Cyber security and risk analysis in power systems

Conference Paper · October 2018

CITATION
1

READS
303

2 authors, including:

Lida Hagh

12 PUBLICATIONS **10** CITATIONS

SEE PROFILE

Proceedings of the 2018 Industrial and Systems Engineering Research Conference
K. Barker, D. Berry, C. Rainwater, eds.

Cyber Security and Risk Analysis in Power Systems

Lida Haghnegahdar, Yong Wang
Department of Systems Science and Industrial Engineering
State University of New York at Binghamton

Abstract

Growing demand for reliable energy has advanced smart power grid development which attempts to merge conventional power grid infrastructures with advanced and modern technologies. To achieve this goal, energy providers need to invest in gaining detailed knowledge of cybersecurity to prevent different types of attacks such as false data injection, intrusion, and load redistribution. Cyber threat vectors and existing vulnerabilities in smart grid technologies need to be well understood and protected against. Successful cyberattacks would most likely result in data loss and equipment damage, and in addition, possible life and other intangible economic losses. A reliable smart grid needs strong and resistant cybersecurity protection, which is capable of detecting and diagnosing these attacks. A security risk analysis model can evaluate the potential risks in power system. This risk analysis will help in reducing costs without increasing the need for power system upgrades. This research introduces the cybersecurity model to assess the risk. This paper uses the PCA procedure in SAS 9.4 to conduct MANOVA analysis on a power system data set including the recordings of 4966 observations and 114 features. The main goal of this analysis is to determine whether a rule can be constructed to discriminate between the two classes of natural and attack events.

Keywords: Cyber Security, Cyberattack, Risk Analysis, Smart Grid, Power System

1. Introduction and Related Work

This paper studies cyber security issues in the power system and smart grid. The power grid is one of the most important networks which could be under different types of cyberattacks. Power system disturbances are complex and can be classified into a wide range of sources that are either natural or man-made. Cyberattacks can have the same effects as natural events and so differentiating between malicious and non-malicious events in a large and interconnected system can be overwhelming if not infeasible for a human. Recently, several studies are done in this area. For example, a classification algorithm [1] can help power system operators make reliable decisions. These studies have shown that an attacker can inject pre-determined false data into smart meters such that it can pass the test of state estimator [2]. Hassan et al. studied the scheme, which is capable of improving the quality of cyber protection with restricted resources [3].

Risk analysis can include vulnerability analysis as well as system and application impacts analysis. The initial step of the risk analysis process is the infrastructure vulnerability analysis [4]. The cyber security disturbance detection is challenging due to the big and complex network data. An effective cyber security system should be able to process large quantity of data in order to detect all malicious activities. A framework has been proposed for building a fast and efficient cyber security intrusion detection system [5]. The performance of the proposed framework is evaluated using Big Data processing tools like Apache Spark and machine learning algorithms. Teixeira et al. used the original dataset for projecting high dimensional dataset to the lower dimensions and classification which resulted in to reduce variance retained [6].

Potential vulnerabilities in a power grid can be diagnosed based on transmission lines on which attacks causes maximum disruption in the power system [7]. Ref. [8] studied the concept of the unrecognized attack, which is a type of attack that cannot be identified by the control center. Kosut et al. studied the detection of malicious attacks against smart grid by using state estimation [4].

The performance of other robust state estimation models has also been studied and a method was used as a recovery model in electric power system [9]. Phasor measurement units (PMUs) are important components in the power grid and they can be applied for monitoring and controlling the power system. According to [10] the attack model and a defense scheme based on the optimal PMU placement strategy is considered for detecting data integrity attacks with low cost. Taha proposed a risk mitigation method on the basis of dynamic state estimation to reduce the false data injection and potential cyberattacks. PMUs can also be used to measures the data that come from fake or false signals

[11]. Using data analytics tools, the fault signaling is classified as real or false (fake) data injection with malicious intentions [12,13,14].

Although such methods may detect a set of very basic cyberattacks, they may fail in the presence of a more intelligent attacker. In [15], Ericsson specified the attacker with a list of objectives and the proposed missions of attacks was analyzed using linear and nonlinear estimators. The cardinality minimization problem for cyberattack detection has also been touched upon in [15, 16].

Unlike the previous research in the literature, this paper studies the problem in a different perspective. We propose a security risk analysis model that can evaluate the potential risks in a power system. This risk analysis will help in reducing costs without increasing the need for power system upgrades. This research introduces the cybersecurity model to assess the risk. More specifically, we use the PCA procedure in SAS 9.4 to conduct MANOVA (multivariate ANOVA) analysis on a power system data set including the recordings of 4966 observations and 114 features. The goal of this analysis is to determine whether a rule can be constructed to discriminate between the two class of natural and attack events.

2. Power System Framework

In this case study 4 breakers were considered by intelligent relays. Pre-built attack scenarios were used with binary data. A total of 37 event scenarios were considered, 28 for attack and 9 for normal operations. Similar to [3], scenarios considered are short circuit, line maintenance, remote command input as an attack, relay setting change as an attack, and false data injection as an attack.

3. Feature Analysis Discussion

The following variables extracted from PMUs are used in this analysis:

- PA1VH-PA3VH, which is phase A-C Voltage Angle
- PAM1V-PM3V, which is phase A-C Voltage Magnitude
- PA4IH-PA6IH, which is phase A-C Current Phase Angle
- PM4I-PM6I, which is considered as Phase A-C Current Magnitude
- PA7VH-PA9VH, which is related to Pos. -Neg. - Zero Voltage Phase Angle
- PM7V-PM9V, which is related to Pos. -Neg. - Zero Voltage Magnitude
- PA10VH-PA12VH, which is related to Pos. -Neg. - Zero Current Phase Angle
- PM10V- PM12V, which is related to Pos. -Neg. - Zero Current Magnitude
- F, which is Frequency for relay
- DF, which is Frequency Delta for relays
- PA: Z, which is Impedance seen by relay
- PA: ZH, which is Impedance Angel for relays and S is Status of relays

The data sets used in this study is provided by Oak Ridge National Lab. There are totally 4966 observations in the data with 114 columns. Main Cyber security issues in power system includes control the low-dimensional structures in high-dimensional data to find the challenges in large data set, storage, and information extraction. And analysis of PMU measurements, missing data recovery and the detection of cyber data attacks. In this work, according to data set, 4 synchro phasors that measured 29 features each for a total of 116 (4*29) PMU measures are considered. We study the data attacks by using the principal component analysis method.

4. Statistical Analysis

4.1. PCA Process

PCA analyzes the variance and reduces the number of variables and tries to decompose all variance into orthogonal components. PCA reproduces the R matrix completely and gives a unique solution and extract as much variance with the least number of factors. For standardized data, it retains only those components whose eigenvalues are greater than one (SAS Default). 17 principles are shown the higher Eigen values and then 13 features are selected for the next step, which is MANOVA analysis. Figure 1 shows the PCA outputs. Figure 2 shows the higher values of the Correlation Matrix for Principal Selection.

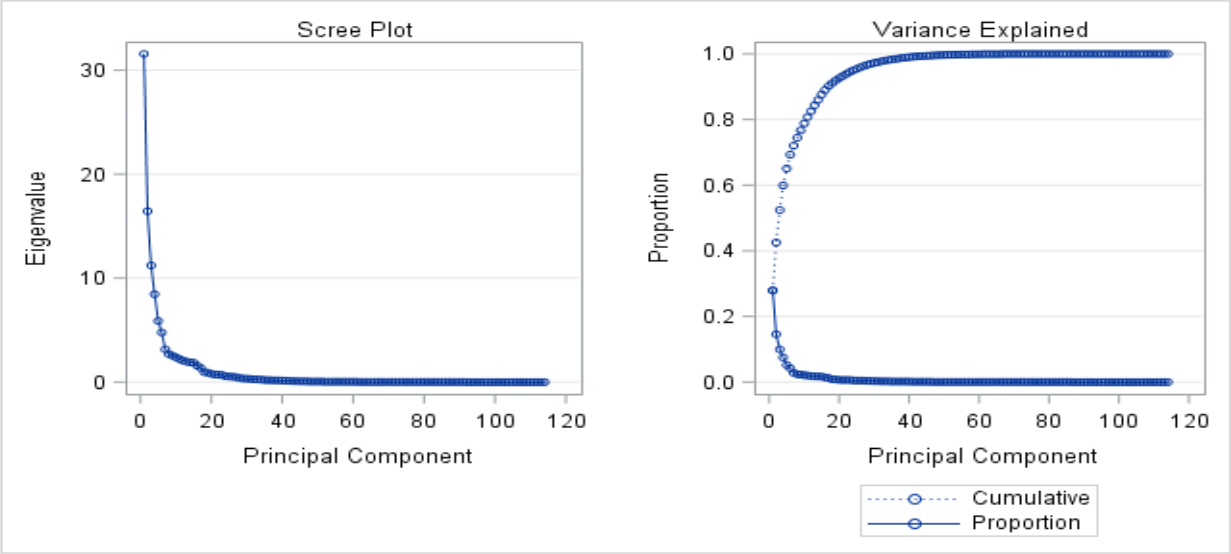

Figure1: PCA outputs

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	31.5720853	15.1334762	0.2794	0.2794
2	16.4386092	5.2187097	0.1455	0.4249
3	11.2198995	2.7818385	0.0993	0.5242
4	8.4380610	2.5723085	0.0747	0.5988
5	5.8657524	1.1048869	0.0519	0.6507
6	4.7608656	1.6271245	0.0421	0.6929
7	3.1337411	0.4494301	0.0277	0.7206
8	2.6843109	0.1481690	0.0238	0.7444
9	2.5361419	0.1535798	0.0224	0.7668
10	2.3825622	0.1649398	0.0211	0.7879
11	2.2176224	0.1550888	0.0196	0.8075
12	2.0625336	0.0870208	0.0183	0.8258
13	1.9755129	0.0876658	0.0175	0.8433
14	1.8878470	0.0252386	0.0167	0.8600
15	1.8626084	0.2802099	0.0165	0.8764
16	1.5823985	0.2175989	0.0140	0.8904
17	1.3647995	0.4030182	0.0121	0.9025
18	0.9617813	0.0559468	0.0085	0.9110

Figure 2: Selected higher values of the Correlation Matrix

4.2. MANOVA Analysis

MANOVA showed the effects of class on the response or dependent variables. We considered two classes, 0 as the natural and 1 as the attack class. The procedure for MANOVA is similar to ANOVA. If there is a significant multivariate effect, then it examines the univariate effects. If there is a significant univariate effect, then it conducts post hoc tests. The main objective in using MANOVA is to determine if the response variables are altered by the observer's manipulation of the independent variables.

In MANOVA, a statistical test will identify the effects of the treatment if it really exists. ANOVA can be used to determine whether there are any statistically significant differences between the means of three or more independent groups. Here, MANOVA results show the effects of independent variable and the dependent variables.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
R1PA1VH	1	0.23771126	0.23771126	1.81	0.1792
R1PM2V	1	0.91659872	0.91659872	6.96	0.0084
R1PM3V	1	0.74093628	0.74093628	5.63	0.0177
R1PA4IH	1	0.23446968	0.23446968	1.78	0.1821
R1PM5I	1	0.55439417	0.55439417	4.21	0.0402
R1PA7VH	1	0.13109207	0.13109207	1.00	0.3185
R2PA1VH	1	29.95249071	29.95249071	227.45	<.0001
R2PA5IH	1	0.00635323	0.00635323	0.05	0.8262
R2PM5I	1	0.38191167	0.38191167	2.90	0.0886
R3PA1VH	1	29.95975800	29.95975800	227.51	<.0001
R4PA1VH	1	0.00114797	0.00114797	0.01	0.9256
R4PA2VH	1	4.22481878	4.22481878	32.08	<.0001
R4PA7VH	1	0.00048348	0.00048348	0.00	0.9517

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	43.1499721	3.3192286	25.21	<.0001
Error	4952	652.1138217	0.1316870		
Corrected Total	4965	695.2637938			

R-Square	Coeff Var	Root MSE	Class Mean
0.062063	43.63430	0.362887	0.831655

Figure 3: The GLM procedure on the variables

The ANOVA analysis for each variable alone presents the effect of the natural and attack event and which one is more significant. The MANOVA results show the correlation between the dependent variables. Since the outputs are large, we selected some of them as a sample to show here. Figure 3 showed the GLM Procedure on the variables. In Figure 4, we see the MANOVA test statistics include: Wilks' Lambda, Hotelling's criterion, Pillai's criterion (Pillai's trace), and Roy's GCR (Greatest Characteristic Root) that measures the differences on only the first discriminant function among the dependent variables, and which is useful when the dependent variables are strongly interrelated on a single dimension. Figure 4 illustrates the hypothesis of no overall effect to each variable.

Figure 5 shows the MANOVA analysis result of the class effect situation on the different observations and events. Figure 6 presents the MANOVA Statistical Hypothesis Test for no overall Class Effect Analysis which is produced by the procedure to represent effects of class and variables. The tables could show the relation between the dependence of the attack and the risk factors. For instance, how many attacks would be occurred if risk factor X is really risky or not really risky.

Characteristic Roots and Vectors of: E Inverse * H, where H = Type III SSCP Matrix for R1PM2V E = Error SSCP Matrix		
		Characteristic Vector V'EV=1
Characteristic Root	Percent	Class
0.00140558	100.00	0.03915960

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall R1PM2V Effect H = Type III SSCP Matrix for R1PM2V E = Error SSCP Matrix S=1 M=-0.5 N=2475					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.99859639	6.96	1	4952	0.0084
Pillai's Trace	0.00140361	6.96	1	4952	0.0084
Hotelling-Lawley Trace	0.00140558	6.96	1	4952	0.0084
Roy's Greatest Root	0.00140558	6.96	1	4952	0.0084

Figure 4: MANOVA statistical hypothesis test (R1PM2V)

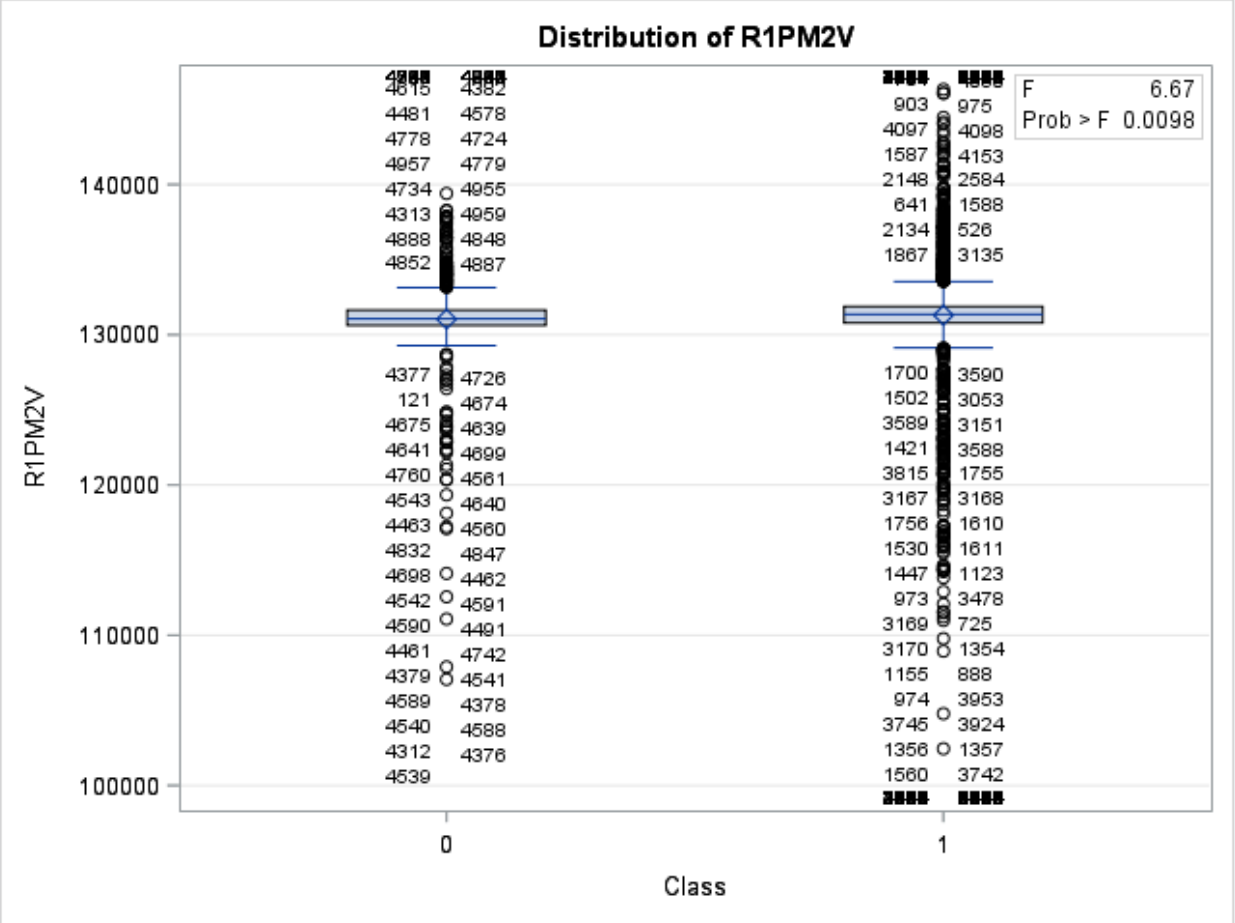

Figure 5: MANOVA analysis result

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Class Effect
H = Type III SSCP Matrix for Class
E = Error SSCP Matrix

S=1 M=5.5 N=2475

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.93793726	25.21	13	4952	<.0001
Pillai's Trace	0.06206274	25.21	13	4952	<.0001
Hotelling-Lawley Trace	0.06616940	25.21	13	4952	<.0001
Roy's Greatest Root	0.06616940	25.21	13	4952	<.0001

Figure 6: MANOVA statistical test for class effects

5. Conclusions and Future Work

This work used PCA for dimension and feature reduction for cyber security and risk analysis in power systems. PCA first reduces the high dimensional cyberattacks dataset, and then the MANOVA analysis was implemented on the data set to differentiate between the natural events and real attacks, and identify the effects of the variables to find which variable is risky for attacks. This research establishes a framework using the multivariate techniques to classify and find patterns to cyber security and its risk analysis for power system. Factor analysis (FA) for dimension reduction and other discriminant analysis such as logistic regression could be considered as a future extension.

References

1. Hink R. et al., 2014, "Machine Learning for Power System Disturbance and Cyber-Attack Discrimination", Resilient Control Systems (ISRCS), 7th International Symposium on IEEE.
2. Liu X., and Zuyi L., 2014, "Local Load Redistribution Attacks in Power Systems with Incomplete Network Information", IEEE Transactions on Smart Grid, 5.4:1665-1676.
3. Hassan M., Hussein M., 2016, "A Study of Resource-constrained Cyber Security Planning for Smart Grid Networks", In Electrical Power and Energy Conference (EPEC), IEEE, 1-6.
4. Kosut O. et al., 2010, "On Malicious Data Attacks on Power System State Estimation", Universities Power Engineering Conference (UPEC), 45th International IEEE.
5. Gupta P., and Manish K., 2016, "A Framework for Fast and Efficient Cyber Security Network Intrusion Detection Using Apache Spark", Procedia Computer Science 93: 824-831.
6. Teixeira A.et al., 2010, "Cyber Security Analysis of State Estimators in Electric Power Systems", Decision and Control (CDC), 49th IEEE Conference.
7. Kim T. et al., 2015, "Vulnerability Analysis of Power Systems", arXiv preprint: 1503.02360.
8. Qin Z. et al., 2012, "Unidentifiable Attacks in Electric Power Systems", IEEE/ACM Third International Conference on Cyber-Physical Systems, IEEE Computer Society.
9. Yang J. et al., 2016, "Performance Analysis of Sparse Recovery Models for Bad Data Detection and State Estimation in Electric Power Networks", Power and Energy Society General Meeting (PESGM), IEEE.
10. Yang Q. et al., 2016, "Towards Optimal PMU Placement against Data Integrity Attacks in Smart Grid", Information Science and Systems (CISS), Annual Conference on. IEEE.
11. Taha A. et al., 2016, "Risk Mitigation for Dynamic State Estimation against Cyber Attacks and Unknown Inputs", IEEE Transactions on Smart Grid.
12. Yu Z. et al., 2015, "Blind False Data Injection Attack Using PCA Approximation Method in Smart Grid", IEEE Transactions on Smart Grid 6.3: 1219-1226.
13. Waghmare S. et al., 2017, "Data Driven Approach to Attack Detection in a Cyber- Physical Smart Grid System", Control Conference (ICC), IEEE.
14. Sou K., Henrik S., and Karl H.,2013, "On the Exact Solution to a Smart Grid Cyber-Security Analysis Problem", IEEE Transactions on Smart Grid, 4(2), 856-865.
15. Ericsson G. N., 2010, "Cyber Security and Power System Communication—Essential Parts of a Smart Grid Infrastructure", IEEE Transactions on Power Delivery, 25(3), 1501-1507.
16. Sridhar S. et al., 2012, "Cyber–Physical System Security for the Electric Power Grid", Proceedings of the IEEE 100.1: 210-224.