Classification of BLE packet according to AGC index with detection of non-ideal reception cases

Morgane Joly
NXP Semiconductor
Caen, France
morgane.joly_1@nxp.com

Éric Renault
LIGM, Univ. Gustave Eiffel, CNRS, ESIEE Paris
Noisy-Le-Grand, France
eric.renault@esiee.fr

Fabian Rivière
NXP Semiconductor
Caen, France
fabian.riviere@nxp.com

***Abstract*—The unlicensed 2.4 GHz band is popular among the personal wireless applications. Even if protocols using this band are implemented with countermeasures, they are no longer robust enough to maintain good quality of service (QoS). Interference between low data rate short-range communication protocols, such as Bluetooth Low Energy (BLE), and high data rate protocols, such as Wi-Fi, is symptomatic of the QoS loss in this band. We present a new way to allow the narrowband protocols receiver to adapt its internal states to the signal power by improving the adaptive gain control (AGC). As a preliminary work, we compare four machine learning algorithms (Decision Tree, Bagged Tree, Support Vector Machine (SVM) and K-Nearest Neighbour (KNN)) to classify BLE received packets subject to an interferer. The receiver robustness can then be improved by combining an optimised AGC index decision with the detection of non-ideal reception conditions. The best model in terms of its performance/complexity trade-off here is the KNN, which classifies the packets with an accuracy of 95.6 %.**

***Index Terms*—ML, BLE, Countermeasure, CR**

I. Introduction

Originally intended for industrial, scientific or medical purposes, the ISM band has became popular for communication at close range devices, offering a good trade-off between hardware cost and wave propagation. Victim of its own success, the 2.4GHz band subject to spatial congestion causing a degradation in the quality of service (QoS) of devices. As highlighted in [4], [8], in crowded context low-power protocols such as BLE suffer from a severe increase in latency during the device discovery process. This behaviour correlates with the number of interfering devices of the same protocol, and large sensor networks are degraded by Wi-Fi, even when coexistence techniques are in place to reduce the impact of interference.

This paper investigate the relevance of artificial intelligence algorithms for the classification of Bluetooth Low Energy (BLE) packets while respecting the constraint of the BLE radio. The aim is to improve the receiver linearity/noise trade-off selected by the Automatic Gain Control (AGC), which is responsible for adapting the radio to the signal strength at the antenna, thus avoiding saturation of the radio to ensure that the payload is received. Here, are exploited internal radio radio metrics linked to the received signal quality produced during packet reception (Received Signal Strength Indicator (RSSI), Signal to Noise Ratio (SNR), Link Quality Indicator (LQI) and Cyclic Redundancy Check (CRC)).

This article shows that it is possible to classify BLE packets according to an optimised AGC index is possible with a good performance/complexity ratio. The potential to avoid retransmission of packets identified as recoverable is highlighted.

This paper is organised as follows. Sec. II presents some related works and Sec. III gives some background. Sec. IV presents the environment setup, and Sec. V presents experimental results, followed by a discussion. Conclusions are summarised in Sec. VI.

II. Related work

Machine learning techniques for classification and detection of non-ideal reception cases is the main subject of this work. This section present the integration of AI algorithm to improve QoS then present few works using internal metrics with ML algorithm are presented to support some RF related purposes.

O'Mahony and al. [6] exploit the in-phase (I) and quadrature-phase (Q) signal of a Zigbee radio receiver with a random forest algorithm to detect malicious interference. A classifier is created in [1] combined with a specific packet to investigate the presence of a harmful interferer. Lee and al. [5] propose the subtraction of ambient noise from the signal to improve its detectability thanks to a neural network. Natively, devices cannot detect nearby protocols to apply the best countermeasure associated with them. Wang and Zhang in [9] use a decision tree to take into account the MAC layer in the spectrum sensing scheme to design a new strategy of emission and improve the quality of service of secondary users by choosing the broadcast time wisely and avoiding the primary user. Intelligent channel assignment is a well-documented research field with promising results based on deep learning [10] and genetic learning algorithms [7]. However, the use of highly energy-consuming algorithms may reduce the interest in these techniques despite promising results. [3] the team enhances the delay of emission and reduces the energy consumption during the discovery phase of available access points thanks to a WiFi scanning manager application.

III. Background

In this part we present briefly some IA algorithm notion and BLE radio particularities that we use in next sections.

979-8-3503-7099-7/24/$31.00

A. Artificial intelligence background

Here we present here the four algorithms compared and give an insight into how they work. The algorithmic complexity of the solution presented is also provided. Let n be the number of samples and p be the number of features.

1) Decision Tree: The decision Tree algorithm is a classic ML algorithm used for classification problems, known for its great generalisation capabilities and for working well with large datasets. The general training complexity of this algorithm is $O(n \times p \times \log(n))$.

2) Bagged tree: This type of approach combines multiple decision trees (called weak learners) using bootstrap aggregation method to form a stronger estimator. The general training complexity of this algorithm is $O(n^2 \times p \times \log(n))$.

3) K-Nearest neighbours (KNN): The classification algorithm uses the vote of the k nearest data in the n-space to determine the class of the new arriving data. The general training complexity of this algorithm is $O(n \times p \times \log(n))$.

4) Support Vector Machine (SVM): SVM is an algorithms family designed originally for discrimination between two classes, it is based on the search of optimal hyper-planes separating data. The general training complexity of this algorithm is $O(n^2 \times p)$.

B. Radio Frequency background

BLE and Wi-Fi are two non-collaborative protocols operating in the same unlicensed 2.4GHz band. Hence, they are affected by each other's communications without any means of detecting and identifying each other's transmissions. In the BLE radio, 3 advertising channels and 9 data channels out of 39 possible are potentially interference-free from Wi-Fi concurrent operations. Internal countermeasure systems are in place to reduce disruption repercussion from other protocols with a balance between energy saving and the complexity of receiver filtering.

The AGC adjusts the receiver gain according to the received signal strength, allowing the radio to avoid saturation during packet reception. The AGC gain index is set based on the packet preamble, and frozen during the payload to avoid corruption of the bit decoding. However, this mechanism cannot preserve the communication from a signal power increase within the receiver after the AGC frozen which creates receiver saturation (see Fig. 1). The higher the interferer, the lower the receiver gain and therefore the lower the AGC index value is. We will use later the terms of under-restricted for an AGC index value higher than required and over-restricted for an AGC index value lower than needed. If the gain is not sufficiently restricted, noise is favoured at the expense of linearity, resulting in better detection of a low power signal but increased sensitivity to the intermodulation product.The terminology of *After* will refer to an interferer that arrives after the AGC gain freeze. Similarly, the terminology of *Before* will refer to an interferer that arrives before the AGC gain freeze.

1) Receiver metrics used as metrics for ML algorithm:

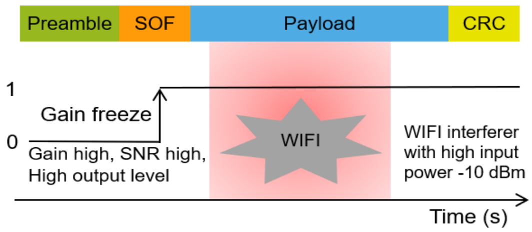

Fig. 1. Interferer arriving after the AGC freeze, perturbing the payload reception. The AGC cannot sustain the power variation.

a) Fast Link estimators : Received Signal Strength Indicator (RSSI), Link Quality Indicator (LQI) and Signal to noise ratio (SNR): These three metrics provide a good indication of the quality of the received signal, and as a consequence, an indication of the link quality in between the transmitter and the receiver. RSSI provides an estimate of the average power in a channel throughout the reception of a packet, all signals included, while LQI and SNR quantify the signal-to-noise ratio, i.e. the ability of the receiver to decode the signal despite ambient noise. SNR gives the margin of the signal over the noise and the ability of the demodulator to recognise the bits. There are good to instant estimation but as exposed in [2], their are poor resources on their own to determine the interferer presence.

b) Cyclic Redundancy Check (CRC): The CRC is a word composed of a few bits calculated by the emitting radio from the payload part of the send packet. It is appended to the end of the packet before it is emitted. When the receiver decodes the packet, this word is recalculated with the received payload and compared with the received CRC to check the validity of the packet.

c) Access address (AA) found : This binary metric provides a confirmation of packet start detection. Two scenarios can lead to an access address not being detected, either the wanted signal strength is too weak for the receiver despite the absence of an interferer, or there is a too powerful interferer at the beginning of the packet reception.

IV. Interference disturbance Classification methodology

A. Scope and environment

The BLE protocol is the core of this study. However, other lightweight protocols (such as Zigbee) could also be investigated using our method. The experimental environment setup consists of a BLE receiver simulated using Matlab (v2020b) and Simulink modelling tools. The seeds used to randomise the packet simulation are controlled to ensure the reproducibility of the experiments.

B. ML Training and Test methodology

The physical restrain of BLE devices exclude the use of deep learning algorithms because of their huge time complexity. In addition, due to the high dynamic of the spectral environment, it is mandatory for the algorithm to achieve all the operation in real time. We have selected four algorithms, namely Bagged Tree, KNN, Decision Tree and SVM, from the twenty-four algorithms proposed by the ML Graphical User Interface (GUI) of Matlab. The selection was made according to their accuracy and precision. The best algorithm in each family was selected. Then, using the function of the ML toolbox, a detailed analysis of their variability is carried out, in Sec. V.

C. Packet labelling : Set the optimal AGC

The reception status is an logic indicator that combine both **Access Address found** and **CRC** features (see Table I).Three designations are deduced: *Good reception*, *Bad Reception*, *No detection*. In Sec. III, the role of the AGC mechanism in the receiver saturation avoidance and its limitation in the case of late interferer arrival scenario has been exposed. For the labelling process, in most of the cases, native AGC gain index decision is taken into account, but some packets are poorly received regardless of the interferer arrival time. Specific classes have been created to identify these packets.

TABLE I
PACKET RECEPTION STATUS CLASSES

	Good reception (GR)	Bad reception (BR)	No detection (ND)
CRC	1	0	0
AA detected	1	1	0

These indicators are combined to sort the packets into classes that put in relation the reception status of the two packet versions according to the following naming convention : The first two letters stand for the reception status for the version when the interferer arrives *Before* the AGC index freeze, while the last two letters indicate the reception status for the version when the interferer arrives *After* the AGC index freeze. The optimal AGC is identify for most packets, the remainder being the packets that cannot be improved by an AGC modification. Three additional labels X, Y and Z are created to distinguish the evolution of reception status (see Tab. II). For these three labels, AGC modification will not lead to an improvement. Other countermeasures must be proposed such as adjusting channel hopping table or changing the communication offset.

- GR_GR: Whatever the timing of arrival of the interferer, the packet is always well received, the packet label is the original AGC index of the packet.
- GR_BR: If the interferer is late, the packet is badly received because the AGC index is not properly adapted to avoid saturation, the packet label correspond to the AGC index of the good received version of packet.
- BR_BR: No matter when the interferer arrives, the packet is never well received.
- ND_GR: The receiver expects a packet but the interferer is too powerful for the receiver to detect the access address. However, the payload can be well received if the interferer arrives after the AGC index freeze.
- ND_BR: The interferer is too powerful for the receiver to detect the access address and the payload cannot be received well.

TABLE II
THE DETERMINATION OF THE PACKET LABELS WITH GR : GOOD RECEPTION, BR : BAD RECEPTION, ND : NO DETECTION

Before_After	Optimal Label
GR_GR	Original packet AGC index
GR_BR	AGC index from *Before* version
ND_BR	X
ND_GR	Y
BR_BR	Z

D. Effect of the forced AGC index on GR_BR class

In Sec. IV-C, packets in class GR_BR leads to AGC index value selection not suitable to sustain interferer power level because it arrives after the AGC index freeze. However, it could have been well received if the AGC index had arrived before the AGC index freeze.

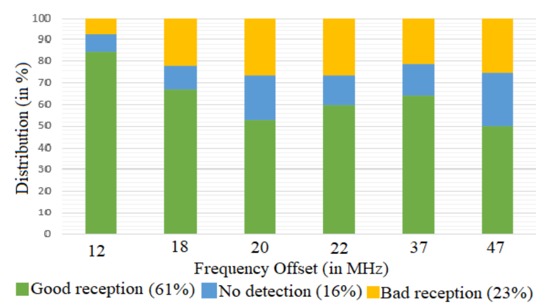

Fig. 2. Recovered packet reception after AGC index forcing (packets originally missed due to interferer arrival after AGC freeze, then AGC forced to value got when interferer occurrence before wanted packet)

By forcing the AGC index of the *After* version packets of this class during the preamble of the packet, we obtain a 61% improvement on the packet subset (see fig. 2).

V. RESULT ANALYSIS OF THE PACKET CLASSIFICATION ACCORDING TO THEIR AGC

A. Data collection

BLE packets are retrieved at six different interferer frequency offsets: 12 MHz, 18 MHz, 20 MHz, 22 MHz, 37 MHz, 47 MHz. A packet is generated by setting the desired signal and the blocker power level at the antenna in the range [-100 dBm, 0 dBm] and [-120 dBm, 0 dBm] range with a 2dB steps. Two types of datasets were generated: packets with the interferer arriving *Before* the AGC index freeze and packets with the interferer arriving *After* the AGC index freeze. Each packet reception scenario exists in two versions. Receiver internal metrics showing the evolution of the reception quality are pre-processed to reduce the computation of the ML-based algorithm. We collect data at three representatives time intervals: at the beginning of the preamble, in the middle of the payload and before the CRC. Therefore, the metrics collected at these time intervals are the wideband and narrowband RSSI, the SNR, the LQI. The AA Found and CRC status indicators, which provide exclusive information for a packet, are stored at the beginning and at the end of the packet reception, respectively.

The dataset consists of 23954 packets, which exist in two versions, i.e. 47908 packets in total. 7-fold cross-validation is used to create balanced training and test data for each of the fifteen training runs.

B. Result

Tab. III compares the accuracy of the four algorithms studied for each of the fifteen training groups. Bagged Tree, KNN and Decision Tree have excellent results, with an accuracy of 96.6% $\pm$ 0.12%, 95.6% $\pm$ 0.117%, 95.2% $\pm$ 0.22% respectively for each retraining. SVM score falls to 93.81% $\pm$ 0.21%.

TABLE III
ACCURACY OF THE TRAINED ALGORITHMS WITH STANDARD DEVIATION (STD)

	Bagged tree	SVM	Decision Tree	KNN
Accuracy	96.6 %	93.81 %	95.2 %	95.6 %
STD	$\pm$ 0.12	$\pm$ 0.21	$\pm$ 0.22	$\pm$ 0.117

C. Algorithms Precision

To evaluate the precision of the algorithms we choose to use on one hand the Mean Square Error (MSE) (see Tab. IV) to determine the mean error of the numerical label (0 to 11), and, on the other hand, use the mode to evaluate the most frequent error of the class X, Y, Z (see Tab. V). Tab. IV presents the Mean Square Error (MSE) between the classification proposed by the algorithm and the expected AGC index of the packet. Bagged Tree shows the smaller errors with the most stability with an MSE of 0.041 $\pm$ 0.004.

TABLE IV
MEAN SQUARE ERROR OF NUMERICAL CLASS WITH STANDARD DEVIATION (STD)

	Bagged tree	SVM	Decision Tree	KNN
MSE	0.041	0.073	0.051	0.044
STD	$\pm$ 0.004	$\pm$ 0.004	$\pm$ 0.005	$\pm$ 0.004

TABLE V
ERROR MODE

	Bagged tree	SVM	Decision Tree	KNN
Mode X	Z	Z	Z	Z
Mode Y	X	11	X	X
Mode Z	X	X	X	X

Figures 3, 4, 5 and 6 highlight the mean misclassification rate matrix. Regarding the numerical class, the Bagged Tree and SVM show a distribution of errors grouped around of the main diagonal with a tendency to AGC index under-restrict the AGC index for class 4 to 11. The Decision Tree is less accurate with a wider error distribution. The KNN presents errors grouped tightly around the main diagonal with errors mainly in the range of $\pm$ 1.

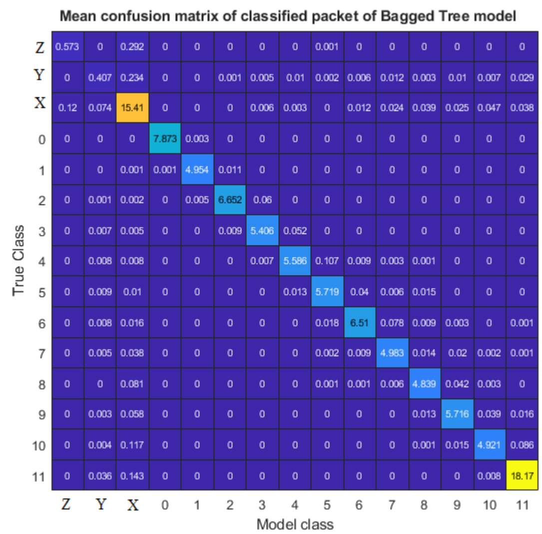

Fig. 3. Confusion Matrix of Bagged Tree

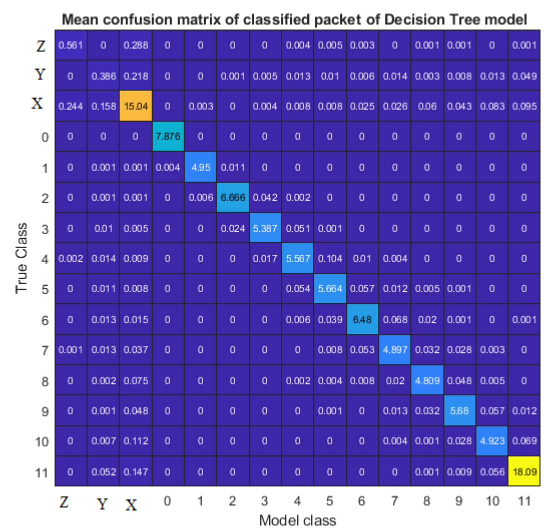

Fig. 4. Confusion Matrix of Decision Tree

D. Discussion

This comparison shows that the classification of the received packets by a BLE receiver according to their AGC index with the recognition of their specific reception status is possible with an accuracy of better than 90 %. BT, KNN and DT show similar results in terms of accuracy. The comparison of numerical MSEs confirms that Bagged Tree has the best precision performance with an error of 0.041 i.e. the error of

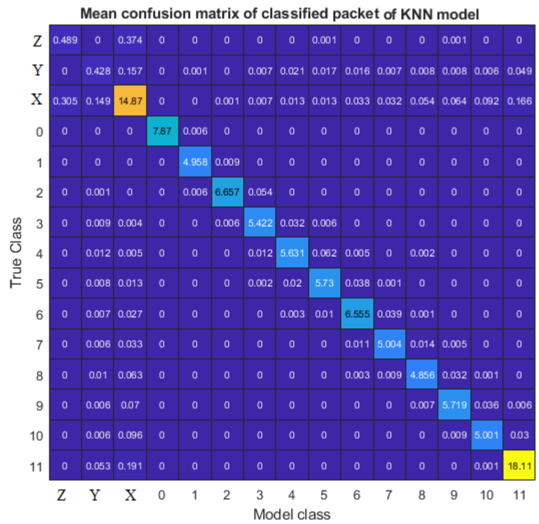

Fig. 5. Confusion Matrix of K nearest neighbour

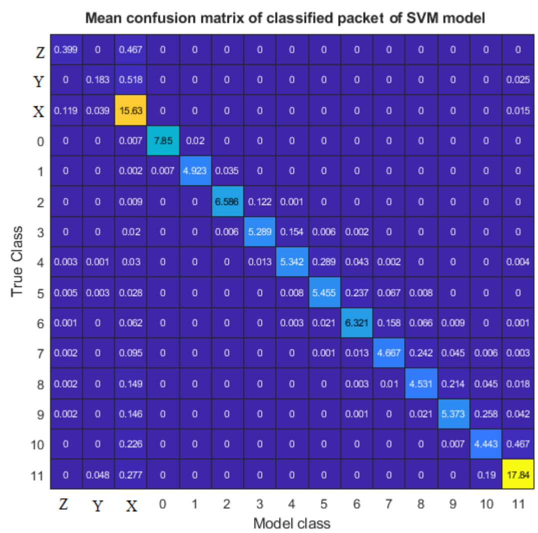

Fig. 6. Confusion Matrix of Support Vector Machine

the label is mainly included between ±1, which are usually sustained by the receiver.

Regarding the class X, Y and Z, according to the confusion matrix, models perform well in X and mode error shows that all models, except SVM, have the majority of their error located in class X or Z. This confusion questions the interest of distinguishing these three cases since they will be managed by countermeasures other than AGC improvement. Nevertheless, the proportion of samples Y and Z is unbalanced compared to the other classes due to the low number of occurrences of these cases in our data collection plan.

The current results show that Bagged Tree and KNN present very similar performance, regarding the need of the minimum complexity the best accuracy/complexity, it is the KNN that shows the best overall performance.

The simulation environment is controlled, which explains the high performance of the models. For the future, we have planned to test our assumptions with more stringent radio coexistence conditions. However, the current coverage provides a good trend regarding classification capability.

VI. Conclusion

The Bluetooth Low Energy is threatened by high-power interferers that cause an over-consumption and loss of quality of service. We have identified new improvement axes by using internal metrics characterising the signal quality (RSSI, SNR, LQI) during packet reception. We has used ML algorithms to classify tuples of wanted signal power, interferer power and frequency offset by their optimal AGC index value and non-ideal reception conditions. Among the four algorithm tested Bagged tree has the best performance with an accuracy of 96.6 % ±0.12 % and KNN has shown the ratio performance/complexity with an accuracy of 95.6 % ±0.117 %.

VII. Further Research

This classification according to an optimal AGC is the beginning of a broader work to design a new type of countermeasure. The aim is to make the radio more resistant to an interferer using a broadband protocol, by using only the radio's regular communication times.

Further developments will focus on predicting the optimal AGC index value needed to guarantee reception of the next packet, and extending our solution to other lightweight protocols such as like Zigbee.

References

[1] Barac, F., Gidlund, M., and Zhang, T. Lped: Channel diagnostics in wsn through channel coding and symbol error statistics. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (2014), pp. 1–6.

[2] Barać, F., Gidlund, M., and Zhang, T. Ubiquitous, yet deceptive: Hardware-based channel metrics on interfered wsn links. *IEEE Transactions on Vehicular Technology 64*, 5 (2015), 1766–1778.

[3] Chung, J., Park, J., Kim, C.-K., and Choi, J. C-scan: Wi-fi scan offloading via collocated low-power radios. *IEEE Internet of Things Journal 5*, 2 (2018), 1142–1155.

[4] Elshabrawy, T. Throughput analysis of ieee 802.15.4 enabled wireless sensor networks under wlan interference. In *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)* (2014), pp. 467–469.

[5] Lee, W.-H., Ozger, M., Challita, U., and Sung, K. W. Noise learning-based denoising autoencoder. *IEEE Communications Letters 25*, 9 (2021), 2983–2987.

[6] O'Mahony, G. D., Harris, P. J., and Murphy, C. C. Investigating supervised machine learning techniques for channel identification in wireless sensor networks. In *2020 31st Irish Signals and Systems Conference (ISSC)* (2020), pp. 1–6.

[7] Tanwongvarl, C., and Chantaraskul, S. Performance comparison of learning techniques for intelligent channel assignment in cognitive wireless sensor networks. In *2015 Seventh International Conference on Ubiquitous and Future Networks* (2015), pp. 503–507.

[8] Treurniet, J. J., Sarkar, C., Prasad, R. V., and De Boer, W. Energy consumption and latency in ble devices under mutual interference: An experimental study. In *2015 3rd International Conference on Future Internet of Things and Cloud* (2015), pp. 333–340.

[9] Wang, D., and Yang, Z. An advanced scheme with decision tree for the improvement of spectrum sensing efficiency in dynamic network. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (2016), pp. 1288–1292.

[10] Zuo, Y., Li, Q., Lu, G., and Xiong, H. Deep reinforcement learning for power controlled channel allocation in wireless avionics intra-communications. *IEEE Access 9* (2021), 106964–106980.