STATE ESTIMATION WITH MALICIOUS ATTACK-

This report was prepared in fulfillment of the requirement for the Class Project in

Course: ECE 692-003
Title: Computational Methods for Electric Power Systems
Instructor: Dr. Fangxing (Fran) Li, University of Tennessee, Knoxville

Submitted by -

Phanikrishna (Krish) Gomatom
Mithat Can Kisacikoglu
Zhe Chen

1 INTRODUCTION

Abstract:

In this paper we have attempted to study and analyze the impact of the multi-dimensional "bad data problem" on the confidence level of power system state estimators. The objective of the research was to study how a power system estimator behaves when the input is a large amount of bad data (relative to the good data). Two scenarios are addressed in the research. The first scenario deals with, the impact (on the confidence interval of the State estimator) due to a larger than normal, error percentages in a few measurements. The second scenario deals with the impact (on the confidence interval of the State estimator) due to unrealistically large error percentages in a large number of measurements, something that is widely assumed to occur during a malicious attack. The results of the study lead us to some interesting conclusions. We think that this study can yield more meaningful and realistic results if a wider scope is considered. A preliminary scoping study to explore the possibility of establishing an "evaluation framework" to identify the dependent, semi-dependent and independent variables, pertaining to this complex problem would be worth considering.

The electric power system is continuously monitored in order to maintain the operating conditions in a normal and secure state. Power system operation in a secure state ensures reliable delivery of energy to consumers. Achieving this goal requires, continuous monitoring of the system conditions, identification of the operating state and determination of preventive action, in the event of abnormal conditions. These actions need real-time information about the system being monitored. Direct measurement of all system state variables is not done today, and even when direct measurements are available, they inevitably have errors of varying magnitude. As a result, an essential software tool called the State estimator (SE) is used for system monitoring, processing a redundant set of measurements, and to obtain the best complete estimate of the current system state. These estimated phasors are then used to calculate other quantities needed by the system operator, such as branch currents, bus power injections, and branch power flows.

System operators need to be confident in the results of the SE, before they use the information for making decisions to control or change the state of the power system. Thus one of the critical requirements of the SE is accurate results, which indeed depends on the ability of the SE to detect, isolate and eliminate the bad data. Isolation and elimination of bad data, requires the knowledge of "exactly, how much is the bad data?", and "specifically, which are the bad data?". This critical requirement of the SE has gained vital significance in the post-911 era, due to the looming threats of physical and cyber attacks. The study presented in this paper, seeks to understand the challenges involved in determining the impact on the confidence level, with large errors and large amounts of bad data. The study also presents the results of research for predicting malicious attacks and some guidelines/standards being used in the industry today for securing SCADA and other systems.

2 OBJECTIVES, TASKS, ASSUMPTIONS & COMMENTS

Objectives

Three objectives were identified

1. Set up a typical AC Power System model[1] and formulate ac power flow, followed by the WLSE formulation (weighted least square estimation) to detect the presence of bad data.

2. Evaluate and report the impact of the confidence of State Estimator with respect to the percentage of bad data.

3. Research and report the results, concerning the existence of a guideline to predict security attacks (on the power system) and to what degree these attacks can be predicted.

Tasks

A. POWER SYSTEM MODEL – AC POWER FLOW AND SE FORMULATION

For the purposes of this project, the 3-bus power system model mentioned in example 6.13 (page 195-199), of the text book[2] was chosen. The model was coded in Matlab to run ac powerflow and evaluate power system states. The WLSE formulation was also executed in Matlab. A copy of the Matlab code is included in the Appendix.

In using this example, we encountered a few challenges, for which we made our "best estimate" assumptions:

Network data - In example 6.13, neither the network data is provided, nor there is a mention to use the same network data of example 3.6 (the author makes a mention concerning the use of network data of example 3.6 for the system described in example 6.9 on page 188. A similar statement for example 6.13 could have been helpful).

Assumption 1: The network data of example 3.6 is used for example 6.13.

[1] Use PowerWorld 7-bus system or IEEE 14-bus test system, or any other if needed.
[2] Computational Methods for Electric Power Systems, Mariesa Crow, 2003 CRC Press

True values of power system quantities - To simulate various error levels, as part of the study, we needed true values of the power system quantities (such as V, I, line flows, real power generation, reactive power generation etc). In the example 6.13, the estimated values of power system states that minimize the weighted measurement errors are calculates as δ_2 = -0.0113, δ_3 = -0.0633 and V_3 =0.9858. These are the estimated values, and using these state values, the respective power system quantities are calculated. Using the calculated values and the given measured values of the power system quantities, the weighted sum of squares of the measurement errors, f = 1.2335 is calculated.

Assumption 2: The converged system state values of δ_2 = - 0.0101, δ_3 = -0.0635, V_3 = 0.9816 obtained in example 3.6 on page 58 are considered as "true" state values. Using these "true" state values, power system quantities V_3, P_{13}, Q_{21}, P_3 and Q_2 are calculated. Using these "true" power system quantities and the measured values of these quantities (given data in example 6.13), the value of the weighted sum of squares of the measurement errors, f = 0.3966. Though f is not equal to zero for the purposes of this study, we accept this value as close enough to zero.

Chi-squared test for bad data detection- The WLSE formulation (weighted least square estimation) calculates the weighted sum of squares of the measurement errors (f) .Chi-squared test is used to detect the presence of bad data, based on the value of f, for a specific value of

- k : degrees of freedom, which is the difference in the number of measurements and the number of states, and
- α: Significance level, which is the level of probability that the measurements are erroneous[1]. For example, α = 0.01, indicates a 1% likelihood that bad data exists. If $f > \chi^2$ then the measurement is bad with the confidence level of 99% and if $f < \chi^2$, then bad data will not be suspected with 99% confidence [2].

The converse statement concerning the confidence level of the "goodness" of the data is most likely, either misleading or untrue. Refer Chi-squared values provided in the appendix. The value of χ^2 increases from left to right as the values of α decrease from left to right. This caused some confusion in the interpretation of the χ^2 values. With the timely advice and counsel of our able class professor/instructor Dr. Fran Li, we were able to understand that the χ^2 values apply to values of f, only to determine the extent of bad data present and not to make any affirmative statement about the confidence level in the "goodness" of the data. If the value of f is greater than the appropriate χ^2 value, the likelihood of the presence of bad data is affirmed to a level of likelihood based on the value of α. This explains why the values of χ^2 increase from left to right as α decreases in the same direction.

Comment: Chi-squared test provides a criterion to detect the probability of the presence of bad data, but may not necessarily imply how much of the data is bad. Thus nothing can be concluded concerning the confidence in the "goodness" of the data.

Assumption 3: We assume α = 0.01, for calculations in the current study.

B. CASE SIMULATION - IMPACT OF MEASUREMENT ERROR AND LARGE PERCENTAGE OF BAD DATA DUE TO MALICIOUS ATTACK

Two scenarios were created to study the impact of measurement error and large percentage of bad data due to malicious attack.

Scenario A – A 3-bus power system model with random measurement error associated with more than one measurement. The error magnitude per measurement is chosen arbitrarily, equal to or marginally larger than the worst-case values obtained from the normal error distribution curve. This scenario attempts to simulate the random error effect phenomenon observed in real time systems mainly due to deterioration of equipment condition by ageing or due to failure.

Scenario B – The same 3-bus power system model with an unreasonably high magnitude of error associated with more than one measurement. The error magnitude per measurement is chosen arbitrarily, as per a uniform error distribution. This scenario attempts to simulate the effect of a malicious attack when measurements received are either “garbage” or “no data” (unreasonably large error).

C. RESULTS AND FINDINGS

Analyze the results and derive conclusions. Report the findings in a meaningful format.

D. GUIDELINE TO PREDICT ATTACKS AND TO WHAT DEGREE

Research security guidelines, applicable to power systems, for prediction of malicious attacks and investigate upto what degree these predictions can be made. Report the summary of findings.

3 CASE SIMULATIONS TO STUDY THE IMPACT OF BAD DATA

We designed two methods to study the impact of bad data on the confidence level of the State estimator – method A and method B. The methods, combinedly address three dimensions[3] of the "bad data" problem.

- Number of bad data measurements (how much is the bad data)
- Extent of bad data measurements (how bad is the bad data)
- Influence of bad data measurements (which bad data is the most influential)

For each of the methods two scenarios are implemented, (mentioned in item B in the previous chapter). Scenario 'A' implements the error effect due to known error phenomenon such as equipment ageing, failure etc and scenario 'B' implements the error effect due to unknown, spurious events such as hacking and other malicious attack.

Method A considers the first two dimensions of the problem - number of bad data measurements and extent of bad data measurements. These dimensions are studied using three case simulations as mentioned below for each of the two scenarios.

Method A

In this method five measurements are considered for the 3-bus system. A Matlab formulation was set up to calculate estimated state variables using power flow, followed by the calculation of the weighted sum of squares of the measurement errors (f). Using the Chi-squared test we attempted to determine the extent of bad data present. Note that pure Chi-squared test, only detects the presence of bad data upto the confidence level of α, which for this study is assumed to 0.01.

Scenario A_Case 1: 2/5 (2 out of 5) measurements are in error

The shaded portion in Table 1 shows two out of five measurements that are in error. Four error percentage values have been considered. Note that, using the Chi-Squared test, if $f < \chi^2$, then bad data will not be suspected with 99% confidence and if $f > \chi^2$ then the measurement is bad with the confidence level of 99%. Thus we detect the suspicion of presence of bad data .No conclusion can be drawn concerning the "goodness" of the data[4] and concerning the confidence level of the SE.

[3] We are not claiming that these are the only three dimensions of this problem. During our simulations we found these three dimensions. Several dimensions possibly exist and different combinations of these dimensions may yield different results.

[4] We definitely know that the data is not 100% good, because we added some error as part of the simulation. Hence the "goodness" has to be less than 100% but the question is how much less?

Case 1 : Error in two measurements					
		Measured values with Error			
MEAS.	**True value**	**e = +3%**	**e = +5%**	**e = -3%**	**e = - 5%**
V3	0.9816	1.011048	1.03068	0.952152	0.93252
P13	0.67	0.6901	0.7035	0.6499	0.6365
Q21	-0.0639	-0.0639	-0.0639	-0.0639	-0.0639
P3	-1.2	-1.2	-1.2	-1.2	-1.2
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	9.6418	25.2179	8.122	22.7142
CL	**99%**	**B.D suspected**	**B.D suspected**	**B.D not suspected**	**B.D suspected**
Result : Confidence level cannot be predicted					

Table 1: Simulation results for Method A - Scenario A- Case 1

Scenario A_Case 2: 3/5 measurements are in error

The shaded portion in Table 2 shows three of five measurements that are in error. No considerable changes are observed in the values of residual error f, and the result is same as above - no conclusion can be drawn concerning the “goodness” of data.

Case 2 : Error in three measurements					
		Measured values with Error			
MEAS.	**True value**	**e = +3%**	**e = +5%**	**e = -3%**	**e = - 5%**
V3	0.9816	1.011048	1.03068	0.952152	0.93252
P13	0.67	0.6901	0.7035	0.6499	0.6365
Q21	-0.0639	-0.065817	-0.067095	-0.061983	-0.060705
P3	-1.2	-1.2	-1.2	-1.2	-1.2
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	9.6455	25.2284	8.1257	22.7246
CL	**99%**	**B.D suspected**	**B.D suspected**	**B.D not suspected**	**B.D suspected**
Result : Confidence level cannot be predicted					

Table 2 : Simulation results for Method A - Scenario A- Case 2

Scenario A_Case 3: 4/5 measurements are in error

The shaded portion in Table 3 shows four out of five measurements that are in error. The result is similar to above - no conclusion can be drawn concerning the "goodness" of data.

Case 3 : Error in four measurements					
		Measured values with Error			
MEAS.	**True value**	**e = +3%**	**e = +5%**	**e = -3%**	**e = - 5%**
V3	0.9816	1.011048	1.03068	0.952152	0.93252
P13	0.67	0.6901	0.7035	0.6499	0.6365
Q21	-0.0639	-0.065817	-0.067095	-0.061983	-0.060705
P3	-1.2	-1.236	-1.26	-1.164	-1.14
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	9.9582	26.1172	8.3293	23.3509
CL	**99%**	**B.D suspected**	**B.D suspected**	**B.D not suspected**	**B.D suspected**
Result : Confidence level cannot be predicted					

Table 3: Simulation results for Method A - Scenario A- Case 3

Scenario B_ Case 1: 2/5 (2 out of 5) measurements are in error

The shaded portion in Table 4 shows two out of five measurements that are in error. In this scenario, observe that higher error percentages are used, to mimic the scenario of "no data" or "garbage" received, which would be the case in the event of a malicious attack. Notice that, using the Chi-Squared test, in this scenario, if $f > \chi^2$ then the measurement is certainly bad with a confidence level of 99%. As in scenario A, no conclusive results can be drawn concerning the confidence level of "goodness" of the data.

Case 1 : Error in two measurements					
		Measured values with Error			
MEAS.	**True value**	**e = + 30%**	**e = + 50 %**	**e = - 30 %**	**e = - 50 %**
V3	0.9816	1.27608	1.4724	0.68712	0.4908
P13	0.6700	0.8710	1.0050	0.4690	0.3350
Q21	-0.0639	-0.0639	-0.0639	-0.0639	-0.0639
P3	-1.2000	-1.2000	-1.2000	-1.2000	-1.2000
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	858.0821	2374.3	841.7754	2348.5
CI	**99%**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**
Result : Confidence level cannot be predicted					

Table 4: Simulation results for Method A - Scenario B- Case 1

Scenario B_Case 2: 3/5 (2 out of 5) measurements are in error

The shaded portion in Table 5 shows errors in three measurements with varying error percentages. Same result as above, no conclusive results can be drawn concerning the confidence level of “goodness” of the data.

Case 2 : Error in three measurements					
		Measured values with Error			
MEAS.	**True value**	**e = + 30%**	**e = + 50 %**	**e = - 30 %**	**e = - 50 %**
V3	0.9816	1.27608	1.4724	0.68712	0.4908
P13	0.6700	0.8710	1.0050	0.4690	0.3350
Q21	-0.0639	-0.08307	-0.09585	-0.04473	-0.03195
P3	-1.2000	-1.2000	-1.2000	-1.2000	-1.2000
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	858.4751	2417.6	842.9569	2349.5
CI	**99%**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**
Result : Confidence level cannot be predicted					

Table 5: Simulation results for Method A - Scenario B- Case 2

Scenario B_Case 3: 4/5 (2 out of 5) measurements are in error

The shaded portion in Table 6 shows errors in four measurements with varying error percentages. Same result as above, no conclusive results can be drawn concerning the confidence level of "goodness" of the data.

Case 3 : Error in four measurements					
		Measured values with Error			
MEAS.	**True value**	**e = + 30%**	**e = + 50 %**	**e = - 30 %**	**e = - 50 %**
V3	0.9816	1.27608	1.4724	0.68712	0.4908
P13	0.6700	0.8710	1.0050	0.4690	0.3350
Q21	-0.0639	-0.08307	-0.09585	-0.04473	-0.03195
P3	-1.2000	-1.5600	-1.8000	-0.8400	-0.6000
Q2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
f	0.3966	884.8606	2492	868.2355	2420.3
CI	**99%**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**	**Certainly B.D**
Result : Confidence level cannot be predicted					

Table 6: Simulation results for Method A - Scenario B- Case 3

Method B

Due to the inadequacy of method A to determine, with certainty, the extent of bad data present, we formulated method B. The algorithm of method B is similar to that of method A, but includes an additional "check and eliminate" strategy. If f is > $\chi^{2,}$ (from Chi-squared test), the algorithm, doesn't stop, but it checks for the largest residual error, one at measurement at a time, starting with the measurement that has the largest residual error. The measurement with the largest residual error is eliminated and the algorithm re-runs power flow and estimates the state variables. Using this re-calculated set of state variables, the desired power system quantities are re-calculated. Using the re-calculated quantities and the measured values, the algorithm re-calculates the weighted sum of squares of the measurement errors (f_2). If f_2 is still > $\chi^{2,}$ (from Chi-squared table) the measurement with the second largest residual error, is eliminated and the steps repeated. Here, it is important to note that, the value of χ^2 changes after eliminating a measurement since k (the degree of freedom) decreases. Our algorithm recalculates the new χ^2 value and compares the residual with the updated χ^2. This continues till the value of f is < χ^2. Thus, this method checks the value of f detects the presence of bad data and eliminates the bad measurement to minimize error.

However, there is a drawback to this method, when the convergence of f < χ^2 is reached, at which point, we can say with a confidence of 99% that there is no bad data. But the re-calculated values, of system states, do not match the true values (because some measurements have been eliminated while calculating power flow).

Thus even if $f < \chi^2$ is satisfied we still cannot be absolutely sure to say that the result does not have bad data, because the estimated values are deviant from the original ones. To overcome this drawback, for the purpose of this study, we established a criterion to deterministically say, that the data is bad.

The criterion is as mentioned below,

Absolute value of (Percent normalized change of the state variable)

= {(PF estimate of variable – true value) / true value} * 100

$\leq$ 10%

If at least one variable violates the above criterion, we conclude data is bad.

Under this method, the same two scenarios mentioned in item B (in the previous chapter) are considered. Each scenario consists of four case simulations.

Scenario A_Case1:

This case considers two measurements P13 (real power flow between lines 1 and 3) and Q21 (reactive power flow between lines 2 and 1) with errors. The respective rows of Table 7 appear in purple color. The errors added are worst-case values according to the normal distribution. Note that in this method, additional steps are introduced to check and then conclude that there is really no bad data, even after the Chi-Squared test No conclusive results can be drawn concerning the confidence level of "goodness" of the data. But, with this test we can conclude that the combination of P13 and Q21 is not "influencing" the error very much.

Case 1 : Error in two measurements (one real line powerflow, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 3%	e = + 5%	e = - 3%	e = - 5%
P13	0.6700	0.6901	0.7035	0.6499	0.6365
P12	0.0387	0.0387	0.0387	0.0387	0.0387
P23	0.5386	0.5386	0.5386	0.5386	0.5386
Q21	-0.0639	-0.0658	-0.0671	-0.0620	-0.0607
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.345	0.345	0.345	0.345
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f	9.00E-07	0.1001	0.2804	0.1013	0.2811
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	No B.D	No B.D	No B.D
PF estimate of δ2	-0.0101	-0.0105	-0.0107	-0.0097	-0.0094
PF estimate of δ3	-0.0635	-0.0643	-0.0648	-0.0628	-0.0622
PF estimate of V3	0.9816	0.9816	0.9816	0.9816	0.9816
% change in δ2	0	3.96	5.94	3.96	6.93
% change in δ3	0	1.26	2.05	1.10	2.05
% change in V3	0	0.00	0.00	0.00	0.00
Conclusion		No B.D	No B.D	No B.D	No B.D

Table 7: Simulation results for Method B - Scenario A- Case 1

Scenario A_Case 2

This case considers three measurements P13 (real power flow between lines 1 and 3), P12 (real power flow between lines 1 and 2) and Q21 (reactive power flow between lines 2 and 1) with errors. The respective rows of Table 8 below appear in purple color. The result is same as above. No conclusive results can be drawn concerning the confidence level of "goodness" of the data. But, with this test we can conclude that the combination of P13, P12 and Q21 is not "influencing" the error very much.

Case 2 : Error in three measurements (two real line powerflows, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 3%	e = + 5%	e = - 3%	e = - 5%
P13	0.6700	0.6901	0.7035	0.6499	0.6365
P12	0.0387	0.0399	0.0406	0.0375	0.0368
P23	0.5386	0.5386	0.5386	0.5386	0.5386
Q21	-0.0639	-0.0658	-0.0671	-0.0620	-0.0607
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f	9.00E-07	0.1001	0.2783	0.1005	0.2790
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	No B.D	No B.D	No B.D
PF estimate of δ2	-0.0101	-0.0105	-0.0108	-0.0097	-0.0094
PF estimate of δ3	-0.0635	-0.0643	-0.0648	-0.0627	-0.0622
PF estimate of V3	0.9816	0.9816	0.9816	0.9816	0.9816
% change in δ2	0.00	3.96	6.93	3.96	6.93
% change in δ3	0.00	1.26	2.05	1.26	2.05
% change in V3	0.00	0.00	0.00	0.00	0.00
Conclusion		No B.D	No B.D	No B.D	No B.D

Table 8: Simulation results for Method B - Scenario A- Case 2

Scenario A_Case 3

This case considers three measurements P13 (real power flow between lines 1 and 3), P12 (real power flow between lines 1 and 2) and P23 (real power flow between lines 2 and 3) with errors. The respective rows of Table 9 appear in purple color. We see that with error of + 5% and -5% bad data is suspected because of the large error (>10%) at the system state δ_2, though pure Chi-Squared test says there is no bad data. No conclusive results can be drawn concerning the confidence level of "goodness" of the data. But, with this test we can conclude that the combination of P13, P12 and P23 is "influencing" the error to some extent. In other words, the choice of the measurement that contains error also affects the confidence of the SE.

Case 3 : Error in three measurements (three real line powerflows)					
		Measured values with Error			
MEAS.	True value	e = + 3%	e = + 5%	e = - 3%	e = - 5%
P13	0.6700	0.6901	0.7035	0.6499	0.6365
P12	0.0387	0.0399	0.0406	0.0375	0.0368
P23	0.5386	0.5548	0.5655	0.5224	0.5117
Q21	-0.0639	-0.0639	-0.0639	-0.0639	-0.0639
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f	9.00E-07	0.0686	0.1902	0.0687	0.1905
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	No B.D	No B.D	No B.D
PF estimate of δ2	-0.0101	-0.0093	-0.0089	-0.0108	-0.0113
PF estimate of δ3	-0.0635	-0.0646	-0.0653	-0.0625	-0.0618
PF estimate of V3	0.9816	0.9815	0.9815	0.9817	0.9817
% change in δ2	0.00	7.92	11.88	6.93	11.88
% change in δ3	0.00	1.73	2.83	1.57	2.68
% change in V3	0.00	0.01	0.01	0.01	0.01
Conclusion		No B.D	B.D suspected	No B.D	B.D suspected

Table 9: Simulation results for Method B - Scenario A- Case 3

Scenario A_Case 4

This case considers four measurements P13 (real power flow between lines 1 and 3), P12 (real power flow between lines 1 and 2), P23 (real power flow between lines 2 and 3) and Q21 (reactive power flow between lines 2 and 1) with errors. The respective rows of Table 10 appear in purple color. We see that with error of + 5% bad data is suspected even after Chi-squared test concludes there is no bad data. No conclusive results can be drawn concerning the confidence level of “goodness” of the data.

Case 4 : Error in four measurements (three real line powerflows, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 3%	e = + 5%	e = - 3%	e = - 5%
P13	0.6700	0.6901	0.7035	0.6499	0.6365
P12	0.0387	0.0399	0.0406	0.0375	0.0368
P23	0.5386	0.5548	0.5655	0.5224	0.5117
Q21	-0.0639	-0.0658	-0.0671	-0.0620	-0.0607
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f	9.00E-07	0.0691	0.1917	0.0692	0.1920
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	No B.D	No B.D	No B.D
PF estimate of δ2	-0.0101	-0.0093	-0.0088	-0.0108	-0.0113
PF estimate of δ3	-0.0635	-0.0646	-0.0653	-0.0625	-0.0618
PF estimate of V3	0.9816	0.9815	0.9815	0.9817	0.9817
% change in δ2	0.00	7.92	12.87	6.93	11.88
% change in δ3	0.00	1.73	2.83	1.57	2.68
% change in V3	0.00	0.01	0.01	0.01	0.01
Conclusion		No B.D	B.D suspected	No B.D	B.D suspected

Table 10: Simulation results for Method B - Scenario A- Case 4

Scenario B_Case 1

Refer Table 11. This case considers two measurements P13and Q21 with errors as highlighted in purple color. The errors added are high values according to an assumed uniform distribution. Note that in this method, we conclude that for + 30% and – 30% bad data is suspected, even after the Chi-Squared test concludes that bad data is not present. For + 50% and – 50% error, Chi-Squared test detects the presence of the bad data. The algorithm checks and eliminates bad data, and using the criterion established we can say that there is no bad data. No conclusive results can be drawn concerning the confidence level of “goodness” of the data. But, with this test we can conclude that the combination of P13 and Q21 at 50% error does “influence” the error.

Case 1 : Error in two measurements (one real line powerflow, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 30%	e = + 50%	e = - 30%	e = - 50%
P13	0.6700	0.871	1.005	0.469	0.335
P12	0.0387	0.0387	0.0387	0.0387	0.0387
P23	0.5386	0.5386	0.5386	0.5386	0.5386
Q21	-0.0639	-0.0831	-0.0959	-0.0447	-0.0320
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f1	9.00E-07	10.1058	28.0745	10.1097	28.0785
CL α=0.01,k=5	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	B.D present	No B.D	B.D present
f2			0.1826		0.1820
CL α=0.01,k=4			99%		99%
Comment			No B.D		No B.D
PF estimate of δ2	-0.0101	-0.0139	-0.0099	-0.0062	-0.0103
PF estimate of δ3	-0.0635	-0.0711	-0.0635	-0.0559	-0.0636
PF estimate of V3	0.9816	0.9814	0.9816	0.9818	0.9816
% change in δ2	0	37.62	1.98	38.61	1.98
% change in δ3	0	11.97	0.00	11.97	0.16
% change in V3	0	0.02	0.00	0.02	0.00
Conclusion		B.D suspected	No B.D	B.D suspected	No B.D

Table 11: Simulation results for Method B - Scenario B- Case 1

Scenario B_Case 2

This case considers measurements P13, P12 and Q21 with errors (as highlighted in Table 12). The errors added are high values according to an assumed uniform distribution. The results are similar to case 1 above. No conclusive results can be drawn concerning the confidence level of “goodness” of the data. But, with this test we can conclude that the combination of P13, P12 and Q21 at 50% error does “influence” the error.

Case 2 : Error in three measurements (two real line powerflows, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 30%	e = + 50%	e = - 30%	e = - 50%
P13	0.6700	0.8710	1.0050	0.4690	0.3350
P12	0.0387	0.0503	0.0581	0.0271	0.0194
P23	0.5386	0.5386	0.5386	0.5386	0.5386
Q21	-0.0639	-0.0831	-0.0959	-0.0447	-0.0320
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f1	9.00E-07	10.0295	27.8622	10.0330	27.8658
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	B.D present	No B.D	B.D present
f2			0.3180		0.3154
CL α=0.01,k=4			99%		99%
Comment			No B.D		No B.D
PF estimate of δ2	-0.0101	-0.0143	-0.0106	-0.0059	-0.0096
PF estimate of δ3	-0.0635	-0.0712	-0.0637	-0.0558	-0.0634
PF estimate of V3	0.9816	0.9814	0.9816	0.9818	0.9816
% change in δ2	0.00	41.58	4.95	41.58	4.95
% change in δ3	0.00	12.13	0.31	12.13	0.16
% change in V3	0.00	0.02	0.00	0.02	0.00
Conclusion		B.D suspected	No B.D	B.D suspected	No B.D

Table 12: Simulation results for Method B - Scenario B- Case 2

Scenario B_Case 3

This case considers measurements P13, P12 and P23 with errors. Refer Table 13 for the simulation results .The errors added are high values according to an assumed uniform distribution. Note that bad data is suspected for all the measurements.
No conclusive results can be drawn concerning the confidence level of “goodness” of the data. But, with this test we can conclude that the combination of P13, P12 and P23 at all the four error levels does “influence” the error very much.

Case 3 : Error in three measurements (three real line powerflows)					
		Measured values with Error			
MEAS.	True value	e = + 30%	e = + 50%	e = - 30%	e = - 50%
P13	0.6700	0.8710	1.0050	0.4690	0.3350
P12	0.0387	0.0503	0.0581	0.0271	0.0194
P23	0.5386	0.7002	0.8079	0.3770	0.2693
Q21	-0.0639	-0.0639	-0.0639	-0.0639	-0.0639
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f1	9.00E-07	6.8549	19.0441	6.8595	19.0564
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	B.D present	No B.D	B.D present
f2			2.8208		2.8329
CL α=0.01,k=4			99%		99%
Comment			No B.D		No B.D
PF estimate of δ2	-0.0101	-0.0026	0.0074	-0.0175	-0.0274
PF estimate of δ3	-0.0635	-0.0741	-0.0715	-0.0529	-0.0556
PF estimate of V3	0.9816	0.981	0.9808	0.9824	0.9828
% change in δ2	0.00	74.26	173.27	73.27	171.29
% change in δ3	0.00	16.69	12.60	16.69	12.44
% change in V3	0.00	0.06	0.08	0.08	0.12
Conclusion		B.D suspected	B.D suspected	B.D suspected	B.D suspected

Table 13 : Simulation results for Method B - Scenario B- Case 3

Scenario B_Case 4

This case considers four measurements P13, P12, P23 and Q21 with errors. Refer Table 14 for the results. The errors added are high values according to an assumed uniform distribution. Bad data is suspected for all the measurements.
No conclusive results can be drawn concerning the confidence level of “goodness” of the data. But, with this test we can conclude that the combination of P13, P12, P23 and Q21 at all the four error levels does “influence” the error very much.

Case 4 : Error in four measurements (three real line powerflows, one reactive line power flow)					
		Measured values with Error			
MEAS.	True value	e = + 30%	e = + 50%	e = - 30%	e = - 50%
P13	0.6700	0.8710	1.0050	0.4690	0.3350
P12	0.0387	0.0503	0.0581	0.0271	0.0194
P23	0.5386	0.7002	0.8079	0.3770	0.2693
Q21	-0.0639	-0.0831	-0.0959	-0.0447	-0.0320
Q23	0.1443	0.1443	0.1443	0.1443	0.1443
Q13	0.3450	0.3450	0.3450	0.3450	0.3450
Qg2	-0.0446	-0.0446	-0.0446	-0.0446	-0.0446
Pg1	0.7087	0.7087	0.7087	0.7087	0.7087
f1	9.00E-07	6.9077	19.1923	6.9111	19.1971
CL α=0.01	100%	99%	99%	99%	99%
Comment	No B.D	No B.D	B.D present	No B.D	B.D present
f2			2.9567		2.9553
CL α=0.01,k=4			99%		99%
Comment			No B.D		No B.D
PF estimate of δ2	-0.0101	-0.0026	0.0075	-0.0176	-0.0276
PF estimate of δ3	-0.0635	-0.0741	-0.0714	-0.0529	-0.02556
PF estimate of V3	0.9816	0.981	0.9808	0.9824	0.9828
% change in δ2	0.00	74.26	174.26	74.26	173.27
% change in δ3	0.00	16.69	12.44	16.69	59.75
% change in V3	0.00	0.06	0.08	0.08	0.12
Conclusion		B.D suspected	B.D suspected	B.D suspected	B.D suspected

Table 14: Simulation results for Method B - Scenario B- Case 4

4 SIMULATION RESULTS AND FINDINGS

Following are the results and findings regarding the confidence interval, based on the simulations performed as part of this project –

1. The Chi-Squared test in general, seems to detect the presence of bad data only, not the extent of bad data present.

2. Using the Chi-Squared criterion alone, no deterministic conclusion can be derived concerning the confidence of the state estimator.

3. Determining the confidence of the state estimator (confidence of the “goodness” of the data) appears to be a multi-dimensional problem with several influencing variables. Certain combinations of variables can produce more influencing impact of error on the result than other combinations.

4. At least, three variables are found to influence the level of error. The variables are number of bad data measurements (relative to the good measurements), extent of bad data measurements (magnitude of error) and the specific measurement itself (some measurements with error have more influence than the others that are in error).

5. Higher number of bad measurements and/or higher magnitude of error, tends to increase the weighted sum of squares of the measurement errors (f) significantly but the result from Chi-Squared test cannot determine much except “bad data certainly present”. Thus, in the event of a malicious attack such as a hacker sending lots of bad data or no data, the SE will only say that bad data is certainly present but there is no way to know the extent of bad data and thus no way to suspect that the system is under attack.

6. The method of “check and eliminate” (method B) could be an effective way to determine which measurements are bad. Thus method B can be used to identify those measurements that have higher “influence” on the overall error.

7. The method of “check and eliminate” (method B) cannot determine how much of the data is bad, and thus cannot help to determine the confidence of the “goodness” of the data.

8. The method of “check and eliminate” (method B) can eliminate bad data one after the other, thus theoretically making it possible to get rid of all the bad data, but practically this is not an accurate approach because, eliminating measurements (depending on how many measurements are eliminated) can

impact the accuracy of the estimated state variables, calculated using power flow.

9. The method of “check and eliminate” along with suitable criterion can be used to validate the result of Chi-Squared test at the cost of lost accuracy of estimated state variables.

10. Real power flow measurement errors have a higher impact on the overall error than reactive power flow measurements of equal error magnitude.

5 SECURITY GUIDELINE TO PREDICT MALICIOUS ATTACKS

Information is a critical business resource for nearly every modern industry. Modern computer networking technologies, such as the Internet, offer new tools for making the communication and sharing of information more efficient, and faster than ever before. Utilities are automating the grid with digital switches and high-tech gear. These improvements are making the system more reliable, but are also making it more vulnerable to cyber attacks. Often, utilities upgrade and program these switches and monitoring gear remotely; if they do so through the Internet, the system immediately becomes more vulnerable. There is a generalized perception (although no big related catastrophe has happened yet) that the probability and potential impact of security breaches have grown heavily in recent years due to the increasing interconnectedness among systems and organizations. [3, 4].

The Presidential Decision Directive PDD63 issued in December 2003, specifically identifies the electric and energy industries as part of the nation's critical infrastructure [5]. During our research we found that there is no single comprehensive guideline or standard that predicts with certainty, an impending security attack on the power system. A secure system framework can be implemented by using the knowledge from different standards. It is difficult to predict a cyber attack [6], though cyber attacks can be deterred and prevented by hardened security controls and practices. An analogy in this regard, can be drawn from power system faults, which as we know cannot be predicted with certainty. However, with technology and experience, power system faults can be detected and isolated rapidly, preventing wide-scale outages, financial loss to property and loss of human life. The same is true for malicious attacks (especially cyber attacks) which cannot be predicted using probabilistic or other objective models and methods. The power system community can however be well-prepared by enhancing their understanding of computer and network vulnerabilities, being aware of technical skills of hackers, and benchmarking the historical trends (or trails) left by cyber criminals [7] such as hackers and cyber intruders.

IEC WG15 TC57:

We have identified several standards and guidelines (listed in the later part of this section) relating to overall power system physical security and (SCADA/EMS/Substation) cyber security. Of most relevance to this report, is IEC WG15 TC57 which is working specifically on the security analysis of power system data communication and control. The International Electrotechnical Commission (IEC), Working group 15, Technical Council (TC) 57 "Power Systems Management and Associated Information Exchange" is responsible for developing widely accepted protocols such as IEC 60870-5, DNP 3.0, IEC 60870-6 (also known as TASE.2 or ICCP) and IEC 61850[8].

The report IEC TR 62210 issued in May 2003, summarizes the IEC 62210 standard "Data and Communication Security". The report recognizes the lack of applicable models of threats and attacks that can readily support security analyses. The report proposes a methodological approach called the CESI-JRC approach that aims to set a conceptual framework for security assessment. The approach is useful to the research presented in this paper because the approach presents a conceptual yet useful, four-dimensional method that might help to predict future attacks with some degree of certainty [9].

- Identification of threats
- Characterizes of vulnerability types such as typical software problems that produce the vulnerability
- Evaluation of malicious attack modes
- Quick estimation of the severity of the vulnerability.

The following is a list of security guidelines and standards that can be applied to secure the power system infrastructure as a whole (mainly SCADA) -:

- NERC CIP Standards[5], CIP-002-001 through CIP-009-001.
- System Protection Profile for Industrial Control Systems (SPP ICS), NIST.
- Guide for developing Security Plans for Information Technology Systems, NIST 800-18.
- Guidelines and Firewalls and Firewall Policy, NIST Special Publication 800-41.
- AGA (American Gas Association)-12, Cryptographic Protection of SCADA Communications general recommendations.
- API (American Petroleum Institute): Security Vulnerability Assessment Methodology for the Petroleum and Petrochemical Industries.
- IEEE 1402: IEEE Guide for Electric Power Substation Physical and Electronic Security.
- IEEE 1686 : Substation Intelligent Electronic Device (IED) Cyber security Standards
- IEEE P1689: Cyber Security for Serial Data Links and IED remote access.
- IEEE P1615 : Recommended Practice for Network Communication in Electric Power Substations
- Good Practice Guide on Firewall Deployment for SCADA and Process Control Networks, NISCC, UK.
- Code of Practice for Information Security Management, ISO/IEC 17799:2005
- Specification for Information Security Management Systems (ISMS), BS 7799-2:2005.
- Security Technologies for Manufacturing and Control Systems, ISA (Instrumentation Systems, and Automation Society) -TR99.00.01-2004.

[5] The North American Electric Reliability Council (NERC) has created many standards, the recent one being the CIP (critical infrastructure protection) standards CIP-002 through CIP-009, passed in May 2006, which became effective June 1, 2006 and with initial compliance auditing started in late 2007. These cyber security standards specify requirements for securing the critical cyber assets of the North American bulk electric system and are intended to ensure that the bulk electric system cannot be severely disabled by a cyber attack [10].

6 LITERATURE SURVEY

As part of this study we did a literature survey to see what approaches are being taken in general, to apply and enhance the performance of State Estimators. We have summarized the five main approaches.

1. 3D Visualization of Power System State Estimation
This paper presents the results of a project which is undertaken in order to develop a flexible visualization tool to be used mainly by power system operators. The project concentrated on two application functions, namely the power flow and the state estimation. One of the essential goals of the envisioned tool is to aid system operators by providing images of the system with easily identifiable characteristics related to violations of various operating limits. In addition, the tool is expected to illustrate deficiencies associated with the existing metering system both from the point of view of metering design as well as accuracy. The result is a user-friendly 3D graphic interface for the stated two power system applications. In power flow analysis mode, animation and bus contouring technique is used for visualizing power flows and bus voltages. In state estimation mode, color contour is used to visualize different observable islands and bad measurements which are identified through the largest normalized residuals test. Effectiveness of the tool is illustrated via different scenarios which are created using the IEEE 118 bus system as an example [11].

2. Modeling Error Effects of State Estimation and Impact on System Operation.

State estimation has been introduced to power systems and implemented in the 60s, using a single frequency, balanced and symmetric power system model under steady state conditions. The single frequency, balanced and symmetric system assumptions have simplified the implementation but have generated practical problems. This paper examines these simplified assumptions and their impact on the state estimation performance. It provides a theoretical basis for the well known fact that the reliability of the state estimator algorithms has been below expectations. Specifically, sensitivity analysis methods are used to quantify the impact of modeling simplifications and measurement schemes on the performance of state estimation. The results clearly illustrate that the traditional state estimation algorithm is biased. These biases affect the accuracy of state estimation and its convergence characteristics. The paper also reviews the traditional state estimation approach against recent technological advances that have enabled synchronized measurements. The implications and possibilities of this new technology are discussed in this paper. Specifically, an example application of the new technology for a three phase State estimator is described. In addition, demonstration of performance with the proposed methods on an actual system (New York Power Authority system) using actual synchronized measurements is shown[12].

3. *State Estimation for the NEPTUNE Power System*.

This paper describes a system of cabled dc network with multiple distributed loads called as the NEPTUNE power delivery system. The study shows how operation under normal operating conditions is challenging for the NEPTUNE power delivery system,. The problem arises because the power management system must control a system that is severely limited in the number and location of measurements. A modified state estimation approach taken to address this challenge is described [13].

4. *Differential Evolution Solution Approach for Power System State Estimation*.

Differential evolution (DE) is a very simple population based evolutionary computation technique used for solving complex optimization problems. This technique is well-suited for many problems in the power systems area, including state estimation. This paper presents an overview of the state estimation problem and the evolutionary computation field. It also presents a comprehensive description of the differential evolution technique. A methodology for solving the power system state estimation problem, based on the differential evolution technique, is presented. The applicability of the proposed method is validated through a case study [14].

5. *State Estimation and Voltage Security Monitoring Using Synchronized Phasor Measurements*.

The phasor measurement unit (PMU) is considered to be one of the most important measuring devices in future power systems. The distinction comes from its unique ability to provide synchronized phasor measurements of voltages and currents from widely dispersed locations in an electric power grid. The commercialization of the global positioning satellite (GPS) with accuracy of timing pulses in the order of 1 microsecond made possible the commercial production of phasor measurement units. Simulations and field experiences suggest that PMUs can revolutionize the way power systems are monitored and controlled. However, it is perceived that costs and communication links will affect the number of PMUs to be installed in any power system. Furthermore, defining the appropriate PMU system application is a utility problem that must be resolved. This thesis will address two key issues in any PMU initiative: placement and system applications. A novel method of PMU placement based on incomplete observability using graph theoretic approach is proposed. The objective is to reduce the required number of PMUs by intentionally creating widely dispersed pockets of unobserved buses in the network. Observable buses enveloped such pockets of unobserved regions thus enabling the interpolation of the unknown voltages. The concept of depth of unobservability is introduced. It is a general measure of the physical distance of unobserved buses from those known. The effects of depth of unobservability on the number of PMU placements and the errors in the estimation of unobserved buses will be shown. The extent and location of communication facilities affects the required number and optimal placement of PMUs [15].

7 OVERALL CONCLUSIONS

Based on the simulation results and the desktop research conducted for this project, following is the summary of conclusions –

Confidence interval:

1. The results of the simulations using the 3-bus power system model of example 6.13 and the Chi-Squared test could not determine the impact of the percentage of bad data on the confidence of the state estimator. We started working on IEEE14 bus test system, due to time constraints we couldn't finish the program coding.

2. The simulation results determined that errors in certain measurements have a larger impact on the error than other measurements.

3. The simulation results determined that Chi-Squared test cannot be trusted completely for making decisions regarding "there is no bad data" and or regarding the confidence level of good data. Additional methods need to be used in conjunction with Chi-Squared test.

4. In this study, we could not determine the confidence interval of the SE, but we think that, if a larger system is considered, along with a systematic analysis strategy to consider different variables and different combination of variables, a meaningful confidence interval might be derived.

5. Based on lessons learnt in this study, we propose the idea of a preliminary scoping study to explore the possibility of establishing an "evaluation framework" to identify the dependent, semi-dependent and independent variables of this multi-dimensional problem. A later effort could include assessment of their level of influence on the confidence level of current state estimation algorithms.

Security guideline:

1. At present there is no single, unified guideline that can be used to predict malicious attacks on the power system.

2. IEC 62210 provides an approach to establish a framework for security analysis of attacks on informational infrastructure related to power systems. This approach can be used to study the "pattern" of attacks, characterize the attacks, and estimate the severity, on a "post-attack" basis. However, we know that the "hacking ability" of criminals is very dynamic in nature – the skill-set of hackers is getting better and better every day, and due to introduction of larger number of

newer computer/networked systems, the vulnerabilities (or potential “holes” to gain access to the systems) are increasing day by day. Thus, wisdom lies in educating, training, learning about new systems with vulnerabilities before installation on secure networks, and improving cyber security as much as possible. Trying to predict the “uncertain” is always a gamble.

3. There are useful standards such as the NERC CIP and ISO ISO/IEC 17799:2005 that can serve as a baseline guide, for securing power system infrastructure both physical and cyber.

Literature survey:

A cursory literature survey shows that not many researchers are working in the area of state estimation. Current research work in this area is focused on application of state estimation to non-traditional power systems such as DC networks. There seems to be some interest in new measurement techniques to improve the effectiveness of State estimation.

8 REFERENCES

1. Mariesa Crow, *Computational methods for electric power systems,* CRC Press, 2003. pp.174-175.

2. Ali Abur and Antonio Exposito, *Power System State Estimation, theory and implementation*, MD press, 2004,pp107.

3. Neil C. Rowe, *Counterplanning deceptions to foil Cyber-Attack Plans*, IEEE Information Assurance workshop (IAW), West Point, NY, June 2003.

4. David. W.Munns, *Cyber dilemma, The sea power*, May 2007, BNET research center.

5. HSPD-7 (Homeland Security Presidential Directive No. 7). December 2003. Directive that was issued by U.S. President George W. Bush in December, 2003 to update policies intended to protect the country from terrorist attacks. This directive superseded the earlier PDD-63 (Presidential Decision Directive No. 63), which was issued by President Clinton in May of 1998.

6. Testimony of FBI Deputy Assistant Director Keith Lourdeau, Cyber Division Senate Judiciary Subcommittee on Terrorism, Technology, and Homeland Security Hearing on Cyber Terrorism,February 24, 2004.

7. Clay Wilson, *Emerging Terrorist Capabilities for Cyber Conflict against the U.S. Homeland*, Technology and National Security Congressional Research Service, November 2005.

8. Frances Cleveland*, IEC TC57 Security Standards for the Power System's Information Infrastructure – Beyond Simple Encryption,* IEC TC57 WG15 Security Standards v5, October 2005

9. G. Dondossola, O. Lamquet, M. Masera, *Emerging standards and methodological issues for the security analysis of Power system information infrastructures, Grenoble, October 2004.*

10. NERC, CIP Standards, CIP-001-1 through CIP-009-1. http://www.nerc.com/~filez/standards/Reliability_Standards.html#Critical_Infrastructure_Protection

11. Bei Xu, Cem Yusel, Ali Abur, Ergun Akleman*, 3D Visualization of Power System State Estimation*. IEEE MELECON 2006*,* May 16-19, Benalmádena (Málaga), Spain*.*

12. A. P. Sakis Meliopoulos, Bruce Fardanesh, Shalom Zelingher*, Power System State Estimation: Modeling Error Effects and Impact on System Operation,* PSERC 2001*.*

13. Chen-Ching Liu, Kevin Schneider, Harold Kirkham, Bruce Howe, *State Estimation for the NEPTUNE Power System*.

14. N.G. Figueroa and J.R. Cedeño, *A Differential Evolution Solution Approach for Power System State Estimation,* Proc. IEEE PES 2004.

15. Reynaldo Francisco Nuqui*, State Estimation and Voltage Security Monitoring Using Synchronized Phasor Measurements, PhD d*issertation, Virginia Polytechnic Institute, July 2001.

9 APPENDICES

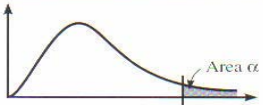
Appendix A

Chi-Squared distribution [Donald. B Owen, Handbook of Statistical tables, 1962]

Table G The Chi-Square Distribution

Degrees of freedom	α									
	0.995	*0.99*	*0.975*	*0.95*	*0.90*	*0.10*	*0.05*	*0.025*	*0.01*	*0.005*
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.262	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Source: Donald B. Owen. *Handbook of Statistics Tables*, © 1962, by Addison-Wesley Publishing Co., Inc.. Reading, Massachusetts. Table A–5. Reprinted with permission of Addison-Wesley Longman Publishing Company, Inc.



Appendix B

Matlab code used for the simulations

```
%%%IN ORDER TO RUN THE PROGRAM, SYMBOLIC MATH TOOLBOX SHOULD BE
INSTALLED IN
%%%MATLAB. THIS PROGRAM IS DEVELOPED IN MATLAB 7.3.0 (R2006b).


%Bad data detection of the 3 bus system of Example 6.13 in the text
book with more number of measurements


%Calculating the bus admittance matrix:
clear;clc;
y=zeros(3,3);
y(1,1)=(0.15i+0.1i)/2;
y(1,2)=1/(0.02+0.3i);
y(1,3)=1/(0.01+0.1i);
y(2,1)=y(1,2);
y(2,2)=(0.15i+0.1i)/2;
y(2,3)=1/(0.01+0.1i);
y(3,1)=y(1,3);
y(3,2)=y(2,3);
y(3,3)=(0.1i+0.1i)/2;
Y=zeros(3,3);
for i=1:1:3;
    for j=1:1:3;
        if i==j
            Y(i,j)= sum(y(i,:));
        else
            Y(i,j)= -y(i,j);
        end
    end
end
%States and given values are given. States will be calculated based on
the
%measurements. All the values given here are taken from the data in
example
%3.6.
delta(2)=sym('d2');% System state 1
delta(3)=sym('d3');% System state 2
V(3)=sym('V3');% System state 3
V(1)=1.02;%Slack bus voltage
V(2)=1.00;%PV bus voltage
delta(1)=0;%Slack bus angle
Pg(1)=sym('Pg(1)');%Genertion 1 will be calculated based on the system
states
Qg(1)=sym('Qg(1)');
Pg(2)=0.5;%Real power generation 2 is given in the example 3.6
Qg(2)=sym('Qg(2)');
Pg(3)=0;%Load bus
Qg(3)=0;
```

```
%Loads are given
Pl(1)=0;
Pl(2)=0;
Pl(3)=1.2;
Ql(1)=0;
Ql(2)=0;
Ql(3)=0.5;

% Active power generation
for i=1;
    sumc1=0;
    for j=1:3;
        Sc1=V(j)*abs(Y(i,j))*cos(delta(i)-delta(j)-angle(Y(i,j)));
        sumc1=sumc1+Sc1;
    end
    Pg(i)=Pl(i)+V(i)*sumc1;
end

% Reactive power generation
for i=1:2;
    sums1=0;
    for j=1:3;
        Ss1=V(j)*abs(Y(i,j))*sin(delta(i)-delta(j)-angle(Y(i,j)));
        sums1=sums1+Ss1;
    end
    Qg(i)=Ql(i)+V(i)*sums1;
end

%Line power flow
for i=1:3;
    for j=1:3;
        P(i,j)=V(i)*V(j)*abs(Y(i,j))*cos(delta(i)-delta(j)-
angle(Y(i,j)))-V(i)^2*abs(Y(i,j))*cos(angle(Y(i,j)));
    end
end

for i=1:3;
    for j=1:3;
        Q(i,j)=V(i)*V(j)*abs(Y(i,j))*sin(delta(i)-delta(j)-
angle(Y(i,j)))+V(i)^2*abs(Y(i,j))*sin(angle(Y(i,j)));
    end
end

%The measurement function and its derivative w.r.t. system states are
%given.
d2=delta(2);
d3=delta(3);
V3=V(3);
h=[P(1,3);P(1,2);P(2,3);Q(2,1);Q(2,3);Q(1,3);Qg(2);Pg(1)];
x=[d2;d3;V3];
H=jacobian(h,x);
%Measurement matrix with bad data is given.
z=[0.3350; 0.0194; 0.2693; -0.032; 0.1443; 0.3450; -0.0446; 0.7087];
% Covariance Matrix is given
R=zeros(8,8);
```

```
R(1,1)=1/(0.05)^2;
R(2,2)=1/(0.05)^2;
R(3,3)=1/(0.05)^2;
R(4,4)=1/(0.075)^2;
R(5,5)=1/(0.075)^2;
R(6,6)=1/(0.075)^2;
R(7,7)=1/(0.075)^2;
R(8,8)=1/(0.05)^2;

%-----------Nonlinear Weighted Least Squares State Estimation----------
---
%Newton Raphson Method to calculate the states
J=transpose(H)*inv(R)*H;
F=transpose(H)*inv(R)*(z-h);
%Flat start
d2=0;
d3=0;
V3=1.00;
x=[0;0;1];
sum=100;%Initial value to fall into the while loop
chi=0;%Initial value to fall into the while loop
alpha=input('Enter the significane level alpha=  ');

while sum>chi
    [n,m]= size(z);%Number of measurements is stored in n.
    [f g]=size(x);%Number of states is stored in f
    k=n-f;%Degree of freedom.
    chi=chi2inv(1-alpha,k);% Chi square value is updated after the
worst
    %measurement is taken out
    e=1;
    ii=1;
    while e>0.00001
        d2=x(1);
        d3=x(2);
        V3=x(3);
        F=transpose(subs(H))*inv(R)*(z-subs(h));
        J=transpose(subs(H))*inv(R)*subs(H);
        dx=inv(J)*F;
        x=x+dx;
        e= norm(F, inf);
        ii=ii+1;% Counter counts the number of iteration.
    end

    h1=subs(h);% Measurement function is updated and stored since it
will
    %be used to calculate the residuals
    sum=0;
    [n,m]= size(z);%Number of meausurements
    Res=zeros(1,n);%Individual residuals for the every measurement will
be
    %stored.
    for i=1:n;
        S=R(i,i)*(z(i)-h1(i))^2;
        Res(i)=S;
```

```
        [t r]=sort(Res);%Sort the residuals from least to greatest.
        sum=sum+S;
    end
    display ('Residual'), sum

    if sum>18.48
        display('Measurements have bad data with a confidence level of
99%')
        display('Worst measurement will be subtracted')
        % The worst meauserment and associated elements of the other
matrices are
        % taken out
        z(r(end))=[];
        h(r(end))=[];
        H(r(end),:)=[];
        R(r(end),:)=[];
        R(:,r(end))=[];
    else
        display('No bad data with a confidence level of 99%')
    end
end

states=[d2 d3 V3]
```