DATA MINING AND FACTS BASED TRANSMISSION LINE PROTECTION

K.S.Panpaliya[1], V.V.Bais[2], S.D.Nimkar[3]

[1] Assistant Professor, Electrical & Electronics Engineering Department, Prof Ram Meghe College of Engineering & Management, Badnera-Amravati, Maharashtra, India
[2,3] Assistant Professor, Innovation & Entrepreneurship Development Center-Electrical & Electronics Engineering Department, Prof Ram Meghe College of Engineering & Management, Badnera-Amravati, Maharashtra, India

ABSTRACT

An electrical power network, as a whole, consists of generation, transmission and distribution. The performance of a power network is frequently affected by the transmission line faults, which give rise to disruption in power flow. Therefore, transmission of electric power and necessary protective measures are the vital issues need to be addressed properly. In demand of electric power and addition of new generation capacity to meet the demand, necessitate enhancement of large transmission capacity between generation and bulk consumption points. This can be achieved either by development of new transmission corridor or by enhancing the power transfer intensity of existing transmission assets Distance protection is used to protect the transmission line against faults by measuring the line voltages and currents at remote end buses using digital fault recorders.

Faults on transmission lines need to be detected, classified, located accurately, and cleared as fast as possible. In power transmission line protection, faulty phase identification and location of fault are the two most important items which need to be addressed in a reliable and accurate manner. Distance relaying techniques based on the measurement of the impedance at the fundamental frequency between the fault location and the relaying point have attracted wide spread attention. The sampled voltage and current data at the relying point are used to locate and classify the fault involving the line with or without fault resistance present in the fault path. In this paper an adaptive approach for FACTS based transmission line protection is proposed and analyzed for different scenario.

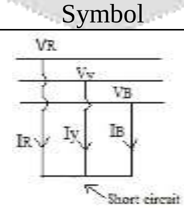
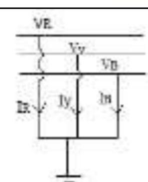
Keyword: - *Data mining distance protection, FACTS, fault identification, transmission line protection*.

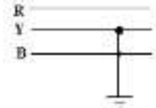
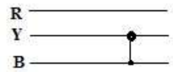
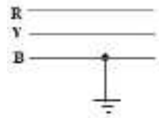
---

1. INTRODUCTION

A fault occurs when two or more conductors that normally operate with a potential difference come in contact with each other. These faults may be caused by sudden failure of a piece of equipment, accidental damage or short-circuit to overhead lines or by insulation failure resulting from lightning surges. Irrespective of the causes, the faults in a 3-phase system can be classified into two main types *viz*. also shown in Table 1.

Table 1:- Types of fault

Nature of fault	Types of fault	Symbol	Effect	% Occurrence
Symmetrical fault	3 PHASE (L-L-L)		Very severe	2-5%
Symmetrical fault	3 PHASE TO GROUND (L-L-L-G)		Very severe	2-5%

Unsymmetrical fault	Double line to ground (L-L-G)		Severe	5-10%
Unsymmetrical fault	Line to Line (L-L)		Less severe	10-15%
Unsymmetrical fault	Line to Ground (L-G)		Very Less severe	75-80 %

- The power quality of transmission line is adversely affected, Failure of industrial loads, due to drop in the voltage of healthy feeders.
- Heating of rotating machines may occur due to unbalancing of currents and supply voltages arising due to short circuit.
- Loss in system stability.
- Affect the transient and steady stare stability
- Severe short circuit current may occur in the system due to fault which may prove fatal to the several equipment's of the power system and lead to the overheating of the system.
- Heavy current is also the reason behind the setting up of very high mechanical stresses
- Continuity of power supply is adversely affected.
- Frequency of power supply is adversely affected.
- Efficiency of transmission line is adversely affected [1].

2. RELATED WORK

A Flexible AC transmission system (FACTS) [2] is an evolving technology to help electric utilities. Its first concept was introduced by N.G Hingorani, in 1988. A flexible alternating current transmission system (FACTS) is a system composed of static equipment used for the AC transmission of electrical energy. It is meant to enhance controllability and increase power transfer capability of the network. The solutions to improve the quality of supply in the electrical networks with go through the applications of the developments in semiconductor power devices, that is to say, the utilization of static power converters in electrical energy networks. The technological advances in power semiconductors are permitting the development of devices that react more like an ideal switch, totally controllable, admitting high frequencies of commutation to major levels of tension and power [3].

In recent years, power demand has increased substantially while the expansion of power generation and transmission has been severely limited due to limited resources and environmental restrictions. Recent development of power electronics introduces the use of FACTS controllers in power systems. FACTS controllers are capable of controlling the network condition in a very fast manner and this feature of FACTS can be exploited to improve the voltage stability, and steady state and transient stabilities of a complex power system. This allows increased utilization of existing network closer to its thermal loading capacity, and thus avoiding the need to construct new transmission lines. The well known FACTS devices are namely SVC, STATCOM, TCSC, and SSSC .Among them Thyristor controlled series compensator. (TCSC) are extensively used for improving the utilization of existing transmission system. TCSC module consists of basically, a series capacitor (C), in parallel with a Thyristor Controlled Reactor (TCR) (Ls). A metal oxide varistor (MOV), essentially a nonlinear resistor, is connected across the series capacitor to prevent the occurrence of high capacitor over voltages. The amount of series compensation in the capacitive operating region is increased (or decreased) by varying the thyristor firing angle α [4].

Transient stability control plays a significant role in ensuring the stable operation of power systems in the event of large disturbances and faults, and is thus a significant area of research. Flexible AC transmission systems (FACTS) controllers have been mainly used for solving various power system steady state control problems. FACTS devices are capable of controlling the active and reactive power flows in a transmission line by controlling its series and shunt parameters [5]. It can have various roles in the operation and control of power systems, such as: Decreasing unsymmetrical components, Providing voltage support, Mitigating subsynchronous resonance (SSR), Damping the power oscillation, Power flow control, Increase of transmission capability, Reactive power compensation, Transient Stability improvement, Power quality improvement, Power conditioning, Rapid continuous control of the transmission line reactance, and Reduce transmission losses.

Advances in high-power, high-efficiency power electronics have led to the development of Thyristor-controlled series compensators in power systems. In contrast to capacitors switched by circuit breakers, TCSC will be more effective because thyristors can offer flexible adjustment, and more advanced control theories can be easily applied. Series capacitor is used in a long distance EHV lines for increasing power transfer. Use of series capacitor is for the most economic enhancing power flow though series have a problem of SSR. Series is only is used for power transfer as compared to shunt. Shunt has the main problem of location not of SSR. To provide variable series compensation thyristor control is used due to which the major problem of SSR is been reduced to much lesser extent.

Transmission lines are very much prone to failures and faults. It is very difficult to fix the fault manually when it occurs. There are many methods available in the past in order for detection of fault diagnosing application. The growth of available data in the electric power industry motivates the adoption of data mining techniques. The companies in this scope face several difficulties that they could not benefit themselves from data mining approach. One of the reasons is that mining power systems data is an Interdisciplinary task. Typically, electrical and computer engineers or scientists need to work together in order to achieve breakthroughs, interfacing power systems and data mining at a mature level of cooperation. Another reason is the lack of freely available and standardized benchmarks. Because of that, most previous research in this area used proprietary datasets, which makes difficult to compare algorithms and reproduce results [6].

In recent years, because of energy, environmental, and regulatory concerns, the growth of electric power transmission facilities has been restricted. Transmission lines can be compensated by Thyristor Controlled Series Compensator (TCSC) to increase power transfer capability, improve transient stability, reduce transmission losses, and dampen power system oscillations.

However, the employment of series compensation creates certain problems for its protective relays and fault locators using conventional techniques because of the rapid changes introduced by the associated TCSC control actions in primary system parameters such as line impedances and load currents. The most important singularity lays in the fact that the positive sequence impedance measured by traditional distance relays is no longer an indicator of the distance to a fault. The apparent reactance and resistance seen by the relay are affected due to the uncertain variation of series compensation voltage during the fault period.

The transition from normal operation to other possible modes does not occur instantly. The transition time is considerable in the time frame of the protection of EHV lines. Hence, ignoring the dynamics of TCSC and modeling it as constant impedance during the fault normally leads to erroneous results that are not in conformity with real systems. For a comprehensive analysis of the impact of TCSC on protection of the lines, it is necessary to consider both aspects of the TCSC operation: (1) dynamics of TCSC; and (2) contribution of the TCSC to power system operation and control without ignoring their linkage. The present of TCSC dynamics have a significant impact on the power system protection and its transition from a mode to another can create serious problems for the conventional relays like forward overreach, reverse overreach, miss-coordination in primary and back-up protection, directional malfunction and adverse effect on distance schemes [5].

The presence of TCSC in fault loop not only affects the steady-state components but also the transient components [1]. The controllable reactance, the metal–oxide varistors (MOVs) protecting the capacitors and the air-gap operation make the protection decision more complex and, therefore, the conventional relaying scheme based on fixed settings finds limitations. In the FACTS-based transmission line, if the fault does not include FACTS device, then the impedance calculation is like an ordinary transmission line, and when the fault includes FACTS, then the impedance calculation accounts for the impedances introduced by FACTS device. The line impedance is compared with the protective zone and if the line impedance is less than the relay setting, then the relay issues a signal to trip the circuit breaker (CB). Further, for similar types of faults, the current level may be of the same order at two different points of the transmission line, (before and after TCSC). Thus, before the apparent impedance to the fault point is computed, a more reliable and accurate fault-zone identification and fault classification technique is necessary for safe and reliable operation of the distance relay. Thus the correct fault-zone identification and fault classification in presence of the FACTS devices, such as TCSC in the transmission line, is one of the critical tasks to be dealt with.

This paper presents a new adaptive approach for FACTS based transmission line protection using ensemble decision tree as decision tool. The random forest based zone identification algorithm takes one cycle post fault current and voltage samples from the fault inception are used as input vectors and gives target output '0' for fault after TCSC and '1' for fault before TCSC for fault zone identification. Similarly, the random forest based classification algorithm takes one cycle data from fault inception of three phase currents along with zero-sequence current and voltage, and constructs the random forest decision tree for classifying all ten types of shunt faults in the transmission line fault process. The results indicate that the proposed method can reliably identify the fault zone and classify faults in the FACTs-based transmission line in large power network.

Extensive simulation studies are carried out using the power system simulation toolbox software MATLAB to evaluate the performance of the proposed algorithm under different conditions. The Thyristor controlled series compensator (TCSC), current transformers (CTs) and capacitor voltage transformers (CVTs) are also modeled precisely. The obtained results confirm that the proposed adaptive approach provides a satisfactory and reliable Protection of FACTS based transmission line.

3. PROPOSED PROTECTION SCHEME

The proposed research work includes the FACTS-based transmission line, if the fault does not include FACTS device, then the impedance calculation is like an ordinary transmission line, and when the fault includes FACTS, then the impedance calculation accounts for the impedances introduced by FACTS device. The line impedance is compared with the protective zone and if the line impedance is less than the relay setting, then the relay issues a signal to trip the circuit breaker (CB). The pre-processing of the fault current and voltage signal samples through the sampler to extract the features of the fault current and voltage phasor information such as amplitude and phase, which are used for impedance calculation to the fault point and for fault zone identification and shunt fault classification using random forest.

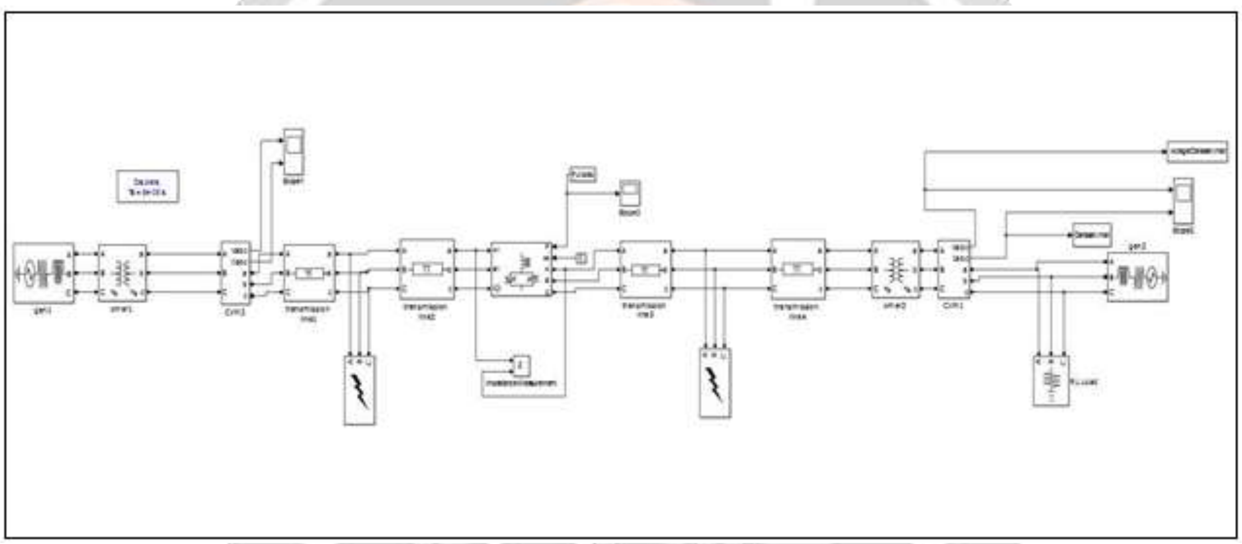


Fig.1:- MATALB Simulation model of proposed approach.

In the Proposed adaptive approached of FACTS-based transmission line, if the fault does not include FACTS device, then the impedance calculation is like an ordinary transmission line, and when the fault includes FACTS, then the impedance calculation accounts for the impedances introduced by FACTS device. The thyristor controlled series compensator (TCSC) is use as FACT device in proposed report. The TCSC is use to maintain the voltage across series reactance i.e impedance of transmission line by varying the firing angel. The line impedance is compared with the protective zone and if the line impedance is less than the relay setting, then the relay issues a signal to trip the circuit breaker (CB). Further, for similar types of faults, the current level may be of the same order at two different points of the transmission line, an accurate and faster decision making tool i.e random forest or ensemble decision tree is use for fault-zone identification in FACTS-based transmission lines and for fault classification in FACTS based transmission line. Since, the protection of FACTS based transmission line is carried out. The proposed method is based on random forest/ ensemble decision tree [7–10] for fault zone identification and fault classification for the TCSC-based line.

Consider that fault is generated after the location of Thyristor controlled series compensator in FACTS based transmission line. so that line impedance is varied which is compare with predetermined impedance relay setting and if it is less then gives signal to circuit breaker to trip. And circuit breaker is isolates the fault part of transmission system. thus before the impedance relay give signal to circuit breaker the one cycle post fault current and fault voltage are measured at impedance relay The one cycle post fault current and voltage are sample at 1khz into ten sample of post fault current i.e. ia=[ia0, ia1, ia2, . . . , ia09], ib = [ib0, ib1, ib2, . . . , ib09], ic = [ic0, ic1,ic2, . . . , ic09], and ten sample of post fault voltage i.e. va = [va0, va1, va2, . . . , va09], vb = [vb0,vb1, vb2, . . . , vb09], vc = [vc0, vc1, vc2, . . . , vc09].. As the sampling frequency is 1.0 kHz, one cycle contains 10 samples. Thus the input vector (one set) contains 60 data points for one target output. These samples are used as input to the random forest decision tool. Random forest decision tree are trained using this fault voltage and current and gives decision against target outputs '1' for faults before the TCSC and '0' for faults after the TCSC.

Consider that shunt fault is generated after the location of Thyristor controlled series compensator in FACTS based transmission line. so that line impedance is varied which is compare with predetermined impedance relay setting and if it is less then gives signal to circuit breaker to trip. And circuit breaker is isolates the fault part of transmission system. thus before the impedance relay give signal to circuit breaker the one cycle post fault current , one cycle zero sequence post fault current and fault voltage are measured at impedance relay The one cycle post fault current and voltage are sample at 1khz into ten sample of post fault current i.e. ia=[ia0, ia1, ia2, . . . , ia09], ib = [ib0, ib1, ib2, . . . , ib09], ic = [ic0, ic1,ic2, . . . , ic09], one cycle zero sequence post fault current into ten sample of io = [io1, io2, io3, . . . , io09], and ten sample of post fault voltage i.e. va = [va0, va1, va2, . . . , va09], vb = [vb0,vb1, vb2, . . . , vb09], vc = [vc0, vc1, vc2, . . . , vc09].. As the sampling frequency is 1.0 kHz, one cycle contains 10 samples. Thus the input vector (one set) contains 70 data points for one target output. These samples are used as input to the random forest decision tool. Random forest decision tree are trained using this fault voltage and current considering various parameter and gives decision against target outputs '1' for classifying various types of shunt fault like (A–G), (B–G), (C–G), (AB–G), (B-C–G), (C-A–G), and (A–B–C-G) as target output. The fault simulations are carried out with various operating conditions. As 70–30% training and testing data set is the generalized one for better classification accuracy.

The proposed system simulation model shown in figure 1 consists of various components with following specifications:

Generators 1, 2: Phase to phase rms voltage = 400e3 V; Frequency = 50 Hz; 3 phase short circuit level = 100e6 V; Base voltage = 400e3 V; X/R=7

Transformer 1: Power= 100e6 VA, Primary voltage = 400e3 V; Secondary voltage = 220e3 V; Frequency = 50Hz

Transformer 2: Power= 100e6 VA Primary voltage = 400e3 V; Secondary voltage = 220e3 V; Frequency = 50Hz

Load: Phase to phase voltage = 400 V; Frequency = 50 Hz; Active power = 10e3 W; Inductive reactive power = 3.5 Positive VAR; Capacitive reactive power = 2.4 Negative VAR; **Transmission line model:** Phase to phase voltage 400 KV; Frequency = 50 Hz; Zone1 (AB) = 150Km; Zone2 (BC) = 150Km;

The following steps are involved for simulation of the proposed fault identification transmission line protections simulation model.

1. Load simulation model in MATLAB.
2. Decide at what time the faults add (start and end time).
3. Load the fault_input file which contains: Total simulation time, Start time and End time
4. Simulate the main simulation model.
5. Simulation finds fault output voltage and current and stored it in dataset file.
6. Load analysis.m file which extract the features of the fault output voltage and current using database file.
7. Features of output fault voltage and current sample are given to Random Forest.
8. Random Forest identify the fault zone & classify the shunt faults.

The random forest is an ensemble approach that can also be thought of as a form of nearest neighbor predictor. Ensembles are a divide-and-conquer approach used to improve performance. The main principle behind ensemble methods is that a group of "weak learners" can come together to form a "strong learner". The random forest starts with a standard machine learning technique called a "decision tree" which, in ensemble terms, corresponds to our weak learner [12]. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. The random forest takes this notion to the next level by combining trees with the notion of an ensemble. Thus, in ensemble terms, the trees are weak learners and the random forest is a strong learner. Consider a system is trained; for some number of trees *T*:

1. Sample *N* cases at random with replacement to create a subset of the data. The subset should be about 66% of the total set. 2 at each node:

- For some number *m*, *m* predictor variables are selected at random from all the predictor variables.
- The predictor variable that provides the best split, according to some objective function, is used to do a binary split on that node.
- At the next node, choose another *m* variable at random from all predictor variables and do the same.

When a new input is entered into the system, it is run down all of the trees. The result may either be an average or weighted average of all of the terminal nodes that are reached, or, in the case of categorical variables, a voting majority.

- With a large number of predictors, the eligible predictor set will be quite different from node to node.
- The greater the inter-tree correlation, the greater the random forest error rate, so one pressure on the model is to have the trees as uncorrelated as possible.
- As *m* goes down, both inter-tree correlation and the strength of individual trees go down. So some optimal value of *m* must be discovered. Random forest runtimes are quite fast, and they are able to deal with unbalanced and missing data. Random Forest weaknesses are that when used for regression they cannot predict beyond the range in the training data, and that they may over-fit data sets that are particularly noisy [13]. The best test of any algorithm is how well it works upon your own data set.

Total 500 fault condition out of which 100 cases are used to train random forest decision trees considering following parameter.

1. Variations in fault resistance = 0 to200 Ω.
2. Variations in fault location 10%, 30%, 55%, and 90% of the line;
3. Different types of fault= 7 (AG, BG, CG, ABG, BCG, ACG, ABCG,)
4. Variations in TCSC firing angle from 180° (minimum compensation) to 150° (maximum compensation).
5. Reverse power flow. 6. Sudden load change.

A 400kv 50 Hz TCSC based power system simulation model is consider in this proposed technique. In this system TCSC is located at mid-point if transmission line. The power system simulation model consists of two source, associated component and 300 km transmission line. Total 100 simulations are carried out on TCSC based transmission line model considering various random forests training parameter. This simulation process is done to acquire the information of training pattern on testing data sets. The Random forest output for fault zone identification and shunt faults classification for protection of TCSC based transmission line are discussed below for various fault condition.

Consider the total simulation time of main model is 1 sec. When the fault is occurred at 150 km of total 300 km transmission line. The start time of fault is 0.5 sec and end time of fault is .80 sec for fault resistance Rf=200Ω. The random forest analysis the features of fault voltage and fault current and gives the decision that phase A-G fault occurred after TCSC. An extensive data set is generated considering various parameters to train and test the RF-based decision tree for developing an accurate and robust classifier for faultzone identification and different shunt fault classification in FACTS-based transmission line. When fault occurred at 150 km having Rf=200Ω. The variation in fault output voltage and fault current during phase 'A-G fault' are also shown in figure 2. From figure 2 we observed that there is large disturbance in voltage and current sample of phase A during the period of .5 sec to .80 sec.

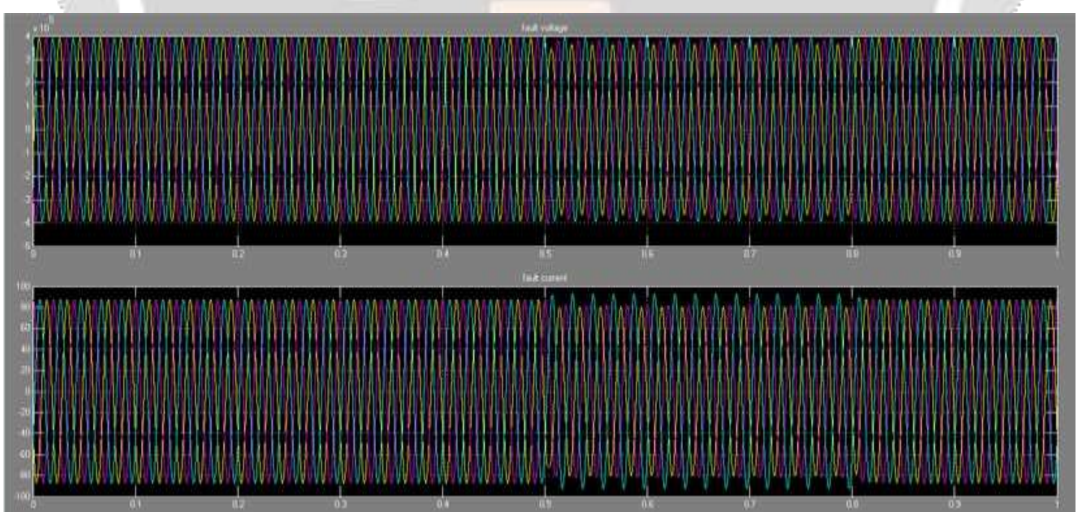


Fig 2:- Variation in fault output voltage and fault current during phase 'A-G fault'.

When the fault is occurred at 150 km of total 300km transmission line. The start time of fault is 0.5 sec and end time of fault is .80 sec for fault resistance Rf=150Ω. The random forest analysis the features of fault voltage and fault current from data base and gives the decision that fault on phase B-G occurred after TCSC. An extensive data set is generated considering various parameters to train and test the RF-based decision tree for developing an accurate and robust classifier for fault-zone identification and different shunt fault classification in FACTS-based transmission line. When fault occurred at 150 km having Rf=150Ω. The variation in fault output voltage and fault current during phase 'B-G fault' can be observed. From the figure 3 we observed that there is large disturbance in voltage and current sample of phase B during the period of .5 sec to .80 sec.

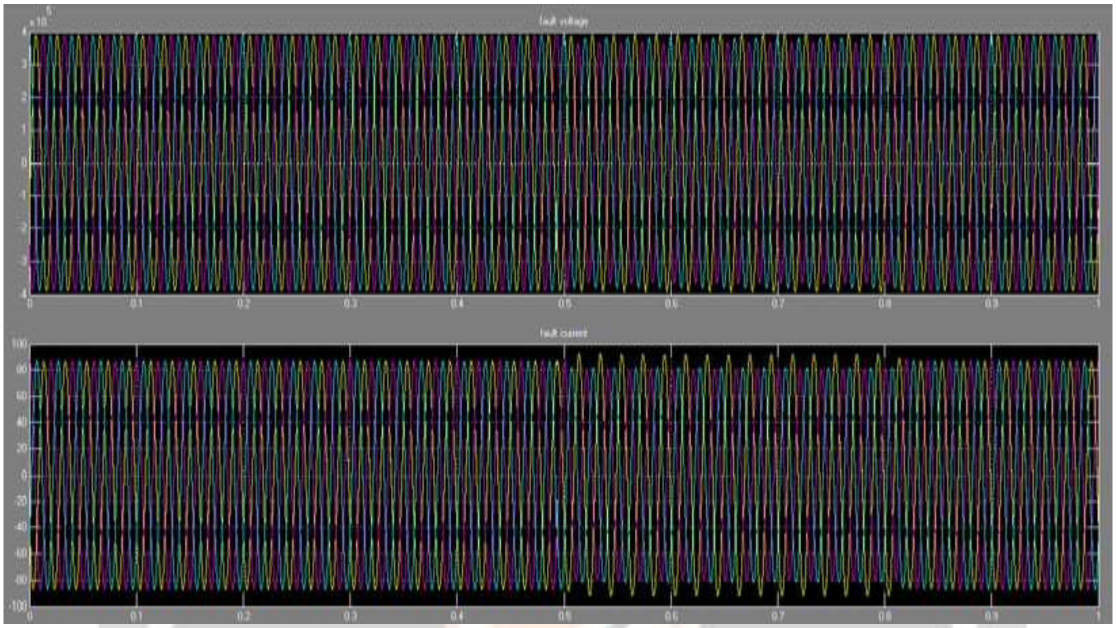


Fig 3:- Variation in fault output voltage and fault current during phase 'B-G' fault after TCSC.

Suppose the total simulation time of main model is 1sec. When the fault is occurred at 150 km of total 300 km transmission line. The start time of fault is 0.5 sec and end time of fault is .80 sec for fault resistance Rf=100Ω. The random forest analysis the features of fault voltage and fault current from data base and gives the decision that fault on phase "C-G" fault occurred before TCSC. An extensive data set is generated considering various parameters to train and test the RF-based decision tree for developing an accurate and robust classifier for fault-zone identification and different shunt fault classification in FACTS-based transmission line.
Suppose the total simulation time of main model is 1sec. When the fault is occurred at 100 km of total 300 km EHV transmission line. The start time of fault is 0.5 sec and end time of fault is .80 sec for fault resistance Rf =200Ω. The random forest analysis the features of fault voltage and fault current from data base and gives the decision that fault on phase "A-B-G" fault occurred before TCSC. An extensive data set is generated considering various parameters to train and test the RF-based decision tree for developing an accurate and robust classifier for fault-zone identification and different shunt fault classification in FACTS-based transmission line.
Consider the total simulation time of main model is 1sec. When the fault is occurred at 75 km of total transmission line. The start time of fault is 0.5 sec and end time of fault is .80 sec for fault resistance Rf=200Ω. The random forest analysis the features of fault voltage and fault current from data base and gives the decision that fault on phase B-C-G occurred after TCSC. An extensive data set is generated considering various parameters to train and test the RF-based decision tree for developing an accurate and robust classifier for fault-zone identification and different shunt fault classification in FACTS-based transmission line.

4. CONCLUSION AND FUTURE SCOPE

The proposed approach provides a new technique for fault zone identification and fault classification for transmission line employing the TCSC using random forest decision tree. The proposed DT is found to be better compared to existing SVM with respect to accuracy and computational burden. The decision tree is train with large data set with wide variation in operating condition of power system. The robustness and accuracy of proposed random forest shows potential of proposed method for protection of TCSC based transmission line as compare to SVM, NN. Random forests are an effective tool in prediction. Because of the Law of Large numbers they do not over fit. Injecting the right kind of randomness makes them accurate classifiers and regressors. The results indicate that the ensemble trees is highly effective and reliable in identifying the fault zone in the FACTS-based transmission line which triggers the next cascaded algorithms, such as an apparent impedance calculation for issuing the tripping signal in distance relaying.

For fault feature extraction and fault diagnosis, this thesis presents a classification method based on RF as a solution for the diagnosis of faults in TCSC based transmission lines. The simulation results show that this algorithm has high accuracy in the classification performance and wider generalization ability even in small group of samples using the learning and testing patterns in the voltage-current domain. Furthermore, in this work, only faults occurring in designated phases were considered and so future validation can be confirmed by verifying the fault analysis results in every possible point. In future work, there are more investigations needed that Deals with current transformer nonlinearity, harmonic and inter harmonic resonances and soon will also require a much larger data base. The presented framework can always be updated to yield better results, as new data becomes available from the same system. In future work, there are more investigations that can be performed among the different data sets needed to ensure best performance. This proposed work can also be extended for Fault classification and fault zone identification for other FACTS devices like UPFC, SSSC etc.

5. REFERENCES

[1] A. A. Girgis, A. A. Sallam, and A. Karim El-din, "An adaptive protection scheme foe advanced series compensated (ASC) transmission line," IEEE Trans. Power Del., vol. 13, no. 2, pp. 414–420, Apr. 1998.

[2] N. G. Hingorani and L. Gyugyi, Understanding FACTS: Concepts and Technology of Flexible AC Transmission Systems. New York: IEEE Press, 2000.

[3] Vireshkumar G. Mathad, Basangouda F. Ronad, Suresh H. Jangamshetti," Review on Comparison of FACTS Controllers for Power System Stability Enhancement", International Journal of Scientific and Research Publications, Volume 3, Issue 3, March 2013 1 ISSN 2250-3153

[4] Atulkumar Shuklaji Nikhade, Vivek Yadav , Rahul P. Dhote, Snusha R. Dharmik," The influence of Thyristor controlled series capacitor on distance protection relay and its mitigation", IJIREEICE Vol. 2, Issue 12, December 2014.

[5] Priyanka Kathal, Arti Bhandakkar," Power Flow Control in Power System using FACTSDevice Thyristor Controlled Series Capacitor (TCSC): A Review", International Journal of Science and Research (IJSR), India Volume 2 Issue 3, March 2013.

[6] J. Morais, Y. Pires, C. Cardoso & A. Klautau," An Overview of Data Mining Techniques Applied to Power Systems" Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, pp. 438, February 2009, I-Tech, Vienna, Austria (2009).

[7] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer,"A comparison of decision tree ensemble creation techniques," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 1, pp. 173–180, Nov.2007.

[8] D. S. Siroky, "Navigating random forests and related advances in algorithmic modeling," Stat. Rev. vol. 3, pp. 147– 163, 2009.

[9] H. Liang, H. Zhang, and Y. Yan, "Decision trees for probability estimation: An empirical study," in Proc. 18th IEEE Int. Conf. Tools with Artif. Intell., 2006. pp.1–9.

[10] Yagang ZHANG, Jing MA, Jinfang ZHANG, Zengping
WANG ," Applications of Data Mining Theory in Electrical
Engineering ", Engineering, 2009, 1, 211-215 doi:10.4236/eng.2009.13025 Published Online November 2009 rp.org/journal/eng).

[11] K. A. Dongre, A. P. Bhagat, "Power flow analysis using optimal power flow method", in Proc. International Conference on Innovations in Information, Embedded and Communication Systems, 2015, pp. 1-6.

[12] Vrushali Y kulkarni, Dr pradeep k sinha." Random Forest Classifiers: A Survey and Future Research Directions" International Journal of Advanced Computing, ISSN:2051-
0845, Vol.36, Issue.1

[13] Hal Daumé, "A course in machine learning" August 2012 http://ciml.info