Acoustic Sensor Network for Surveillance and Target Acquisition of Terrestrial Military Operations

Gonçalo Atanásio
Academia Militar
IST / UL
Lisboa, Portugal
Email: goncalo.atanasio@ist.utl.pt

António Serralheiro
CINAMIL, Academia Militar
INESC-ID
Lisboa, Portugal
Email: antonio.serralheiro@academiamilitar.pt

António Grilo
IST / UL
INESC-ID
Lisboa, Portugal
Email: antonio.grilo@inesc-id.pt

Abstract—**The main objective of this work is to present a system that detects and identifies combat vehicles with a wireless sensor network that contains acoustic sensors. The objective is to gather the sound from the combat vehicles, with acoustic sensors in the nodes of the wireless sensor network, perform the classification of the vehicle in the node and then send the packet to the base station using packet routing and data aggregation. The data base of vehicle sounds, drawn from combat vehicles of the Armed Forces, was made by training Gaussian mixture models (GMM) with data that explores the spectral characteristics of the sound, Mel Frequency Cepstral Coefficients (MFCC) and Beats per minute (BPM) were used as observation data. Vehicle detection and identification is performed with a GMM classifier. The NS3 simulator was used to simulate the wireless sensor network and the respective data aggregation and routing. The routing protocol used is DSDV, and it was tested over two different technologies, IEEE 802.15.4 and IEEE 802.11g. For the tests of the classifier, the number of Gaussian mixtures was optimized and the tests were made with different time windows for the MFCC. It was concluded that the classifier is better without the rhythm detection, that is, using only the MFCC it can obtain an F-score of 0.875 with eight combat vehicles on the data base. The simulations done using the NS3 simulator indicate that there must be a trade-off between the delays, the radio range and power consumption. The 802.15.4 technology, with aggregation, would be the most appropriate choice to the case study. However if it is desired to cover big areas, because of tactical restrictions, the most suitable technology is 802.11g.**

1. Introduction

With the evolution of the technology, wireless sensor networks have expanded their potential application domains. In many applications, a very high data traffic must be provided despite the limited hardware resources that leads to an unsustainable power consumption. Thus, low power consumption routing algorithms are needed as well as data aggregation, for the purpose of increasing the nodes lifetime. These nodes allow the acquisition, processing and sending/receiving data.

When inserted in a military context, wireless sensor networks requires certain concerns such as resistance to changes in the network, redundancy, scalability, low power consumption and low latency for the identification and classification of vehicles as close to real time as possible.

Since the technology is available to both conflict sides, the advantage will go to those who best understand its scope and limitations. Today, the ability of detecting and classifying vehicles in an operational context have great tactic importance, since, increasingly, there is a technological evolution from the adversary and therefore, a higher speed of information flow is required for decision-making so, this technology allows you to optimize the command and control. Thus, with the proper application of this technology it is possible to increase the protection of military forces on the battlefield. Even with the existence of acoustic sensors in the Portuguese Army [1], the development of this technology becomes important because the current systems do not allow detection and identification of military vehicles in an autonomous way which is a current limitation.

In this particular case it will be applied a wireless sensor network to a military environment, whose purpose is the detection and classification of military vehicles in operational context. These sensor networks usually consist in a set of stationary nodes distributed in a particular area. The implementation objectives are: Vehicle Sound Collection; Vehicle classification by processing the data in the node; Routing the result of classification to the base station; Using data aggregation in intermediate nodes in order to reduce network energy consumption.

The system to be developed is summarily shown in Figure 1. In this figure is represented a vehicle approaching the wireless sensor network, which itself emits a certain sound that is collected by the closer nodes of the network. When a node collects a sound fragment through a microphone it performs its pre-processing by extracting spectral features. Each node has a database containing GMM models representing various vehicles, which is used to perform classification of the collected sound fragment.

After the classification the node sends a packet with the result to the base station. Since the nodes are distant from the base station and do not have enough power to communicate directly with this, it is necessary to use other nodes to forward these packets. The nodes, in addition to the routing task, will also perform data aggregation (to eliminate redundant data), limiting the total packets sent which leads to less energy consumption.

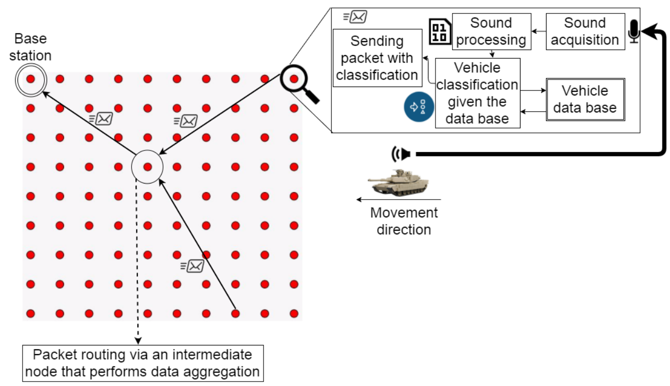


Figure 1. System description.

1.1 Document Structure

In this document it is made a literature review required for this work, which is contained in Section 2. The detection and classification of military vehicles starts with the characterization of the vehicle given its acoustic signature, created with the spectral characteristic and the BPM of the sound.

In this particular application, the nodes begin to acquire sound and subsequently process them using the Mel-Frequency Cepstral Coefficients (MFCC) and the beats per minute (BPM) (Section 2.1.2).

The MFCC and BPM are represented in a set of probability density functions, called Gaussian Mixture Model (GMM) [2]. It is intended to create a database in which are included signatures of certain types of vehicles (represented in GMM) (Section 2.1.4). With the data base it is possible to make vehicle sound classification (Section 2.1.3).

In Section 2.2 are addressed aspects related to the sensor network itself. In this section it will be studied the following topics: node architecture (Section 2.2.1); some existing network architectures (Section 2.2.2); data aggregation and its mechanisms (Section 2.2.3); communication technologies in wireless sensor networks (Section 2.2.4); and also the network layer protocols (Section 2.2.5).

In Section 3 is described the developed implementation in order to make the vehicle identification and classification.

In Section 4 are presented the results of the tests performed on synthetic and real data associated with the GMM training and classification (Section 4.1) and the results of the wireless sensor network simulations with the respective aggregation and packet routing (Section 4.2).

Finally, in Section 5, the conclusions related to this work are drawn as well as indications of future work that can be done.

2. Background

2.1 Signal Processing

2.1.1 Sound components

As by [3], the sound is compost of nine components. For this application it will be studied more than one sound component to optimize the classification.

It is intended to perform a sound analysis across the range of frequencies - timbre - and an analysis of one rhythm component. To achieve this goal, MFCC and BPM are performed.

2.1.2 Parameter extraction

In this section it will be addressed the methods used for the extraction of sound parameters, MFCC and BPM, which respectively study the timbre and rhythm. These methods will produce data that later will be used for GMM training to generate the models associated with each vehicle.

The MFCC are typically used in speech recognition applications and takes advantage of the human auditory system capacity to create a vector containing information about linguistic message, while eliminating certain harmonics that are not relevant [4].

Although it is mainly used in speech recognition, the MFCC can be adapted to other types of applications. Their use in a military component proves to be advantageous, especially in this case where it is necessary to perform frequency analysis and then make a classification from a frequency signature.

In order to perform the beat sound detection it is necessary to satisfy two conditions [5]: the selected instants should correspond to the moments when an audio beat is indicated, for example, by the appearance of a note generated by one of the instruments; an interval between beats should be locally constant, since the regular spacing between beats is what defines the musical rhythm.

The beat detection system which first estimates a global tempo, uses this tempo to construct a transition cost function and finally uses dynamic programming to find the best-scoring set of beat times that reflect the tempo as well as corresponding to moments of high ”onset strength”. [5] also developed a Matlab code to apply this system that will be useful to calculate the BPM for our application.

2.1.3 Sound classification

In this section we present the GMMs that are trained using MFCC and BPM. This training is described in Section 2.1.4.

The GMM are a pattern recognition system used to model and represent the probability density function of the observed samples with mixtures of Gaussian densities. Given a data vector, the GMM defines the weight of each distribution by algorithms such as Expectation-Maximization algorithm (EM) [2].

The probability density function of one Gaussian associated with various dimensions is [6]:

$$N(x|\mu_x, \Sigma_x) = \\ = p(x|\mu_x, \Sigma_x) \\ = \frac{1}{\sqrt{2\pi|\Sigma_x|}} e^{(-\frac{1}{2}(x-\mu_x)\sum_x^{-1}(x-\mu_x)^T)} \quad (1)$$

where μ_x is a vector of means, Σ_x a covariance matrix and x the samples vector which in this case is associated with the MFCC and BPM.

For multiple Gaussian mixtures is defined as follows [6]:

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_x, \Sigma_x) \quad (2)$$

where π_k is the mixture coefficient, that is, the variable that assigns the weight of each Gaussian mixture, under the restriction [6]:

$$\sum_{k=1}^{K} \pi_k = 1 \quad (3)$$

Given a GMM model that represents the data it is possible to classify a piece of sound through the use of Equation 4. The log-likelihood, associated with one Gaussian distribution, given a set of data is defined as $-ln\, p(X|\mu, \Sigma, \pi)$ [6]. To perform a classification of a sound passage for all existing vehicle models, the log-likelihood is computed as:

$$ln\, p(X|\mu, \Sigma, \pi) = \\ = -\sum_{n=1}^{n} ln \left\{ \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k) \right\} \quad (4)$$

The model that generates the minimum log-likelihood is the model that the classifier will return as the correct one.

This type of classification becomes very advantageous to this study because the only element required to transmit on the network is the resulting vehicle number of the classification performed by the node. Subsequently, the base station evaluates the various classifications made by the nodes. As a result, this alternative allows to send fewer bits and thus spend less energy in transmission.

2.1.4 Collected data training

As mentioned above the GMM models generated are used to perform the classification.

The generation of GMM models is a procedure denominated: *training*. Each GMM is defined by the means, covariances and mixing coefficients of each Gaussian mixture. These parameters can be obtained using the EM algorithm, defined as an iterative method that calculates the Maximum-Likelihood Estimation (MLE) [7].

The MLE is a method used to estimate the parameters of a statistical model, and seeks to find the mean values, covariances and mixture coefficients maximizing the probability of the sampled data [7]. In this case the used probability density function is Gaussian.

After completing the EM algorithm, we obtain the means vector, the covariances matrix and the mixture coefficients vector associated to one GMM given each set of MFCC and BPM. Thus, it is generated a model for each vehicle capable of being stored in a database.

2.1.5 Performance measures

It is good design practice to check the classifier proper function through performance measures after training. The performance of a system is normally tested using precision indicators like *precision* (p), *recall* (r) and *F-score* (F_β) [8].

This measures are defined as follows [8]:

$$p = \frac{TP}{TP+FP} \quad (5)$$

$$r = \frac{TP}{TP+FN} \quad (6)$$

$$F_\beta = (1+\beta^2)\frac{pr}{r+\beta^2 p} = \frac{(1+\beta^2)TP}{(1+\beta^2)TP+\beta^2 FN+FP} \quad (7)$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

In the Equation 7, the value of β defines the weight given to the precision or to the recall. That is, $\beta < 1$ places more weight on precision and $\beta > 1$ places more weight on recall. If $\beta = 0$ is only considered the precision, if $\beta = \infty$ is only considered the recall.

2.2 Wireless sensor network

Typically, wireless sensor networks consist of nodes that contain sensors with the function of collecting data from the physical environment. Then, these data are processed and forwarded to other nodes or to a base station, for example via radio communication, depending on the scenario in question.

2.2.1 Node architecture

General purpose nodes should be versatile so that they can adapt to a variety of situations [9]. In the application under study the node must contain at least: GPS - the use of the GPS receiver allows the identification of the node's position, which can be used in a motion tracking algorithm; Microphone - allows the collection of sound; Sender/Receiver - to send and/or receive packets; Processor - used in this study, to perform data processing at the nodes, before sending data to another; Random Access Memory (RAM) - fast-access memory that helps the processing; Storage memory - designed to save the vehicles databases and other relevant data; Battery - used to store energy and supply the components of the node; Power collection module (optional) - used to collect energy from the environment, to increase the autonomy of nodes and as a consequence the network in its entirety.

2.2.2 Wireless sensor network architecture

The sensor networks architectures can be grouped into three major groups [10]: Flat; Hierarchical and Location based.

In flat networks all nodes have the same role, while in hierarchical networks the nodes are distributed in clusters or levels so that the "father" of each cluster can make some aggregation leading to less transmission of data in order to save energy. Location-based protocols use the node's position to request the wanted data in certain regions [10].

For the application in study it is intended that all the nodes have the ability to perform data aggregation so, all of them will be equal but this does not mean that they can't have different roles. Thus, the network can be hierarchical.

It is intended that some of the nodes wait for packets that others send after collecting sound, aggregate and then forward a packet with data from multiple nodes instead of route all the packets they receive. Thus, if a node is included in the route to the base station it can receive data from multiple nodes. The nodes that are on the route are automatically the "father" nodes of the clusters, which are designed to send the packet to the base station. If the routes change, the "father" node of each cluster automatically changes.

With this methodology there is only one base station, which is different from all the other nodes, giving the network a hierarchical structure. The "father" nodes are elected because they belong to the route to the base station. These nodes will receive packets from other nodes, aggregate and send the result to the base station. Thus, these nodes will be equal to all other nodes, except the base station, but they will have different roles, which also gives the network a hierarchical structure.

2.2.3 Data aggregation

Since these networks have constraints related to energy sources and bandwidth, it may be necessary to perform processing in intermediate nodes and only a few transmit data to the base station. Once the goal is to minimize data transmissions, one node collects data from multiple ones and performs some processing eliminating some redundancies. Subsequently, the data is sent to the base station or to another node. This methodology is known as data aggregation [11].

With no aggregation a node just forward all the received packets. If ten packets are received in a node, it has to forward ten separate packets, making ten transmissions, spending a certain energy. If these data are all equal, the node may contain a time of aggregation at which it receive, for example, these same ten packets to which redundancy are eliminated in order to send only one packet. This way, energy is only spent in a transmission. This can be done easily in this application because all the nodes' destination is the base station. In other cases it would be necessary to check the destination of the various packets and perform aggregation with this condition.

The use of data aggregation is useful in this project even if it requires additional processing in intermediate nodes. This happens because the energy spent in the additional processing is much smaller than the energy spent in a high number of transmissions. Thus, since the collected sound samples require some processing to signature identification, the nodes that collect these samples can perform processing and transmit only the data referred to in Section 2.1.3.

In many sensor networks exists a high number of nodes and in the case studied it is also aimed to use a lot of nodes, what makes data aggregation necessary [11].

2.2.4 Communication Technologies in Wireless Sensor Networks

Nowadays exist several communication technologies and a comparison between Bluetooth using IEEE 802.15.1, Ultra-wideband using IEEE 802.15.3, ZigBee using IEEE 802.15.4 and Wi-Fi using IEEE 802.11 are made by [12].

The Ultra-wideband is intended for indoor networks, typically distances of 10m between nodes. This is not recommended because it is intended to cover the battlefield area with a small number of nodes [12].

The Bluetooth is also designed for very short range communications, also 10m, typically for networks of personal devices. Bluetooth was developed to replace the cables of peripheral devices such as computers keyboards or mouses [12]. This technology only allows the creation of clusters up to eight nodes, which is a limitation when it is intended to develop multi-hop networks with a high number of nodes covering large areas.

Technologies IEEE 802.15.4 and IEEE 802.11 are the most suited to the intended implementation.

The protocol IEEE 802.15.4 is designed for Low-Rate Wireless Personal Area Networks (LR-WPAN) and is typically used for networks where nodes can be distanced up to 100 meters. With this protocol the sensor networks support nodes with more complex functions and simpler ones that just collect small amounts of data [12]. In this particular case it is intended that all nodes have the same characteristics, whereby this advantage will not be exploited. This technology is also associated with very low energy consumption, which means that the network has a better autonomy [12]. The protocol IEEE 802.15.4 has the following advantages over the protocol IEEE 802.11g [12]: Higher number of channels (but smaller width per channel); Higher number of nodes per basic cell.

The protocol IEEE 802.11 was designed for wireless local area networks, but not necessarily with low transmission rate. It allows very mobile users to navigate on the Internet at high speeds [12]. It also allows great distances between nodes for lower frequency bands but it has a high energy consumption. In this work this technology can be applied because it is possible for the sensors to be relatively expensive, given the nature of the military application, and so they can have a battery with more capacity. It is also desired, in this application, to use a data aggregation mechanism that can reduce the high energy consumption of the technology.

There are several versions of the technology IEEE 802.11 but the latest as ”g,” ”n” and ”c” allows greater compatibility with today’s devices and have greater range and robustness in communications. The chosen version was ”g” because it is compatible with the version ”n”. The version ”c”, operating at 5GHz, it is not compatible with the other two. The protocol IEEE 802.11g has the following advantages over the protocol IEEE 802.15.4 [12]: Higher data transfer rate; Extended reach and so a wider area can be covered; Greater amount of packet data.

A comparison between the performances of the two protocols in the sensor network will be made in Section 4.2.

2.2.5 Network layer protocols

IPv4 and IPv6 are two versions of the IP protocol that are included in the network layer of the TCP/IP model. The [13] states that over the years IP-based sensor networks have evolved in particular with the introduction of IPv6 adaptation layer (6LoWPAN) for the protocol IEEE 802.15.4. Therefore, some mechanisms needed to be adapted from IPv4 to IPv6, which led to many incompatible devices. Through the results that these authors achieved, it was found that the IPv6-based architecture meets the requirements of a sensor network with low power consumption.

Increasingly there is a trend towards the ”Internet of Things”, which leads to the existence of large numbers of sensor nodes that need to communicate between them. Currently, IPv6 is more suitable to this kind of systems than the IPv4 and therefore should be a priority when developing new applications for wireless sensor networks.

In this work it was used the IPv4 combined with the technology IEEE 802.11g and IPv6 combined with the technology IEEE 802.15.4. Finally, it was made a comparison between these in Section 4.2.

The packet routing protocols also belong to the network layer. As mentioned above, the sensor networks are composed of a large number of nodes. These, after receiving and/or processing the data of interest, need to send the data to another node or to the base station. In many cases, the nodes that are the destination are not within the range of the sending node, which makes it necessary to forward the data via other nodes. These are not reachable since the power required to cover certain distances would be too much compared to the resources that the node has. The nodes have limited energy resources and limited communication capacity. Thus, routing protocols need to take into account the energy available in the network, in order to increase the lifetime of the nodes [10].

Some routing protocols were studied and their characteristics are shown in Table 1.

After analyzing the various routing protocols and their characteristics, different options can be discussed. Using a modified protocol is considered a simple and effective approach in order to adapt it to the application under consideration.

TABLE 1. ROUTING PROTOCOLS CHARACTERISTICS.

	DD	LEACH	DSDV	RPL	GEAR
Clustering		X			
Location use					X
Data requests	X				X
Multipath	X			X	
Data aggregation	X	X		X	
Shortest path			X	X	

The DSDV protocol can obtain a high efficiency, since it is desired that the network nodes are static and there is only one base station. We just need to insert the route to the base station on all tables, and the routing data updates will not flood the network. The nodes are static so, the routing updates can have a very long period, or in the extreme case, the routing tables are filled when the network starts and after that, there will not be any further updates.

The DSDV protocol itself does not include data aggregation and elimination of redundancy, but it is intended and possible to implement these features at the application level. The nodes send the data to the base station and if the node is not in the range of the base station, the node that will forward the packet waits for more packets, and if it receives more than one, it performs aggregation to eliminate any redundancies and sends the packet again to the base station.

The RPL protocol is, among the studied ones, the most developed and most widely used. This protocol allows you to choose different metrics and so the routing can be more efficient.

Given the application under consideration, initially there is a DODAG with ”father” nodes well-defined, for example, for the metric ”shortest path to the base station”. Thus, when the nodes detect vehicles, it is possible for them to capture, process and send the data to the ”parent” node, which forwards the data through the shortest path to the base station. An adaptation that can be made to this algorithm is: if there are two routes with the same number of hops, it is chosen the one that has a ”father” node also receiving data, so that aggregation will be made more efficient.

To all the protocols referred to in this section, the NS-3 simulator [14] has only the DSDV protocol, but the RPL protocol is in current development and in future applications RPL may be an option.

In conclusion, the protocol used is the DSDV in two different versions, IPv4 and IPv6, with the amendments referred to in this section. The implementation and results associated with simulations within this protocol are described in Sections 3.2 and 4.2 respectively.

3. Implementation of the System

3.1 Vehicle Characterization and classification

In this Section it will be described the creation of the data base and the development of vehicle classification.

3.1.1 Audio signals acquisition

In order to develop the code for characterization and classification of military vehicles using their spectral signature it was acquired sound from various military vehicles (Army and Air Force), using the digital recorder Zoom H6 [15]: Panhard M11 (Figure 2(a)), Mitsubishi 4x4 L200 (Figure 2(b)), Leopard-2A6 (Figure 2(c)), M60A3 (Figure 2(d)), M113 (Figure 2(e)), Condor (Figure 2(f)), Protect Fire (Figure 2(g)) and Unimog 1300 L (Figure 2(h)).

(a) Panhard M11. (b) L200. (c) Leopard-2A6. (d) M60A3. (e) M113. (f) Condor. (g) Protect Fire. (h) Unimog 1300 L.

Figure 2. Vehicles used for sound acquisition.

Samples of each vehicle were acquired on standard roads and in dirt environment at different speeds.

3.1.2 Audio files processing

The recordings made with the Zoom H6 comprised two channels, sampled at 48000Hz. To facilitate sound analysis in Matlab, the recordings were converted to one channel, downsampled to 16000Hz. Additionally, the gain of the sounds of all the vehicles was normalized.

Some sound fragments were removed from the recordings which related to: planes, animal sounds, shooting, human speech and civilian vehicles. These were stored separately so that, together with the sounds of other vehicles, they can form a model of *garbage*. Thus, if the sound sample is any of these garbage sounds, it will not be wrongly classified as one of the vehicles that are in the database. The amount of sound files and the total duration of the sound recorded for the garbage and for each vehicle without the surrounding noise, are shown in Table 2.

3.1.3 Choosing the number of Gaussian mixtures

The number of Gaussian mixtures intended for each model must be properly found. Since the number of samples is limited and not infinite, it is not possible to find the "true" model, thus, methods for selection of the best model are used, given a set of samples [16].

TABLE 2. AMOUNT OF SOUND FILES AND TOTAL DURATION OF THE SOUND RECORDED FOR THE GARBAGE AND FOR EACH VEHICLE.

	Number of recordings	Total duration of the sound recorded (s)
PanhardM11	11	458,4
Mitsubishi 4x4 L200	2	55,1
Leopard-2A6	5	600,4
M60A3	2	88,1
M113	9	423,0
Condor	4	202,4
Protect Fire	5	166,6
Unimog 1300 L	4	139,2
Garabage	15	478,5

There were used two criteria, Bayesian information criterion (BIC) and Akaike information criterion (AIC). The BIC is consistent, in the sense that if the "real" model exists between the candidate, the probability of selecting the "true" model approaches 100%. Moreover AIC can have similar performance for parametric cases (based on parameterized family of probability density functions) and non-parametric. In general, the AIC is not consistent, while BIC can be in the parametric case [16].

According to [16], these criteria must be used to find the best model in which the BIC is to be preferred since, as mentioned above, is more consistent and also due to the fact that in the application under consideration it is being used a Gaussian distribution, which is a parametric case.

The AIC can complement the BIC since, in both cases, the smaller the value, the more suitable is the model, and if there are several BIC minimum, the AIC can be used as a tiebreaker.

3.1.4 Vehicle classification

In GMM training the used sample vectors include only the first 85% and the last 15% are used for testing.

Initially the MFCC are evaluated in one-second frames. Then, for each GMM model in the database (vehicles in the data base) the Equation 4 is applied to calculate the log-likelihood for each second of the testing sound, and finally it is made an average of all log-likelihood values respecting all the seconds.

Having an average log-likelihood value associated with each vehicle given the testing sound, it is possible to find the minimum one and classify the sound as being the vehicle associated with this minimum.

3.1.5 Calculation and integration of BPM in the classification

It is not intended that the weight of the BPM in the classification is the same as that of MFCC, whereby the audio windows analyzed are four times higher than the windows used by the MFCC. That is, for four MFCC values, the BPM value is repeated four times and while there are eight cepstral coefficients in the MFCC, BPM will be only one value for each window.

To insert the BPM in the classification, the BPM will be included in the MFCC already generated and the classification is made as mentioned in the previous section. Thus, the BPM are concatenated in the MFCC, adding a column in the matrix and since there are used eight cepstral coefficients, the matrix containing eight columns is now a matrix of nine columns. This new column corresponds to the calculated BPM.

3.2 Wireless sensor network

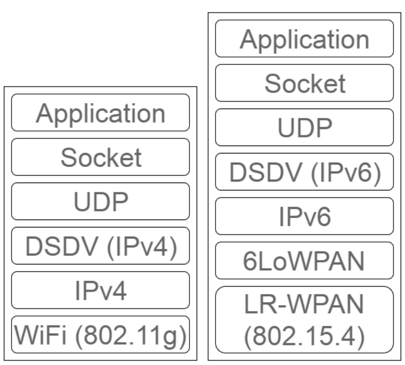
In this section it will be described all the details concerning the implementation of wireless sensor network, using the NS-3 simulator [14], which is developed in C++ programming language.

As mentioned above, the objective of this study is to collect sound samples of military vehicles through acoustic sensors in network nodes, to make their classification and forward it to the base station.

Since it will be used the NS-3 simulator to create this network and this implementation will not be done on the field, it is intended to use Matlab to perform all tasks related to the classification. The NS-3 will only be used to simulate the wireless sensor network, which means that these two components are being developed separately. Therefore, it will be included in a file values associated with the number of a vehicle corresponding to the result of the classification performed on each node.

The only goal of the sensor network simulations is to route the data resulting from the classification, efficiently, to the base station, using data aggregation on intermediate nodes. As mentioned in Section 2.2.5 it will be used two technologies thus, two versions were developed: Version 1: Uses the DSDV routing protocol, the IPv4 protocol and the standard IEEE 802.11g (Wifi); Version 2: Uses the DSDV routing protocol, the IPv6 protocol, the 6LoWPAN and the standard IEEE 802.15.4 (LR-WPAN).

The protocol layers associated with Version 1 and 2 are represented in the Figures 3(a) and 3(b), respectively.



(a) Protocol stack of Version 1. (b) Protocol stack of Version 2.

Figure 3. Protocol stack of the two Versions.

3.2.1 Version 1 implementation

The objective is to have only one base station and all the nodes static except one, that will simulate a vehicle. Therefore, an original example of NS-3 was modified to this end.

As mentioned before, it is necessary to read from a file the number of a vehicle that resulted from classification and subsequently include this number in packets before sending to the base station. It was developed a class that contains a method that given the file name returns the contents.

After being able to get the information of the vehicle, a message was developed, containing two variables: `cardata` - This variable contains the imported value of the file; `ID` - This vector will store the identifiers of each node. Initially, only one respecting the sending node, but after aggregation, if multiple nodes send the same data, several ID's will be associated to that data.

This message will be used in the application and added to the packets before they are sent (Figure 4).

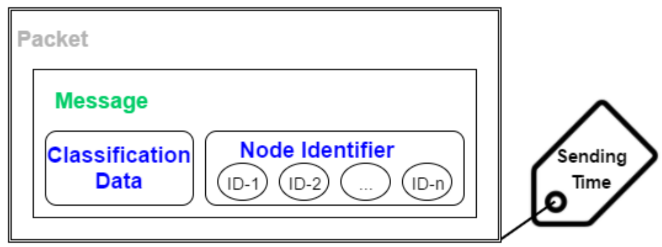


Figure 4. Sending message.

For the delay calculation a tag with the sending time is added to the packet. This tag is analyzed in aggregations, in order to always maintain the maximum delay, and in the base station, in which the calculation of the delay is done.

Before developing the application it must be ensured that packets can be intercepted by it. Thus, it was added to the Internet module, specifically to the IPv4, one packet interceptor developed by Professor Antonio Grilo. The interceptor makes possible the interception of packets that don't have that node as destination, in order to make the aggregation at the application level. The packet is anticipated by the application at the IP layer before it reaches the DSDV layer. Thus, packets are intercepted before being forwarded and are sent by the application after aggregation has been performed. When the application sends the packet it goes through DSDV layer and the routing is done normally.

With the above-mentioned aspects is possible to develop the application that will run on all leaf nodes of the network. This application sends packets from time to time, but only if the vehicle is distanced less than 300m. It's not ideal but it was one way to simulate the sound degradation with distance. As mentioned before, if the node receives some packet during aggregation time, it is intercepted, aggregated and then sent after the specified time has passed. For this end, it was created a timer where it is possible to specify the aggregation time and after that timed has passed the aggregation is sent and the timer is restarted. The base station application only receives the packets and makes the calculation of performance measures.

As mentioned in Section 2.2.5 in order to make the network more efficient a modification to the DSDV module was made. This modification aims to propagate only the address of the base stations to the network nodes, which implies that all nodes will contain only the route to the base stations. It can be chosen which nodes introduce the route in the update, making this modification suitable to the application in question but also for other applications containing more than one base station.

3.2.2 Version 2 implementation

The Version 2 implementation is identical to the Version 1, the main differences are: the conversions from IPv4 to IPv6 Classes and the developing of an DSDV IPv6 module, because in NS-3 there was not any IPv6 routing protocol.

The problem of incompatibility with IPv6 module has to do with the Neighbor Discovery Protocol (NDP), similar to Address Resolution Protocol (ARP), which exists in IPv4. The NDP has specific addresses called solicited-node multicast addresses, which are used to make requests to the nodes in order to get the MAC's address. These requests are made using the Internet Control Message Protocol version 6 (ICMPv6), which works in the simulator at the transport layer. The ICMPv6 contains different types of packets, one of them is Neighbor Solicitation, whose function, as mentioned before, is to determine the MAC's address of the neighboring nodes, sending the packets to the solicited-node's multicast address. It is important to note that first, the ICMPv6 sends requests for Router Solicitation, to all routers multicast address and the Neighbor Solicitation is only sent after response.

In Version 1 ARP was not a problem, since the broadcast addresses were transparent to the DSDV layer. In NDP case, the addresses used are different and the DSDV routing cannot ignore them, or consult the routing tables and return a route. Therefore, there is no transparency in DSDV to ICMPv6 packets, which means that they are discarded.

It is then necessary to make the ICMPv6 packets transparent for DSDV layer. Thus, a destination address verification is performed and a route is returned manually to the layers below. In the transport layer, ICMPv6 protocol decides whether the node that received the packet is the destination or not. If the node is not the destination, the packet is simply dropped.

In practice, a verification of the destination address as solicited-node multicast address or all routers multicast was added to the function that handles the sending in the DSDV module (IPv6 version). On DSDV receive function, the ICMPv6 packets can have as destination: the multicast address or the node global address. If the packet reaches the right destination, the protocol discards these packets because these addresses are not in the routing tables. Accordingly, it is necessary to introduce a condition to send to the above layer the ICMPv6 packets to be properly handled.

4. Results

4.1 Characterization and classification of vehicles using GMM

Sets of random synthetic data were created with Gaussian distribution, having a given mean and standard deviation. These tests allowed to conclude: a number of components equal to the number of distinct groups allows GMM to make the correct representation of the data; If the data groups are not distinct, increasing the number of components may not improve the classification; When exist fewer components than the number of different data groups, the means of the components are close to the data groups with more samples and denser (less distance between them) and the covariances increase; With only one distinct group, using a GMM with more components may be redundant or can represent the data incorrectly; If there are few samples, even with the same mean and standard deviation, that doesn't define a separate group, the GMM can not correctly represent this data.

Given these findings it was possible to start the tests with the vehicle data, using eight cepstral coefficients for the MFCC and the windows: 50ms, 100ms, 150ms, 250ms, 300ms, 350ms, 400ms, 450ms, 500ms, 550ms, 600ms, 800ms and 1000ms with 50% overlap.

The performed tests were: Test 1 - MFCC with eight cepstral coefficients for all the windows and their classification; Test 2 - MFCC with eight cepstral coefficients and BPM for all the windows and their classification.

In Test 1 it was found that the results were, in general, better than those of Test 2 i.e., by adding the BPM the classification worsens. This may be due to the fact that this rhythm component is not the best feature to distinguish vehicles. It is important to state that in both tests you can achieve an F-score of 0.875, making the correct classification of seven vehicles in eight but, overall, the Test 1 produces better results.

The execution time is a key factor when choosing the best option because, in the real case, we have to process the sound and classify the vehicle before sending the result to the base station. By analyzing the execution time it is possible to conclude that the smaller the covariance matrix and the means vector and the less the MFCC, the faster the classification. However, the MFCC calculation time is higher for larger windows because the MFCC uses the fast Fourier transform that has a computational time proportional to $N * log(N)$ [17].

In Test 1 it is possible to confirm the following: for very small windows the computational time is very high, since even if the computation of MFCC is faster, there is a large amount of data. As the window increases, the computational time begins to decrease, but in 500ms, the time increases (MFCC computational time is higher), which indicates that the ideal window in this case is 300ms.

In Test 2 it is also possible to obtain an F-score of 0.875, but with a window of 150ms, which implies the following drawbacks: This window can lead to a non- invertible matrix in the classification; It has a longer computational time than the 300ms window of Test 1.

4.2 Wireless sensor network

This section presents the results obtained using the implementation described in Section 3.2.

There are several independent variables of interest to the study of the sensor network performance: Time that the station waits to analyze the data; Simulation time; Density of the nodes in the battlefield; Number of nodes; Disposition of nodes; Developed versions with IEEE 802.11g and IEEE 802.15.4 technologies; Existence, or not, of the developed aggregation; Aggregation time; Distance limit from the node to the vehicle, in which packets are sent.

Given the independent variables, the measuring of the network performance could be done with the following dependent variables: Losses - To calculate the losses of the network. Since the tests can be done with or without aggregation, the number of packets lost is not important. Therefore, identifiers that the base station has not received and should have received are the important measure; Delays - In the case of aggregation it is always kept the minimum time that is, is always calculated the maximum delay. Therefore, it is calculated the maximum delay average of all the packets. In order to obtain the confidence intervals the absolute delay, corresponding to the maximum delay in receiving a packet during the entire simulation, and the minimum delay, corresponding to the minimum delay in receiving a packet during the entire simulation, were also measured; Energy - It is important to account for the energy consumed by the network and the energy spent on each node because, in the application under consideration, one node may never send information about vehicles, while others can send and receive lots of information. Goodput - This parameter determines the number of bytes/s, related to the sound, received in the base station. Goodput without node processing - This parameter determines the number of bytes/s, related to the sound that are required to be received in the base station if there was no processing on the nodes.

Network tests were then carried out in both versions, always oriented to obtain the performance measures of the specified network. To not exist a high number of simulations some independent variables were fixed, with the effort of keeping the simulation consistent with the reality.

It was fixed the time that the base station waits to analyze data so that the identification and vehicle classification is not too far from the real time (10s), the simulation time (2000s), the number of nodes (100) and the distribution of nodes in the field, putting them equally spaced between them thus, allowing to change the spacing and therefore the density of the network. The threshold distance of a node to the vehicle, which enables packets to be sent to the base station was fixed at 300m and this can be seen in Figure 5.

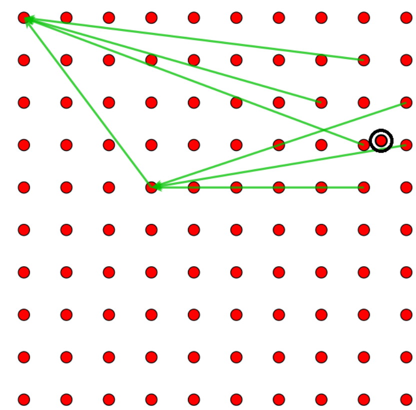


Figure 5. Nodes distanced less than 300m form the vehicle sending packets.

To be able to generate the network performance measures with variable aggregation time and density, the simulations were divided into two groups. In the first group the aggregation time was fixed and the density of the nodes varied. In the second group the density was fixed and the aggregation time varied.

For each density, version, with and without aggregation in the first group and for each aggregation time, version, with or without aggregation, in the second group ten simulations were performed, in each one them different vehicle paths were used.

The simulations of the first group allowed to verify that the aggregation always introduces delays, but on the other hand, leads to energy saving. When comparing the two technologies, it was found that the 802.11b performed better regarding to losses and it allowed spacing higher than 100m which as a consequence led to less nodes needed to cover a larger area. Technology 802.15.4 has lower power consumption, even less with aggregation, but it can't have 0% losses and cover areas as large as the 802.11g.

Second group simulations allowed to find, in addition to the previous findings, that increasing aggregation time leads to lower Goodput and less energy consumption, but always lead to higher losses and delay therefore, there were packets delivered overtime.

Choosing the density of nodes and aggregation time is always a compromise between delays, network range and power consumption. In this particular case, it is intended that the network has a certain autonomy with some margin for losses below 1%, since the packets should not have long delays for the identifying to be as close to real time as possible. With these constraints and given the abrupt difference in the power consumption of the two technologies, the 802.15.4 with aggregation would be the most appropriate to the case study. Unless, given some tactic restrictions, it is desired to cover large areas with few nodes and in that case the technology 802.11g would be the most suitable one.

5. Conclusion

The proposed objective was to develop a wireless sensor network in a military environment, whose purpose is to detect and classify military vehicles in operational context. To this end, tools to extract components of the sound like frequency analysis (MFCC) and rhythm analysis (BPM) were used. These tools enabled the generation of data that can be represented by GMM.

The GMM were used to represent data from different vehicles in the Armed Forces thus, leading to the creation of a database. With this database, it became possible to classify the vehicle audio segments. Subsequently, the data resulting from this classification was sent to a base station where the data was analyzed. To this purpose, a wireless sensor network was developed, in order to route packets to the base station using the DSDV routing protocol. In order to reduce the power consumption of the network, some modifications were implemented to this protocol, such as data aggregation. Two different technologies were tested, the 802.15.4 and 802.11g in order to be able to see which one would be better for the application under consideration.

After the analysis of the results it appears that there are sound components, in this case BPM, which do not produce a good representation of vehicle sound. The MFCC individually proved to be the best way to analyze sound in order to achieve the best classification with GMM.

When sending the packets containing the classification it was expected that the technology 802.15.4 along with aggregation would provide lower power consumption, which would lead to greater autonomy for the network. This was proven right by the simulations at different distances and with different times of aggregation.

It was found that for smaller spacing between nodes, the technology 802.15.4 behaves generally better than 802.11g. It is important to state that the technology 802.11g leads to lower delays and losses and that this difference can be significant, even if it has a high energy consumption. In scenarios where a very large area needs to be covered using few nodes, the technology 802.11g is a better choice.

The developed system has some limitations, and many can be addressed in future work. The sound collected to create the database has not enough time and would be recommended to develop a more consistent database. In developing this database, it is also important, to make simulations more real, to develop a sound degradation model related to distance.

The microphone used for recording the sounds was better than the microphones that a node will contain so, it is important to make tests with sound captured by microphones that can be used on the nodes. The system developed is only prepared to identify a different vehicle at a time, that is, if there are two or more vehicles simultaneously to be recorded, it is unclear what will be the system behavior. It is necessary to develop a mechanism that can overcome this problem.

In wireless sensor network, only two technologies were tested in the physical layer (802.15.4 and 802.11g). It is possible that this network can get better performance with other technologies, or with other versions of IEEE 802.11.

Overall, the results obtained were the expected ones and it is considered that, even with some limitations that may be addressed in future work, the stipulated objectives were achieved.

Acknowledgements

We acknowledge all the support provided by the Officers, Sergeants, Corporals and Privates of Escola das Armas, the Military Camp of Santa Margarida and the Alcochete Air Force Shooting Range, for the help provided in sound collection of vehicles belonging to the Army and the Air Force.

References

[1] V. H. D. de Almeida *et al.*, "A aquisição de objectivos e a artilharia na vanguarda da tecnologia," http://www.exercito.pt/sites/EPA/Publicacoes/Documents/Boletim2009-web.pdf, Dec. 2009, consultado em 19 de Setembro de 2016.

[2] Y.-C. Wu, "Gaussian mixture model," *Connexions*, pp. 1–3, 2005.

[3] R. Mott, *Sound Effects: Radio, TV, and Film*. Focal Press, 1990. [Online]. Available: https://books.google.pt/books?id=lytTAAAAMAAJ

[4] M. Lutter, "Mel-frequency cepstral coefficients," http://recognize-speech.com/feature-extraction/mfcc, Nov. 2014, consultado em 20 de Novembro de 2015.

[5] D. P. W. Ellis, "Beat tracking by dynamic programming," *J. New Music Research*, vol. 2007, pp. 51–60, 2007.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[7] J. H. Plasse, "The em algorithm in multivariate gaussian mixture models using anderson acceleration," 2013.

[8] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *in: Proceedings of the 27th European Conference on Information Retrieval*, 2005, pp. 345–359.

[9] S. A. Chris Townsend, "Wireless sensor networks: Principles and applications," *MicroStrain, Inc*, pp. 439–449, 2004.

[10] A. E. K. Jamal N. Al-Karaki, "Routing techniques in wireless sensor networks: A survey," *Wireless Communications, IEEE*, pp. 1–36, 2004.

[11] G. Spandan, A. Patel, C. R. Manjunath, and N. GS, "Data aggregation protocols in wireless sensor networks," *International Journal of computational Engineering Resaerch*, pp. 18–24, 2013.

[12] J.-S. Lee, Y.-W. Su, and C.-C. Shen, "A comparative study of wireless protocols: Bluetooth, uwb, zigbee, and wi-fi," in *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*. IEEE, 2007, pp. 46–51.

[13] J. W. Hui and D. E. Culler, "Ip is dead, long live ip for wireless sensor networks," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 15–28.

[14] NS-3, "Ns-3," https://www.nsnam.org/docs/release/3.24/tutorial/ns-3-tutorial.pdf, 2016, consultado em 8 de Fevereiro de 2016.

[15] M. Nozokido, "Zoom h6," https://www.zoom-na.com/products/field-video-recording/field-recording/h6-handy-recorder/specs, 2013, consultado em 5 de Janeiro de 2016.

[16] Y. Yang, "Can the strengths of aic and bic be shared? a conflict between model indentification and regression estimation," *Biometrika*, vol. 92, no. 4, pp. 937–950, 2005.

[17] S. G. Johnson and M. Frigo, "A modified split-radix fft with fewer arithmetic operations," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 111–119, 2007.