Developing Methods for the Detection of High Impedance Faults in Distribution Power Grids

Tomasz Nowicki
IBM Watson Research Center
1101 Kitchawan Road
Yorktown Heights NY 10598, USA
Email: tnowicki@us.ibm.com

Grzegorz Świrszcz
IBM Watson Research Center
1101 Kitchawan Road
Yorktown Heights NY 10598, USA
Email: swirszcz@us.ibm.com

Mark Yao
IBM Watson Research Center
1101 Kitchawan Road
Yorktown Heights NY 10598, USA
Email: markyao@us.ibm.com

***Abstract*—High Impedance Faults are known to be extremely hard to detect. At the same time the efficient detection of them is very important to electric power providers. One of the challenges belonging to a Smart Grid paradigm is how to use new data not available before to improve methodology of High Impedance Fault detection. In the paper we describe our experiences, insights and lessons learnt from actual field experiments. A series of experiments was conducted that simulated most common High Impedance Faults. The data was recorded at 256 samples per cycle at 4 stations located at different places on a feeder. We describe the experiments, the process of acquiring the data and how we used it. Also, we discuss challenges we encountered and what we have learned in the process.**

***Keywords*-High Impedance Fault; PMU; sensors; time series; experiments;**

I. INTRODUCTION

"High Impedance Faults (HIFs) on distribution systems create unique challenges for the protection engineer. HIFs that occur do not produce enough fault current to be detectable by conventional overcurrent relays or fuses." [10]. The difficulty in detecting High Impedance Faults is due mainly to two facts. Firstly they generate weak signal, very often resembling a change in the load caused by a legitimate usage. Secondly many diverse phenomena with very diverse physical properties are classified as HIFs. A downed line on a non-conducting surface (continuous flow), tree branch tapping on a wire (erratic, random series of faults depending on a wind of a high frequency), cracked insulator that causes HIF whenever water fills the cracks due to humidity (very weak flow, semi random nature, low frequency) are very different events all falling into the category of HIFs.

At the same time HIFs pose a threat to human life, to the equipment and can lead to high costs for the provider if not detected in time. Thus developing methods of detection of HIFs is an item of high priority. Many years of experience show that it is practically impossible to detect HIFs using conventional methods and on the level of substations. Therefore HIF detection fits perfectly as an essential component into the Smart Grid paradigm.

Currently HIFs are often detected using methods like breakers tripping, readout from meters at the substation (human) or a phone call from someone who noticed a fault. Needless to say, accuracy of those methods is highly insufficient and not acceptable in a modern power grid. The concept of Smart Grid provides an access to data that was not available neither in such quantity (number of measurement points, sampling rates), nor quality (accuracy of in situ sensors, resolution, time stamps, cross validation) nor in data management (transmission, preprocessing). This data can be used for HIF detection. As with every new paradigm this leads to new challenges. There is no sufficient textbook knowledge about how to use readouts of multiple sensors on the level of multiple samples per cycle to detect faults. Industrial standards how to use such new data are not yet established and the users are not yet trained. There are several approaches to model HIFs using neural networks, wavelets, Fourier transforms [2], [9], [3], [6], [8], [7] none of these models being satisfactory nor becoming a standard. Often such models are only partially validated by simulations and not field experiments. In general HIFs are not yet sufficiently well understood and characterized neither from theoretical nor practical point of view. Advances in physics and establishing technological methodology and procedures require experiments and HIF detection is not an exception.

We had a unique opportunity to actually be able to perform a series of field experiments that provided us with real (not simulated) data. In this paper we describe our experiences, insights and the lessons we have learnt.

The paper is organized as follows. In Section II we describe the setup of the experiments we performed, the equipment used and how the data was recorded. In Section III we discuss the outcomes of the experiments. Finally, in Section IV we present our insights and conclusions from the work we did.

II. EXPERIMENTS

The experiments were performed thanks to a collaboration with a customer of IBM - a major electric energy provider. In this section we present an overview of what was done

 ISBN: 978-1-61208-136-6

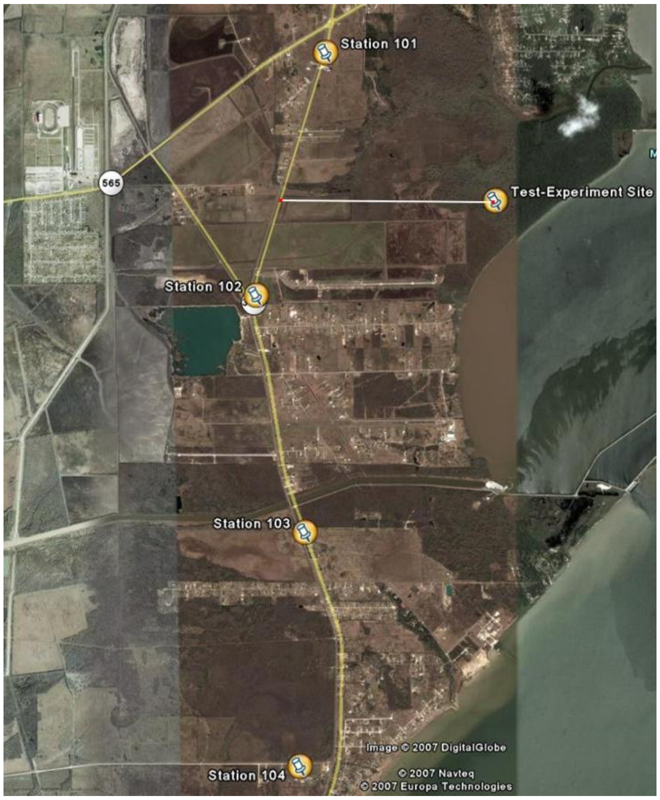

Figure 1. Experiment site and 4 recording stations.

focusing on the general setup, equipment used, experiments performed and data collection and transmission.

A. Experiment setup

The experiments were performed on a lateral that was under construction and its end was in a remote location that could be easily secured. That way safety measures were met and generated faults did not cause interruption in the delivery of the electric power. The main feeder was leading to a residential area, thanks to which the requirement of having HIFs generated on top of a real load was met.

The feeder had a standard 4-wire configuration with grounded neutral wire. A wire was connected to phase B that was used to generate various types of faults. Due to security reasons experiments could not be 100% realistic - for example throwing a powered wire on the ground would pose too much danger for the crew. What was done instead - there was a switch installed on phase B before the wire used for the experiments. All the time except for when the trials were done the switch was in an "off" position. The wire was secured in a position like on a block of concrete or on wet grass. Then the switch would be moved to "on" position and the fault would be generated. See IV for insights and remarks related to the influence of that aspect of a setup on the data.

There were four recording stations named 101, 102, 103 and 104 (see Fig. 1) located on the main feeder.

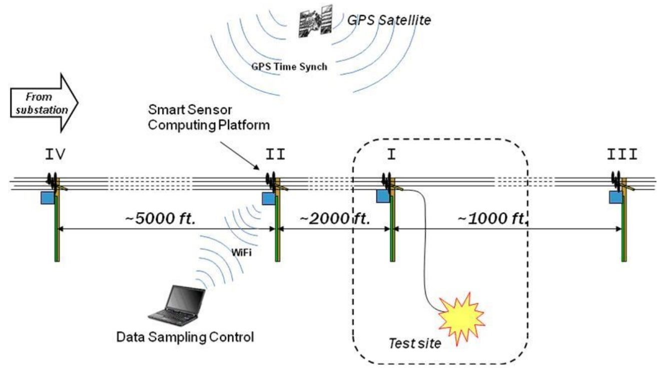

Figure 2. Experiment setup.

Station 101 was located downstream from the experiment site, stations 102, 103 and 104 were upstream. The distances from the experiment site to recording substations were

experiment site $\leftrightarrow$ station 101 = 2.38 miles,
experiment site $\leftrightarrow$ station 102 = 2.07 miles,
experiment site $\leftrightarrow$ station 103 = 3.66 miles,
experiment site $\leftrightarrow$ station 104 = 5.44 miles.

The signal was sampled 256*60 times per second. On each of four wires both current and voltage were recorded making a total of 8 channels recorded per station.

B. Equipment used

Each of four sampling stations selected along the distribution feeder had following hardware equipments and devices installed on the electric pole:

- Four pre-calibrated Lindsey Current & Voltage sensor for 3 line phase plus neutral line.
- Sensor wire terminator and power supply (for RTU) with enclosure.
- Intelligent RTU with following components in a weather proof enclosure.

Key hardware components and their functions within each intelligent RTU were

- **High speed A/D board** A 16-Bit Analog I/O board was selected for its compatibility with underlying system board and for its high speed data sampling rate to meet experiment requirement. This I/O board can provide 250,000 samples per second maximum sampling rate and, most importantly, 1024-sample FIFO for reduced interrupt overhead.
- **GPS** A GPS card was used to provide high resolution reference-time signal to synchronize all distributed RTU in time. Time synchronization was a critical data collection requirement to allow accurate time coordination for later data analysis and modeling. In the

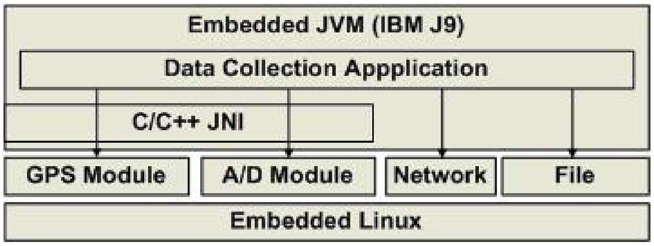

Figure 3. Data Collection Application.

experiment, the resolution of time stamp for the data record was set to micro-second.

- **Wifi Card** At the time of setting up the experiment there was no wide area communication link between RTU and workstation that would configure and control the RTU and the test site. Once the RTU was mounted at pole top, the communication interface available to control and configure the RTU was through either wired serial line communication with system terminal program or through Wifi communication through socket-based application program. The capability of control/configuration of RTU at runtime was critical to the experiment. For example, we turned on the data collection process through Wifi network at each RTU location only when the experiment had been ready to be conducted so the data storage would not overflow.
- **Storage card** An 8 GB USB storage card was used as a local storage for all data collected during the experiment time. Given the rate of data collection (256x60x8 samples/per second), the storage had a capacity to accommodate 4-5 hour time window of experiment.

The RTU system was an ultra-low power, high-performance, single board computer based on the 520 MHz PXA270 processor. The PXA270 was an implementation of the Intel XScal micro architecture, combined with a comprehensive set of integrated peripherals, including interrupt controller, real time clock and various serial/USB interfaces, and network interfaces. Fig. 3 displays the software stack installed for each of 4 intelligent RTUs. The data collection application was implemented in JAVA programming language. Once the experiment setting was ready, the application initialized and configured A/D board and on timer (through GPS module) based on pre-loaded configuration file, started the data collection process through hardware interrupt setting and serialized the data with pre-defined format to storage files. The application could be started, stopped and reconfigured through local Wifi communication interface.

C. Experiments performed

Experiments were performed in 9 categories corresponding to 9 most common types of High Impedance Faults.

1) Concrete Drops
2a) Sand Drops (Dry Sand)
2b) Sand Drops (Wet Sand)
3) Asphalt Drops
4) Grass/Soil Drops (note: grass and soil were very wet)
5) Capacitor Switching (Operate capacitor to see switching transient)
6) Tree Taps
7) Axial Pinhole Insulator (salt water sprayed into pinhole drilled in the insulator)
8) Dirty Insulator (Observations: No obvious activity, no sound)

It should be noted that Experiment 5) consisted of switching capacitor banks on and off, therefore it was not a simulation of a HIF but of a legitimate activity that is know to be a source of false positives.

In each of the categories 10 trials were planned to be performed, but in numerous cases less trials were done because of observed phenomena of the sort of:

Too high current flow that was causing breakers to trip (in case of wet soil and grass).
Difficulty in conducting the experiment and safety issues (in cases of Tree Tap, Axial Pinhole Insulator).
No observed activity (Dirty Insulator).
High repeatability (Capacitor switching).

Every trial was described by (manually) recording:

TRIAL NO: (*for example* 1)
Start Time: (*for example* 9:59:19)
Finish Time: (*for example* 9:59:30)
Arcing Visible: Y/N
Observations: (*for example* Small arc on surface)

D. Data recording and transmission

RTUs were recording the data in a custom-designed format (see Table I). The data for each channel was recorded in an interleaved way. This format was used for recording purposes. For working with the data we designed a format basing on .wav standard. Pure .wav format was not sufficient for our purposes as it does not contain information about timestamps containing information about when the data was recorded. We used a format DataStream that consisted of .wav files and a text file .cfg containing metadata such as time of the beginning of recording, time of recording end, Station ID, Phase, Current/Voltage and possible comments and remarks. Each DataStream could be built of one or multiple .wav files.

One of the main challenges we were facing was the proper synchronization of the equipment. For most part the use of GPS was the solution, but there still remained an issue of mapping the experiments happening to timestamps recorded in the data (see IV). We used a two-way approach. Our first source of data were manually made notes by an engineer. We used his handwatch that was synchronized with GPS

Table I
RTU RECORDING FORMAT

Block	Field	Size	Comment
Header block	sample_station_id	4 byte	Data collection station unit id
	sample_size	2 bytes	Data sample unit size, in number of bytes, for each data sampling. 2 bytes for this project
	sample_edian	2 bytes	1 for little and 0 for big endian
	chan_start_num	2 bytes	Starting data channel number
	chan_end_num	2 bytes	Ending data channel number
	chan_scan_rate	4 bytes	Data channel scan rate in Hz. Each scan converts all channel data specified by start and end channel number
	chan_time_interval	4 bytes	Time interval between each channel data in micro-seconds (default 4 micro-sec)
	time_stamp_sec	4 bytes	Time stamp in sec when first data scan happened
	time_stamp_usec	4 bytes	Time stamp in micro-seconds for when first data scan happened
Data block	<start_chan_data> ... <end_chan_data> <start_chan_data> ... <end_chan_data>		

timers before the experiments to properly record times and short descriptions of each experiment (see II-C). The second source of data was a video recording of the experiments with a timestamp of the recording embedded in each frame. We did not want to rely on any of those sources solely. The reason was that if something went wrong with one of them the experimental data would lose most of it's value without a way of identifying its portions corresponding to experiment versus the ones corresponding to faultless state of the feeder. Luckily we ended up having two good sources of that data, nevertheless we are convinced that having redundant sources of information is the right approach if possible.

It is worth mentioning that the manual way of recording experiments turned out to be surprisingly accurate. It also

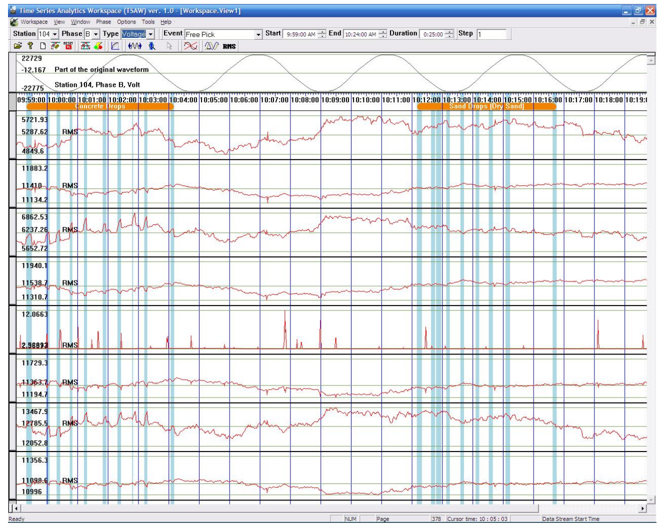

Figure 4. Current and voltage waveforms recorded on Phase B at stations 101, 102, 104 and 104 respectively (top to bottom).

could not have been eliminated completely as we needed not only times but also descriptions of experiments.

III. AFTER THE EXPERIMENTS

A. Experimental data

A primary database of labeled waveforms was created, organized into 4 main folders corresponding to 4 recording stations. It contained over 200 files. A consistent naming convention for the files was used and they were arranged by phase, type (Voltage/Current) and time.

A secondary database of labeled waveforms was made of 1720 wavefiles in 261 subfolders. Files in .wav format were accompanied by configuration files containing metadata.

A list of experiment descriptions (from handwritten notes) was used to generate DataStream objects. The format has been defined that can be used as an input by the code to populate data structures appropriately, namely .cfg files for storing metadata of the analytics/development modules allowing the access to multiple wavefiles as a continuous data stream (compare II-D).

Over 12 Gigabytes of data have been collected corresponding to over 3.5 hours of experiments.

B. ...and what we did with it

With the amount of data we were facing standard tools for manipulating and visualizing time series were inadequate (Compare lesson eleven in IV). Envisioning upcoming paradigm of Smart Grid and related challenges we have developed a platform called Time Series Analytics Workspace (TSAW).

TSAW is a software environment for working with multiple large-size time series. It combines various visualization tools (see Fig.s 4, 6), manipulation utilities allowing various

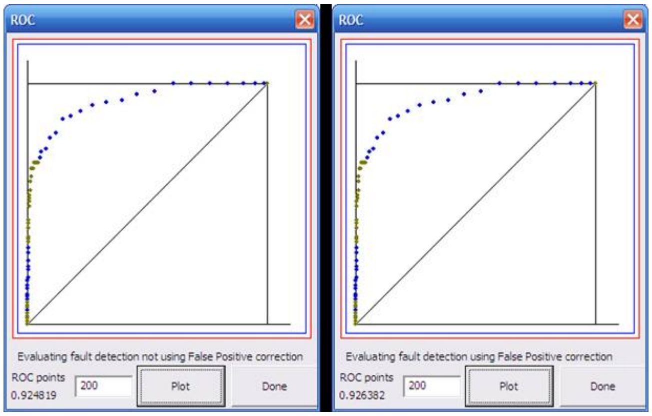

Figure 5. ROCs obtained on the experimental data using novel algorithms.

ways of working with the data and powerful analytics inbuilt. TSAW has been designed and built with broad array of applications in Smart Grid related projects in mind, not limited to the scope of HIF detection.

Thanks to working with the data using TSAW numerous properties of HIFs were understood. 6 new invention disclosures (see [4], [5]) were filed related to the work done and innovative algorithms were created based on the knowledge acquired. The performance of the algorithms was evaluated using the standard metric - AuC (see for example [11]). Performance efficiency yielding AuC of 0.924 has been achieved using anomaly detection combined with false positive elimination algorithm. (see Fig. 5).

IV. Lessons learnt

The first lesson learnt would be the confirmation of how great the importance of real experiments is. This was very much expected, but we still find this insight to be of an extreme value.

The second lesson we learnt was that the choice of recording data format was not the best possible. In one of the files one of the AD converter drivers inserted one 0 sample at the beginning of data. As the data was recorded in an interleaved fashion this lead to shifting of channels - the first channel became the second etc., so for example current of phase A was being recorded where the data for voltage of phase B should have been. This was discovered when gluing waveforms coming from different files showed discontinuity. Enhancing the recording format with markers containing the information about the data being recorded inside the channel and/or writing each channel to a different file might be a good idea to be considered.

Third lesson learnt was that data recording can be more efficient and economic if a remote communication was available between the test site and all RTU data recording station. The total time of data recording time for the experiment lasted about 4-5 hours, while each experiment conducted was only lasted about 5-10 minutes. Most of data recorded was within non-experiment time window and was irrelevant to the later data analysis and wasted storage space. The non-experimental data also added additional work to the post experimental work since experimental data need to be allocated by time filtering out these non-experiment-data. All these disadvantages can be resolved if a remote communication to each RTU was available from the test site so that each experiment can recorded separately with different time reference point. It is with pointing out though that some amount of non-experimental data is as important and the experimental data, as it allows to train models properly providing a baseline for true negatives.

Lesson four - installation of physical RTU units should meet following requirements:

a) Security: the unit should be sealed with a secure lock and installed at certain height to prevent any theft and unwanted access. The installation height should be also convenient enough so that unit could be physically reached by the experimenter for different purposes such as physical repair or removal if the USB data storage after experiments.
b) Weather proof: one of RTU was damaged due to leakage of rain after the experiment and this can be prevented if enclosure is well sealed.

Lesson five - the cost of installation of sensor and RTU can be high. With a crew of 2 workers, it took 1.5 full days to install and test 4 units of RTU and wire them to the pre-installed sensors.

Lesson six - our experiments confirmed that real (not only simulated) experimental data is essential to progress in physics in general and to get an understanding of the nature of HIFs in particular. Smart Grid and access to the readouts of meters capturing what phenomena accompany real events will provide whole new universe of knowledge that will clarify many unknowns and misconceptions about HIFs. Thanks to the access to the real data we were able to determine that on a contrary to a common belief that phase shifts should accompany HIFs it is not the case. We have established that not only this is not true in most cases, but also that the occurrence of phase shift provides an excellent tool for eliminating false positives [5].

Lesson seven - if we knew the exact setup of experiments ahead of time we could have attached a device to the switch starting/ending each trial (see II-A to increase the accuracy of recording experiments beginnings/endings (as compared to the handwatch we used).

Lesson eight - even best experiments are still only experiments (compare II-A) and need to be treated as such. They cannot either replace or be replace by a real data from the Smart Grid. The benefit of experimental data is having the full information of what happened and exact timestamps of

 ISBN: 978-1-61208-136-6

what happened. The drawback is that experiments cannot be 100% realistic for numerous reasons such as for example the fact that due to related hazard it would not be feasible to break a wire and make it fall on the ground, which is one very common scenario of HIF occurring. Such data will become available once Smart Grid starts becoming a reality. Also, equipment used during experiments (like the switch) generates a signal of its own that is not present in a real life scenario. All those aspects have to be taken under consideration while working with the data and building models.

Lesson nine - it would have been of a great value to have had a sensor station installed also on the lateral and closer to the experiment site. It was not possible during our experiments due to technical reasons and we had at out disposal only the HIF-generated signal mixed with a strong signal on the main feeder. Once Smart Grid becomes a reality such data will become available and we are very thrilled about the prospect of working with it.

Lesson ten - our experiments show consistency with the results of a paper [1] about determining the direction to a flicker source. At the Fig. 6 the first two rows are current and voltage at phase B recorded at station 101 which was located downstream from the experiments. Both current and voltage change in the same direction when the fault is generated. The last two rows are current and voltage at phase B recorded at station 102 which was located upstream from the experiments. Their perturbations have different directions, as described in [1].

Lesson eleven - tools for visualization of the data need to be be developed. With numerous sensors generating streams of data the need for an ability to display the readouts in a human-readable manner becomes essential. The challenge is to develop means of efficiently visualizing enormous amounts of data, orders of magnitude above what a typical software is able to handle effectively. TSAW (see III-B) is an example of such tool that allowed us to work effectively with the data.

ACKNOWLEDGMENT

The authors would like to thank Ron Ambrosio, Charlie Arteaga and members of the crew at the experiments site for their work and help with this project.

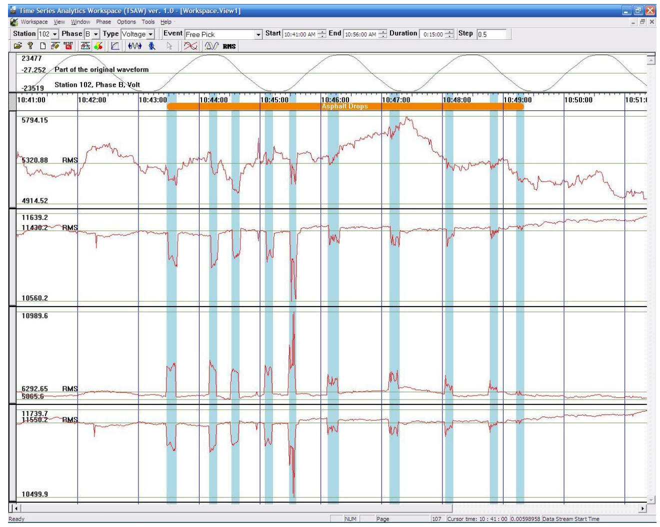

Figure 6. Effect of direction to the source on current and voltage - top to bottom: I at 101, V at 101 (downstream), I at 102, V at 102 (upstream).

REFERENCES

[1] P. G. H. Axelberg and M. H. J. Bollen, *An Algorithm for Determining the Direction to a Flicker Source*, IEEE Transactions on Power Delivery, Vol.21, NO. 2, April 2006.

[2] C. Benner and P. Carswell and B. Don Russell, *Improved algorithm for detecting arcing faults using random fault behavior*. Electric Power Systems Research **17**.1, 49 — 56, (1989)

[3] T. M. Lai and L. A. Snider and E. Lo *Wavelet transform based relay algorithm for the detection of stochastic high impedance faults*. Electric Power Systems Research, **76**.8, 626 — 633, (2006)

[4] T. Nowicki and G. Swirszcz, *A Method for Detecting High Impedance Faults on Distribution Systems using Method of Deviation from a Local Model*, preprint 2010.

[5] T. Nowicki and G. Swirszcz, *Detecting and Eliminating False Positives in High Impedance Fault Detection on Distribution Systems*, preprint 2010.

[6] R. Patterson, *Signatures and software find high impedance faults*. IEEE Computer Applications in Power, **8**.3, 12—15 (1995) (updated 2002)

[7] B. D. Russell, *Computer relaying and expert systems: New tools for detecting high impedance faults* Electric Power Systems Research, **20**.1, 31—37, (1990)

[8] A. R. Sedighi and M. R. Haghifam and O. P. Malik and M. H. Ghassemian *High impedance fault detection based on wavelet transform and statistical pattern recognition* IEEE Transactions on Power Delivery, **20**.4, 2414—2421, (2005)

[9] A. R. Sedighi and M. R. Haghifam and O. P. Malik, *Soft computing applications in high impedance fault detection in distribution systems*. Electric Power Systems Research **76**.1-3, 136—144, (2005),

[10] J. Tengdin et. al.,High Impedance Fault Detection Technology Report of PSRC Working Group D15, http://grouper.ieee.org/groups/td/dist/documents/highz.pdf

[11] http://en.wikipedia.org/wiki/Receiver_operating_characteristic