A Novel Method for Transmission Network Fault Location Using Genetic Algorithms and Sparse Field Recordings

Mladen Kezunovic, *Fellow*, *IEEE*, Shanshan Luo, Donald R. Sevcik, *member*, *IEEE*

Abstract—**The paper presents an approach to locate a fault in a transmission network based on waveform matching. Matching during-fault recorded phasor with the during-fault simulated phasor is used to determine the fault location. The search process to find the best waveform match is actually an optimization problem. The genetic algorithm (GA) is introduced to find the optimal solution. The proposed approach is suitable for the situations where only the data recorded sparsely is available. Under such circumstances, it can offer more accurate results than other known techniques.**

Index Terms—**Fault Location, Phasor Matching, Genetic Algorithm, Simulation, Digital Fault Recorders.**

I. INTRODUCTION

THERE are many fault location methods for a specific transmission line applications, such as one-end[1], two-end[2], three-end[3], utilizing the voltages and currents measured at the line end(s) to calculate the impedance. Another method, based on a stand-alone recording device to capture the high frequency transient signal generated by faults [4], is utilizing traveling wave-based algorithm to locate the fault. However, it is difficult to locate a fault in a transmission network when data obtained from only a limited number of recording devices is available. This paper proposes an approach using waveform matching to locate fault even when only limited data is available.

The proposed approach utilizes the data obtained from the recording devices installed in the power system. Here, the recording devices may include digital fault recorders (DFRs), digital relay or other intelligent electronic device. The recorded data may include analog quantities (voltages and currents), as well as digital quantities (breaker and relay operation status). When a fault occurs in the system, some devices are triggered and corresponding records are sent to the central office. The data may be limited but it can still be used to locate the fault based on the proposed approach.

In our work, we define the limited data as "sparse data". Sparse data may result for two reasons: 1. DFR or digital relays may not be installed at every substation or bus for monitoring purpose. 2. Every DFR may not always be triggered under fault condition. Whatever the case, for most fault cases, only limited measurements may be available. Under this circumstance, the mentioned methods to locate a fault in a transmission network cannot be utilized. This paper aims at proposing a more flexible fault location method, which matches the recorded during-fault waveform with the simulated during-fault waveform.

First, the waveform matching idea is explained. Next, the algorithm implementation is discussed. Test results and conclusions are given at the end.

II. WAVEFORM MATCHING

In the "waveform matching" approach, a power system model is used to carry out simulation study. A fault needs to be posed to obtain the during-fault simulated waveforms. Then the matching is made between the simulated waveforms and the recorded waveforms to determine the fault location based on the degree of matching. Theoretically, the simulated fault waveform will match completely with the recorded fault waveform if the assumed fault location and fault resistance correspond to the real fault condition.

The process to determine the fault location is iterative because several lines in the system and variety of possible fault resistances should be searched to obtain the optimal matching. In the practical operation, most probable fault locations are searched firstly by selecting a certain fault resistance. Changing the fault resistance according to a specific increment, fault locations are searched thoroughly. The process will proceed till the selected sections in power system and possible fault resistance range are exhausted.

After the search is completed, the fault location is determined based on the optimal matching scheme. There are two possible schemes - the phasor matching and transient matching. In the "phasor matching", short circuit studies are carried out to obtain phasors under fault condition. In the "transient matching", transient simulations are carried out to obtain transient waveform. We utilize the phasors for matching. The matching degree can be represented by a value obtained from the following criterion.

$$f_c\left(x,R_f\right)=\sum_{k=1}^{N_v} r_{kv}\left|V_{ks}-V_{kr}\right|+\sum_{k=1}^{N_i} r_{ki}\left|I_{ks}-I_{kr}\right| \qquad (1)$$

where,

$f_c\left(x,R_f\right)$-the cost function using phasors for matching, it is a non-negative number

x, R_f -the fault location and fault resistance.

This work was supported by PSerc Consortium, an NSF I/UCRC, and in part by Texas A&M University.

Mladen Kezunovic is with Dept. of Electrical Engineering, Texas A&M University, College Station, TX 77843 (e-mail: kezunov@ee.tamu.edu)

Shanshan Luo works presently as a research associate with Dept. of Electrical Engineering, Texas A&M University, College Station, TX 77843. (e-mail: shanshan@ee.tamu.edu)

Donald R. Sevcik is with Reliant Energy HL&P, P.O. Box 1700, Houston, TX 77251-1700, (e-mail: don-sevcik@reliantenergy.com)

r_{kv}, r_{ki} -weights for the errors of the voltages and currents respectively

V_{ks}, V_{kr} -the during-fault voltage phasors obtained from the short-circuit studies and from recorded waveforms respectively

I_{ks}, I_{kr} -the during-fault current phasors obtained from the short-circuit studies and from recorded waveforms respectively

N_v, N_i -the number of the selected voltage and current phasors respectively

k - the index of the voltage or current phasors

The cost function will be zero when the phasors obtained from the simulation studies exactly match those obtained from the recorded waveforms. The best fault location estimation will be achieved when the cost function is at a minimum. Therefore, the problem of fault location estimation is actually the optimization problem.

Since there may be several local minimum and maximum points, it is difficult to use the gradient-based method to find the global minimum. The GA based optimization approach is a good choice to search for the global optimal solution. We have to convert the minimization problem to maximization problem in order to utilize the GA. That requires us to convert the cost function to a fitness function of GA. The simplest conversion is to multiply the cost function by a minus one. We have to add a constant to make the corresponding fitness function positive. The fitness function is as follows:

$$f(x, R_f) = C_{\max} - f_c(x, R_f) \tag{2}$$

where,

$f(x, R_f)$ is fitness function

$C_{\max}$ is the maximum fitness value in the current population.

In Equation (2), fault location and resistance are selected as two variables. They are represented as binary strings in GA. Three GA operators are generally used: selection, crossover and mutation. The selection operator mimics the process of natural selection of strings to create a new generation, where the fittest members reproduce most often. The crossover, applied with a probability, acts on a pair of selected members providing the exchange of binary string. The mutation, applied with a probability, randomly affects the single bit in a member. The GA search process is as follows: at the beginning, the initial population is generated randomly. Then posing the fault according to the initial population, the short circuit study is carried out to obtain the simulated during-fault phasors and further calculate the fitness value for each individual. The next generation is produced by applying the three steps as described above. The process is repeated until the best match is found.

III. THE IMPLEMENTATION OF THE FAULT LOCATION SYSTEM

A. Overall Architecture

The overall architecture of the fault location solution is shown in Fig.1. Two commercial software packages, represented as dotted line in Fig.1, are utilized. One is DFR Assistant [5], utilized to analyze the fault waveform, relay breaker and communication channel data based on an expert system. It also converts the DFR raw data into COMTRADE format [6]. DFR Assistant can generate an analysis report including the fault type and possible faulted line. Another is PSS/E (PTI Power System Simulator) [7]. It can calculate the power flow and carry out the short circuit study.

The input modules represented by broken lines include DFR data files, the interpretation files, fault information entered by user, and power system model. The first three items are necessary for each monitored substation. The last one is used as the input to PSS/E.

The main modules of the software are discussed next.

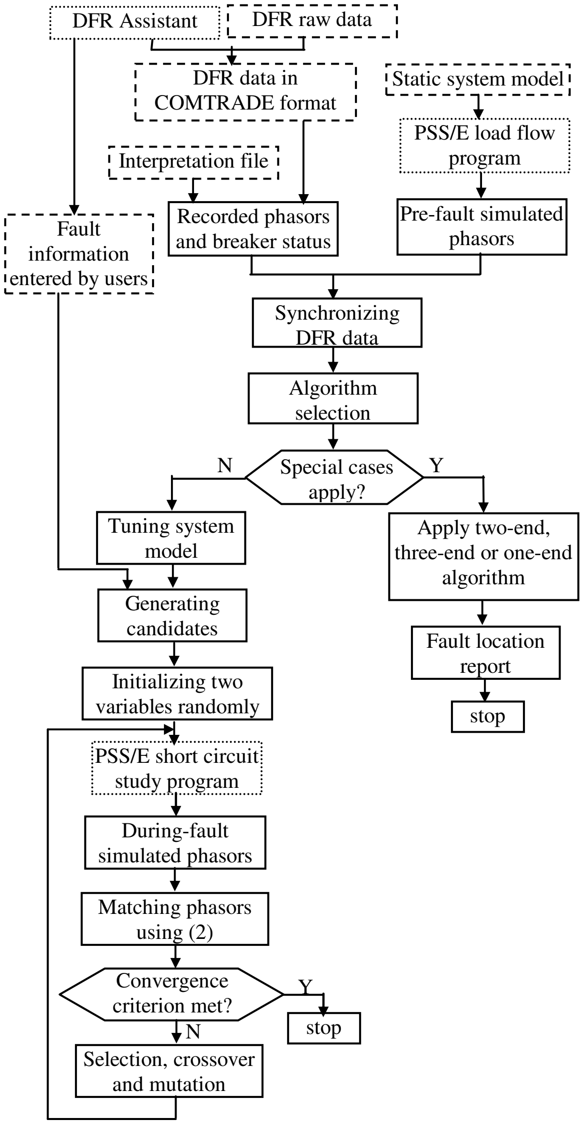

Fig. 1 Architecture of the fault location software

B. Data Requirement

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

The static system model refers to the saved case of PSS/E. It should contain the power flow raw data, sequence impedance data and system topology. The model is static since it may not reflect the prevailing system conditions. This may affect the accuracy of the algorithm and some measures overcoming the shortcoming should be taken.

Fault data refers to the data captured by Digital Fault Recorders (DFRs). Current software reads fault data provided in COMTRADE format [6]. The COMTRADE file should include two files: COMTRADE configuration file, which contains information for interpreting the data file, and COMTRADE data file, which contains analog (current and voltage) and digital values (breaker contacts and relay status) for all input channels for a specific substation.

The substation interpretation data contains information that relates the channel numbers to the monitored signals. It also represents the correspondence between the designations used in the DFR files and those used in the PSS/E file. Each substation should have one interpretation file and the interpretation file needs to be modified to reflect the DFR configuration or the system model changes. The information should be provided by the user in advance.

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

C. Synchronizing Phasors Obtained from DFR Recordings

In order to apply equation (2) to calculate the fitness value, the voltage and current phasors for different substations are obtained by applying the Fourier algorithm. The DFR data from different channels in the same substation or in different substation may lack synchronization. In order to reduce the error of matching, the phasors calculated from DFR recordings should be synchronized. Fig. 2 shows the relationship between the phasors obtained from the load flow study and from the recorded waveforms. S_{na}, S_{nb}, S_{nc} represent pre-fault phasor of phases a, b, c respectively obtained by the load flow study. R_{na}, R_{nb}, R_{nc} represent pre-fault phasors of phases a, b, c respectively obtained from the recorded waveform. R_{fa}, R_{fb}, R_{fc} represent during-fault phasors of phases a, b, c respectively obtained from the recorded waveform. α is the angle difference between the pre-fault phasor obtained by the load flow study and the pre-fault phasor obtained from the recorded waveform.

The synchronization is done by rotating counterclockwise the pre-fault phasors R_{na}, R_{nb}, R_{nc} by an angle of α. Consider that the angle difference between pre-fault phasor and during-fault phasor for a specific phase and current (voltage) is fixed; during-fault phasor R_{fa}, R_{fb}, R_{fc} is also rotated by an angle of α. The DFR data are synchronized to the simulated phasors in the same way. Note that α may be different for each monitored signal.

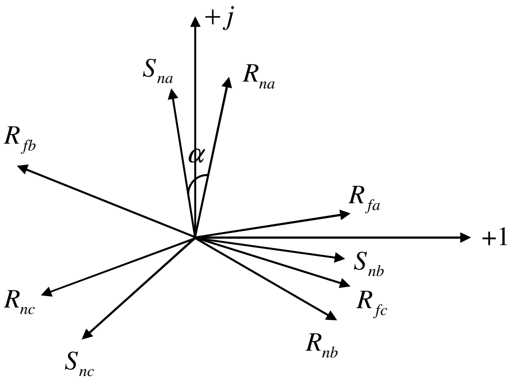

Fig.2 The relationship between the phasors obtained by load flow study and by recorded waveform

D. Tuning the Static System Model

As mentioned earlier, the given static system model, used in the simulation studies, may not reflect the prevailing operation conditions of the system when the fault occurs. The generator power output and load power may not always keep the same value and may vary with time. To match exactly the phasor extracted from DFRs and those obtained from simulation studies, it may be beneficial that the system model used in simulation studies is updated by utilizing the information captured close to the moment before the fault occurs. The process of updating the system is called “tuning of the static system model”.

Tuning the static system model includes two aspects [8]: tuning the topology of the system model and tuning the static parameters such as generator and load data.

The updating of the system topology relates to updating the service status of the circuits in the system. The pre-fault breaker status of the circuit or current through the circuit is used to update the system topology. There are two possible situations. In first situation, where both the circuit’s breaker status and currents (or only the currents) are monitored by DFR, the current magnitude will be utilized to update the service status of the circuit. If the current magnitude is smaller than a pre-specified threshold, the circuit will be designated as being out of service. Otherwise, the circuit will be designated as being in service. In the second situation, where only the breaker status of the circuits is monitored, the pre-fault breaker status will be utilized to determine the service status of the circuit. If the pre-fault breaker status indicates an open circuit, then the circuit will be determined as out of service. Otherwise, the circuit will be determined as being in service.

The goal of updating the generator output power and load power is to bring the static system model closer to the real life system. To reach the goal, the waveform-matching based approach is utilized again. The matching is made between the voltage and current waveforms obtained by DFRs and those generated by load flow studies. The equation (1) is applied as the objective function to evaluate the matching degree of the simulated and recorded waveforms. Here, the corresponding V_{ks},V_{kr} and I_{ks},I_{kr} have different meaning. They are the pre-fault voltage or current phasors obtained from the load flow study and recorded waveforms respectively. The static parameters that provide the best match are the ones that minimize the objective function. The flowchart for tuning the static system parameters is shown in Fig. 3. Table 1 shows the effect of tuning the static parameters. The fault case is selected for which two DFRs located at Limstone and THW locations respectively, were triggered. In table 1, the first column represents the different combinations of DFR data. The second column shows the fitness value calculated using the pre-fault phasor. The fitness value decreases significantly after tuning the system based on the strategy mentioned above. The updated fitness value is shown in the third column. These results prove the tuning strategy is effective. The more accurate tuning depends on the region of tuning and more real life data.

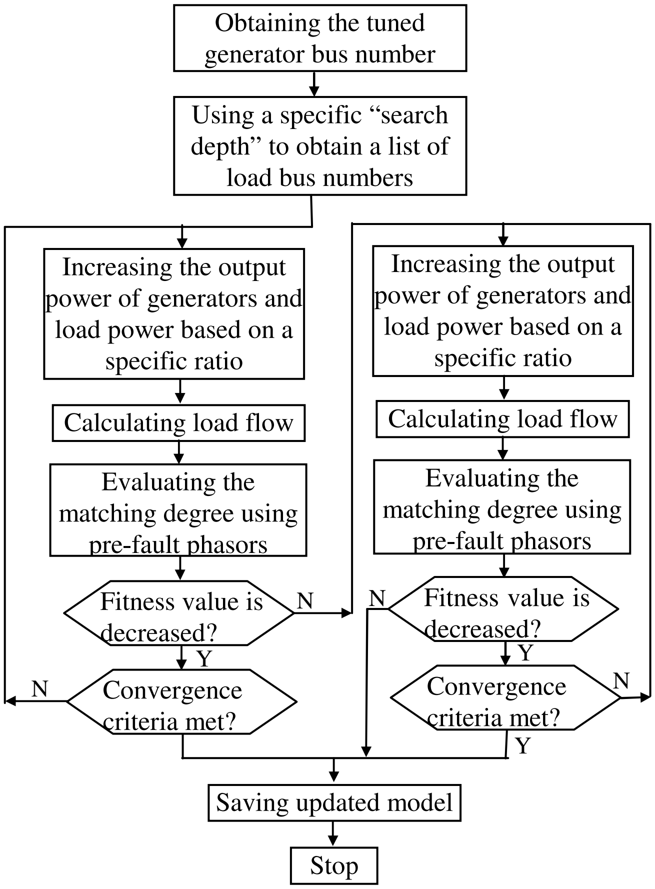

Fig.3 The flowchart of tuning the static parameters

TABLE 1 THE CHANGE OF FITNESS VALUE USING PRE-FAULT PHASOR BEFORE AND AFTER TUNING

DFR Utilized	Fitness value using pre-fault Phasor before model tuning	Fitness value using pre-fault phasor before model tuning	Quantities matched
Limestone	15.50	9.29	All monitored currents
Limestone THW	5.34	0.53	Currents on affected Ckt.74
Limestone THW	35.49	26.43	All currents
Limestone THW	7.36	0.91	Currents on affected Ckt.74, Ckt.98
Limestone	15.71	9.50	All currents and voltages
Limestone THW	35.81	26.75	All currents and voltages

E. Generating a List of Possible Fault Branch Candidates

The purpose of building a list of faulty branch candidates is to limit the search range and save the run time. The candidates for posing faults are selected based on the DFR data and the list of faulty branches obtained from DFR Assistant software package. Fig. 4 is used to illustrate the approach.

Suppose that the DFRs are located at substations Angleton, W_Col and Webster. Assume that for a given fault, DFRs installed at bus 1 and 16 are triggered. Using data recorded at bus 1, DFR Assistant finds that the fault is on the line between bus 1 and 2 or beyond bus 2. Using data recorded at bus 16, the fault is found to be on the line between bus 16 and 15 or beyond bus 15. For GA-based approach, a list of faulted lines needs to be prepared and "search depth" selected.

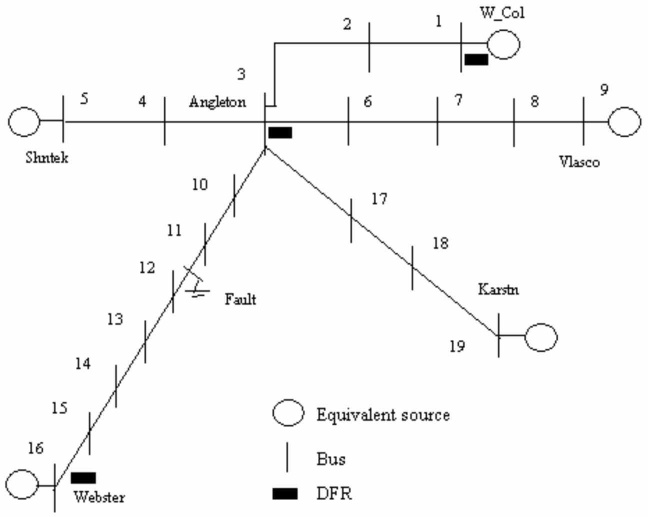

Fig. 4 Illustration of the building of the initial faulty branch candidates

The "search depth" is defined as the number of lines on a certain search path starting from a specified bus. For example, starting from bus 1, a search depth of 2 will result in two lines: 1-2 and 2-3. "1-2" represents the line from bus 1 to 2. A search depth of 4 will result in the following lines: 1-2, 2-3, 3-4, 4-5, 3-10, 10-11, 3-17, 17-18, 3-6, and 6-7. Starting from bus 16, a search depth of 4 will result in the following lines: 16-15, 15-14, 14-13, and 13-12.

F. Implementation of Fault Location Algorithm

Based on the system topology data and the number of monitored voltages and currents, the type of the fault location algorithm is determined.

The fault location software compares the actual system topology data and the available measurements to determine whether the special algorithms, such as one-end, two-end or three-end, should be used. Further disscussion is realted to the GA approach when only the sparse data is available.

In order to speed the search and convergence, the small population is used. This may result in converging prematuraly or losing the diversity. To overcome the default, fitness scalling is introduced [9]. So-called fitness scaling is actually a linear scaling. Let us define the raw fitness function f (obtained from (2)) and the scaled f'. The relationship between f and f' as follows:

$$f' = a f + b \tag{3}$$

where

$$a = \frac{(C_{mult} - 1) \cdot f_{avg}}{f_{\max} - f_{avg}} \tag{4}$$

$$b = \frac{f_{\max} - C_{mult} \cdot f_{avg}}{f_{\max} - f_{avg}} \cdot f_{avg} \tag{5}$$

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

In genetic algorithm, the multiple point crossover and weak parent replacement is applied. The GA parameters are as follows: population size 30, crossover probability, 0.85, mutation probability 0.05, coding binary string length for fault location 8, coding binary string length for fault resistance 8, r_{kv} and r_{ki} in (1) are set as one. Fault location ranges from 0 to the sum of lengths of all fault branch candidates. Fault resistance ranges from 0 to 0.4 p.u. The criterion for stopping iteration is predefined as the maximum generations number.

IV. TEST RESULTS

This section presents simulation studies utilizing the Reliant Energy HL&P transmission system. The results presented here go well beyond the initial tests results provided when the concept was introduced first [10]. The system has 4840 buses, 5895 lines, 375 generators, 3788 loads and 907 transformers. Over 30 DFRs are installed in the system.

A. Case I

Fig. 6 shows the one-line diagram in which three DFRs, located at SRB, Cedar Bayou and South Channel substation respectively, were triggered when the fault occurred. The fault occurs on 138KV system. The actual fault location, determined and provided by Reliant Energy HL&P, is on Exxon Ckt. 03, 2.5 miles from SRB 138. No taps are located in between DFR and fault location.

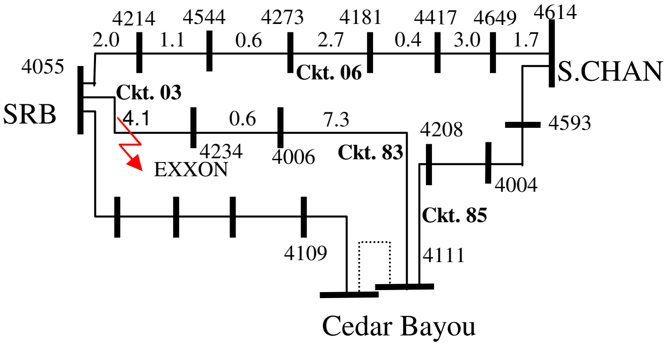

Fig. 6 One-line Diagram for case I

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; the affected circuit based on South Channel data is Tenneco poly Ckt. 06; the affected circuit based on Cedar Bayou data is Exxon Ckt. 83. Based on the information, we select different combination of these DFR files and quantities for matching shown in table 2 to estimate the fault location. Whatever DFR files are selected, the estimated results are close to the calculated fault location provided by Reliant Energy. The estimated errors are smaller than 1.0 mile. In this case, the choice of particular quantities selected for matching does not affect the result.

TABLE 2 TEST RESULT FOR CASE I

DFR Used	**Estimated fault location**	**|Error| (miles)**	**Quantities matched**
SRB	SRB-EXXON, 45% from SRB	0.6	Ckt. 03 currents
SRB Cedar Bayou	SRB-EXXON, 55% from SRB	0.2	Ckt. 03 currents Ckt. 83 currents
SRB Cedar Bayou S. CHAN	SRB-EXXON, 49% from SRB	0.5	Ckt. 03 currents Ckt. 83 currents Ckt. 06 currents
SRB Cedar Bayou S. CHAN	SRB-EXXON, 76% from SRB	0.5	Ckt. 03 voltages Ckt. 83 voltages Ckt. 06 voltages
SRB Cedar Bayou S. CHAN	SRB-EXXON, 57% from SRB	0.2	All recorded currents

B. Case II

Fig.7 shows the one-line diagram in which three DFR located at Green Bayou 138, White Oak 138 and King 345 substation were triggered when the fault occurred. The first two are located in 138KV system; the last DFR is located in 345 KV system. The fault occurs in 138KV system. The actual fault location provided by Reliant Energy HL&P is on Ckt. 21, 3.32 miles from Witter to Green Bayou.

According to the analysis report of DFR Assistant, the fault type is B to ground; the affected circuits are Hardy Ckt. 21 for Green Bayou, Hardy Ckt.95 for White Oak and North Belt Ckt. 97 for King respectively. Table 3 shows the results obtained from different combinations of the DFR files and quantities for matching. In all test cases, the three affected circuits are specified for GA search.

FL software indicates Hardy Ckt.21 as the faulted one.

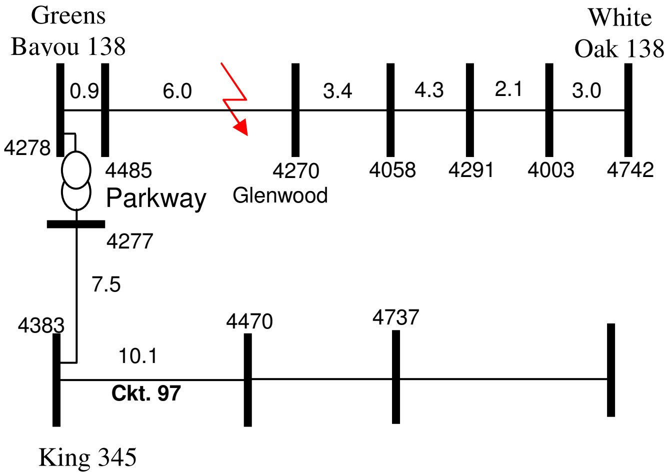

Fig.7 One-line Diagram for case II

TABLE 3 TEST RESULT FOR CASE II

DFR Used	Estimated fault location	|Error| (miles)	Quantities matched
Green Bayou	Glenwd-Parkwy, 39% from Glenwd	<0.1	All recorded currents
Green Bayou White Oak	Glenwd-Parkwy, 37% from Glenwd	0.1	All recorded currents
Green Bayou White Oak King 345	Glenwd-Parkwy, 39% from Glenwd	0.2	All recorded currents and voltages
Green Bayou King 345	Glenwd-Parkwy, 65% from Glenwd	1.5	All recorded currents
White Oak King 345	Glenwd-Parkwy, 89% from Glenwd	3.0	All recorded currents

V. CONCLUSION

The paper presents a novel fault location approach using "waveform matching" based on Genetic algorithm. The biggest advantage is to utilize the sparse data to locate fault without a need for additional 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, the known system model including the static parameters and topology is assumed.

The recently improved software package has been tested using 15 Reliant Energy HL&P fault cases. The test results show that the approach is quite promising.

VI. ACKNOWLEDGEMENTS

The authors thank the industrial advisors representing PSerc industry members: ABB, Entergy Services, Mitsubishi ITA, Reliant Energy HL&P, TXU Electric for their support. Mr. Yuan Liao and Zijad Galijasevic are recognized for their research contribution made on this subject while working as graduate students at TAMU.

VII. REFERENCES

[1] M. S. Eriksson, and G. D. Rockefeller, "An accurate fault locator with compensation for apparent reactance in the fault resistance resulting from remote-end infeed," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-104, no. 2, Feb. 1985, pp. 424-436.

[2] A. A. Girgis, D. G. Hart, and W. L. Peterson, "A new fault location technique for two- and three-terminal lines," *IEEE Trans. on Power Delivery*, vol. 7, no. 1, Jan. 1992, pp. 98 –107.

[3] D. Novosel, D. Hart, E. Udren, and J. Garitty, "Unsynchronized two terminal fault location estimation," *IEEE Trans. on Power Delivery*, vol. 11, no. 1, Jan. 1996, pp. 130-138.

[4] Z.Q.Bo, G.Weller, M.A.Redfern, "Accurate fault location technique for distribution system using fault-generated high-frequency transient voltage signals", IEE Proc.-Gener. Trandm. Distrib. Vol.146, No. 1, January 1999

[5] Test Laboratories International, Inc. "DFR Assistant, Product Description", http://www.tli-inc.com

[6] IEEE Std C37.111-1999, "IEEE Standard Common Format for Transient Data Exchange (COMTRADE) for Power Systems", Approved 18 March 1999

[7] Power Technologies, Inc., "PSS/E 26 Program Operation and Application Manuals", Dec. 1998

[8] P. D. Yehsakul, and I. Dabbaghchi, "A topology-based algorithm for tracking network connectivity," *IEEE Trans. on Power Delivery*, vol. 10, no. 1, Feb. 1995, pp. 339-346.

[9] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Addison Wesley, Reading, MA, 1989.

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

VIII. BIOGRAPHIES

Mladen Kezunovic (S'77, M'80, SM'85, F'99) received his Dipl. Ing. Degree from the University of Sarajevo, the MS and PhD degrees from the University of Kansas, all in electrical engineering, in 1974, 1977 and 1980, respectively.

He has been with Texas A&M University since 1987 where he is the Eugene E. Webb Professor and Director of Electric Power and Power Electronics Institute. His main research interests are digital simulators and simulation methods for equipment evaluation and testing as well as application of intelligent methods to control, protection and power quality monitoring. Dr. Kezunovic is a registered professional engineer in Texas and a Fellow of IEEE.

Shanshan Luo received her BS from the Taiyuan industry university, the MS and PhD degrees from the Tianjin University, all in electrical engineering, in 1984, 1993 and 1996, respectively.

She worked for chinese electrical science institute as a post doc. during 1996-1998 and for Tianjin University as a faculty during 1999-2000. She has been with Texas A&M University as a research associate since Sept. 2000. Her research interests are digital protection, communication applications in power system and substation control.

Donald R. Sevcik (M'81) received his BS degree in electrical engineering from Texas A&M University.

He has been with Reliant Energy HL&P since 1975 and is currently a consulting engineer, substation projects section. He has worked in following areas: power plant electrical systems, system studies, as well as transmission and generation protection. He is registered professional engineer in the state of Texas.