Incorporation of Data-Mining in Protection Technology for High Impedance Fault Detection

A. Valero Masa, *Student Member, IEEE*, S. Werben, *Siemens AG*, and J. C Maun, *Member, IEEE*.

***Abstract*-- Modernizing the power distribution system implies improving the reliability and performance of protection devices. By incorporating data-mining in the process of designing protection functions, the limits of performance are extended. We propose a method that uses data-mining, able to detect high impedance faults (HIFs) in multi-grounded distribution networks when conventional devices are insufficient. HIFs are produced when overhead lines contact a quasi-isolated surface, such as a tree or the ground. The fault current can be lower than the residual current under normal conditions; hence overcurrent devices do not detect this fault. We describe a set of indicators that characterize HIFs and that can be used in data-mining to distinguish fault situations from other situations. The result is a HIF detection function whose development is based on pattern recognition analysis. The presented methodology can be applied to other fault detection problems to achieve more reliable protection devices.**

***Index Terms*— Classification algorithms, data-mining, fault detection, grounding, high impedance faults, pattern recognition, power distribution lines.**

I. Introduction

Nowadays, the main goal of electrical engineers is to modernize the power system by making it more intelligent, reliable and sustainable. An important field of the power system engineering that has to be improved is fault detection. Some faults cannot be detected by the current protection technology, so new functions have to be implemented to enhance the performance of the protection devices. Progress in artificial intelligence and availability of digital information coming from the network make it possible to face fault detection in an innovative way. This new approach is likely to achieve results that improve the limitations of the present protection technology.

Some complex faults in multi-grounded distribution networks are the High Impedance Faults (HIFs) [8]. The difficulty in detecting these faults is due to the low fault current that they produce, which is usually lower than the residual current present in the network under normal conditions. Hence overcurrent protection is not capable to detect them. Certain relay manufacturers [2], [4], [5] commercialized protection devices with HIF detection functions, but utilities that have tested them agree that the performance is not satisfactory [6]. Therefore, other approaches different to the conventional one, have to be used for solving the detection problem of HIFs.

The objective of this work is to characterize HIFs, find indicators allowing differentiating them from other events, and applying data-mining for obtaining a classification method. This classification method will be part of the design of a HIF protection function, which will contribute to the improvement of the reliability and security of the electric grid.

The methodology presented in this paper for dealing with the HIF detection could be apply to other challenging faults that cannot be detected by conventional protection devices.

II. Addressing the Problem of the HIF Detection

High Impedance Faults (HIFs) are undetectable by overcurrent protection devices under certain conditions. The difficulty of the detection depends on the configuration of the network, being the worst scenario the multi-grounded distribution systems. We attempt to overcome the limitation of the present protection devices regarding HIFs by developing a HIF detection function for multi-grounded distribution systems using an advanced approach. This approach is based on the accurate characterization of the fault and the use of data-mining for recognizing HIF patterns.

A. Definition of HIF

A HIF occurs when an energized conductor makes undesired contact with a quasi-insulating object, such as a tree or the ground. This contact restricts the level of the current to a very low value, from a few mA up to 75A [6]. In networks where the value of the residual current under normal conditions is higher than that value, overcurrent devices do not protect against HIFs. However, this protection is essential for guaranteeing public security, because HIFs represent a public security hazard and risk of fire.

B. Difficulties in Detecting HIFs

The problem of the HIF detection is different depending on the configuration of the network. The difficulty in detecting the fault is directly related to the existence of residual current in the network under normal conditions. In multi-grounded networks the normal value of the residual

This work was supported by Siemens A.G under doctorate student scholarship.

A. Valero Masa is PhD student in the Department of BEAMS (Bio-, Electro- And Mechanical Systems) Energy, Université Libre de Bruxelles, Brussels, Belgium. (e-mail: avalerom@ulb.ac.be).

J. C. Maun is Professor in the Department of BEAMS Energy, Université Libre de Bruxelles, Brussels, Belgium. (e-mail: jcmaun@ulb.ac.be).

S. Werben is with Siemens AG, Energy Sector, Power Distribution Division, 59 Humboldtstr. 90459 Nurnberg, Germany. (e-mail: stefan.werben@siemens.com).

978-1-4673-2729-9/12/$31.00 ©2012 IEEE

current is higher than the typical value of HIF current, thus these faults are beyond the capability of overcurrent protection.

Solidly grounded distribution systems in Europe are grounded at a single point, at the substation. This practice together with the use of three-phase transformers in the MV/LV substations means that the neutral conductor under normal conditions carries only the current of the unbalance of the capacitive charging of the three phases, that is, barely a few amperes. In contrast, the typical configuration in American countries is multi-grounded system using single-phase distribution transformers for rural and residential areas. This practice means that the current unbalance due to the load switching activities is transferred to the primary distribution system, producing important neutral current.

The neutral current in wye-connected power systems is a consequence of single-phase loads connected line to neutral (Fig.1). If the connection of the primary of the distribution transformers is other than delta, the neutral conductor in the feeder can carry a surprising level of current. But in multi-grounded systems the neutral current is not the only component of the residual current, there is also the contribution of the stray current (Fig.1). The stray current is the continuous flow of current, excluding the momentary fault current, over the earth. In multi-grounded systems there are numerous neutral connections creating paths for stray currents, thus the current flows uncontrolled with an important magnitude. [1]

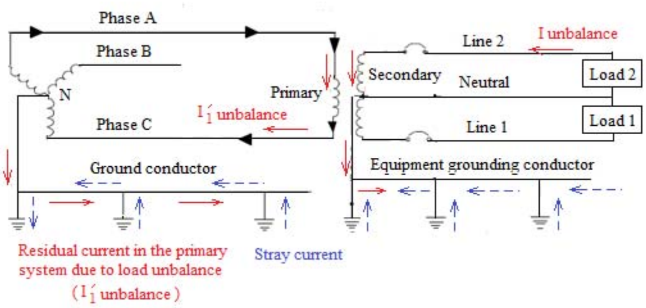


Fig. 1. Scheme of the flow of residual current under normal conditions in multi-grounded 4-wire wye electrical systems.

The residual current in multiple-grounded systems, which is composed of the neutral current and the stray current, is higher than in other configurations. Hence the settings of the overcurrent protections are 10 or 50 times less sensitive than in single-point-grounded systems, thus the HIF detection is more difficult. The demand for a detection method comes mainly from utilities operating multiple-grounded systems. This work is focused on responding to this demand.

C. *Proposed Method for HIF Detection*

Even though important improvements have been made in the last two decades [6] there is no satisfactory solution for HIF detection. The reasons are the intrinsic difficulties in detecting the fault and also the lack of knowledge of the fault, which is not documented.

The method we propose is based on the accurate characterization of the fault and on the application of modern computing techniques for identifying and distinguishing patterns. The procedure for developing the method can be explained in four general steps (represented in Fig.2):

- Collecting data. It is essential to have a proper formulation of the problem. In this step we study the state-of-the-art. We gain knowledge about HIFs and we obtain recordings of residual current ($3I_0$) under fault and normal conditions.

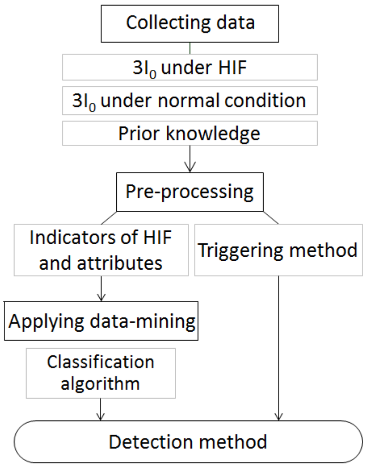


Fig.2. Proposed procedure for developing the HIF detection method based on an accurate description of HIFs and the use of data-mining techniques.

- Pre-processing. We have to extract information from the current recordings using the prior knowledge. This information is a list of magnitudes that represent relevant and distinctive characteristics of HIFs (indicators and attributes) and a method that indicates the beginning of the fault (triggering method).
- Applying data-mining. We use Weka (Waikato Environment for Knowledge Analysis) [11] for developing a classifier able to differentiate HIFs from other events. The events we consider are fluctuating loads, switching new loads, production of inrush current and any other event that produces a current similar to HIF currents. In our research we denominate "critical residual currents under normal conditions" to the residual currents produced by these events, because they are the most challenging currents for differentiating them from HIF currents.

The classifier is learnt and tested using the database object of the study. The database is previously created and it contains the attributes of each residual current recording. The technique used for developing the classifier is the one that best distinguishes the instances labelled as *HIF* from the ones labelled as *normal*.

- Describing the detection method. Once the trigger method and the classification algorithm have been obtained, the last step is to assemble them and to test the complete detection method.

III. Knowledge Discovery in Databases Applied to HIF Detection

Knowledge Discovery in Databases (KDD) is, according to the American Association of Artificial Intelligence, the nontrivial extraction of implicit, previously unknown, and potential information from data [10]. In other words, it is the process of turning data into knowledge, discovering meaningful patterns by the application of data-mining techniques. Nowadays, with recordings from the network at our disposal, we should take advantage of these new techniques by incorporating them in the process of designing protection functions. As a result, the performance of the detection functions would be optimized and their limitations reduced.

The process of KDD is interactive and iterative. It is commonly defined in five steps [10]: Data selection, pre-processing, incorporation of appropriate prior knowledge, data-mining and interpretation and evaluation of the results.

This section of the paper is aimed to explain how we applied the KDD for developing a solution to the problem of HIF detection.

A. Data Selection for HIF Detection

Before anything else, we had to understand the domain of application and the aim of the KKD process. The application is fault detection and the aim is to classify residual currents in *HIF situation* or *normal situation*. Therefore, the two first complementary tasks are getting knowledge of HIFs and obtaining recordings.

Once we had knowledge and recordings we considered the other necessary inputs for the data-mining process. These inputs take the form of concepts, instances, attributes and knowledge. The concept is the residual current; we want to find a criterion for considering residual currents and relating them to *HIF situation* or to *normal situation*. The instances are individual and independent examples of the concept, in this case, recordings of HIF currents and residual currents. And the attributes are the magnitudes that measure different aspects of the instance. The third task of the data selection step is searching for and determining the attributes.

1) Getting Knowledge of HIFs: We got prior knowledge of HIFs and the difficulty of its detection by literature study and by direct contact with distribution system operators. From the literature study [6], [8] we learnt about the generally accepted characteristics of HIFs and about the methods that have been developed for detecting HIFs. The distribution system operators that we contacted (Iberdrola, Electricaribe, Emcali, CEEE-d, Edenor and HydroQuebec) provided us with their experience with HIFs, information about how they deal with them, and prospects for the future regarding HIF detection. In general, utilities in America are aware of the inexistence of methods that detect HIFs in a practical and secure way, and they are interested and willing to collaborate for making improvement in this area.

2) Obtaining Recordings: The wanted recordings are residual current under HIF conditions and under normal conditions. The main problem is the unavailability of these recordings.

HIFs in multi-grounded networks, where the residual current is higher than the HIF current, are not detected; hence there are no recordings of residual current under HIF situation coming from multi-grounded networks. We overcame this limitation by making the assumption that the signal can be obtained by superposition of a HIF current and a residual current under normal conditions. Therefore, we designed and performed HIF laboratory tests for obtaining HIF current recordings [9].

The other difficulty is to get recordings of residual current under normal conditions, given that it is unusual to record the current in distribution networks if there is no fault. Nevertheless, some of the distribution system operators contacted, Emcali and Electrocaribe, provided us with some examples, as the one shown in Fig 5.

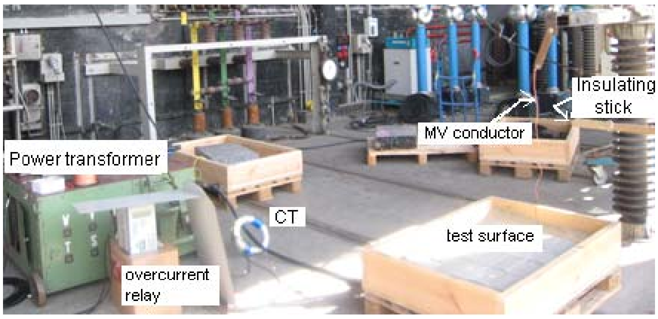


Fig. 3: Set-up for the HIF tests in the MV testing Laboratory of Siemens.

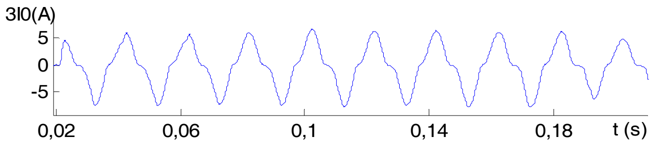


Fig. 4. Current of a HIF test on sidewalk surface at 12kV.

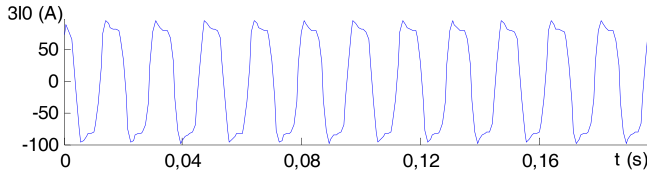


Fig. 5. Recording of residual current under normal conditions provided by Electricaribe, presenting some features typical of HIF currents, at 11 kV.

The laboratory tests were performed in the MV Testing Laboratory of Siemens in Berlin. The laboratory set-up and an example of the resulting current are shown in Fig.3 and Fig.4. As a result we created a database of HIF currents under different conditions of contacting surface (sand, paving-stone, tree, soil, asphalt, concrete and sidewalk surface), damp level and voltage level (12kV, 24kV, 36kV).

3) Searching for and Determining the Attributes: Characterizing the fault is determining the relevant features of HIFs and finding indicators capable to measure these features. Distinguishing HIFs and other similar events should be done in parallel with the characterization of HIFs, since the attributes useful for developing the classifier algorithm are the indicators able to reveal the difference between HIFs and other situations.

Indicators of HIFs are quantitative or qualitative variables that provide information about distinguishing features of these faults. These distinguishing features of the current have been presented and explained in previous work [9], they are:

- Low current, between 0.01A and 100A.
- Intermittent arc, due to the imperfect contact between the conductor and the surface.
- Randomness, arbitrary changes in the amplitude of the fundamental and the harmonic currents component.
- Waveform asymmetry.
- Presence of low harmonics, mainly the 3rd harmonic.
- Non-harmonic components.
- Single-phase fault.

TABLE I
CHARACTERISTICS OF HIFS, INDICATORS OF HIFS AND ANALYSIS TECHNIQUES USED FOR THIS CALCULATION

Characteristics	Analysis Techniques	Indicators
Very low current		Amplitude of the current
Low harmonics	FFT	Correlation 3rd harmonic and fundamental amplitude
		Phase of the 3rd harmonic
Intermittent arc		Main harmonic
	Waveform analysis	Constant current at zero crossing
Randomness	Difference cycle per cycle	Difference cycle per cycle
Non-harmonics		Changes of the accumulated difference cycle per cycle
Waveform asymmetry	Decomposition of cycle	Asymmetry ratios
Single phase fault	Comparison	One phase fault indicator

Table I lists the characteristics that we need to measure, the proposed indicators for measuring them, and the analysis techniques used for calculating the indicators. These techniques are: frequency analysis, waveform analysis focused on detecting the effect of the arc at the current-zero-crossing, calculation of the difference cycle per cycle, and specific functions for studying the asymmetry of the waveform. We settled the indicators after a recursive process of calculating magnitudes, applying them to the recordings, and optimizing the extraction of information.

The indicators calculated by frequency analysis are the identification of the main harmonic, the phase of the 3rd harmonic, and the correlation between the 3rd harmonic and the fundamental component amplitude. The frequency analysis is performed using the Fast Fourier Transform (FFT) applied cycle per cycle to the current recordings.

Our first impression was that the 3rd harmonic was useful for the HIF characterization, so we identified the main harmonic of the current and we searched for the amplitude and the phase of the 3rd harmonic. Based on the observation of the current waveform it was expected that the 3rd harmonic was the main component in most of the recordings and its phase was constant and close to 180°. These expectations were validated with some examples (see Fig. 6), so these magnitudes were chosen as indicators.

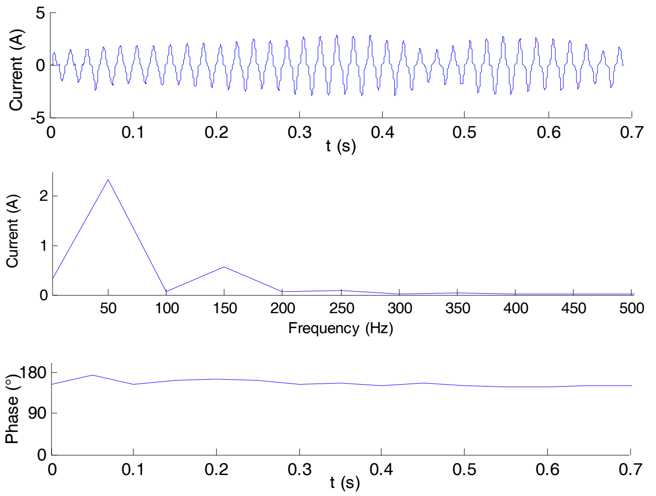


Fig. 6. Current from a HIF test using a piece of sidewalk as test surface, frequency analysis of the current showing the 3rd harmonic dominance and phase of the 3rd harmonic, which is constant and close to 180°.

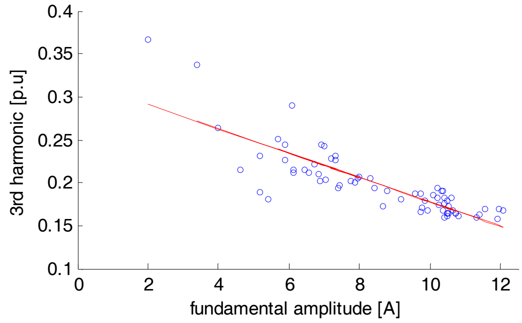


Fig. 7. Correlation between the 3rd harmonic amplitude in values per unit of fundamental current (p.u) and the fundamental current amplitude of the current of a HIF test on sand. The correlation is measured by the slope of the linear-curve-fitting using least-squares method.

The best way we found for using the amplitude of the 3rd harmonic is studying the correlation between this amplitude and the fundamental amplitude. It was noticed that in low currents the effect of the arc was more important and so, the relative amplitude of the 3rd harmonic. Therefore, if the amplitude of the current changes with time, it is typical of HIFs that the relative amplitude of the 3rd harmonic increases when the fundamental amplitude decreases. Fig.6 illustrates this correlation in the recording of a HIF test on earth surface. The last indicator of this group is the slope of the linear-curve-fitting using the least-squares method.

The indicator calculated by studying the waveform of the current is the ratio of the half-cycles presenting the arc effect at the current-zero-crossing. This effect, which is visible in the upper plot in Fig.6, is distinctive of HIF currents. The function developed for identifying this effect stands on the derivative function of the current. As shown in Fig. 8 the derivative function of the fundamental component of the current (dI_{50Hz}/dt) is, obviously, sinusoidal. However, the derivative function of the HIF current (dI/dt) is not sinusoidal; it has a minimum close to the maximum of the dI_{50Hz}/dt. The use of derivative functions amplifies the effect of the arc at the current-zero-crossing so it is simpler to detect it. The function counts the number of half-cycles presenting this effect over a fixed window, and calculates the ratio that is an indicator of HIF.

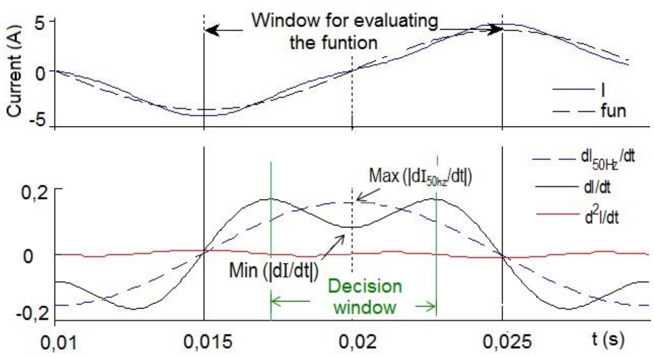


Fig. 8. The zero-crossing-effect detection function is applied in a window centered in the current-zero-crossing. If the function dI/dt has a minimum in the window of length quarter-cycle that is centered in the maximum of dI_{50Hz}/dt, there is presence of arc effect in this half-cycle.

The technique of the differences cycle per cycle is known as indicator of non-periodicity, randomness and presence of non-harmonic components. Hence it is a suitable and necessary technique in the characterization of HIFs. The function we built computes the difference of the current from cycle to cycle and accumulates the absolute values of these differences during n_{acc} samples. The resulting magnitude is the Accumulated Absolute Difference (AAD). We decided to consider two indicators regarding the HIF randomness: the value of AAD and the detection of important irregularities in the dynamic of the AAD.

We concluded that a current is random if the value of AAD is higher than the *normal_AAD*, which is fixed during the normal operation of the network and is an adaptive value. Besides that, the randomness is recognized as the one typical of HIFs if the values of AAD present frequent and sudden changes. AAD under HIF conditions is highly irregular and shows recurrently sudden peaks. We count a peak when AAD is higher than n times the average of AAD during a fixed window of 5 cycles (see Fig.9). The indicator of dynamic of AAD turns positive when peaks repeatedly occur.

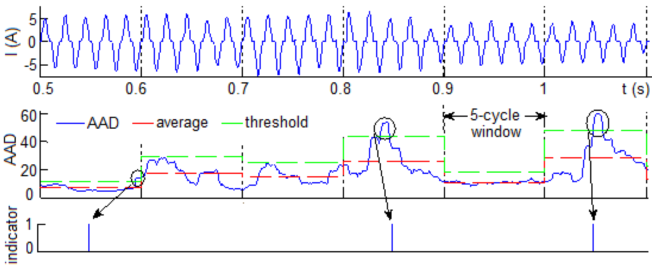


Fig. 9. Current and AAD of a HIF recording, illustrating the repetition of peaks that represent the typical random of HIFs.

We measure the asymmetry of the current waveform by two indicators, each one related to the two different types of asymmetry that we distinguish.

The first type is the asymmetry between the positive and negative half-cycle (*P/N_Asym*). It was remarked by observing that several HIF currents had positive peak values different to the negative one. The HIF test current shown in Fig.10 illustrates this characteristic, which is evident when comparing the current with the fundamental component. We developed the positive/ negative asymmetry function for detecting this behavior.

The other type of asymmetry was called even and odd quarter-cycle asymmetry (*E/O_Asym*). We decided to have an indicator for this asymmetry when we noted that the arc effect at the zero-crossing is, most of the times, not exactly at the current zero, so the waveform becomes asymmetric as presented in Fig.11. Based on this fact, we say that the first and third quarter-cycles are different from the second and fourth quarter-cycles. Hence the developed function builds one signal using the information of the first and third quarter- cycles, and other signal using the information of the second and fourth ones. After comparing them, it determines if the indicator of the even/odd asymmetry must be positive or not.

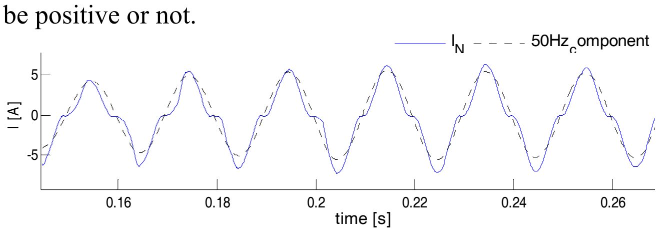


Fig. 10. Positive and negative half-cycle asymmetry, visible by comparing the HIF current with the fundamental component.

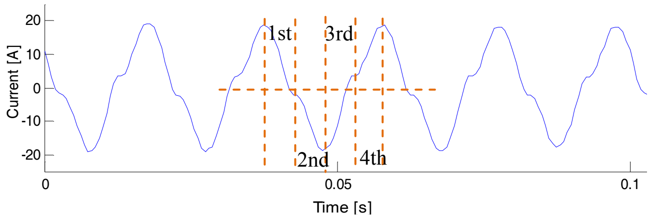


Fig. 11. Even and odd quarter-cycle asymmetry of HIF current, due to the arc effect at a moment different to the current-zero-crossing.

Two last indicators related to the definition of HIF are aimed to verify the low amplitude of the current and the condition of single-phase fault, respectively.

We consider that the search for indicators is the central step of the procedure. The mentioned indicators are capable of describing HIFs as accurate as possible, presenting an advantage compared to previous research on HIFs. Moreover, the subsequent data-mining process uses the indicators as attributes for performing the classification. For achieving a good classification it is essential to have a correct formulation of the problem, therefore, a correct list of attributes.

B. *Pre-processing: Creating and Cleaning the Relational Database*

Data for KDD is organized in a relational database, which are tables depicting the values of some attributes for each instance. We found it convenient to use the HIF indicators as attributes for the data-mining process. Therefore, the values of the magnitudes chosen as indicators were calculated for each instance of the residual current database, both HIF currents and residual currents under normal conditions. After assembling the data coming from different sources (laboratory tests and distribution recordings), we obtained the relational database, and we structured it as Attribute-Relation File Format (ARFF), as shown in Fig.12.

```
@RELATION Iresidual

@ATTRIBUTE FaultorNot  {HIF,normalcond}
@ATTRIBUTE AverageAAD  NUMERIC
@ATTRIBUTE Fch_indicator  {yes,no}
@ATTRIBUTE Main_H   NUMERIC
@ATTRIBUTE Mean_AH3   NUMERIC
@ATTRIBUTE Slope_3Hpu_vs_fundA  NUMERIC
@ATTRIBUTE Mean_Ph3H  NUMERIC
@ATTRIBUTE sigma_Ph3H  NUMERIC
@ATTRIBUTE Ratio_0Cross  NUMERIC
@ATTRIBUTE Ratio_P/N_Asym  NUMERIC
@ATTRIBUTE Ratio_E/O_Asym  NUMERIC

@DATA
HIF,  22.48524968,     no,  no,  3,    0.318625163,
      0.067647732,     178.8531549,    165.9168325,
      194.6660645,     100, 2.3334,    4.8498,    85.294,
      0.983626574
HIF,  11.06583701,     no,  no,  3,    0.318625163,    -
0.098653746,     174.7149821,    157.0468851,     197.6864988
      97.101,    0.5577,    0.8600,    70.588,    0.294063211
normalcond,      13.523,     no,  no,   3,    0.208,
      0.023,           22.456,          -16.237,   59.720,
      0.000,     0.203,     0.286,      68.919,    1.187
```

Fig. 12. Extract of the ARFF file containing the relational database of residual currents used for the HIF detection.

C. *Incorporation of Prior Knowledge*

This step consisted in finding a data-mining method and useful features to represent the data in order to achieve the goal of the KDD.

Each way of representing the result dictates a kind of data-mining technique that can be used for inferring the output structure. Our objective is to get a criterion for differentiating data in two pre-defined classes: fault and normal condition. Consequently, the representation of the result must be classification rules or decision trees, so these will be the data-mining techniques.

Through examining the database seeking for a reduced number of variables and invariant representations for the data, the data-mining process become more robust and simpler.

D. *Data-mining: Classification Techniques*

In the data-mining step we searched for patterns of HIF for predicting the class of future inputs of residual current and for describing HIFs in a precise form.

The data-mining methods are based on machine learning, using also techniques of pattern recognition, expert systems and statistics. The most common methods are: classification, clustering, regression, summarization, dependency modelling, and change and deviation detection.

Our problem of HIF detection is formulated as a classification problem, and the data-mining techniques we used are decision trees and rules. These techniques express the classification decision based on conditions of some selected attributes. Thus, the use of the output in the complete HIF detection algorithm is almost direct.

We used the software WEKA (Waikato Environment for Knowledge Analysis) for performing the data-mining in our database [11]. It supports standard data-mining tasks, and among them, the ones that we required: pre-processing, attribute selection, classification and visualization.

E. *Interpretation and Evaluation of the Results*

The data-mining process is interactive and iterative, so the results were interpreted applying the prior knowledge and repeated when finding some convenient modification for improving the results. Several different models of trees and rules were applied, so we could evaluate the results and chose the more appropriate ones.

IV. Results

In the first step we obtained criteria for describing HIFs through the calculation of the value of the indicators of the HIF recordings. Once those values were organized in a relation database, we performed data-mining and got different appropriate classifiers.

A. *Description of HIFs*

The first result is the description of HIFs built on the application of the proposed indicators to our HIF database composed by 73 current recordings. The main harmonic of HIF currents is the 3rd harmonic, and the phase of this component is 180° ± 30°. If the amplitude of the current varies, the correlation between the amplitude of the 3rd harmonic and the amplitude of the fundamental component is negative in most cases, the 82.5% of the HIF currents of our database. The effect of the arc at the zero-crossing is identified in the current waveform. It was detected in all the recordings, being the ratio of zero-crossing-effect higher than 0.9 in 92% of the cases, and higher than 0.8 in the 96%. Therefore, accepting the inaccuracy of the 4%, we say that HIF currents present a zero-crossing-effect ratio higher than 80%. The 87.5% of the HIF currents are recognized as random currents, and the 78.5% of them present changes in the AAD magnitude. HIF current may have a certain asymmetry, 30.5% of the cases presented positive and negative half-cycle asymmetry, and 45% of the cases presented even and odd quarter-cycle asymmetry.

TABLE II
HIF Description based on the Characterization of our Database

Characteristics	Indicators	HIF Description	
Very low current	Amplitude of the current	I_{HIF} = 0.01A - 100A	100 %
Low harmonics	Slope of the fitting curve 3rd harmonic vs fundamental	slope < 0	82.5 %
	Phase of the 3rd harmonic	Ph_{3rdH} = 180°± 30°	96 %
Intermittent arc	Main harmonic	Main harmonic = 3rd	100 %
	Ratio of zero-crossing-effect ($R_{0\text{-cross}}$)	$R_{0\text{-cross}} > 0.8$	96 %
Randomness	Accumulated Absolute Difference cycle per cycle (AAD)	AAD>normal_AAD+5(nacc/200)	87.5 %
Non-harmonics	Dynamic of AAD	Dynamic AAD (0/1) = 1	78.5 %
Asymmetry	P/N_Asym E/O_Asym	P/N_Asym (0/1) = 1 E/O_Asym (0/1) = 1	30.5 % 45 %

The proposed set of indicators precisely presents the characteristics of HIFs. This is an improvement in HIF detection since the essential problem for HIF detection is the lack of documentation and knowledge of the fault.

B. *Differentiating between HIF and loads*

The second result is the classifier for distinguishing HIF currents and currents under normal conditions. The last objective of the work is to achieve a method for detecting HIFs, namely, a method able to recognize HIFs and to differentiate them from other events, even from critical loads that create a residual current similar to the one due to HIFs.

Considering the problem formulated as "differentiating between HIF and loads", the solution is highly satisfactory. Most decision tree models and classification rule models applied to the database create classifiers with very good performance. One example is the tree model *id3* [7], which generates the tree based on the selection of attributes in order to minimize the entropy of this selection. Using cross-validation for training the algorithm, the resulting tree is the one shown in Fig. 13. The detection rate calculated for our database is 98.3% and the false alarm rate is 10%, although this last rate has been calculated using loads that produce a current similar to HIF current; the more difficult cases for classifying.

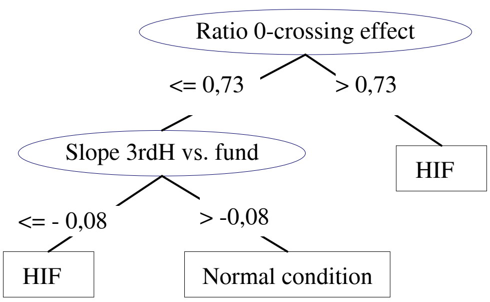


Fig. 13. Decision tree generated by the model *id3* for classifying HIF currents and residual current under normal conditions.

Using different models of decision trees or classification rules, the attributes chosen for generating the classifier can change. Fig. 14 is a tree representing the rule produced by the *One Rule* method [3]. This method generates rules that test a single attribute and every attribute, producing branches that correspond to each value of the attributes. The best classification is the one presenting the lower error. This classifier showed a 100% of detection rate and 8% of false alarm.

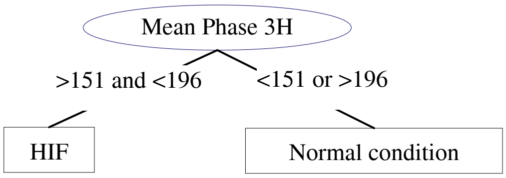


Fig. 14. Tree representing the classification rule generated by the *One Rule* model for classifying HIF currents and residual current under normal conditions.

The attribute chosen by this method was the phase of the 3rd harmonic, while other methods chose the ratio of the zero-crossing-effect, alone or complemented by other attributes. The fact is that some of the selected attributes are very powerful in distinguishing between HIF current and residual current under normal conditions. These attributes are those related to the waveform of the current due to the effect of the arc. The problem of classifying HIFs and loads in normal conditions presents no special difficulties once the attributes have been selected.

V. Future Work

The difficulty that we are currently facing is that the input current we have to consider is not HIF current and residual current under normal conditions, but residual current under HIF conditions and residual current under normal conditions. The classification problem formulated in this paper is the simplified one, and the only one that has a satisfactory solution. When a HIF occurs in multi-grounded networks, its current will be superposed to the residual current present in the system. The critical situation for the HIF detection, the one that may cause false alarms, occurs when, apart from the no-random normal load residual current, a component of the residual current similar to HIF current but due to load is present in the system.

There is no method that can differentiate these two types of currents (residual current under HIF conditions and residual current under load conditions similar to HIF current) in a practical and effective way. Fig. 15 and Fig. 17 illustrate that the distinctive characteristics of HIF currents (seen in Fig. 16) are hidden when HIF current and residual current under normal situations are superposed. Other researchers working in HIF detection have tried to differentiate them, resulting in an inherent limitation on the performance of their methods. However, we will overcome that limitation by removing the no-random normal load residual current component from the residual current, so the remaining classification problem is the one presented in this paper. We are working at the moment on developing and optimizing the method for removing the no-random normal load residual current component and so, after this last task, the detection algorithm will be complete.

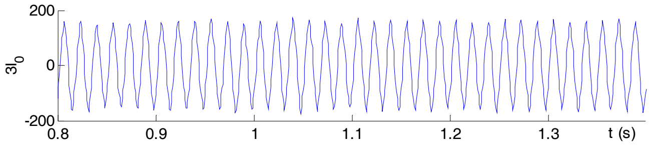


Fig. 15. Recording of residual current under normal conditions coming from a distribution network of Emacali.

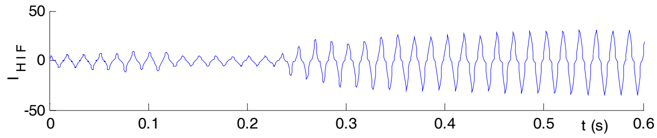


Fig. 16. Recording of HIF current coming from the laboratory tests.

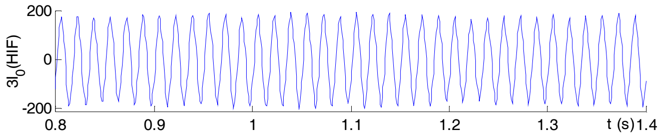


Fig. 17. Superposition of the residual current under normal conditions and the HIF current, resulting in a current that is expected to be a residual current under HIF conditions.

VI. Conclusions

This paper presents a new approach for developing advance detection algorithms for protection devices, consisting in incorporating data-mining techniques in the designing process. The application of data-mining techniques makes it possible to overcome the present limitations inherent to the conventional protection algorithms. One of these limitations, and the problem faced in this research, is the high impedance fault (HIF) detection.

The proposed HIF detection method involves two main tasks: accurate characterization of the fault and distinction between HIF currents and critical residual current under normal conditions. The HIF characterization, which is based on the list of HIF indicators, is the first important result. The lack of knowledge of this fault has been a difficulty for the HIF detection, so such a precise description of HIF was required. Besides that, the characterization is the fundament for performing the distinction between HIFs and critical residual current under normal conditions. The indicators used in the characterization are the attributes input to the data-mining techniques. Given that the attributes have been selected for indicating the relevant and distinguishing characteristics of HIFs, the output of the data-mining is highly satisfactory. Several classifiers, resulting through applying different learning methods, recognize HIF patterns and differentiate them from other events.

The explained HIF detection method has advantages compared to others because through the use of data-mining we are capable of distinguishing HIF patterns. This capability is the requirement that the protection devices need for improving its performance.

VII. Acknowledgment

The authors gratefully acknowledge the support and assistance of Siemens A.G, and particularly to M. Kereit, and to J. Oemisch and the team of the Medium Voltage Testing Laboratory. They also express their gratitude to I. Ojanguren from Iberdrola Distribución Eléctrica S.A.U., to M. A. Pulice and O. N. Rivera from Edenor, to J.F. Aguirre from Emcali and to R. T. Amador and A. Arregoces from Electricaribe.

VIII. References

[1] S. Benchluch, J. Burke, "European vs. U.S. Distribution System Design" . N.Y.C. : WPM, 1999.

[2] C. L. Benner, B. D. Russell, "Practical high impedance fault detection on distribution feeders", in *Proc. 1997 IEEE Transactions on Industry Applications Conf.*

[3] G. Buddhinath, D. Derry, "A Simple Enhancement to One Rule Classification". University of Melbourne, Tech. Rep.

[4] R. Das, B.ayoumin, " System for detection of high impedance faults", in Proc. 2007 *19th International Conference on Electricity Distribution.*

[5] D. Hou, *"Detection of High-Impedance Faults in Power",*Schweitzer Engineering Laboratories*,* Tech. Rep., 2006

[6] I. Ojanguren, G. Steynberg, J-L Channelière, L. Karsenti, R. Das, "High Impedance Faults". WG B5.94 Cigre, Tech. Rep., 2009

[7] J.R. Quinlan, "Induction of decision trees", in *Machine Learning,* Kluwer Academic Publishers 1986, pp. 81-106.

[8] J. Tengdin, R. Westfall, K. Stephan, "High Impedance Fault detection Technology", PSRC Working Group D15, Tech. Rep., 1996.

[9] A. Valero Masa, J.-C. Maun and S. Werben," Methodology to Describe High Impedance Faults in Solidly Grounded MV Networks", in *Proc. 2011 Cired Conf.*

[10] J. William, Frawley, G. Piatetsky-Shapiro, J. Matheus, "Knowledge Discovery in Databases: An Overview", American Association for Artificial Intelligence, 1992.

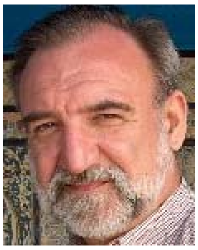
[11] I. H. Witten, E. Frank, *Data Mining: Practical Learning Machine Tools and Techniques,*.2nd ed., Elsevier, 2005.

IX. Biographies

Ing. Alicia Valero Masa was born in Valladolid, Spain, on July 8, 1984. She graduated from the High Polytechnic School of Engineering of Valladolid in 2008.

She staged an internship in the researching center of Arcelor Mittal in Gent (Belgium) and one year later she become PhD student in the Electrical Engineering Department of the Université Libre de Bruxelles of Brussels (ULB), in Brussels, Belgium. Her special fields of interest include control and protection of electrical networks and industrial applications.

Dipl.-Ing. Stefan Werben was born in Nijmegen, Netherlands on June 1, 1964. He graduated from high school Goethe Gymnasium, Einbeck, Germany in 1983. He studied Electrical Engineering at the Technical University Braunschweig, Germany and graduated with his Diploma in 1990. Post-Graduate studies in Business Management at the Southern Illinois University, Carbondale, USA tool place in year 1991.
1992 he joined Siemens AG as a software developer for digital protection devices. After R&D project management he changed into the Product Management where he worked as technical specialist for differential protection. Since 2001 his job is the Product Lifecycle Management for Overcurrent protection devices.



Prof. Jean-Claude Maun Jean-Claude Maun received the M.Sc degree in mechanical and electrical engineering in 1976 and the Ph.D. degree in Applied Sciences in 1981, both from the Université Libre de Bruxelles (ULB), Brussels, Belgium. He joined the Electrical Engineering Department of this university in 1976 and he is now professor and Dean of the ULB Engineering School.

He has been leading research projects in the field of the design of digital protections for Siemens as a consultant. An emeritus member of the Société des Electriciens et Electroniciens français and a recipient of many academic awards, Prof. Maun is an recognised expert in electrical networks' safety and protection systems as well as, more generally, in power transport and distribution networks and in decentralized electricity generation.