

Alarm tree analysis using statistical techniques

Omar Aizpurúa, Humberto Álvarez

*Department of Electrical Engineering, Technological University of Panamá
omar.aizpurua@utp.ac.pa, humberto.alvarez@utp.ac.pa*

Ramón Galán, Agustín Jiménez, Alfonso Romero

*School of Industrial Engineers Polytechnic University of Madrid
ramon.galan@upm.es, agustin.jimenez@upm.es
a.romero@alumnos.upm.es*

Abstract- *This work presents a systematic method for the generation and treatment of the alarms' graphs, being its final object to find the Alarm Root Cause of the Massive Alarms that are produced in the dispatching centers. Although many works about this matter have been already developed, the problem about the alarm management in the industry is still completely unsolved. In this paper, a simple statistic analysis of the historical data base is conducted. The results obtained by the acquisition alarm systems, are used to generate a directed graph from which the more significant alarms are extracted, previously analyzing any possible case in which a great quantity of alarms are produced.*

Keywords- *Avalanche of alarms, Root Cause Analysis, Alarm Tree Analysis, Electrical Power Administration*

Resumen- *Este trabajo presenta un método sistemático para la generación y tratamiento de los gráficos de las alarmas, siendo su objetivo final el de encontrar la causa principal de las alarmas masivas que se producen en los centros de distribución. Aunque se han desarrollado muchos trabajos sobre este tema, el problema de la gestión de alarmas en la industria no está todavía completamente resuelto. En este trabajo, se lleva a cabo un análisis estadístico simple de la base de datos histórica. Los resultados obtenidos por el sistema de adquisición de las alarmas, se utilizan para generar un gráfico dirigido en el que se extraen las alarmas más importantes, analizando previamente todos los posibles casos en los que se produjeron una gran cantidad de alarmas.*

Palabras Claves- *Avalancha de alarmas, Análisis de Causa Raíz, Análisis del árbol de alarmas, Administración de Alarmas.*

Paper type: Original

Received: January 24, 2011

Accepted: July 13, 2011

1. INTRODUCTION

With today's digital control systems it is possible to configure a large number of alarms in industrial processes. This leads to a large number of alarms generated under normal conditions in many centers (2,000 alarms per day and operator), multiplying these when there is an abnormality in the process (coming to overcome 40,000 daily alarms) [1]. This large number of alarms in times of failure are called "avalanche of alarms". In addition, few alarms are the "Root Cause" of more than 90% of all alarms [2].

A group of researchers in intelligent control at the Polytechnic University of Madrid and the Technological University of Panama, has been working on the analysis of avalanches of alarms occurring in electrical power transmission and distribution centers. This issue can be addressed with different methods such as Bayesian Networks [3], neural networks [4], the expert systems [5], decision theory [6], etc., or even by a mixture of several of these methods, such as Neuro-fuzzy models [7] and Integrated Methodologies [1].

In this study, using statistical concepts, a directed graph of the alarms of an industrial process is to be developed using information from the database provided by the data acquisition system and identifying the most significant alarms from those that occur in a given time. That is, in an avalanche of alarms produced at any time, the system gets roots alarms that generate the rest of the alarms, reducing alarms received by the operator responsible for the supervision.

A file of alarms is generated, allowing to carry out certain operations, from the basis of available data of an electric power distribution center. These operations allow the development of a directed graph adjacency matrix, i.e. get to know how the network alarms that can occur are structured. Once matrix this is obtained the most significant alarms that occurred at an interval of time (temporal window) can be deduced by Boolean operations that filter the alarms objective of the problem.

The second section of this document develops the processing of data, presenting calculated correlations, partial correlations and conditional probabilities among all alarms resulting in the adjacency matrix of the directed graph of alarms. The third section explains the adjacency matrix and some of its characteristics and properties. The fourth section performs analysis of significant alarms, i.e. removal of alarms that are root cause in a case of avalanche. The fifth paragraph is a brief description of the program software created for the generation and analysis of alarms. Finally, outlining conclusions and possible future work used references. To illustrate the application, there is an example with its arrays and results (see Fig. 1)

2. DATA PROCESSING

From an electrical energy distribution center data acquisition system, a set of encoded data with a format of 0's and 1's is obtained. In this matrix, columns correspond to the database acquisition system alarms and the rows are the time intervals considered when creating the encrypted file. When a 1 appears in a row or temporary window, it indicates that the alarm corresponding to the column in the matrix is active at that time. This last file will provide the necessary relationships between the alarms needed to obtain the adjacency matrix of a directed graph.

The example presented (Fig. 1) is part of a given directed graph for which that alarm file is built. The file, in this case, consists of 10

columns that correspond to the 10 alarms to be studied and by a large number of lines that correspond to the same amount of temporary windows obtained from the database of events in a more general case.

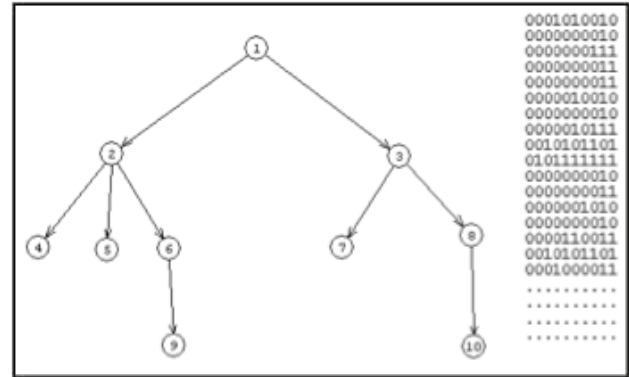


Figure 1. Example

A. Correlation Coefficients

Correlation is a dimensionless measure of the linear relationship between two variables. Correlation between two alarms x and y is given by the ratio between the covariance and product or the standard deviations of the two variables, i.e.:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad (1)$$

Thus the correlation is a measure on the degree of relationship between two variables, no matter what the cause or the effects are.

For this analysis, only the positive correlation greater than certain value (threshold), are needed to draw a graph of relations between alarms.

For the example, the correlations between alarms i, j, are presented in the following table, considering a convenient threshold value of 0.2. In this case, the correlation between alarm i and j, is given by the corresponding i, j position of the matrix.

Table 1. Correlation coefficients

Matriz de relaciones: (solamente aparecen las correlaciones mayores que el Umbral)

1	0.56078	0.33765	0.37982	0.27243	0.28214	0.22861	0.20677	*	*
0.56078	1	*	0.67730	0.48580	0.50312	*	*	0.27469	*
0.33765	*	1	*	*	*	0.67707	0.61238	*	0.42169
0.37982	0.67730	*	1	0.24861	0.36247	*	*	0.20092	*
0.27243	0.48580	*	0.24861	1	0.28794	*	*	*	*
0.28214	0.50312	*	0.36247	0.28794	1	*	*	0.54597	*
0.22861	*	0.67707	*	*	*	1	0.39030	*	0.26654
0.20677	*	0.61238	*	*	*	0.39030	1	*	0.68861
*	0.27469	*	0.20092	*	0.54597	*	*	1	*
*	*	0.42169	*	*	*	0.26654	0.68861	*	1

As seen in (Fig 2), it is possible to build a correlation graph from the resulting matrix, with the graph consisting of several redundant connections.

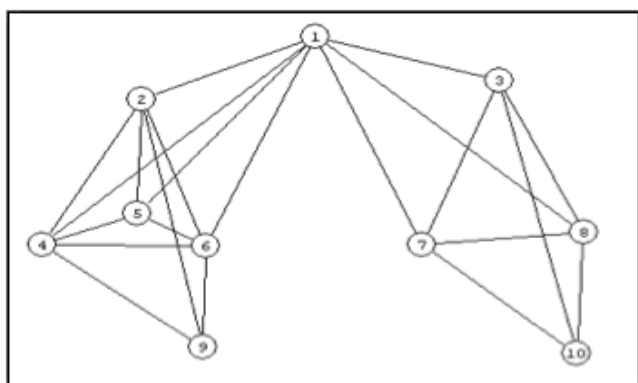


Figure 2. Alarm tree correlations

B. Partial Correlations

Excessive connections found with correlations in the previous paragraph can be simplified using partial correlations between two alarms and holding a third alarm constant. Namely, an existing correlation between two alarms (x and y) can be positive and greater than the threshold, but its partial correlation, maintaining alarm constant z, can be less than the threshold; Therefore, the relationship between the two alarms (x and y) would be explained by the relationship between the two alarms with the third (z). Partial correlations will help deleting redundant connections between nodes separated by more than one edge since they explain their common relationship with nodes between them. This process will delete connections between grandchildren and grandparents, because they are explained by the common parent-child relationship of the two alarms.

Partial correlations between all alarms are calculated to eliminate redundant or excessive connections of the graph of the figure above. The partial correlation between two alarms (x and y) keeping constant a third (z) are calculated based on correlations between three alarms, pairwise, follows:

$$r_{xy.z} = \frac{(r_{xy} - r_{xz}r_{yz})}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}} \quad (2)$$

As with previous correlations, only values greater than the threshold are considered for the analysis.

For the example discussed, a portion of their calculated partial correlations is shown below.

Table 2. Partial Correlations

Parcial [1][2].[3] = 0.53579	Parcial [2][4].[3] = -0.00000
Parcial [1][2].[4] = 0.44600	Parcial [2][4].[5] = 0.36566
Parcial [1][2].[5] = 0.50943	Parcial [2][4].[6] = 0.33487
Parcial [1][2].[6] = 0.50516	Parcial [2][4].[7] = 0.31041
Parcial [1][2].[7] = 0.54257	Parcial [2][4].[8] = 0.35805
Parcial [1][2].[8] = 0.55113	Parcial [2][4].[9] = 0.37656
Parcial [1][2].[9] = 0.54572	Parcial [2][4].[10] = 0.36043
Parcial [1][2].[10] = 0.55264	Parcial [2][4].[10] = 0.37347
Parcial [1][3].[2] = 0.28033	Parcial [1][5].[2] = -0.00000
Parcial [1][3].[4] = 0.32106	Parcial [1][5].[3] = 0.26238
Parcial [1][3].[5] = 0.32995	Parcial [1][5].[4] = 0.19866
Parcial [1][3].[6] = 0.31119	Parcial [1][5].[6] = 0.20809
Parcial [1][3].[7] = 0.29524	Parcial [1][5].[7] = 0.23525
Parcial [1][3].[8] = 0.27283	Parcial [1][5].[8] = 0.26553
Parcial [1][3].[9] = 0.33024	Parcial [1][5].[9] = 0.25495
Parcial [1][3].[10] = 0.30931	Parcial [1][5].[10] = 0.26561

For example, correlations between alarms 1 and 2 for a given z, are all greater than the threshold of 0.2. It has been shown experimentally that a threshold equal to 0.2 is a good number to discard whether or not correlated variables.

The same is shown for correlations between 1 and 3. On the other hand, between alarms 1 and 4 there are values that are less than the threshold. That is, the partial correlation for 1 and 4, given as constant alarm 2, has a value of 0.000. Alarm 2 is a son of alarm 1 and father of alarm 3; thus, the relationship between grandfather (1) and grandson (3) is explained by the common father-son (2). The same is seen in the relationship between alarms 1 and 5. The resulting graph is shown in figure 3, after eliminating the redundant relationships.

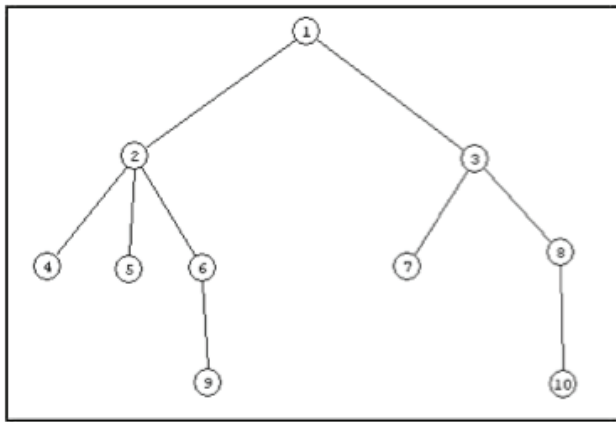


Figure 3. Not directed alarm tree

C. Conditional Probability

Conditional probabilities between alarms, which are also calculated from alarm files, are used to direct the graph. The conditional probability of alarm x given alarm y is:

$$P(A_x / A_y) = \frac{P(A_x \cap A_y)}{P(A_y)} \quad (3)$$

For example, the array of probabilities shown below reflects the conditional probabilities of alarm i given alarm j in the corresponding position (i, j) for i different of j, and the probability of occurrence of the considered alarm when i is equal to j (diagonal).

Table 3. Matrix odds

Matriz de Probabilidades:									
0.012	0.347	0.156	0.185	0.118	0.123	0.097	0.088	0.070	0.067
1.000	0.035	0.239	0.533	0.340	0.355	0.200	0.176	0.202	0.165
1.000	0.531	0.077	0.391	0.347	0.391	0.623	0.565	0.329	0.427
1.000	1.000	0.330	0.065	0.389	0.457	0.326	0.280	0.317	0.278
1.000	1.000	0.459	0.609	0.102	0.580	0.457	0.430	0.453	0.424
1.000	1.000	0.495	0.685	0.556	0.098	0.457	0.425	0.568	0.412
1.000	0.714	1.000	0.620	0.556	0.580	0.124	0.668	0.531	0.573
1.000	0.694	1.000	0.587	0.576	0.594	0.737	0.136	0.551	0.757
1.000	1.000	0.734	0.837	0.764	1.000	0.737	0.694	0.172	0.702
1.000	0.857	1.000	0.772	0.750	0.761	0.834	1.000	0.737	0.180

By hypothesis, when an alarm occurs descendants also should activate your alarm.

The probability of the father given the occurrence of the son will then be lower than the probability of the son given the occurrence of the father. Therefore, with these conditional probabilities the directed graph is constructed. Therefore, when probability of alarm i given alarm j or $P(A_i/A_j)$ is less than the probability of alarm j given i or $P(A_j/A_i)$, then the link is from i to alarm j.

If this applies to the previous undirected graph, it achieves the directed graph that was assumed at first, as you can see in the next figure (Fig. 4):

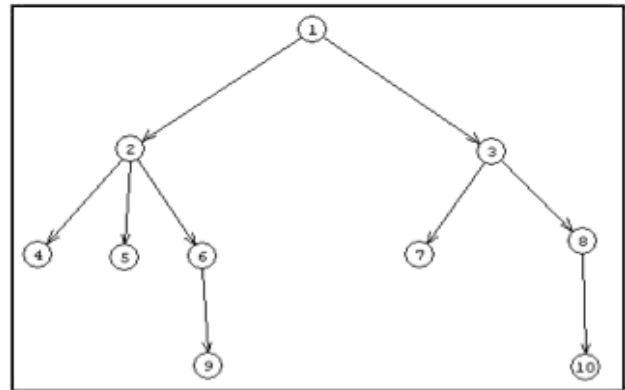


Figure 4. Prospect alarms tree

3. ADJACENCY MATRIX

A directed graph adjacency matrix is a squared matrix of dimension n, for n being the number of nodes (alarms). The matrix elements are 0, 1 for 1 when there is a link or edge between node i, j or 0 otherwise. For the analyzed example the next adjacency matrix was obtained:

Table 4. Adjacency matrix of directed graph

Matriz de Adyacencia del grafo dirigido:									
0	1	1	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

As shown, it is possible to find a simple way to calculate the adjacency matrix file of a set of data alarm, if the following three conditions are met,

$$r_{ij} > Threshold \quad (\forall k \neq i,j)$$

$$r_{ijk} > Threshold$$

$$P\left(\frac{A_i}{A_j}\right) < P\left(\frac{A_j}{A_i}\right)$$

If all the conditions are fulfilled, then there are links between alarms i and j.

The adjacency matrix has an important feature. Elevating adjacency matrix to different

powers, it results in several matrixes indicating connections between nodes in different generations. Elevating the adjacency matrix to the nth power, it is possible to obtain the connections of nodes separated by n edges. For example, the following connection matrices are obtained:

Table 5. Connection matrices

Matriz de conexiones de grado 2:									
Alarma 1:	0	0	0	1	1	1	1	0	0
Alarma 2:	0	0	0	0	0	0	0	1	0
Alarma 3:	0	0	0	0	0	0	0	0	1
Alarma 4:	0	0	0	0	0	0	0	0	0
Alarma 5:	0	0	0	0	0	0	0	0	0
Alarma 6:	0	0	0	0	0	0	0	0	0
Alarma 7:	0	0	0	0	0	0	0	0	0
Alarma 8:	0	0	0	0	0	0	0	0	0
Alarma 9:	0	0	0	0	0	0	0	0	0
Alarma 10:	0	0	0	0	0	0	0	0	0
Matriz de conexiones de grado 3:									
Alarma 1:	0	0	0	0	0	0	0	1	1
Alarma 2:	0	0	0	0	0	0	0	0	0
Alarma 3:	0	0	0	0	0	0	0	0	0
Alarma 4:	0	0	0	0	0	0	0	0	0
Alarma 5:	0	0	0	0	0	0	0	0	0
Alarma 6:	0	0	0	0	0	0	0	0	0
Alarma 7:	0	0	0	0	0	0	0	0	0
Alarma 8:	0	0	0	0	0	0	0	0	0
Alarma 9:	0	0	0	0	0	0	0	0	0
Alarma 10:	0	0	0	0	0	0	0	0	0
Matriz de conexiones de grado mayor que 3:									
Alarma 1:	0	0	0	0	0	0	0	0	0
Alarma 2:	0	0	0	0	0	0	0	0	0
Alarma 3:	0	0	0	0	0	0	0	0	0
Alarma 4:	0	0	0	0	0	0	0	0	0
Alarma 5:	0	0	0	0	0	0	0	0	0
Alarma 6:	0	0	0	0	0	0	0	0	0
Alarma 7:	0	0	0	0	0	0	0	0	0
Alarma 8:	0	0	0	0	0	0	0	0	0
Alarma 9:	0	0	0	0	0	0	0	0	0
Alarma 10:	0	0	0	0	0	0	0	0	0

As shown, at row i of any n order matrix, it is possible to observe the descendants of order n or alarm i. On the contrary, if column i is seen it is possible to know the ascendants of order n in this alarm.

4. DISCUSSION

For the analysis of any particular case, the Problem Vector approach is utilized. The Problem Vector is a n dimension vector, for n being the number of alarms, formed by 1 and 0's being 1 if the alarm is set or 0 otherwise. Pre-multiplying this vector by the distinct adjacency matrices, values different from 0's are obtained in the alarm positions that are nth grade descendant of the Problem Vector alarms. A series of vector with zeros in the nth order descendants is obtained using the NOT statement with all these products. Thus, it is possible to logically multiply the previous results to filter all the descendant alarms from the Problem Vector. Finally, the previous product is logically

multiplied by the Vector itself to obtain only the significant alarms belonging to the problem.

For example, assuming a vector problem with alarms 2, 8 and 10 the following Problem Vectors are:

For the example:

$$\text{Problem} * [A] = (0\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 0\ 1)$$

$$\text{Problem} * [A] = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0)$$

$$\text{Problem} * [A] = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

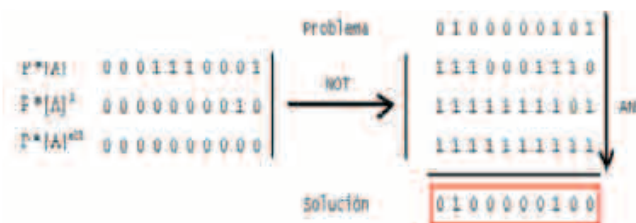
$$\text{Problem} * [A] = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

Applying the Boolean logical operations, the solution of significant alarms in a given time is:

$$\text{Solution} = (P) \text{ and } \{\text{not}(P * [A])\} \text{ and } \{\text{not}(P * [A]^2)\} \dots \text{ and } \{\text{not}(P * [A]^{n \geq 3})\}$$

being n such that $[A]^{n \geq 3}$ is the null array.

For the example:



Note that for the used problem vector the most significant alarms are 2 and 8, as seen in the following graph (Fig. 5).

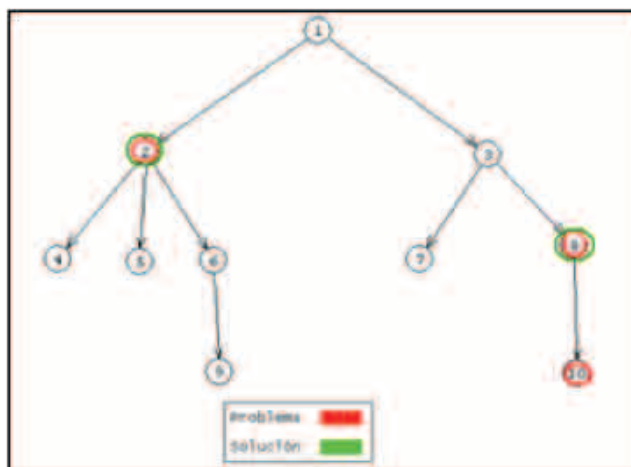


Figure 5. Example of analysis

5. SOFTWARE

To make all the calculations required for the creation of the directed graph of alarms, as well as for the analysis of any possible case of avalanche a software application was developed. This customized software ("Alarm Tree") was programmed in Visual C++. It allows the user to open a file of alarms, consisting of 0's and 1's, and to calculate the adjacency matrix, that allows the analysis of different alarm avalanches. As seen in the Fig.6, the graphical interface of the program is very simple, being formed by three clearly differentiated blocks: alarm file, calculation of adjacency matrix and analysis avalanche alarms.

In the first block, it opens the file alarm to be studied. It is read in order to count both the number of lines (temporary windows) and columns (alarms).

The second block, the adjacency matrix calculation, calculates the correlations between alarms, partial correlations between all them and their conditional probabilities. The adjacency matrix is built with all these calculations. Once done the calculation process, the matrices all data are displayed.

The last block allows the analysis or the active alarms to obtain the significant root alarms of the systems.

In addition, the software has an Exit key and contact information about the author of the program.



Figure 6. Program "Árbol de Alarmas"

6. CONCLUSIONS AND FUTURE WORK

Systematic treatment of massive alarms in electricity distribution and control centers is composed, typically, of a set of different methods that work together to solve the problem. Fig. 7 shows this methodology proposed by [1]. This work is intended to simplify the problem of alarm avalanches, achieving an easy way to obtain significant alarms in any case, starting with only data provided by the data acquisition system.

Future work is concerned with improving the software application. Among them it is possible to consider aspects such as incorporating a database with specific data from alarms and a response suggestion for them. Such improvement will help to analyze any avalanche and the decision-making process in terms of the root causes and possible failures in the system. Moreover, an application that allows the construction of the directed graphs and sub-graphs will be incorporated. Finally, other functions facilitating calculation handling, modifications or simulation of different conditions, can be added.

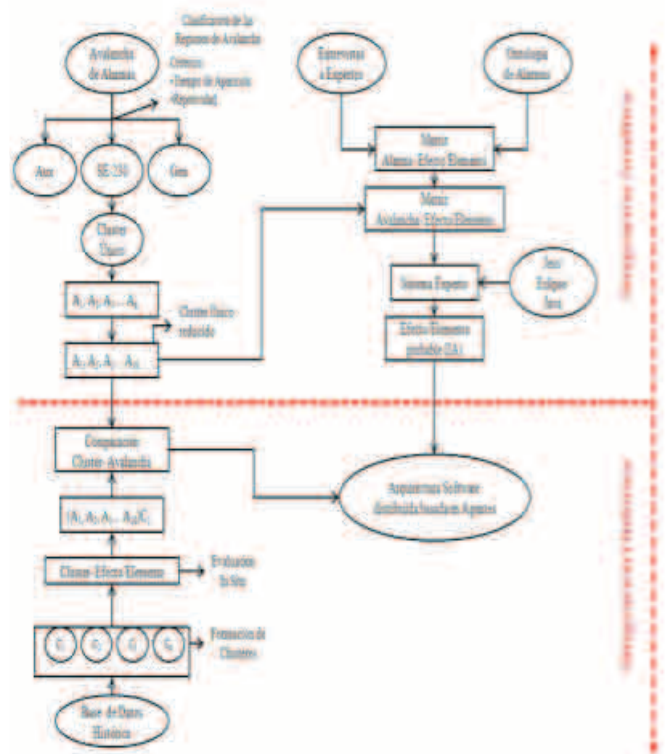


Figure 7. Usual methodology in the processing of massive alarms.

7. REFERENCES

- [1] O. Aizpurúa, et al., "A New Methodology of Massive Alarm System, in Electrical Power Administration". 7th Latin American and Caribbean Conference for Engineering and Technology, Laccet 2009.
- [2] K. Julisch, "Clustering Intrusion Detection Alarms to Support Root Cause Analysis". ACM Transactions on Information and System Security 6. 2003.
- [3] I. A. Beinlich, H. J Suermondt, R. M. Chavez, & Cooper, G. F. "The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks". 1989 Pearl, Judea. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann. 1988.
- [4] J. Orr, y D. Westenskow, Journal of Clinical Monitoring. "A Breathing Circuits Alarm System Based on Neural Networks". Vol. 10, n°2, p. 101-109. Orr, J., et al., 1994 & Windsor, C.G., 1993
- [5] B. Buchanan, and E. Short Life Rule-based expert systems, Readings (MA). USA: Addison-Wesley. 1984.
- [6] Horvitz, E. and Barry, M. Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence. Display of Information for Time-Critical Decision Making. p., 296-305. 1995
- [7] L. Blázquez, and L. De Miguel, "Diagnóstico Automático de Fallos para Sistemas Dinámicos No-Lineales". Universidad de Nuevo León y Universidad de Valladolid. 1999.