U.P.B. Sci. Bull., Series C, Vol. 82, Iss. 1, 2020 ISSN 2286-3540

BAYESIAN ALGORITHM ANALYSIS FOR BIOGAS LEAKAGES DETECTIONS ON FOG COMPUTING

Dragos Enache[1], Dan POPESCU[2]

The trend of SCADA (Supervisory Control and Data Acquisition) is to deploy in Cloud Computing domain and have as main advantage reducing the costs and removing the local hardware and software infrastructure. The most common method of communication between SCADA sites and Cloud servers is GSM (Global System for Mobile Communication) network.

Process plant placed in the field involve sometimes quick decisions in order to avoid hazard or malfunctioning processes. In these cases, the GSM communication support is not a convenient solution, due to the latency of the data transfer. When the result of data analysis requires immediate action and higher levels must be accessed as Cloud, a server installing into local architecture, able to communicate to the Cloud server is a solution called Fog Computing. The concept of Fog Computing has emerged due to data latency in Cloud Computing servers.

The article presents a solution for detecting hazardous situations, possible to occur in the biogas production area of a Wastewater Treatment Plant. The information is sent from the sensors via PLCs to the Fog Computing server as tags on OPC protocol. The data are analyzed using the Bayesian algorithm and has the effect warning the operating personnel about the imminence of a dangerous situation. A field maintenance control is requiring.

Keywords: data transmission, SCADA, remote control, FOG computing, biogas leakage, Bayesian algorithm

1. Introduction

In order to optimize performance, automation processes require adoption of internet technologies as Cloud Computing, IoT (Internet of Things) and subsidiary Fog Computing.

Cloud Computing offer a centralized access while Fog Computing provides a decentralized local access, as main difference. The decentralized approach is adopted because the centralized model hasn’t been able to cope resources to the devices located to the edge of the wide network. The decentralization concept has been developed as a response to the geographically

[1] Eng., Dept. of Automation and Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: enache.drago@gmail.com

[2] Prof. Eng., Dept. of Automation and Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: dan_popescu_2002@yahoo.com

spread infrastructure, grows of mobile architecture and an important increase in the use of IoT devices.

Fog Computing idea was also to reduce latency that plague centralized Cloud Computing. It emerged because Cloud Computing wasn't able to follow the growth of IoT devices. Rather than sending data inefficiently from IoT devices to the cloud, fog nodes analyses data on the network edge and avoid transferring the result back to the center of the network.

The article presents a solution to avoid hazardous situations that may occur in the biogas production area of a Wastewater Treatment Plant. Due the PLCs were not originally designed for collecting, archiving and analyzing data, a local solution as Fog computing is reliable into hazardous area when is mandatory a rapid decision.

SCADA servers, PLC's process and integrated server acting as Fog computing are connected into local network. The data from field sensors is retrieved via PLCs and transmitted throw local network to the Fog servers. SCADA servers use the data into technological process, while Fog servers collect data and base on specific software, evaluate situation and take a rapid decision, alerting the personnel, in order to prevent a hazardous situation.

Sending information as tags from PLC, to more computers, is possible due to the OPC protocol installed as clients on different local servers. The OPC protocol can be configured to receive information into client servers at different time interval. The interval between two receiving tags is known as "pulling". SCADA system use tags into technological process and the pulling is set in seconds. This interval is enough for technological process and also avoid a rapid fulfil of database. The pulling for OPC protocol installing into Fog servers is set in millisecond due the necessity of a rapid decision. This is one of the reasons to use Fog servers, instead to install the algorithm on SCADA servers.

Bayesian algorithm software is based on real situation data, which are archived and analyzed in order to determine the effect of other random events occurrence. The effect is immediate and warning the operating personnel about the imminence of a hazard situation. A site control of maintenance team is required. This method has been chosen because it can take into account events with different characteristics, like status (False-True) and predictors (Low-Medium-High), in order to determine through posterior probabilities a hazard situation.

As is shown in [1] even from the designing phase, a hazard analysis is performed, but is not possible to cover all possible situation. During the technological process operation, are identify new issues able to produce hazard. In exploitation is difficult to modify or adding new systems to controlling the issues and for that an indirect monitoring sensors become a solution to prevent dangerous situation. The risk management represents also the base for new control algorithms and strategies while process control and optimization are used to increase the safety and efficiency of industrial plants [2].

Bayesian algorithm is a method used in many other domains as medicine, industry, internet security, earthquake prediction and even by NASA [3] for probabilistic risk. Bayesian network provide effective and concise method for inference of causality, build with a probability graph model which represent the dependence between variables [4]. The Bayesian method implemented into the computer programed algorithm, offer a rapid result, 500ms, even the network points has over 400 nodes as show in [5]. Reliability of fault diagnoses analysis for complex engineering system is based traditionally on reliability of fault tree diagram which is limited methods. It has been demonstrated that Bayesian network method, has a great flexibility and it has been introduced into reliability engineering category [6]. In [7] are studied cause of gas leakage and the accidents triggered by gas leakage, using bow-tie analysis and Bayesian network, in order to confirm critical nodes of accidents introducing three measures: Birnbaum measure, risk achievement worth and Fussel-Vesely. Bow-tie analysis is a quantitative method which includes fault three analysis and event tree analysis in view of insufficient field failure data. Failure data was partly obtained from standard reliability data sources or fuzzy method based on expert judgement. Simplification to the standard Bayesian network model called Noisy-OR is described in [8]. There are explained how probabilistic risk importance measures can be applied to Bayesian networks in order to calculate supplier risk measures. Noisy-OR, treating the relationship between disruptions at parents and child nodes. Bow-tie and Bayesian network algorithms are use in [9] for the leakage failure of natural gas pipelines. Compared to the Bow-tie method, fault tree, and event tree method, Bayesian network carries out two-way analysis, not only to find the results from the causes, but also to find causes from the results.

The rest of this paper is organized as follows: Section 2 presents the biogas leakage detection, components and biogas leakage situation. Section 3 details the implementation of Bayes theorem and its practical application. In Section 4 – an example evaluates the situation. Conclusions are summarized in Section 5.

2. Biogas leakage detection

For testing, a biogas plant, Fig.1, located into WWTP (Wastewater Treatment Plant) has been used. The biogas is used to produce electricity for own WWTP consumption and for burners, to heat the biogas digester. A risk of explosion exists, due the leakage into the complex piping system. To prevent hazard situation, a biogas sensors are spread throw different places, in order to detect the leakages. Each sensor provide three alarm levels, depends of the biogas quantity detected.

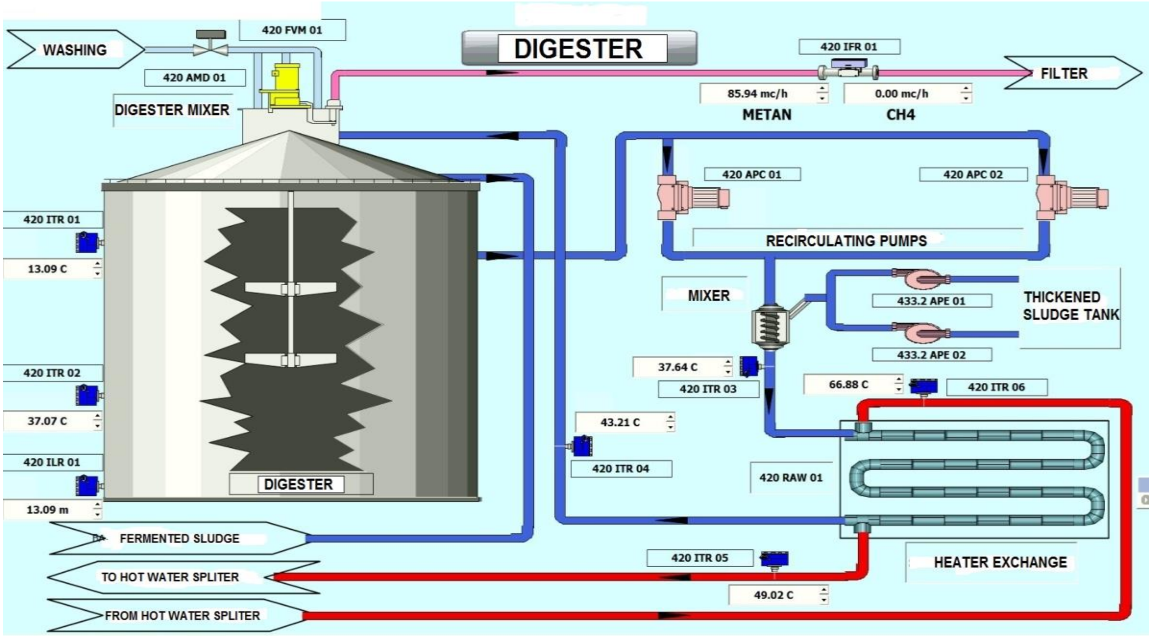

Fig. 1. Biogas system production in WWTP Campina City (HMI SCADA screen)

The local Ethernet network uses an Ethernet fiber optic ring support, to connect Siemens PLCs with WinCC SCADA redundant servers. Each PLC is connected to the network by fiber optic/wire switch convertor. The switch collect also the local HMI (Human Monitor Interfaces) and other PLC from dedicated equipment. The second Ethernet network Fig.2, is present also and collect data from biogas equipment's producers, throw four PLC, provided by Phoenix Contact. Each of this PLC sends also data to WinCC local SCADA.

In biogas control network, PLC's survey also four biogas sensors, situated into the biogas building production. Alarms levels are classified from Low to Medium and High. The Medium and High alarms have as effect closing the main emergency biogas valve and warning the personnel in local SCADA. The alarm is sent also into Regional Cloud Server. Closing the biogas valve, all installation in charge with biogas process and also adjacent equipment, are shutting down. In this case the personnel intervention is required, to evaluate situation and restart the whole process. Restarting biogas system, takes time, involve complex procedure and could affect the installations piping sludge.

The Low-level biogas sensors often indicated alarms for a short time period and the personnel has been alerted. A reason for this false alarm was found to be the sludge smell presented into the area. These leakages are also possible without biogas sensor detection due to the multiple piping joints. The main reasons for this situation are external temperature, pressure and temperature inside the biogas piping. These parameters are use in Bayesian algorithm presented in this paper.

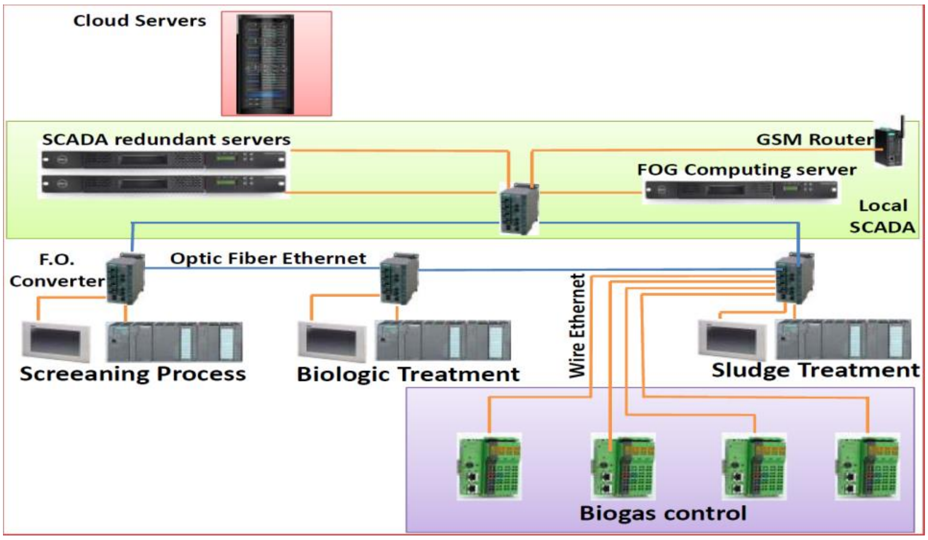

Fig.2. SCADA system architecture including biogas control PLC

Fig. 2 presents the system architecture structure including biogas control PLC and defines the transfer of information between the corresponding subsystems. The system structure follows a multilevel hierarchical model that allows the decomposition of the implemented functions.

The Low-level alarms more often appear and has been decided do not close the main biogas emergency valve, in order to avoid the process disturbance. This information is not ignored, because it can offer a clue for future dangerous biogas leakage and it was taken into account in Bayesian algorithm.

To analyzing data, software compatible with PLC is installed on Fog server. Developed software warns the personnel about imminent hazardous situation and sends the alarm in Regional Cloud.

3. Proposed methodology for alarm classifications

For grouping multi-dimensional data into groups (clusters) defined algorithmically, classification algorithms are used. The method is useful for large quantity information quantification and as main feature each group has more points with similar characteristics.

Considering Ω, a space with *n* elementary elements equally probable and events $P(A \cap B)$ and B having *m* respective *p* elementary elements. Certainly, event B occurred and $P(B) \neq 0$, probability to happened event A is:

$$P(A|B) = \frac{m}{p} \tag{1}$$

Dividing (1) to *n* obtains equation (2), called the "probability of event A, conditioned by event B" [10] and is noted:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2}$$

$P(A|B)$ - The probability of event A is fulfilled, when the event B is known to already been occurred.

$P(A \cap B)$ -The probability that events A and B to take place simultaneously.

$P(B)$ - The probability of event B to fulfil.

According with probability properties is known that:

$$P(A \cap B) = P(B \cap A) \tag{3}$$

According with (2) can conclude:

$$P(A|B) \times P(B) = P(B|A) \times P(A) \tag{4}$$

From (4) results:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{5}$$

In practice the events of Ω space can be partitioned into disjunctives events with probability other than zero. Otherwise exists disjunctive elements $A_1, A_2, \ldots A_n$ with $P(A_1), P(A_2), \ldots P(A_n) \neq 0$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$. In this case the events $A_1, A_2, \ldots A_n$, constitutes a complete systems of events. If $A_1, A_2, \ldots A_n$ is a complete system of events, whatever would be the event B, results total probabilities formula:

$$P(B) = P(B|A_1) \times P(A_1) + \ldots .P(B|A_n)P(A_n) \tag{6}$$

If $P(A) > 0$, replace $P(B)$ in (5) and result Bayes Formula:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \times P(A_i)}{\sum_{j=1}^{k} P(B \mid A_j) \times P(A_j)} \tag{7}$$

4. Alarms analysis

In our applications we made the following notation for applying the algorithm:

$P(c)$ - The probability of *class*.

$P(x)$ - The prior probability of *predictor*.

$P(c \mid x)$ - The posterior probability of *class* (c-as target) given *predictor* (features or attribute).

$P(x \mid c)$ *)* - The likelihood is the probability of *predictor given class*.

$$P(c \mid x) = \frac{P(x \mid c) \times P(c)}{P(x)} \tag{8}$$

Into biogas producing system, some different features has been identify as factors for biogas leakage. Notations for different status events are define:

Low – L
Normal – N
High – H
False – F
True – T

The possible status of these factors is presented in the Table 1, according with the notations.

Table 1

Possible events for analyzed factors into biogas system

Factors	Temp. External	Temp. Process	Pressure	Low Alarm	Control
Events	L	L	H	T	YES
	N	N	N	F	NO
	H	H			

In order to prevent a major leakage, pressure, external temperature and process temperature are monitoring, combining with the Low alarm biogas leakage monitoring. Events combination observed on site for monitored factors are presented in table 2. The list of possible factors can be extended.

Base on this, a frequency tables for each attributes, versus the targets are presented in Fig.3. Transforming the results into likelihood table, posterior probability can be calculated. Naive Bayes, suppose that effect of the predictor "*x*" value's, belong to a specific class "*c*", is independent from the other predictor's values.

Table 2

Experimental events occurrence

No.	Temp. External	Temp. Process	Pressure	Low Alarm	Control	No.	Temp. External	Temp. Process	Pressure	Low Alarm	Control
1	L	H	H	F	NO	8	L	N	H	F	YES
2	L	H	H	T	YES	9	L	L	N	F	NO
3	N	H	H	F	YES	10	H	N	N	F	NO
4	H	N	H	F	NO	11	L	N	N	T	NO
5	H	L	N	F	NO	12	N	N	H	T	YES
6	H	L	N	T	NO	13	N	H	N	F	YES
7	N	L	N	T	NO	14	H	N	H	T	YES

Table 3 present an example of the likelihood for "*Temperature External*" factor with *"L"* predictor and *"Yes"* and *"No"* classes. These are necessary to be computed for all factors, in order to fulfill the likelihood table in Fig.3.

Table 3

Likelihood example

		Control		
	x	**P(x|YES)**	**P(x|NO)**	**P(x)**
Temp. External	L	2/6	3/8	5/14
	N	3/6	1/8	4/14
	H	1/6	4/8	5/14
		6/14	8/14	
		P(Yes)	**P(No)**	

The likelihood for *"L|Yes"* - affirmative probability is in (9):

$$P(x\,|\,c) = P(L\,|\,Yes) = \frac{2}{6} = 0.333 \tag{9}$$

The prior probability of *"L"* predictor is in (10):

$$P(x) = P(L) = \frac{5}{14} = 0.357 \tag{10}$$

The probability of *"Yes"* class is in (11):

$$P(c) = P(Yes) = \frac{6}{14} = 0.428 \tag{11}$$

Posterior Probability for “*Yes\L*” is in (12):

$$P(c \mid x) = P(Yes \mid L) = \frac{P(L \mid Yes) \times P(Yes)}{P(L)} = 0.380 \tag{12}$$

The likelihood computing for “*L|No*” is similar with “*L|Yes*”:

$$P(x \mid c) = P(L \mid No) = \frac{3}{8} = 0.375 \tag{13}$$

$$P(x) = P(L) = \frac{5}{14} = 0.357 \tag{14}$$

$$P(c) = P(No) = \frac{8}{14} = 0.571 \tag{15}$$

Posterior Probability for “*No\L*” is similar with (12):

$$P(c \mid x) = P(No \mid L) = \frac{P(L \mid No) \times P(No)}{P(L)} = 0.342 \tag{16}$$

Frequency Table

		Control	
		YES	NO
Temp. External	L	2	3
	N	3	1
	H	1	4

		Control	
		YES	NO
Temp. Process	L	0	4
	N	3	3
	H	3	1

		Control	
		YES	NO
Pressure	N	1	6
	H	5	2

		Control	
		YES	NO
L Alarm	T	3	3
	F	3	5

Likelihood Table

		Control	
		YES	NO
Temp. External	L	2/6	3/8
	N	3/6	1/8
	H	1/6	4/8
		6/14	8/14

		Control	
		YES	NO
Temp. Process	L	0/6	4/8
	N	3/6	3/8
	H	3/6	1/8
		6/14	8/14

		Control	
		YES	NO
Pressure	N	1/6	6/8
	H	5/6	2/8
		6/14	8/14

		Control	
		YES	NO
L Alarm	T	3/6	3/8
	F	3/6	5/8
		6/14	8/14

Fig. 3. Frequency table vs Likelihood table

5. Computing of unconsidered situation

Using the final results of posterior probabilities, we can compute if a combination of factor, other than initial monitored events, is possible to occur, according with Table 4:

Supposing occurrence of events, excepting experimental table (Table 2), we intend to determine if it's necessary a control, if *"Low Alarm"* is *"F"* or *"T"*. In Table 4 we consider *"Low Alarm"* as *"F"*.

Table 4

Supposing "Low Alarm" as "F"

Temp. External	Temp. Process	Pressure	Low Alarm	Control
H	H	H	F	?

According with the example from Table 3, the posterior probability calculation for classes "*Yes*" in (17) and "*No*" in (19) are performed for all events predictors: "*Temperature External*|*H*" "*Temperature Process*|*H*", "*Pressure*|*H*" and "*L Alarm*|*F*".

$$P(Yes \mid X) = P(H \mid Yes) \times P(H \mid Yes) \times P(H \mid Yes) \times P(F \mid Yes) \times P(Control \mid Yes) \tag{17}$$

Replace in (17) according with likelihood table:

$$P(Yes \mid X) = \frac{1}{6} \times \frac{3}{6} \times \frac{5}{6} \times \frac{3}{6} \times \frac{6}{14} = 0.01488 \tag{18}$$

$$P(No \mid X) = P(H \mid No) \times P(H \mid No) \times P(H \mid No) \times P(F \mid No) \times P(Control \mid No) \tag{19}$$

Replace in (19) according with likelihood table:

$$P(No \mid X) = \frac{4}{8} \times \frac{1}{8} \times \frac{2}{8} \times \frac{5}{8} \times \frac{8}{14} = 0.00558 \tag{20}$$

The final posterior probabilities can be standardized between 0 and 1 performing arithmetic average, in order to have percentual comparing of the probability. If the result of class "*Yes*" is more than 50%, a control of biogas system is necessary.

$$P(X \mid Yes) = \frac{P(Yes \mid X)}{P(Yes \mid X) + P(No \mid X)} = 0.7272 \tag{21}$$

$$P(X \mid No) = \frac{P(No \mid X)}{P(Yes \mid X) + P(No \mid X)} = 0.2727 \quad (22)$$

Can be concluded that in this situation, exist a 72.72% risk of biogas leakages and a control on site is necessary.

For *"T"* of *"Low Alarm"* events, in Table 5, consider predictors for events, other than Table 2. Compute then the posterior probability for classes *"Yes"* in (23) and *"No"* in (25).

Table 5

Supposing "Low Alarm" as "T"

Temp. External	Temp. Process	Pressure	Low Alarm	Control
L	H	N	T	?

$$P(Yes \mid X) = P(L \mid Yes) \times P(H \mid Yes) \times P(N \mid Yes) \times P(T \mid Yes) \times P(Control \mid Yes) \quad (23)$$

Replace in (23) according with likelihood table:

$$P(Yes \mid X) = \frac{2}{6} \times \frac{3}{6} \times \frac{1}{6} \times \frac{3}{6} \times \frac{6}{14} = 0.0059 \quad (24)$$

$$P(No \mid X) = P(L \mid No) \times P(H \mid No) \times P(N \mid No) \times P(T \mid No) \times P(Control \mid No) \quad (25)$$

Replace in (25) according with likelihood table:

$$P(No \mid X) = \frac{3}{8} \times \frac{1}{8} \times \frac{6}{8} \times \frac{3}{8} \times \frac{8}{14} = 0.0075 \quad (26)$$

For percentual result an arithmetic media is performed:

$$P(X \mid Yes) = \frac{P(Yes \mid X)}{P(Yes \mid X) + P(No \mid X)} = 0.4402 \quad (27)$$

$$P(X \mid No) = \frac{P(No \mid X)}{P(Yes \mid X) + P(No \mid X)} = 0.5597 \quad (28)$$

Base on monitored events of different factors, this method can evaluate through posterior probabilities, if an incidences of events, other than experimental occurrences, results in a possible biogas leakage into monitored area.

For analyses, Table 4 take into account *"Low Alarm"* as "*F*" and Table 5, "*T*" for *"Low Alarm"*, but different predictor for the rest of events.

A result for class *"Yes"* over 50% of biogas leakage probability, as was evaluated in Table 4, generates an alarm for SCADA operator's personnel. This alarm indicates a possible hazard into biogas area and requires a mandatory control of biogas equipment's, on site. This information is sent as alarm in Regional cloud server for archiving. In (28), the class*"No"* is less than 50% and the alarm is not generated.

Even in case of "*T*" predictor for *"Low Alarm"*, presented in Tabel 4, the algorithm indicates for class *"Yes"*, 44.03% probability risk of biogas leakage. This case doesn't generate an alarm and is not mandatory site equipment's control. The class *"No"* indicates 55.97% risk of biogas leakage. It is considering a false alarm, because according to the complementary monitored factors, do not present an imminent leakage conditions.

The algorithm is implemented into Fog server, developed under SCADA software, dedicated for biogas leakage detection. The alarms are sending via local network, to local process SCADA server by OPC UA protocol and also to Regional Cloud Server via VPN connection.

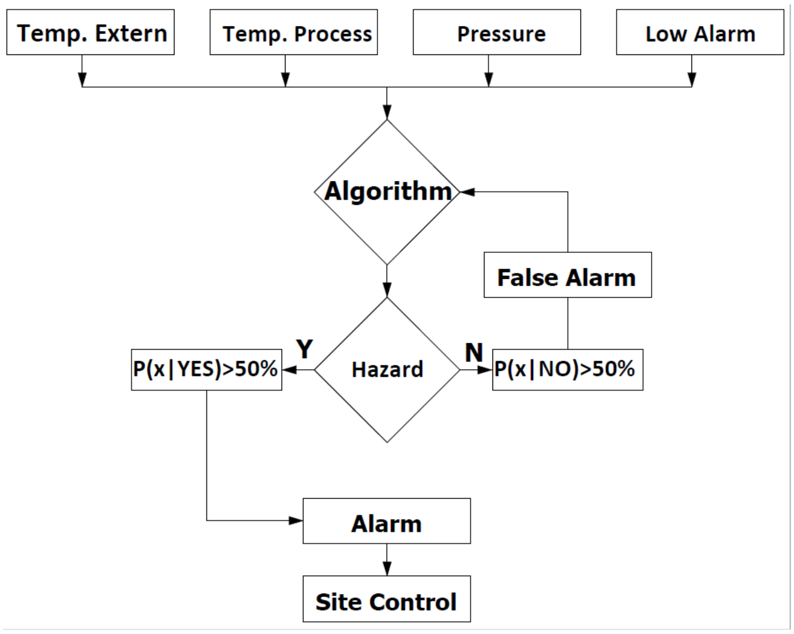

Fig. 4. Control site decision tree

6. Conclusions

With this algorithm, the SCADA personnel, is advice about an imminent biogas leakage, even a *"Low Alarm*" from biogas sensors is not active. In this way is avoiding a hazardous situation.

The sensors send information simultaneous, via PLC to SCADA and Fog servers. While SCADA use the information for technological process, Fog computing take rapid decision in order to prevent a hazard, evaluating situation according to the implemented algorithms.

The events table can be extended in future with other possible factors which can be introduced into algorithm in order to have a detailed evidence of effects which can produce biogas leakages.

R E F E R E N C E S

[1] *L. Ocheana, D. Popescu, O.Rohat,* Remote risk and hazard control. International Journal of Emerging Trends in Engineering and Development Issue 3, Vol.5 (September 2013) .

[2] *G. Florea, L. Ocheana, D. Popescu, O.Rohat*, Emerging tehnologies – the base for the next goal of Process Control - Risk and Hazard Control, Recent Advances in Signal Processing, Computational Geometry and Systems Theory. Florence Italy August 23-25 2011

[3] *H. Dezfuli, D. Kelly, K. Vedros, W. Galyean*, Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis, National Aeronautics and Space Administration, June 2009

[4] *X. Kong, J. Yao, Z. E, X. Wang,* Generation Expansion Planning Based on Dynamic Bayesian network Considering the Uncertainty of Renewable Energy Resources. Energies 2019, 12, 2492; doi:10.3390/en12132492

[5] *O.G. Mengshoel, A.Wiche, S. Uckun,* Sensor Validation using Bayesian Networks, 9th International Symposium on Artificial Intelligence, Robotics, and Automation in Space (iSAIRAS-08), February 25-29, 2008, Los Angeles, CA.

[6] *S.Chen, Z.Qi, D. Chen, L. Guo, W. Peng,* Investigation of Bayesian network for reliability analysis and fault diagnosis of complex systems with real case applications, Advances in Mechanical Engineering 2017, Vol. 9(10) 1–18

[7] *Fang Yan, Kaili Xu, Xiwen Yao, Yang Li,* Fuzzy Bayesian Network-Bow-Tie Analysis of Gas Leakage during Biomass Gasification, July 27 2016, https://doi.org/10.1371/journal.pone.0160045

[8] *Anssi Käki, Ahti Salo, Sri Talluri*, Disruptions in Supply Networks: A Probabilistic Risk Assessment Approach, Journal of Business Logistics · June 2015 https://www.researchgate.net/publication/272563945

[9] *Xian Shan, Kang Liu, Pei-Liang Sun,* Risk Analysis on Leakage Failure of Natural Gas Pipelines by Fuzzy Bayesian Network with a Bow-Tie Model, Volume 2017, Article ID 3639524, https://doi.org/10.1155/2017/3639524

[10] *L. Florin* Artificial Intelligence: probabilistic reasoning, classification techniques, Editura Tehnopress Iasi 2012xx