A Dimensional Transformation Scheme for Power Grid Cyber Event Detection

Nathan Wallace
Louisiana Tech University
Ruston, Louisiana USA
nsw004@latech.edu

Stanislav Ponomarev
Louisiana Tech University
Ruston, Louisiana USA
spo013@latech.edu

Travis Atkison
Louisiana Tech University
Ruston, Louisiana USA
atkison@latech.edu

ABSTRACT

As the technologies used for the safe and efficient delivery of power become more sophisticated, the amount of system state parameters being recorded increases. This data not only provides an opportunity for monitoring and diagnostics of the power system but also creates an environment wherein security can be maintained. Being able to extract relevant information from this pool of data in a reasonable amount of time is one of the key challenges still yet to be obtained in the smart grid. New power grid security applications can be created that use the statistical patterns in the reported data as a metric for security. Anomalies detected by the developed security metrics can then be alerted upon as a possible cyber intrusion. This article is an examination into the utilization of principal component analysis along with a Naive-Bayes classifier for the identification of spoofed power system state parameters. Examination targets a 5 Bus power system with results indicating successful classification of simulated cyber attacks or true positive classification at a rate of 92%. These findings also indicate a dependency on which variables were compromised, providing an initial formalization into the stealthiness of state estimation attacks.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Network**]: Security and Protection; B.8.1 [**Performance and Reliability**]: Reliability, Testing, and Fault-Tolerance

General Terms

Security

Keywords

SCADA, PLC, control systems, state estimation, intrusion detection

1. INTRODUCTION

One of the biggest challenges of securing cyberspace is the ability to secure the critical infrastructure power grid. For prudent reasons, this meshed network of geographically distributed industrial control systems (ICS) has recently been interlaced with network capable devices. Such interlacement, unbeknownst to the utility provider or independent system operator (ISO), can provide an individual or city-state with malintent direct access to the control local area network. Once the control LAN has been breached, control decisions can be made that are outside the intended operation specifications, the most harsh being a full denial of service attack. In a recent effort known as Project Shine, over 7,200 control devices were found to be directly connected to the World Wide Web [7]. These startling results indicate that critical control devices have and will continue to be accidentally connected in a manner that is inconsistent with the so called 'air-gap' separation. Other possible and in some cases historically documented breaches into power systems are conducted via insider threat, the use of a zero-day attacks, or unpatched system attacks.

To solve the problem of detecting power system cyber intrusions, a context specific approach is developed that utilizes the physical perspective of the power system in order to derive a cyber conclusion. Creation of the physical perspective is based on the extraction of information contained within historical or contingency power system states. By examining these states or instances, the power flow dynamics can be extracted and new patterns formed that offer a signature of the trusted system being observed. Once the trusted model is created utilizing the physical perspective, all incoming power system state parameters transported via the cyber infrastructure can be checked and verified against the trusted model. The approach this article takes utilizes a dimensional transformation that guarantees the decrease of redundant information and once in the new dimensional space applies a Naive-Bayes classifier for the detection of cyber events. Specifically, principal component analysis is used as the approach for transforming power system instances and the Hotelling T^2 metric is used for the classification of each newly observed instance. Once an instance is labeled as suspect, the state parameters contained within that instance are compared against the variances of previously observed or trusted state parameters in an effort that identifies the node or control device that was the target of the intrusion.

Details and model assumptions of the power grid are described in Section 2. An overview of the dimensional transformation technique, principal component analysis (PCA), is given in Section 4. Section 4.1, applies principal component analysis to the 5 bus power systems. The cyber-event model outlined in Section 3 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 detection scheme are presented in Section 4.3 followed by conclusions.

CISR '14, Apr 08-10, Oak Ridge, TN, USA
ACM 978-1-4503-2812-8/14/04.
http://dx.doi.org/10.1145/2602087.2602108

2. 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 [1]. The power flow dynamics throughout a system are 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.

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

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

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.

2.1 State Estimation and Power Flow

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. Due to computational requirements, state estimation calculations are conducted only at periodic intervals. The state estimator is based on m imperfect telemetered data points from the power system application. 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 detection presented in this article is designed to be implemented on top of the existing SCADA infrastructure utilizing only the historical data and newly observed state parameters. Other approaches for security focus on reliability and do not consider the full power of historical data [2].

Power flow analysis uses an iterative method, in most cases the Newton-Rhapshon method [1], for solving the nonlinear algebraic power flow equations, Equations 1 and 2 [5]. Convergence is said to happen when the error or mismatch drops below a certain threshold. 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 5 bus systems 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 for that given set of inputs the system cannot exist. The fact of non-convergence corresponds to the likelihood that the power system being observed does not exist at that given set of inputs. 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, 5, 8].

2.2 Power System Models

In order to demonstrate the implementation of PCA for cyber-event detection a 5 Bus power system [8] with 6 transmission lines operating at 135 kV with a base power of 100 MVA was simulated. Multiple instances of this power system were determined via a pseudorandom nonsequential Monte Carlo simulation about a converging power system state. This and other similar power flow analysis techniques are common place, offering a probabilistic perspective of power flow based on system and topological constraints [1]. Each input variable was changed about a snapshot according to a probability density function of a uniform distribution between zero and two. Upon completion of the Monte Carlo power flow simulation, a data matrix $\mathbf{X}$ is constructed where each row x_i represents an observable power system instance with each element $x_i[j]$ representing a state variable contained within that instance. Therefore each state variable is denoted by a column vector x_j contained within the data matrix $\mathbf{X}$.

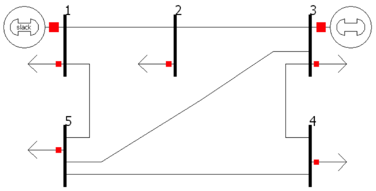

Figure 1: 5 Bus Power System Oneline Diagram [8]

3. CYBER-EVENT MODEL

The cyber-event model used for this detection approach is two-fold in that it represents two possibilities that can occur in a power system. Event #1 can be considered 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. Event #2 can be classified as an actual malicious event in which an attacker purposely launches an attack against the control system. Examples of this include the falsification or spoofing of data values reported from a smart meter as revealed by Brinkhaus et al [3]. This work currently makes no distinction of the two events only that it is able to determine that an event occurred. Once detection has occurred that instance then can be further investigated and the actual cause of the event can be determined.

The approach presented in this article assumes that both Event#1 and Event #2 will produce a state value of zero at that origin of the event. This assumption provides an initial starting point for the development of the detection scheme presented in this article. Furthermore the cyber-event model assumes that only one cyber-event occurs per observed power system instance. To simulate these types of events a random instance from data matrix $\mathbf{X}$ was selected. This random instance vector $\overrightarrow{X_r}$ serves as the basis for the cyber-event simulation. Next an element of $\overrightarrow{X_r}$ is changed in a manner that reflects the desired simulated cyber-event and added as a row to a suspicious data set $\mathbf{X}'$. Each row in the newly created suspicious data set represents a possible power system observation where a cyber-event has occurred. A total of 50 instances were selected at random from the original data set and each of the 47 power system sate parameters were changed to zero in an iterative fashion. This approach produced an attack matrix or suspicious set that contains a total of 2350 simulated cyber-events.

4. PRINCIPAL COMPONENT ANALYSIS

In any determinable system there generally are a finite number of driving forces which governs how the system behaves. By observing grouping phenomenon in reported power system states it is possible to replace a group of power system variables with a single new variable, greatly reducing the redundancy in the data. Principal component analysis (PCA) is a quantitative process for achieving a dimensional simplification [4]. The dimensional simplification provides a decrease in redundancy 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 [4]. 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 $\mathbf{Y} = \mathbf{WX}$ where $\mathbf{X}$ and $\mathbf{Y}$ are mxn matrices related by a transformation $\mathbf{W}$. The row vectors of $\mathbf{W}$ $\{w_1, ..., w_m\}$ are called the principal components of $\mathbf{x}$.

The linear transformation produced by PCA selects a transformation $\mathbf{W}$ such that the principal 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. The solution to PCA seeks a covariance matrix $\mathbf{S_Y}$ such that the off-diagonal terms are zero where, $\mathbf{S_Y} = \frac{1}{n-1}\mathbf{YY}^T$. It can be shown through the resulting solution, $\mathbf{S_Y} = \frac{1}{n-1}\mathbf{W}(\mathbf{XX}^T)\mathbf{W}^T$, that the principal components of data matrix $\mathbf{X}$ are the eigenvectors of $\mathbf{XX}^T$ 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 $\mathbf{p_i}$. Before PCA can be applied to a data set it is customary to first preform a certain amount of sanitization on the data. This sanitization eliminates any unintended biassing of the new components and is accomplished through centering and normalizing the dataset. Centering of the data matrix is accomplished by using the mean vector $\mu = E[\mathbf{x}] = [E[x_1], E[x_2], ..., E[x_{16}]]$ where $E[x_1]$ is the mean of the first column of the data matrix $\mathbf{X}$. These values are then subtracted from the associated columns to produce a newly centered data matrix $\mathbf{X}$. From here the normalized covariance $\mathbf{S_X}$ was determined using the unbiased estimator for normalization. This produced a covariance matrix with dimensions mxm with the diagonal terms representing the variances and off-diagonal terms representing the covariances of the matrix $\mathbf{X}$.

4.1 Transformation Results

Using the power system instances contained within the data matrix $\mathbf{X}$, PCA was performed transforming the observed power system instances into a new dimensional space. The first axis or dimensional basis vector in this new space accounted for over 35% of the variance found in the data matrix $\mathbf{X}$ and the second axis (principal component) accounted for $\approx$ 20% of the variance. Figure 2 shows each power system instance plotted as single points (scores) in the new dimensional space provided by PCA along with the vector projections of the original power system state variable features. To prevent clutter, only a select few of the vector projections are labeled. A total of 47 power system features were maintained in the transformation and include: real and reactive powers generated, real and reactive powers consumed, real and reactive powers injected, bus voltages, and bus angles. The real powers injected into the the transmission lines are labeled as *PLij* where *i* is the source bus and *j* is the destination bus. Using this notation directionality is maintained and transmission line losses can be accounted for in the detection scheme. Other vector projection labels include *PGei* and *Voli* representing the power generated and voltage at bus *i*.

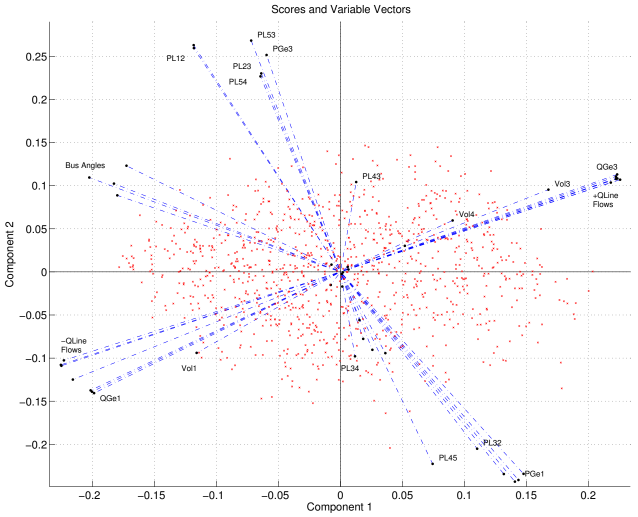

Figure 2: PCA Scree Plot

4.2 Detection Using PCA

The Naive Bayes classifier Hotelling T^2 metric, $T^2 = n(\mathbf{X} - \mu)'\mathbf{S}^{-1}(\mathbf{X} - \mu)$, is utilized for detection and 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 multivarite distance of each instance from the center of the data set in the reduced dimensional space. The result allows for the detection of instances that occur at far distances from the data center as defined by data matrix $\mathbf{X}$. The detection 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 3 [6].

$$T^2 \sim \frac{(n-1)p}{(n-p)}\mathcal{F}_{p,n-p}(x) \tag{3}$$

where p is the number of principal components retained and n is the number of instances in the sample space. Because over 90% of the variance is accounted for by the first 8 principal components, an value of $p = 8$ was used. The $\mathcal{F}$ cumulative probability distribution function returns the cumulative probability of obtaining a value x for given parameters p and n. Rearranging Equation 3 we can calculate that the probability of observing at least T^2 is $P(\geq T^2) = 1 - F_{p,n-p}(z)$ where,

$$z = T^2 \frac{(n-p)}{p(n-1)}$$

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 $\mathbf{X}$. Using the maximum Hotelling T^2 value as a threshold all newly observed power system instances are classified as either suspect or non suspect. The smaller the value the closer the power system instance aligns with the dynamics of the trusted model. Then upon classification a control engineer can perform further analysis to determine the root cause of the cyber-event.

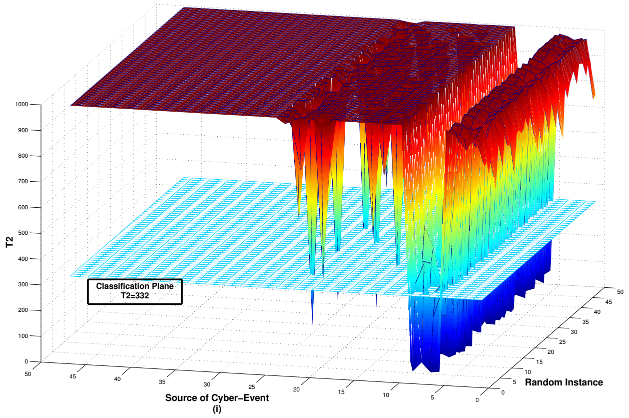

Figure 3: Cyber-Event T^2 Values and Classification Plane

4.3 Detection Results

If a cyber event has occurred it is desired to detect such an event and be able to alert on intrusion or failure. This immediate feedback will allow the trigger of some 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 cannot interpret new instances of the power system. With a trusted model derived from known instances a threshold value, T^2_{thr} was utilized to classify newly observed power system instances and is based on the maximum Hotelling T^2 of the trusted model in the transformed dimensional space. When each of the 2,350 simulated cyber-events were mapped to the new dimensional space as a single score, that event's Hotelling T^2 value was calculated. If the T^2 value is above the threshold the corresponding power system instance is alerted upon and classified as containing a cyber-event. Lower threshold values can be utilized, however this may increase the rate of false positive classification. Using a trusted model containing 1,000 simulated power system instances, the maximum threshold value was determined to be $T^2_{thr} = 332$. Though the instance that corresponds to this maximum threshold value was contained within the trusted model the probability of the next power system instance occurring at a greater distance is approximately zero.

Figure 3 shows the T^2 for each of the 2,350 simulated cyber instances plotted as a contour. The 'Random Instance' denotes the random power system instance that was selected from the trusted model and malformed to represent a cyber-event. The 'Source of the Cyber-Event' denotes which power system state variable *i* was malformed, or in the case of this particular analysis changed to zero. A description for each of the sources, power system state variables *i*, is presented in Table 1. Also, provided in Table 1 are the average T^2 and $P(\geq T^2)$ for each grouping of variables and is denoted as $\bar{T}^2$ and $\bar{P}(\geq T^2)$. For reference the threshold plane is also shown in Figure 3. Any instance whose corresponding Hotelling T^2 value was determined to be above this plane is classified as containing a cyber-event. A high T^2 value corresponds to a low probability and therefore can be classified as a cyber-event. Out of the 2,350 simulated cyber events a total 2,167 where found to be at a distance greater than T^2_{thr} for a successful classification rate of 92.2%. Both Figure 3 and Table 1 reveal that cyber-events occurring at the voltage angles are more likely to be misclassified. Furthermore, on average the probabilities of observing a Hotelling T^2 value greater than the associated cyber-events were all determined to be $P(\geq T_i^2) \approx 0$.

Table 1: Source Description and Simulation Results

Source i	Description	T^2	% $P(\geq T^2)$
1-5	Bus Voltages	939.6	0
6-9	Bus Angles	153.0	0
10-13	Powers Generated	992.4	0
14-23	Powers Consumed	961.4	0
24-47	Powers Injected	995	0

5. CONCLUSION

Using a dimensional transformation to transform observed power system instances a probabilistic metric is created to successfully classify simulated power system cyber-events. Principal component analysis (PCA) was used as a feature reduction method reducing 47 power system state variables into 8 principal components. The process of PCA was applied to a 5 bus power system and a detection scheme was developed based on the Hotelling's T^2 values of suspect and non-suspect instances of the power system. Trusted 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. Power system cyber-events were simulated by extracting 50 trusted power system observations and changing each power system state variable in an iterative fashion to zero. A change of zero may be the result of faulty equipment or an individual spoofing power system variables in an effort to lower his utility bill. This approach resulted in a suspicious testing set containing a total of 2,350 simulated power system cyber-events. A total of 92.2% of the simulated cyber events were found to be over the maximum Hotelling's T^2 and all events were found to have an occurance probability of $P(\geq T_i^2) \approx 0$.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] *Power Systems (The Electric Power Engineering Hbk, Second Edition)*. CRC Press, 2007.

[2] George Anders. *Probability concepts in electric power systems*. Wiley, New York, 1990.

[3] Carluccio D. Brinkhaus S. Smart hacking for privacy, 2011.

[4] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, and Lukasz A. Kurgan. *Data Mining: A Knowledge Discovery Approach*. Springer, 2007.

[5] J. Duncan Glover, Mulukutla S. Sarma, and Thomas Overbye. *Power System Analysis and Design, Fifth Edition*. Cengage Learning, 2011.

[6] Wolfgang Karl Hardle and Leopold Simar. *Applied Multivariate Statistical Analysis*. Springer, 2012.

[7] ICS-CERT. Project shine. *ICS-CERT Newsletter Monthly Monitor*, October-December, 2012.

[8] William D. Stevenson. *Elements of Power System Analysis (Mcgraw Hill Series in Electrical and Computer Engineering)*. Mcgraw-Hill College, 1982.