Electrical Power and Energy Systems 122 (2020) 106179

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Probabilistic extension of flexible hybrid state estimation for cyber-physical systems

Vanja G. Švenda[a,*], Aleksandar M. Stanković[a], Andrija T. Sarić[b], Mark K. Transtrum[c]

[a] Department of Electrical and Computer Science, Tufts University, Medford, MA 02155, USA
[b] Department of Power, Electronic and Communication Eng., University of Novi Sad, Novi Sad, Serbia
[c] Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA

ARTICLE INFO

Keywords:
Co-simulation
Cyber-physical systems
ICT irregularities
Probabilistic observability
State estimation

ABSTRACT

This paper proposes a probabilistic extension to flexible hybrid state estimation (FHSE) for cyber-physical systems (CPSs). The main goal of the algorithm is improvement of the system state tracking when realistic communications are taken into account, by optimizing information and communication technology (ICT) usage. These advancements result in: 1) coping with ICT outages and inevitable irregularities (delay, packet drop and bad measurements); 2) determining the optimized state estimation execution frequencies based on expected measurement refresh times. Additionally, information about CPSs is gathered from both the phasor measurement units (PMU) and SCADA-based measurements. This measurement transfer introduces two network observability types, which split the system into observable (White) and unobservable (Grey) areas, based on 1) deployed measuring instruments (MIs) and 2) received measurements. A two-step bad data detection (BDD) method is introduced for ICT irregularities and outages. The proposed algorithm benefits are shown on two IEEE test cases with time-varying load/generation: 14-bus and 300-bus.

1. Introduction

With persistent advancements in communication systems and an ever-increasing demand for reliable and efficient electric energy, ICT is set to become a vital part of every power system. For electric utilities ICT is a key enabler for system monitoring, control, protection and data processing [1–3]. However, this increased reliance also has to acknowledge possible ICT irregularities. Thus, a reshaping of whole power system perception is required, by modeling and operating them as heavily intertwined two-layer cyber-physical systems (CPSs), consisting of:

1. Cyber parts—primarily used for information exchange, formed by ICTs [4].
2. Physical parts—primarily used for electric power generation, transmission and consumption, formed by power elements [5].

This transition offers opportunities to modify and enhance operations of Energy Management System (EMS). One important component is state estimation (SE), which provides vital information for utilities—the system state [6]. A straightforward addition for SE is CPS observability, which defines at what system's areas SE can be executed based on available measurements. The paper aims at:

1. forming a probabilistic CPS observability model, and
2. acquiring the best possible state tracking when ICT irregularities and outages are taken into account.

The first objective can be viewed as enhancing the observability models by incorporating information provided by CPS—formation of a probabilistic observability model. Conventionally, observability was derived from deployed measuring instruments (MIs), which represent devices that transform a physical variable of interest (e.g. voltage magnitude) into a form that is suitable for recording (measurement) [7]. Or rather, the availability of a measurement was determined by the existence of a corresponding MI [8]. Here, a more comprehensive model is proposed, by additionally taking into account the measurement transfer details. This gives more thorough ICT information, as the focus is both on what MI exist and when their sent packets will be available. One such research idea is presented in [9], where a statistical model of observability was formed, but for systems with solely PMUs. This paper follows on this idea by trying to do the same for systems with remote terminal units (RTUs) and PMUs (hybrid systems), and use such a model to improve state tracking. Initial exploration of this idea is

* Corresponding author.
E-mail address: vanja.svenda@tufts.edu (V.G. Švenda).

https://doi.org/10.1016/j.ijepes.2020.106179
Received 20 February 2020; Received in revised form 2 April 2020; Accepted 8 May 2020
Available online 25 May 2020
0142-0615/

presented in the authors' recent work [8] which focuses on static SE (SSE). The approach is extended here by:

- Incorporating the probabilistic observability model into a more complex state estimator which can tackle ICT irregularities and outages.
- Developing a two-tail significance test used for testing the probabilistic observability models which are first viewed as a hypothesis. This test is also adjusted to detect ICT outages.

With this observability platform, the focus can move on to the second objective—SE with ICT irregularities and outages. Several recent research strands address various ICT issues, their consequences and how to deal with them when performing SE. The authors' work in [11,12] examined packet delays and drops, their sensitivity to communication parameters, and their (negative) effects on SE, where the lessons learned will be further used in this paper when developing an appropriate state estimator and its corresponding tools. An Iterated Extended Kalman Filter (KF)-based SE for cases of observation and innovation outliers and non-Gaussian PMU noise is shown in [13]. While [14] proposes an estimator for the limited number of measurements and uncertain system parameters, [15] addresses for correlated SCADA and PMU measurements, non-Gaussian errors and non-synchronized measurements. Cubature KF for tackling temporary communication failure or packet loss is given in [16]. Performance degradation due to observation delays and irregular sampling is analyzed in [17]. Reference [1] examines approaches, scopes and major advancements in CPS security and reliability. Quantifying the impacts of ICT element failure and transmission interference on operation reliability is shown in [18]. Finally, the IEEE Task Force on Power System Dynamic State and Parameter Estimation recently published an overview of today's state-of-the-art estimators [19].

The starting point for the second objective (the best achievable state tracking with ICT irregularities and outages) will be Flexible Hybrid State Estimation (FHSE), which can deal with measurement packet delays and drops as shown in [20]. With the derived probabilistic CPS observability model, FHSE will be further enhanced to:

- Calculate the optimal SE frequency in terms of expected measurement refresh times.
- Detect, identify and deal with ICT irregularities and outages via a two-step Bad Data Detection (BDD) model, developed for this paper.

The remainder of the paper is as follows. Section 2 formulates the problem and presents the mathematical model of FHSE. Section 3 discusses CPS observability in terms of deployed MI and available measurements. Section 4 expands on this idea by including ICT irregularities and forming a probabilistic extension to observability, which is further used to calculate the optimal SE frequency. Section 5 discusses cyber-physical reliability and defines the two-step BDD method. The proposed algorithm and numerical results are shown in Sections 6 and 7, respectively. Conclusions in Section 8 are followed by the list of references.

2. Problem formulation

The following CPS information is assumed available:

- Measurements—RTU- and PMU-based (hybrid system). Note that it is assumed that each measurement has a timestamp specifying when it was taken.
- Measurement transfer details—set of transfer time values based on a large number of simulations executed by the Network Simulator 2 (NS-2) tool, a discrete-event, object-oriented simulator, targeted at communication network simulation and examination [21]. Different communication irregularities may be simulated through classes, such as packet delays and drops through *LinkDelay Class* and *ErrorModel Class*, respectively. For further information about simulations in NS-2, please refer to [20].
- Model for slow dynamics—bus power injection pseudo-measurements derived from load/generation forecast or daily profiles. Note that it is assumed at least one such measurement is available at each bus.

To have a good overview of ICTs, a system partitioning into appropriate areas is proposed based on 1) deployed MI, and 2) received measurements.

Deployed Measuring Instruments (MIs)

In terms of the deployed MI, the CPS can be split into 1) observable (White), and 2) unobservable (Grey) areas (see Section 3.A). Additionally, by incorporating this with the measurement transfer time details, the optimal SE frequency can be derived (see Section 4).

Received Measurements

The CPS may also be split in the same way concerning the received measurements (Section 3.B). Then, the state is estimated using one of the two FHSE components [20]:

- SSE (Section 2.A)—the state of the entire CPS (or their part) may be estimated using the SSE if enough measurements are provided for the entire CPS (or their part) to be observable; or
- Forecasting-Aided State Estimation (FASE) (Section 2.B)—to overcome measurement deficiency, the state transition matrix ($\mathbf{F}_n$) is used. This matrix is driven by slow stochastic power injection (load/generation) changes [19], which may be depicted by load/generation forecast, or daily load/generation curves.

Note that if appropriate measurements are not available, and certain abrupt changes which cannot be tracked by $\mathbf{F}_n$ have occurred (e.g. short circuits, power line outages, etc.) this model cannot be used, as it will lack appropriate information about the system.

Additionally, based on received measurements and FHSE results, the proposed two-step BDD is executed (Section 5).

When forming any type of Hybrid SE, the following issues arise (and are addressed later):

- SE reference bus [22]—since PMU angle measurements exist, there is no need to define the reference bus, and voltage angles at all buses are estimated.
- PMU-based measurement buffering [23,24]—to fully utilize PMU fast sampling rates. In this paper, buffer length is set as the time between two consecutive state estimations.
- Time skew of SCADA [25] and PMU [26] measurements—as all measurements have time-stamps, SCADA measurements corresponding to a high time skew may simply be disregarded. For PMUs, ideal sampling clocks are assumed as a practical solution whose details (local vs. GPS-derived) go beyond the scope of this paper.

A. Static State Estimation (SSE)

The SSE model for the n-th time instant is [6]:

$$\mathbf{z}(t_n) = \mathbf{h}(\mathbf{x}(t_n)) + \mathbf{e}(t_n) \tag{1}$$

where:

$\mathbf{z}(\cdot)$, $\mathbf{x}(\cdot)$ – measurement and state vectors, respectively;
$\mathbf{h}(\cdot)$ – vector of nonlinear functions relating the measurement and state vectors;
$\mathbf{e}(\cdot)$ – measurement error vector.

B. Forecasting-Aided State Estimation (FASE)

The state transition matrix for the n-th time instant ($\mathbf{F}_n$) is derived in [20]. To calculate this matrix, appropriate (pseudo) measurements from which bus injection increments can be calculated are needed. These can

be:

- Real-time measurements of load/generation (if available).
- When that is not the case, appropriate pseudo-measurements may be 1) load/generation forecast, or 2) daily load/generation profiles.

Once matrix $\mathbf{F}_n$ is formed, the system state can be calculated with the Discrete Kalman Filter (DKF) [27]:

$$\mathbf{x}(t_n) = \mathbf{F}_n\mathbf{x}(t_{n-1}) + \mathbf{w}(t_{n-1}) \quad (2)$$

$$\mathbf{z}(t_n) = \mathbf{h}(\mathbf{x}(t_n)) + \mathbf{e}(t_n) \quad (3)$$

where $\mathbf{z}(\cdot)$, $\mathbf{x}(\cdot)$, $\mathbf{h}(\cdot)$ are the same as in Section 2.A, and:

$\mathbf{w}(\cdot)$ − process noise, assumed as a zero mean multivariate normal distribution with covariance $\boldsymbol{Q}$;
$\mathbf{e}(\cdot)$ − measurement noise, assumed as a zero mean multivariate normal distribution, with covariance $\boldsymbol{R}$.

Lastly, even though DKF will provide the states, in this paper it is proposed that the final results be derived from SSE as [20]:

- SSE, which differs to the DKF in inputs/outputs, convergence criteria and mathematical modeling assumptions, still remains the main tool for state estimation in todays' power utilities.
- DKF uses pseudo-measurements in form of system slow dynamics (daily load/generation profiles), which are of low accuracy. The additional SSE step is hence used to improve the state estimation accuracy.

Thus, following DKF, the SSE-based state and measurement alignment is executed, where results from DKF are used as additional pseudo-measurement for SSE (1). Note that these measurements have a slightly better accuracy than the ones in form of system slow dynamics, as they contain information which went through the DKF-based estimator. Nevertheless, also note that these measurements should have a lower weighting factor, and they are included only to the degree needed to achieve observability.

3. Observability of a cyber-physical system (CPS)

Information used to observe a CPS is derived from the cyber layer formed by ICTs. In this paper, the focus is on the real-time operational communications [4]—MI, their corresponding communication channels and information (measurements) they provide. These can be used to define two types of observabilities, exploited in the proposed algorithm:

- Measuring Instrument observability (denoted by InstrO, Section 3.A)—conventional observability based on the MI deployment.
- Measurement observability (denoted by MeasO, Section 3.B)—observability based on the available measurements when the system state is to be estimated.

Note that both of above defined observabilities will fall under the category of numerical observability models [6].

A. Measuring Instrument Observability (InstrO)

Here, CPS is divided into observable and unobservable areas based on the deployed MI (note that this convention is used in the rest of this sub-section with regards to observability). A set of all non-redundant MI in the system (denoted by $\xi_{\mathrm{NR}}^{\mathrm{instr}}$) is formed first, together with its corresponding Jacobian matrix:

$$\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}} = \left[\frac{\partial \mathbf{h}_{\mathrm{NR}}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{T}}}\right], \quad (4)$$

where $\mathbf{h}_{\mathrm{NR}}(\cdot)$is the vector of nonlinear functions relating the measurements from $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ and the state vector $\mathbf{x}(\cdot)$.

In this context, the non-redundant MI correspond to the linearly independent rows in $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$ (for details of this procedure see [8]). It is important to note that, for a given CPS, $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ ($\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$) is generally not unique as redundant measurements are usually needed in transmission systems [28]. This is exploited here, as the proposed algorithm provides an optimized $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ ($\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$) for a stated optimization goal (see Section 4).

Next, by utilizing $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$, the unobservable (and consequently observable) areas may be defined in a given CPS. By definition of a Jacobian matrix, if a zero i-th column exists in $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$, it implies that changes in x_i have no effect on measurements from MI which form $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ (4). The following proposition further argues that this conclusion can be expanded to measurements from all systems' MI.

Proposition 1. *If a zero i-th column exists in $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$, it implies that changes in x_i have no effect on measurements from any MI in the system.*

Proof.. If changes in x_i have an effect on measurements from an MI that is not in $\xi_{\mathrm{NR}}^{\mathrm{instr}}$, by adding the corresponding row to $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$ a non-zero value in the i-th column would arise. But, as all other values in that column are zero, the added MI must be non-redundant with the rest of $\xi_{\mathrm{NR}}^{\mathrm{instr}}$, which is inconsistent with its definition.

From Proposition 1 the following may be concluded:

1. A zero i-th column in implies that there is no MI which directly or indirectly measures x_i. By examining the analytical forms of the Jacobian matrix $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$ [6], it may be concluded that no parameters are measured at: 1) i-th bus; 2) any bus directly connected to i-th bus, and 3) any branch directly connected to i-th bus. Thus, this implies that the i-th bus and all branches directly connected to it are unobservable—these elements form an unobservable area of the CPS.
2. Expanding on the first conclusion, even though $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ ($\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$) is usually not unique for a given system, unobservable (and observable) areas of the system are indeed unique. That is, measurements from $\xi_{\mathrm{NR}}^{\mathrm{instr}}$ provide observability over the largest possible part of the CPS.

Note that various other methods for finding unobservable areas exist (see Section 4 in [6]). Nevertheless, the highlighted set of information [$\xi_{\mathrm{NR}}^{\mathrm{instr}}$, $\mathbf{H}_{\mathrm{NR}}^{\mathrm{instr}}$] is obtained through other steps of the proposed algorithm, which is why using this specific method becomes natural (further explained in Section 4).

Thus, InstrO is used to separate the CPS into the following areas (simplified example is shown in Fig. 1):

- Measuring Instrument White (InstrW)—observable considering the available MI.
- Measuring Instrument Grey (InstrG)—unobservable considering the available MI.

B. Measurement Observability (MeasO)

Analogously, the CPS may be split into observable and unobservable areas based on received measurements at the time the system state is to be estimated. To do so, a set of non-redundant measurements (denoted

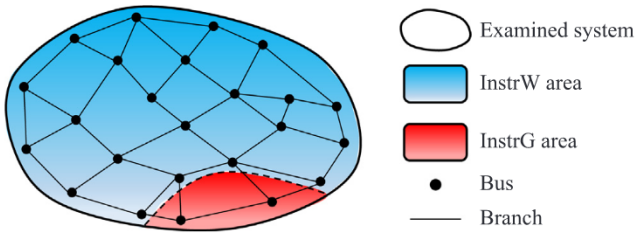

Fig. 1. Single line diagram of simplified CPS separated into InstrO areas.

by ξ_{NR}^{instr}) and its corresponding Jacobian matrix (denoted by $\mathbf{H}_{NR}^{meas}$) are formed, which are then used the same way as explained in Section 3.A. Thus, the measurement observability (denoted by MeasO) is used to separate the CPS into following areas:

- Measurement White (MeasW)—observable considering the available measurements.
- Measurement Grey (MeasG)—unobservable considering the available measurements.

4. Probabilistic extension to observability

To assign probabilities to data transfer times, the concept of measurement refresh times (t^{ref}) is introduced, which denotes the time between two consecutive measurements of a certain MI received at the EMS. These depend on a vast number of ICT parameters and irregularities, such as type of channel, bandwidth, queue limit, send interval, packet size, packet delay and drop probability [12]. Thus t^{ref} present stochastic variables with probability models that are difficult to explicitly describe.

To overcome this, this paper relies on communication simulation tools, like NS-2 [21], to observe measurement transfer over a realistically modeled communication system. Thus measurement refresh times observations, denoted $t_{i,j}^{ref}$ (j-th observation for i-th MI), can be obtained which represent:

- the time between measurements taken at time instances k and $k + 1$ received at the EMS, if neither of them have been dropped; or
- the time between measurements taken at time instances k and $k + 2$ (or k–1 and $k + 1$) received at the EMS, if measurement taken at time instance $k + 1$ (k) has been dropped.

Thus, both packet delays and drops will be taken into account.

Based on such observations, for every t_i^{ref} an estimate of their cumulative distribution functions (CDF), the empirical CDF (ECDF), can be derived [29]:

$$\hat{\mathscr{F}}_{T\,i}(t) = \frac{\text{Number of elements in observation set} \leqslant t}{N_i} \tag{5}$$

where N_i is the number of observations for the i-th MI (i.e. $j = 1, \ldots, N_i$ for $t_{i,j}^{ref}$). Thus, by simulating data transfer over a long, but practically feasible time period, a large number of observations is collected and a good estimate of the underlying CDF can be numerically derived [8,28]. As a result, the ECDF will provide a one-on-one relationship between refresh time values and their probabilities:

$$\hat{\mathscr{F}}_{T\,i}(t^{ref}) = \wp[T \leqslant t^{ref}] = \wp_T \tag{6a}$$

$$\hat{\mathscr{F}}^{-1}_{T(N)}(\wp_T) = \inf\{t^{ref}: \hat{\mathscr{F}}_{T(N)}(t^{ref}) \geqslant \wp_T\} \tag{6b}$$

where:

$\wp_T \in (0, 1)$ – corresponding probability;
$\hat{\mathscr{F}}^{-1}_{T(N)}(\cdot)$ – generalized inverse distribution function of $\hat{\mathscr{F}}_{T\,i}(\cdot)$.

Initially, these models are not proclaimed correct, but rather viewed as hypotheses which have to be tested. Thus, hypothesis H_0^i (ECDF is an appropriate estimate for CDF of t_i^{ref}) is defined for each i-th MI. These hypotheses are then evaluated using the two-tail significance testing [10], as follows. First, the sample mean for each MI is calculated based on refresh time observations:

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} t_{i,j}^{ref} \tag{7}$$

Using this information, it should be defined in which limits ($\mu_i \pm \Delta t_i^{break}$) does an actual observation of t_i^{ref} have to be in order for us to accept the given H_0^i. To do so, the significance level α, which defines the possibility of a false rejection error [10], has to be set:

$$\alpha = \wp[|\mu_i - t_i^{ref}| \geqslant \Delta t_i^{break}] \tag{8}$$

Note that the same value of α will be assigned to all MI. For a low false rejection error possibility, α is set to 0.03 (3%).

Next, Δt_i^{break} can be calculated from (8):

$$\alpha = \wp[\mu_i - t_i^{ref} \geqslant \Delta t_i^{break} \text{ OR } \mu_i - t_i^{ref} \leqslant -\Delta t_i^{break}] \Rightarrow$$

$$\alpha = \wp[t_i^{ref} \leqslant \mu_i - \Delta t_i^{break} \text{ OR } t_i^{ref} \geqslant \mu_i + \Delta t_i^{break}] \Rightarrow$$

$$\alpha = \hat{\mathscr{F}}_{T\,i}(\mu_i - \Delta t_i^{break}) + 1 - \hat{\mathscr{F}}_{T\,i}(\mu_i + \Delta t_i^{break}) \tag{9}$$

After Δt_i^{break} is calculated, hypothesis H_0^i is evaluated as:

$$\begin{aligned} &H_0^i \text{ is accepted;} \quad \mu_i - \Delta t_i^{break} \leqslant t_{i,j}^{ref} \leqslant \mu_i + \Delta t_i^{break}; \\ &\qquad H_0^i \text{ is rejected;} \quad \text{otherwise.} \end{aligned} \tag{10}$$

Once adequate probability models are available, they can be further used to improve ICT usage. One important aim is to derive the optimal frequency for SE (denoted by Δt^{SE}) with regards to measurement refresh times. As seen in Proposition 1, the MI in ξ_{NR}^{instr} proclaim the largest possible part of the CPS as an InstrW area. To optimize the proposed algorithm, the likelihood of this entire area being MeasW when trying to estimate the system state should be maximized. This will indeed be true if all measurements from ξ_{NR}^{instr} are refreshed once MeasO is checked. Thus, the following methodology is proposed:

1. Define ξ_{NR}^{instr}.
2. For a certain calculate MI from ξ_{NR}^{instr} (6); denote the largest value as Δt^{ref}; and set $\Delta t^{SE} = \Delta t^{ref}$.

Two issues can be observed from this pseudo-algorithm:

- As ξ_{NR}^{instr} is not unique (see Section 3.A), so is Δt^{ref}.
- Even for a certain ξ_{NR}^{instr}, by changing $\wp_T$ the resulting Δt^{ref} will change too. Accordingly, if a higher certainty is requested (larger $\wp_T$), more time should be allowed for the measurement to arrive (larger Δt^{ref}), and vice versa. This will represent a trade-off between Δt^{ref} (want to minimize) and $\wp_T$ (want to maximize).

A question may be raised now whether an optimal pair of ξ_{NR}^{instr} and $\wp_T$ exists (yielding an optimal Δt^{ref}), with a goal of maximizing the number of times MeasO proclaims the entire InstrW area as MeasW over a certain time period t. If so, this will produce an optimal Δt^{SE}, which can be formalized as:

$$\Delta t^{SE} = \max_{\xi_{NR}^{instr},\, \wp_T\,(\Delta t^{ref})} \left\{ \wp_T \cdot \frac{t}{\Delta t^{ref}} \right\} \tag{11a}$$

subject to:

$$\Delta t^{ref} = \max\{t_i^{ref} \mid \text{instrument } i \in \xi_{NR}^{instr}\}; \tag{11b}$$

$$eq.\ (6); \tag{11c}$$

$$\wp_T \in (0, 1). \tag{11d}$$

Note that Δt^{SE} is to be determined only once at the beginning of the algorithm, or when ICT outages occur, which is why (11) is solved by a direct (exhaustive search) method [8].

Finally, through the procedure of acquiring Δt^{SE} (11), corresponding ξ_{NR}^{instr} and $\mathbf{H}_{NR}^{instr}$ needed for InstrO are found (see Section 3.A). In the text bellow, the set of these parameters [ξ_{NR}^{instr}, $\mathbf{H}_{NR}^{instr}$, Δt^{SE}] will be denoted by $\mathbf{\Psi}$.

5. Reliability of information and communication technologies (ICTs)

Bad data from ICTs (CPS cyber portion) may have a significant influence on SE results, making an improved and efficient BDD model necessary. Note that the focus is on ICT reliability (component failures), rather than security. A two-step procedure, fully integrated into the proposed SE algorithm, is thus proposed:

1. Pre-Estimation BDD (Section 5.A), executed before estimating the state, based on the received measurements.
2. Post-Estimation BDD (Section 5.B), executed after estimating the state, based on the SE results.

The following ICT bad data are thus examined:

- Measurement packet delays and drops—certain time has to pass between a measurement being sent and received, with a possibility of it being lost (dropped). This is taken into account with ECDF models of t^{ref} (see Section 4).
- Measurement errors—measured and actual parameter values often differ. If this difference is significant, measurement errors are detected. This is examined in the Post-Estimation BDD step.
- Component outages—due to failure of MI or communication channels. This is examined in the Pre-Estimation BDD step.

A. Pre-Estimation BDD

This step detects and identifies ICT outages using the two-tail significance test, by checking if a measurement has not been received for more than the predefined $\Delta t_i^{\text{break}}$ (10). It is important to note that outages will be distinguished from packet drops, as the ECDF models and thus $\Delta t_i^{\text{break}}$ are formed by taking into account possible packet drops, which extends the waiting time before declaring ICT outages (see Section 4).

Two types of such outages can be observed:

- ICT outage from $\xi_{\text{NR}}^{\text{instr}}$—a new $\boldsymbol{\Psi}$ has to be defined ($\xi_{\text{NR}}^{\text{instr}}$ can no longer be the same).
- ICT outage not from $\xi_{\text{NR}}^{\text{instr}}$—even though this will generally result in SE result degradation, defining a new $\boldsymbol{\Psi}$ is not required ($\xi_{\text{NR}}^{\text{instr}}$, $\mathbf{H}_{\text{NR}}^{\text{instr}}$ and Δt^{SE} will not change).

B. Post-Estimation BDD

This step utilizes SE results to further examine if measurement errors exist in MeasW areas, or rather if the measured values deviate significantly from the true parameter values. Note that as true parameter values are practically unattainable, this is tested by comparing values calculated using the estimates and corresponding measurements. One such well-known method, used in this paper, is the Largest Normalized Residual (LNR) test [6].

If the LNR is greater than the predefined border value:

$$r_{\max}^{\text{N}} \geqslant \beta \tag{12}$$

the test has failed, and the corresponding measurement is disregarded from all future calculations. β is set to 3 for a high confidence level [20].

6. The overall proposed algorithm

Steps of the proposed algorithm are described as:

Step 1: Initialize CPS

- Physical part: network parameters, load/generation parameters (forecasts or daily profiles) and topology.
- Cyber part: the set of available MI (the initial system is assumed entirely InstrW) and ICT parameters, such as channel types, bandwidths, queue limits, send intervals, packet sizes, etc.

Step 2: Form the set of measurement transfer times

Using the NS-2 or an analogous tool, simulate the packet transfer through the observed CPS, and form a set of measurement refresh time observations for every MI (see Section 4). Note that in test examples these simulations are run for one hour (3600 s).

Step 3: Define the transfer time probability models

This step is done as explained in Section 4—based on information from ***Step 2*** define ECDFs and set up corresponding hypotheses H_0. If it is assumed ICT outages will not occur immediately, the hypotheses are accepted (or rejected) after the first few arrived measurements (10).

Step 4: *Define the initial* $\boldsymbol{\Psi}$

This step is done as explained in Section 4, (11).

Step 5: Measurement reinstatement

If measurements have been received from MI previously lost (denoted as an outage), reinstate them into the set of available MI. Check InstrO and form InstrW(G) areas (see Section 3.A). Finally, define $\boldsymbol{\Psi}$ over InstrW area (see Section 4).

Step 6: Pre-estimation BDD

This step is done as explained in Section 5.A. If outage from $\xi_{\text{NR}}^{\text{instr}}$ has occurred: check InstrO, form InstrW(G) areas, and define $\boldsymbol{\Psi}$ over InstrW area.

Step 7: Creating the set of available measurements

To create the measurement set, information is obtained from two different MI types used in this paper:

- RTU—latest received measurements are taken from each RTU, which is common practice in real-life CPS [20]. However, in order to overcome possible SCADA time-skew problems (see Section 2), measurement considered as "old" are disregarded. In this paper, to detect old measurements, it is observed how long ago they have been taken—if this time is more than twice the sampling rate of corresponding instruments, they are considered old.
- PMU—as measurement buffering is used, the average measurement values are taken, where the buffer length is the time between two consecutive state estimations (see Section 2).

Step 8: The FHSE

Based on the measurement set formulated in ***Step 7***, check MeasO, form MeasW(G) areas, and estimate the system state by utilizing:

- SSE (Section 2.A)—if the entire system is MeasW.
- FASE (Section 2.B)—if MeasG areas exist.

Step 9: Post-estimation BDD

This step is explained in Section 5.B. If measurement errors exist, check InstrO, form InstrW(G) areas, and define $\boldsymbol{\Psi}$ over InstrW area.

Step 10: BDD imposed FHSE

If no measurement errors are detected in ***Step 9***, the system state is as calculated in ***Step 8***. Otherwise, first disregard the corresponding MI and their measurements, and then estimate the system state, as explained in ***Step 8***.

7. Application

For the proposed algorithm, physical and cyber system parts are modeled in MATLAB and NS-2 [21], respectively. These two parts are then co-simulated using PiccSIM [30], providing the final CPS models—14-bus (Section 7.A) and 300-bus (Section 7.B). For all examples note the following:

- CPSs are simulated for one hour (3600 s).
- Measurements are formed as random variables, with power flow solution means and variances as following percentage of their measured values: 1) 10^{-4} for voltage magnitudes and angles; and 2)

10^{-3} for power injections and branch flows. These values are later denoted in figures as 'Measured'.

- Measurement weights for SSE are set as the reciprocal value of corresponding measurement variances.
- RTU and PMU sampling rates are 2 s and 0.02 s, respectively. Note that for both measurement types, corresponding measurements are assumed to have been taken taken at the same moment (same snapshot).
- To quantify the proposed algorithm, its results (denoted in tables and figures as 'FHSE') are compared with SSE results (denoted as 'SSE') and measured values.
- SSE is attempted every 2 s only over InstrW areas as it cannot estimate the states where appropriate MI, and thus measurements, are not available (InstrG areas). On the other hand, FHSE can overcome this by utilizing the proposed FASE (see Section 2.B).

C. IEEE 14-Bus Test System

Details of the physical and cyber (ICT) parts are given in [31] and [20], respectively. Note that two measurement transfer types are examined, depending on used MI:

- PMU—measurements travel directly to EMS [phasor data concentrator (PDC) and EMS are physically at the same location].
- RTU—measurements are first collected at a data center and then forwarded to EMS (SCADA and EMS are not physically at the same location).

Finally, details of available measurements are given in Table 1. Note that measurement types and corresponding MI are chosen arbitrarily, and that the proposed algorithm will work for all other combinations.

An example is demonstrated here when ICT outages occur at $t = 1000$ s, and remain persistent for the rest of the simulations, for measurements at:

- Bus 12—measurement # 13, denoted by PQf_{12-6} (Table 1).
- Bus 13—measurement # 11, denoted by PQf_{13-14} (Table 1).

Utilizing the algorithm provided in Section 6, the CPS is first initialized (***Step 1***). Next, through the NS-2 tool, packet transfer may be simulated, whose results are used to calculate ECDFs for all measurement refresh times and set up corresponding two-tail significance tests (***Steps 2–3***). For example, critical transfer times [Δt^{break} (9)] for PQf_{12-6} and PQf_{13-14} are 2.53 s and 2.45 s, respectively.

The initial $\boldsymbol{\Psi}$ is then calculated (***Step 4***):

- $\Delta t^{\text{SE}} = 2.4$ s;
- $\xi_{\text{NR}}^{\text{instr}} = [1; 2; 3; 4; 5; 7; 8; 10; 11; 12; 13; 14; 15; 16]$.

Table 1
Available measurements.

#	Type	MI	Sending bus	Receiving bus	Denoted by
1	P_{INJ} (Q_{INJ})	PMU	2	/	PQi_2
2	P_{INJ} (Q_{INJ})	PMU	3	/	PQi_3
3	P_{INJ} (Q_{INJ})	PMU	6	/	PQi_6
4	P_{INJ} (Q_{INJ})	RTU	8	/	PQi_8
5	P_{INJ} (Q_{INJ})	RTU	9	/	PQi_9
6	P_{INJ} (Q_{INJ})	RTU	11	/	PQi_{11}
7	P_{INJ} (Q_{INJ})	RTU	14	/	PQi_{14}
8	P_{FLOW} (Q_{FLOW})	PMU	4	7	PQf_{4-7}
9	P_{FLOW} (Q_{FLOW})	RTU	8	7	PQf_{8-7}
10	P_{FLOW} (Q_{FLOW})	RTU	9	10	PQf_{9-10}
11	P_{FLOW} (Q_{FLOW})	RTU	13	14	PQf_{13-14}
12	P_{FLOW} (Q_{FLOW})	RTU	11	6	PQf_{11-6}
13	P_{FLOW} (Q_{FLOW})	RTU	12	6	PQf_{12-6}
14	θ (V)	PMU	1	/	θV_1
15	θ (V)	PMU	4	/	θV_4
16	θ (V)	PMU	5	/	θV_5

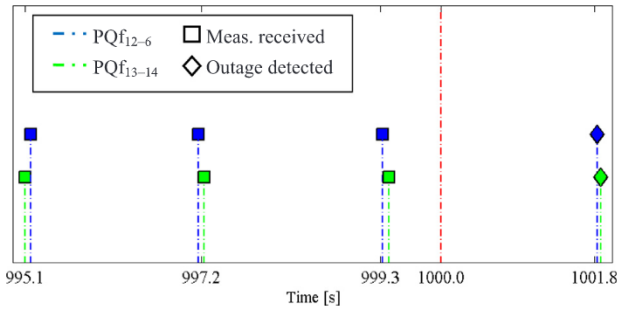

Fig. 2. Measurement arrivals and outage detection.

Finally, the set of available measurements is formed (***Step 7***) and the system state is calculated (***Step 8***) every Δt^{SE}, while bad data is monitored. Once outages occur, they are detected and identified by the Pre-estimation BDD (***Step 6***), as shown in Fig. 2. Note how measurement refresh times differ, due to ICT irregularities (e.g. packet delays). Once measurements have not been received for the predetermined Δt^{break}, outages are detected and identified.

As outage of measurement from initial $\boldsymbol{\Psi}$ has occurred, InstrO is checked and InstrW(G) areas formed, as demonstrated on Fig. 3. Additionally, recalculating of $\boldsymbol{\Psi}$ is needed:

- $\Delta t^{\text{SE}} = 2.3$ s;
- $\xi_{\text{NR}}^{\text{instr}} = [1; 2; 3; 4; 5; 7; 8; 10; 12; 14; 15; 16]$.

Due to ICT outages, and thus fewer available measurements, a decline in estimation results quality can be observed, as shown in Fig. 4. This is most expressed in:

- Buses 12 and 13—they form InstrG area.
- Buses 6, 11 and 14—effect of InstrG area is conveyed by measurements 12, 13 and 6 (see Table 1 and Fig. 3).

Note that even though ICT outages have occurred, good state tracking still exists as the proposed algorithm will switch to FASE, and continue estimating even the InstrG area states.

Finally, the errors of average voltage magnitudes and angles for

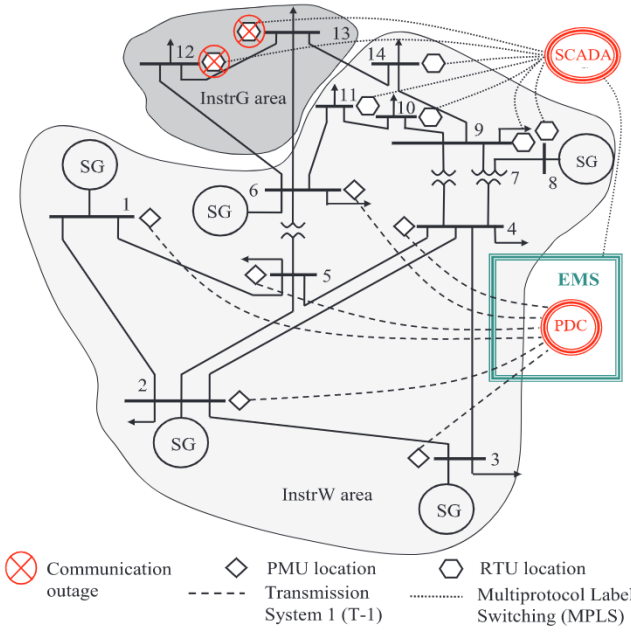

Fig. 3. IEEE 14-bus test system split into observable and unobservable areas based on available MI.

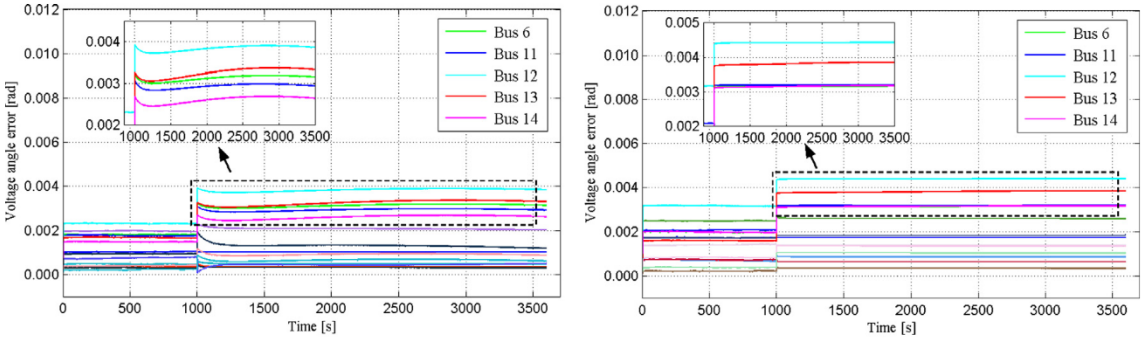

Fig. 4. Errors of bus voltage angle (a) and magnitude (b).

Table 2
Errors of average voltage magnitudes and angles.

Time interval		Pre-outage	Post-outage		Entire interval
Area		InstrW + InstrG	InstrW	InstrG	InstrW + InstrG
FHSE	V [p.u.]	0.0014	0.0017	0.0041	0.0024
	θ [rad]	0.0011	0.0011	0.0032	0.0018
SSE	V [p.u.]	0.0016	0.0024	0.0063	0.0034
	θ [rad]	0.0014	0.0018	0.0048	0.0027

both FHSE and SSE are shown in Table 2.

The following can be concluded from results in Table 2:

- FHSE gives slightly better results before ICT outages. This is due to existing ICT irregularities which split the network into MeasW(G) areas. Unlike the SSE, the FHSE can estimate states in MeasG areas utilizing FASE.
- Once outages occur, the FHSE gives slightly better results again in InstrW area, due to existing ICT irregularities. But, it gives significantly better results for InstrG areas, where MI, and thus measurements, are no longer available—SSE cannot estimate the states without appropriate measurements, whereas FHSE may utilize the proposed FASE (see Section 4.B).
- Better overall state tracking is achieved when using FHSE.

D. IEEE 300-Bus Test System

While the physical part details are given in [32], cyber part (ICT) details are given the Appendix.

An example is demonstrated here when measurement errors are introduced in buses 197, 198, 203, and 211 at $t = 2000$ s until the end of the simulations (this is done by increasing their variances 10^3 times). In order to visualize this, the corresponding area is shown in Fig. 5.

To demonstrate the effectiveness of the proposed algorithm, the state is estimated with 1) SSE; 2) FHSE with non-optimal $\Delta t^{SE} = 2.0$ s, and 3) FHSE with optimal $\Delta t^{SE} = 2.2$ s. Note that the optimal $\Delta t^{SE} = 2.2$ s remains the same after bad data is detected and corresponding measurements are disregarded.

The results may be summarized as follows:

- The advantage of using the optimal frequency for SE is observed in Fig. 6, as the number of possible SSE executions (entire InstrW area is also MeasW, Section 3). Note that the filled and empty dots in the zoomed-in area denote the time intervals where SSE execution was possible or not, respectively. Even though SSE will be attempted more often when $\Delta t^{SE} = 2.0$ s (zoomed-in area), more SSE executions are possible for $\Delta t^{SE} = 2.2$ s, as it is derived while taking into account inevitable ICT irregularities (Section 4).
- Numerical results are provided in Table 3. Notice the advantages of using the proposed algorithm instead of SSE, due to the ICT irregularities and the existence of InstrG area (all of which are not dealt with in SSE). Slightly better results are achieved for $\Delta t^{SE} = 2.2$ s than $\Delta t^{SE} = 2.0$ s, as measurement usage is optimized (Fig. 6).
- Good state tracking can be observed for the proposed algorithm in Fig. 7 for buses 198, 211 and 50. Note that once bad measurements are detected, the degradation of results is observed in buses 198 and 211, due to loss of observability—InstrG area (see Fig. 5). However, this is not the case for bus 50, as it remains in an InstrW area.

Finally, a few notes on the practical implementation of the proposed algorithm:

- Due to insufficient modern ICT equipment (e.g. fast-sampling PMUs), many of todays' real-life power systems still operate their state estimators at low refresh rates. This might be even slower than RTU refresh times, limiting the effectiveness of optimal state estimation frequencies (Section 4). Nevertheless, this paper strives to motivate future practical implementations, especially in power systems with modern ICT.

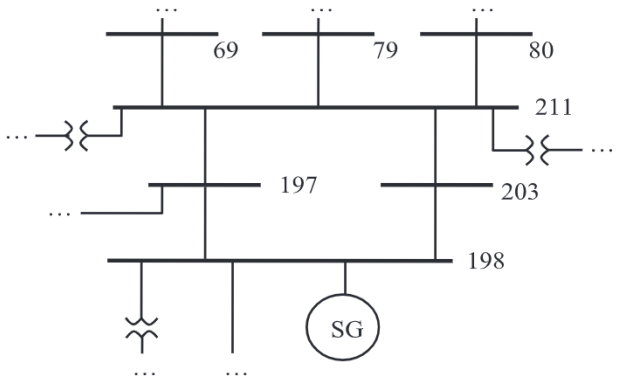

Fig. 5. System area affected by bad measurements.

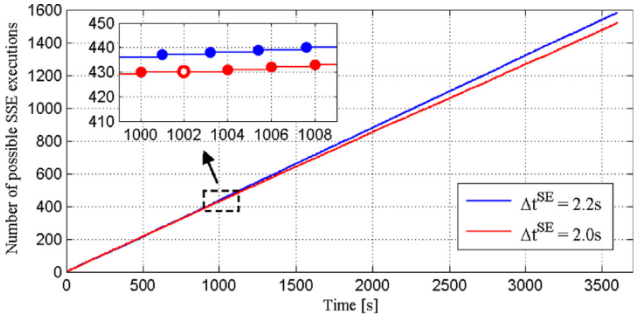

Fig. 6. Possibility of executing SSE over a time.

Table 3
Errors of average voltage magnitudes and angles.

	SSE	FHSE ($\Delta t^{SE} = 2$ s)	FHSE ($\Delta t^{SE} = 2.2$ s)
V [p.u.]	0.0045	0.0035	0.0031
θ [rad]	0.0053	0.0041	0.0035

- On the other hand, the proposed two-step BDD may be integrated into various types of state estimators applied in real-life power systems, regardless of available ICT equipment, as the probabilistic network observability extension is not constrained by it (Section 5).

8. Conclusion

This paper proposes a probabilistic extension of flexible hybrid state estimation for cyber-physical systems (CPS) with realistically modeled information and communication technologies (ICT). Such ICTs do experience difficulties, including irregularities (packet delays, drops and measurement errors) and component outages. To deal with such systems, two distinct network observabilities are formulated, based on deployed MI and actual received measurements. They are then further enhanced by a probabilistic extension to CPS's observability model, used to optimize measurement usage. These concepts are used for: 1) ICT outage detection and identification, based on the two-tail significance test; and 2) finding the optimal frequency for state estimation based on measurement refresh times. Additionally, to robustly deal with identified ICT issues, a two-step bad data detection model is fully integrated in the proposed algorithm.

CRediT authorship contribution statement

Vanja G. Švenda: Methodology, Software, Validation, Formal analysis, Writing - original draft. **Aleksandar M. Stanković:**

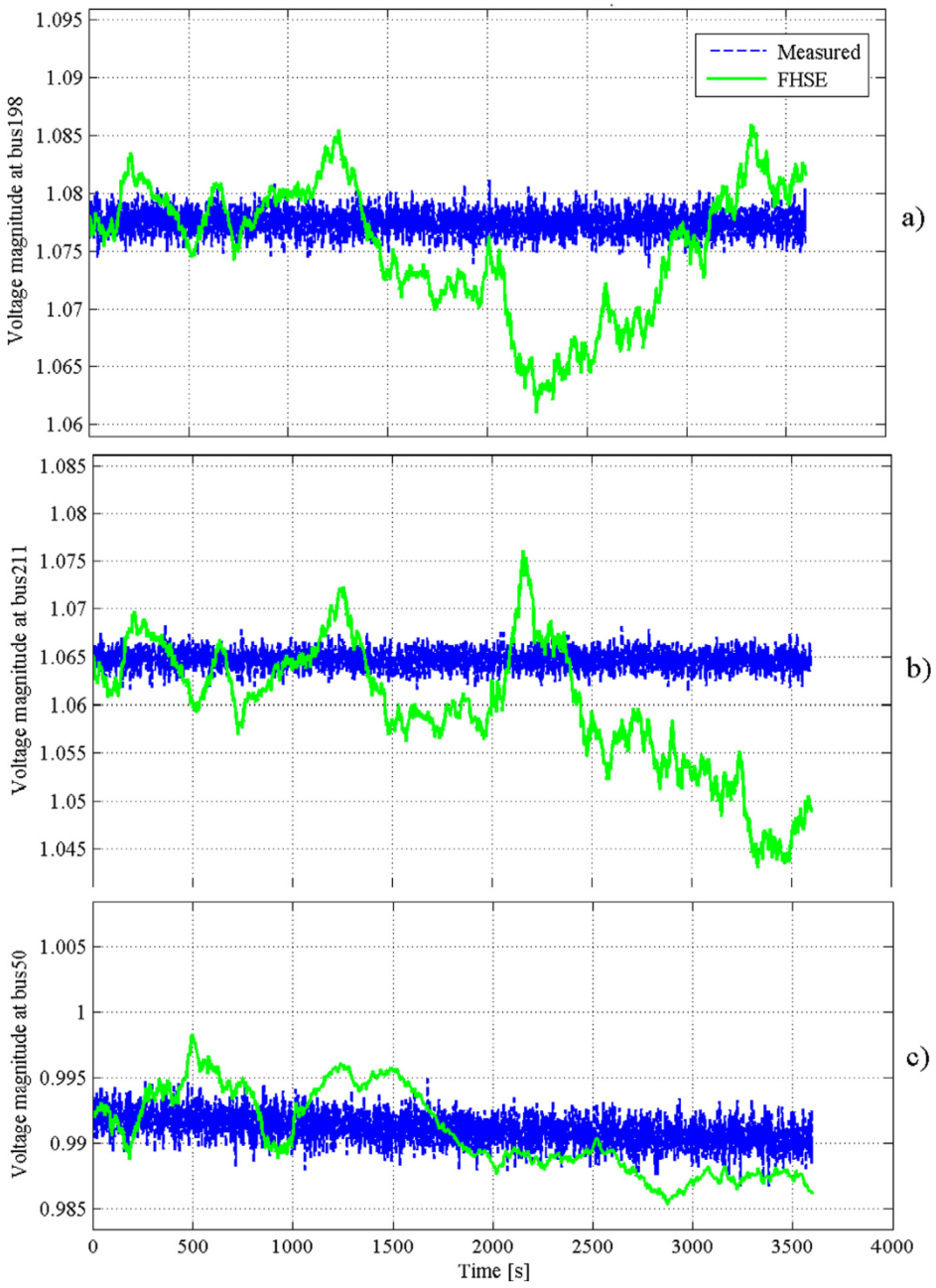

Fig. 7. Voltage magnitudes at buses 198 (a), 211 (b) and 50 (c).

Conceptualization, Writing - review & editing, Supervision, Funding acquisition. **Andrija T. Sarić:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Mark K. Transtrum:** Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been supported by NSF under grant ECCS- 1710944, by CURENT Engineering Research Center of the National Science Foundation and the Department of Energy under NSF Award Number EEC-1041877, by ONR under grant N00014-16-1-3028, and in part by the Ministry of Education and Science of the Republic of Serbia, under project III-42004.

Appendix A. 300-Bus system cyber layer parameters

The entire CPS is split into three parts (denoted as *CPS-section 1*, *2* and *3*). To cover all relevant cases, the different measurement parameters are set for the three sections [20]:

- Measuring instrument (MI) and channel types –

CPS-section 1 has RTU measurements at every bus, which are sent directly to the EMS. These measure both power flows and injections.

CPS-section 2 has PMU measurements at every bus, which are first collected at four different PDCs and then forwarded to the EMS. These measure complex bus voltages.

CPS-section 3 has RTU measurements at every bus, which are first collected at 3 different SCADAs and then forwarded to the EMS. These measure both power flows and injections.

- Type of link – duplex link assumed for each channel, which is a two-way communication link.
- Traffic type – Constant Bitrate (CBR) for each channel is assumed, meaning that traffic moves at a constant rate.
- Packet size – 1000 bits for each channel.
- Bandwidth – 2 Mb assumed for each channel.
- Default delay distribution – normal distribution assumed for each channel.
- Send interval – PMU measurements are sent 50 times per second, while RTU measurements are sent every 2 s.
- Queue type – drop-tail for each channel, which operates through a first in first out (FIFO) mechanism.
- Default delay rate –

CPS-section 1: RTU-EMS: mean 500 ms, standard deviation 50 ms.

CPS-section 2: 1) PMU-PDC: mean 1000 ms, standard deviation 150 ms; 2) PDC-EMS: 100 ms.

CPS-section 3: 1) RTU-SCADA: mean 400 ms, standard deviation 50 ms; 2) SCADA-EMS: 100 ms.

- Queue Limit –

CPS-section 1: RTU-EMS: 2; *CPS-section 2*: 1) PMU-PDC: 100; 2) PDC-EMS: 1000;

CPS-section 3: 1) RTU-SCADA: 10; 2) SCADA-EMS: 100.

- Drop model rate – *CPS-section 1*: RTU-EMS: 2%; *CPS-section 1*: 1) PMU-PDC: 2%; 2) PDC-EMS: 1%;

CPS-section 1: 1) RTU-SCADA 2%; 2) SCADA-EMS: 1%.

References

[1] Lei H, Chen B, Butler-Purry KL, Singh C. Security and reliability perspectives in cyber-physical smart grids. IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Singapore. 2018. p. 42–7.
[2] Cintuglu MH, Mohammed OA, Akkaya K, Uluagac AS. A survey on smart grid cyber-physical system testbeds. IEEE Com Surv Tutorials 2017;19(1):446–64.
[3] Salman SK. Smart grid communication system and its cyber security. Introduction to the Smart Grid: Concepts, Technologies and Evolution, The Institution of Engineering and Technology (IET), ch. 5. 2017.
[4] Ericsson GN. Cyber security and power system communication—Essential parts of a smart grid infrastructure. IEEE Trans. Power Del. Jul. 2010;25(3):1501–7.
[5] Grainger JJ, Stevenson Jr. WD. Power System Analysis. McGraw-Hill Education; 1994.
[6] Abur A, Exposito AG. Power System State Estimation – Theory and Implementation. 1st ed. CRC Press; 2004.
[7] Webster JG. Measurement Characteristics ch. 1 The Measurement, Instrumentation and Sensors Handbook. CRC Press; 1998.
[8] Švenda VG, Stanković AM, Sarić AT, Transtrum MK. Probabilistic approach to network observability of a hybrid power system with communication irregularities. 2019 North American Power Symp. (NAPS), Wichita, KS, USA. 2019.
[9] You M, Jiang J, Tonello AM, Doukoglou T, Sun H. On statistical power grid observability under communication constraints (invited paper). IET Smart Grid Jul. 2018;1(2):40–7.
[10] Yates RD, Goodman DJ. Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers. 3rd ed. Wiley; 2014.
[11] Stanković AM, Švenda V, Sarić AT, Transtrum MK. Hybrid power system state estimation with irregular sampling. IEEE PES General Meeting, Chicago, IL, USA. 2017.
[12] Švenda VG, Stanković AM, Sarić AT, Transtrum MK. Influence of communication irregularities and co-simulation on hybrid power system state estimation. IEEE PES Innovative Smart Grid Tech. Conf. Europe, Sarajevo, Bosnia and Herzegovina. 2018.
[13] Zhao J, Netto M, Mili L. A robust iterated extended Kalman filter for power system dynamic state estimation. IEEE Trans Power Syst Jul. 2017;32(4):3205–16.
[14] Bilil H, Gharavi H. MMSE-based analytical estimator for uncertain power system with limited number of measurements. IEEE Trans Power Syst Sep. 2018;33(5):5236–47.
[15] Zhao J, Wang S, Mili L, Amidan B, Huang R, Huang Z. A robust state estimation framework considering measurement correlations and imperfect synchronization. IEEE Trans Power Syst Jul. 2018;33(4):4604–13.
[16] Sharma A, Samantaray SR. Power system tracking state estimator for smart grid under unreliable PMU data communication network. IEEE Sens J Mar. 2018;18(5):2107–16.
[17] Yan B, Lev-Ari H, Stankovic AM. Networked state estimation with delayed and irregularly spaced time-stamped observations. IEEE Trans Control Netw Syst Sep. 2018;5(3):888–900.
[18] Wang C, Zhang T, Luo F, Li F, Liu Y. Impacts of cyber system on microgrid operational reliability. IEEE Trans Smart Grid Jan. 2019;10(1):105–15.
[19] Zhao J, et al. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. IEEE Trans Power Syst Jul. 2019;34(4):3188–98.
[20] Švenda VG, Stanković AM, Sarić AT, Transtrum MK. Flexible hybrid state estimation for power systems with communication irregularities. IET Generation, Transmission & Distribution, accepted for publication. http://doi.org/10.1049/iet-gtd.2019.1148.
[21] Fall K, Varadhan K. The ns Manual (formerly ns Notes and Documentation). The VINT project; 2011, Nov. [Online]. Available: https://www.isi.edu/nsnam/ns/doc/ns_doc.pdf.
[22] Zhu J, Abur A. Effect of phasor measurements on the choice of reference bus for state estimation. IEEE PES General Meeting, Tampa, FL. 2007.
[23] Zhang Q, Chakhchoukh Y, Vittal V, Heydt GT, Logic N, Sturgill S. Impact of PMU measurement buffer length on state estimation and its optimization. IEEE Trans Power Syst May 2013;28(2):1657–65.
[24] Murugesan V, Chakhchoukh Y, Vittal V, Heydt GT, Logic N, Sturgill S. "PMU data buffering for power system state estimators. IEEE Power Energy Tech Syst J Sep. 2015;2(3):94–102.
[25] Dabbaghchi I, VanSlyck LS. Inter-utility data exchange for state estimation. IEEE Trans Power Syst Aug. 1988;3(3):1254–62.
[26] Zhang Q, Vittal V, Heydt G, Chakhchoukh Y, Logic N, Sturgill S, et al. The time skew problem in PMU measurements. The 2012 IEEE Power and Energy Society General Meeting, San Diego, CA. 2012. p. 1–6.
[27] Welch G, Bishop G. An introduction to the Kalman filter. University of North Carolina. Chapel Hill, NC, USA; 2006, Jul. [Online]. Available: https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.
[28] Wu FF, Monticelli A. Network observability: Theory. IEEE Trans Power App Syst 1985;PAS-104(5):1042–8.
[29] van der Vaart AW. Empirical Processes. Asymptotic Statistics. Cambridge University Press; 2000. p. 265–90.
[30] Bjorkbom M, Nethai S, Kohtamaki T. PiccSIM Manual. Wireless Sensor Systems Group, School of Electrical Engineering, Aalto University. Espoo, Finland; 2013, Dec. [Online]. Available: http://wsn.aalto.fi/en/tools/piccsim/piccsim_manual_1.1.pdf.
[31] Christie R. 14 bus power flow test case. University of Washington. Seattle, WA, USA; 1993. [Online]. Available: https://www2.ee.washington.edu/research/pstca/pf14/pg_tca14bus.htm.
[32] Christie R. 300 bus power flow test case. University of Washington. Seattle, WA, USA; 1993. [Online]. Available: https://www2.ee.washington.edu/research/pstca/pf300/pg_tca300bus.htm.