

Traffic characterization in a communications channel for monitoring and control in real-time systems

Leonardo Serna-Guarin¹, Luis J. Morantes-Guzman², Edilson Delgado-Trejos³, Miguel A. Becerra⁴

^{1,2}AEyCC Research Group, Instituto Tecnológico Metropolitano ITM, Medellín, Colombia

³AMYSOD Lab, Instituto Tecnológico Metropolitano ITM, Medellín, Colombia

⁴Institucion Universitaria Pascual Bravo, Medellín, Colombia

Article Info

Article history:

Received Oct 10, 2019

Revised Dec 20, 2020

Accepted Dec 22, 2020

Keywords:

Communication networks

Machine learning

Real-time systems

Remote control

Traffic analysis

ABSTRACT

The response time for remote monitoring and control in real-time systems is a sensitive issue in device interconnection elements. Therefore, it is necessary to analyze the traffic of the communication system in pre-established time windows. In this paper, a methodology based on computational intelligence is proposed for identifying the availability of a data channel and the variables or characteristics that affect the performance and data transfer, which is made up of four stages: *a*) integration of a communication system with an acquisition module and a final control structure; *b*) communication channel characterization by means of traffic variables; and *c*) relevance analysis from the characterization space using SFFS (sequential forward oating selection); *d*) Channel congestion classification as Low or High using a classifier based on Naive Bayes algorithm. The experimental setup emulates a real process using an on/off remote control of a DC motor on an Ethernet network. The communication time between the client and server was integrated with the operation and control times, to study the whole response time. This proposed approach allows support decisions about channel availability, to establish predictions about the length of the time window when the availability conditions are unknown.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Leonardo Serna Guarin,
AEyCC Research Group,
Instituto Tecnológico Metropolitano ITM,
CL 54A No 30-01, ITM Campus Fraternidad, Medellín, Colombia.
Email: leonardoserna@itm.edu.co

1. INTRODUCTION

Monitoring and remote-control systems are used in applications where access is limited or when the user is exposed to physical risk (e.g., extreme temperatures, polluted air, chemical substances among others) [1][2]; consequently, critical conditions embedded in a control system are managed or operated remotely [3]. These systems are supported by communication channels for data transfer where performance conditions become essential factors in remote monitoring system design for industrial control [4][5]. In this sense, statistical conditions of traffic and channel characteristics determine the minimum time required for a data flow throughout a network from its origin to its destination [6]. In this process, the network latency derived from the communication protocols is involved, which are used in industrial communication systems [7]. It is important to note that Ethernet networks are applied in industrial environments and represent a smaller cost in terms of implementation and maintenance, and consequently, they facilitate the implementation of solutions in data transfer processes [8][9]. In data transfer solutions the response time for a control system is critical, therefore it is necessary to analyze the traffic of the network to determine the restrictions and conditions of the

communication system (hardware and software) that must be evaluated in order to monitor and control a plant (element to control) at pre-established time windows. In this way, a delay time limit of 1 *ms* to 100 *ms* is defined in generic scenarios in real-time depending on the application, and a time between 10 *ms* and 100 *ms* is applied for real-time software in monitoring and automation processes, while factory automation conditions requires timings between 0.25 *ms* and 10 *ms* [10]. Despite multiple studies, it is considered an open research field.

In this study, a methodology for identifying variables that affect a communication system is proposed. First, a channel identification was carried out applying statistical analysis and machine learning procedures to analyze the traffic behavior in communication networks for decisions making as expressed in other research [11][12][13]. In this work, 32 traffic characteristics of a business network communication systems were measured, and a database of the circulating traffic in the network collected for three months. The database was analyzed by applying machine learning to identify redundant and relevant characteristics for establishing the congestion of the channel. Once the variables of the system were selected, a statistical analysis was carried out in a laboratory prototype included in an embedded device for sampling a channel in any communications system. In this way, a business network was emulated in a campus university for controlling a DC motor 6v12v-140RA using an ON/OFF control. It was tested through the channel. The control operation was performed in different states of network congestion to determine the relationship between the congestion of the channel and the response capacity supported by it to operate the said engine.

In the section 2, state of the art is presented from three perspectives: *i*) relevant features of a communication channel, *ii*) influence of the network architecture on traffic behavior, *iii*) characterization of the system for remote monitoring and *iv*) classification of the congestion as High and Low. In section 3 is presented the experimental setup. Then, results and discussion are presented in section 4, followed by conclusions and future work.

2. RESEARCH METHOD

2.1. Background

Relevant features of a communication channel

LAN environments have been widely used (specifically Ethernet), but they present problems of latency, reliability, and synchronization for implementing applications that require real-time support [14]. In practice, Ethernet technology operates in a non-deterministic way to access the medium and presents difficulties for real-time systems, especially where the topology involves concentrators and repeaters; despite that the networks evolved to the switched topologies, the increase in data traffic in the medium generates more significant latency in the channel due to the high flow of packets in the network[15]. Therefore, the standard parameters have been modified for some applications in the industry [16][17][18], but the high cost of the industrial protocols, that enables the interconnection from devices to the network, it has lead to the exploration of new alternatives, applying Ethernet and extending this application through the Internet and supported with the protocol synchronization NTP (Network Time Protocol) [19]. Similarly, ethernet features have been studied for industrial applications [20] as maximum transfer delay, jitter in transmission, available bandwidth, packet loss [21] and the network latency should also be into account, as well as the time demanded by the servers to answer the customers [22], since this affects the real-time conditions of the system [23]. The authors [24] relate three kinds of real-time: the first is real-time software with scalable cycle time for factory and process automation, the second is for more demanding responses where the cycle time ranges between 1 *ms* and 10 *ms* for control processes and finally the third class requires an isochronous (constant rhythm, with periods of equal duration) real-time with cycle time between 250 *us* and 1 *ms* and a phase fluctuation of less than 1 *us* for motion control. In [25] was analyzed the computing stations influence, due to generated times which are added to the communication systems, considering delays of controls, sensors, and actuators. These studies require tools for data capture and information analysis comprise mathematical models and proofs of real data that relate with these models, which can be executed in simulated mode, so results should evidence the presence of errors that can be distinguished from the real data [26]. A feature in a communication channel is the rate of transmission [27], where the measure of the data flow of the channel is obtained using the ping tool, which determines the transmission delay of the packets. In the same way, in [28] the Round Trip Time (RTT) measure is applied, which details the time required by the packets to reach their destination and return to their origin. Nevertheless, it is essential to bear in mind the cache phenomenon of operating systems, which does not allow a reliable measure of RTT [29]. Another features applied to measure the quality of the data is the packet loss and the number of packets that travel across the channel. On the other hand, in the last decade, software and hardware

tools have frequently been used to determine relevant parameters like jitters, delays in transmission, throughput, and periods without connectivity of the network [30]. Likewise, other variables that are part of the communication system must be considered, such as processing, memory, applications, and the number of available or occupied resources.

Influence of the network architecture on traffic behavior

Automation architectures consist of intelligent devices connected by a local or global communication network, which simplifies their complexity and expands the development of network control systems. Initially, it was essential to reduce the system wirings to ease the information exchange among different components of the system. Later, the trend was to use the same network technology at all levels in the industrial organizations [31][32]. Mainly, these network architectures are based on the Ethernet standard, which is used for its advantages in interconnecting industrial devices and giving the possibility of improving performance in real-time applications [33][34].

Improve or change the performance of a communications network requires to be analyzed from three levels: architecture, technology, and services [35]. The communication architecture consists of the physical organization and logical functioning of nodes linked by switches, where these links are modeled by buffers. However, every architecture requires a communication protocol and interfaces made up by applications and users [36]. The network calculus, introduced by Rene L. Cruz in [37][38][39], only assumes that the number of bytes sent on the network links does not exceed an arrival curve service (traditionally, a leaky bucket) which can represent both periodic and aperiodic traffic. The network calculus then determines the packet maximum end-to-end delay or jitters from exchange matrices which model the traffic between the industrial devices. A widely accepted model of real-time industrial communications [40][41] establishes that network infrastructure changes considerably and affect the real-time performance and traffic conditions [42][43]. Thus, the impact of varying parameters such as communication mechanism, network architecture, network traffic, data size, number of connections and processor load on the peer-to-peer interlocking performance must be analyzed as a real-time operating condition [44]. The technology refers to the characteristics of the devices and the channel and, similarly, the services involve the kind of traffic and the applications that the communications system could support [45].

Characterization of the system for remote monitoring

Remote access for controlling and monitoring various devices in an industrial environment is of value to engineers and automation plants [46][47]. Although Ethernet is the dominant technology in communications, in industrial applications, it requires meeting real-time performance requirements, traffic restrictions, and information security. Therefore, tunnels are incorporated into the network platform for monitoring and control via the Internet [48][49], demonstrating the sensitivity of the conditions of availability, integrality, and confidentiality to the channel and data traffic. To all this, the combination of technologies to remotely monitor and control processes is added through the use of the GSM network [50][51] which incorporates, together with Ethernet and TCP/IP, solutions in the control plants exploiting new technologies in automation and process control. In monitoring, remote access to the industrial network requires security [52], because Internet-based accessibility increases the vulnerability of these systems.

Remote communication is based on the synchronization between local control and remote monitoring, and the quality of the data transmission depending on the characteristics of the channel and the configuration of the system. These configurations, such as quality of service [53] and temporary memory technologies, provide compensation for the time delay required to minimize the uncertainty in the transmission of data in a critical time interval. Likewise, the implementation of a robust network [54][55] protects against data distortion during transmission, taking into account that industrial networks are specially designed for critical data transmission in multiple industrial applications [23].

The identification of the communication system involves analyzing all the possible characteristics involved in the mentioned aspects of security and architecture where the related variables include traffic behavior and the response of the network [56]. This monitoring process seeks to identify latency of the network, loss of data, identification of routes, bandwidth, percentage of the occupation, and use of resources. This process is done actively when additional traffic is injected into the network or passively when it is based only on the observed resources [57]. For both cases, the SNMP (Simple Network Management Protocol) protocol is used, which allows access to the operation conditions of each of the system's devices [58]. In the case study of the experiment, characteristics related to the channel were taken as transfer rate, bandwidth and packet loss, and aspects of the devices related to its performance as memory and processor use.

2.2. Experimental setup

In the selected study case, critical variables of the channel were considered and the response time emitted by a controller when a control signal is sent through the network that communicates it. With this, the relationship between the congestion of the channel and the response times in it was observed, and it was found how the behavior of the system altered the operating requirements in a control operation [13].

In this work, a business network was used, in where measurements of different the parameters of the device and channel traffic were considered. In the proposed methodology, four steps are used to examine the availability of the channel: in the first step, the variables used are evaluated by Pearson's correlation to identify the redundant system parameters. Multivariate data have a dependency structure between the variables, and these relationships are evaluated between peers and between all. The dependence between variables is evaluated using the correlation coefficient, for two variables x_j and x_k , with the properties: $0 \leq |r_{jk}| \leq 1$, $x_{ij} = a + bx_{ik}$ with $|r_{jk}| = 1$ for an exact linear relationship between the variables and r_{jk} is invariant before linear transformations of the variables. It is expressed:

$$r_{jk} = \frac{S_{JK}}{S_J S_K} \quad (1)$$

Pearson's correlation computes a coefficient that measures the degree of the linear relationship between two variables; the coefficient is an index that measures the degree of covariation or the structure of dependence between different linearly related variables. It is represented by r and is obtained by typing the average of the products of the differential scores of each case in the two correlated variables:

$$r_{xy} = \frac{\sum x_i y_i}{n S_x S_y} \quad (2)$$

where x_i and y_i refer to the differential scores of each pair, n is the number of cases and S_x and S_y to the deviations typical of each variable. The absolute values of the correlation coefficients range between 0 and 1, although the numerical result fluctuates between the ranges of +1 to -1 [59]. It is defined as the ratio between covariation and the square root of the product of the variation in X and the variation in Y [60]:

$$r = \frac{N \sum(XY) - (\sum X \sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (3)$$

In the second step, the feature selector sequential forward floating selection (SFFS) is used to find the features that contain the most system information. In floating search methods, sequential forward floating selection (SFFS) has been commonly used [61], where the number of added features can also change at each step, and these wrapper routines search for a considerably smaller number of subsets.

Sequential forward selection *SFS* is a bottom-up search procedure which adds new features to a set of functions one at a time until the final feature set is reached. The characteristics of the database are enabled or selected beforehand for the design of the classifier, and the *SFS* (*Scalar Feature Selection*) allows to determine a sufficient number of characteristics. Likewise, the good selection of information-rich in characteristics, Simplify the design of the classifier. In addition to removing the redundant variables and choosing the most representative characteristics, the evaluation of their performance is important in the design of the classification system, in which the probability of classification error is considered [62]. With *SFS* the characteristics are treated individually, any criterion of the measure of class separability can be adopted as *ROC* (*Receiver Operating Characteristic*), *FDR* (*Fisher's Discriminant Ratio*), *one-dimensional divergence*, among others. The value of the criterion $C(R)$ is calculated for each of the characteristics, $R = 1, 2, \dots, m$. The characteristics are sorted in descending order of values of $C(R)$. The l feature corresponds to the l best value of $C(R)$ and shown in a feature vector. The *one-dimensional divergence* d_{ij} is used and calculated by each pair of classes, where the corresponding $C(R)$ for each characteristic is expressed as the minimum divergence value overall peer classes [63]

$$C(R) = \min_{ij} d_{ij} \quad (4)$$

Through steps 1 to 3 of the following algorithm, the cross-validation operation is detailed for the selection of the best feature set with the sequential search procedure:

Step 1. Divide the data into training and test sets.

Step 2. Specify the search strategy: *SFS*, *GSFS*, At each stage of the algorithm: Generate subsets of features for evaluation.

Cross validation procedure:

- Divide the training data into 10 equal parts, ensuring that all classes are representative in each part; It uses nine parts for training and the remaining part for testing.
- Train the classifier model for each subset of variables, h , in each subset, k , of the training data, in turn perform the tests of the remaining parts. Get the performance, $CV(h, k)$.
- Average result

$$CV(h) = \frac{1}{10} \sum_k CV(h, k) \tag{5}$$

Select the smallest subset of features, S_{h^*} so that $CV(h)$ is optimal for the next stage of the search.

Step 3. Evaluate the test data set with the smallest subset of features, S_{h^*} , from the search procedure that offers the best performance, learning throughout the training set and performance evaluation with the test to the set.

In the next step, the Bayes classifier is used for training and learning the system with the characteristics found in the previous two steps.

This classifier is based on Bayes' theorem and allows a general solution to the problem when the parameters are known [64][65]. Given the classes $w_1, w_2 \dots, w_M$ assigned to M and known the pattern, which is represented by a characteristic of the vector x , the M conditional probability $P(w_i|x), i = 1, 2, \dots, M$ is formed. This is known as a posterior probability. Having both classes w_1, w_2 assumes that the prior probabilities $P(w_1), P(w_2)$ are known. N is the total workouts and N_1 and N_2 belong to w_1 and w_2 respectively, then $P(w_1) \approx \frac{N_1}{N}$ and $P(w_2) \approx \frac{N_2}{N}$ and the probability density function is assumed to be known $p(x|w_i), i = 1, 2$, the basic Bayes rule is defined

$$P(w_i | x) = \frac{p(x|w_i)P(w_i)}{\sum_{i=1}^2 p(x|w_i)P(w_i)} \tag{6}$$

This classification study is validated using a cross-validation procedure is performed to validate the classification method. The *Cross-Validation* strategy consists of dividing a set of samples of size n that can be analyzed into two data sets. One of these sets train the samples it contains, and the results obtained are applied to the other set that is used for the classification of samples or the estimation of the error [66]. Validation is the average overall possible training sets of size $n - 1$. In order to determine an appropriate model, cross-validation is calculated by each member of the candidate family of models, $M_k, k = 1, \dots, K$, and the M_k model is selected where

$$CV(\hat{k}) \leq CV(k) \text{ para todo } k \tag{7}$$

Cross-validation allows seeing and comparing the performance of a classifier in a given problem. The objective is to minimize the error. The classifier is trained until the minimum validation of the error is reached, keeping in mind not to include training points in the validation.

Proposed procedure

Seen the above then, the most relevant parameters were tested through a practical laboratory implementation. The proposed procedure is shown in Figure 1 alongside graphic the control on the web system using as plant a DC motor.

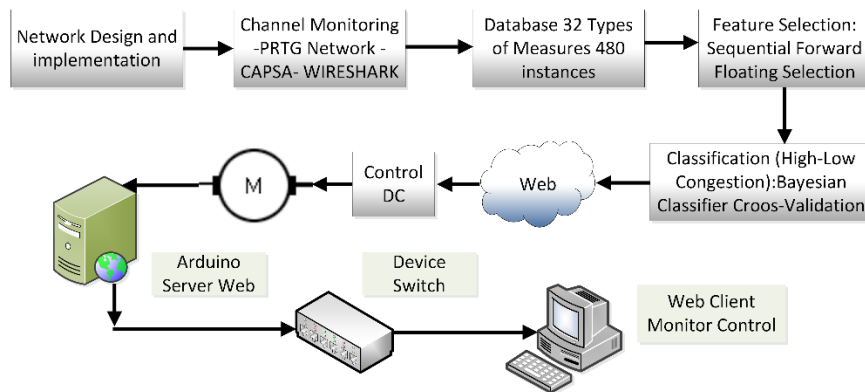


Figure 1. Control Motor DC via web

Two scenarios were analyzed: the first was the collection of data in the business network and the second was an emulation in a laboratory network (A P2P connection was implemented using an Ethernet channel built into a corporate network), in this traffic was generated in several states to simulate the high and low congestion of the channel. In both cases, the data was passed by the Characteristic Selector, the Bayes classifier, the cross-validation, and the control operation of an engine as described in the following sections.

Network structure and data acquisition

The data network was made up based on wireless technology heterogeneous architecture [67][68], which is depicted in Figure 2. Measurements were acquired from a channel, and then, they were processed and represented in a specific scenario to validate the methodology used (see Figure 1).

The communication channel availability was characterized using monitoring equipment in the border router (connection to the Internet Service Provider –ISP) and in an internal router which accesses the corporate LAN using wireless techniques.

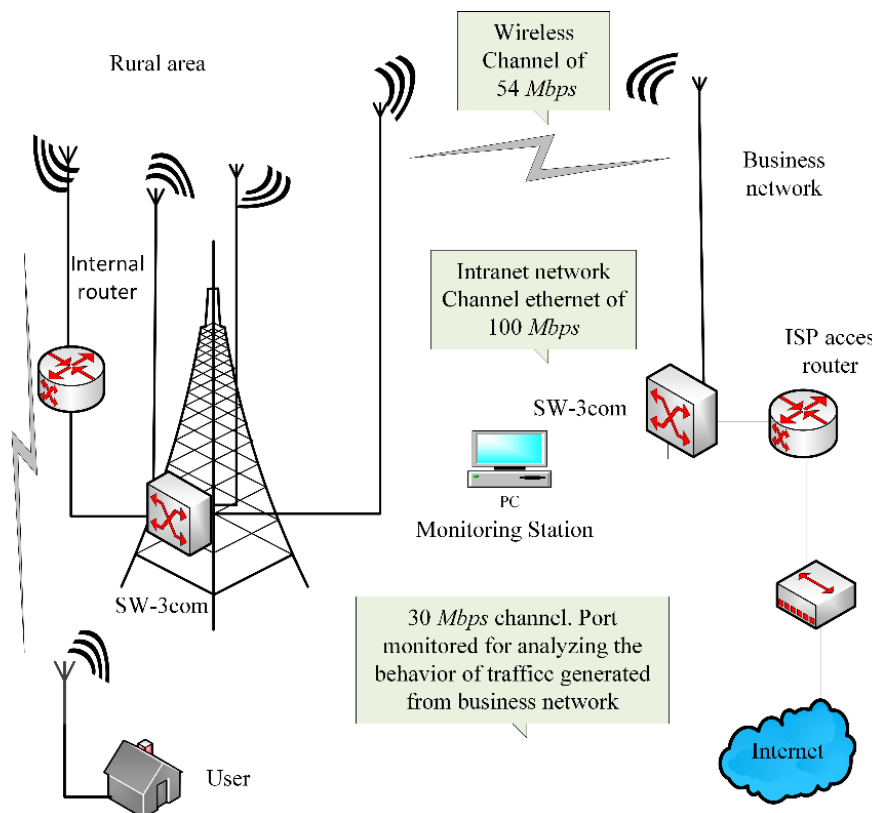


Figure 2. Remote monitoring in a real-time system

The data acquisition of traffic measurements was performed using the PRTG Network Monitor software, version 13.01, by Paessler. In order to execute different tests for different congestion states, two opposite levels of congestion were established as labels (i.e., low and high congestion). Channel congestion reduced the quality of the service when the network carries more data than it can support. These effects generated a delay in the queue, loss of packets, blocking of new connections, and decreased the network performance using the Traffic Jam of Quest Free Network Tools and the Ping application for packet transfer of variable size and low/high file transfer between the monitoring and control stations. The applications used for this data analysis were Wireshark 1.10.1 GNU and Capsa 7.72, both freely distributed. For each congestion level, 31 repetitions were performed to average the results and reduce noise effects.

Channel status characterization

The characterization space was made up using 32 features (see Table 1), $X = \{x_1, x_2, \dots, x_{32}\}$, related to traffic, jitter, processing time, and hardware state (like CPU performance and memory). These features were acquired in 480 instances, divided into two groups: group 1 (high congestion, 240×32) and group 2 (low congestion, 240×32).

Table 1. List of the 32 features with PRTG of the business network

Variable	Name of the variable (PRTG)	Device
x_1	Traffic sum (volume - kB)	Edge router (Rb)
x_2	Traffic sum (rate - kbps)	
x_3	Traffic in (volume - kB)	
x_4	Traffic in (rate - kbps)	
x_5	Traffic out (volume - kB)	
x_6	Traffic out (rate - kbps)	
x_7	Processor 2 (%)	Router/Cpu
x_8	Jitter	Jitter remote
x_9	Runtime (ms)	
x_{10}	Ping Time (ms)	Ping
x_{11}	Traffic sum (volume - kB)	Internal router (Ri)
x_{12}	Traffic sum (rate - kbps)	
x_{13}	Traffic in (volume - kB)	
x_{14}	Traffic in (rate - kbps)	
x_{15}	Traffic out (volume - kB)	
x_{16}	Traffic out (rate - kbps)	
x_{17}	Unicast in (volume - paq)	
x_{18}	Unicast in (rate - paq/s)	
x_{19}	Unicast out (volume - paq)	
x_{20}	Unicast out (rate - paq/s)	
x_{21}	State (%)	Probe/core
x_{22}	Virtual memory available (MB)	
x_{23}	Memory of paging file available (MB)	
x_{24}	Physical memory available (MB)	
x_{25}	Dedicated memory (MB)	
x_{26}	Processor load (%)	
x_{27}	Handles	
x_{28}	Threads	
x_{29}	Processor load (%)	Probe/system
x_{30}	Percentage of available memory (%)	
x_{31}	Available memory (MB)	
x_{32}	Utilization of total processing (%)	PC/Processor

Feature selector

In order to determine the redundant features, Pearson correlation was calculated for each group. The results are shown in Figure 3, evidencing high correlation among features $\{x_1, x_2\}$ and $\{x_1, x_2, x_3\}$.

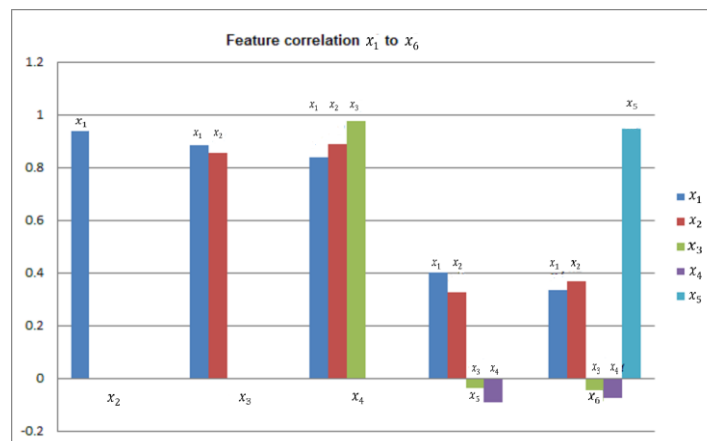


Figure 3. Redundant feature analysis

The whole feature set x from each group (low and high congestion group) was studied in order to test the linear dependence among variables, and the representation degree of each feature. The intra-class analysis allowed discard redundant features, and the number of features was reduced to 14 linearly independent features. The variables of the intraclass analysis were listed in the Table 2 where the set of variables reduced to 14 is observed, and the next row represents the dynamics of the system eliminating the correlated variables. Then, a relevant feature selection was applied to preserve most of the relevant information of the original data according to an optimality criterion that directs the representation context. For this work, the dataset was normalized (using mean and standard deviation estimations), and each group was taken as a class, where the optimality criterion was the classification accuracy of a Bayesian classifier [69][70]. According to the classification capacity, weights were assigned to each feature. In Figure 4, this feature weighting is shown, where x_2 has a representation capacity of 51.7% (the highest weight), followed by x_6 with 44.14%. Likewise, features x_{12} , x_7 , and x_{22} also represent in a lesser degree.

Table 2. Relevant variables

Relevant variables	Traffic sum (rate) Rb	Traffic out (rate)	Processor 2	Jitter	Execution time	Ping time	Traffic sum (rate) Ri	Virtual memory available	Physical memory available	Core processor load	Handles	Threads	Processor load system	Total Processing Utilization
	x_2	x_6	x_7	x_8	x_9	x_{10}	x_{12}	x_{22}	x_{24}	x_{26}	x_{27}	x_{28}	x_{29}	x_{32}
Redundant variables	x_1	x_5					x_{11}	x_{25}	x_{21}					
	x_3						x_{13}		x_{23}					
	x_4						x_{14}		x_{30}					
							x_{15}		x_{31}					
							x_{16}							
							x_{17}							
							x_{18}							
							x_{19}							
							x_{20}							

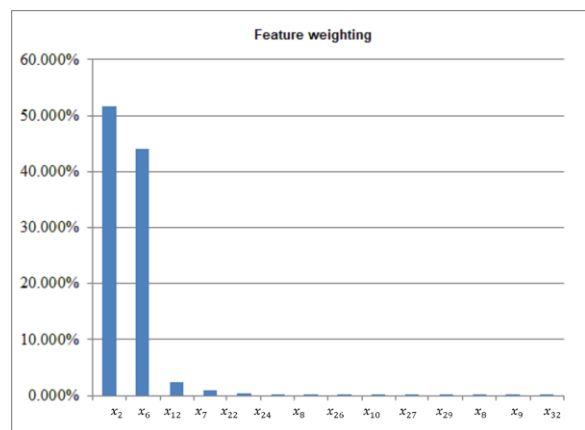


Figure 4. Feature weighting results

Training, validation and implementation

In order to determine the most representative features, the accuracy of a Bayesian classifier was used as optimizing criterion. This classifier was validated using a 30-fold cross-validation strategy (70% for training and 30% for validation). Experimental tests were performed in the laboratory using the three most relevant features. The communication channel was characterized, and with the trained classifier, the traffic conditions in terms of the instantaneous channel availability were determined. The response of DC motor was suitable.

3. RESULTS AND DISCUSSION

Using the three most relevant features, x_2 , x_6 , x_{12} a classification accuracy result of 99.9% was achieved. This indicates a high degree of separability between classes (low between 0% to 50% and high between 51% to 100% congestion or occupation of the channel), and therefore, as shown in Figure 5, the location of a 3-dimensional point in this characterization space can determine the availability of a communication channel for remote monitoring and control tasks in a real-time system.

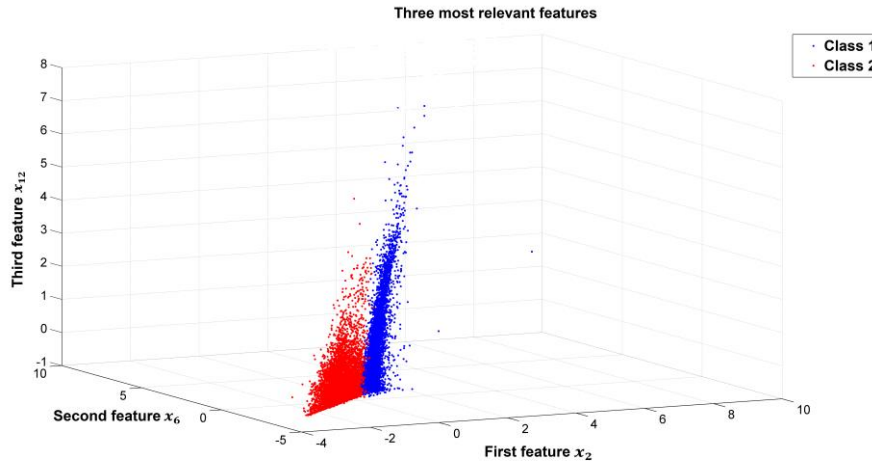


Figure 5. Decision space made up by the three most relevant features

The estimated time interval for a motor to reach 99% of an assigned velocity is $4\tau m$, where the mechanical time constant, τm , expressed in ms , is the time for the rotor to reach 63% of its velocity. Here, the effects of friction, load, and load inertia were not considered. In order to reduce operative costs even more, in the implementation, two of the three most relevant features were tested (i.e., x_2 and x_6) which achieved a classification accuracy of 99.63%.

For an Ethernet connection of 100 Mb, in low traffic conditions, the control system via web required a time interval of 14 ms. The data transfer took 9 ms. As shown in Figure 6, the data transfer with the client represents the longest time (the notation 00: 00: 00: 000000 indicates the number of hours, minutes, seconds and microseconds). Likewise, the connection time, the three-way handshake time and the server response time represent 58.7% (7 ms) of the time used for control operations, while the server transfer time is of 80 μs , and therefore, the protocol for this connection must be longer in terms of execution time.

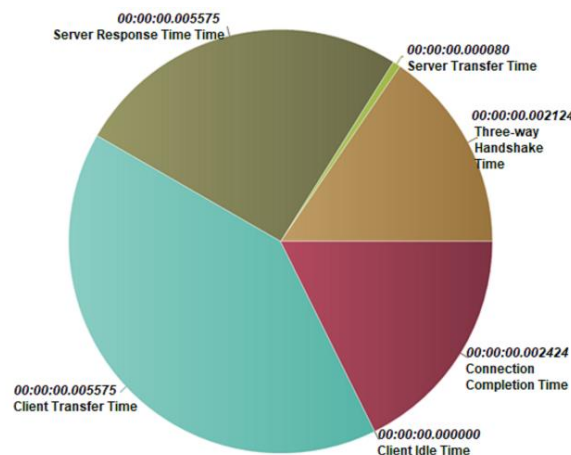


Figure 6. Time of control operations with congestion

Another test was performed with an Ethernet channel of 10 MB and P2P connection. In Figure 7, it can be seen that while data quantity was increased, the network latency proportionally incremented, which means loss of packets. Thus, this access form by the Ethernet standard provoked delay times in the channel mainly due to the number of packets and simultaneous access.

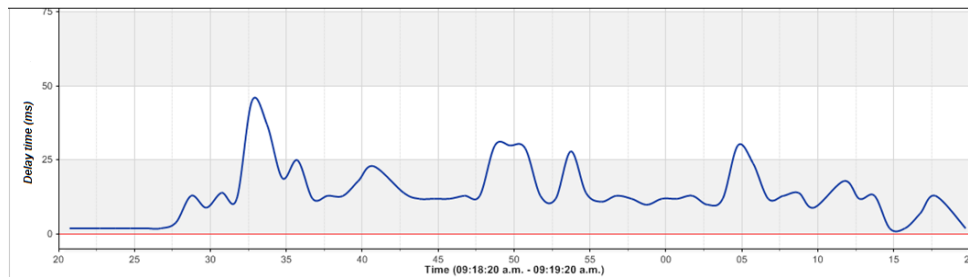


Figure 7. Network Latency

When incrementing the channel's congestion to 10%, there is a packet loss of 18% and a control time for the DC motor is obtained that goes from 14 ms (without congestion) to a delay of 272 ms. In this way, when congestion increases up to 90%, the control time increases to 707 ms, as summarized in Table 3

Experimentally, different tests (many congestion levels) were performed, for validating the 2-dimensional characterization space (using x_2 and x_6) and determining a time threshold for control operations via web on the DC motor, a time of 20 ms was the maximum time delay for a successful control process via web associated with the congestion level of the communication channel. This time threshold indirectly defined the class boundary in the characterization space for the classifier training and the validation process.

Table 3. Control response in high and low channel congestion

Congestion level (%)	Control time interval (ms)	Lost packages (%)
0	14	0
10	272	18
90	707	27

4. CONCLUSION

Although the methodology of [13] is similar to characterize channel traffic, it is not comparable to the method proposed due to the variables involved of the devices and the identification and elimination of parameters that do not represent information relevant to the system. For this, an approach for the characterization of a communication channel was discussed and implemented in a real-time system for remote monitoring and control. In this study, a feature set made up using 32 attributes extracted from a communication channel, which was analyzed using multivariate data analysis and machine learning techniques. The redundant features were discarded, and the relevant feature set was selected. The used optimization criterion was the classification accuracy of a Bayesian classifier. The results obtained using multivariate data analysis coincide with the expert intuition, as a form of validating the reduced characterization space. Thus, there can be stronger reliability for this proposed approach when the nature of features is not previously known. Additionally, in this study, a time threshold was determined after several experiments via Web on a real-time system (i.e., control and monitoring of a DC motor) with different congestion levels. This time threshold was a key factor for defining the plane between classes in the feature space. Experimental results determined the terms and conditions of use for control functions and variables that require critical response times in a particular process although this approach allows automatic decisions about the channel availability, in order to establish predictions about the length of the time window when the availability conditions are favorable, as future work it is necessary to involve alternatives such as evolutionary algorithms and forecasting routines to make those predictions

ACKNOWLEDGEMENTS

This work was supported by the Instituto Tecnológico Metropolitano ITM of Medellín (PM10255 and P17202 projects).

REFERENCES

- [1] N. Shanmathi and M. Jagannath, "Computerised Decision Support System for Remote Health Monitoring: A Systematic Review," *Irbm*, vol. 39, no. 5, pp. 359–367, 2018.
- [2] L. Lampani and P. Gaudenzi, "Innovative composite material component with embedded self-powered wireless sensor device for structural monitoring," *Compos. Struct.*, vol. 202, pp. 136–141, 2018.
- [3] M.-H. Liang, Y.-F. He, L.-J. Chen, and S.-F. Du, "Greenhouse Environment dynamic Monitoring system based on WIFI," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 736–740, 2018.
- [4] A. Bahnasse, F. E. Louhab, H. A. Oulahyane, M. Talea, and A. Bakali, "Smart bandwidth allocation for next

- generation networks adopting software-defined network approach,” *Data Br.*, vol. 20, pp. 840–845, 2018.
- [5] P. G. V. Naranjo, Z. Pooranian, M. Shojafar, M. Conti, and R. Buyya, “FOCAN: A Fog-supported smart city network architecture for management of applications in the Internet of Everything environments,” *J. Parallel Distrib. Comput.*, 2018.
- [6] A. A. Amaral, L. de Souza Mendes, B. B. Zarpelão, and M. L. P. Junior, “Deep IP flow inspection to detect beyond network anomalies,” *Comput. Commun.*, vol. 98, pp. 80–96, 2017.
- [7] F. A. Fernandes, G. S. Sestito, A. L. Dias, D. Brandão, and P. Ferrari, “Influence of network parameters on the recovery time of a ring topology PROFINET network,” *IFAC-PapersOnLine*, vol. 49, no. 30, pp. 278–283, 2016.
- [8] P. Zeng, Z. Wang, Z. Jia, L. Kong, D. Li, and X. Jin, “Time-slotted software-defined Industrial Ethernet for real-time Quality of Service in Industry 4.0,” *Futur. Gener. Comput. Syst.*, vol. 99, pp. 1–10, 2019.
- [9] N. Varis and J. Manner, “Evaluation of link layer mobility in Ethernet networks,” *Comput. Commun.*, vol. 103, pp. 193–209, 2017.
- [10] P. M. Rodriguez, A. Lizeaga, M. Mendicute, and I. Val, “Spectrum handoff strategy for cognitive radio-based MAC for real-time industrial wireless sensor and actuator networks,” *Comput. Networks*, vol. 152, pp. 186–198, 2019.
- [11] P. Singhal, R. Mathur, and H. Vyas, “State of the Art Review of Network Traffic Classification based on Machine Learning Approach,” *IJCA Proc. Int. Conf. Recent Trends Eng. Technol. 2013*, vol. ICRTET, no. 1, pp. 12–16, 2013.
- [12] Z. A. G. H. Shaikh, “An Overview of Network Traffic Classification Methods,” *Int. J. Recent Innov. Trends Comput. Commun. IJRITCC*, vol. 3, no. 2, pp. 482–488, 2015.
- [13] K. L. Dias, M. A. Pongelupe, W. M. Caminhas, and L. de Errico, “An innovative approach for real-time network traffic classification,” *Comput. Networks*, vol. 158, pp. 143–157, 2019.
- [14] A. Przybył, “Hard real-time communication solution for mechatronic systems,” *Robot. Comput. Integr. Manuf.*, vol. 49, pp. 309–316, 2018.
- [15] M. Manderscheid, G. Weiss, and R. Knorr, “Verification of network end-to-end latencies for adaptive ethernet-based cyber-physical systems,” *J. Syst. Archit.*, vol. 88, pp. 23–32, 2018.
- [16] J. Barrios-Avilés, A. Rosado-Muñoz, T. Iakymchuk, and M. García-Chulbi, “POWERLINK and Ethernet/IP Comparison as Robust Industrial Ethernet Protocols,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 363–368, 2017.
- [17] X. Li and L. George, “1 - A Survey of Switched Ethernet Solutions for Real-time Audio/Video Communications,” in *Building Wireless Sensor Networks*, S. Femmam, Ed. Elsevier, 2017, pp. 1–30.
- [18] F. Prinz, M. Schoeffler, A. Lechler, and A. Verl, “Dynamic Real-time Orchestration of I4.0 Components based on Time-Sensitive Networking,” *Procedia CIRP*, vol. 72, pp. 910–915, 2018.
- [19] *IEEE P1722/D14, August 2015: IEEE Draft Standard for a Transport Protocol for Time Sensitive Applications in a Bridged Local Area Network*. IEEE, 2015.
- [20] X. Wu and L. Xie, “Performance evaluation of industrial Ethernet protocols for networked control application,” *Control Eng. Pract.*, vol. 84, pp. 208–217, 2019.
- [21] M. Sollfrank, E. Trunzer, and B. Vogel-Heuser, “A testbed for evaluating QoS of different classes of industrial Ethernet protocols based on raspberry Pi,” *Proc. IECON 2018 - 44th Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 1, pp. 2955–2962, 2018.
- [22] R. Toulson and T. Wilmshurst, “Chapter 12 - Internet Communication and Control,” in *Fast and Effective Embedded Systems Design (Second Edition)*, Second Edi., R. Toulson and T. Wilmshurst, Eds. Newnes, 2017, pp. 291–321.
- [23] K. L. S. Sharma, “16 - Data Communication and Networking,” in *Overview of Industrial Process Automation (Second Edition)*, Second Edi., K. L. S. Sharma, Ed. Elsevier, 2017, pp. 237–261.
- [24] P. Neumann and A. Pöschmann, “Ethernet-based real-time communications with PROFINET IO,” *WSEAS Trans. Commun.*, vol. 4, no. 5, pp. 235–245, 2005.
- [25] B. R. Xia Jiqiang Cao Quan, “Modeling and quantitative real-time analysis of Ethernet based on CSMA/CD,” in *IEEE Conference Anthology*, 2013, pp. 1–5.
- [26] P. Axer, D. Thiele, and R. Ernst, “Formal timing analysis of automatic repeat request for switched real-time networks,” in *Proceedings of the 9th IEEE International Symposium on Industrial Embedded Systems (SIES 2014)*, 2014, pp. 78–87.
- [27] S. Kangovi, “3 - The Ethernet Landscape,” in *Peering Carrier Ethernet Networks*, S. Kangovi, Ed. Morgan Kaufmann, 2017, pp. 55–98.
- [28] J. J. Lee, S. M. Chung, B. Lee, K. T. Kim, and H. Y. Youn, “Round Trip Time Based Adaptive Congestion Control with CoAP for Sensor Network,” in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, May 2016, pp. 113–115.
- [29] N. H. Shahapure and P. Jayarekha, “Distance and Traffic Based Virtual Machine Migration for Scalability in Cloud Computing,” *Procedia Comput. Sci.*, vol. 132, pp. 728–737, 2018.
- [30] B. Yi, X. Wang, K. Li, S. k. Das, and M. Huang, “A comprehensive survey of Network Function Virtualization,” *Comput. Networks*, vol. 133, pp. 212–262, 2018.
- [31] A. A. Yaseen and M. Bayart, “A Model-Free Approach to Networked Control System with Time-Varying Communication Delay,” *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 558–563, 2018.
- [32] A. J. C. Trappey, C. V Trappey, U. H. Govindarajan, A. C. Chuang, and J. J. Sun, “A review of essential standards and patent landscapes for the Internet of Things: A key enabler for Industry 4.0,” *Adv. Eng. Informatics*, vol. 33, pp. 208–229, 2017.
- [33] X. Wu and L. Xie, “Performance evaluation of industrial Ethernet protocols for networked control application,”

- Control Eng. Pract.*, vol. 84, no. November 2018, pp. 208–217, 2019.
- [34] F. Civerchia, S. Bocchino, C. Salvadori, E. Rossi, L. Maggiani, and M. Petracca, “Industrial Internet of Things monitoring solution for advanced predictive maintenance applications,” *J. Ind. Inf. Integr.*, vol. 7, pp. 4–12, 2017.
- [35] I. F. Akyildiz, A. Kak, E. Khorov, A. Krasilov, and A. Kureev, “ARBAT: A flexible network architecture for QoE-aware communications in 5G systems,” *Comput. Networks*, vol. 147, pp. 262–279, 2018.
- [36] M. Habiba, M. R. Islam, and A. B. M. S. Ali, “A component based unified architecture for utility service in cloud,” *Futur. Gener. Comput. Syst.*, vol. 87, pp. 725–742, 2018.
- [37] Z. Li, Y. Gao, P. Li, L. Sang, and D. Yang, “Packet delay analysis in wireless multi-channel networks: A network calculus perspective,” in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, Oct. 2015, pp. 1–5.
- [38] S. Ren *et al.*, “A Delay Estimation in HPC Wormhole Switching Based on Network Calculus,” in *2012 Fourth International Conference on Computational and Information Sciences*, 2012, pp. 955–957.
- [39] Q. Liu, X. Jiang, and Y. Zhou, “Per-flow end-to-end delay bounds in heterogeneous wireless networks,” in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 295–299.
- [40] Z. Fernandez, C. Cruces, I. Val, and M. Mendicute, “Deterministic real-time access point concepts for industrial hybrid Ethernet/IEEE 802.11 networks,” in *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*, May 2017, pp. 1–6.
- [41] M. Thibaud, H. Chi, W. Zhou, and S. Piramuthu, “Internet of Things (IoT) in high-risk Environment, Health and Safety (EHS) industries: A comprehensive review,” *Decis. Support Syst.*, vol. 108, pp. 79–95, 2018.
- [42] X. Ge, F. Yang, and Q.-L. Han, “Distributed networked control systems: A brief overview,” *Inf. Sci. (Ny)*, vol. 380, pp. 117–131, 2017.
- [43] T. Posewsky and D. Ziener, “Throughput optimizations for FPGA-based deep neural network inference,” *Microprocess. Microsyst.*, vol. 60, pp. 151–161, 2018.
- [44] W. Goralski, “Chapter 1 - Protocols and Layers,” in *The Illustrated Network (Second Edition)*, Second Edi., W. Goralski, Ed. Boston: Morgan Kaufmann, 2017, pp. 3–46.
- [45] S. Noye, R. North, and D. Fisk, “A wireless sensor network prototype for post-occupancy troubleshooting of building systems,” *Autom. Constr.*, vol. 89, pp. 225–234, 2018.
- [46] L. Zhou, C. Su, Z. Li, Z. Liu, and G. P. Hancke, “Automatic fine-grained access control in SCADA by machine learning,” *Futur. Gener. Comput. Syst.*, vol. 93, pp. 548–559, 2019.
- [47] S. P. Mertikas, P. Partsinevelos, C. Mavrocordatos, and N. A. Maximenko, “Environmental applications of remote sensing,” in *Pollution Assessment for Sustainable Practices in Applied Sciences and Engineering*, Elsevier, 2021, pp. 107–163.
- [48] J. Rosenberg, “Chapter e6 - Embedded security,” in *Rugged Embedded Systems*, A. Vega, P. Bose, and A. Buyuktosunoglu, Eds. Boston: Morgan Kaufmann, 2017, pp. e1–e74.
- [49] P. K. Malik *et al.*, “Industrial Internet of Things and its applications in industry 4.0: State of the art,” *Comput. Commun.*, vol. 166, pp. 125–139, Nov. 2020, doi: 10.1016/j.comcom.2020.11.016.
- [50] D. F. Merchán, J. A. Peralta, A. Vazquez-Rodas, L. I. Minchala, and D. Astudillo-Salinas, “Open Source SCADA System for Advanced Monitoring of Industrial Processes,” in *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, Nov. 2017, pp. 160–165.
- [51] T. Oksanen, R. Linkolehto, and I. Seilonen, “Adapting an industrial automation protocol to remote monitoring of mobile agricultural machinery: a combine harvester with IoT,” *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 127–131, 2016.
- [52] X. Fan, K. Fan, Y. Wang, and R. Zhou, “Overview of cyber-security of industrial control system,” in *2015 International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC)*, 2015, pp. 1–7.
- [53] N. Based and T. Classification, “A Survey of Network Based Traffic Classification Methods,” *Database Syst. J.*, vol. 7, no. 4, pp. 24–31, 2017.
- [54] A. Al-Shdifat and C. Emmanouilidis, “Development of a Context-aware framework for the Integration of Internet of Things and Cloud Computing for Remote Monitoring Services,” *Procedia Manuf.*, vol. 16, pp. 31–38, 2018.
- [55] G. Sinnapolu and S. Alawneh, “Integrating wearables with cloud-based communication for health monitoring and emergency assistance,” *Internet of Things*, vol. 1–2, pp. 40–54, 2018.
- [56] J.-F. Wang, X. He, S.-Z. Si, H. Zhao, C. Zheng, and H. Yu, “Using complex network theory for temporal locality in network traffic flows,” *Phys. A Stat. Mech. its Appl.*, vol. 524, pp. 722–736, 2019.
- [57] R. Khan and S. U. Khan, “Design and implementation of an automated network monitoring and reporting back system,” *J. Ind. Inf. Integr.*, vol. 9, pp. 24–34, 2018.
- [58] W. Cerroni, G. Moro, R. Pasolini, and M. Ramilli, “Decentralized detection of network attacks through P2P data clustering of SNMP data,” *Comput. Secur.*, vol. 52, pp. 1–16, 2015.
- [59] D. Pena, *Análisis de datos multivariantes*. {McGraw-Hill} Interamericana de Espana {S.L.}, 2002.
- [60] G. Santiago Briones and C. I. de E. a D. Chile, *Metodología de la investigación cualitativa*. Centro Iberoamericano de Educacion a Distancia, 2001.
- [61] Z. Qiu, J. Jin, H.-K. Lam, Y. Zhang, X. Wang, and A. Cichocki, “Improved SFFS method for channel selection in motor imagery based BCI,” *Neurocomputing*, vol. 207, pp. 519–527, 2016.
- [62] R. M. Golden, “Statistical Pattern Recognition,” in *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, John Wiley & Sons, 2015, pp. 411–417.
- [63] *Pattern Recognition*. Elsevier, 2009.

- [64] R. G. Martins, A. S. Martins, L. A. Neves, L. V. Lima, E. L. Flores, and M. Z. do Nascimento, "Exploring polynomial classifier to predict match results in football championships," *Expert Syst. Appl.*, vol. 83, pp. 79–93, Oct. 2017, doi: 10.1016/j.eswa.2017.04.040.
- [65] C. Chitu and H. Song, "Data analytics and processing platforms in CPS," in *Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things*, Elsevier, 2019, pp. 1–24.
- [66] F. Van der Heijden, R. P. W. Duin, D. De Ridder, and D. M. J. Tax, "Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB," *Classif. Param. Estim. State Estim. An Eng. Approach Using MATLAB*, pp. 1–423, Jun. 2005, doi: 10.1002/0470090154.
- [67] V. Narasimha Swamy *et al.*, "Real-time cooperative communication for automation over wireless," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 11, pp. 7168–7183, 2017.
- [68] F. Ullah and C.-M. Pun, "The role of Internet of Things for adaptive traffic prioritization in wireless body area networks," in *Healthcare Paradigms in the Internet of Things Ecosystem*, Elsevier, 2021, pp. 63–82.
- [69] H. Bouhamed, A. Masmoudi, and A. Rebai, "Bayesian Classifier Structure-learning Using Several General Algorithms," *Procedia Comput. Sci.*, vol. 46, pp. 476–482, 2015.
- [70] L. Jiang, C. Li, and S. Wang, "Cost-sensitive Bayesian network classifiers," *Pattern Recognit. Lett.*, vol. 45, pp. 211–216, 2014.

BIOGRAPHY OF AUTHORS



Leonardo Serna-Guarin professor assigned to the Faculty of Engineering of the Metropolitan Technological Institute - ITM, in the city of Medellín. He develops his scientific activities in the Automatic, Electronic and Computer Science research group of the ITM. He received a degree in Electronics Technologist, Systems Engineer, Data Network Specialist and Master in Automation and Industrial Control from ITM, in 2001, 2004, 2008 and 2014 respectively. Cisco Acedemia Instructor and certified in Data Cabling System and Fiber Optics. His academic and research interests are oriented to issues related to teletraffic engineering, performance of industrial networks, communications infrastructure, network services, communications networks and security in communications systems.



Luis J. Morantes-Guzman received a B.S. degree in Electronic Engineering and a M.Eng. in Industrial Automation from the Universidad Nacional de Colombia, in 2005, and 2009, respectively. He is currently working toward the Ph.D. degree with the Universidad de Antioquia, Colombia. Since July 2009, he has been a Full-time Lecturer at the Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia. His current research interests include computer vision, pattern recognition, and machine learning.



Edilson Delgado-Trejos received a B.S. degree in Electronic Engineering, a M.S. degree in Industrial Automation, and a Ph.D. degree in Engineering Sciences from the Universidad Nacional de Colombia, in 2000, 2003, and 2008, respectively. Since August 2008, he has been a Full-time Lecturer and Senior Researcher at the Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia. Currently, he is an Associate Professor and Vice-Chancellor of Research and Academic Extension at the same institute. He has published more than 50 papers, 10 book chapters and 3 books in indexed scientific journals and editorials. His current research interests include pattern recognition, machine learning, multivariate data analysis, nonlinear analysis, signal processing and soft metrology.



Miguel A. Becerra received a B.S. degree in Electronic Engineering from the Universidad Nacional de Colombia, a Master in Virtual pedagogy, an M.S. degree in Industrial Automation from ITM 2012. He is currently working toward the Ph.D. degree. Nowadays, he is a Researcher at the Institución Universitaria Pascual Bravo, Medellín, Colombia. He has published more than 60 publications among papers and chapters in indexed scientific journals and editorials. His current research interests include data fusion, information quality, machine learning, and signal processing.