Decision Tree Based Anomaly Detection for Remedial Action Scheme in Smart Grid using PMU Data

Vivek Kumar Singh, Manimaran Govindarasu
Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011.
(Email: *vsingh@iastate.edu, gmani@iastate.edu*)

***Abstract*—The advanced and persistent cyber threats facing the critical infrastructure such as the smart grid are exponentially rising which require sophisticated defense strategy. Remedial Action Scheme (RAS), also known as Special Protection Scheme (SPS), relies on the interconnected cyber physical system for automated protection which is exposed to the multitude of vulnerabilities. In this paper, we propose an innovative approach to develop an Intelligent Remedial Action Scheme (IRAS) which can detect and distinguish cyber attacks from the physical disturbances in smart grid and later take smart corrective actions as required to minimize the impact on system reliability and economy. Specifically, we have proposed the decision tree based anomaly detection methodology which can distinguish between the normal tripping during power line faults and malicious tripping attack on the physical relays in the context of RAS. The classification model is developed using differential features of voltage and current phasors. Next, as a proof of concept, we have implemented and validated the proposed methodology in cyber physical environment at Iowa State's PowerCyber testbed. Finally, the proposed methodology is tested on modified IEEE 39 bus system in offline and real-time mode. Our experimental results show that the proposed method is efficient in detecting attacks and performing corrective actions within an acceptable time frame.**

***Index Terms*- remedial action scheme, cyber attack, phasor measurement units, decision tree.**

I. INTRODUCTION

With the motive of making the grid modernized, more advanced technologies have been introduced in the infrastructures including smart meters, IEDs and phasor measurement units (PMUs). The installed infrastructures require advance communication to meet the growing demand and further maintaining the stability and reliability of the system. The latest advancement in the communication is making cyber security more challenging. It has allowed increased attacked surfaces and many cyber intrusion success points for the attackers. The malicious tripping attack is one of the possible phenomenon which can happen by injecting trip commands to the circuit breakers connected to the transmission lines. The malicious tripping can cause the unnecessary opening of healthy transmission lines, even though the system is not experiencing any physical disturbances. In 2011, due to lack of secure N-1 operating state, the loss of single 500 KV transmission line has initiated cascading outages leading to severe widespread blackout which has affected more than 2 million people in the U.S. Based on the combined NERC and FERC report [1], the line outage happened due to the human error leading to the line fault (phase-to-phase), however, switching of lines due to cyber-attacks (malicious tripping) may also lead to similar consequences. Ukraine's power grid hack is one of the known cyber attacks on the power grid, which causes shutdown of multiple substations and affected 0.225 million customers [2].

Remedial Action Scheme (RAS), also known as Special Protection Scheme (SPS), is widely used as autonomous protection scheme which takes corrective action during disturbances. There exists the limited research works which address the cyber vulnerabilities in the RAS. Researchers in [3], [4] have shown how the malicious tripping of breakers can be orchestrated in developing the coordinated attack targeting the RAS. The previous research works provide a clear motivation to develop more advanced, sophisticated RAS scheme in the context of cyber physical security. The deployment of real time PMU and advanced wide area communication have enhanced the situational awareness for wide area protection. Using PMU, It is possible to develop the sophisticated RAS scheme which can handle credible disturbances in a very short time allowing less severe impact on the system [5].

Generally, the RAS scheme is designed to protect the system from abnormal events like tripping of breakers during line faults or maintaining transient stability once fault is cleared. It is not designed to handle failures due to cyber attacks [6]. For example, the malicious tripping attacks can open circuit breakers at an inappropriate time, which trigger the RAS controller to take unnecessary actions like shed the generation or load which may affect the operational reliability and the market economy. In this paper, we propose an intelligent remedial action scheme which relies on PMUs to obtain system information and data mining based decision tree approach has been employed to classify the normal, malicious operations and further takes automated corrective actions based on the classification. We have considered the normal operation of relays when the line is out due to physical faults (3phase to ground) and malicious operation during tripping of line without any disturbances. We have computed the differential features of PMU data and generated a library of datasets capturing the system dynamics which is used for training and building the decision tree rules. Furthermore, the trained decision tree is used for offline and real-time testing

Acknowledgement: This research is funded in part by a grant from the DOE CEDS program.

using Hardware-In-the-Loop (HIL) simulation platform in cyber physical environment at Iowa State's PowerCyber testbed.

II. RELATED WORKS

In this section, we provide a quick background on RAS, decision tree related applications and discuss about the research works done by other researchers. As the number of PMUs installed in substation keep increasing, their applications related to wide area monitoring and protection are gaining more popularity. The synchrophasor based Remedial Action Schemes (RASs) have been implemented in the real world to perform autonomous corrective actions and maintain the system stability during the component failures. The PSERC report in [7] talks about PMU based RAS schemes deployed in utilities and companies such as BPA, SCE, etc. Decision tree is one of the prominent data mining tools which has the potential to classify the complex, sophisticated problems into a set of simple rules, making it easier to comprehend and interpret. Due to high availability of PMUs data, decision tree based data mining algorithm are widely employed for dynamic events classification and, in general, solving power science engineering related problems including developing sophisticated RAS. Kamwa et al. talks about PMU based RAS for predicting the catastrophic power system events using ensemble decision trees based classification model [8]. Since the accuracy of the classification model depends on the input features, different methods and techniques have been proposed with a motive of selecting the significant features [9], [10]. Samantaray et al. shows how the computed differential features of PMUs can be used as inputs to build the decision tree based boundaries [11]. Although very useful, none of these works completely address the challenge in the context of cyber physical security in RAS. In this work, we are considering the possible cyber attack on circuit breaker through the malicious tripping as an another event apart from the normal breaker tripping due to line faults and integrate the classification model developed for detecting cyber attack with existing infrastructure.

III. PROPOSED METHODOLOGY

The proposed intelligent RAS (IRAS) scheme addresses the limitation of current RAS scheme by incorporating data mining based decision tree (DT) technique for predicting malicious and legitimate behavior of relays. Once the specific event is detected, appropriate corrective actions are performed based on the scenarios. Fig. 1 shows the flow chart of the proposed IRAS methodology. We have considered generation and load rejection based RAS scheme which allows generation reduction and appropriate load shedding during abnormal conditions. Initially, It collects data from PMUs at regular interval in terms of relays status and line flows calculated from current and voltage phasor measurements. Once the relay is tripped, the RAS controller is activated. It checks with the decision tree based classification model to detect the malicious and normal tripping. During normal tripping of relays which generally happen during the line faults, the RAS controller sheds the generation and load if the power flows in the remaining lines exceed the operational transfer limit (OTC). The OTC maximum limit (OTC_max limit) during contingency is computed based on the thermal limit of lines and other factors, including angular and voltage stabilities are not considered in this work. It is provided to the RAS controller (RASc) through the predefined action table computed through offline contingency analysis. During the malicious tripping, the RASc is allowed to perform two functions. First, it recloses the tripped relays to avoid overloading on other lines instead of shedding the generation and load. Second, it sends the alarm alertness to the operator to provide situational awareness. The reclosing of tripped relay in a shorter time frame allows the system to retain its initial stable state as well as minimizing impact on system economy.

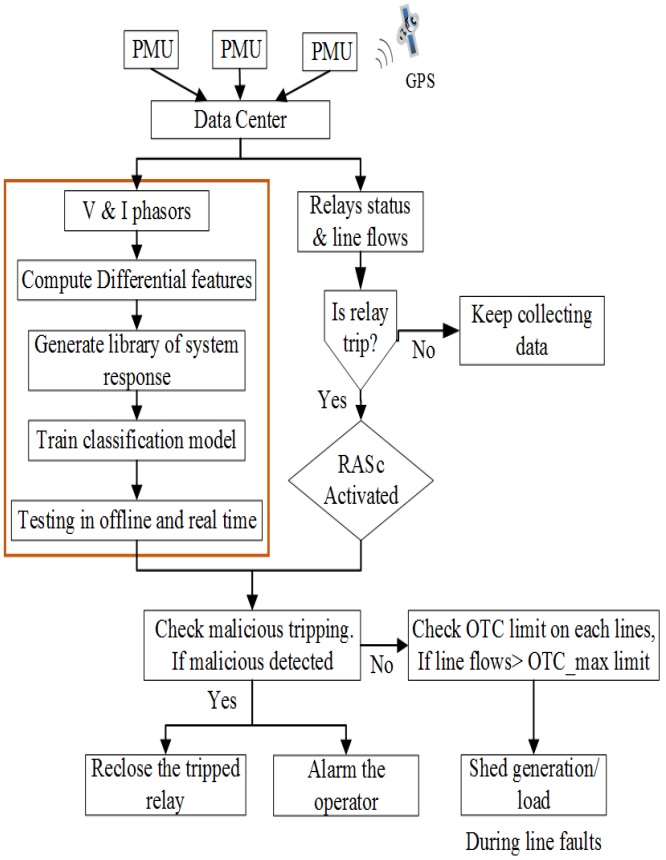

Fig. 1. Proposed methodology for Intelligent Remedial Action Scheme (IRAS)

The colored box of Fig. 1 shows the different steps involved in the input selection, building, training and testing of DT. The voltage and current phasors are collected on both sides (sending and receiving end) of the transmission lines from the PMUs which are used for computing differential features as shown in Table 1. Table 1 shows 8 different differential features extracted for building the classification model where subscripts s and r represent the sending and receiving end of the transmission lines. From the power system's perspective, sudden line outages due to the malicious tripping and normal tripping during the faults look similar making it difficult to classify using raw PMU data. However, they leave a unique footprint on voltage and current waveform of the system and the differential PMU features can assist in identifying the

unique events happening in cyber and physical layer. The differential feature's values deviate sharply during the line fault as compared to the line outage without any faults and thus assists the mining process.

We have generated a library of system database by capturing the system dynamic performance through multiple contingency simulations. The modeled sample system is characterized by generating capacity, load levels, system topology, etc. For different legitimate contingencies, we have considered 3 phase to ground fault where fault locations, fault durations and system operating condition are varied to generate datasets. We have also varied the operating conditions during the malicious tripping of relay where the line is out without any physical faults. Finally, the library of the generated database is employed for training and building the DT. The DT removes the unnecessary features and generate decision rules using most relevant features for classifying the events which is further used for offline and real-time testing.

It is worth mentioning that the system may experience temporary and permanent faults and It is required to follow certain guidelines while performing reclosing when temporary fault is cleared as discussed in [2] which is beyond the scope of this paper. We have not considered the planned outages where transmission lines are out during maintenance by operators. Since RAS is not required during planned outages, it can be disabled or other appropriate actions can be taken by operators.

IV. ATTACK VECTOR

The malicious tripping attack can be performed in multiple different ways. Attackers can perform the tripping attack by getting unauthorized access to the control center. At substation level, setting of physical relays can be altered to cause tripping of breakers. Attackers can also perform the attack through the SCADA communication network. In this work, we have implemented the tripping attack by eavesdropping the network packets going between the substation and control center. Once the attacker has internal access to wide area network, he/she can easily learn about the network packets used to trip the relays and eventually replay the tripping packet to perform the attack.

V. EXPERIMENTAL SET UP AND CASE STUDY

Fig. 2 shows hardware in the loop based experimental setup for implementing the attack, as well as offline and real-time testing of the proposed approach. The modified IEEE 39 bus system is modeled in ePHASORSIM and simulated in OPAL-RT, real time digital simulator. The simulator is integrated with two physical relays which are connected to the remote terminal unit (RTU) inside the substation and are monitored, controlled by the control center. For attack implementation, we have captured the tripping packet going from the control center to the substation RTU using Wireshark [14] and then replayed the captured packet to theremote terminal unit (RTU) using python script to trip the relays. We have used virtual PMUs modeled inside the OPAL-RT for generating phasors at 60 samples per second. The SEL 2407, the satellite synchronized clock, is providing time synchronization to the virtual PMUs in the simulator. The virtual PMUs are sending phasors and their differential features using IEEE C37.118 protocol to the iPDC [15], the phase data concentrator, in real-time. The iPDC is saving data to the MySQL database. Initially, the generated database is used for training and building the decision tree and further performing offline testing for different cases using rattle [16]. During real-time testing, the RASc is running in the python script which pulls the data coming from the simulator to the MySQL database in terms of line flows, relays status and differential PMU features. When the line is out, it checks with the decision tree rules to identify events and performs control action by sending control signal directly to the simulator using DNP3 OPC server client protocol.

We have employed the modified IEEE 39 bus system which is divided into two major areas. The outlined Area1 is the primarily generation area which is supplying power to the rest of the system through the tie lines 15-16 (L15-16) and 16-17 (L16-17). To prevent thermal overloading on line L15-16, during the line outage L16-17, RASc sheds the generation at bus 35 and the equal amount of load is shed at bus 18 to maintain the load generation balance. We have computed the differential PMU features using bus 16 and 17 as sending and receiving end. We have created miscellaneous operating points through generation and load scaling. The generation at bus 35 is varied from 610 MW to 700 MW and load is varied from 118MW to 208 MW in equal step increase of 10 MW to maintain the generation and demand balance. For each

TABLE1. Differential PMU features

Vars	Features	Description
X1	VMs − VMr	Positive sequence voltage magnitude (VM) difference
X2	∂(VMs − VMr)/ ∂t	Rate of change of positive sequence VM difference
X3	VAs − VAr	Positive sequence voltage angle (VA) difference
X4	∂(VAs − VAr)/ ∂t	Rate of change of positive sequence VA difference
X5	IMs – IMr	Positive sequence current magnitude (IM) difference
X6	∂(IMs – IMr)/∂t	Rate of change of positive sequence IM difference
X7	(IAs – IAr)	Positive sequence current angle (IA) difference
X8	∂(IAs – IAr)/∂t	Rate of change of positive sequence IA difference

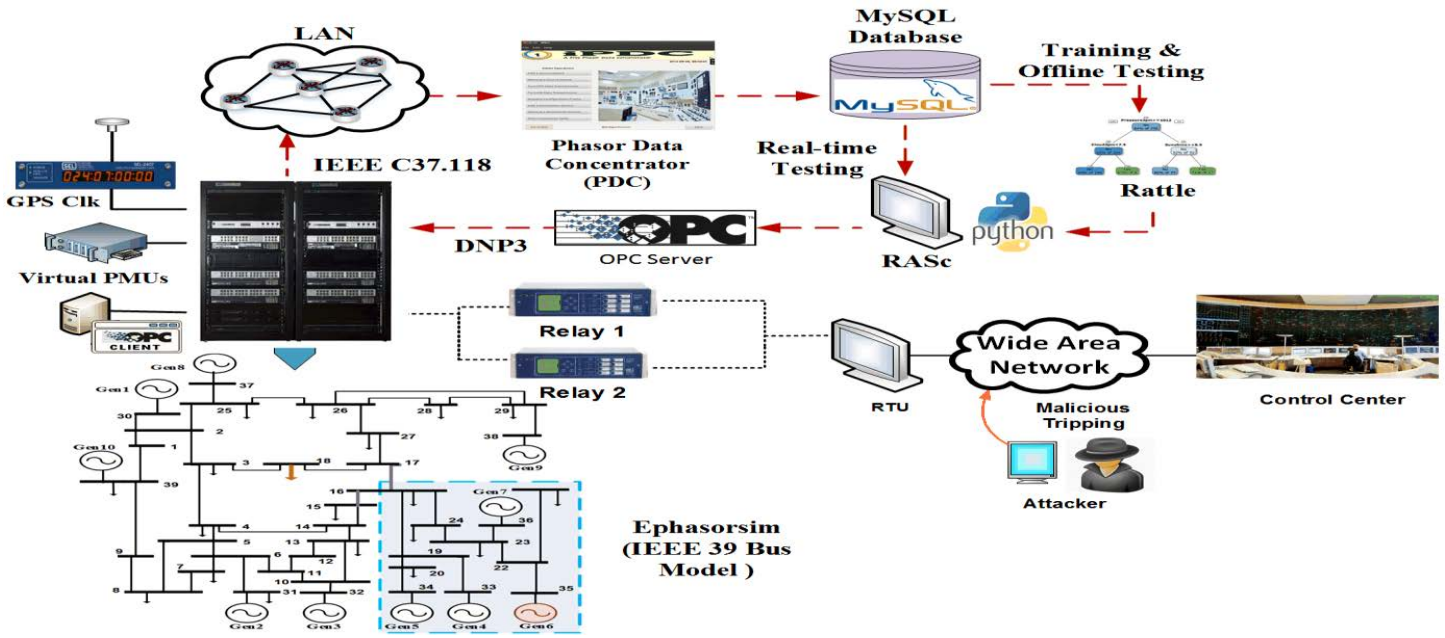

Fig. 2 Experimental setup for the proposed methodology

operating point, we have simulated a 3 phase to ground fault followed by line tripping as normal tripping event and sudden line outage as a cyber attack event at line 16-17. We have varied the duration of the fault with mean values of 6 cycles and 0.667 cycles standard deviation. The fault location distance factor is varied from 0 to 1 along the length of line with step size of 0.1, excluding the limits (0 and 1). Total number of fault cases are 50 fault durations * 9 fault locations *10 operating points = 4500. We have also simulated 10 line outages as the tripping attack, one for each operating point, and finally, total 4510 are simulated for the proposed method.

VI. RESULTS AND DISCUSSIONS

A. Offline Testing

Table 1 shows the performance of the decision tree in terms of accuracy and processing time during offline testing. It shows the effect of training datasets on the performance parameters and can be observed that 60% training of data is sufficient to obtain 100% accuracy. There is a slight increase in computational time for building the decision rules from 20% to 60% training dataset: i.e., 0.01 sec. Fig3. shows the generated decision tree rules for 60% training data sets. It can be shown that only 2 features (X1, X3) are involved in the decision process which shows the important capability of DT to select optimal features for effective decision process. The target outputs are categorized as 0 for malicious tripping and 1 for normal tripping.

B. Real-time testing

Fig. 7 shows the performance of the proposed intelligent RASc and conventional RASc during real time testing. It shows how the RAScs are identifying the events, either line faulty or malicious tripping event, and taking automated corrective actions in real-time. Fig 7 (a) and (c) show the behavior of conventional RAS, without classification model,

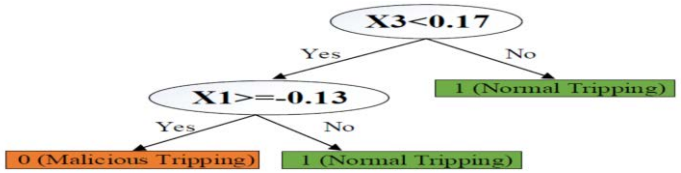

Fig. 3. Generated Decision Tree (DT) for 60% training data (case 4).

TABLE II. Training and Testing for different datasets

Cases	Training	Testing	Accuracy	Processing Time (sec)
1	20	80	98.6	0.01
2	40	60	98.4	0.01
3	50	50	99.6	0.02
4	60	40	100	0.02
5	80	20	100	0.02

during the malicious tripping attack. Initially, the power flow in line 15-16 (Line flows 15-16), as shown in Fig. 7 (c), is 2.14 pu. When the line L16-17 (Line Status 16-17) is out, as shown in Fig. 7 (b), power flow in the line L15-16 exceeds the OTC limit. We have considered the OTP limit to be 4.5 pu. It triggers the RAS generation controller (Gen. Controller 35) which sheds 90 MW (0.9 pu) of generation at bus 35 (Power generation 35) from 0.65 to 0.56 pu, and equal amount of load is shed at bus 18 (load 18) from158 to 68 MW. Fig 7 (b) and (d) shows the intelligent RAS performance during cyber attack which triggers when the line is out at 175.5 sec. It checks with the DT rules to detect the malicious tripping attack and line is reclosed in a short period at 183 sec. In this case, RASc takes the smart decision of reclosing the line instead of shedding the generation and load. Therefore, we observe a flat line for load 18 and Gen. Controller 35. Although the plot shows reclosing of line L16-17 in 7.5 secs, it can be operated in few cycles depending on

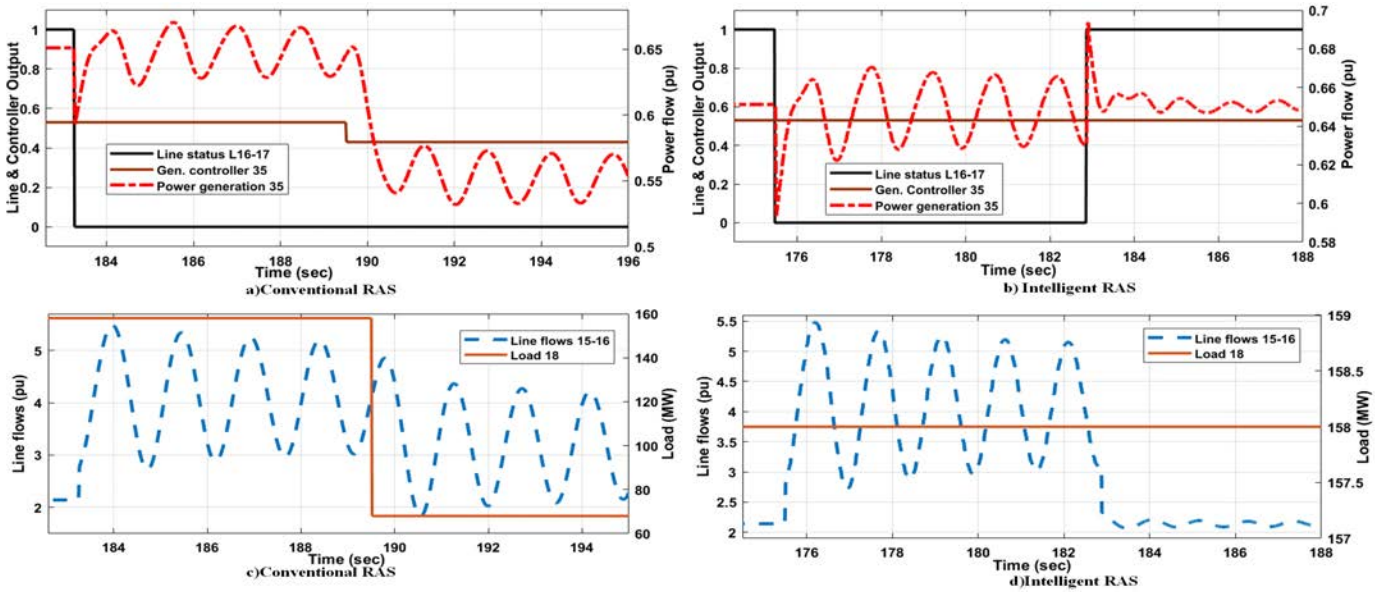

Fig. 4. Real-time testing of Conventional (a, c) and intelligent RAS (b, d).

the RASc computation capability. It is to be noted that during the line fault event, intelligent RASc is performing similar to the conventional RAS of shedding the generation and load. Due to the space limit, we have not plotted it separately.

VII. CONCLUSION

In this paper, we propose the intelligent remedial action scheme (IRAS), also known as Special Protection Scheme (SPS), for detecting the cyber attacks and physical disturbances and take corrective actions as required. We have considered the cyber attack in terms of malicious line tripping and line fault as a physical disturbance in power system. In this approach, we have computed the different features of voltage and current phasors for better and faster system observability during events. The differential features are used as inputs for training the data-mining based decision tree for different cases. We have then performed the offline and real-time testing in a cyber physical environment. We have also discussed the experimental set up for attack implementation and real-time testing of the proposed methodology on the ISU's PowerCyber testbed. It can be inferred from the preliminary experimental results that the proposed methodology provides promising solution in accurately distinguishing cyber attacks from the power system fault in the context of RAS.

For future studies, we are planning to perform more detailed analysis when system is subjected to measurement errors and communication delay during real-time testing. It also opens several avenues for further research that include developing stealthy coordinated attacks and look up possible detection and mitigation using system and network information.

VIII. REFERENCES

[1] Arizona-Southern California Outages on September 8, 2011, Causes and Recommendations, NERC and FERC, April 2012.

[2] Peter Fairley,"A December attack on Ukraine's grid was a wake-up call", IEEE Spectrum, april 20,2016.

[3] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid", Smart Grid, IEEE Transactions on, vol. 4, no. 2, pp. 847855, 2013.

[4] V. Kumar Singh, A. Ozen and M. Govindarasu, "Stealthy cyber attacks and impact analysis on wide-area protection of smart grid," *2016 North American Power Symposium (NAPS)*, Denver, CO, 2016, pp. 1-6.

[5] M. G. Adamiak *et al.*, "Wide Area Protection—Technology and Infrastructures," in *IEEE Transactions on Power Delivery*, vol. 21, no. 2, pp. 601-609, April 2006.

[6] A. Ashok, A. Hahn, and M. Govindarasu, Cyber-physical security of Wide-Area Monitoring, Protection and Control in a smart grid environment. Journal of Advanced Research, vol.5, pp.481-489, 2014.

[7] J. McCalley et al., "System Protection Schemes: Limitations, Risks, and Management " PSERC, Dec. 2010.

[8] I. Kamwa, S. R. Samantaray and G. Joos, "Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices,"in *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 144-158, Sept. 2010.

[9] K. El-Arroudi, G. Joos, I. Kamwa and D. T. McGillis, "Intelligent-Based Approach to Islanding Detection in Distributed Generation," in *IEEE Transactions on Power Delivery*, vol. 22, no. 2, pp. 828-835, April 2007.

[10] A. Samui and S. R. Samantaray, "Assessment of ROCPAD Relay for Islanding Detection in Distributed Generation," in *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 391-398, June 2011.

[11] S. R. Samantaray, K. El-Arroudi, G. Joos and I. Kamwa, "A Fuzzy Rule-Based Approach for Islanding Detection in Distributed Generation," in *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1427-1433, July 2010.

[12] A. Orebaugh, G. Ramirez, J. Beale, Wireshark and Ethereal Network Protocol Analyzer Toolkit, Rockland, MA, USA:, Feb. 2007

[13] iPDC/PMUSimulator, http://ipdc.codeplex.com

[14] Rattle (the R Analytical Tool to Learn Easily), by D.Williams: ver. 2.3, May 2008. [Online]. Available: http://rattle.togaware.com