Identification of State Parameters for Stealthy Cyber-Events in the Power Grid Using PCA

Nathan Wallace, Sean Semple, and Travis Atkison
Deptartments of Electrical Engineering, Cyber Engineering, and Computer Science
Louisiana Tech University
Ruston, LA 71270
Email: dcsl@latech.edu

***Abstract*—One of the biggest efforts of securing cyberspace is the ability to secure the critical infrastructure power grid. This meshed network of geographically distributed control systems has recently been interlaced with network capable devices. This article manipulates power system state parameters to identify stealthy cyber-events based on the feature identification method principal component analysis. Principle component analysis is used to examine power system instances contained in an in-control set. To develop a set of observable power system instances the Newton-Rhapson method is used to solve the power flow equations. Cyber-events were created by changing the in-control instances of the power system parameters over a range of $\pm 40\%$. If a changed instance, simulated malicious instance, was mapped via principle component analysis within close range of the average Hotelling's T^2 value that instance was deemed stealthy. Results indicate certain features have a higher likelihood of remaining in the stealthy regions.**

I. Introduction

The critical infrastructure power grid has recently seen an increase in the implementation of networked solid state devices. The key goal of such influxes is to increase the number of reporting nodes in the Wide Area Measurement System (WAMS) for the purpose of billing, state estimation, grid health, and for the efficient delivery of electricity to its consumers. However, security becomes a concern when the control decisions being implemented in the power system is based on the values being reported by the nodes in the WAMS.

The approach presented in this article uses a technique known as principle component analysis (PCA) to classify stealthy cyber-events based on the probability of occurrence. PCA is applied to the data resulting from an iterative solution to the power flow equations. The iterative solution used for the development of system data is the Newton-Rhapson method and is known to be the most common approach for solving the power flow equations [1]. Once system data was obtained, PCA was used to transform the data into a new vector space in order to better understand the dynamics of the power system. A cyber event model was then created to simulate either a malicious attack on the system or a failing sensor. The classification scheme developed in this article shows a statistical deviation in the values observed during a cyber event and the values observed under normal operating conditions.

The complexities and of the power grid is described in Section II along with the details of the power system used for this examination. Section III provides an overview of the PCA method along with details regarding sanitization of the data and classification. This section also contains the results of the transformation in the new dimensional space. The cyber-event model is developed in Section IV and describes how the instances are created such that they represent a possible malicious attack on the power system or a failed sensor. Lastly, the results of the cyber-event classification scheme is presented in Section V, followed by conclusions.

II. The Power Grid

The primary steady state algorithms that are used to ensure the stability and reliability of the critical infrastructure power grid are: 1) power flow, 2) optimal power flow, and 3) state estimation [3]. The power flow dynamics throughout a system is determined by using a computational method to solve the power equations, Equations 1 and 2, which are obtained by applying Kirchoff's law at each Bus of the system in question.

$$0 = \Delta P_i = P_i^{injec} - V_i \sum_{i=1}^{n} V_j Y_{ij} cos(\theta_i - \theta_j - \varphi_{ij}) \quad (1)$$

$$0 = \Delta Q_i = Q_i^{injec} - V_i \sum_{i=1}^{n} V_j Y_{ij} sin(\theta_i - \theta_j - \varphi_{ij}) \quad (2)$$

where, P_i^{injec} and Q_i^{injec} are the injected powers into each Bus, V_i is the voltage on Bus i and Y_{ij} is element ij of the admittance matrix. Optimal power flow takes into account the economics of the system in that a cost function is used to determine how much and when to generate power. State estimation is a more realistic approach to understanding the system as conditions in the field may not match a purely analytical solution of the power system. With state estimation, system parameters are measured using intelligent electronic devices (IEDs) and are reported back to a centralized location.

A. State Estimation

The supervisory control and data acquisition (SCADA) system gathers all the sensor data from field intelligent

electronic devices (IEDs) and then according to the system architecture derives a state estimation in order to obtain a complete understanding of the system at that state. The data collected is stored in database format on a server known as the Historian. The state is a function of n system state variables including Bus voltages, phase angles, circuit breaker status, and tap changing transformer position amongst others. The approach for cyber-event classification presented in this article is designed to be implemented on top of the existing SCADA infrastructure. This proposed classification scheme uses the data stored in the Historian to develop a near complete understanding of the power system dynamics where statistical inferences then can be made.

B. Power Flow

The goal of the Newton-Rhapson method when applied to power system analysis is to provide an iterative method for solving the nonlinear algebraic power flow equations, Equations 1 and 2 [4]. This method is known to be the most common method used [1]. The goal of the iterative method is to decrease the vector of errors produced by taking the difference of powers to a certain point that is declared acceptable. For instance the error stopping point used in this approach is $\varepsilon_s = 0.01$. This means that the absolute values of both the active and reactive power mismatches all had to be below 0.01 to be considered a converging instance. Also, for this examination convergence had to occur within 15 iterations or the instance was declared a non-converging instance. On average the five Bus system converged within 4 iterations. The extreme of 15 iterations was selected as a stopping point given that if the system did not converge within 15 iterations it is likely that for that given set of inputs the system can not exists. A flat start means that for simulation purposes all non-voltage controlled Busses are assumed to have a voltage of 1 per unit while all angles are assumed to be zero. For a more detailed description of the iterative solutions to the power flow problem the reader is encouraged to view the following referenced text [1], [4], [5].

C. System Model

To demonstrate the classification of stealthy cyber-events a relatively simple power system was selected. Multiple instances of this model were conducted using the Newton-Rhapson method to solve the nonlinear algebraic power flow equations. Using the 5 Bus power system [5] shown in Figure 1 a series of power flow simulations were conducted. The system shown is a 100 MVA 138 kV system with the swing Bus positioned at Bus #1 or the Slack Bus. Generators are connected at Bus #1 and Bus #2. Loads are connected to every Bus in this model and are identified by that Bus's number. Table I shows the impedances used for the six transmission lines considered in this system model. A snapshot of the Bus input data is shown in Table II. This information serves as the input parameters to the power flow equations and with the successful convergence

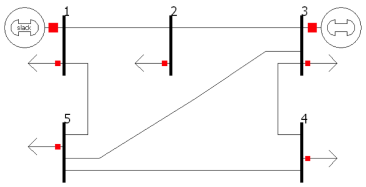

Fig. 1. Five Bus One Line Diagram [5]

of the Newton-Rhapson method the other variables can be determined. Bus #3 is a voltage controlled Bus and is part of the input variable set. The slack Bus is simulated in such a way that given the inputs shown in Figure 1 it picks up the remaining slack to supply the required load.

TABLE I
5 BUS TRANSMISSION LINE PARAMETERS [5]

Bus - Bus	Line Length (mi)	R	X	B
1 - 2	40	0.042	0.168	0.041
2 - 5	30	0.031	0.126	0.031
2 - 3	30	0.031	0.126	0.031
3 - 4	80	0.084	0.336	0.082
3 - 5	50	0.053	0.210	0.051
4 - 5	60	0.063	0.252	0.061

TABLE II
5 BUS INPUT SNAPSHOT

Bus #	Type	V	Delta	PG	QG	PL	QL
1	0	-	0	-	-	0.65	0.3
2	1	-	-	0	0	1.150	0.6
3	2	1.020	0	1.8	-	0.7	0.4
4	1	-	-	0	0	0.7	0.3
5	1	-	-	0	0	0.850	0.4

III. PRINCIPAL COMPONENT ANALYSIS

In any determinable system there is a finite number of driving forces which governs how the system behaves. By observing grouping phenomenon in the data it is possible to replace a group of variables with a single new variable, greatly reducing the redundancy in the data. Principle component analysis (PCA) is a quantitative process for achieving a system simplification. A decrease in redundancy and an overall simplification of the data is made possible through a transformation into a new vector space where all the basis vectors are independent of each other. The basis vectors in the new dimensional space are called principal components [6].

Perhaps one of the most commonly used statistical analysis tools for feature extraction, PCA is based on the statistics of a training set to linearly transform the set in such a way that the new primary basis are independent of each other. The linear transformation used is based on a covariance matrix which is defined by the patterns found in the training set. PCA finds a linear transformation such that

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \tag{3}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are mxn matrices related by a transformation $\mathbf{W}$. Based on Equation 3 the following variables can be defined: $\mathbf{w_i}$ are the rows of $\mathbf{W}$, $\mathbf{x_i}$ are the columns of $\mathbf{X}$, and $\mathbf{y_i}$ are the columns of $\mathbf{Y}$.

The row vectors of $\mathbf{W}$ $\{w_1, ..., w_m\}$ are called the principal components of $\mathbf{x}$. Before PCA can be applied to a data set it is customary to first preform sanitization on the data. This sanitization guarantees any unintended biassing of the new components. After centering the normalized covariance $\mathbf{S_X}$ was determined using the unbiased estimator for normalization.

$$\mathbf{S_X} = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T \tag{4}$$

This produced a covariance matrix with dimensions mxm with the diagonal terms representing the variances and off-diagonal terms representing the covariaces of data matrix $\mathbf{X}$. The closer the off-diagonal terms are to zero the closer the variables represented by the indices of $\mathbf{S_X}$ are to being completely uncorrelated. Conversely, the higher these off-diagonal terms are the more correlated the two variables are. Also the higher the off diagonal terms are the higher the redundancy is in the data matrix $\mathbf{X}$.

The linear transformation produced by PCA selects a transformation $\mathbf{W}$ such that the principle components or basis vectors w_i produced are completely orthonomal. Orthonomality is ensured due to the fact that the dot product of each basis vector with another produces the Kronker delta function, $w_i \cdot w_j = \delta_{ij}$. In addition to being orthonormal, the basis vectors are ordered based on the amount of variance that is being accounted for by that basis vector or principal component. This corresponds to the fact that PCA will produce a transformation matrix $\mathbf{W}$ such that the variance of data matrix $\mathbf{X}$ is mostly accounted for by principal component w_1. As hinted at in the previous section the lower the diagonal terms of the covariance matrix are the lower the redundancy is in the data. Therefore the solution to PCA seeks a covariance matrix $\mathbf{S_Y}$ such that the off-diagonal terms are zero where,

$$\mathbf{S_Y} = \frac{1}{n-1}\mathbf{Y}\mathbf{Y}^T \tag{5}$$

Plugging Equation 3 into Equation 5 we have

$$\mathbf{S_Y} = \frac{1}{n-1}\mathbf{W}(\mathbf{X}\mathbf{X}^T)\mathbf{W}^T \tag{6}$$

With this solution to PCA it can be shown that the principal components of data matrix $\mathbf{X}$ are the eigenvectors of $\mathbf{X}\mathbf{X}^T$ or are the rows of $\mathbf{W}$. Also, the i^{th} diagonal term of $\mathbf{S_Y}$ is the variance of $\mathbf{X}$ projected onto $\mathbf{p_i}$.

A. Classification of Stealthy Attacks

To use PCA for classification we seek a method where for a suspicious data matrix $\mathbf{X'}$, the data can be projected and the successful identification of malicious instance $\overrightarrow{\mathbf{X}'}_m$ is made possible. If malicious instance $\overrightarrow{\mathbf{X}'}_m$ does not follow the statistical trends identified by PCA using the training set $\mathbf{X}$ it can be classified as a potential malicious instance. Once classified the instance then can be further investigated to identify the cause of the compromised reading $\mathbf{x'}_i$. This would allow the investigator or control engineer to isolate the intrusion.

The Hotelling's T^2 value, Equation 7, is an extension of the t-test used to determine the difference between means of two independent variables. This extension allows for a statistical measure of the multivariate distance of each instance from the center of a data set. An instance is labeled stealthy if it is both a malicious instance and occurs close to the multivariate center.

$$T^2 = n(\mathbf{X} - \mu)'\mathbf{S}^{-1}(\mathbf{X} - \mu) \tag{7}$$

The classification approach presented in this article is a probabilistic approach in describing how likely an instance is to occur. Instances that fit to the dynamics of the data matrix $\mathbf{X}$ or control set have a high likelihood of occurring while instances that lie on the boundaries are less likely to occur.

It can also be shown that the Hotelling's T^2 value follows the $\mathcal{F}$ distribution as defined by Equation 8 [7].

$$T^2 \sim \frac{(n-1)p}{(n-p)}\mathcal{F}_{p,n-p}(x) \tag{8}$$

where p is the number of principal components retained and n is the number of instances in the sample space. The $\mathcal{F}$ cumulative probability distribution function returns the cumulative probability of obtaining a value x for given parameters p and n. Rearranging Equation 8 we can calculate that the probability of observing at least T^2 is

$$P(\geq T^2) = 1 - F_{p,n-p}(z) \tag{9}$$

This allows for a probabilistic metric to determine whether or not an instance is in-control. If the instance is in-control then it follows the dynamics as defined by the data matrix $\mathbf{X}$. A low probability, as defined by Equation 9, for observing at least that T^2 corresponds to a high T^2 value. This means that the instance is far away from the multivariate center and therefore is least likely to occur. Conversely, a high probability corresponds to a low T^2 value and is therefore closer to the center of the data.

Any in-control instance can be considered an instance whose variables follow the dynamics of the system. These instances can be considered instances that would occur under normal operation. Because these instances are likely to occur under normal operation any malicious instance contain in this set is classified a stealthy event. An out-of-control instance would be an instance whose dynamics do not fit uniformly in with the dynamics of the in-control instances. Out-of-control instances are not considered normal operation and therefore any operation that exists outside of normal operation can be classified as an out-of-control instance. If a malicious instances is located in this out-of-control set then it is not classified a stealthy event. The stealthy set $\mathbf{S}$ can be classified into the following three

regions: slightly, reasonably, and extremely stealthy based on how far that instance occurs from center. Table III shows the classification metric used in this article to classify the stealthiness of cyber-events. This metric is based on the standard deviation σ of the average Hotelling's T^2 value, $\bar{T}^2$, and is therefore also based on the probability of occurrence as defined by the $\mathcal{F}$ distribution, Equation 9.

TABLE III
CLASSIFICATION METRIC FOR STEALTHYNESS

Notation	Description	Region
$\mathbf{S}_s$	Slightly	$\frac{1}{2}\sigma \leq \bar{T}^2 < \frac{3}{4}\sigma$
$\mathbf{S}_r$	Reasonably	$\frac{1}{4}\sigma \leq \bar{T}^2 < \frac{1}{2}\sigma$
$\mathbf{S}_e$	Extremely	$\bar{T}^2 < \frac{1}{4}\sigma$

B. PCA Transformaiton

Once mapped to the new dimensional space it was determined that the first principal component or basis vector accounted for over 15% and the second accounted for over 10% of the variance found in the data set. As noted in the previous section, the Hotelling T^2 value is the multivariate distance from the center of data. This value is calculated using Equation 7 for each of the 13,741 converging instances of the data matrix $\mathbf{X}$. By calculating the T^2 value for every instance it was determined that the power system data when plotted in the new dimensional space has a max distance of $\bar{T}^2_{max} = 4717$ and a mean distance of $\bar{T}^2 = 18.99$ from the data center with a standard deiviation $\sigma = 88.06$.

IV. CYBER-EVENT MODEL

The cyber-event model used for classification is two-fold in that it represents two possibilities that can occur in power system WAMS. Event #1 can be considered to be a non-malicious incident in which the controller or sensor in the field making the measurement breaks or becomes damaged as a result of natural causes. Some examples of this may include natural disasters, faulty equipment, or wear on the device over the years. Such readings may fluctuate within a certain percent of its actual measured value. Event #2 can be classified as an actual malicious event in which an attacker purposely launches a data injection attack against the control system. Examples of this include the falsification or spoofing of data values reported by a smart sensor as revealed by Brinkhaus et al [8]. This work currently makes no distinction of the two events only that it is able to classify the event that did occur.

To simulate these types of cyber-events being reported by the WAMS 100 instances were selected from the data matrix $\overrightarrow{\mathbf{X}_i}$ such that each subsequent event has the highest probability of occurring, i.e $P(\geq T^2(\overrightarrow{\mathbf{X}}_1)) \geq P(\geq T^2(\overrightarrow{\mathbf{X}}_2)) \cdots \geq P(\geq T^2(\overrightarrow{\mathbf{X}}_{100}))$. These vectors provide a starting point to simulate several thousand cyber-events. The 19 features of each of the 100 instances were then changed individually between $\pm 40\%$ at 1% intervals of its original value. After each change, PCA was performed and the new instances was plotted against all instances of the original in-control set. This iterative process of changing each element of each instance produced a suspicious set $\mathbf{X}'$ whose size after sanitization was 13741x19. Sanitization of the suspicious set included the deletion of all instances $\overrightarrow{\mathbf{X}_i}'$ whose $T^2(\overrightarrow{\mathbf{X}_i}') > \frac{8}{7}\sigma \simeq 100$ as this falls well outside the region of slightly stealthy.

V. EVENT CLASSIFICATION

If a cyber event has occurred, it is desired to detect and classify such an event and be able to alert on intrusion or failure. This immediate feedback will allow the trigger of an alarm allowing a security analyst or control engineer to further investigate the event. Given that we now have defined a transformation matrix $\mathbf{W}$ such that this transformation has eliminated all redundancy when mapped to the dimensional space we can now interpret new instances of the power system. With the suspicious set $\mathbf{X}'$ and the classification regions described in Table III event mapping can now begin. This process of mapping small malicious changes based on the T^2 and comparing it to an average under steady state conditions will produce a mapping of malicious instances that can occur in the bounded regions of an average. An example of this would be if the power system instances naturally observed average around a given T^2 point in the multivariate system then by creating an event, be it an attack or not, such that its T^2 occurs in close proximity then that malicious instances can be labeled suspicious. The results show that by changing the values of different features of the same class by the same amount, i.e changing the voltage 1 and voltage 2 by 10%, will have a different T^2 and therefore a different level of stealthiness.

For each of the 13,741 instances the 19 features were changed across a boundary of $\pm 40\%$. After each simulated event, that event's T^2 was compared to the the average using a metric based on the standard deviation σ. This allows for a classification of stealthiness based on how close it occurs to the average T^2 value. After each simulated instance a count was kept expressing how many simulated instances occurred within each of the regions based on a change in percent of its original value. The results for the voltages and powers consumed by the loads are shown in Figure 2. Recall that $S_s, S_r,$ andS_e refer to the following regions of steeliness: slightly, reasonably, and extremely. The top two graphs show a count of instances occurring in the slightly region, the middle two show a count in the reasonably region and the bottom two graphs show the extremely stealthy region. Though this information only reveals a count of instance occurrence for each feature and percent change it offers valuable comparative information.

From Figure 2 it can be determined that there is a higher count of V5 changes in the S_s region, slightly stealthy region, than there is V4 changes. In other words the same percent change in voltage at Bus 5 is more likely to occur in the slightly region than it would for the voltage at

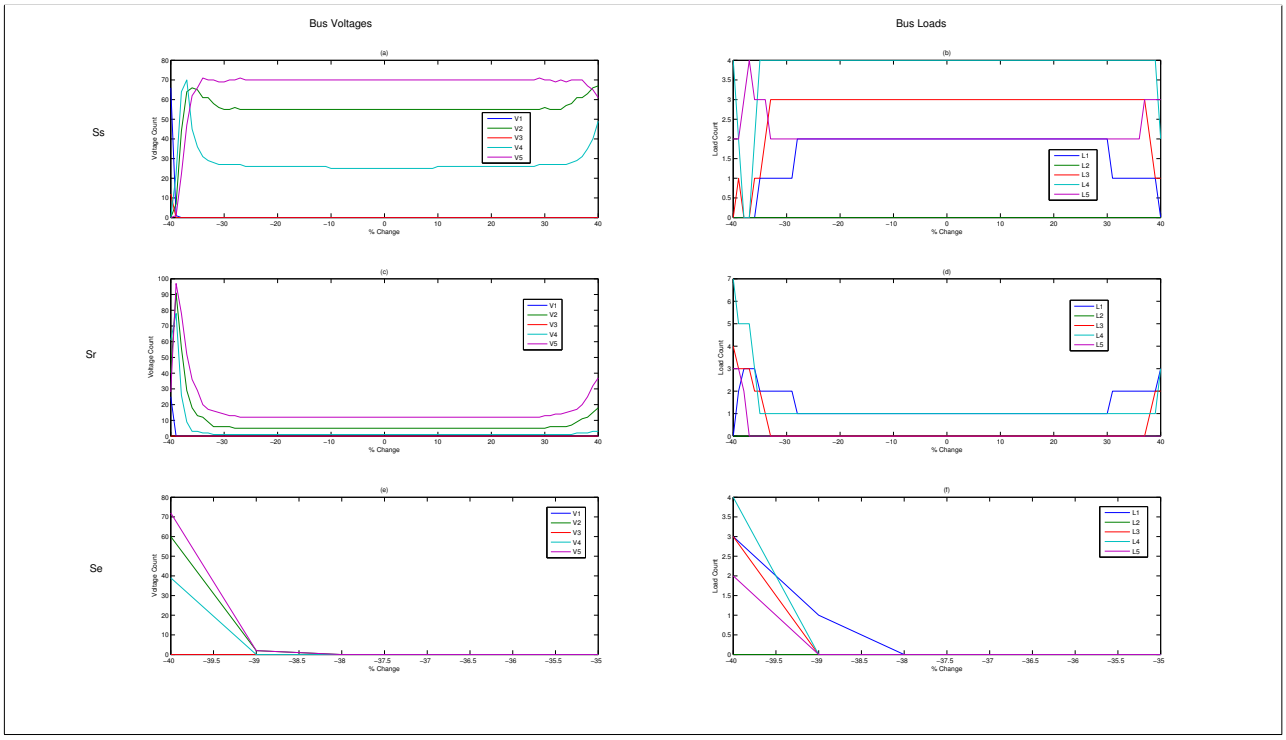

Fig. 2. Malicious Event Classification Count of Voltage and Load Events in Each Region.

Bus 4. In fact the count reveals a count of double that of the voltage at Bus 4. Also, this information reveals that small changes in the load values have drastic effects on the stealthiness of the overall instance occurring. Of all the malicious state parameters simulated, the voltage parameters remained amongst the highest count found to be contained within the declared regions of stealthiness. Loads, however, were observed to have the lowest count as shown in Figure 2. This information reveals that it is harder to spoof load values and remain in close proximity of the average than it is to make voltage changes and remain in the regions of stealthiness.

VI. Conclusion

The dimensional reduction provided by PCA offers a way to observe system dynamics that otherwise would remain hidden. The transformation decreases the redundancy of the data allowing for a better understanding of the data set. The process of PCA was applied to a 5 Bus power system and a cyber-event identification scheme was developed based on the Hotelling's T^2 values of suspect and non-suspect instances of the power system. Instances of the power system were determined using the Newton-Rhapson method of mismatch error less than 0.01 and convergence was required within 15 iterations. The instances defined by this iterative method was then transformed to the new space using PCA where certain inferences can take place.

By extracting in-control instances of the 5 Bus power system and changing features one at a time over a range of $\pm 40\%$ and mapping that instances to the new dimensional space it was determined that some changes fall close to the norm while others occur at far distances. Based on this information three regions of stealthiness were labeled such that the standard deviation σ of the Hotelling's T^2 value was used has a identification metric for stealthiness. If a change, simulated malicious instance, was mapped within close range of the average Hotelling's T^2 value that instance was deemed stealthy. Results indicate that in order to remain stealthy, within $\frac{1}{2}\sigma$ of $\bar{T}^2$, it is better to change the Bus voltages than it is to change the Bus load values. Similarly it was determined that by changing voltage at Bus 1 it is more likely to remain in the stealthy region then by changing the voltage at Bus 4.

Acknowledgment

This research was supported by a Louisiana Board of Regents Graduate Fellowship.

References

[1] *Power Systems (The Electric Power Engineering Hbk, Second Edition)*. CRC Press, 2007.
[2] S. P. Pool, "Spp 101," Nov 2012.
[3] *Power Systems (The Electric Power Engineering Hbk, Second Edition)*. CRC Press, 2007.
[4] J. D. Glover, M. S. Sarma, and T. Overbye, *Power System Analysis and Design, Fifth Edition*. Cengage Learning, 2011.
[5] W. D. Stevenson, *Elements of Power System Analysis (Mcgraw Hill Series in Electrical and Computer Engineering)*. Mcgraw-Hill College, 1982.
[6] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*. Springer, 2007.
[7] W. K. Hrdle and L. Simar, *Applied Multivariate Statistical Analysis*. Springer, 2012.
[8] C. D. Brinkhaus S., "Smart hacking for privacy." 28th Chaos Communication Congress, 2011.