Available online at www.sciencedirect.com
SCIENCE DIRECT®
Electric Power Systems Research 71 (2004) 169–177
ELECTRIC POWER SYSTEMS RESEARCH
www.elsevier.com/locate/epsr

Locating faults in the transmission network using sparse field measurements, simulation data and genetic algorithm

Shanshan Luo [a], Mladen Kezunovic [b,*], Don R. Sevick [c]

[a] *Shanghai Tengen Future Electric Automation Co., Shanghai, PR China*
[b] *Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128, USA*
[c] *CenterPoint Energy, Houston, TX, USA*

Received 4 December 2003; accepted 28 January 2004

Abstract

The paper presents a modeling and simulation approach to locate a fault in a transmission network. The basic concept is to match phasors recorded during fault with the corresponding phasors obtained by simulating the same fault. For the simulation, it is necessary to assume a fault location in a power system model, and then carry out short circuit study. The matching degree can be calculated by a pre-set criterion. The operation is repeated till the best match is found. The process of finding the best match is an optimization problem, therefore, the genetic algorithm (GA) is introduced to find the optimal solution. The proposed approach is suitable for situations where only sparsely recorded field data is available. Under such circumstances, the proposed approach can offer more accurate results than other known techniques.

Keywords: Fault location; Genetic algorithm; Simulation; Digital fault recorders; Short circuit study

1. Introduction

The fault location in the transmission network is an important issue since identifying an accurate fault location can facilitate repairing the damage and restoring the transmission line rapidly. If a fault location cannot be identified quickly and this makes prolonged transmission line outage during a period of peak load, severe economic losses may occur and reliability of service may be questioned.

A lot of efforts were spent on the topic and several solutions were proposed in the literature. In Refs. [1,2], expert systems utilize both the binary quantities (relay and breaker status) and analog quantities (voltage and current measurements) to locate a fault. Various one-, two-, or three-end algorithms utilizing voltages and currents for estimating the fault location have been proposed [3–8]. In order to obtain a more accurate result, the use of global positioning system (GPS) of satellites was introduced for performing synchronized data sampling [9]. In Ref. [10], fault location is based on the measurement of the fault-generated traveling wave component. The mentioned methods are applied successfully in many transmission network fault scenarios.

These applications have a common requirement: the measurement must be obtained from one or both ends of a faulted line. In some systems, only sparsely recorded data at limited substation locations are available. When a fault occurs in such systems, only a few (two or three) recording devices are triggered. The most likely case is that the measurements could not be obtained at either or both ends of the faulted transmission line. Under this situation, the mentioned methods could not be applied correctly.

The proposed method gives a solution even when only the sparse measurements are available. The first part describes the basic concept of the proposed method, the second presents the genetic algorithm application, the third shows details of implementation, and the last gives several typical fault cases and results.

Abbreviations: DFR, digital fault recorders; GA, genetic algorithms; GPS, global positioning system; IED, intelligent electronic devices

* Corresponding author. Tel.: +1-979-845-7509; fax: +1-979-845-9887.
E-mail address: kezunov@ee.tamu.edu (M. Kezunovic).

0378-7796/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.epsr.2004.01.011

2. Fault location approach

2.1. Sparse data case

In the paper, sparse data means the data obtained from recording devices sparsely located at substations in a power system network. Examples of recording devices may include digital fault recorders (DFR), digital relays, or other intelligent electronic devices (IED).

Fig. 1 illustrates the sparse data case. The system represents a part of the 138 kV CenterPoint Energy transmission system. While the system part has a total of 19 buses, DFRs are installed at three buses only. Clearly, the system is sparsely monitored. When a fault occurs on the line between buses 11 and 12, the DFRs located at bus 1, 3, or 16 may be triggered to record specified quantities during that fault. In certain cases, some of the DFRs at bus 1, 3, 16 may not be triggered. Then, even fewer measurements will become available for locating the fault. The data obtained in these cases may be designated as “sparse data”. The fault may be several buses away from DFR locations. Therefore, none of the common algorithms, such as one-, two-, or three-end is applicable for locating the faults. To solve this problem, a phasor matching approach is proposed as follows.

2.1.1. Phasor matching

In the phasor matching approach, the model of the power system is utilized to carry out simulation studies. The matching is made between the voltage or current phasors obtained by recording devices and those generated in the corresponding simulation studies. The fault location is placed in the system model and simulations are carried out in an iterative way. First, an initial fault location is assumed and the simulation study is set up according to the specified fault location conditions. Next, the simulation study corresponding to the specified fault is carried out and simulated phasors of the signals of interest are obtained. Then, the simulated phasors are compared with the recorded ones, and the matching degree between the simulated and recorded phasors is evaluated by using an appropriate criterion. The initial fault location is modified and the above steps are iterated until the best match between simulated and recorded phasors is produced. The fault location is then determined as the one specified in the simulation study generating the simulated phasors that best match the recorded ones.

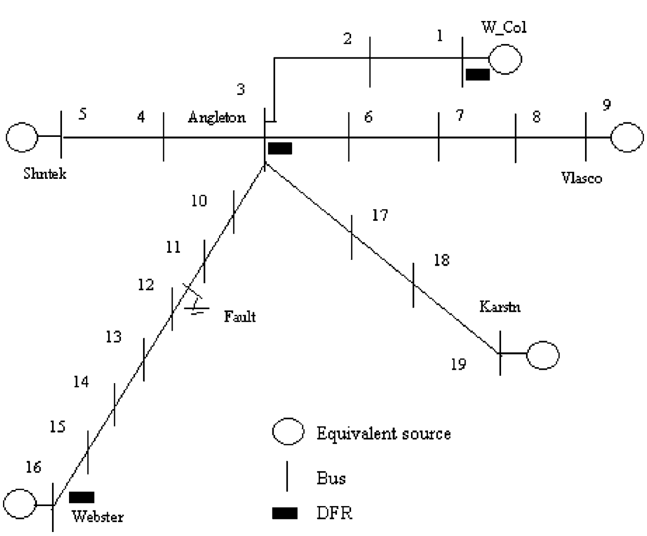

Fig. 1. The sample system for illustrating data sparsity.

2.1.2. Degree of matching

To evaluate the matching degree of the simulated and recorded waveforms, phasors are used. To perform the phasor matching, short circuit model of the system is needed. Short circuit studies can usually directly generate simulation results in the phasor format. To extract phasors from the recorded fault transients, appropriate signal processing technique needs to be applied. Fourier transform may be used for this purpose [11]. For this study, CenterPoint Energy provided the short circuit model in PSS/E [12].

In order to determine the matching degree between the simulated and recorded phasors and find out the best match, a criterion for determining the matching degree is necessary. First, the variables should be determined. When posing a fault in PSS/E, a fault location, and fault resistance should be specified. The matching degree can be formulated as follows:

$$f_c(x, R_f) = \sum_{k=1}^{N_v}\{r_{kv}|\dot{V}_{ks} - \dot{V}_{kr}|\} + \sum_{k=1}^{N_i}\{r_{ki}|\dot{I}_{ks} - \dot{I}_{kr}|\} \quad (1)$$

or

$$f_c(x, R_f) = \sum_{k=1}^{N_v}\{r_{kv}||\dot{V}_{ks}| - |\dot{V}_{kr}||\} + \sum_{k=1}^{N_i}\{r_{ki}||\dot{I}_{ks}| - |\dot{I}_{kr}||\} \quad (2)$$

where $f_c(x, R_f)$ is the defined cost function using either both the phasor angle and magnitude or the magnitude only for matching, x the fault location, R_f the fault resistance, r_{kv} and r_{ki} the weights for the errors of the voltages and currents, respectively, V_{ks} and V_{kr} the during-fault voltage phasors obtained from the short circuit simulation studies and recorded phasors, respectively, I_{ks} and I_{kr} the during-fault current phasors obtained from the short circuit studies and recorded phasors, respectively, k the index of the voltage or current phasors match, and N_v and N_i the total number of voltage and current phasors to be matched.

The cost function $f_c(x, R_f)$ theoretically equals to zero when the phasors obtained from a simulation study exactly match those obtained from the field recordings. Therefore, the best fault location estimate would be the one that minimizes the cost function. An appropriate optimization needs to be selected to solve the problem of finding a minimum. Formula (1) and (2) can be converted into the problem of finding a maximum

$$f_f(x, R_f) = -f_c(x, R_f) \quad (3)$$

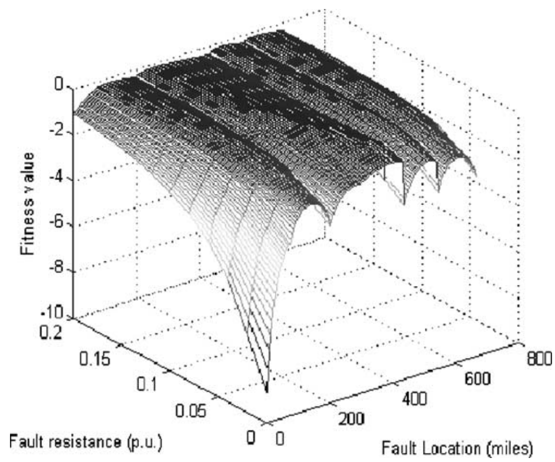

Fig. 2. The fitness surface for an A–G fault.

where $f_f(x, R_f)$ represents the fitness function using phasor magnitude for matching.

2.2. Properties of the fitness function

To investigate the nature of the fitness function, various simulation studies have been carried out to obtain fitness value using the sample power system shown in Fig. 1. The fitness value is obtained by specifying the fault location and varying fault resistance on each line throughout the system, and then running the simulations and applying formulae (2) and (3). When posing faults, the fault location was changed in the steps of 4 miles and fault resistance in the steps of 0.02 pu. For each fault location and fault resistance, a corresponding fitness value was obtained. To select fault locations, a search sequence for all transmission lines in the system had to be defined. If the recorded data are only available at bus 1, the fitness value versus the fault location and fault resistance for this specific fault is depicted in Fig. 2. The maximum fitness value occurs at point (312.7, 0.1), which is the optimal solution for the phase A to ground fault.

As it may be noted from Fig. 2, the surface is not regular since it contains some local maxima and saddles. The research shows that the genetic algorithm (GA)-based optimization approach is good at finding the globally optimal solution while avoiding the local optima and it is more robust than the conventional optimization methods. The GA is selected as a tool for finding the global optimum in our study [13].

3. Genetic algorithm application

3.1. The genetic algorithm solution

Genetic algorithms were initially developed using binary strings to encode parameters of an optimization problem. Binary encoding is a standard GA representation that can be employed for many problems: a string of bits can encode integers, real values, sets or whatever is appropriate. For our case, the binary encoding representation is selected. The variables we use are the fault location and fault resistance.

Encoding is used to map the parameters of an optimization problem into a binary string of length l. Suppose the variable x ($x_{\min} \le x \le x_{\max}$) is a decimal value that is a positive non-integer number) and it is to be represented by a binary string of length l. The encoded value x for the variable will be

$$x_b = \left[\mathrm{round}\left(\frac{(x - x_{\min})(2^l - 1)}{x_{\max} - x_{\min}} \right) \right]_b \tag{4}$$

where the value identified by round( ) gives the nearest integer of the argument, x_b and $[\,]_b$ represents binary number.

Decoding is used to convert the binary string into a meaningful decimal parameter employed in the GA. The decoding process is given by

$$x = \frac{x_{\max} - x_{\min}}{2^l - 1} x_b + x_{\min} \tag{5}$$

Based on above formulae, fault location and fault resistance should be known in advance. For the fault location, the upper limit is determined as the sum of all the candidate lines.

3.2. Fitness scaling

In order to speed up the search and convergence, a small population is used. This may result in converging prematuraly or losing the diversity. To overcome this, fitness scalling is introduced [13]. So-called fitness scaling is actually a linear scaling. Let us define the raw fitness function f (obtained from (2)) and the scaled one f'. The relationship between f and f' is as follows:

$$f' = af + b \tag{6}$$

where

$$a = \frac{(C_{\mathrm{mult}} - 1) f_{\mathrm{avg}}}{f_{\max} - f_{\mathrm{avg}}} \tag{7}$$

$$b = \frac{f_{\max} - C_{\mathrm{mult}} f_{\mathrm{avg}}}{f_{\max} - f_{\mathrm{avg}}} f_{\mathrm{avg}} \tag{8}$$

where C_{mult} is the number of expected copies desired for the best population member, f_{avg} an average of the fitness values for a specific generation, and $f_{\max}$ the maximum of all fiteness values for a specific generation.

3.3. Mapping the objective function into the fitness function

In formula (3), the fitness function is a negative value. This character does not meet the GA requirement [13]. In GA, a fitness function must be a non-negative figure of merit [13]. It is necessary to map the function to a desired fitness

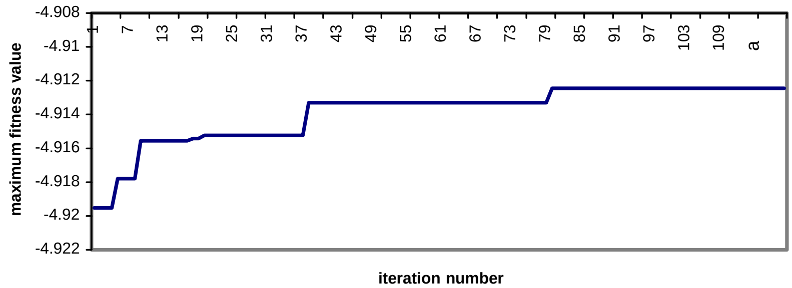

Fig. 3. Maximum fitness value for the resistance range 0–0.8.

function form through one or more mappings, which then meets the requirement of GA.

The following cost-to-fitness transformation is used

$$f(x) = C_{\max} - f_c(x, R_f) \tag{9}$$

where $C_{\max}$ is the maximum $f_c(x, R_f)$ value in the current population, and $f_c(x, R_f)$ has the same meaning as the one in the formula (3).

3.4. Updating the population

There are several approaches for updating the population. The most utilized approaches known are the general and steady state. For the general approach, the population is replaced by offspring created by reproduction, cross-over and mutation. When creating new population by GA process, the best chromosome might be lost since the selection of chromosome is more or less done at random. Elitism is the name of the method, which first copies the best chromosome (or a few best chromosomes) to new population for further evolution. Elitism can very rapidly increase performance of GA, because it prevents losing the best-found solution.

For our case, this approach is utilized. The best individual in the population pool is generally retained (these individuals are elitists). In this case, individuals can only be recombined with those from the same generation.

For some cases, the elitists are almost unchanged throughout the iterations. In order to enlarge the search region, some strategies such as using the multi-point cross-over to replace the simple cross-over, and adopting variable mutation possibility may be used as described next.

3.5. Convergence

It has been observed that GA-based FL software does not always converge to the same solution when number of generation is set as the criterion to stop iteration. In some cases, differences may be significant. Several alternative methods are investigated in an attempt to obtain further improvements. Only two of them are listed as follows.

3.5.1. Using maximum fitness value as the iteration stop criterion

Maximum fitness value refers to the maximum of all population's fitness values within a specific generation. In order to observe the changing regularity of the maximum fitness value, some cases were tested. Fig. 3 shows one example in which the maximum fitness value varies with the iteration and may remain the same within many generations. The tendency of hill climbing is visible. *Note:* The fitness value in Fig. 3 corresponds to the one listed in formula (3). When the maximum fitness value approaches a specified threshold, GA iteration is stopped. Therefore, using maximum fitness value as a stop criterion may be a choice. However, sometimes it is difficult to determine the threshold.

3.5.2. Using average fitness value as the iteration stop criterion

Average fitness value refers to the average of all populations' fitness values among a specific generation. Fig. 4 shows that average fitness value gradually approaches the maximum fitness value with the number of iterations increasing. *Note:* The fitness value here corresponds to formula (9). This property can be utilized. When the relative

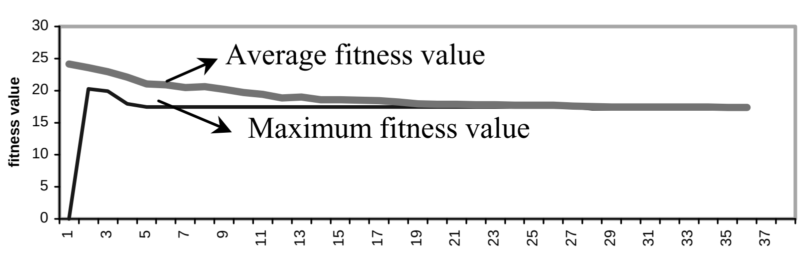

Fig. 4. Average fitness value for the resistance range 0–0.8.

error between the average fitness value and the maximum meet a specific threshold or a relative ratio of maximum fitness value to average fitness value, the iteration stops.

According to the above analysis, the second criterion is suggested. It may require longer computation time than when only using a fixed generation number as the stop criterion. An alternative is to limit the search range using other information and then continue by using GA optimal method to find out exact final solution. For example, using the post-fault breaker status or SOE records as well as the oscillography produced by protection relays located in different substations, one can estimate the fault location.

4. Implementation

4.1. Architecture of fault location software

The overall architecture of the fault location solution is shown in Fig. 5. Two commercial software packages, represented by dotted lines in Fig. 5, are utilized. One is DFR Assistant [14]. It analyzes field-recorded fault waveforms as well as relay breaker and communication channel status based on an expert system. It also converts the DFR raw data into COMTRADE format [15]. DFR Assistant can generate an analysis report including identification of the fault type and a suspected faulted line. Another is PTI Power System Simulator (PSS/E) [12]. It can calculate the power flow and carry out the short circuit study. The main modules of the software are discussed next.

4.2. Data requirement

The data requirement includes: static power system model, fault data, substation interpretation data, and fault information entered by the user.

The static system model refers to the saved data case of PSS/E. It should contain the power flow raw data, sequence impedance data, and system topology. The model only reflects a specific system operating condition.

Fault data refers to the data captured by digital fault recorders (DFRs) and data file should be in the COMTRADE format since the software is designed to read fault data provided in that format.

The substation interpretation data contains information that relates the channel index numbers to the monitored signals and represents correspondence between the monitored circuits and PSS/E numbering scheme. Each substation should have one interpretation file. The interpretation file needs to be modified to reflect the DFR configuration or the system model changes.

The data provided by the user includes necessary fault information, matching options, and selected fault data. The necessary fault information relates to the estimated fault type and faulted circuit that can help in limiting the GA search range. The matching options are used for specifying currents through the circuits or voltages at buses for phasor matching. Selected fault data refers to a choice in the use of different combinations of DFR records under the situation where multiple DFRs are triggered.

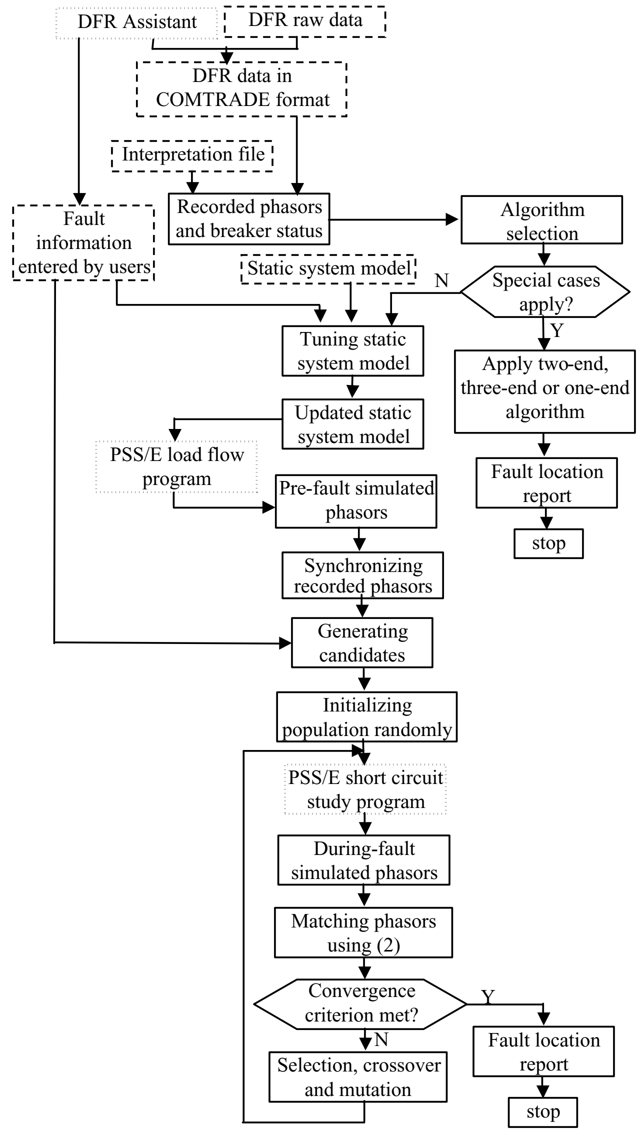

Fig. 5. The flowchart of fault location software.

4.3. Calculation of phasors from recorded data

Considering that phasors are needed to represent during-fault conditions, two processing steps are taken. A specified low-pass filter is utilized to remove high-frequency noises. An improved Fourier algorithm is implemented to effectively remove decaying dc-offset component and obtain the pre-fault and during-fault phasors of voltages and currents [11].

The pre-fault phasor can be calculated using the first cycle of the recorded phasor. The during-fault phasor can be calculated using any fault cycle following the fault inception and prior to fault clearance. Therefore, identification of the moment of fault inception is required before calculating the during-fault phasors. The fault cycle is determined from the waveforms recorded by DFR.

It is possible to select different fault cycle to calculate the during-fault phasors from the different DFR recordings. This may introduce fault location error, especially for the arcing faults during which the fault resistance is changing. Under this situation, selecting different fault cycles means experiencing different impacts of fault resistance. An alternative is that the user checks the waveforms manually and specifies the same matching fault cycles across all recordings.

4.4. Synchronization of phasor

It is well known that the phase angle of a selected signal in a three-phase system, calculated by a Fourier algorithm, may be constant or it may rotate $360/N$ (N is sample times per cycle) degrees with each new sample depending upon selection of the time reference. For a specific fault occurring in a power system, several DFRs may be triggered. Phasors calculated by using waveforms recorded in different substations may lack synchronization since the time reference in different substation may not be the same. This will introduce phase angle difference among phasors calculated using waveforms from the substation where DFRs are triggered. This factor will introduce an error in calculating the fault location using during-fault phasor. The synchronization of the phasors obtained from different DFR recordings is necessary.

The PSS/E load flow study based on the modified system model, described in Section 4.5, is carried out to obtain the pre-fault simulated phasors. The phasors calculated from each DFR recording are angle-synchronized by rotating them in reference to the phasors obtained by the load flow study. As a result, each recorded pre-fault phasor has the same phase angle as the corresponding one obtained through simulation. It is assumed that the angle difference between the pre- and during-fault phasor, for the corresponding recorded current or voltage, is fixed. Hence, the during-fault phasor should also be rotated for the same angle. All recorded pre- and post-fault phasors are synchronized using the same reference.

4.5. Model tuning

A given static system model topology, used in simulation studies, may not reflect the prevailing system topology when the fault occurs. To match the phasors extracted from DFR recordings and those obtained from simulation studies, the model used in the study should be updated by utilizing the topology information captured close to the moment when the fault occurs. Tuning of the system model may include updating the topology as well as generation and load data of the system. The pre-fault data, including the pre-fault phasors and breaker status, contained in DFR records, may be used for the tuning process.

4.5.1. Topology modification

Based on the pre-fault breaker status and current magnitudes of the monitored branches derived from the pre-fault data recorded by DFR, the service status (saved in the PSS/E file) of the branches will be updated. A zero (or smaller than a pre-set value) magnitude of the current through a monitored branch indicates an out-of-service status of the branch. If both the current and breaker status of a branch are available, the current measurement will be used instead of the breaker status for the topology update.

The topology modification is realized using the IPLAN language, which is part of the PSS/E package [12]. The user is able to program a sequence to modify the system topology, control the load flow and short circuit studies, and control the reporting of the results of the PSS/E activities. The IPLAN language also facilitates interaction between the PSS/E activities and C++ program used to implement the fault location algorithm.

4.5.2. Scaling the generator and load power

There are two ways to update the generator and load data. When additional data such as power for generators and loads in a specific area are available, this can be utilized to tune the static system model. When additional field data are not available, the fault location software will use the probing method to produce updated models and finally pick the best one. For each model, the load and generator scaling feature, SCAL, which enables the user to uniformly increase or decrease any or all quantities for a selected grouping of buses, loads and machines, is utilized. The following rule is applied in determining the individual bus, load, or machine power: adjust the power such that the ratio of individual bus, load, or machine power to the respective total power of all buses, loads, or machines being processed remains unchanged. The activity SCAL can be invoked with the suffix, such as “ALL”,“AREA” and ”ZONE” indicating if the whole system or only a subsystem is to be considered. In our study, we use “AREA” to limit the tuning range.

Based on pre-set strategy, a series of updated models are produced. For each model, the simulated load flow is calculated and then the recorded load flow is matched with simulated load flow. The matching criterion is similar to (1) and (2). The difference is as follows:

$V_{k\mathrm{s}}$ and $V_{k\mathrm{r}}$ should be the pre-fault voltage phasors obtained from the short circuit simulation studies and recorded waveforms, respectively;

$I_{k\mathrm{s}}$ and $I_{k\mathrm{r}}$ should be the pre-fault current phasors obtained from the short circuit simulation studies and recorded waveforms, respectively.

Based on the matching criterion, the updated model is determined before running the fault location software by

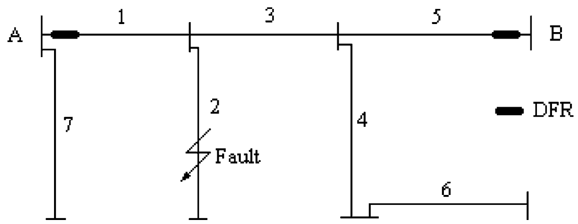

Fig. 6. Illustration of the building of the initial faulted branch candidates.

changing the generator power in the whole network or some specific areas. This is done based on real time measurements obtained half an hour or even closer in time before the fault occurs. Then the activity “SCAL” is executed in PSS/E. The total change of the load data should equal to the total change of the generator data. The updated system model replaces the static system model and PSS/E uses the updated one to simulate faults.

4.6. Determination of branch candidates

Based on DFR records and fault information obtained from digital relays (such as estimated fault section and fault type), a list of faulted branch candidates is produced for the purpose of posing faults. Fig. 6 is used to illustrate the approach.

Suppose that DFRs located at substation A and B are triggered by a fault on the line 2. Then a list of possible faulted branch candidates based on each DFR location can be obtained. In order to do so, the depth of the search layers needs to be specified. The lines directly connected to the triggered DFR are considered belonging to the first search layer. The lines connected to the other side of the lines selected in the first search layer are considered as the second layer. For example, the first search layer starting from substation A consists of lines 1 and 7 and the second layer consists of lines 3 and 2. If the search layer is the third, a list of possible faulted branches can be obtained based on each DFR as shown below.

List 1 (from A): 1, 7, 2, 3, 4, 5;
List 2 (from B): 5, 4, 3, 6, 1, 2.

To produce a list of faulted branch candidates, two set operators may be used: union and intersection.

Union: 1, 2, 3, 4, 5, 6, 7;
Intersection: 1, 2, 3, 4, 5.

4.7. User interface

The information that may be provided by the user is shown in Fig. 7.

Main information in Fig. 7 includes

- DFR data interpretation;
- Fault type;

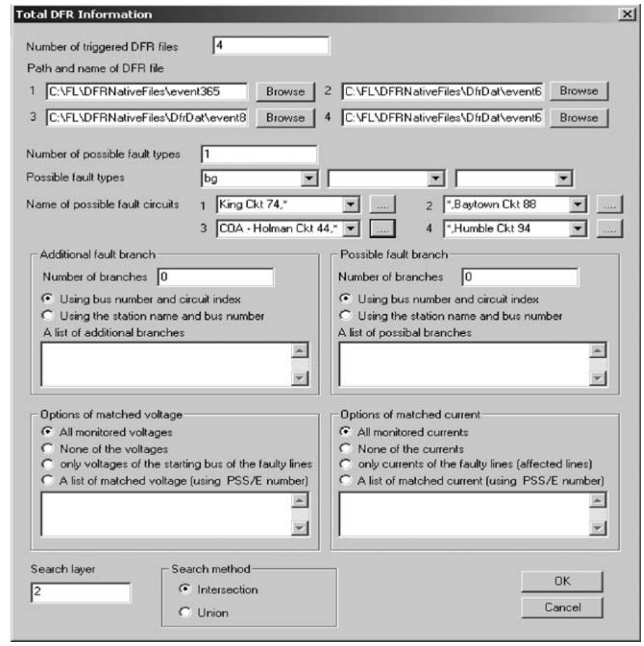

Fig. 7. Illustration of the building of the initial faulted branch candidates.

- Possible faulted circuit, search depth, and the set operators;
- Quantities used for matching.

5. Test results

This section presents test results. CenterPoint Energy provided detailed description of fault cases. The version of the utilized PSS/E is 28.0. In the new version, information on tapped transmission lines is included. The system has 5437 buses, 5544 lines, 528 generators, 3263 loads, and 1068 transformers. Over 30 DFRs are installed in the system. The results presented here go well beyond the initial tests results provided when the concept was first introduced [16,17].

Fig. 8 shows one-line diagram in which three DFRs, located at the SRB, Cedar Bayou and South Channel substation, respectively, were triggered when the fault occurred. The fault occurred on 138 kV system. The actual fault location, determined and provided by CenterPoint Energy, is on

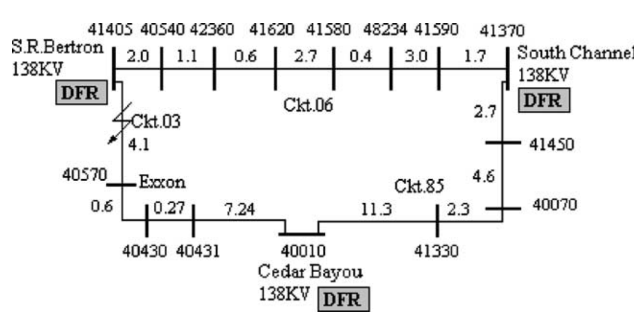

Fig. 8. Section of CenterPoint Energy Transmission for case I.

Table 1
Sensitivity study results for case I

DFR utilized	Fault location	|Error| (%)	Quantities marched
Event 316	40570–41450, 37.2% from 40570	1.8	Current of Ckt. 03 in SRB
Event 316, Event 318	40570–41450, 39.3% from 40570	0.3	All recorded currents and voltages in SRB and Cedar Bayou
All three	40570–41450, 39.6% from 40570	0.6	All recorded currents and voltages

Table 2
The calculation result

Case #	Number of DFR triggered	Fault type	Actual fault location	Estimated fault location	Error
1	2	BG	48402–40590 3.32 miles	48402–40590 3.63 miles	0.3
2	3	BG	40570–41405 2.50 miles	40570–41405 2.47 miles	0.0
3	1	AB	46570–48412 1 mile	46570–48412 0.1 mile	0.9
4	1	AG	5915–9073 66.0 miles	5915–9073 66.9 miles	1.0
5	3	BG	40620–48295 2.36 miles	40620–48295 2.13 miles	0.2
6	2	CG	46020–3390 7.77 miles	46020–3390 7.09 miles	0.7

Exxon Ckt. 03, 2.5 miles from SRB 138. No line taps are located in between DFR and fault location.

According to the analysis report of DFR Assistant, the fault type is B to ground; the affected circuit based on SRB data is Exxon Ckt. 03; based on South Channel data, it is Tenneco poly Ckt. 06; and based on Cedar Bayou data, it is Exxon Ckt. 83.

Based on the information, we select different combinations of the DFR files and quantities for matching shown in Table 1 to estimate the fault location. Whatever DFR files are selected, the estimated results are close to the actual fault location provided by CenterPoint Energy. The estimated errors are smaller than 0.1 mile. In this case, the choice of particular quantities selected for matching does not affect the result.

More fault cases were tested. The detailed results are listed in Table 2.

6. Conclusion

The paper presents a new fault location approach using “phasor matching” based on genetic algorithm. The biggest advantage is to utilize the sparse data to locate fault without necessarily a need for additional recording device or more monitored data. It is suitable for the situation in which the conventional algorithms cannot be applied. The approach does not refer to a specific section or line; it is based on a system view. However, a system model including the static parameters and topology is required.

The recently improved software package has been tested using some fault cases in the CenterPoint Energy system. The test results show that the approach is quite promising as illustrated by the test cases given in the paper.

Acknowledgements

Drs. Yuan Liao and Zijad Galijasevic are recognized for their original research contribution made on this subject while working as graduate students at TAMU.

Appendix A. List of symbols

C_{max}	maximum $f_c(x, R_f)$ value in the current population
C_{mult}	number of expected copies desired for the best population member
f	raw fitness function
f'	scaled value of f
f_{avg}	average of the fitness values for a specific generation
$f_c(x, R_f)$	defined cost function using either both phasor angle and magnitude or magnitude only for matching
$f_f(x, R_f)$	fitness function using phasor magnitude for matching
f_{max}	maximum of all fiteness values for a specific generation
I_{ks} and I_{kr}	during-fault current phasors obtained from short circuit studies and recorded phasors, respectively
I_{ks} and I_{kr}	pre-fault current phasors obtained

	from short circuit studies and recorded waveforms, respectively
k	index of the voltage or current phasors match
l	binary string of length
N_v and N_i	total number of voltage and current phasors to be matched, respectively
round( )	nearest integer of the argument
r_{kv} and r_{ki}	weights for the errors of the voltages and currents, respectively
R_f	fault resistance
V_{ks} and V_{kr}	during-fault voltage phasors obtained from the short circuit simulation studies and recorded phasors, respectively
V_{ks} and V_{kr}	pre-fault voltage phasors obtained from short circuit simulation studies and recorded waveforms, respectively
x	fault location
x_b and $[\,]_b$	binary number

References

[1] A.A. Girgis, M.B. Johns, A hybrid expert system for faulted section identification, fault type classification and selection of fault location algorithms, IEEE Transactions on Power Delivery, vol. 4 (2), April 1989.

[2] M. Kezunovic, I. Rikalo, C.W. Fromen, D.R. Sevik, S.M. McKenna, D. Hamai, W.M. Carpenter, S.L. Goiffon, Automated fault analysis using intelligent techniques and synchronized sampling, in: Proceedings of the CIGRE General Session, Paris, France, September 1998.

[3] M.S. Eriksson, G.D. Rockfeller, An accurate fault locator with compensation for apparent reactance in the fault resistance resulting from remote-end infeed, IEEE Transactions on Power Apparatus and Systems, vol. PAS-104 (2), February 1985.

[4] Q. Zhang, Y. Zhang, W. Song, D. Fang, Transmission line fault location for single-phase-to-earth fault on non-direct-ground neutral system, IEEE Transactions on Power Delivery, vol. 13 (4), October 1998.

[5] D. Novosel, D.G. Hart, E. Udren, J. Garitty, Unsynchronized two-terminal fault location estimation, IEEE Transactions on Power Delivery, vol. 7 (1), January 1992.

[6] A.A. Girgis, D.G. Hart, W.L. Peterson, A new fault location technique for two- and three-terminal lines, IEEE Transactions on Power Delivery, vol. 7 (1), January 1992.

[7] D.L. Waikar, S. Elangovan, A.C. Liew, Fault impedance estimation algorithm for digital distance relaying, IEEE Transactions on Power Delivery, vol. 9 (3), July 1994.

[8] Y. Liao, S. Elangovan, Digital distance relaying algorithm for first-zone protection for parallel transmission lines, IEE Proceedings-Part C: Generation, Transmission and Distribution, vol. 145 (5), September 1998.

[9] M. Kezunovic, B. Perunicic, Automated transmission line fault analysis using synchronized sampling at two ends, IEEE Transactions on Power Systems, vol. 11 (1), February 1996.

[10] Z.Q. Bo, A.T. Johns, R.K. Aggarwal, A novel fault locator based on the detection of fault generated high frequency transients, in: Proceedings of the Sixth International Conference on Developments in Power System Protection, Nottingham, England, March 1997.

[11] Y. Guo, M. Kezunovic, D. Chen, Simplified algorithms for removal of the effect of exponentially decaying dc-offset on the Fourier algorithm, IEEE Transaction on Power Delivery, vol. 18 (3), July 2003.

[12] Power Technologies Inc., PSS/E-27-Program Operation Manual, PTI Inc., December 2000.

[13] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

[14] Test Laboratories International Inc. DFR Assistant, Product Description, www.tli-inc.com.

[15] IEEE Std. C37.111-1999, IEEE Standard Common Format for Transient Data Exchange (COMTRADE) for Power Systems, IEEE, March 1999.

[16] M. Kezunovic, Y. Liao, Fault location estimation based on matching the simulated and recorded waveforms using genetic algorithm, in: Proceedings of the Development in Power System Protection Conference, Amsterdam, The Netherlands, April 2001.

[17] M. Kezunovic, S. Luo, A novel method for transmission network fault location using genetic algorithms and sparse field recordings, in: Proceedings of the PES IEEE Summer Meeting, July 2002.