From Data Mart to Information Smart: Substation Automated Analysis Implementation

M. Kezunovic, *Fellow Member, IEEE*, S. Sternfeld, B. Fardanesh, *Senior Member, IEEE*, and T. Popovic, *Senior Member, IEEE*

***Abstract* - The paper discusses substation IED data integration and its importance for implementation of automated analysis solutions. Recorded data collected from various substation IEDs is stored into a substation data mart that utilizes standardized file formats and a database interface. The data mart provides a foundation for multiple uses of substation data. Utilities are faced with a challenge of how to bring together internal needs and requirements, available standards, and recommendations developed by professional organizations, as well as the requirements imposed externally (i.e. NERC standards). Sometimes, there is a challenge of understanding what to expect from solutions that implement all those requirements.**

In this paper a data mart implementation concept for automated integration and analysis of substation is given first. The solution and implementation requirements are discussed through realistic examples and evaluation of the solution with a variety of test data. Evaluation examples are based on both simulated fault data as well as the actual fault disturbances captured in field by substation IEDs. The presented test examples illustrate what can be expected from substation data mart based solution for automated analysis and how such a solution can be further expanded.

***Index Terms* — smart grid, substation automation, power system faults, power system monitoring, substation data mart, automated fault analysis, fault location**

I. Introduction

To achieve the goal of the smart grid, integration of substation IED data often called “data warehouse” or “data mart” is necessary. Converting data to information to allow improved decision making requires automation, which is referred to as an “information smart” approach. The paper discusses automated integration of data from substation IEDs and automated analysis of recorded events [1].

Recorded event data collected from various substation IEDs, often called non-operational data, is stored into a substation data warehouse that utilizes standardized file formats and database interface [2]. The paper considers implementation requirements to meet the internal (company) and external (oversight) level standards and recommendations (IEEE, IEC, NERC). The data warehouse or data mart allows different software modules to convert raw data to useful information allowing for integration of different applications such as disturbance detection and classification, relay and breaker operation evaluation, fault location calculation, etc.

The automation allows the use of non-operational data for operational purposes. It enables access to event and configuration data and provides variety of analysis reports and user interfaces. The open software design allows a transparent interface to applications from variety of vendors, an important feature of future information smart systems. Handling the system configuration data at device, substation, and system level is one of the biggest challenges. Being able to access configuration information that corresponds to particular IED event data for a given time stamp enables efficient automation of the analysis functions. The configuration data also helps the integration of different analysis results into a comprehensive analysis report. A good example is a two-end fault location calculation that is based on processed data stored in the data mart.

Data mart implementation concept and implementation requirements are discussed through an example solution. The solution has been evaluated using data from both simulations and field recorded fault cases.

II. Data Mart: Implementation Concept

Dealing with data coming from IEDs from different vendors typically implies use of vendor-specific data communication and collection software. It is not uncommon that a single vendor may have multiple software packages for different equipment vintages. In most cases it is possible to configure IEDs for automated data retrieval and to assume that within a certain time frame all newly recorded event files will be downloaded and available on a file server on a corporate network. NERC requirements for handling substation disturbance data as well as for cyber security should be consulted for proper implementation of communication and data collection [3,4]. An important issue is that all substation event data should be properly time stamped so that all of

Mladen Kezunovic is with Texas A&M University, College Station, Texas, USA (e-mail: kezunov@ece.tamu.edu)

Scott Sternfeld is with FirstEnergy, Akron, OH, USA (e-mail: ssternfeld@firstenergy.com)

Bruce Fardanesh is with New York Power Administration (NYPA), White Plains, New York, and New York and Manhattan College, New York, USA

Tomo Popovic is with Test Laboratories International, Inc. (TLI), College Station, Texas, USA (e-mail: tomo@tli-inc.com)

the event records can be sorted in historical orders as well as grouped into clusters of event files that correspond to a single event [5].

Having the data collection process fully automated and making IED data available on the file server is only the first step for proper data integration. The second step is to bring all the data into same format and make sure that a quick search and retrieval is possible. This process typically includes file conversion into unified non-proprietary standardized file format [6,7] and also require knowledge of current system configuration [2]. The file conversion process for each type of IED can also be accompanied with device level analysis functions for more efficient implementation.

A concept for substation automated analysis implementation based on data mart is depicted in Fig. 1. Substation IED data such as event records from digital fault recorders (DFRs), digital protective relays (DPRs), and other types are automatically collected and made available to the solution [8].

The solution consists of:

- Processing and Analysis Client
- Data Manager
- Substation Data Mart (Data Warehouse)
- User interface

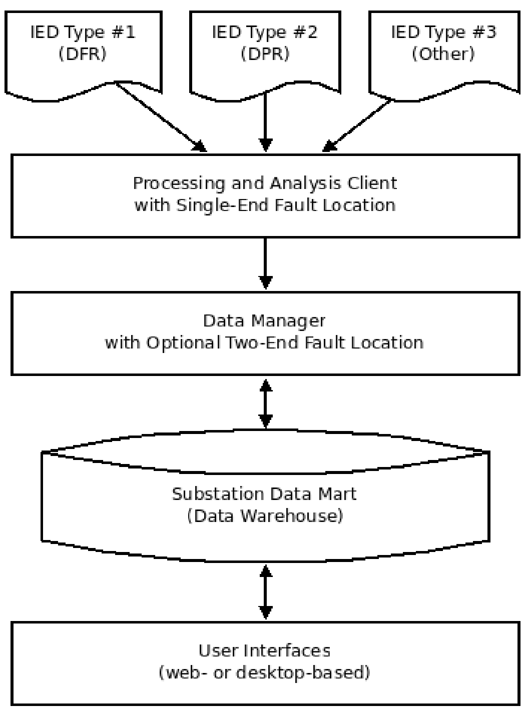


Fig. 1. Data mart based solution

A. Processing and Analysis Client

All the IED records are being automatically converted into a unified non-proprietary file format [6,7]. Signal waveforms, namely voltages and currents, are being mapped to monitored power system components (transmission lines, transformers) and the processing includes fault type detection, disturbance classification, and single-end fault location calculation [9]. There could also be a processing and analysis software module for each IED type group (DFR, DPR, CBR, etc.) and these modules typically correspond to IED level functionalities described in the previous section.

B. Data Manager

The Data Manager is responsible for broadcasting reports and proper maintenance of the data warehouse and centralized repository. In addition, it provides functions for broadcasting notifications to selected users via email and/or pager service. Optionally, the two-end fault location function can be attached to the data manager. The module demonstrated later in the paper utilizes phasor-based algorithm for two-end fault location calculation [9,10]. Every time the data manager receives a new event file from one end of the line, it checks the data warehouse for an event file from the remote end that falls into predefined time frame. If such a file is located, the two-end calculation is executed.

C. Substation Data Mart (Data Warehouse)

All the converted event data, analysis reports, and system configuration are kept in the data warehouse. The data warehouse implementation utilizes a standard database engine and a centralized file repository. Keeping the event data in non-proprietary standard file format [7,8] enables direct access and use of fault data even with standard file managers such as Windows Explorer or Nautilus. The data base implementation should utilize standard database interfaces such as the SQL subset that will work with most database engines [11]. The solution discussed implements the database interface using Java technologies and JDBC, which allows easy interfacing to a majority of modern database engines [12].

D. User Interface

The main user interface is web-based and enables users to access data warehouse utilizing just a standard web browser (for example, Internet Explorer or FireFox). The layout is displayed in Fig. 2 below. It is important to note that the interface is using only standard and "clean" HTML/CSS technology [13].

For more elaborate user interfaces (not only web-based), there are variety of technologies available. In this solution, for example, one of the latest Java technologies, Web Start, is used for implementing the waveform and report viewer (also shown in Fig. 2) that runs as a desktop application on a client computer, but starts via web server and does not need to be installed locally on the user's computer [12].

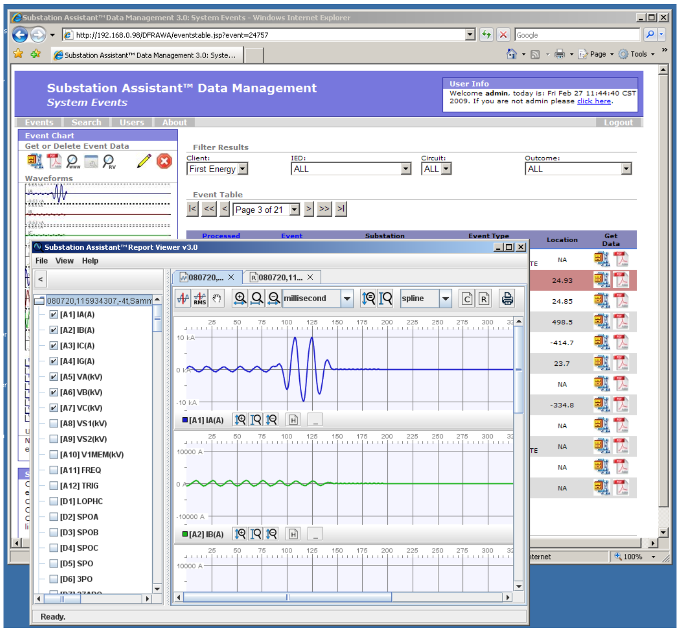


Fig. 2. Web and desktop based user interfaces

III. In-House Evaluation

The described solution has been implemented and currently being deployed in New York Power Authority (NYPA) and FirstEnergy (FE). The initial evaluation has been done using historical IED data (DFR and DPR data), but in order to provide additional "good" data sets, several fault events have been simulated using an ATP model (Fig. 3) [14].

Four different types of faults have been simulated: a) phase-to-ground, b) phase-to-phase, c) phase-to-phase-to-ground, and d) three-phase (Table I below). Each fault type has been simulated at 50, 60, 70, 80, and 90 percent from one end of the line (50, 40, 30, 20, 10 as "seen" from the other end respectively) .

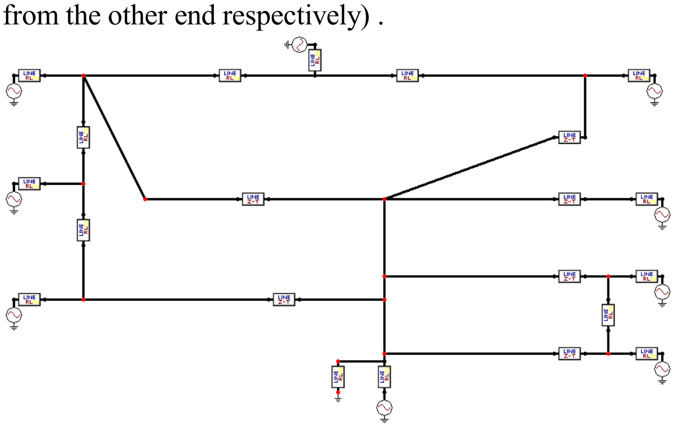


Fig. 3. ATP model used for in-house evaluation

The ATP output PL4 files have been exported automatically into COMTRADE using digital simulator software and a batch generator (BGEN) simulator tool [15]. Each simulation step produced test data quantities from both ends of the simulated transmission line. The time stamps inside generated COMTRADE files were adjusted to allow pairing of simulated two end fault data, thus allowing automated analysis software to recognize those event files as if they were coming from a real system.

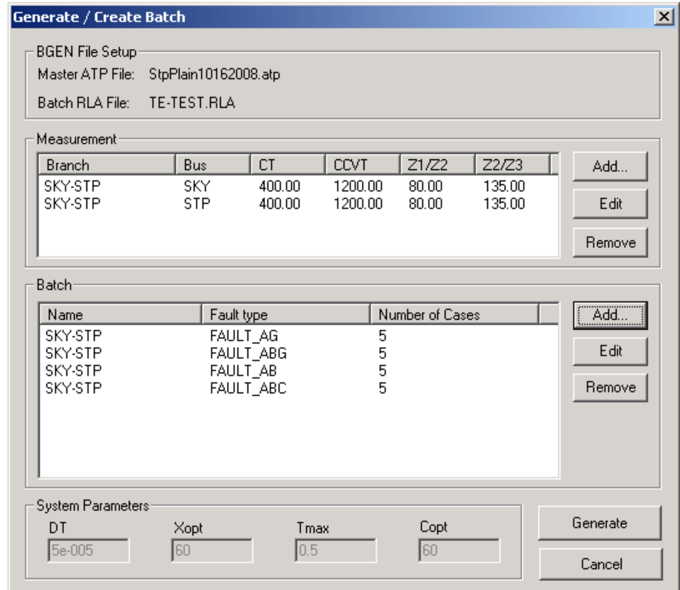


Fig. 4. Use of the batch generator tool for fault simulation

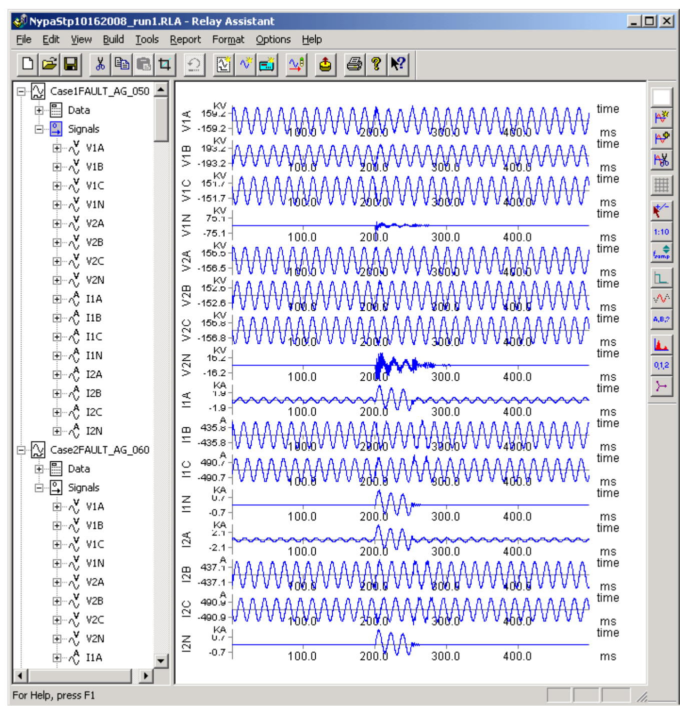


Fig. 5. Exporting simulated events to COMTRADE

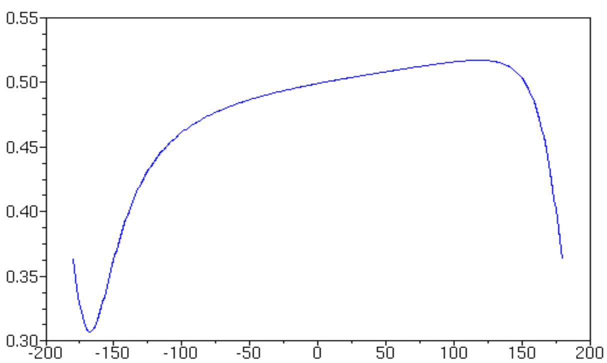
All of the simulated IED data was processed automatically with correct fault type and classification outcome. The testing provided good insight into single-and two-end fault location algorithm implementation as well (Table I).

TABLE I
IN-HOUSE EVALUATION RESULTS

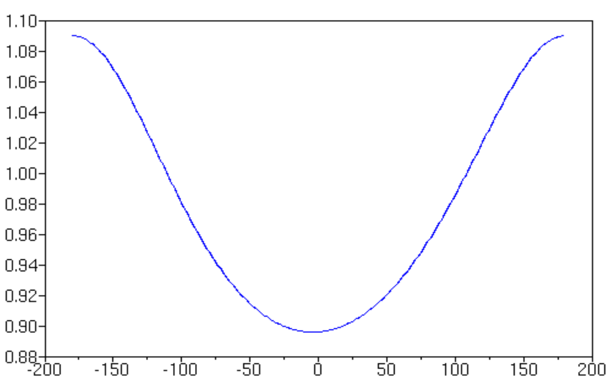
#	Fault Type	Number of cases	Single-end Error [%]	Two-End Error [%]
1	A-G	10	0.61 - 3.75	0.05 - 0.27
2	AB	10	0.59 - 2.57	0.01 – 0.48
3	AB-G	10	0.61 -2.57	0.05 – 0.21
4	ABC	10	0.42 – 2.57	0.05 - 0.44

Note: error % relative to line length

As expected, the two-end fault location calculation displayed better accuracy for the simulated test data. For successful two-end data analysis it is crucial that the IED data coming from two ends of the same line are properly synchronized. This can be a challenge even when GPS clocks are being used since the devices on both ends may be different IED types, versions, or vintage.



a) A-G fault at 0.50 (50%)



b) A-G fault at 0.90 (90%)

Fig. 6. Two-end fault location vs. angle difference

A simple "sensitivity study" with the simulated data has been done to emulate two end fault data that is not fully synchronized. With this set of simulated data it has been identified that the implemented phasor-based two-end fault location calculation was providing good estimates even when the angle difference between phasors taken from two ends was 30 degrees (Fig. 6) . However, each case may have its own sensitivity curve and these observations were done primarily so that the authors get "a feel" for what happens if there is a lack of excellent synchronization between IED data obtained from two ends.

IV. FIELD DATA EVALUATION

Field data examples discussed in this section are obtained from an actual solution installation (Fig. 7). IED data are collected from two substations (#1 and #2): a) one DFR and seven DPRs; and b) one DFR and one DPR (monitoring line towards substation #1) in substations #1 and #2 respectively. The distance between the two substations is around 49 miles.

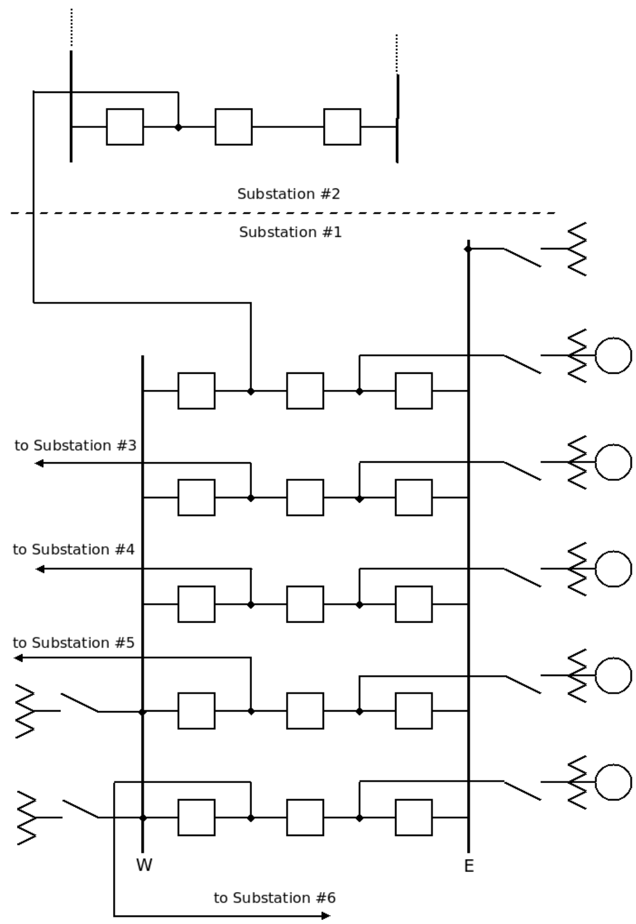


Fig. 7. Field setup configuration

Around 250 IED records have been processed and in four instances there were nice examples of the same fault or disturbance event being captured by multiple IEDs.

A. Field Case One

In this case the actual fault was located on the transmission line between substations #1 and #4 at 23.7

miles from substation #1. The fault event was captured by six IEDs (Table II) including primary and backup distance relays on that line and DFRs in both substations. All the IED data was properly integrated. DPR data is being converted into same revision of COMTRADE file format and DPR reports are automatically parsed for fault information. Both relays on the faulted line tripped and calculated fault at around 24.9 miles. DFR data was automatically converted and analyzed by the automated processing function that correctly calculated the actual fault location at 23.7 miles. It is important to note that the data from other relays, although not providing the fault location, can be useful for system protection engineers to get a better assessment of what happened.

TABLE II
FIELD DATA EVALUATION – CASE ONE

#	IED/ Line	Time Stamp	Fault Type	Fault Loc
1	DPR Line Primary Sub #1 to #4	11:59.34.302	A-G	24.85
2	DPR Line Backup Sub #1 to #4	11:59.34.307	A-G	24.93
3	DPR Line Primary Sub #1 to #2	11:59.34.343	A-G	-414.7
4	DPR Line Primary Sub #2 to #1	11:59.34.305	A-G	498.5
5	DFR at line Sub #1 to #4	11:59.34.207	A-G	23.7
6	DFR at line Sub #2 to #1	11:59.34.295	Not Fault	N/A

B. *Field Case Two*

The actual fault was B-G on the transmission line between substations #1 and #6 just outside substation #1 (Table III).

TABLE III
FIELD DATA EVALUATION – CASE TWO

#	IED/ Line	Time Stamp	Fault Type	Fault Loc
1	DPR Line Primary Sub #1 to #2	07:15:02.825	B-G	-7.49
2	DPR Line Primary Sub #1 to #4	07:15:02.826	B-G	-4.32
3	DPR Line Primary Sub #2 to #1	07:15:02.782	B-G	49.25
4	DFR at line Sub #1 to #6	07:15:02.704	B-G	0.2

The data in the third and fourth row in Table III correctly points the actual fault location, while other two relays on lines towards substations #2 and #4 correctly "saw" the fault behind (inverse fault).

C. *Field Case Three*

The actual fault was B-G, but beyond the transmission line between substations #1 and #4. Both the relay calculation and DFR analysis calculation "saw" the fault at the same location (Table IV).

TABLE IV
FIELD DATA EVALUATION – CASE THREE

#	IED/ Line	Time Stamp	Fault Type	Fault Loc
1	DPR Line Primary Sub #1 to #4	19:55:44.772	B-G	63.7
2	DPR Line Backup Sub #1 to #3	19:55:44.933	B-G	375.5
3	DFR at line Sub #1 to #4	19:55:44.925	B-G	63.8

D. *Field Case Four*

The actual fault was C-G on the transmission line between substations #1 and #3 at around 8 miles from substation #1 (Table V).

TABLE V
FIELD DATA EVALUATION – CASE FOUR

#	IED/ Line	Time Stamp	Fault Type	Fault Loc
1	DPR Line Primary Sub #1 to #4	03:05:50.833	C-G	NA
2	DPR Line Backup Sub #1 to #3	03:05:50.796	C-G	8.06
3	DPR Line Primary Sub #2 to #1	03:05:50.797	C-G	146.3
4	DFR at line Sub #1 to #3	03:05:50.806	C-G	8.4

V. CONCLUSIONS

The paper discusses substation IED data integration and its importance for implementation of automated analysis solutions. The following is being pointed out:

- Solution specifications have to meet requirements defined by a given utility, available standards and recommendations developed by professional organizations (IEEE, IEC, etc.), and the external requirements such as standards defined by NERC.
- Communication with substation IEDs should be configured to enable automated and quick event data collection while conforming to internal and external networking security standards and polices.
- Automated data integration in a form of substation data mart is a foundation for an automated fault analysis solution. The data warehouse should conform to available data format standards (IEEE, IEC) and provide an interface for easy data access and retrieval.

- Testing and evaluation of data mart based solutions for automated analysis may benefit from use of advanced tools for fault simulation and test waveform generation.
- Benefits and "what to expect" discussion was illustrated through several simulated and field obtained fault data cases. Presented examples focus on fault type detection and fault location calculation, but should provide the readers with ideas on how such solutions can be customized and grow to meet needs of different utilities and variety of user groups.

VI. Acknowledgments

Authors would like to thank several individuals involved in current solution deployment projects:

Mr. Paul Myrda from Electric Power Research Institute (EPRI);

Mr. Robert Foster from FirstEnergy, Akron, Ohio;

Mr. Deepak Maragal from New York Power Authority (NYPA), White Plains, New York;

Mr. Matthew Kuhn from Test Laboratories International, Inc. (TLI).

VII. References

[1] J. D. McDonald, *"Substation Automation, IED integration and availability of information"*, IEEE Power&Energy, Vol. 1, No. 2, March/April 2003.

[2] M. Kezunovic, T. Popovic, *"Substation Data Integration for Automated Data Analysis Systems"*, IEEE PES 2007 General Meeting, Tampa, Florida, June 2007

[3] *NERC Standard, Protection and Control,PRC-002-1 and PRC-018-1*, North American Electric Reliability Corp., 2006-2007

[4] *NERC Standard, Cyber Security, CIP-002 through CIP-009*, North American Electric Reliability Corp., 2008.

[5] Global Positioning System, Wikipedia, [online]: http://en.wikipedia.org/wiki/Global_Positioning_System

[6] *IEEE Standard Common Format for Transient Data Exchange*, IEEE Standard C37.111-1999, 1999.

[7] *IEEE Recommended Practice for Naming Time Sequence Data Files,* IEEE PC37.232-2007, 2007.

[8] Substation Assistant™: *"Data Integration and Automated Analysis"*, [online]: http://www.tli-inc.com

[9] *IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines*, IEEE Standard C37.114-2004, 2004.

[10] J. G. Webster, et al., *Wiley Encyclopedia of Electrical and Electronics Engineering*, A Wiley-Interscience Publications, John Wiley and Sons, Inc., 1999.

[11] Database Language SQL, ISO/IEC 9075, [online]: http://www.iso.org

[12] Java programming language and related technologies [online]: http://java.sun.com

[13] World Wide Web Consortium, HTML and CSS specification, [online]: http://www.w3.org/

[14] Another Transient Program, [online]: http://www.emtp.org/

[15] M. Kezunovic, T. Popovic, D. Sevcik, H. DoCarmo, *"Transient Testing of Protection Relays: Results, Methodology and Tools",* International Conference on Power System Transients - IPST 2003, New Orleans, Louisiana, September-October 2003.

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 Site Director of Power Engineering Research Center (PSerc), and NSF I/UCRC. 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 system monitoring. Dr. Kezunovic is a registered professional engineer in Texas and a Fellow of the IEEE.

Scott Sternfeld received his B.S. in Mechanical Engineering from the University of Illinois, Urbana-Champaign in 2001.

He worked for the Naval Surface Warfare Center, Carderock Division from 2003-2007 as the Towed Array Test Equipment Acquisition Engineering Agent for submarine sonar systems.

Since 2007, he has been working for FirstEnergy on the Integrated Grid Communications and Automation (IGCA) initiative as part of the Energy Efficiency department, where he is currently a Senior Engineer. His role has been to coordinate and facilitate various data integration projects, as well as modeling Smart Grid interfaces and applications in both laboratory and field environments.

Bruce Fardanesh received his B.S. in Electrical Engineering from Sharif University of Technology in Tehran, Iran in 1979, and his M.S. and Doctor of Engineering degree both in Electrical Engineering from the University of Missouri-Rolla and Cleveland State University in 1981 and 1985 respectively.

Since 1985 he has been teaching at Manhattan College where he holds the rank of Associate Professor of Electrical Engineering. Currently, he is working as a Program Manager at the New York Power Authority. His areas of interest are power system dynamics, control and operation. Mr. Fardanesh is Senior Member of the IEEE.

Tomo Popovic (M'99, S'06) received his BS degree in electrical engineering in 1994 from University of Novi Sad, Serbia and his MS degree from Texas A&M University, USA.

He has been with Test Laboratories International Inc. since 1998 where he is a Sr. Systems Engineer. His prior positions were with NIS-GAS, part of Petroleum Industry of Serbia, and University of Novi Sad, both in Novi Sad, Serbia. His main professional interest is developing and implementing software and hardware solutions for industrial applications, especially in the field of electric power system engineering: substation automation, analysis of fault records, transient testing of protective relays and digital simulators. Mr. Popovic is Senior Member of the IEEE.