535–3548.

[62] Sichitiu, M. L. and C. Veerarittiphan (2003). "Simple, accurate time synchronization for wireless sensor networks". In: *2003 IEEE Wireless Communications and Networking, 2003. WCNC 2003.* Vol. 2. 1266–1273.

[63] Sterzbach, B. (1997). "GPS-based clock synchronization in a mobile, distributed real-time system". *Real-Time Systems*. 12(1): 63–75.

[64] Sun, K., P. Ning, and C. Wang (2006). "TinySeRSync: Secure and resilient time synchronization in wireless sensor networks". In: *Proceedings of the 13th ACM Conference on Computer and Communications Security. CCS '06*. Alexandria, VA, USA: ACM. 264–277. DOI: 10.1145/1180405.1180439.

[65] Symmetricom (2009). "The importance of network time synchronization". *Tech. rep.*

[66] Tellabs (2014). "LTE stretches synchronization to new limits". *Tech. rep.*

[67] "Timing and synchronization aspects in packet networks" (2008). *Telecommunication Standardization Sector.* Apr.: 1–300. DOI: 10.1109/IEEESTD.2008.4579760.

[68] Ullmann, M. and M. Vögeler (2009). "Delay attacks — Implication on NTP and PTP time synchronization". In: *2009 International Symposium on Precision Clock Synchronization for Measurement, Control and Communication.* 1–6. DOI: 10.1109/ISPCS.2009.5340224.

[69] van Eeden, C. (2006). *Restricted Parameter Space Estimation Problems.* Springer.

[70] Van Greunen, J. and J. Rabaey (2003). "Lightweight time synchronization for sensor networks". In: *Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications. WSNA '03.* San Diego, CA, USA: ACM. 11–19.

[71] Vig, J. R. (1992). "Introduction to quatz frequency standards". *Army Research Laboratory Electronics and Power Sources Directorate Tech. Rep. SLCET-TR-92-1.* Oct.

[72] Wang, X., D. Jeske, and E. Serpedin (2015). "An overview of a class of clock synchronization algorithms for wireless sensor networks: A statistical signal processing perspective". *Algorithms.* 8: 590–620.

[73] Wolter, K., P. Reinecke, and A. Mittermaier (2011). "Model-based evaluation and improvement of PTP syntonisation accuracy in packet-switched backhaul networks for mobile applications". In: *Computer Performance Engineering.* Ed. by N. Thomas. Berlin, Heidelberg: Springer. 219–234.

[74] Wu, Y., Q. Chaudhari, and E. Serpedin (2011). "Clock synchronization of wireless sensor networks". *IEEE Signal Processing Magazine.* 28(1): 124–138.

[75] Zhang, J., R. S. Blum, L. M. Kaplan, and X. Lu (2017). "Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks". *IEEE Transactions on Signal Processing*. 65(3): 705–720.