A Mechanism for Detecting Data Manipulation Attacks on PMU Data

Seemita Pal and Biplab Sikdar
Department of ECSE, Rensselaer Polytechnic Institute, Troy, NY, USA
Department of ECE, National University of Singapore, Singapore

***Abstract*—The fundamental role of the measurement and control information in the normal operation of smart grids makes cyber-security a critical necessity for existing and future power systems. This paper addresses the problem of detecting data manipulation attacks on smart grids, in the particular context of the data from Phasor Measurement Units (PMUs). The proposed methodology is based on comparing the estimates of the transmission line parameters as obtained from the PMU data with their known values. Data modification attacks are then detected when a statistically significant deviation is observed in the estimated and nominal values. In this proof-of-concept, work-in-progress paper, we verify the proposed detection methodology using mathematical analysis.**

***Keywords*—Cyber-security, smart grid, synchrophasor network**

I. Introduction

A Phasor Measurement Unit or synchrophasor provides real-time, highly accurate, precision time-stamped data about the frequency, voltage and current phasors of the buses on which it is located. PMU data is particularly useful for power system state-estimation, real-time monitoring, analyzing disturbances and power swings, and for power system operation, control and planning. Consequently, PMUs are considered to be an important part of current and future smart grids. However, as a result of their importance in the monitoring and control of power systems, PMUs and the data they generate are attractive targets for malicious attackers who seek to disrupt the normal operation of the power system. For example, the attacker may drop, modify or delay the PMU packets, any of which may create errors in the applications that are dependent on accurate and timely availability of PMU data, leading to outages, damage to equipment equipment, and economic losses. Cyber-attacks on PMUs and their data thus constitute a serious threat of immediate concern to the power infrastructure.

In this paper, we address the problem of detecting the presence of data modification attacks on PMU data. In such attacks, the attacker gains control of one or more links or routers in the network and then modifies the data in packets generated by the PMUs as they pass through the network. Data manipulation can adversely affect the operation of a power grid by introducing errors in the monitoring and control applications (e.g. state estimators) that use synchrophasor data. The problem of detecting data manipulation attacks has some similarities with the bad data detection problem in SCADA (supervisory control and data acquisition) based measurements. However, detecting data modified by an attacker is more difficult to detect since the attacker may exploit information about the power system to modify the values in such a way that it passes the methodologies to detect bad data. For example, it has been shown that attackers with information about the grid configuration can successfully inject arbitrary errors into certain state variables without being detected by the conventional bad data processing techniques [1].

This paper addresses the problem of detecting data manipulation attacks on PMU data by using the estimates of the transmission line parameters as the discriminant. This is a work-in-progress paper that aims to present the proof-of-concept of the proposed detection mechanisms. Our objective here is to present the details of the detection methodology and validate it mathematically. Experimental validation using real-life PMU traces is left as future works. The proposed detection methodology is based on exploiting the electrical properties (e.g. resistance, inductance and capacitance) of the transmission line. To detect the presence of modified data, we first use the PMU data to estimate the state of the line parameters for the buses in the electrical network. The estimated line parameters are then compared against the nominal values, and any significant statistical variation between them is taken as an indication of data manipulation. We mathematically show that the proposed mechanism is able to detect data manipulations, and is also effective against data modifications that will be missed by traditional bad data detection algorithms.

The rest of this paper is organized as follow. In Section II, we present an overview of PMUs, data modification attacks and their impact, and our system model. Section III presents the proposed mechanism for detecting data modification attacks and its mathematical validation. Finally, Section IV concludes the paper.

II. Background, Related Work and Assumptions

This section presents the background material related to the use of PMUs in the monitoring and control of power systems and the possible impacts of cyber-attacks on PMU data. Related work on detecting bad data in power systems when using SCADA-based measurement system is discussed. In this section, we also present the assumptions and system model used for our analysis.

A. PMU Data Attacks and Possible Impacts

A PMU is a power system device that measures frequency, voltage phasor and current phasors at the node where it is installed. The data are accurately time-stamped based on a common time source of the Global Positioning System (GPS)

This work was supported primarily by the ERC Program of the National Science Foundation and the Department of Energy under NSF Award Number EEC-1041877 and the CURENT Industry Partnership Program.

and provides measurements with accuracy better than 1%. These phasor data are usually sampled at a rate of 30, 50 or 60 samples per second and sent to the Phasor Data Concentrator (PDC) in the form of data packets at regular intervals. At the PDC, the data received from all its assigned PMUs are correlated into a single data set based on the associated time-tags. Grid-wide time-aligned PMU data are fed to the state estimator for determining the accurate states of the power system.

The classical methods of state estimation use active and reactive power measurements as inputs and provide a state estimation solution obtained in an iterative manner. However, state estimation performed using PMU data is more accurate and much quicker. Thus, the state estimator provides accurate snapshots of the power systems conditions at shorter intervals, sufficient for maintaining proper operation. The system state information may be used by the Energy Management System (EMS) to carry out functions such as automatic generation control (AGC), optimal power flow analysis and contingency analysis (CA).

The power grid is a critical infrastructure of any nation an thus an attractive target of attack by adversaries. Given their importance to a number of functionalities in smart grids, PMUs and the data they generate need to be secured against all forms of cyber attacks. Our interest is data manipulation attacks where an attacker may corrupt the data in three possible ways: by attacking the PMUs, by tampering with the communication network or by breaking into the synchrophasor system through the control center office LAN [2]. If an adversary is able to fake PMU data causing biasing of the power system state estimates without being detected, the operator may take erroneous control actions that are detrimental to the system. It can cause uneconomic dispatch choices, congestion, failure of generators, failures of transmission lines, as well as cascading failures leading to blackout. At the very least, if the operator is suspicious of the derived states, the distrust will create confusion regarding the true states of the system and observability will be hampered.

B. Related Work

While the problem of detecting malicious changes in the PMU data has received some attention recently, a significant body of work exists in the area of detection bad data in power systems. This section reviews the literature related to the problem of detecting data manipulations in power systems.

Traditional bad data detection techniques typically use redundant measurement data to compute measurement residuals in order to detect gross errors caused by sensor problems and/or telemetry failures. In the bad data detection techniques described in [3], [4], the 2-norm of the difference between the observed measurement vector and the estimated states is compared against a threshold to detect the presence of bad measurements. Although such conventional techniques are quite effective against random interacting measurement noises, they fail to detect highly structured manipulated data that conform to the network topology and some applicable physical laws.

The authors of [1] were the first to show that an attacker, armed with the current grid configuration information, may successfully inject arbitrary errors into certain state variables without being detected by the conventional bad data processing techniques. They presented a new class of attacks, called false data injection attacks from the attacker's perspective and performed analysis. In such attacks, highly-structured and coordinated data tampering can mislead the state estimation process without raising an alarm [1]. Based on the findings of [1], indices that quantify the least effort needed by attackers to achieve attack goals while avoiding bad data detection were introduced in [2]. These indices are related to the critical measurements without which observability is lost, and are thus most vulnerable as well as sensitive to data modification attacks [2]. In [5], the authors looked at false data injection attacks from an operator's point of view in order to determine how to defend against such attacks, in the context of smart meters. The authors provide a lower bound on the number of meters that need to be protected to thwart false data injection attacks in absence as well as presence of certain verifiable state variables. A Bayesian framework that leverages the knowledge of prior distribution on the states to detect false data injection attacks is proposed in [6]. The smallest number of meters that need to be tampered with by the attacker is modeled as an optimization problem in [7]. Similar to the attacker's strategy, a defender's strategy is proposed to develop an optimized set of secure measurements that can help in detection.

The results of the papers listed above are based on the same basic assumptions. The defenders in most of the cases use the original bad data processing technique with slight enhancement. For enabling detection of highly-structured data attacks it is assumed that either some of the PMUs are secured (i.e. their data cannot be tampered with) or some verifiable state variables are created. However, no practical as well as effective new or modified techniques have been proposed so far. In this paper, we propose a simple technique for the detection of data modification attacks in the PMU network. In this detection method, we have no requirements to have a set of PMUs that are immune from attacks, or that some of the state variables will be always available for verification.

C. Threat Model

The threat model assumed in this paper is that the adversary has compromised one or more of the PMUs, PDC, network routers and links or the communication system LAN at the control center. At each of the compromised nodes, the adversary is assumed to have the ability to manipulate or inject PMU measurement data in order to bias the power system state estimation. We do not make any assumption on the encryption of the data generated by the PMUs. Even if the data is encrypted, a data modification or data injection attack assumes that the encryption has been broken.

Under the adversary model described above, this paper considers the following cyber attack. We consider a scenario where power system measurement data is carried in data packets from the the PMU to the PDC, then on to the Super PDC, and finally to the control center, via a number of intermediate routers. It is assumed that an adversary compromises one or more of these mentioned nodes or links in the network and manipulates the PMU data in the packets. To maximize the damage, the objective of the adversary is to manipulate data to

the maximum extent possible without detection. The data manipulated by the attacker changes the estimated system states from their true values and larger deviations are more likely to lead to erroneous actions of greater consequence. However, even relatively small changes can cause uneconomic dispatch choices or billing manipulation over time. Our objective is to develop a mechanism that will effectively detect PMU data manipulation attacks that are performed by compromising one or more nodes of the communication system delivering the data.

III. PMU Data Modification Attack Detection Mechanism

In this section, we propose a data modification attack detection scheme based on verification of the line parameters in a distributed manner at the regional PDCs. First, the various transmission line parameters, their significance and possible variations are discussed. Additional benefits of estimating and constantly updating the line parameters and are also touched upon. Next, the detection mechanism is described in details.

A. Transmission Line Parameters

Transmission line parameters in a power system play an important role in relay-setting, accurate state estimation, location of line faults and dynamic estimation of the maximum load of a line. Distance relays use impedance values of the lines for setting the proper zone. Also, state estimation techniques use the values of the line parameters for estimating the system states and accurate information is required for error-free and reliable solutions. Fault locating algorithms utilize transmission line parameter information for determining the location of faults [11].

Electrical transmission lines can be represented by four parameters: resistance (R), inductance (L), capacitance (C) and conductance (G). Resistance and inductance are uniformly distributed along the line length and constitute the series impedance. Capacitance and conductance, on the other hand, exist between the conductors and/or between conductor and neutral, and constitute the shunt impedance of the transmission line.

The resistance is affected by temperature as well as skin effect and thus exhibits considerable variations. With change in the operating temperature, the resistivity of the conductive material varies and hence causes the resistance to change. In case of alternating current (ac), current distribution is not uniform throughout the conductor and this non-uniformity increases at progressively at higher frequencies. This in turn leads to higher current density near the surface than the center and ultimately results in raising the effective resistance. Even at the working frequency of the power system, sufficient change is observed due to this phenomenon which is called the skin effect.

The inductance is due to the voltage induced by the magnetic field produced by the changing conductor current. It is the most dominant line parameter. The line reactance which depends on the working frequency and the inductance is, therefore, much larger than the line resistance. Thus, the resistance is often neglected in the study of transmission line behavior [13].

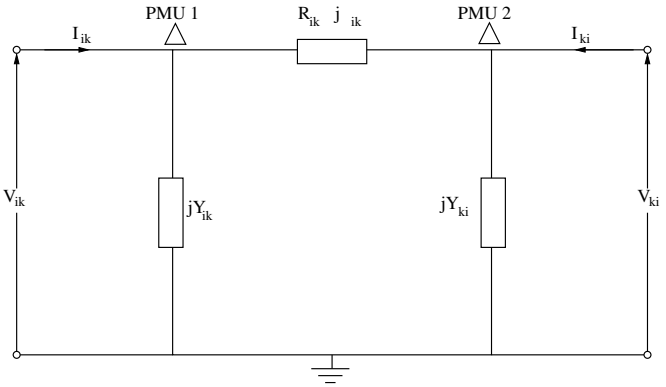

Fig. 1. Example topology.

The capacitance in transmission lines is present due to the effects of the electric fields around the conductor. It is almost constant depending on the size and the spacing of the conductors. For short transmission lines (<80 Km or 50 miles), the effect of the capacitance is negligible but it becomes increasingly significant for longer lines. The susceptance of the transmission line depends on the working frequency and the capacitance.

The fourth parameter, conductance, is caused by the leakage current over the surface of the insulators. Since this leakage current is negligible in case of overhead lines, the conductance between conductors of an overhead line is almost always neglected. In our detection mechanism, we have neglected the conductance.

B. Detection Method

Our objective is to develop a mechanism to detect the integrity of the voltage and current phasor measurements taken from PMUs. In power systems, PMUs are typically present on a number of high-voltage buses. The PMUs deployed over a specific region send their data to the regional PDC where these data are assimilated, time-aligned according to the GPS time-stamp and then sent off to the Super PDC. The proposed detection method can be performed at the regional PDCs or at the Super PDC.

In the proposed methodology, PMU measurements are taken as the input and Gauss-Newton iterative method is performed to obtain the transmission line parameters by minimizing the residuals. The nominal values of the line parameters are assumed to be known. The chi-square test is performed based on the known values and the estimated ones to detect possible anomalies. When data is manipulated by the attacker, the estimated values are expected to be statistically different from the normal values.

Consider a power system with N buses and let these buses be labeled as $i = 1, 2, \cdots, N$. In order to analyze the interaction of the bad data and the effects on the estimates, we assume that all the buses are equipped with PMUs so that estimates can be compared to the measured values. The measured bus voltage magnitudes and their corresponding voltage phase angles are represented by Vm_i and θm_i respectively. The PMUs also measure line current flows on L branches between the buses. The magnitudes and the phase angles

of the currents flowing from any bus i to another bus k, measured by the PMU are denoted by I_{ik} and δ_{ik} respectively. The total number of available measurements in the model are $2N$ voltage magnitudes and angles and $4L$ line current magnitudes and angles, in case current flows in both directions are available. Let the total number of measurements be m, i.e., $m = 2N + 4L$. This measurement vector is arranged as:

$$z = [Vm_i \quad \theta m_i \quad I_{ik} \quad \delta_{ik}]^T \tag{1}$$

The network model is developed based on circuit equations, using the equivalent circuit in Figure 1 showing the connection between any two buses in the network. The branch connecting buses i and k is represented by the transmission line with impedance $Z_{ik} = R_{ik} + jX_{ik}$ and line charging $Y_{ik} = jB_{ik}$ where R_{ik}, X_{ik} and B_{ik} are the series resistance, series reactance and shunt susceptance respectively. The shunt conductance of the transmission line is very small and is therefore neglected. We do not use the estimate of the line resistances in the detection mechanism since they vary considerably due to temperature variations and skin effect. We assume that the nominal values of line reactance and susceptance are known and are given by Xm_{ik} and Bm_{ik} respectively.

Our state estimation problem requires solving for n states, which consist of the N voltage magnitudes, N voltage phase angles, L line reactances and L line susceptances, i.e. $n = 2N + 2L$. The objective is to solve for the states while minimizing the measurement errors. Here, since $m > n$, the problem can be formulated as a nonlinear weighted least-squares (WLS) problem. Redundancy of measurements also ensures greater accuracy. Even if all $2N + 4M$ measurements are not available, the formulation can be conveniently done as long as $m > (2N + 2L)$. In the case $m = n$, the number of unknown variables is equal to the number of constrained equations and a unique solution for the unknown variables exists. In this specific case of equal number of states and constraints, however, the measurement errors are taken to be zero.

Let x represent the values of the n states, and be given by

$$x = [V_i \quad \theta_i \quad X_{ik} \quad B_{ik}]^T \tag{2}$$

where $m > n$. The relationship between the system states x, and the measurements z is given by:

$$z = h(x) + e \tag{3}$$

where $h(x)$ is a measurement function relating the measurements to the state vector and e is the vector of measurement errors. The partial derivatives of the residual equations with respect to the states is computed to obtain the measurement Jacobian matrix H. The WLS problem can then be formulated as:

$$\begin{aligned} \underset{x}{\text{Minimize}} \quad & {R_i e_i}^2 \\ \text{subject to} \quad & e_i = z_i - h(x_i),\ i = 1, \cdots, m. \end{aligned}$$

In the optimization function, a weight R_i is applied in order to account for the variance of each measurement and solve the optimization using the Gauss-Newton iterative method. R is a diagonal matrix with weighting factors inversely proportional to the square of the measurement accuracy of each PMU.

To solve the WLS problem, we start by initializing the states: the voltage magnitudes are set to 1 and the phase angle biases are set to 0. In each iteration, the increment of the state values is obtained using

$$\Delta x = G^{-1}(RH)^T Re \tag{4}$$

where G is the gain matrix and is usually chosen as,

$$G = (RH)^T(RH). \tag{5}$$

The new state is updated to $(x + \Delta x)$ and the Gauss-Newton iteration is repeated until the solution converges. The results provide the estimated voltage magnitudes, associated voltage phase angles, line reactances and susceptances. The voltage magnitudes are expected to be near 1 (normalized). The voltage drop in a practical transmission line is generally limited to about 5% and therefore the estimated voltage can be used as an indicator of convergence. The voltage phase angles and line parameters are used for comparing against the known nominal values and verifying the authenticity of the PMU data.

After the state estimates are obtained, we test the results for presence of modified data. Let v_i be the nominal value of the n^{th} variable whose corresponding estimated value is x_i. For each of these variables, the statistical distance of the estimated value from the nominal values are computed for determining a distance based chi-square statistic as follows:

$$X = \sum_{i=1}^{n} \frac{(x_i - v_i)^2}{v_i}. \tag{6}$$

X^2 is small if an observation of the variables is close to its expectation. The mean $\bar{X}^2$ and standard deviation ${S_X}^2$ of X^2 of the population can be estimated from the sample data. The in-control limits to detect data modifications can be set to obtain 3-sigma control limits. Only the upper control limit of $\bar{X}^2 + 3{S_X}^2$ is of our interest as significantly large X^2 values indicate sufficient deviation of the observed values from the nominal values for being able to conclude possible tampering of measurements. If the computed X^2 for an observation is greater than $\bar{X}^2 + 3{S_X}^2$, we raise the alarm for PMU data attack.

The proposed data manipulation detection algorithm is shown in Algorithm 1. One of the additional advantages of this method is that the line parameters estimated in this manner will serve as the updated values and can be used in the other power system applications mentioned earlier in this paper for more accurate results.

C. Mathematical Validation

In this section we demonstrate the effectiveness of the proposed detection mechanism using mathematical analysis. To show the effectiveness of the proposed mechanism, we show that any data modification attempted by an attacker will cause the line parameter estimates to deviate beyond the expected limits and thereby indicate possible attack. Note that such attacks will be missed by traditional bad data detection mechanisms.

We assume that the attacker has access to t PMUs in the network. The attacker can modify the measurements of all these PMUs. Let z_a represent the vector of the observed measurements where $z = (z_1, \cdots, (z_m))^T$ is the vector of

Algorithm 1 PMU Data Manipulation Attack Detection

loop
  for Arrival of PMU packets with time-stamp 't' **do**
    Update the measurement set z^t;
    Count number of available measurements m;
    Initialize the values of the states x^t;
    Count the number of state variables n;
    if $m > n$ **then**
      STATE ESTIMATION(t)
    end if
  end for
end loop Session is terminated

function STATE ESTIMATION(t)
  Initialize tolerance: $\epsilon = 1$;
  Initialize iteration number: $k = 0$;
  Calculate the weight matrix R;
  while $\epsilon > 1\text{E} - 5$ **do**
    Update iteration number: $k + 1$;
    Compute Hessian matrix $H(x)$;
    Calculate error vector: $e^t = z^t - H(x^t)$;
    Calculate gain matrix: $G = (RH)^T(RH)$;
    Calculate: $\Delta x^t = G^{-1}(RH)^T Re^t$;
    Update state vector: $x^t = \Delta x^t + x^t$;
    Calculate $\epsilon = norm(\Delta x^t)$;
  end while
  Estimated state vector for time-stamp 't': x^t;
  Number of required iterations: k;
  CHI-SQUARE TEST(t)
end function

function CHI-SQUARE TEST(t)
  Vector of nominal values: v;
  Calculate: $X^t = \sum_{i=1}^{n} \frac{(x_i^t - v_i)^2}{v_i}$;
  Mean and standard deviation of statistic: $\bar{X}^2$ and ${S_X}^2$
  if $X^t > \bar{X}^2 + 3{S_X}^2$ **then**
    Generate alarm for "Data manipulation attack";
  else
    Update $\bar{X}^2$ and ${S_X}^2$;
  end if
end function

real measurements and $a = (a_1, \cdots, (a_m))^T$ is the vector of malicious data injected by the attacker or the attack vector. Therefore,

$$z_a = z + a \tag{7}$$

In the traditional bad data processing techniques, the maliciously modified measurements will not raise any alarm if a is a linear combination of the column vectors of H, that is, $a = Hc$. Let $\hat{x}$ be the vector of true estimates of the states when the measurements are not manipulated, and let $\hat{x_{bad}}$ be the corresponding estimates obtained with the manipulated measurements. Therefore, if $a = Hc$,

$$||z_a - H\hat{x_{bad}}|| \leq \tau \tag{8}$$

where, τ is the detection threshold. Thus, traditional bad data detection methods fail to detect such coordinated attacks. Also, $\hat{x_{bad}} - \hat{x} = c$. Thus, c is the error injected in the estimated states, thereby, biasing them. It is a non-zero vector, each of whose elements can be an arbitrary number [1].

However, in the proposed data manipulation attack detection method, the nominal values of the line parameters are already known and hence the elements of c corresponding to those states should always be zero. Also, it is well known that the values of the voltage magnitudes are typically close to 1. Thus, manipulation of the voltage magnitude states is difficult and for evading detection, the elements of vector c corresponding to the voltage magnitude states should be as small as possible, ideally zero. Further, the farther the value of any voltage magnitude is away from 1, the more the alarm that it will raise regarding the grid operations. Thus if the attacker succeeds in manipulating the measurements and biasing the states, the obtained states can be compared against the nominal values and the attack can be easily detected.

IV. CONCLUSIONS

This paper proposed a mechanism for detecting the presence of data manipulation attacks on PMU data. The proposed mechanism can detect attacks irrespective of the specific PMUs targeted by the attacker and unlike many existing mechanisms for detecting bad data, it does not require the assumption that either some of the PMUs are absolutely secure or that some of the states are verifiable. Mathematical verification of the method has been provided.

REFERENCES

[1] Y. Liu, P. Ning and M. Reiter, "False data injection attacks against state estimation in electric power grids", *Proceedings of ACM CCS,* Chicago, IL, November 2009.

[2] H. Sandberg, A. Teixeira, and K. Johansson, "On Security Indices for State Estimators in Power Networks", *First Workshop on Secure Control Systems,* Stockholm, Sweden, 2010.

[3] E. Handschin, F. Schweppe, J. Kohlas and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems,* vol. 94, no.2, pp. 329-337, March 1975.

[4] M. Baran and A. Abur, "Power System State Estimation," *Wiley Encyclopedia of Electrical and Electronics Engineering,* 1999.

[5] R. B. Bobba, K. M. Rogers, Q. Wang, and H. Khurana, "Detecting false data injection attacks on DC state estimation", *In Proceedings of the First Workshop on Secure Control Systems,* 2010.

[6] Kosut, O.; Liyan Jia; Thomas, R.J.; Lang Tong, "Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures," *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on* , vol., no., pp.220,225, 4-6 Oct. 2010.

[7] Kim, T.T.; Poor, H.V., "Strategic Protection Against Data Injection Attacks on Power Grids," *Smart Grid, IEEE Transactions on* , vol.2, no.2, pp.326,333, June 2011.

[8] http://www.nbcnews.com/tech/security/hackers-launched- cyberattack-u-s-public-utility-government-says-n110476, Date accessed 06/30/2014.

[9] J. Meserve, "Sources: Staged cyber attack reveals vulnerability in power grid," *CNN,* Sept 26, 2007.

[10] A. Wood and B. Wollenberg, *Power Generation, Operation and Control,* 2nd ed. John Wiley and Sons, 1996.

[11] Dasgupta, K.; Soman, S.A, "Line parameter estimation using phasor measurements by the total least squares approach," *Power and Energy Society General Meeting (PES), 2013 IEEE*, pp.1,5, 21-25 July 2013.

[12] Ghiocel, S., "Applications of Synchronized Phasor Measurements for State Estimation, Voltage Stability and Damping Control", May 2013.

[13] J. Grainger and W. Stevenson, Power System Analysis. New York: McGraw-Hill, 1994.