Malicious Data Detection in State Estimation Leveraging System Losses & Estimation of Perturbed Parameters

William Niemira Rakesh B. Bobba Peter Sauer William H. Sanders
University of Illinois at Urbana-Champaign
Email: {niemira2, rbobba, psauer, whs}@illinois.edu

***Abstract*—It is critical that state estimators used in the power grid output accurate results even in the presence of erroneous measurement data. Traditional bad data detection is designed to perform well against isolated random errors. Interacting bad measurements, such as malicious data injection attacks, may be difficult to detect. In this work, we analyze the sensitivities of specific power system quantities to attacks. We compare real and reactive flow and injection measurements as potential indicators of attack. The use of parameter estimation as a means of detecting attack is also investigated. For this the state vector is augmented with known system parameters, allowing both to be estimated simultaneously. Perturbing the system topology is shown to enhance detectability through parameter estimation.**

I. Introduction

Operators of the power grid rely on large networks of sensors, known as SCADA (Supervisory Control and Data Acquisition) systems, to monitor conditions on the grid. To incorporate the many redundant measurements available into a single coherent picture, the grid state is estimated from the output of these sensors periodically. The state estimates affect both operational and economic functions, so it is critical that state estimates accurately reflect the grid state.

State estimators are classified as DC state estimators, utilizing a linear system model, or AC state estimators, using a nonlinear model. Under the DC model, typical measurements consist of real power flows and injections and states consist of bus angles [1].

Traditional bad data detectors are designed primarily to deal with random sensor noise or errors. Potential sources for noisy or erroneous measurements could be non-simultaneity of measurement sampling, incorrect installation of or damage to a sensor, or communication errors. These schemes are designed to detect isolated, random bad data. Interacting bad data is more difficult to detect. A subclass of interacting bad data is malicious data injection attacks.

Malicious data injection attacks against state estimators in power systems gained attention after they were written about in [2]. These attacks consist of coordinated modifications to measurements, such that modifications are coherent with the linearized model and unmodified measurements. This conceals the presence of attacks from DC state estimators completely. The effect of linear model attacks on nonlinear estimators was shown to be diminished in [3]; however [4] showed the potential for DC attacks to succeed on real EMS software using a nonlinear model.

This paper examines the sensitivity of real and reactive power measurement residuals in a nonlinear state estimator to false data injection attacks based on a linearized model. The effects of malicious data injection attacks are shown to have varying degrees of effect on different measurement types and that residuals of some measurement types are a much better indicator of attacks. This paper also presents a detector based on parameter estimation combined with topology perturbation of estimated parameters.

II. Prior Work

A. Bad Data Detection

Traditional bad data detection generally assumes that errors will be large and isolated, resulting in a poor fit to the model. Detection of isolated bad data has been addressed in many works. As examples, a method using normalized residuals is addressed in [5] and another based on the coherency between measurements with large residuals and the other measurements in [6].

Detection of multiple bad data using various methods has also been investigated. Examples include [7], where a linear program was used to detect multiple bad data sources that did not fit a Gaussian noise model, and [8], where hypothesis testing was used for multiple and interacting bad data detection.

B. Malicious Data Attacks and Detection

Data injected deliberately by an adversary would not be expected to follow the same patterns as random bad data. A method of injecting bad data into linearized state estimation (DC model) without changing measurement residuals was demonstrated in [2] and [9]. These attacks are not detectable by traditional bad data detection schemes accompanying linear state estimators and are dubbed stealth attacks. To construct these attacks, the attacker needs some knowledge of the system topology. It was shown in [10] that full knowledge of system topology is not necessary, and also that the attack based on the linearized model can be effective against a non-linear estimator. Although [3] shows decreased effectiveness of linear attacks when used against non-linear estimators, [4] shows that real EMS software is indeed vulnerable to attack.

A protection scheme against data injection attacks, based on hardening of sets of measurements to make them invulnerable to attack, is shown in [11]. Measurements are selected such that observability is guaranteed for the operator, which is shown to be necessary and sufficient to prevent stealth attacks.

A Bayesian detector for malicious data injection attacks is formulated in [12]. The efficient use of encryption to thwart attacks is investigated in [13], where an algorithm to maximize the utility of encrypted measurement placement is developed.

A financial incentive for attacks on state estimators is described in [14], where it is shown that the use of malicious data injection by an adversary can manipulate energy markets such that an attacker can guarantee trading profits.

C. Topology Perturbation

A topology perturbation based detector for detection of malicious data is developed in [15]. That approach uses known topology perturbations (specifically changes to impedance) applied to a power system as probes. The expected system response is predicted and compared against the actual measurements. The topology perturbations are hidden from an attacker, so the attacker is unable to correct the attack to remain stealthy. If measurement values do not change as computed beforehand, an attack is assumed. Although effective, this method has the disadvantage that it has to assume no changes in load and generation between probes or account for such changes making it difficult. The approach here is to estimate the parameter changes with measurement data rather than to predict measurement changes. Furthermore instead of probing the system, the estimate of the parameters can be used for detection by leveraging regularly occurring perturbations.

III. Approach

A. Attacker Model

The attacker is assumed to have access to baseline topology information necessary to formulate a DC attack offline i.e. at least one of the columns of the sensitivity matrix used for linear state estimation. This matrix is H in Equation (1), where z is the vector of measurements and x is the vector of states. This information is assumed to be static, and will not reflect changes such as line outages, tap changes, or other topology or parameter changes. The attacker is also assumed to have the ability to change a subset of measurements needed for an attack. The change could occur by corrupting the measurement device at the substation, interfering with communication between the substation and control center, or by installation of malware at the control center. The attacker is not assumed to have the ability to observe conditions across the entire system.

$$z = Hx \tag{1}$$

B. Nonlinear Sensitivity Analysis

1) Overview: The sum of squared residuals of a state estimators output quantifies how well the measurement data fits the model. This is useful for bad data detection, as bad data would not be expected to fit the model well. Unlike bad data, malicious data is designed to fit a simplified model, decreasing the impact on measurement residues. Using knowledge of the DC model, we can predict what measurements would tend to be impacted most by an attack. This work focuses on real and reactive power measurements. Two classes of measurements, flows and injections, are considered here.

The DC model neglects the effects of losses, uneven voltage profiles, and reactive power on the power system. When an AC state estimator is attacked using attacks based on a DC model, the residues increase due to these simplifications. The magnitude of effect on real and reactive flows and injections varies by measurement type.

For real flows, the DC model tends to underestimate their magnitude due to neglecting real power losses. The impact of this simplification increases with the square of current on a line. Reactive flows also have losses that increase with the square of line current, however the high X/R ratio of power lines means that reactive power transmission is inherently lossier than real power transmission. When combined with the DC models neglect of reactive power, the expected effect is for malicious data injection attacks to affect reactive flow residues more than real flow residues.

The error from neglecting losses becomes more apparent when generator power output is monitored. Losses have to be supplied by a generator, so errors on flows due to neglecting losses will be accumulated at generators.

To verify these intuitions, and to examine the detectability of malicious data injection attacks, the effect of attacks of various magnitudes on a composite weighted residual, and on residues for real flows, reactive flows, real generator injections, and reactive generator injections was investigated.

2) Establishing Baseline: Baseline residual are needed for comparison purposes. Measurement residues are expected to vary under normal circumstances due to noise. Monte Carlo trials were used to establish the expected distribution of residuals. Noise vector n was added to noiseless system quantity vector z (obtained through simulation), resulting in measurement vector z^* as seen below in Equation (2).

$$z^* = z + n \tag{2}$$

State estimation was conducted for many such random noise vectors, and sums of squared residues of real and reactive flows and generator injections were recorded separately and together as a weighted composite.

Using the distributions generated in this fashion, a maximum expected residue can be established. For values higher than this, malicious data is assumed. The cutoff is selected based on the acceptable proportion of false alarms. For example, if 1% false alarms are acceptable, the cutoff chosen would be a residue value higher than the residue for 99% of trials.

3) Comparison of Detectability: To determine the detectability of a particular attack vector at a particular magnitude, the Monte Carlo procedure used in the previous section was repeated for a system under attack. To each z^* generated previously, attack vector a was added, resulting in z^{**} as seen below in Equation (3).

$$z^{**} = z^* + a = z + n + a \tag{3}$$

The detectability of an attack by a residual can be quantified

by the proportion of measurements above the baseline cutoff established earlier. For example, if 80% of the distribution of residuals for the attacked system lies above the cutoff, that attack would be considered 80% detectable. This allows comparison of the effectiveness of residuals of specific measurement types for detecting bad data, as well as comparing the detectability of different attacks.

The total detectability is the non-overlapping proportion of attacks detected by any of the metrics. This may be larger than detectability based on any single residual type.

C. Parameter Estimation

1) Overview: The use of parameter estimation to detect maliciously injected data was also investigated. The state estimator was modified for this purpose. State vector x was augmented with parameters p to form $\tilde{x}$ as in Equation (4). The equation used to relate measured system values to states becomes a function of parameters as well, as shown in Equation (5).

$$\tilde{x} = \begin{pmatrix} x \\ p \end{pmatrix} \tag{4}$$

$$\tilde{z} = h(\tilde{x}) \tag{5}$$

And where $\tilde{z}$ is the measurement vector z augmented with parameter estimates $\bar{p}$ as in Equation (6).

$$\tilde{z} = \begin{pmatrix} z \\ \bar{p} \end{pmatrix} \tag{6}$$

The sensitivity matrix H, which is the function h as evaluated at each iteration is also modified as shown in Equation (7) (6).

$$\tilde{H} = \begin{pmatrix} H_x & H_p \\ 0 & I \end{pmatrix} \tag{7}$$

The parameters estimated could be any system parameter.

In this work, line series reactance was considered as a parameter to estimate. Distributed flexible AC transmission system (D-FACTS) devices were assumed to have been installed on the system. This allows line impedance to be controlled by the control center, and might be done for congestion management and loss minimization.

Installation of D-FACTS devices turns line reactances into variable parameters. If parameter values change, i.e. the D-FACTS setting is altered, attacks based on pre-change values will perform poorly. If an attack is constructed for a system using the wrong parameter values, the interactions between attacked measurements tends to drive parameter estimates toward the values used to construct the attack vector. This increases parameter estimate residues.

The residues on parameter estimates are more meaningful for the purpose of malicious data detection than measurement residuals because the true value of the parameter can be known with more precision. The parameter value, unlike measurements, should not be affected by grid state. This makes comparison of estimated parameters with their known values useful for malicious data detection.

Detection of malicious data injection attacks is possible whenever the parameters are altered for control purposes and the alteration can be concealed from the attacker at least for a while.

2) Establishing Baseline: As with measurement residuals in Section III-B2, baseline parameter residual values had to be established for comparison purposes. To do this, noise vector n was added to noiseless system quantity vector $\tilde{z}$ (obtained through simulation), resulting in $\tilde{z}^*$ as seen below in Equation (8). Noise was limited to measurement values (not parameters).

$$\tilde{z}^* = \tilde{z} + n \tag{8}$$

3) Determining Detectability: The procedure of Section III-B3 was repeated to generate distributions of parameter residuals. To each $\tilde{z}^*$ generated previously, attack vector a was added, resulting in $\tilde{z}^{**}$ as seen below in Equation (9).

$$\tilde{z}^{**} = \tilde{z}^* + a = \tilde{z} + n + a \tag{9}$$

The effects of perturbations on detectability can be determined by comparing distributions generated for the same attack vectors under different system perturbations to each other and to the base case.

IV. Evaluation

A. Nonlinear Sensitivity Analysis

1) Setup: Analysis was conducted on the IEEE 14-bus test case. State estimation was executed using MATPOWER, a MATLAB power system simulation package. Each noise vector n used for generating distributions in Monte Carlo trials had entries consisting of a normally distributed random variable with zero mean and standard deviation of 1% of the measurement value corresponding to the entry in n. Measurements used were real and reactive flows and generator injections. Distributions of measurement residuals for determining baseline values and attack detectability were generated by executing state estimation using 500 different noise vectors. The cutoff used to establish the baseline as described in the previous section was the 99th percentile (i.e. 1% false alarms).

2) Establishing Baseline: Data consisting of sums of squared residues of real and reactive flows and injections were recorded separately and as a weighted composite. Histograms of the result were generated from this data. An example for real power flows, is shown below in Figure 1.

From these histogram bin counts, a cumulative density function plot was created by normalizing the histogram to have an area of 1 and computing the cumulative sum of bin counts. The 99% cutoff was established by finding the sum of squared residue value of the CDF at the 99th percentile. An example for reactive power flows is shown below in Figure 2.

3) Determining Detectability: Distributions of sums of squared residues were constructed for the system under various attack regimes. In theory any linear combination of columns of the DC H-matrix can serve as a stealth attack vector. For our

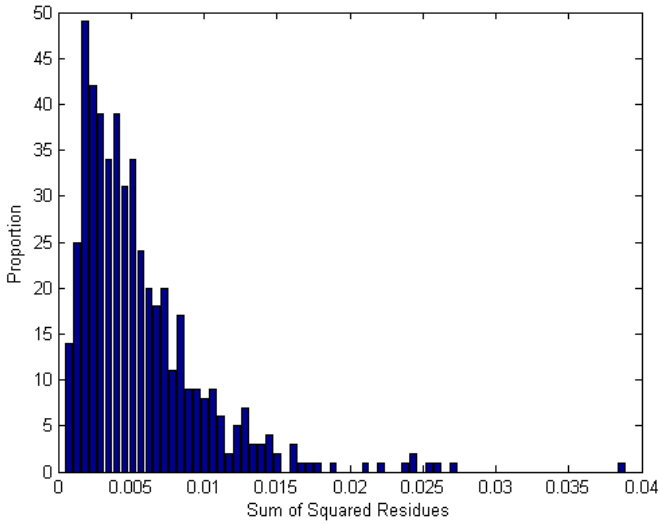

Fig. 1. Histogram of sum of squared residues of reactive power flows for 500 random noise vectors.

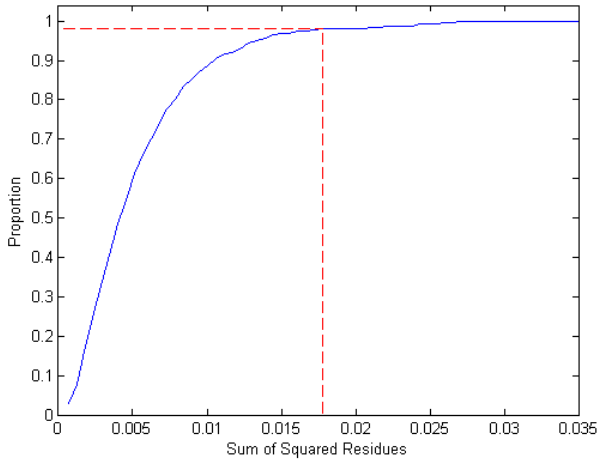

Fig. 2. CDF of sum of squared residues of reactive power flows for 500 random noise vectors. Horizontal line indicates 99th percentile. Vertical line indicates value for sum of squared residues corresponding to 99th percentile.

experiments we used individual columns of the DC H-matrix albeit scaled as attack vectors. Each column vector contains entries corresponding to measurements of quantities at a bus and all incident branches to that bus. For example, the first column of the H- matrix has entries corresponding to power injected at bus 1 and power flows on all lines incident to bus 1.

Attack vectors were scaled such that the largest entry in each attack vector corresponded to a 10 MW injection (0.1 p.u. in 100 MVA base). Subsequent trials increased the scale of attack vectors in 10 MW steps up to 100 MW, so that each of the 14 columns of the DC H-matrix was used at each of 10 different power levels (10 MW to 100 MW in 10 MW increments). For every distribution generated, the percentage of measurements falling above the previously determined cutoff was calculated.

It was predicted that generator injections, due to having to supply losses, would be impacted most by malicious data

TABLE I
DETECTION BY RESIDUAL TYPE

Residual Type	Attacks Detected Best
Weighted Composite	2
Real Power Flows	8
Real Power Injections	60
Reactive Power Flows	17
Reactive Power Injections	53

injection attacks. Reactive flows were also expected to be impacted more than real flows. To compare the relative sensitivities to attack, the residual with the highest proportion of detectable attacks was recorded for each of the 14 attack columns at each magnitude (140 in total). Table I contains counts of the number of attacks for which a residual type is best, based on having the highest detectability for that attack.

The grouped bar graph of Figure 3 shows the proportion of attacks detected by each of the residual types at the 10 MW attack level. The total detectability is also shown, which indicates the proportion of attacks that caused any one of the residuals to exceed its cutoff value. This is higher than any residual type taken singly. As expected both Table I and Figure 3 show that there is significant difference in the residuals based on measurement type and that residuals of real and reactive power injections are better at indicating the presence of attack data.

Figure 4 shows the total detectability sampled at attack levels of 30 MW, 50 MW, 80 MW, and 100 MW. At 30 MW, most of the attacks exceed one of their residue thresholds and are detectable. Attacks based on columns 7, 10, and 14 in the DC H-matrix have the poorest detectability. And while attacks based on column 7 begin to have good detectability from 50MW upward, attacks based on columns 10 and 14 were still undetectable even at attack magnitude of 80 MW. As a reference the total injections in the system are about 275 MW. Attack based on column 14 reamind only partially detectable even at attack magnitude of 100 MW. Buses 10 and 14 were sparsely connected, each only connecting to two other buses, and were not connected to any buses with generation. These attacks required the least amount of measurement modifications, and their distance from generation minimized the impact on generator residuals.

B. Parameter Estimation

1) Setup: As was done previously, analysis was conducted on the IEEE 14-bus test case. State estimation was executed using a modified version of MATPOWER. The modified program estimates parameters simultaneously with states using state augmentation as described in Section III-C1, and was used in this case to estimate line series reactances. Each noise vector n used for generating distributions in Monte Carlo trials had entries consisting of a normally distributed random variable with zero mean and standard deviation of 1% of the measurement value corresponding to the entry in n. Distributions of measurement residuals for determining baseline values and attack detectability were generated by

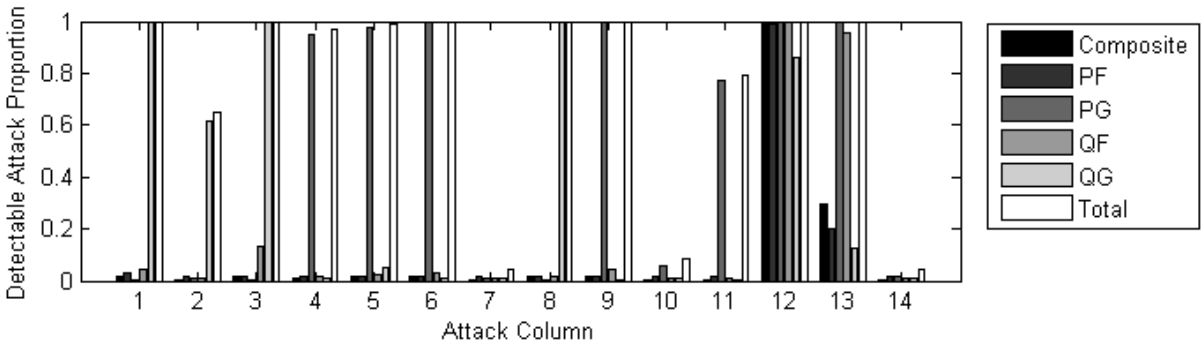

Fig. 3. Grouped bars indicating the proportion of each attack detected at the 10MW attack level. Bars in each group from left to right are residuals of: the weighted composite residual, real power flow, real power generation, reactive power flow, reactive power generation, and total detectability.

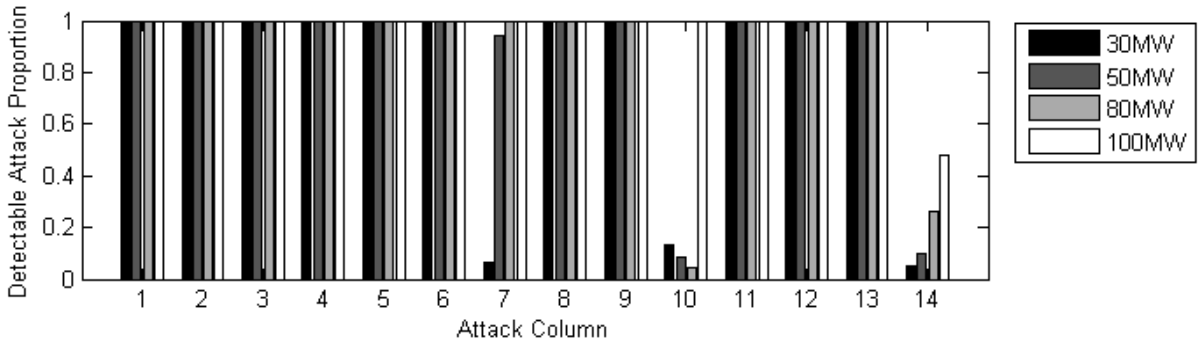

Fig. 4. Grouped bars indicating the total proportion of each attack column type detected at 30MW, 50MW, 80MW, and 100MW, attack levels.

executing state estimation using 500 different noise vectors. The cutoff used to establish the baseline as described in the previous section was the 99th percentile (i.e. 1% false alarms). D-FACTS devices were added to branches 4, 8, and 12, allowing the series reactance to be modified by 5%. As before, attacks consist of the columns of the DC H-matrix used in DC state estimation. Attack magnitude was not varied. Instead, attacks were all normalized at the 100 MW level.

2) Establishing Baseline: To establish expected distributions of parameter estimate residues due to measurement noise, the state estimation was performed with the addition of random noise vectors on measurements. Each noise vector n was composed such that each entry was a normally distributed random variable with zero mean and standard deviation of 1% of the measurement value. Distributions of parameter residuals for determining baseline values and attack detectability were generated by executing state estimation using 500 different noise vectors. The cutoff used to establish the baseline as described in the previous section was 99%.

Histograms of the sum of squared errors of parameter estimates were generated from this data. From these histogram bin counts, a cumulative density function plot was created by normalizing the histogram to have an area of 1 and computing the cumulative sum of bin counts. The 99% cutoff was established by finding the sum of squared residue value of the CDF at the 99th percentile.

3) Determining Detectability: Distributions of residuals were created for parameter residues when the system was unperturbed and under attack, and for attacks against a perturbed system. These distributions were compared to the baseline residuals to determine the effects of attacks on parameter residuals, and to see whether topology perturbation in combination with parameter estimation enhances the detectability of malicious data injection attacks.

Figure 5 shows a comparison of parameter residue distributions. The unattacked system has the lowest residues, and is shown at the far left of the plot. For this attack, the parameter estimates for an unperturbed system would be a good indicator, as we see that most of the area under the CDF for the attacked, unperturbed system (dot-dash) falls above the cutoff established for normal data. The effect of perturbation in this case is to shift the CDF even higher, indicating improved detection of attacks. This effect is also present at lower attack levels, albeit to a lesser extent, as seen for an attack at the 50 MW level in Figure 6.

V. Conclusion

The simplifying assumptions made by an attacker using a linear model to construct attacks against a non-linear state estimator do not affect all measurement types equally. The DC model was shown to introduce very little residue on real power flow measurements in comparison to generators. Generators, which must supply losses, accumulate the effects of attacks. We established in this work that malicious data injection attacks should be expected to impact some classes of measurements greater than others and that such differences can be leveraged to detect maliciously injected data. Bad data detectors already implemented in EMS systems could be augmented to include detectors specifically designed to detect perturbations caused by false data injection.

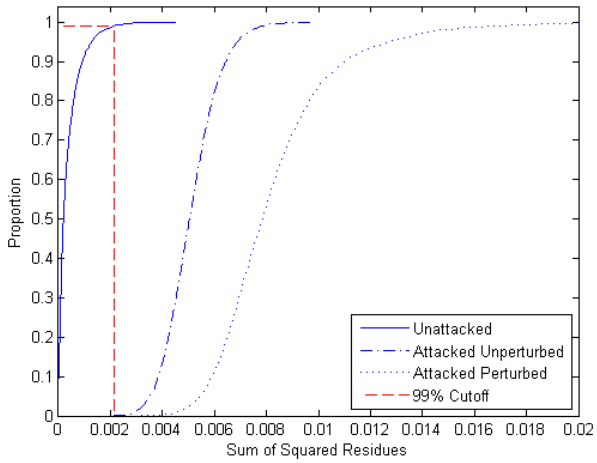

Fig. 5. CDF of sum of squared residues of parameter estimates for unattacked system (solid), attacked system without topology perturbation (dot-dash), and with perturbation (dotted). Horizontal line (dashed) indicates 99th percentile under normal conditions. Vertical line (dashed) indicates value for sum of squared residues corresponding to 99th percentile. Attacks were at 100MW level.

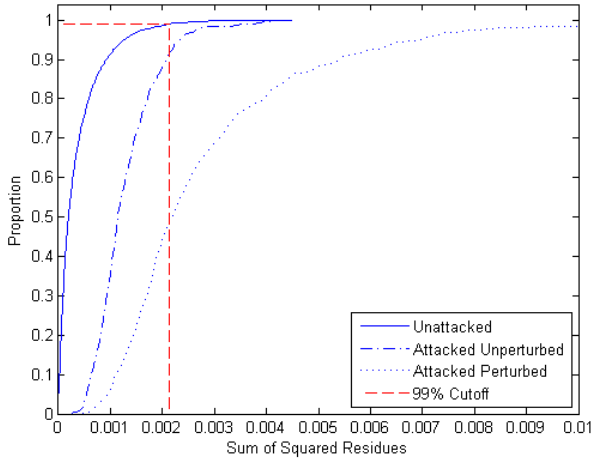

Fig. 6. CDF of sum of squared residues of parameter estimates for unattacked system (solid), attacked system without topology perturbation (dot-dash), and with perturbation (dotted). Horizontal line (dashed) indicates 99th percentile under normal conditions. Vertical line (dashed) indicates value for sum of squared residues corresponding to 99th percentile. Attacks were at 50 MW level.

Parameter residues were also found to increase with data injection attacks. Perturbation of the system in conjunction with parameter estimation was shown to cause further increase. This increase enhances detectability of malicious data injection attacks.

We also found that some attacks are hard to detect even at high energy levels. We intend to study these types of attacks further for other systems and identify ways to detect them. Future work could include comparisons of residual distributions to ordinary bad data to distinguish between bad and malicious data. Other areas for future research are locating which measurements specifically are being attacked, and studying attacks and attack effects to determine which subsets of the attack space would be most attractive to an adversary.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000097 [1]. The authors thank György Dán for his valuable suggestions.

REFERENCES

[1] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.

[2] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. of the 16th ACM Conference on Computer and Communications Security*, 2009.

[3] L. Jia, R. Thomas, and L. Tong, "On the nonlinearity effects on malicious data attack on power system," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1–8.

[4] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "Cyber security study of a scada energy management system: Stealthy deception attacks on the state estimator," in *18th IFAC World Congress, Milan, Italy*, 2011.

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

[6] A. Monticelli and A. Garcia, "Reliable bad data processing for real-time state estimation," *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-102, no. 5, pp. 1126–1139, 1983.

[7] W. Peterson and A. Girgis, "Multiple bad data detection in power system state estimation using linear programming," in *System Theory, 1988., Proceedings of the Twentieth Southeastern Symposium on*, 1988, pp. 405–409.

[8] L. Mili, T. Van Cutsem, and M. Ribbens-Pavella, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation," *Power Engineering Review, IEEE*, vol. PER-4, no. 11, pp. 31–32, 1984.

[9] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *ACM Trans. in Information and Systems Security (TISSEC)*, 2011.

[10] A. Teixeira, S. Amin, H. Sandberg, K. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010, pp. 5991–5998.

[11] R. Bobba, K. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. Overbye, "Detecting false data injection attacks on dc state estimation," in *Proceedings of the First Workshop on Secure Control Systems*, 2010.

[12] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, 2010, pp. 1–6.

[13] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 214–219.

[14] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 226–231.

[15] K. Morrow, E. Heine, K. Rogers, R. Bobba, and T. Overbye, "Topology perturbation for detecting malicious data injection," in *System Science (HICSS), 2012 45th Hawaii International Conference on*, 2012, pp. 2104–2113.

[1]Disclaimer: 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.