

An investigation of the bits corruption in the IEEE 802.11p*

Sébastien Bindel[†] Dorine Tabary Soumia Bourebia
Frédéric Drouhin Benoît Hilt

University of Haute Alsace

Data rate management algorithms aim to perform a proper selection of the signal modulation and the coding rate to avoid the corruption of data bits. This paper describes a preliminary investigation on the bit corruption pattern related to the IEEE 802.11p standard. Measurements have been acquired with an experimental test-bed made up with a couple of software radios to perform white-box tests. Software radios are stationary and operate on the same channel without disturbances coming from concurrent communication. The aim of this experimental test-bed is to represent a static scenario where vehicles are stationary such as a crossroad situation. The data analysis shows that a data length reduction has an impact as much as a decrease of the data rate. A deeper analysis of the data bit corruption distribution highlights that some bits are more corrupted than others, rejecting the independent and identically distributed assumption for some situations. This opens a perspective to design algorithms dealing with multiple constraints, even if they are **NP**-complete.

I. Introduction

Vehicular communications are a building block for automotive applications. Each vehicle across the network can broadcast information to the neighbourhood or send messages to a target destination. Communications are not limited to vehicles but include also the infrastructures, the pedestrians, leading to introduce the concept of Vehicle-to-Everything (V2X). The IEEE 802.11p is an amendment of the IEEE 802.11 standard related to Dedicated Short Range Communication (DSRC) and define communication between Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) [778, 2016]. The related physical layer is closed to the one defined from the IEEE 802.11a, with a change on communication operating in the 5.9GHz band and a bandwidth of 10MHz. However, the IEEE 802.11a has been developed for indoor environments with relatively stationary entities, as the result the performance of the 802.11p might suffer since vehicles are moving in the outdoor.

The main causes of these impairments can be related either to the settings of the materials or to the implemented physical layer. The first one includes the gains of antennas, the transmission/reception power, but the transmission power represents the major challenge, since it determines the amount of energy to send the signal. However, only the receiver is able to determine if the transmission power was sufficient to receive and decode the signal, leading to develop a costly feedback mechanism due to the time varying nature of the channel. The second one covers all process related to the physical

*This article is an extension of the published article presented at the Nets4Workshop 2020

[†]Corresponding author : sebastien.bindel@uha.fr

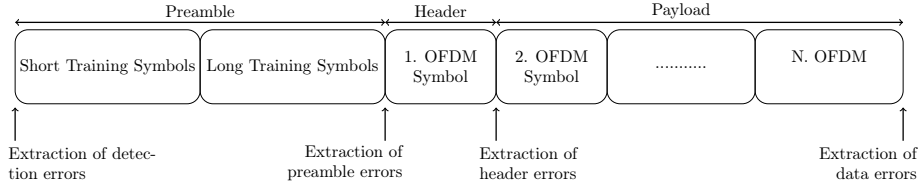


Figure 1: OFDM packet structure

layer from the signal coding to its parsing to decode data. Before the transmission, the 802.11p standard let to the transmitter the proper selection of the modulation and the coding rate which have an impact on the data rate and the probability on the data bit corruption. This selection is performed without taking into account the data length since, the physical layer performs a channel estimation for decoding data and can be quickly outdated due to the time-varying nature of the channel. Therefore, the primary challenge is to determine the impact of the following couple, data rate and data length to the packet error rate. This study goes deeper by pointing out the bit error patterns and determines the computational complexity of the related combinatorial problem to deal both with the data rate and the data length.

The investigation details in this article is based on some articles described in Section II. The Section III details the experimental test-bed which aims to perform measurements on the IEEE 802.11p, made up with software radios (USRP B210) to perform white-box tests and avoid side effects generated by proprietary radio chipsets. A discussion is opened on the impact of the data rate and data length and is extended to the bit error pattern in the Section IV. From this discussion, an investigation of the related combinatorial problems is developed in the Section V and shows the **NP**-completeness when some constraints arise. The Section VI ending this article, opens a perspective on online algorithms related to the related combinatorial problem.

II. Overview of the 802.11p performance

The physical layer of the 802.11p uses as transmission technique the Orthogonal Frequency-Division Multiplexing (OFDM). The OFDM frame structure is composed of three parts, the preamble, the signal and data, as illustrated in the figure 1. The preamble is split in a couple of parts, the short training symbols for the signal detection and the long training symbols for the channel estimation. The signal part gives information for decoding the payload including the signal modulation and the coding rate. At the end, the payload¹ is divided and coded into OFDM symbols according to the coding scheme detailed in the table 1.

The reception process being sequential, each part must be decoded without error to trigger the next one. Major impairments of the reception process are related to the long training symbols used by the channel estimation process. Fernandez et al. [Fernandez et al., 2012] point out the performance of the channel estimation algorithm used for the demodulation. They argue that in a dynamic environment, the channel estimation is outdated before the complete reception of data especially with the Least-Square (LS) method, as the result they have designed a new one called Spectral Temporal Averaging (STA). Such a method require to perform modifications in the physical layer meaning, the development of a specific hardware prototype. Authors have hinted, that the data length has a significant impact, since a shorter length means a shorter reception time and so an higher chance for the channel estimation to not be outdated.

A second strategy is a proper selection of the modulation and the coding rate, because they impact on the Bit Error Rate (BER) and the data rate. The figure 2 depicts the theoretical relationship between the signal to noise ratio and the bit error probability in an Additive White Gaussian Noise

¹In the IEEE 802.11p the payload is the whole MAC frame including the header, data and the tail.

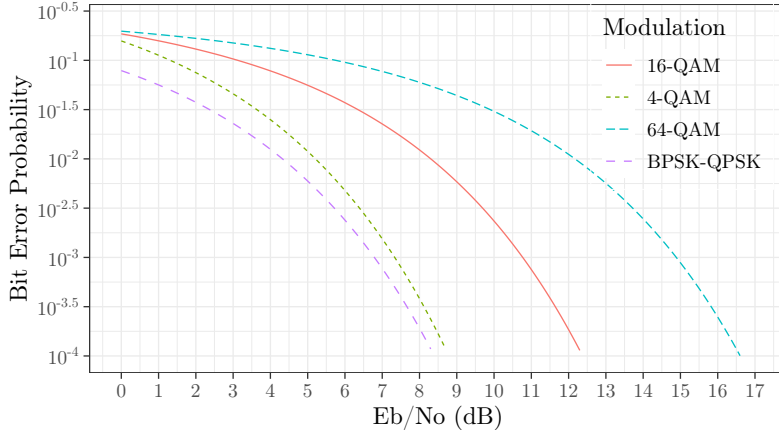


Figure 2: Relationship between the SNR and bit error probability in a AWGN

(AWGN) channel. This plot and the table 1 give an overview of the selection modulation problem. It highlights that more a modulation a scheme offers a faster data rate more bits are prone to the corruption. This opened challenges on the design of data rate management algorithms. These ones are based on a request-answer scheme to determine the proper level of data rate according to feedback sent by the receiver. This feedback is a piece of data related to the incoming signal, frame loss ratio, etc [Yin et al., 2020]. More complex approaches attempt to deal with additional constraints such as the power transmission [Chevillat et al., 2005]. Some further investigations have been performed on the bit corruption to understand the distribution, find some patterns and open strategies, including the important study performed by Han et al. [Han et al., 2012]. In this article is described some preliminary results on the impact of the data rate and the packet size on the loss ratio with an investigation on the bit corruption. The results illustrated in this article confirm partially those obtained by Han et al. [Han et al., 2012].

The proposed approach is a preliminary study and continue the investigations performed by Han et al. [Han et al., 2012]. In contrast with their study, the 802.11p is taking into account and only software radios are used to avoid side effects generating by proprietary radio chipsets. The approach is also motivated by the observations of Fernandez et al. [Fernandez et al., 2012] pointing out the impact of the couple data rate and data length on the packet error rate. This one is confirmed by the study of Bourebia et al. [Bourebia et al., 2020] which have investigated the rate of OFDM errors triggered during the reception process and show that the errors related to the decoding payload are the most important. These ones are confirmed by our results detailed in Section IV. This kind of errors is related to the data bit corruption detected at the MAC layer with the Frame Check Sequence (FCS) algorithm using the Cyclic Redundancy Check (CRC) in the tail. The proposed study focuses on the current version of the 802.11p standard with an experimental test-bed embedding only an open source physical layer within a static environment representing some static vehicular scenarios.

III. Experimental test-bed

The experimental outdoor test-bed aims to represent a scenario wherein entities are static and the communication channel is stable. In order to perform white-box testings, software radios are used and embed an open source IEEE 802.11p transmitter. They have been customised and configured to log all incoming MAC frames, including the corrupted ones.

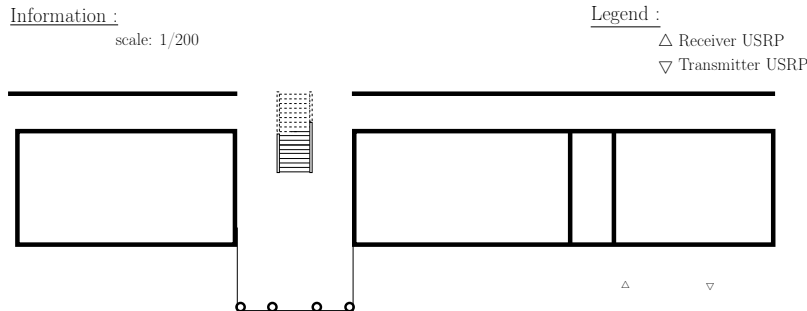


Figure 3: Map of the test-bed

III.1. Materials

All experiments have been performed with a couple of USRP (Universal Software Radio Peripheral) B210 managed by laptops under ArchLinux, one being the transmitter, the other being the receiver. Each radio is equipped with an omnidirectional antenna supporting two frequency slots, 2.4 – 2.5GHz and 4.9 – 5.9GHz with a gain of 3dBi. Software radios have been programmed with the GNURadio 3.8 and embed the IEEE 802.11 a/g/p transmitter module developed by Bloessel et al. [Bloessel et al., 2013]. This one, called gr-ieee802-11, details all physical components of the OFDM transmitter from the signal detection to the MAC frame retrieval. Once MAC frame data are retrieved, these ones are logged into a pcap file thanks to a companion module, called gr-foo, plugged after the gr-ieee802-11 module.

III.2. Material modifications

The source code of the GNURadio software has been modified and compiled manually to set the preprocessor macro `GR_FIXED_BUFFER_SIZE`² to 128Kb instead of 64Kb to not overflow the communication buffers between successive blocks in the gr-ieee802-11 module. This one has been also modified to log corrupted MAC frames detected by the FCS algorithm³. The last but not the least was the log of the SNR value through a floating point number instead as a natural number in pcap files⁴. This modification takes place in the radiotap header wherein the antenna signal field (unsigned 8 bits) contains the integral part of the SNR and the floating part is encoded in the antenna noise field (unsigned 8 bits).

III.3. Experiment settings

The test-bed is made up of a couple of nodes situated along the administrative building as illustrated in the figure 3. Each experiment has been scheduled during the daytime on weekdays and radio have operated on the same channel, 5860MHz, without any interference generated by a wireless device. The transmitter sends periodically MAC frames containing either 42 bytes or 1500 bytes of data, corresponding to the minimum length without bit stuffing and the maximum length regardless the Jumbo frames. Data bytes enclosed in MAC frames are set to 0x00 and each field in the MAC header have a fixed value, except the sequence number, increased by one after each sending. The periodical sending has been defined at 100ms which is lower than the coherence time which represents the time duration over the channel is not varying. The FCS algorithm identifies the presence of errors thanks to a CRC field. This one is a 32 bits CRC value positioned at the tail of the MAC frame. Its computation is based on the XOR operation which is the simplest scheme for error

²Defined in the gnuradio-runtime/lib/flat_flowgraph.cc file

³The source code is available at <https://github.com/sbindel/gr-ieee802-11>

⁴The source code is available at <https://github.com/sbindel/gr-foo>

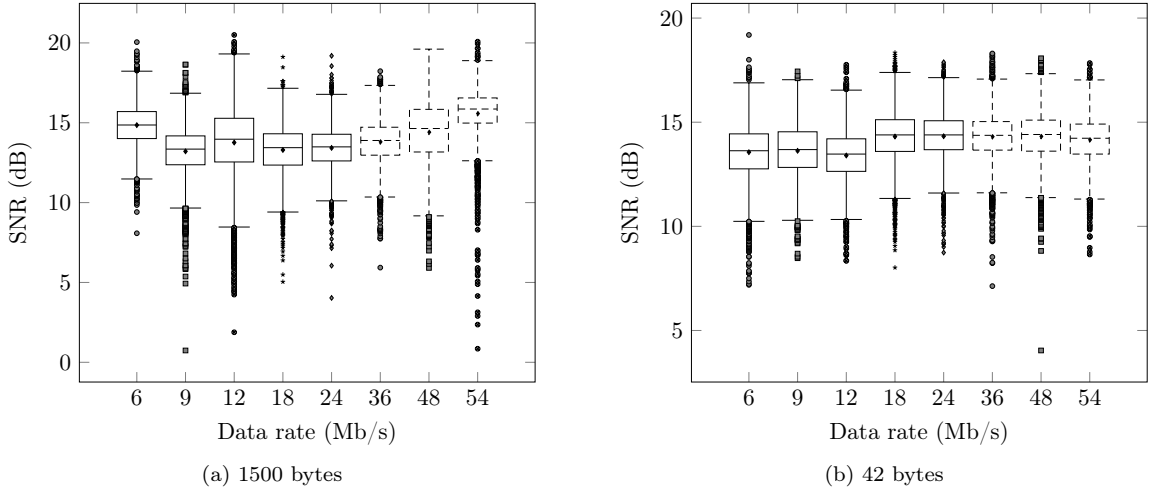


Figure 4: Distribution of the SNR

handling, without a correction possibility. The transmitter and the receiver being software radios, they embed only the physical layer of the IEEE 802.11p, as a result any additional services are provided including the MAC re-transmission scheme when a frame is lost. The channel equalisation in the IEEE 802.11p hardware is responsible to perform a channel estimation for improving the signal demodulation. The time varying nature of the channel tends to disturb the channel equalisation process since the estimation can become outdated. The couple of software radios implement the Least Squares (LS) estimator and uses the long training symbols (depicted in the figure 1) to estimate the channel [Fernandez et al., 2012]. The LS estimator has the benefit to have a low computational complexity but the estimation can be outdated before the complete packet reception. However, the couple of radios being static, the channel should be enough stable to not impair the channel estimation.

IV. Data Analysis

Each experiment has been achieved with all modulations and coding rates, summarised in the table 1. The transmitter broadcasting periodically a fixed bits sequence, the receiver is able to determine if data are lost or received. In the last case, the receiver compares the received bits sequence with the reference to find out the position of corrupted bits. This analysis aims to point out the impact of the data rate and the length of data on the received rate of MAC frames by investigating the first kind of data. It also shows that some data sets related to selected data rate are more interesting to investigate the corruption of data bits. A surprising observation is the bit corruption distribution is not uniform and open a perspective on the minimisation of MAC frame lost without reduce the data rate.

IV.1. Data set Selection

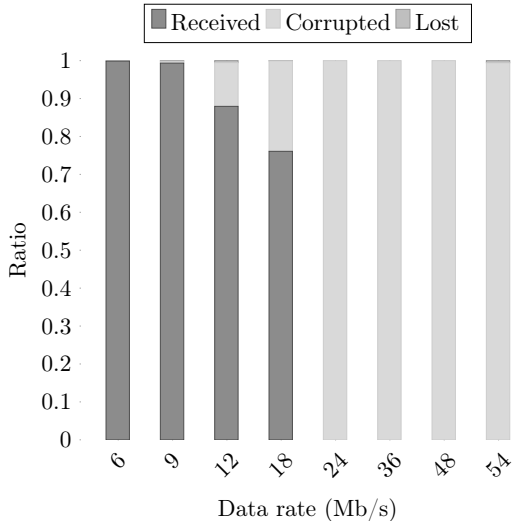
Regardless the bit corruption, the receiver computes for each incoming frame the SNR, the ratio between the power of the incoming signal and the power of the background noise. By definition, it depends on the environment but not on the modulation, the coding rate and the length of data, as a result, it is used to determine if a couple of data sets have been getting within a same environment. The figure 4 depicts the distribution of the SNR for a couple of scenarios within the data length is different and confirms the previous assumption. This couple of scenarios have been selected, since

Data rate	Receivid	Corrupted	Lost	Total
6	4617	6	0	4623
9	3280	18	4	3302
12	4564	599	27	5190
18	3177	997	2	4176
24	0	4073	0	4073
36	0	4217	0	4217
48	0	2994	0	2994
54	0	3735	21	3756

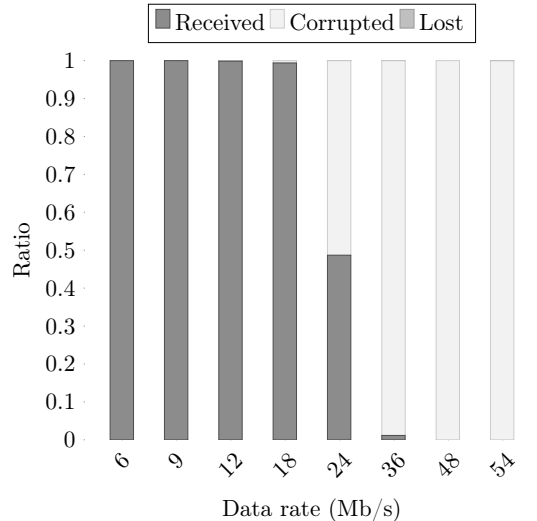
(a) Statistics of reception states (1500 bytes)

Data rate	Received	Corrupted	Lost	Total
6	4385	0	0	4385
9	1890	0	0	1890
12	2734	2	2	2738
18	4170	26	0	4196
24	2543	2680	1	5224
36	55	4722	1	4778
48	0	4092	0	4092
54	0	3962	2	3964

(b) Statistics of reception states (42 bytes)



(c) Reception state ratio (1500 bytes)



(d) Reception state ratio (42 bytes)

Figure 5: The impact of packet size and the data rate on the reception state

their data set contain the most relevant data for this analysis. The SNR distribution appears barely different with a few outliers since the median (central line of the box plot) and the mean (diamond symbol) are approximately equal.

IV.2. Impact of data length and the data rate

The figure 5 depicts the statics and the ratios of the reception states of the incoming packets. The state *Received* covers the MAC frames which have been well received, beside if data are corrupted the frame is considered as *Corrupted*, the rest is declared as *Lost*. In both scenarios, few data are considered as lost since the receiver stays in the transmission range of the transmitter. As expected, when the data rate increases the number of corrupted packets increases too, since there is a positive relationship between the SNR and the Bit Error Rate (BER). However, it differs according the data length. The most significant difference is at 24 Mb/s where the ratio of received frames is ~ 0.48 with a data length of 42 bytes, meanwhile almost all frames with a length of 1500 bytes are corrupted. MAC frame embedding less data than a bigger one are more robust to corruption for a same modulation. It is more surprising to note that the ratio of received frames for a smaller frame is often higher than the bigger sent with the previous lower data rate. This highlight the importance of data length on the reception rate even more to decrease to a lower data rate.

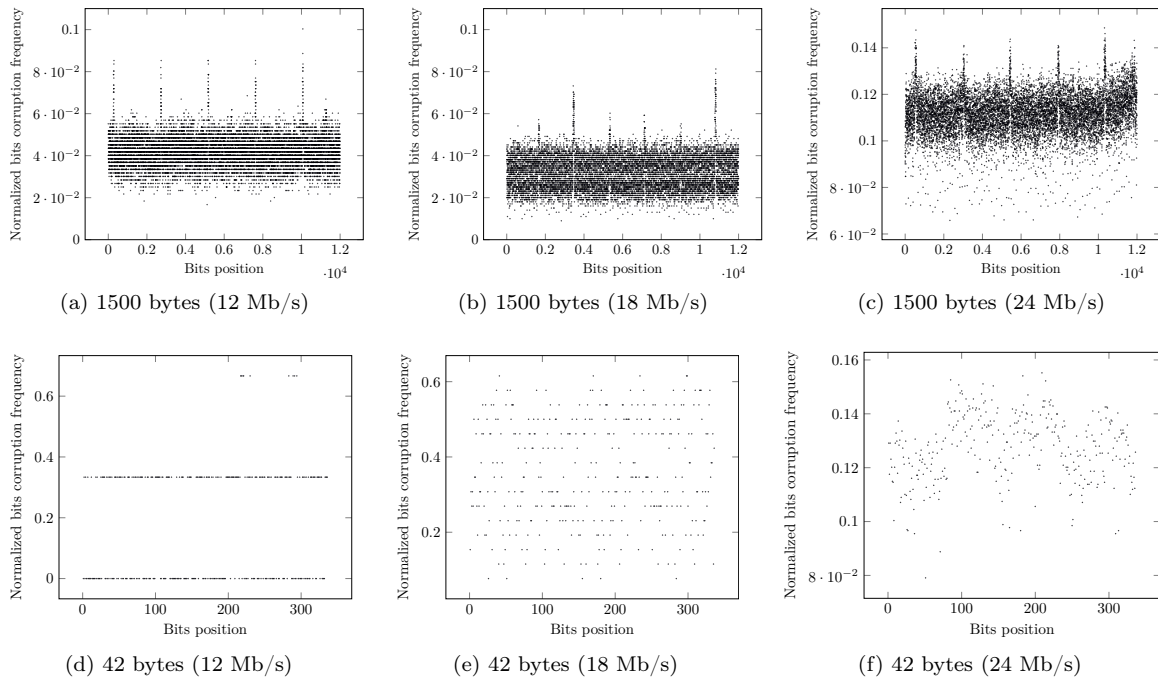


Figure 6: Bit corruption frequency

IV.3. The bits corruption

The investigation of the bit corruption is performed by plotting the normalised bit corruption frequency. Depicted in the figure 6, it is computed for each bit as the ratio between the number of corruption occurrence and the total number of corrupted frames. The bit corruption frequency for the frames with a length of 1500 bytes is clearly not uniform a reveal the presence of peaks, regardless the data rate. These peaks indicate that some bits are more sensible to the corruption, but their positions and their strengths differ between the couple of scenarios. However, the bits position can be incorporated to a pool of bits according to these peaks. For the 18Mb/s scenario, a first removal from the max corrupted bit situated on the last peak (bit position at 10821) to the last bit, the number of corrupted MAC frames decreases by 3.5%. Another removal from the second peak (bit position at 3445) to the last bit, the number of received MAC frames without corruptions increased by 52%. This confirms that a proper reduction of the data length can improve significantly the rate of received MAC frames without corrupted bits. However, such a strategy has a limit, since plots related to the corruption of MAC frames with a length of 42 bytes do not highlight the presence of peaks.

V. Theoretical framework for data rate selection problems

In the previous section, experimental results have shown that a proper selection of the data rate and the data length have a significant impact. However, a theoretical framework is required to determine the hardness of the computation related to the data rate and the data length selection problem under some constraints. In this article, the problem is modelled with a complete k-partite graph, each partite being an occurrence of transmission. Each one shares the same size and represents the number of available data rates. An illustration of the modelled problem is depicted in the figure 7. Two variants of the problem can be considered, one without constraint and another with constraints.

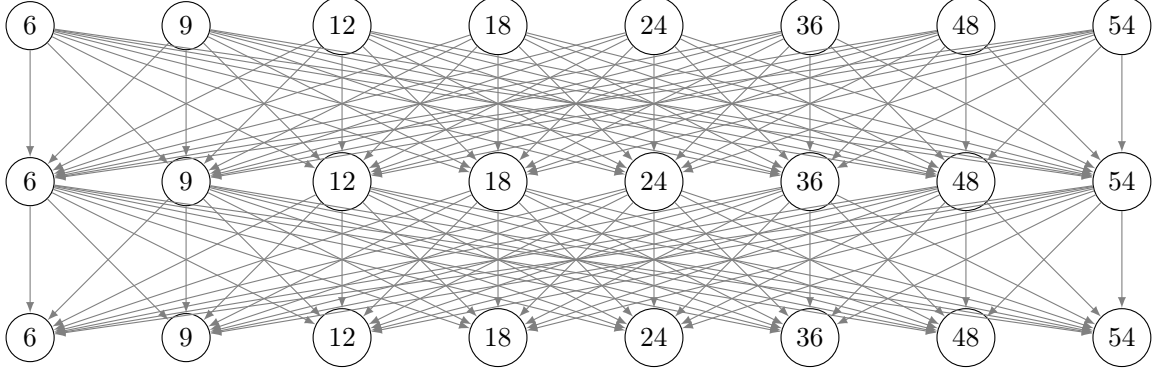


Figure 7: An instance of the data rate selection problem

V.1. Combinatorial optimisation without restriction

Let G a complete k -partite graph, where each partite V_i have the same size n such as $V_1 \cup \dots \cup V_k = V(G)$. A couple of vertices (u, v) are connected by an arc, $A_{u,v} \in A(G)$ if $\forall u \in V_i, \forall v \in V_j, j = i + 1$. Each vertex in the same partite have a weight, $w : v \rightarrow \mathbb{R}^+$, such as $w(v_1) < \dots < w(v_n)$, $\{v_1, \dots, v_n\} \in V_i$. In addition, $\forall u \in V_i, \exists y \in V_j, j = i + 1$ such as $w(u) = w(y)$. The presence (e.g. absence) of an arc is the result of the following function $\Gamma : A \times x \rightarrow [0, 1]$ with $x \in [0, 1]$. Each arc, A , has also a weight such as $w(A_{u,v}) = w(v)$. A path, p , in a the k -partite graph has a weight denoted $w(p)$. This weight is a function of its individual arc is given by a function, such as $f : (w(A_{i,j}), w(A_{j,k}), \dots, w(A_{y,z})) \rightarrow \mathbb{R}^+$. The aim of an algorithm is to find a path p , crossing all partites, which maximises its weight and takes into account the presence (e.g. absence) of its arc denoted A_i , such as :

$$\max w(p) = \sum_{i=1}^k (w(A_i) \times \Gamma(A_i)).$$

Consider now a complete k -partite graph and a Bernoulli random variable, X , used by the Γ function such $\Gamma : A \times X$. Now the time-complexity of an algorithm which aims to find an optimal path⁵ can be expressed.

Theorem V.1. *The time complexity of the maximisation path computation where the presence of arcs is determined by a Bernoulli random variable is $\mathcal{O}(nk)$.*

Proof. The proof is trivial. The algorithm has to find sequentially the best weight of an available arc, however in the worst case, the latest checked has the best weight and is available. \square

V.2. Combinatorial optimisation with transmission restrictions

The previous analysis determines the time complexity of an algorithm that computes the path which has the maximal weight. In practise, an algorithms has to deal with a sequence of bits and has to choose a packet size for each transmission in order to minimise the number of transmissions. Most of the time, the data rate must satisfy a minimal rate to satisfy an application. As a result, the number of transmissions through a path, p is denoted $t(p) \leq Th_t$ with Th_t a given threshold. Beside, the weight of the path $w(p)$ must satisfy $\frac{1}{w(p)} \leq Th_w$ with Th_p a given threshold. We now define the

⁵An optimal path is a path with a maximal weight

Selection of data Rate and data Length problem, denoted *SRL*, which aim to satisfy the couple of constraints on the minimal data rate, Th_w and data length Th_t given by a path p .

Lemma V.2. *SRL* \in **NP**.

Proof. Given a path p with a length denoted $|p|$. The weight of the path is the sum of the weight of its arcs and compared to Th_w . The length of p determines the number of transmissions and can be also compared to Th_t . All of these operations can be done in $\Theta(|p|)$. \square

Lemma V.3. *OPT SRL* is **NP-hard**.

Proof. This proof is based on a reduction to the n additive metrics with LEQ constraints problem detailed in [Wang, 1999]. Let $(G(V, E), d_1(p), d_2(p), Th_1, Th_2)$ an instance of the n additive metrics with LEQ constraint problem with $d_1(p)$ and $d_2(p)$ a couple of additives metrics and their related constraints given by Th_1 and Th_2 . The instance of the *SRL* problem can be performed as following. Each adjacent vertices of V are placed in a different partite. Two nodes are in a same partite if they are adjacent with a same node. Since all partites must have the same given size, dummy vertices are added to satisfy this condition, this part can be done in polynomial time. All edges $e_{(i,j)} \in E$ between two node are split into a couple of arc such as the first metric is the weight of arc $d_1(e_{i,j}) = w(A_{i,j}) = w(A_{j,i})$. The same process is used for the second metric, this part can be done in polynomial time. At the end, all edges with dummy nodes are notified absent with the Γ function, the rest is notified present, this part can be done in polynomial time. To find the path with a minimum number of transmission and maximise the data rate implies to find a path p with the costs $d_1(p)$ and $d_2(p)$ and satisfying the minimum of the couple of thresholds Th_1 and Th_2 . This completes this proof. \square

Theorem V.4. *The SRL problem is NP-complete.*

Proof. According to the couple of lemmas V.2 and V.3, the *SRL* problem is **NP-complete**. \square

VI. Conclusion

In this paper some preliminary results about the impact of the data rate and the length of data have been discussed. They also show that the decrease of the data rate is not the only way to increase the packet reception rate, but the data length reduction can be also considered. Fortunately, the distribution related to the bit corruption is not uniform and some clusters arises. As a result, the design of an algorithm performing on the data length do not have to inspect each bit but clusters to split data. Even if all algorithms which attempts to maximize the packet reception rate dealing the data rate and additional constraints are **NP-complete**, the split of data can be performed according to a pool of bits and not on the bit itself.

This study must be extended to additional experiments both on the length of data and in different scenario to confirm the observations detailed in this article. However, obtained results are sufficiently close to the observations made by Han et al. [Han et al., 2012] to confirm that our data set is not a outlier. Concerning the algorithm part, the study of the optimal strategy opens the perspective to study if c -competitive online algorithms can be designed.

References

- [778, 2016] (2016). IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements - part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pages 1–3534.

- [Bloessl et al., 2013] Bloessl, B., Segata, M., Sommer, C., and Dressler, F. (2013). An IEEE 802.11a/g/p OFDM Receiver for GNU Radio. In *ACM SIGCOMM 2013, 2nd ACM SIGCOMM Workshop of Software Radio Implementation Forum (SRIF 2013)*, pages 9–16, Hong Kong, China.
- [Bourebia et al., 2020] Bourebia, S., Laghmara, H., Hilt, B., Drouhin, F., Bindel, S., Ledy, J., Lauf-fenburger, J., and Lorenz, P. (2020). A belief function-based forecasting link breakage indicator for vanets. *Wirel. Networks*, 26(4):2433–2448.
- [Chevillat et al., 2005] Chevillat, P. R., Jelitto, J., and Truong, H. L. (2005). Dynamic data rate and transmit power adjustment in IEEE 802.11 wireless lans. *Int. J. Wirel. Inf. Networks*, 12(3):123–145.
- [Fernandez et al., 2012] Fernandez, J. A., Borries, K., Cheng, L., Vijaya Kumar, B. V. K., Stan-cil, D. D., and Bai, F. (2012). Performance of the 802.11p physical layer in vehicle-to-vehicle environments. *IEEE Transactions on Vehicular Technology*, 61(1):3–14.
- [Han et al., 2012] Han, B., Ji, L., Lee, S., Bhattacharjee, B., and Miller, R. R. (2012). Are all bits equal?—experimental study of IEEE 802.11 communication bit errors. *IEEE/ACM Transactions on Networking*, 20(6):1695–1706.
- [Wang, 1999] Wang, Z. (1999). On the complexity of quality of service routing. *Information Pro-cessing Letters*, 69(3):111 – 114.
- [Yin et al., 2020] Yin, W., Hu, P., Indulska, J., Portmann, M., and Mao, Y. (2020). Mac-layer rate control for 802.11 networks: a survey. *Wireless Networks*, 26:3793–3830.

A. IEEE 802.11p coding scheme

Table 1 lists all modulations and the coding rates defined by the IEEE 802.11p and the resulting data rates.

Coded bit rate (Mb/s)	Data rate (Mb/s)	Modulation	Coding rate	Coded bits per sub-carrier	Data bits per OFDM symbol
6	3	BPSK	1/2	1	24
9	4.5	BPSK	3/4	1	36
12	6	QPSK	1/2	2	48
18	9	QPSK	3/4	2	72
24	12	16-QAM	1/2	4	96
36	18	16-QAM	3/4	4	144
48	24	64-QAM	2/3	6	192
54	27	64-QAM	3/4	6	216

Table 1: IEEE 802.11p coding scheme