Fast, Approximate State Estimation-Based Approach for Cyber Threat Detection in Power Systems

A. Chattopadhyay, S. Dasgupta, R. Macwan,
A. Valdes, G. Gross, P. W. Sauer
University of Illinois at Urbana-Champaign
Urbana, IL
{ac33, sujayd2, rmacwan, avaldes, gross, sauer}@ilinois.edu

R. Nuqui
ABB Corporate Research Inc., Rayleigh, NC
reynaldo.nuqui@us.abb.com

Abstract

Increasing reliance on cyber-components for communication and control has made cybersecurity in power systems an increasing concern. While Information technology (IT) based detection and prevention methods are deployed to detect cyber threats, leveraging of the system physics provides a complementary detection scheme. Here, we consider malicious power order threats directed at a high-voltage direct current (HVDC) line in a large AC network. A fast, approximate tracking state estimation method is investigated that uses a reduced iteration count and measurement prioritization using power transfer distribution factors (PTDF) to rapidly compute the approximate injections at the AC buses of the HVDC line as a power order is executed. The algorithm's accuracy in tracking the system's change is investigated. It is observed that with the above methods, the estimator can achieve results within 5% of the true injection. Deviations from the expected injection can be understood to be indicative of a compromised power order.

1. Introduction

There is concern that increasing reliance of power systems on cyber components for communication and control exposes a cyber attack surface that may be exploited to cause adverse system impacts. While it is possible, and advisable, to include intrusion detection/prevention systems (IDS/IPS) to verify the source, destination, and syntax of the protocol messages, leveraging knowledge of the power system network provides a complementary detection capability. Transmission operators can use knowledge of the power network topology and the underlying physical laws that govern the system to their advantage to detect and potentially mitigate suspicious system operation. In this study, we propose such a complementary detection mechanism and use it to examine the scenario of a large AC power system with a high-voltage direct current (HVDC) link. Specifically, we seek to detect malicious cyber threats directed at the power order command in conjunction with spoofed measurements from the HVDC stations.

In a large power system, such an HVDC link can be used to transfer power from one region of the network to another during times of high demand to help ease congestion. Typically, in such instances, the system operator's control center (SOCC) sends a power order command over a communication channel to the sending end converter control station (CCS) for a certain amount of power to be exported on the HVDC line. This order is then executed by the operator at the CCS. At the receiving end, the inverter control station (ICS) receives the power, converts it back into AC power, and injects it into the AC network. In this situation, there exists a possibility that an attacker will intercept the power order command and alter it to a malicious value, while also spoofing the feedback measurements from the HVDC station, so that the power order execution appears correct to the control center. An execution of this altered power order can jeopardize the safe operation of the system. Thus, it is critical to develop means to ensure that such an intrusion and attack can be detected before it is too late. That will require an ability to rapidly track a system's evolution in time as a power order is executed. That way, if the order has been maliciously altered, it can be detected in a short time span so that swift action can be taken to stop the continued execution of the malicious power order.

The detection scheme being proposed utilizes real and reactive line flow measurements to perform state estimation based on weighted least squares (WLS). The resulting state can then be used to estimate the expected power injections at the AC buses connected to the HVDC converter and inverter stations. If the injections are found to be significantly different—accounting for measurement noise, communication latency, and estimation error—from the expected value for a given power order execution, they can be flagged as potentially

URI: https://hdl.handle.net/10125/64129
978-0-9981331-3-3

compromised executions. As part of the the scope of this study, we make a few assumptions:

1. The communication links between the control center and the HVDC stations are considered compromised, while the communication from the larger AC grid is assumed to be trustworthy, but with non-malicious measurement errors that are normally distributed with zero bias.

2. The measurements from the CCS and ICS are compromised and none of them can be trusted.

3. The system topology does not change in the time period in which the power order is executed.

4. The system has the generating capability to provide for this power export. At every *set point* of the HVDC line, the system does not exhibit any small-time scale dynamic behavior.

Those assumptions define the scope of our work.

The contributions of this paper are the development of approximation techniques to enable sufficiently accurate state estimation, and quantitative evaluation of these techniques as applied to cyber threat detection in power systems. While our analysis is in the context of an HVDC system in a larger AC system, our techniques are applicable to any power system situation, such as generator dispatch for which an expected impact of a major operation on a wider transmission system is physically modeled and the operation itself is potentially subject to cyber attack.

2. Our Approach

To emulate realistic conditions of an operating power system that sends data to a control center, which then receives and processes the information appropriately, it is desirrable to have one platform that serves as a proxy of the real power system and another that serves the role of the computational platform at the SOCC. In our modeling, we use the network model on the PowerWorld Simulator platform [1] as a stand-in for the real system. For a DC power order executed on this system, the power flow is computed for a sequence of points along the power injection ramp that are equally spaced in time as the power order evolves, progressively going from zero to full rating. For brevity, these points will be referred to as *set points* of the HVDC line. The *set points* are defined in terms of the DC current (I_{DC}) that results in a desired DC power flow on the HVDC line. Real and reactive line flow measurements obtained from the power flow for each *set point* are then exported in comma-separated values (.csv) format to a computational platform—in this case, MATLAB—which serves as a stand-in for the computational platform at the SOCC. The measurements are used in a weighted least squares (WLS) formulation to determine the state of the system. We utilize the MATPOWER set of open-source scripts [2] for the power system computational aspects, with modifications as needed to implement the fast, approximate tracking state estimation method. Once an estimate for the state is found, it can be used to compute the real power injections at the AC buses connected to the converter and inverter station. For an approved power order, the estimated injections should conform—within a margin of error—to those expected for the power order being executed. Any deviations from the expected injection profile would be indicative of a power order that is either incorrectly received or incorrectly executed.

Given that a large system has the ability to output a correspondingly large set of measurements at every instant in time, and that full state estimation may not be computationally feasible in the available time budget, some approximations and means of prioritization of measurements are required in order to achieve an estimate of sufficient accuracy within a short computational time for detection purposes. The prioritization of measurements also reflects the real-life scenario where measurements are received asynchronously and thus only a small subset of measurements can be received in a specific time instant that corresponds approximately to a static system operating point. The contributions of this paper are the identification of approximation techniques to enable sufficiently accurate state estimation and quantitative evaluation of these techniques as applied to cyber threat detection in power systems.

The two modifications we have made to the state estimation algorithm are capping of the maximum number of iterations and the prioritization of measurements, according to the power transfer distribution factor (PTDF). The approach of capping the maximum number of iterations for power system state estimation for the purpose of faster cyber threat detection in real time system operation was studied in [3]. Leveraging on that work, we have extended the approach by adding prioritization of measurements that uses PTDF analysis to account for issues caused by communication latency and slower computational times due to large system state estimations in real time. PowerWorld was also used for calculation of the PTDFs used in this study.

3. Methodology

In this section, we discuss the specific case chosen to investigate our approach. The steps required for the implementation of the fast, approximate state estimation-based approach are also elaborated upon.

3.1. Case Under Study

The system under study is a synthetic system that approximates the US part of the Western Interconnection, managed by the Western Electrical Coordinating Council (WECC). The synthetic case file is the 10,000-bus case created and maintained at the Electric Grid Test Case Repository at TAMU, available online at [4]. The cases maintained in the repository were created based on an approach laid out in [5]. An HVDC line that is comparable to the Pacific DC intertie is modeled in the 10,000-bus case at synthetic buses that correspond closely to the geographical location of the converter and inverter stations in the physical WECC system. The HVDC converter station is modeled as current-controlled, while the HVDC inverter station is modeled as voltage-controlled. The capacity of the line is 3100 MW, which corresponds to a DC *set point* I_{DC} of 6200 A. We consider a linear ramp rate for the DC power injection under which the line is ramped from zero to full rating in 20 minutes, resulting in a ramp rate of 155 MW/min. The system is capable of a faster ramp rate, but we believe the operation we chose is typical, based on discussions with a utility operating a similar system. Our techniques do not require a linear ramp rate. Since the real Pacific DC intertie is connected to the physical AC grid at 230 kV, the 10,000-bus system is equivalenced to retain transmission lines of voltage levels 230 kV and above. That results in a reduced system of approximately 2500 buses. External generators are retained. The equivalencing was done in PowerWorld. Table 1 lists the system characteristics for the original as well as the reduced network.

Table 1. Result of reduction operation on the original system

System Parameter	Original WECC case	Reduced WECC case
Number of Buses	10,000	2,470
Number of Lines	12,706	5,375

3.2. AC State Estimation Theory

The classical state estimation technique adapted for use in power systems, first proposed by Schweppe [6][7][8], is a weighted least squares (WLS) problem formulation. The formulation below closely follows the one described by Abur and Expósito in [9]. Given a set of m measurements and n states (where $m > n$), considering the vector of measurements $\mathbf{z} \in \mathbb{R}^m$ and its associated measurement error vector $\mathbf{e}$, we can write the corresponding relation between the two:

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

Here, $\mathbf{h}(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ is the nonlinear function that relates the measurements to the vector of states denoted by $\mathbf{x} \in \mathbb{R}^n$. The assumptions made for the errors in the measurements are:

1. The expected value of the error is 0, i.e, $E(e_i) = 0, i = 1, \ldots, m.$
2. Measurement errors are independent, i.e. $E[e_i e_j] = 0 \implies$ $\mathbf{R} = Cov(e) = E[\mathbf{e} \cdot \mathbf{e}^\top]$ $= \textbf{diag}\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2\}$,

where σ_i is the standard deviation of the ith measurement, which reflects the expected accuracy of the corresponding meter used to take the measurement. The WLS estimator minimizes the following objective function:

$$J(x) = \sum_{i=1}^{m} \frac{(z_i - h_i(x))^2}{\sigma_i^2}, \tag{2}$$

which can be written in matrix notation as

$$J(\mathbf{x}) = [\mathbf{z} - \mathbf{h}(\mathbf{x})]^\top \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \tag{3}$$

At the function minimum, the first-order optimality condition needs to be satisfied. This means that

$$\mathbf{g}(\mathbf{x}) = \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{H}(\mathbf{x})^\top \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] = 0 \tag{4}$$

Here, $\mathbf{H}(\mathbf{x}) = \left[\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}\right]$. Expansion of the left side of the nonlinear function $\mathbf{g}(\mathbf{x})$ in (4) using the Taylor series around the state vector $\mathbf{x}^k$ yields:

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x}^k) + \mathbf{G}(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \cdots = 0 \tag{5}$$

Here, $\mathbf{G}(\mathbf{x}) = \left[\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\right]$. From (4), we can write $\mathbf{G}(\mathbf{x}^k)$ as:

$$\mathbf{G}(\mathbf{x}^k) = \frac{\partial \mathbf{g}(\mathbf{x}^k)}{\partial \mathbf{x}} = \frac{\partial^2 J(\mathbf{x}^k)}{\partial \mathbf{x}^2} = \mathbf{H}(\mathbf{x}^k)^\top \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}^k) \tag{6}$$

Substituting (6) into (5) and neglecting the higher-order terms, we obtain:

$$\mathbf{g}(\mathbf{x}^k) + \mathbf{H}(\mathbf{x}^k)^\top \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) = 0 \tag{7}$$

Upon substitution of (4) into (7) and rearranging, we have:

$$\mathbf{H}(\mathbf{x}^k)^\top \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) = \mathbf{H}(\mathbf{x}^k)^\top \mathbf{R}^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x}^k)] \tag{8}$$

The above equation can be solved iteratively, in the form as shown below:

$$\mathbf{G}(\mathbf{x}^k)(\boldsymbol{\Delta}\mathbf{x}^{k+1}) = \mathbf{H}(\mathbf{x}^k)^\top \mathbf{R}^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x}^k)] \tag{9}$$

where $\boldsymbol{\Delta}\mathbf{x}^{k+1} = \mathbf{x}^{k+1} - \mathbf{x}^k$ and k is the iteration index. Equation (9) is typically solved using sparse matrix methods, along with triangular factorization followed by forward and backward substitution, as described in [10].

3.3. Tracking AC State Estimation Theory

The above formulation is the time-invariant case in which the state estimates are extracted from a single scan of measurements. It can be modified to reflect operations in real-time systems based on a time sequence of "snapshots" of systems measurements, referred to as a "tracking state estimator" by Monticelli [11]. Equation (1) can then be modified to

$$\mathbf{z}(t_i) = \mathbf{h}(\mathbf{x}(t_i)) + \mathbf{e}(t_i) \tag{10}$$

for $i = 0, 1, 2, \ldots,$ where $\mathbf{z}(t_i)$ is the measurement vector corresponding to time t_i; $\mathbf{x}(t_i)$ is the state vector at t_i; and $\mathbf{e}(t_i)$ is the zero mean, orthogonal (uncorrelated in time) vector with variance given by

$$E\{\mathbf{e}(t_i)\mathbf{e}^\top(t_j)\} = \begin{cases} \mathbf{R}(t_i), & t_i = t_j \\ \mathbf{0}, & t_i \neq t_j \end{cases} \tag{11}$$

We executed the DC power order on the HVDC line for equidistant I_{DC} *set points* spaced at 77.5 A, monotonically increasing from 0 A (zero current flow) to 6200 A (maximum current flow)—corresponding to the real power flow on the line increasing from 0 MW to 3100 MW. For a ramp time of 20 minutes and linear ramp rate, the power order execution is equivalent to solving the power flow of the network at 15-second intervals, for a total of 81 equidistant "snapshots" of the system as the power flow varies. Thus, for every time instant $i = 0, 1, 2, \ldots, T$, there is an associated DC *set point* I_{DC_i} that is determined by the ramp rate. This way of solving the power flow also means that for $i = 0, 1, 2, \ldots, T$, we generate a sequence of measurement snapshots $\mathbf{z}(t_1), \mathbf{z}(t_2), \ldots, \mathbf{z}(t_T)$. Each measurement snapshot, consisting of the real and reactive power flow from the sending and receiving end of each line, is exported in a .csv format, thus generating a set of measurements of the quasi-steady state of the system at each *set point*. We also recorded the real power injection at the AC buses connected to the converter and inverter buses. That serves as the result of the power flow against which the state estimate values are compared.

For each time interval $\Delta t = t_i - t_{i-1}$, we limit the number of iterations in the state estimation to k^{max}. For measurements at each time step, the iterative process indicated by (9) is continued as long as $k \leq k^{max}$, after which the time step is advanced to t_{i+1}, and the process is repeated.

The tracking state estimation process is now modified to incorporate the prioritization technique. The measurement vector $\mathbf{z}(t_i)$ at each time step t_i is considered to be made up of two non-intersecting subsets of measurements, referred to as *prioritized measurements* (of size $\mathbb{R}^{m_p}$) and *non-prioritized measurements* (of size $\mathbb{R}^{m_{\bar{p}}}$). At a given time instant t_i, the elements $z_j(t_i)$, where $j \in m_p$ of $\mathbf{z}(t_i)$, are received in real time, while the elements $z_j(t_i)$, where $j \in m_{\bar{p}}$, are computed based on the state estimate $\hat{\mathbf{x}}(t_{i-1})$ from the previous time instant t_{i-1}. As these *non-prioritized measurements* are not true measurements received from the system, we refer to them as *pseudo-measurements*.

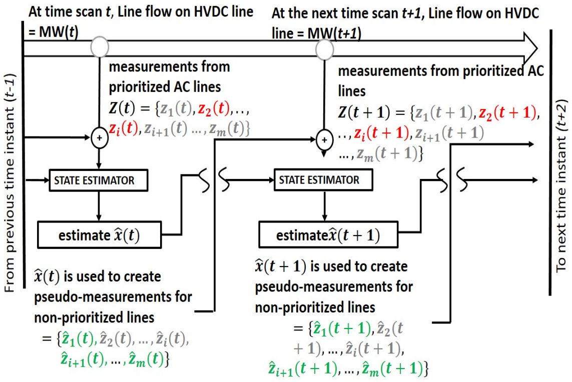
At time scan t, Line flow on HVDC line = MW(t)
At the next time scan t+1, Line flow on HVDC line = MW(t+1)
measurements from prioritized AC lines
$Z(t) = \{z_1(t), z_2(t), .., z_i(t), z_{i+1}(t) \ldots, z_m(t)\}$
measurements from prioritized AC lines
$Z(t+1) = \{z_1(t+1), z_2(t+1), .., z_i(t+1), z_{i+1}(t+1) \ldots, z_m(t+1)\}$
+
+
STATE ESTIMATOR
STATE ESTIMATOR
estimate $\hat{x}(t)$
estimate$\hat{x}(t+1)$
$\hat{x}(t)$ is used to create pseudo-measurements for non-prioritized lines $= \{\hat{z}_1(t), \hat{z}_2(t), \ldots, \hat{z}_i(t), \hat{z}_{i+1}(t), \ldots, \hat{z}_m(t)\}$
$\hat{x}(t+1)$ is used to create pseudo-measurements for non-prioritized lines $= \{\hat{z}_1(t+1), \hat{z}_2(t+1), \ldots, \hat{z}_i(t+1), \hat{z}_{i+1}(t+1), \ldots, \hat{z}_m(t+1)\}$
From previous time instant (t-1)
To next time instant (t+2)

Figure 1. Graphical depiction of the process of incorporating pseudo-measurements with real-time measurements.

Figure 1 provides a graphical representation of the process. The components of the measurement vector $\mathbf{z}(t)$, indicated in red, are the prioritized transmission lines for which measurements are received in real-time. The previous estimate of the state is utilized to generate *pseudo-measurements* for all other transmission lines, indicated in green. The combined hybrid measurement vector is then passed on to the state estimator to generate $\hat{\mathbf{x}}(t_i)$ at time step t_i. The process is then repeated at the next time step.

In effect, this process models the real-life scenario wherein the SOCC only has the ability to receive a very small set of measurements from the large system because of a limited number of metered nodes and lines where measurements are available.

3.4. The Use of Power Transfer Distribution Factors (PTDF) for Prioritization

The prioritization of measurements should be conducted using a metric that reflects the relative importance of elements in the power system network. The metrics considered for use in this prioritization scheme are a particular class of linearized sensitivity factors, outlined by Wollenberg and Wood in [12], called the power transfer distribution factors (PTDF), as investigated by Sauer [13]. PTDF is one among a class of distribution factors that are used for rapid contingency screening. PTDFs are a measure of how power injection at a specific bus in the network is distributed across the various transmission lines in the network. The formulation of a PTDF matrix is specific to a particular bus i with respect to a choice of reference bus r. Through the use of DC approximations in power systems [14], the following linear relation can be written:

$$\Delta\mathbf{P}_l = \mathbf{\Psi}_{i,r}\Delta\mathbf{P}_{inj}, \tag{12}$$

where $\Delta\mathbf{P}_{inj} \in \mathbb{R}^{N_b}$ is the vector of real power injection at the N_b buses in the system, $\Delta\mathbf{P}_l \in \mathbb{R}^{N_l}$ is the vector of changes in the real power flow across the N_l lines in the system, and $\mathbf{\Psi}_{i,r}$ is the PTDF matrix for the given choice of reference bus. Equation (12) describes the real power flow change in a line flow for a unit of real power exchange between bus i and the reference bus. To understand the distribution of line flow changes between two buses i and j in the network, the PTDF matrix $\mathbf{\Psi}_{i,j}$ can be computed by:

$$\mathbf{\Psi}_{i,j} = \mathbf{\Psi}_{i,r} - \mathbf{\Psi}_{j,r} \tag{13}$$

For the specific case under study, the buses i and j are the AC buses to which the converter and inverter stations of the HVDC system are connected. As a result, only the ith and jth element in $\Delta\mathbf{P}_{inj}$ are nonzero, while all other entries are zero. The product $\mathbf{\Psi}_{i,j}\Delta\mathbf{P}_{inj}$ then gives a vector of line flow changes $\Delta\mathbf{P}_l$ that can be normalized and sorted. Upon sorting those values for the 2470-bus, 5400-line reduced WECC case, we observed that a very large number of lines have a PTDF value near 0, and only about 50 lines have PTDF over 10%. This can be explained by understanding that for a given power transfer occurring between a set of two buses in the system, a small subset of lines will end up carrying a majority of the exchanged power, while the line flow of a large number of lines will be minimally affected.

4. Results

4.1. Measurement Prioritization

The PTDF analysis was done in PowerWorld. The lines were then sorted in descending order by the PTDF value. The determination of the acceptable threshold value was done by examining the trade-off in including fewer lines and the resulting estimates for the power injections at either end of the HVDC line. It was noted that as the set of *prioritized measurements* was increased, the estimates were generally found to become more accurate, at the cost of having to meter more lines, which in a physical system corresponds to increased costs for metering and larger data sets to be processed at every time step. Table 2 shows the number of lines that need to be considered for a particular threshold of PTDF value.

Table 2. Size of prioritized measurement set as a function of PTDF threshold

PTDF Threshold [%]	Number of Selected Lines
5%	135
10%	51
15%	28
20%	21
25%	14
30%	6
35%	5
40%	4
45%	1
50%	1

The errors in the estimate results were computed as the PTDF threshold was varied. The error metric used here is the relative error, defined as follows:

$$Error = \left| \frac{P_{inj,i}^{LF} - P_{inj,i}^{SE}}{P_{inj,i}^{LF}} \right| \quad (14)$$

Here, $P_{inj,i}^{LF}$ is the injection as computed from the load flow results, and $P_{inj,i}^{SE}$ is the injection computed based on the state estimate at the time instant t_i. It is also often expressed as a percentage.

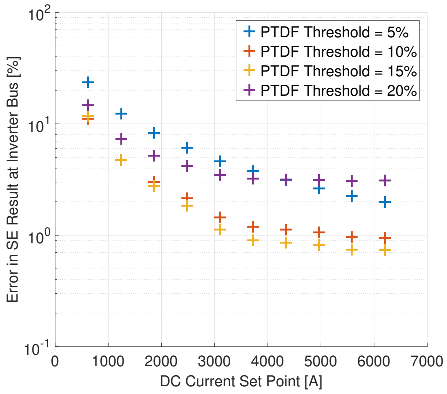

Figure 2. Comparison of the errors in the inverter end power as the PTDF threshold was varied.

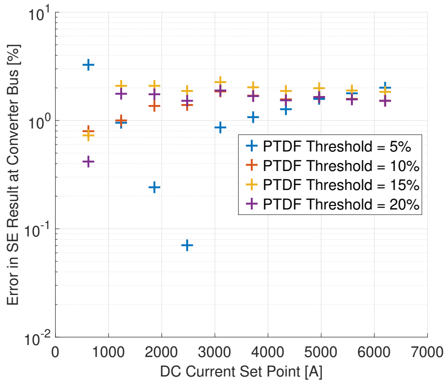

Figure 3. Comparison of the errors in the converter end power as the PTDF threshold was varied.

From the variation in the estimates due to changes in the threshold as seen in Figures 2 and 3, the error is found to be most well-behaved in both the converter and inverter cases when the PTDF threshold is 10%. This provided for sufficient accuracy and minimal variability across *set points*, while also keeping the number of prioritized (and therefore metered) lines to a minimum. Recall that the system has approximately 5400 lines, so fewer than 1% of the lines are considered for prioritization.

As this point, we should make a short note regarding the choice of using DC PTDFs instead of linearized AC PTDFs. DC PTDFs are computed based on the system topology and the use of the DC approximations, while the linearized AC PTDF is computed based on the system topology and the system's operating point. Thus, the computation of the AC PTDF requires real-time computation, while the DC PTDF can be a truly offline analysis that is invariant with respect to the system's operating point as long as the network topology does not change. DC PTDF therefore minimizes the amount of real-time computation that needs to be performed. There is some difference between the distribution factors with changes in system loading, as mentioned by Baldick in [15], who also provides certain criteria that ensure that the variability in the distribution factors is low. Furthermore, our use of PTDF in the study is not for the purpose of understanding line flow changes, but to inform our decision regarding which lines to prioritize for measurements. We are more concerned with the identification of the important lines than with the PTDF values of those lines.

Table 3. Size of prioritized subset of lines for AC vs. DC PTDF

PTDF [%]	AC PTDF	DC PTDF	Common Lines
5%	143	135	129
10%	56	51	49
15%	32	28	26
20%	24	21	20
25%	17	14	13
30%	8	6	6
35%	5	5	5
40%	5	4	4
45%	3	1	1
50%	1	1	1

For a chosen threshold value, we noticed a slight discrepancy in the number of lines that need to be considered depending on whether the AC or DC PTDF is used. Table 3 shows the number of lines that are shortlisted for a given PTDF threshold depending on whether the DC or AC PTDF is used. It can also be noted that the size of the set of lines common to both DC and AC PTDF sets is slightly smaller than the size of either the DC PTDF or the AC PTDF set. The use of the DC PTDF results in a selection of a slightly lower number of lines being selected than the AC PTDF would produce. Keeping the longer-term implementation in mind, the DC PTDF identifies a fixed subset of lines (invariant of the system's operating point) for which the communication channels would need to be upgraded and hardened to resist cyberattacks. Therefore, we used the DC PTDFs.

4.2. Determination of the Reduced Iteration Count

The determination of the reduced iteration count k^{max} was based on an analysis similar to that conducted for the variation of the results with the PTDF threshold. For a specific value of k^{max}, the relative error was calculated at every current *set point*. The error was expected to reduce as the value of k^{max} was increased. The error was observed to converge to a nonzero value for every *set point*. We investigated the marginal reduction in error as k^{max} was increased.

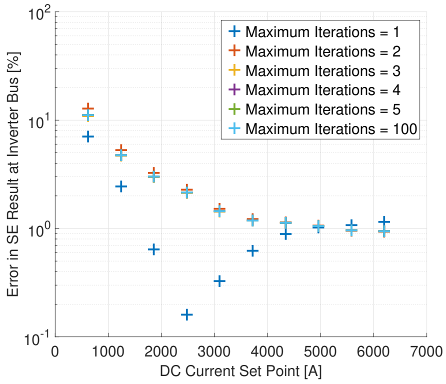

Figure 4. Variation of the error magnitude at the inverter end as a function of the maximum iterations allowed.

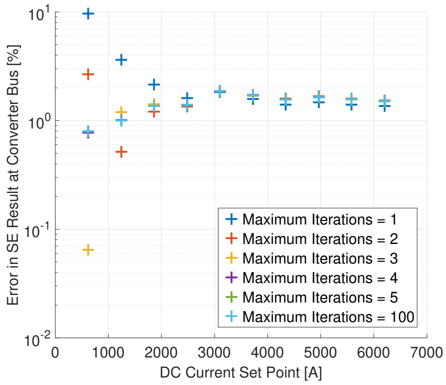

Figure 5. Variation of the error magnitude at the converter end as a function of the maximum iterations allowed.

Figures 4 and 5 show the variation of the state estimation error at the converter end (sending end) and the inverter end (receiving end) as k^{max} is varied. It can be noted that for some *set point* values, the error is found to be lower at lower values of k^{max}. That indicates that there are ranges of the DC *set point* for which convergence is reached a lot sooner. It can also be seen that within a relatively small iteration count, the estimates converge to a steady value, which is reflected by the overlap of several data points, particularly for higher values of real power flow on the HVDC line.

Another factor that influences the selection of k^{max} and requires attention is the time interval between time steps of measurements (which we refer to as the *polling interval*). If the time required per iteration of the state estimation algorithm is denoted by τ, a k^{max} should be chosen such that $\tau k^{max} < \Delta t$. A smaller polling interval means that the differences in system state between successive time steps is small, which results in a lower error in estimates, but also means that less time is available for state estimation computation. Therefore, while a shorter polling interval is desirable, it needs to be balanced to ensure that it does not significantly reduce the maximum number of allowed iterations. The observed value of τ was found to be bounded above by 0.1 s across all injection values.

We can observe that for values of $k^{max} > 5$, the marginal reduction in error is almost negligible. In Figures 4 and 5, the results for k^{max}=5 are indistinguishable from the results for k^{max}=100. Therefore, the lowest number of iterations that result in an acceptable error margin in the neighborhood of the true power flow solution is used to set the value of k^{max}.

4.3. Tracking State Estimator Accuracy

In this section, we investigate the accuracy of the tracking state estimator for a PTDF threshold and k^{max} chosen based on the analysis in the previous sections.

Based on the results shown in Figures 6 and 7, it can be seen that the tracking state estimator provides good agreement with the results of the power flow as the flow on the DC line is changed. It can be seen that the state estimator is able to track the changes in power injection with a good degree of accuracy. One thing to note is that the estimates are consistently lower than the injections computed from the load flow. That can be attributed to the correspondence of the *prioritized measurements* to a time instant t_i, while the *pseudo-measurements* are computed based on the state estimate at the time instant t_{i-1}. The state estimation is therefore attempting to compute one single state that minimizes the weighted sum of square errors, where the measurements are

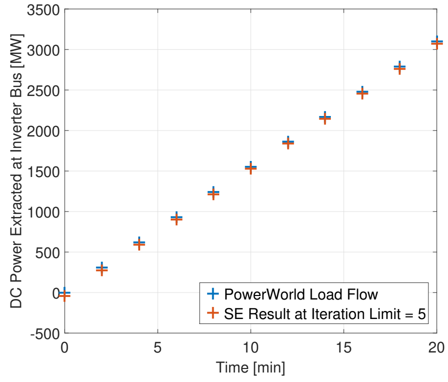

Figure 6. Comparison of the state estimation results with the load flow results at the inverter end of the HVDC line.

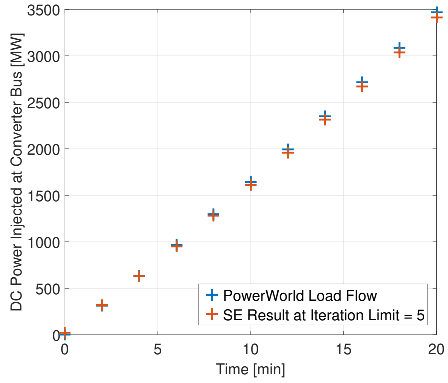

Figure 7. Comparison of the state estimation results with the load flow results at the converter end of the HVDC line.

pieced together from two different states.

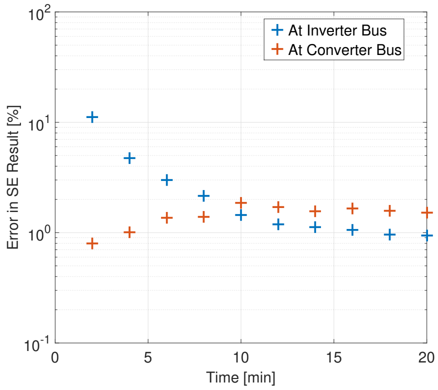

Figure 8. The relative error between the state estimate and load flow result at the converter and inverter end of the HVDC line.

Figure 8 shows results of the relative error of the estimates. The error at the converter end is approximately 2%, while the error at the inverter end is approximately 10% at the start of the injection, eventually reducing to about 1%.

4.4. Tracking Estimator under Simulation of Threat

Based on the results obtained from the tracking state estimator for the fast, approximate state estimation process, we performed a preliminary investigation of its ability to track a compromised power order. The initial power flow along the HVDC line was set to 1000 MW, and a power order of 1000 MW was set to be executed, for a final expected line flow of 2000 MW. That information allowed the SOCC to ascertain the expected injection and withdrawal profile at the converter and inverter buses, respectively. At execution, that order was corrupted to a higher value chosen to be between the intended final line flow and the maximum line capacity. The output of the state estimator was then observed as the corrupted power order was executed.

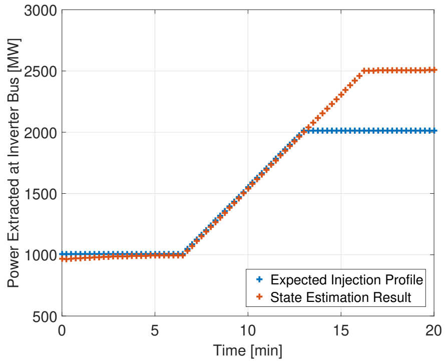

Figure 9. State estimation tracking for a corrupted power order of 1500MW. Polling interval = 15 s, and k^{max} = 5.

From Figure 9, we see that that state estimation tracked the state of the power system very closely. Having gained an understanding of the behavior of the errors in the previous section, we can gain insight on how to identify the improper execution of a power order based on the behavior of the error as the system state changes. In Figure 10, we see that as the state corresponding to the intended final flow on the line is passed, the error in the estimate begins to grow.

An increase in the errors as seen in Figure 10 is highly suggestive of a compromised or improperly executed power order.

A related profile that can be studied to track the system's evolution in time is the successive estimate

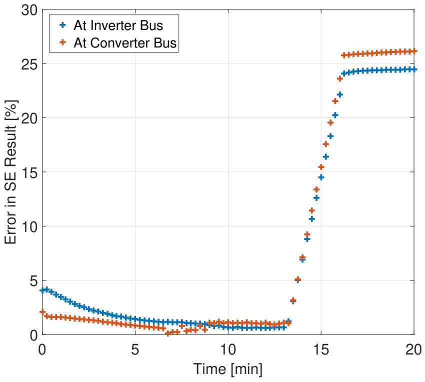

Figure 10. Observed error in state estimation results for the corrupted power order.

difference rate (SEDR), which is the difference between the power injection estimates computed at t_i and t_{i-1} divided by the time interval $t_i - t_{i-1}$. Figure 11 shows the SEDR profile for the simulated corrupt power order.

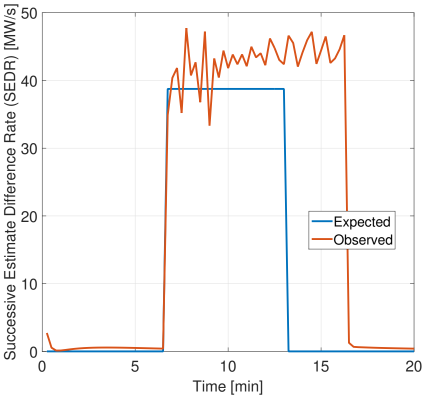

Figure 11. Expected and observed SEDR for the corrupted power order of Fig. 9.

In our linear ramp rate consideration, the expected and observed SEDR should appear pulse-shaped, displaying a constant nonzero value only when the power order is being executed. The observed SEDR profile shows noise, which can be attributed to the error introduced by the approximations in the state estimate process. For an uncompromised power order, the expected and observed SEDR profiles should match closely. This can be mathematically expressed as an SEDR integral criterion:

$$\left| \int_{t_0}^{t_f} \text{SEDR}_{\text{observed}}\, dt - \int_{t_0}^{t_f} \text{SEDR}_{\text{expected}}\, dt \right| \leq \varepsilon, \tag{15}$$

where t_0 and t_f are the limits on the time period of interest and ε signifies the acceptable margin of error. If we set the upper limit t_f to the time step t_i, we can compute the integral in real time as the state estimate computations are completed, which allows a detection time threshold to be assigned for a chosen value of ε. The SEDR integral criterion allows the extension of the detection strategy even when the ramp under consideration is not linear.

One caveat is that if the power order is altered to a value that is within the error bound of the state estimate, it cannot be distinguished from an error. However, if the error is maintained in a small interval, such a compromised power order will result in a comparatively less detrimental outcome. In effect, our approach bounds the degree to which an adversary can corrupt a power order and evade detection.

5. Conclusion and Future Work

We investigated and verified our assumptions on the fast, approximate state estimation process. Our assumptions about measurement prioritization and reduced iteration count were found to provide good agreement with the true state of the system. A simulation of the execution of the compromised power order was done, and the state estimation errors involved were used to qualitatively study the detection criteria.

The results of this paper show that the fast, approximate state estimation process can be utilized with promising results to detect malicious tampering with power order commands—within a margin of error—and make a large AC system with HVDC lines more resilient to attack. Our results also provide a means to identify the critical lines in a system for which communication channels should be hardened to be more resilient to a wider cyber-physical attack. Our methodology can be further extended to other problems of a similar nature that involve the time evolution of a power system's operating point.

Ongoing and future work will include implementation in a real-time, simulation-based hardware-in-loop environment to better reflect the realities of cyber-physical systems in terms of communication latency and asynchronous measurement acquisition through SCADA protocols such as DNP3. The sensitivity of the detection algorithm to topology changes is another direction for future work. It will be followed by investigation of mitigation strategies for increasing the cyber-resiliency of HVDC systems. Receiver operating characteristic (ROC) curves will also be studied to investigate the diagnostic ability of detection thresholds.

6. Acknowledgments

The work described here was performed with funding from the Dept. of Energy (DOE) under Cooperative Agreement DE-OE0000824, under subcontract to ABB US Corporate Research Center and in collaboration with the Bonneville Power Administration. The views expressed are those of the authors. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

The authors would also like to acknowledge support received from the University of Illinois at Urbana-Champaign. A. Valdes and R. Macwan are with the University of Illinois Information Trust Institute (ITI).

References

[1] "PowerWorld Simulator." [Online]. Available: https://www.powerworld.com/products/simulator/overview

[2] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[3] H. Lin, A. Slagell, Z. Kalbarczyk, P. W. Sauer, and R. K. Iyer, "Runtime Semantic Security Analysis of SCADA Networks to Detect and Mitigate Control-Related Attacks in Power Grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 1290–1300, 2009.

[4] "ACTIVSg10k: 10000-bus synthetic grid on footprint of western United States." [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg10k/

[5] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid Structural Characteristics as Validation Criteria for Synthetic Networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[6] F. C. Schweppe, "Power System Static-State Estimation, Part III: Implementation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 130–135, 1970.

[7] F. C. Schweppe and D. B. Rom, "Power System Static-State Estimation, Part II: Approximate Model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 125–130, 1970.

[8] F. C. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120–125, 1970.

[9] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[10] W. F. Tinney, V. Brandwajn, and S. M. Chan, "Sparse Vector Methods," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 2, pp. 295–301, 1985.

[11] A. Monticelli, *State Estimation in Electric Power Systems - A Generalized Approach.* Springer, 1999.

[12] A. Wood and B. Wollenberg, *Power Generation Operation and Control,*. Elsevier BV, 1996.

[13] P. W. Sauer, "On the Formulation of Power Distribution Factors for Linear Load Flow Methods," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 2, pp. 764–770, 1981.

[14] B. Stott, J. Jardim, and O. Alsac, "DC Power Flow Revisited," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1290–1300, 2009.

[15] R. Baldick, "Variation of Distribution Factors with Loading," *IEEE Transactions on Power Systems*, vol. 18, no. 4, pp. 1316–1323, 2003.