A Comparative Study of Smart Grid Security Based on Unsupervised Learning and Load Ranking

Shuva Paul, Md. Rashedul Haq, Avijit Das, Zhen Ni
Electrical Engineering and Computer Science
South Dakota State University
Brookings, South Dakota, United States
{shuva.paul, md.rashedul.haq, avijit.das, zhen.ni}@sdstate.edu

***Abstract*—Due to the increasing number of risk factors, energy sector has been experiencing interruptions (attack) in the normal operation both externally and internally. Different methods are used for the identification and evaluation of vulnerabilities due to these interruption in the complex and critical infrastructures like the smart grid. Based on the objective of the attack, the performance and effectiveness of the learning-based approaches may vary when compared with other approaches to identify critical components of the smart grid. In this work, we adopted two target selection strategies (one is an unsupervised learning algorithm and the other is a load ranking based approach) for attack and measured system performances based on two evaluation metrics. We conducted the experiments on four different standard power system test cases and compared the performances of the aforementioned target selection strategies by two evaluation metrics. We used K-means clustering as the unsupervised learning method for the target selection of contingencies. To evaluate the system damage, we used generation loss and number of transmission line outages. For different attack orders, with two different attack objectives (evaluation metrics), experiments were conducted on $W\&W$ 6 bus system, IEEE 7 bus system, IEEE 8 bus system, and IEEE 300 bus system. We showed that, a clustering based attack performs better when the system is relatively large (highly dense in terms of connection to other buses) and the objective is to achieve a high number of transmission line outages. On the other hand, load ranking based attack outperforms clustering based attack when the attack objective is to achieve higher generation loss, regardless of the size of the system.**

***Index Terms*—K-means clustering, clustering based attack, load ranking based attack, generation loss, and smart grid security.**

I. Introduction

The application of different machine learning techniques enhanced the fragile electric power grid to include several benefits, such as enhanced security, multidimensional communication and power flow, advanced monitoring and control, and converted it into a smart grid [1]–[3]. In other words, a smart grid is the modernization of electric power transmission and distribution systems. Recently, advancement of the grid structure exposed the whole power system to severe security threats. The number of events related to power outages in the smart grid infrastructure is increasing rapidly. The nature and the number of these events in last few years clearly indicates that we need strong protection schemes to reduce these threats and damages. Almost $3,526$ power outage events were recorded in 2017 in the United States affecting almost 36.7 million people [4]. The reason behind these outage events include human error, weather related error, mechanical error, and cyber-attacks etc. Regardless of the reason, the damage caused by these events creates a severe impact on society, including huge financial losses [5]. Responsible authorities are creating different policies and standards to reduce the risk and loss in the energy sector [6], [7]. Hence, grid vulnerability analysis techniques are being used to identify the patterns of the anomalies, event signatures, critical elements of the power system, and so on. Among the learning based techniques, different supervised, unsupervised, and semi-supervised learning techniques are used in the smart grid for vulnerability identification [8]. Among these strategic endeavours, detection of the critical components of a power system bears a huge significance in protecting the energy sector.

Several research efforts have been made to identify the critical contingencies and vulnerabilities for smart grid which includes several learning based approaches [9]–[20]. In [9]–[11], the authors used reinforcement learning and game theory approaches to identify the critical transmission lines and their sequences to avoid large scale blackout. Instead of using clustering or classification techniques, they adopted a reinforcement learning based online learning technique to find the sequence of critical contingencies. The authors used a clustering-based method for unsupervised intrusion detections in [12]. The authors developed a novel incremental clustering algorithm using least distance principle to divide the data-set into hyper spheres with almost the same radius. In [13], the authors used a different cluster-based learning approach for vulnerability analysis in the Smart Grid. In [14], the authors identified the vulnerable nodes against false data injection attacks in an AMI based smart grid by implementing an improved Constriction Factor Particle Swarm Optimization (CF-PSO) based hybrid clustering technique. In [15], the authors used a clustering-based approach to detect cyber-attacks in process control systems adopting Gaussian mixture clustering. In [16], the authors used an unsupervised clustering method on PMU data for event characterization on the smart grid. The authors in [17], compared the power grid security studies with network connectivity and power flow information using unsupervised learning (self-organizing map (SOM)). In [18], the authors analyzed smart grid vulnerability by identifying the critical contingencies using time to reach blackout and number of transmission line outage as the evaluation metric. The authors in [19] analyzed the risk to power system for multi-timescale cascading outages using Markovian tree search. In [20], the authors proposed a comprehensive unsupervised clustering method (hierarchical, partitioning, and density-based approach) to classify 2226 disturbances stored in the Public Service Company of New Mexico (PNM) from 2007 to 2010. In the aforementioned literature, the authors adopted different approaches to classify the events and faults and analyze the vulnerabilities. Target selection (identification of the critical contingencies) using an unsupervised learning algorithm has rarely considered which is one of the most significant and critical factors while conducting grid vulnerability analysis. Moreover, a proper comparison between the approaches is needed to identify the appropriate approach to use for a specific attack strategy. The evaluation of the performances needs to be compared for different approaches with different evaluation metrics.

To provide a clear explanation of the approaches, we identify the vulnerable elements of a power system by proposing a comparative study of power grid vulnerability analysis between two target selection strategies, load ranking based and unsupervised learning based. In terms of unsupervised learning, we use K-means clustering [21]. The advantage of using K-means clustering over other clustering algorithm is its simplicity and faster operation even if the variable size is huge. Generation loss and the number of transmission line outages due to the attacks are used as evaluation metrics. We also conduct experiments with different orders of attack to test the feasibility of the target selection strategies. Finally, it is shown that the load ranking based attack (LRBA) strategy causes higher damage than the clustering based attack (CBA) for cases where generation loss is the evaluation index. However, for larger systems with higher transmission line outages as the attack objective, CBA outperforms LRBA. The outcome of this comparative study will help the engineering community to select appropriate target selection strategy (learning scheme or others) to identify the vulnerabilities of a CPPS.

The rest of the paper is organized as follows. Section II gives a brief introduction of the two attack strategies and the attack model. Section III provides benchmark information; how the targets are selected and the attack is executed, the evaluation metrics that are used to evaluate the damage caused by the attacks, simulation results, observations from the results, and a brief discussion on the result from a theoretical point of view. Finally Section IV, concludes with a summary of the work.

II. Attack strategies and evaluation metrics

In this section we introduce the attack model, overall process flowchart, target selection strategies used in this research work, and the evaluation metrics to evaluate the losses caused by the attacks adopting the aforementioned target selection strategies.

A. Attack model

First, we consider that the intruder/terrorists already gained the access to the control center of the smart power system via cyber-intrusion (phishing attack, DDoS attack, brute force attack, etc.). Then we move forward with the threat and attack model. The attack model is adopted from [22]–[24]. We consider that the attacker is capable of causing line switching attacks by cyber intrusion. The attack model is initialized with the pre-contingency power flow to make sure that the system is $(n-1)$ contingency secured. Then we initiate the attack by applying $(n-k)$ contingencies. After execution of the attack, the system may be separated into multiple islands, due to the application of $n-k$ contingency, where k is the order of the contingencies. To adjust the demand and supply, the generators are ramped up or down. After re-dispatching the generators, the generation, $\sum_{g\in G} P_g$ is compared with the load demand, $\sum_{d\in D} P_d$ which is defined by $Z = (\sum_{g\in G} P_g - \sum_{d\in D} P_d)$, where G and D represents the set of generator and load buses. If $Z > 0$, generators in the islands are tripped to balance the demand. After this, if $Z < 0$, the load is shed as a multiplication of the loads in that island by a scalar quantity λ, where $\lambda = \frac{\sum_{g\in G} P_g}{\sum_{d\in D} P_d}$ Then, a standard DC power flow is applied in the power system. After that, the relay settings are updated. To identify the transmission lines to be tripped due to the overcurrent/overload, time delayed overcurrent relays are used. Generally, the threshold for overcurrent is fixed by the system operator and termed as $\bar{o_j}$. During the simulation for transmission line j, if the power flow is f_j, flow limit is $\bar{f_j}$, the cascaded outage occurs when the associated overload o_j exceeds the limit $\bar{o_j}$. The overload can be calculated as follows:

$$\Delta o_j(t,\Delta t) = \begin{cases} \int_t^{t+\Delta t} (f_j(t) - \bar{f_j})dt & \text{if} f_j(t) > \bar{f_j} \\ 0 & otherwise \end{cases} \tag{1}$$

The model then finds the minimum time for tripping the next transmission line. This time is termed as ΔT. Then this time is updated with the addition of ΔT. After this, if the relay trips due to overcurrent, they will trip the associated active transmission lines to offline.

B. Overall flowchart

Figure 1, provides an algorithmic flowchart of the processes of this research work. For a selected power system test case, we start with the CBA as the target selection strategy. Then we select the number of K (order of the attack). We select the attack order with the aim to interpret high-impact, low-frequency (HILF) events. So, we start with a smallest attack order and increase gradually. Then we perform the clustering based on the geographical coordinates (topological information) of the buses. Thus, we divide all the buses of a power system into K groups of densely populated buses. Next, we conduct the power flow for the whole system, and select the highly loaded buses from the clusters as targets. Finally, the transmission lines connected with these target buses are grouped together to initiate the attack.

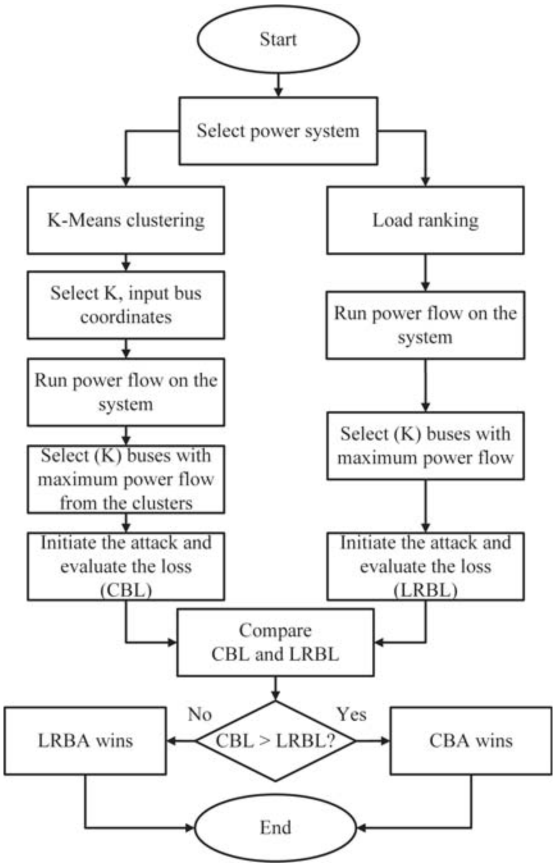
Start
Select power system
K-Means clustering
Load ranking
Select K, input bus coordinates
Run power flow on the system
Run power flow on the system
Select (K) buses with maximum power flow
Select (K) buses with maximum power flow from the clusters
Initiate the attack and evaluate the loss (CBL)
Initiate the attack and evaluate the loss (LRBL)
Compare CBL and LRBL
No
Yes
LRBA wins
CBL > LRBL?
CBA wins
End

Fig. 1. The overall flowchart to compare the two methods of target selection. CBL, represents clustering based loss, and LRBL represents load ranking based loss. CBA represents clustering based attack and LRBA represents load ranking based attack. The whole process is repeated 100 times to get an overall idea about the superior target selection strategy.

After executing the attack, we calculate the loss or evaluate the damage caused by the attack based on generation loss and number of total outages. The process of calculation of the losses and overloads are explained in detail in Section II-A. The damage or loss caused by the CBA is termed as clustering based loss (CBL). Then we select LRBA as a target selection strategy. In LRBA, we select the target buses with high loading capacity. Then we group the connected transmission lines from the target buses for initiating the attack. After executing the attack, we calculate the loss caused by the LRBA, which we termed as load ranking based loss (LRBL). After measuring the loss, we compare the losses caused by the attacks adopting the aforementioned two target selection strategies.

C. *Clustering based attack (CBA)*

To select the targets based on clustering, K-means clustering is used. K-means clustering tries to group data samples based on Euclidean distance between samples. Clustering is also called data segmentation. It gives back clusters of locations close to each other. K-means clustering is unsupervised learning, which is used when there are unlabeled data (i.e., data without defined labels). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable k. The algorithm works iteratively to assign each data point to one of k groups based on the features that are provided. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because a different location causes a different result. So, the better choice is to place them, as much as possible, far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, we may notice that the k centroids change their location step by step until no more changes occur. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{k} \|x_i^{(j)} - c_j\|^2, \tag{2}$$

where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measured between a data point $x_i^{(j)}$ and the cluster centre c_j. The algorithm is composed of the following steps:

1) Place k points into the space represented by the objects being clustered. These points represent the initial group centroids.
2) Assign each object to the group that has the closest centroid.
3) When all objects have been assigned, recalculate the positions of the k centroids.
4) Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The data points used here are the coordinates of the bus locations. The coordinates are clustered into different groups of bus locations closer to each other. From those groups, the buses with higher power flow are selected as the targets based on the attack order and the number of the clusters.

D. *Load ranking based attack (LRBA)*

In order to select the target buses based on load ranking, the amount of power flow through the buses is used. Total load flow is calculated for all the buses of a power system. The bus with higher amount of load flow is ranked higher than the buses with lower amount of load flow. Let us consider bus B_1 and B_2 when power P_1, and P_2 are flowing through them, respectively. If $P_1 > P_2$, then B_1 will be selected as the target. This is how the target buses are selected for LRBA. The transmission lines associated with the target buses are the actual targets which need to be switched in order to initiate the attack.

III. Simulation studies

The simulation is conducted using MATLAB R2018a on a standard PC with an Intel Core i7-6700 CPU running at 3.40-GHz and 24.0-GB RAM. To conduct the experiments and comparative studies between the previously mentioned target selection strategies, some of the available test cases are selected.

A. Benchmarks

To conduct the simulation studies, $W\&W$ 6 bus system, IEEE 7 bus system, IEEE 8 bus system, and IEEE 300 bus system are used. For selecting targets based on unsupervised learning, coordinate data of the nodes of the system is used; and for selecting targets based on load ranking, power flow information is used.

B. Target selection and attack execution

For the two attack methods, target selections are different. For example, while using CBA, if we use $K = 2$ for $W\&W$ 6 bus system, it divides the buses of the system into two groups. Among these two groups, buses with higher load flow are selected as targets which are 1 and 4. Similarly, clustering is executed for IEEE 7 bus system, IEEE 8 bus system, and IEEE 300 bus system. Before clustering, the bus coordinates are normalized. Normalization is one of the data pre-processing techniques used for re-scaling the data in the range of zero to one. For example, the set of bus coordinates is, $Q = [(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)]$. The normalization of the x coordinate is done using the formula below:

$$x'_t = \frac{x_t - min(x_1, x_2, x_3, \dots, x_n)}{max(x_1, x_2, x_3, \dots, x_n) - min(x_1, x_2, x_3, \dots, x_n)}$$

where, x' represents the normalized coordinates and t represents any components from $1, 2, 3, \dots, n$ for both of the x and y coordinates. y coordinate is normalized in the similar way. After normalization clustering is done based on the attack order (K).

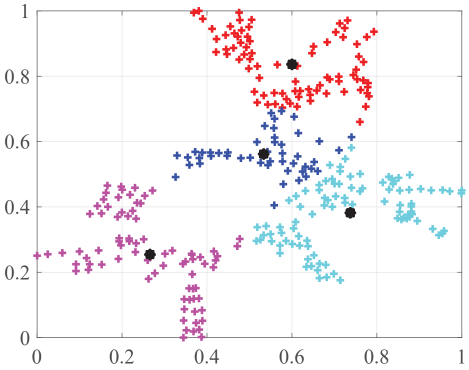

Fig. 2. K-means clustering for IEEE 300 bus system bus coordinates. The X and Y axis are representing the normalized bus coordinates. And the four different colors represent the clustered buses for four different clusters. Among these four clusters, we select four buses with higher power flow as the targets.

Figure 2 shows the clustering of the target buses using K-means clustering for IEEE 300 bus system. There are four clusters in this clustering. From these four clusters (groups of buses), targets with high loading capability are selected. So, from four clusters, four target buses are selected. Now, to execute the attacks on these targets, we simply disconnect the transmission lines connected to these buses. Similarly, for the LRBA, the mostly loaded buses are bus 1 and 2. So, we select these buses as the targets for LRBA. To execute the attack we disconnect the transmission lines connected to the target buses. Then we observe the damage to the system using two different evaluation metrics. The same method is applied to select the targets for the other bus systems as well.

C. Evaluation metrics

To evaluate the damage caused by the above mentioned two attack strategies, we employed two different evaluation indices. We used the total generation loss and the number of total outages (both the attacked lines and the cascaded outages combined) to measure the success of the attack. Generation loss, as evaluation metric, provides insight into how the power system is losing its generation power either due to load shedding or due to generation and demand imbalance. The number of total outages refers to the number of transmission line failures caused by the attack and as a consequence of cascading failure. It provides insight about damage in the power system from the topological perspective. With higher number of total outages, more area will be deprived of power. Subsection II-A, provides the process of calculation of generation loss and the number of total outages due to the attacks.

D. Simulation results

We use 3 different attack orders $(2, 3, \text{and } 4)$. So we use the clustering for three different values of K. For three different values of K, we conduct the experiment for four different power system test cases. Each of the experiments with unique settings is repeated 100 times to check the performance of these two attack strategies.

TABLE I
Comparison of the attack performances for attack order 2 and total transmission line outage as the evaluation index

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	2.13
IEEE 7 bus system	0	100	2.20
IEEE 8 bus system	0	100	2.32
IEEE 300 bus system	0	100	30.71

TABLE II
Comparison of the attack performances for attack order 3 and total transmission line outage as the evaluation index

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	1.68
IEEE 7 bus system	11	89	1.84
IEEE 8 bus system	11	89	1.76
IEEE 300 bus system	14	86	34.69

Table I - VI show the number of times where CBA outperforms LRBA and vice versa.

TABLE III
COMPARISON OF THE ATTACK PERFORMANCES FOR ATTACK ORDER 4 AND TOTAL TRANSMISSION LINE OUTAGE AS THE EVALUATION INDEX

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	1.75
IEEE 7 bus system	16	84	1.93
IEEE 8 bus system	30	70	1.85
IEEE 300 bus system	63	37	35.44

TABLE IV
COMPARISON OF THE ATTACK PERFORMANCES FOR ATTACK ORDER 2 AND TOTAL GENERATION LOSS AS THE EVALUATION INDEX

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	1.69
IEEE 7 bus system	0	100	1.81
IEEE 8 bus system	0	100	1.78
IEEE 300 bus system	0	100	27.08

TABLE V
COMPARISON OF THE ATTACK PERFORMANCES FOR ATTACK ORDER 3 AND TOTAL GENERATION LOSS AS THE EVALUATION INDEX

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	1.80
IEEE 7 bus system	0	100	1.76
IEEE 8 bus system	0	100	1.93
IEEE 300 bus system	10	90	34.80

TABLE VI
COMPARISON OF THE ATTACK PERFORMANCES FOR ATTACK ORDER 4 AND TOTAL GENERATION LOSS AS THE EVALUATION INDEX

Test cases	CBA	LRBA	Total time (seconds)
$W\&W$ 6 bus system	0	100	1.64
IEEE 7 bus system	0	100	1.84
IEEE 8 bus system	0	100	1.82
IEEE 300 bus system	16	84	32.29

E. Observation

From Table I, we can see that for the second order attack, for all the test cases, LRBA outperforms CBA when the number of total transmission line outages is the evaluation index. But, as we increase the attack order from $K = 2$ to $K = 4$, we can see the number of times where the CBA outperforms the LRBA increases. For the IEEE 300 bus system and attack order 4, CBA outperforms LRBA most of the time. Table VII, shows the target buses for 3^{rd} and 4^{th} order attacks for both CBA and LRBA. It also provides the number of total transmission line outages associated with the target buses and attack orders. From Table VII, we can also see that, for $K = 3$, the number of total transmission line outages is 32 for CBA with the target set of [3 170 98]. The number of total transmission line outages is 55 for LRBA with the target set of [98 109 170]. However, when we increase the attack order to $K = 4$, the number of total transmission line outages is 63 for CBA with the target set of [170 36 166 98]. On the other hand, for LRBA, the number of total transmission line outages is 43 with the target set of [98 109 170 3]. Therefore, we can conclude from Table VII that, for higher order attacks and for larger systems CBA causes more damage if the target is to cause a higher number of transmission line outages.

TABLE VII
TOTAL NUMBER OF TRANSMISSION LINE OUTAGE DUE TO DIFFERENT ATTACK ORDER FOR CBA AND LRBA ON IEEE 300 BUS SYSTEM.

Attack order	Attack strategy	Target buses	Number of total outages
$K = 3$	CBA	[3 170 98]	32
	LRBA	[98 109 170]	55
$K = 4$	CBA	[170 36 166 98]	63
	LRBA	[98 109 170 3]	43

Similarly, from Table IV, V, and VI, we can see that for all the test cases LRBA outperforms CBA when the generation loss is the evaluation index. Occasionally CBA outperforms the LRBA, but this is only for the larger systems, like IEEE 300 bus system.

TABLE VIII
GENERATION LOSS DUE TO DIFFERENT ATTACK ORDER FOR CBA AND LRBA ON IEEE 300 BUS SYSTEM.

Attack order	Attack strategy	Target buses	Generation loss (MW)
$K = 3$	CBA	[3 170 98]	5867.20
	LRBA	[98 109 170]	12029
$K = 4$	CBA	[98 3 166 170]	6864.55
	LRBA	[98 109 170 3]	9907.61

Table VIII shows the target buses for 3^{rd} and 4^{th} order attacks for both CBA and LRBA when generation loss is considered as the evaluation metric. It also provides the amount of generation loss for both attack orders and both attack types. Although we increase the attack order to $K = 4$, the generation loss is still higher for LRBA than CBA. So, we can conclude from Table VIII that if the target is to achieve higher generation loss, regardless of the system size, LRBA outperforms the CBA. These observations can be supported by providing the explanation from a theoretical point of view. In CBA, the clustering is done based on the topological connection. The targets selected from the clusters are densely populated buses having a high order of connectivity. In that case, after attacking (disconnecting) these targets, a large number of transmission line outages occurs. But, in small power system test cases, observation showed that LRBA performs well compared to CBA. This is because, for small systems, like the W & W 6 bus system, IEEE 7 bus, and IEEE 8 bus system, the topological connections between the buses are not highly dense. So CBA does not perform accurate target selection in small test cases. That is why, in case of LRBA in small power system test cases, the selected targets carry both the properties to cause high generation losses and high transmission line outages. So, LRBA performs better than the CBA for small systems with transmission line outage as attack objective. On the other hand, according to the observations, if the generation loss is considered as the attack objective, LRBA performs better than CBA. This is because, when selecting targets for the LRBA, buses with higher loading capability are selected as targets. As these target selections do not depend on the topological density, LRBA performs better than CBA with generation loss as the attack objective.

From a machine learning point of view, most feature-based clustering methods (e.g., K-means, GMM, etc) fail to scale high-dimensional data well due to the curse of dimensionality. For high dimensional data, considering compact and representative features is more reasonable and feasible instead of the whole feature space [25]. Different deep learning techniques learns a significant and strong representation from the raw data using high-level non-linear mapping [26]. Uses of these deep learning techniques for clustering might add several benefits for clustering, improving the efficiency and performance.

IV. Conclusion

The use of machine learning techniques is becoming popular for the identification of vulnerabilities in the modern power system along with the increasing complexity of the infrastructures. Smart attackers possess different attack schemes for different attack objectives. Proper use of machine learning techniques and understanding the vulnerabilities of the power grid from the attacker's perspective could help the system operators or utilities to make stronger protection plans for the vulnerable elements. This research provides insight into appropriate target selection strategies for different system types and different attack objectives. It concludes that, for larger systems with higher attack order, if the attack objective is to cause higher transmission line outage, CBA outperforms LRBA scheme. If the attack objective is to cause higher generation loss, then LRBA scheme outperforms CBA scheme regardless of the system size.

Acknowledgement

This work is partially supported by the National Science Foundation under grant #OIA-1833005 and the South Dakota Board of Regents Competitive Research Program FY2018.

References

[1] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, vol. 11, pp. 1644–1652, Sep. 2017.

[2] D. S. Terzi, B. Arslan, and S. Sagiroglu, "Smart grid security evaluation with a big data use case," in *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, pp. 1–6, April 2018.

[3] Z. Zheng, Y. Yang, X. Niu, H. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 1606–1615, April 2018.

[4] *Blackout Tracker: United States Annual Report 2017*, 2017 (accessed November 2, 2018). Available at: http://electricalsector.eaton.com/forms/BlackoutTrackerAnnualReport.

[5] *Cost of Hourly Downtime Soards: 81% of Enterprises Say it Exceeds $300K On Average*, August 2016 (accessed November 2, 2018). Available at: https://itic-corp.com/blog/2016/08/cost-of-hourly-downtime-soars-81-of-enterprises-say-it-exceeds-300k-on-average/.

[6] J. A. Harer, L. Y. Kim, R. L. Russell, O. Ozdemir, L. R. Kosta, A. Rangamani, L. H. Hamilton, G. I. Centeno, J. R. Key, P. M. Ellingwood, M. W. McConley, J. M. Opper, S. P. Chin, and T. Lazovich, "Automated software vulnerability detection with machine learning," *CoRR*, vol. abs/1803.04497, 2018.

[7] M. Mylrea, S. N. G. Gourisetti, R. Bishop, and M. Johnson, "Keyless signature blockchain infrastructure: Facilitating nerc cip compliance and responding to evolving cyber threats and vulnerabilities to energy infrastructure," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, pp. 1–9, April 2018.

[8] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, pp. 1773–1786, Aug 2016.

[9] Z. Ni and S. Paul, "A multi-stage game in smart grid security: A reinforcement learning solution," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2018. doi: 10.1109/TNNLS.2018.2885530.

[10] S. Paul and Z. Ni, "A study of linear programming and reinforcement learning for one-shot game in smart grid security," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2018.

[11] Z. Ni, S. Paul, X. Zhong, and Q. Wei, "A reinforcement learning approach for sequential decision-making process of attacks in smart grid," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, Nov 2017.

[12] S. Jiang, X. Song, H. Wang, J.-J. Han, and Q.-H. Li, "A clustering-based method for unsupervised intrusion detections," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 802 – 810, 2006.

[13] E. D. Santis, A. Rizzi, and A. Sadeghian, "A cluster-based dissimilarity learning approach for localized fault classification in smart grids," *Swarm and Evolutionary Computation*, vol. 39, pp. 267 – 278, 2018.

[14] A. Anwar, A. N. Mahmood, and Z. Tari, "Identification of vulnerable node clusters against false data injection attack in an ami based smart grid," *Information Systems*, vol. 53, pp. 201 – 212, 2015.

[15] I. Kiss, B. Genge, and P. Haller, "A clustering-based approach to detect cyber attacks in process control systems," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, pp. 142–148, July 2015.

[16] E. Klinginsmith, R. Barella, X. Zhao, and S. Wallace, "Unsupervised clustering on pmu data for event characterization on smart grid," in *2016 5th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS)*, pp. 1–8, April 2016.

[17] S. Poudel, Z. Ni, X. Zhong, and H. He, "Comparative studies of power grid security with network connectivity and power flow information using unsupervised learning," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2730–2737, July 2016.

[18] S. Paul and Z. Ni, "Vulnerability analysis for simultaneous attack in smart grid security," in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, April 2017.

[19] R. Yao, S. Huang, K. Sun, F. Liu, X. Zhang, S. Mei, W. Wei, and L. Ding, "Risk assessment of multi-timescale cascading outages based on markovian tree search," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2887–2900, 2017.

[20] O. P. Dahal, S. M. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Transactions on Power Delivery*, vol. 29, pp. 1390–1397, June 2014.

[21] A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data," *Data & Knowledge Engineering*, vol. 63, no. 2, pp. 503 – 527, 2007.

[22] M. J. Eppstein and P. D. H. Hines, "A random chemistry algorithm for identifying collections of multiple contingencies that initiate cascading failure," *IEEE Transactions on Power Systems*, vol. 27, pp. 1698–1705, Aug 2012.

[23] P. Rezaei, P. D. H. Hines, and M. J. Eppstein, "Estimating cascading failure risk with random chemistry," *IEEE Transactions on Power Systems*, vol. 30, pp. 2726–2735, Sep. 2015.

[24] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 3, p. 033122, 2010.

[25] K. Tian, S. Zhou, and J. Guan, "Deepcluster: A general clustering framework based on deep learning," in *Machine Learning and Knowledge Discovery in Databases* (M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, eds.), (Cham), pp. 809–825, Springer International Publishing, 2017.

[26] E. Aljalbout, V. Golkov, Y. Siddiqui, and D. Cremers, "Clustering with deep learning: Taxonomy and new methods," *CoRR*, vol. abs/1801.07648, 2018.