



HAL
open science

Métrologie des réseaux sans-fil: A la frontière du signal et du numérique

Guillaume Kremer

► **To cite this version:**

Guillaume Kremer. Métrologie des réseaux sans-fil: A la frontière du signal et du numérique. Réseaux et télécommunications [cs.NI]. Université Toulouse III Paul Sabatier, 2014. Français. NNT: . tel-01104964

HAL Id: tel-01104964

<https://theses.hal.science/tel-01104964v1>

Submitted on 19 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *20/10/2014* par :
Guillaume Kremer

**Métrieologie des réseaux sans-fil:
À la frontière du signal et du numérique**

JURY

THIERRY GAYRAUD	Président du Jury
ANDRZEJ DUDA	Rapporteur
MARCELO DIAS DE AMORIM	Rapporteur
THIERRY TURLETTI	Examineur
DANIELA DRAGOMIRESCU	Invitée
PHILIPPE OWEZARSKI	Directeur de thèse
PASCAL BERTHOU	Co-encadrant de thèse

École doctorale et spécialité :

MITT : Domaine STIC : Réseaux, Télécoms, Systèmes et Architecture

Unité de Recherche :

Laboratoire d'Analyse et d'Architecture des Systèmes - LAAS-CNRS

Résumé

La multiplication des réseaux sans-fil aura pour conséquence l'augmentation des problèmes d'accès au médium et une forte pollution spectrale. Pour maintenir la qualité des communications, les nœuds devront devenir plus adaptatifs et exploiter un maximum d'informations sur l'état du médium et des communications. Dans cette optique, cette thèse présente 3 contributions ayant respectivement trait à la mesure des réseaux sans-fil, à leur modélisation et à la recherche de solutions d'adaptabilité accessibles pour les nœuds de ces futurs réseaux.

La première contribution concerne la conception et la mise en œuvre d'un banc de mesure expérimental compatible avec les besoins de ces études (*i.e.* fournir des possibilités de mesures inter-couches à partir du niveau physique et un contrôle maximal des perturbations du médium). La solution développée est celle d'un banc conçu à l'intérieur d'une chambre anéchoïque dans laquelle du bruit perturbateur est injecté à l'aide d'une antenne directionnelle. À l'intérieur de la chambre, les mesures sont effectuées à l'aide d'équipements WIFI et d'équipements de mesure RF.

La seconde contribution vise à prendre en compte les mesures issues de ce banc expérimental dans le but de tester et d'améliorer le réalisme du simulateur ns-3 et de ses modèles. En effet, malgré leur peu de réalisme, les simulateurs de réseau comme ns-3 sont utilisés pour tester de nouveaux protocoles ou de nouvelles applications sans-fil. Dans cette optique, nous proposons un modèle d'analyse de cause racine (RCA) conçu pour détecter les différences de configuration, d'implémentation ou de modélisation entre le simulateur et le banc de mesure à partir de données issues de traces expérimentales. L'application de ce modèle a conduit à une amélioration importante du réalisme des modèles WIFI du simulateur.

La troisième contribution consiste à appliquer les algorithmes d'apprentissage SVR, k -nn et DT en vue de l'estimation prédictive du débit IP mesuré sur un lien sans-fil. Les estimations se font respectivement à partir des valeurs de SNR, de RSS et de bruit mesurées au niveau du nœud récepteur. Les différents algorithmes sont évalués selon la précision de leurs estimations mais aussi sur leurs caractéristiques fonctionnelles (*e.g.* taille des modèles, ...). Les résultats indiquent que les algorithmes SVR et DT utilisés avec le SNR permettent les estimations les plus précises. De plus, ces 2 algorithmes offrent les meilleures performances respectivement en termes de mémoire utilisée par le modèle et de temps de calcul.

Mots clé : performance et fiabilité des réseaux sans-fil, apprentissage, instrumentation, métrologie des réseaux sans-fil, adaptabilité des nœuds

Abstract

The multiplication of wireless networks may result in more medium contention and stronger spectral pollution. To maintain the communications quality, nodes will have to be more adaptive and take benefits of the available informations about the state of the medium and of the communications. In that optic, this thesis presents 3 contributions which respectively tackle the following fields of wireless networks research : measurements, modeling and development of new solutions for nodes adaptability.

The first contribution concerns the design and implementation of a measurement test bench which complies with the requirements of these studies (*i.e.* it must allow cross-layer measurements starting from the physical layer and offer maximal control over the environment and its perturbations). Our solution is to perform the communications inside the RF protected environment of an anechoic room. To interfere with these communications, a modulated AWGN noise is injected inside the room with the help of a directive antenna. Inside the room, measurements are made using WIFI devices and RF equipments.

The second contribution aims to take advantage of the experimental platform measurements for improving the realism of ns-3 simulations. In fact, despite a recognized lack of realism, network simulators such as ns-3 are massively used to test new wireless protocols and applications. We then propose a root cause analysis (RCA) method designed for detecting configuration, implementation or modeling anomalies between the simulator and the experimental bench. The application of this method results in a major improvement of the realism of the ns-3 WIFI model.

The third contribution consists in the application of three machine learning algorithms, namely SVR, k-nn and DT, in order to predict the IP throughput on a wireless link. The estimations are based on SNR, RSS and noise measurements at the receiver side. The different algorithms are evaluated according to the accuracy of their estimations but also according to their functional characteristics (*e.g.* model size, training time, ...). The results show that the SVR and DT algorithms used in conjunction of the SNR metric obtain the most accurate estimates. Moreover the SVR and DT algorithms are respectively the most efficient in terms of memory and time.

Keywords : performance and reliability of wireless networks, machine learning, instrumentation, wireless metrology, node adaptability

Table des matières

1	La couche physique des réseaux sans-fil	7
1.1	Introduction	7
1.2	Contraintes physiques et spécificités du médium sans-fil	8
1.2.1	Propagation du signal radio	8
1.2.2	Interférences et bruit	8
1.2.3	Impact du médium sur les piles de communication sans-fil	9
1.3	Couche physique pour les réseaux sans-fil	11
1.3.1	Codage canal et binaire des données	11
1.3.2	Modification de la forme d'onde et filtrage du signal	12
1.3.3	Conversion du signal bande-de-base vers le signal RF	14
1.4	Conclusion	14
2	Le standard IEEE 802.11	17
2.1	Introduction	17
2.2	Couche MAC	18
2.2.1	Architecture d'un réseau 802.11	18
2.2.2	Accès au médium	19
2.2.2.1	Méthode DCF	19
2.2.3	Tramage	23
2.2.3.1	Type de trames	23
2.2.3.2	Format d'une trame de donnée	24
2.3	Couche physique	25
2.3.1	Bandes ISM et interférences	25
2.3.2	Modes de transmission	26
2.3.2.1	Algorithme de choix du mode de transmission	26
2.4	Conclusion	27
3	État de l'art sur les moyens de mesure de réseaux sans-fil	29
3.1	Introduction	29
3.2	Plates-formes de mesure pour réseaux sans-fil	30
3.2.1	Plates-formes de type grille	30
3.2.1.1	Utilisation des plates-formes de type grille	31
3.2.2	Plates-formes en environnements de production	32

3.2.2.1	Utilisation des plates-formes de production	32
3.2.3	Plates-formes de laboratoire	32
3.2.3.1	Utilisation des plates-formes de laboratoire	33
3.3	Simulateurs, validation des modèles et comparaison de performances	33
3.4	Conclusion	35
4	Banc de mesure sans-fil inter-couche pour la mesure RF	37
4.1	Introduction	38
4.2	Contributions	39
4.3	Problématique de la mesure RF	39
4.3.1	Spécificités	39
4.3.2	Environnement de mesure	40
4.3.3	Contrôle des perturbations	40
4.4	Problématiques fonctionnelles associées à la conception d'un banc de test distribué	41
4.4.1	Synchronisation des nœuds	41
4.4.2	Récolte des données et commande du banc de test	41
4.5	Techniques de mesure RF	42
4.5.1	Échantillonnage	42
4.5.2	Filtrage	43
4.5.3	Indicateurs de mesure physique	43
4.5.3.1	RSS	43
4.5.3.2	SNR	44
4.5.3.3	CSI	45
4.5.3.4	EVM	45
4.5.3.5	LQI	46
4.5.3.6	$\frac{E_b}{N_o}$	46
4.5.3.7	ACPR	47
4.5.4	Composants pour la mesure	47
4.5.4.1	Diviseur de puissance	47
4.5.5	Instrumentation	47
4.5.5.1	Oscilloscope LeCroy	47
4.5.5.2	Radio logicielle	49
4.5.5.3	Matériel générique 802.11	51
4.6	Techniques de mesure de trafic et de mesure MAC et réseau	51
4.6.1	Équipements	51
4.6.2	Logiciels pour la mesure et la génération de trafic	52
4.7	Implémentation du banc de test	52
4.7.1	Implémentation de la synchronisation NTP	55
4.7.1.1	Validation	55
4.7.2	Filtrage hors-ligne du signal	57
4.7.3	Impact du diviseur de puissance	57
4.8	Conclusion du chapitre	58

5	Validation expérimentale et amélioration des modèles de simulation ns-3 vis à vis d'un banc de test expérimental	59
5.1	Introduction	60
5.2	Contributions	61
5.3	Description et implémentations du scénario commun pour la comparaison entre traces expérimentales et traces de simulation	62
5.3.1	Mise en œuvre expérimentale	62
5.3.2	Mise en œuvre sous simulateur ns-3	63
5.3.2.1	Injection de bruit dans le modèle YANS	64
5.4	Métriques de comparaison	65
5.4.1	Débit IP	65
5.4.2	Taux de trames erronées au niveau du récepteur	65
5.4.3	Nombre de paquets UDP abandonnés par l'émetteur	66
5.4.4	Nombre de retransmissions nécessaires par datagramme UDP	66
5.4.5	Temps de transmission nécessaire par datagramme UDP	66
5.4.6	Temps d'inter-arrivées des trames 802.11	66
5.5	Méthode d'appairage des traces	67
5.6	Méthodes de comparaison des données	67
5.6.1	Modèle de Gilbert-Elliott pour la caractérisation des motifs d'erreurs	67
5.6.2	Modèle RCA pour la comparaison d'environnements	68
5.6.2.1	Introduction aux modèles d'arbre de comparaison	68
5.6.2.2	Définition d'un arbre pour la comparaison d'environnements	70
5.7	Application de la méthodologie d'association et de comparaison sur les données générées	73
5.7.1	Appairage des traces et validation de la configuration par méthode de Gilbert-Elliott	73
5.7.2	Application du modèle RCA	75
5.8	Analyse des résultats expérimentaux	80
5.8.1	Supplément de diagnostic	80
5.8.2	Modélisation des paramètres BEB	81
5.8.2.1	Facteurs environnementaux	81
5.8.2.2	Régression et calcul des paramètres	82
5.8.2.3	Vérification	84
5.8.3	Améliorations obtenues	84
5.9	Conclusion	85
6	Estimations prédictives de performance d'un lien sans-fil par la mesure RF et l'utilisation d'algorithmes d'apprentissage	89
6.1	Introduction	90
6.2	Contributions	91
6.3	Algorithmes d'apprentissage automatique pour la régression	92

6.3.1	Méthode SVR	92
6.3.2	Méthode k -nn	93
6.3.3	Méthode DT	94
6.4	Données expérimentales	95
6.4.1	Protocole expérimentale pour la génération de données . . .	95
6.4.1.1	Constitution des jeux de données	96
6.4.2	Caractéristiques de mesure de performance et du canal sans fil	97
6.4.2.1	Caractéristique de performance	97
6.4.2.2	Caractéristiques de mesure physiques du canal . . .	97
6.4.3	Notations alternatives des caractéristiques	98
6.5	Méthodes d'évaluation de performances	98
6.5.1	Méthodes d'évaluation de la précision des estimations	98
6.5.1.1	Erreur quadratique absolue	98
6.5.1.2	Pourcentage de 'bonnes estimations'	99
6.5.2	Méthodes d'évaluation de la pertinence des caractéristiques de mesure physique	99
6.5.3	Méthodes d'évaluation de performance des algorithmes . . .	100
6.5.3.1	Efficacité en temps de calcul	100
6.5.3.2	Efficacité mémoire	100
6.6	Méthodologie de mesure et d'apprentissage	100
6.6.1	Configuration des algorithmes de ML	100
6.6.1.1	SVR	100
6.6.1.2	k -NN	101
6.6.1.3	DT	101
6.6.2	Mesure de performance des algorithmes	101
6.6.2.1	Mesure des délais d'apprentissage et d'estimation .	101
6.6.2.2	Mesure de la taille du modèle	101
6.7	Résultats	101
6.7.1	Précision des estimations	101
6.7.2	Pertinence des caractéristiques de mesure RF	105
6.7.3	Fonctionnement des algorithmes	107
6.7.3.1	Efficacité en temps de calcul	107
6.7.3.2	Efficacité mémoire	109
6.8	Conclusion	109

A Graphiques des estimations par apprentissage automatique 133

A.1	Estimations par SVR	135
A.1.1	Estimation de SVR pour la métrique <i>ATH_RSS</i>	135
A.1.2	Estimation de SVR pour la métrique <i>LECR_RSS</i>	136
A.1.3	Estimation de SVR pour la métrique <i>LECR_NOISE</i>	137
A.1.4	Estimation de SVR pour la métrique <i>LECR_SNR</i>	138
A.1.5	Estimation de SVR pour la métrique <i>ATH_RSS+LECR_NOISE</i> .	139
A.1.6	Estimation de SVR pour la métrique <i>LECR_RSS+LECR_NOISE</i> .	140

A.2	Estimations par k -nn	141
A.2.1	Estimation de k -nn pour la métrique ATH_RSS	141
A.2.2	Estimation de k -nn pour la métrique $LECR_RSS$	142
A.2.3	Estimation de k -nn pour la métrique $LECR_NOISE$	143
A.2.4	Estimation de k -nn pour la métrique $LECR_SNR$	144
A.2.5	Estimation de k -nn pour la métrique $ATH_RSS+LECR_NOISE$	145
A.2.6	Estimation de k -nn pour la métrique $LECR_RSS+LECR_NOISE$	146
A.3	Estimations par DT	147
A.3.1	Estimation de DT pour la métrique ATH_RSS	147
A.3.2	Estimation de DT pour la métrique $LECR_RSS$	148
A.3.3	Estimation de DT pour la métrique $LECR_NOISE$	149
A.3.4	Estimation de DT pour la métrique $LECR_SNR$	150
A.3.5	Estimation de DT pour la métrique $ATH_RSS+LECR_NOISE$	151
A.3.6	Estimation de DT pour la métrique $LECR_RSS+LECR_NOISE$	152

Table des figures

1.1	Mesures de performance des couches application 1.1(a), réseau 1.1(b) et physique 1.1(c) sans-fil lors de la diffusion d'une application vidéo. Graphiques extraits de [72].	10
1.2	Chaînes de transmission analogique/numérique.	12
1.3	Une représentation graphique des codages binaires est celle des diagrammes de constellation dans laquelle chaque point représente un symbole c_k . Les diagrammes de constellation théoriques des schémas de modulation numérique à 2 et à 4 états sont respectivement présentés sur les figures 1.3(a) et 1.3(b) (points rouges). Lors de la phase de réception d'un signal, les symboles sont examinés. Ces symboles peuvent avoir été altérés. Le récepteur estime alors le point de la constellation idéale le plus proche du symbole reçu. Un exemple est donné sur la figure 1.3(a) dans lequel un symbole S est reçu déformé. Dans ce cas, le récepteur choisira la valeur 0 lors du décodage du symbol.	13
1.4	Exemples de signaux bande-de-base et passe-bande.	14
2.1	Organisation des méthodes d'accès 802.11. Figure reproduite de [42].	19
2.2	Séquence d'émission DCF. Figure extraite de [42].	20
2.3	Mécanisme de réservation RTS/CTS. Figure extraite de [42].	22
2.4	Valeurs prises par CW en fonction du numéro de retransmission [9]. À partir de la 7 ^{ème} retransmission, le nombre de créneaux ne varie plus.	23
2.5	Format d'une trame de donnée 802.11.	24
2.6	Représentation des différentes bandes ISM de fréquence inférieure à 6 GHz.	26
4.1	La chambre anéchoïque du LAAS-CNRS nous permet d'effectuer des mesures dans un environnement protégé, <i>i.e.</i> sans les perturbations physiques naturelles qui existent dans un milieu non-protégé.	40
4.2	Illustration des différentes représentations du signal.	43
4.3	Le point A est le signal du diagramme constellation d'origine. Le point B est le signal reçu. Le vecteur d'erreur se calcule comme la différence des coordonnées des deux signaux.	46

4.4	Schématique d'un diviseur de signal 2 voies.	47
4.5	Utilisations d'un diviseur de puissance.	48
4.6	Équipements utilisés pour la mesure physique.	49
4.7	Description complète du banc de test et des différents réseaux utilisés. La notation $\{a_x\}$ correspond à l'ensemble des interfaces de l'équipement A . La notation $a \longleftrightarrow b$ signifie qu'il existe une connexion entre l'interface a et l'interface b . Pour ne pas surcharger les boîtiers, le trafic est généré sur la machine $M1$ et réceptionné sur la machine $M2$. Les boîtiers $P1$ et $P2$ sont donc paramétrés comme des ponts entre leur connexion filaire et leur connexion sans-fil. Les nœuds $S1$ et $S2$ sont les nœuds écouteurs facultatifs mis en place pour ne pas biaiser les mesures. L'ensemble des équipements est commandé par l'intermédiaire du réseau de contrôle et d'un logiciel situé sur la machine $M1$	54
4.8	Montage expérimental permettant la validation de la synchronisation NTP utilisée sur le banc de test.	55
4.9	Différence entre les estampillages des requêtes ICMP capturées par les boîtiers B et C dans les cas respectifs d'une synchronisation unique au démarrage de l'expérience (a) et d'une synchronisation répétée chaque seconde (b).	56
4.10	Montage expérimental utilisé pour la caractérisation des sorties du diviseur de signal.	57
5.1	Disposition des équipements pour la mesures expérimentales.	63
5.2	Modèle de Gilbert-Elliott.	67
5.3	Exemple d'arbre de comparaison. Les valeurs V indiquées sur les nœuds sont les valeurs évaluées lors du parcours de l'arbre.	69
5.4	Arbre de déduction défini pour la comparaison de traces. La notation $M(a)$ utilisée pour évaluer la valeur des noeuds constitue le médian pour la métrique a (ex : $M(BW)$ est le médian des valeurs de débit mesurées chaque seconde pour chacune des traces).	71
5.5	Mesures de FER en fonction des différents niveaux de bruits appliqués expérimentalement 5.5(a) et en simulations 5.5(b).	73
5.6	Profils des mesures de FER après association et différences absolues mesurées entre les FER des traces associées.	74
5.7	Evolutions des différents coefficients de Gilbert-Elliott mesurés en simulation et expérimentalement	75
5.8	Comparaisons des données de simulation et expérimentales pour les métriques BW (a), $DROP$ (b) et TT (c).	77
5.9	Comparaisons des données de simulation et expérimentales pour les métriques TI (a) et NR (b).	78

5.10	Arbre de déduction pour la comparaison de traces avec indication de parcours lors de l'évaluation des nœuds. Les nœuds non barrés sont les nœuds évalués comme vrais, les nœuds barrés par des lignes entières sont évalués comme faux et les nœuds barrés par des tirets ne sont pas évalués pendant le parcours de l'arbre.	79
5.11	Valeurs de TIR mesurées expérimentalement et en simulation.	80
5.12	Sommes cumulées et maximum des différences entre les débits mesurés en simulation et en expérimentation en fonction de la taille maximale de la fenêtre de contention Cw_{max}	83
5.13	Résultats des débits obtenus avec les valeurs optimales.	85
5.14	Différences observées entre les débits obtenus avec les différentes configurations d'algorithmes de temporisation et le débit expérimental.	86
6.1	Exemple d'arbre de décision.	94
6.2	Estimations obtenues avec SVR et le jeu d'entraînement <i>Dataset1</i> pour les caractéristiques <i>LECR_SNR (C4)</i> (figure 6.2(a)) et <i>ATH_RSS (C1)</i> (figure 6.2(b)).	102
6.3	Précision moyenne des estimations obtenues pour les différents algorithmes de ML, en fonction des différents jeux d'apprentissage. L'unité de précision est ici le MSE.	103
6.4	Précision moyenne des estimations obtenues pour les différents algorithmes de ML, en fonction des différents jeux d'apprentissage. L'unité de précision est ici le pourcentage de bonnes estimations $P(e < 1Mbps)$	104
6.5	Pertinence des différentes métriques pour l'estimation de débit sur un lien sans-fil vis-à-vis des différents algorithmes de ML. Le classement est exprimé en fonction des résultats mesurés en MSE.	106
6.6	Pertinence des différentes métriques pour l'estimation de débit sur un lien sans-fil vis-à-vis des différents algorithmes de ML. Le classement est exprimé en fonction des résultats mesurés en $P(e < 1Mbps)$	107
6.7	Indices de pertinence calculés selon le critère d'impureté par l'algorithme CART. Les indices sont normalisés, <i>i.e</i> pour chaque jeu d'apprentissage, la somme des indices est égale à 1.	108
6.8	Temps d'apprentissage et d'estimation des différents algorithmes SVR, <i>k</i> -nn et DT pour les différentes métriques.	111
6.9	Performances de calcul des différents algorithmes SVR, <i>k</i> -nn et DT pour les différentes caractéristiques.	112
6.10	Taille moyenne des modèles générés et facteurs multiplicatifs par rapport aux données brutes pour les différents algorithmes et pour chaque jeux d'apprentissage.	113
A.1	Graphiques des estimations de SVR pour la métrique <i>ATH_RSS</i>	135
A.2	Graphiques des estimations de SVR pour la métrique <i>LECR_RSS</i>	136

A.3	Graphiques des estimations de SVR pour la métrique <i>LECR_NOISE</i> .	137
A.4	Graphiques des estimations de SVR pour la métrique <i>LECR_SNR</i> .	138
A.5	Graphiques des estimations de SVR pour la métrique <i>ATH_RSS + LECR_NOISE</i> .	139
A.6	Graphiques des estimations de SVR pour la métrique <i>LECR_RSS + LECR_NOISE</i> .	140
A.7	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>ATH_RSS</i> .	141
A.8	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>LECR_RSS</i> .	142
A.9	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>LECR_NOISE</i> .	143
A.10	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>LECR_SNR</i> .	144
A.11	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>ATH_RSS + LECR_NOISE</i> .	145
A.12	Graphiques des estimations de <i>k</i> -nn pour la métrique <i>LECR_RSS + LECR_NOISE</i> .	146
A.13	Graphiques des estimations de DT pour la métrique <i>ATH_RSS</i> .	147
A.14	Graphiques des estimations de DT pour la métrique <i>LECR_RSS</i> .	148
A.15	Graphiques des estimations de DT pour la métrique <i>LECR_NOISE</i> .	149
A.16	Graphiques des estimations de DT pour la métrique <i>LECR_SNR</i> .	150
A.17	Graphiques des estimations de DT pour la métrique <i>ATH_RSS + LECR_NOISE</i> .	151
A.18	Graphiques des estimations de DT pour la métrique <i>LECR_RSS + LECR_NOISE</i> .	152

Liste des tableaux

1.1	Caractéristiques des schémas de modulation disponibles pour <i>802.11a,g</i> . 13
4.1	Aperçu des données récupérables à l'aide de l'analyseur National Instruments. 50
4.2	Exemples de données de niveau physique disponibles avec le pilote ath5k. 51
4.3	Exemples de données de niveau MAC disponibles avec le pilote ath5k. 52
4.4	Champs disponibles dans les trames 802.11 capturées par l'outil TCPDump. 53
4.5	Résultats des différences de synchronisation dans les deux cas de test. 56
5.1	Paramètres utilisés dans les scénarios et expérimentations. 62
5.2	Notation après le processus d'appairage des traces expérimentales et de simulation. Pour une question de commodité, seule une partie des associations est présentée. 73
5.3	Évaluation logique des différents nœuds de l'arbre de déduction. . . 76
5.4	Valeurs des coefficients et des fenêtres de contention pour les différents niveaux de retransmission. 82
5.5	Meilleurs résultats obtenus pour la recherche du paramètre k optimum. 83
5.6	Valeurs de k et m utilisées pour les 5 meilleurs résultats des tests. La valeur entre parenthèses est le classement du paramètre k dans la section précédente. 84
6.1	Constitutions et caractéristiques des jeux d'apprentissage (a) et jeu de test (b). Chaque vecteur représente 1 seconde de mesure. 96
6.2	Notations alternatives des différentes caractéristiques d'apprentis- sage automatique. 98

Glossaire

Symboles

4G 4^e Génération (mobile).

5G 5^e Génération (mobile).

A

ACK Acquittement dans le vocable 802.11.

ACPR *Adjacent Channel Power Ratio*, traduction française *Rapport de puissance émise du canal adjacent*.

ANR *Agence Nationale de la Recherche*.

AP *Access Point*, traduction française *Point d'accès dans le vocable 802.11*.

AWGN *Additive White Gaussian Noise*, traduction française *Bruit blanc gaussien additif*.

B

BEB *Binary Exponential Backoff*, traduction française *Temporisation exponentielle binaire*.

BPSK *Binary Phase Shift Keying*.

BSS *Basic Service Set*.

C

CART *Classification And Regression Trees*, traduction française *Arbres de classification et régression*.

CDMA *Code division multiple access*, traduction française *Accès multiple par répartition en code*.

CSI *Channel State Information*, traduction française *Information d'état du canal*.

CSMA/CA *Carrier Sense Multiple Access with Collision Avoidance*, traduction française *Accès multiple par écoute de porteuse avec évitement de collisions*.

CTS *Clear To Send (802.11)*.

CW *Contention Window*.

CW_{max} Valeur maximale de la fenêtre de contention MAC 802.11.

D

dBm *Décibel.*

DCF *Distributed Coordination Function*, traduction française *Fonction de coordination distribuée.*

DIFS *DCF InterFrame Space.*

DSSS *Direct-Sequence Spread Spectrum*, traduction française *Spectre étalé en séquence directe.*

DT *Decision Tree*, traduction française *Arbre de décision.*

E

EIFS *Extended InterFrame Space.*

EVM *Error Vector magnitude*, traduction française *Magnitude du vecteur d'erreur.*

F

FHSS *Frequency-Hopping Spread Spectrum.*

FSK *Frequency Shift Keying.*

G

GHz *Gigahertz.*

GPS *Global Positioning System.*

GS *GigaSample.*

GSM *Global System for Mobile Communications.*

H

HCF *Hybrid Contention Function*, traduction française *Fonction de contention hybride.*

I

ICMP *Internet Control Message Protocol.*

IEEE *Institute of Electrical and Electronics Engineers.*

IFS *InterFrame Space.*

IoT *Internet of Things*, traduction française *Internet des objets.*

ISM *Bandes Industrielles, Scientifiques et Médicales.* Bandes de fréquences libres.

K

K-nn *k-nearest neighbors*, traduction française *Méthode des k plus proches voisins*.

L

LQI *Link Quality Indication*.

LTE *Long Term Evolution*.

M

MBPS *MégaBits Par Seconde*.

MHz *MégaHertz*.

MIMO *Multiple Input Multiple Output*.

ML *Machine Learning*, traduction française *Apprentissage automatique*.

MS *MégaSample*.

MSE *Mean Squared Error*, traduction française *Erreur quadratique absolue*.

N

NAV *Network Allocation Vector*.

NTP *Network Time Protocol*.

O

OFDM *Orthogonal Frequency Division Multiplexing*, traduction française *Multiplexage par répartition en fréquences orthogonales*.

OSI *Open Systems Interconnection*, traduction française *Interconnexion des Systèmes Ouverts*.

P

PCF *Point Coordination Function*, traduction française *Fonction de coordination centralisée*.

PHY *Couche physique*.

PIFS *PCF InterFrame Space*.

PTP *Precise Time Protocol*.

Q

QAM *Quadrature Amplitude Modulation*.

QoS *Quality of Service*, traduction française *Qualité de Service*.

QPSK *Quadrature Phase Shift Keying*.

R

RCA *Root Cause Analysis*, traduction française *Analyse de cause racine*.

RESCUE *RESeau Coordonné de sUbstitution mobilE.*

RF *Radio Frequency*, traduction française *Fréquence radio.*

Roaming *Itinérance.*

RSS *Received Signal Strength*, traduction française *Puissance du signal reçu.*

RSSI *Received Signal Strength Indicator*, traduction française *Indicateur de puissance du signal reçu.*

RTS *Request To Send (802.11).*

S

Seqctl Nom du champ contenant le numéro de séquence en 802.11.

SIFS *Short InterFrame Space.*

SISO *Single Input Single Output.*

Smartphone *Téléphone intelligent.*

SNIR *Signal to Noise Ratio plus Interferences*, traduction française *Ratio signal sur bruit plus interférences.*

SNR *Signal to Noise Ratio*, traduction française *Ratio signal sur bruit.*

STA *Station.*

SVR *Support Vector Regression*, traduction française *Régression à support vecteur.*

T

TCP *Transmission Control Protocol.*

U

UDP *User Datagram Protocol.*

UMTS *Universal Mobile Telecommunications System.*

V

V2I *Vehicle-to-infrastructure.*

V2V *Vehicle-to-vehicle.*

W

WIMAX *Worldwide Interoperability for Microwave Access.*

WLAN *Wireless Local Area Network*, traduction française *Réseau local sans-fil.*

WPAN *Wireless Personal Area Network*, traduction française *Réseau sans-fil personnel.*

Introduction

Les réseaux sans-fil sont aujourd'hui omniprésents dans nos vies et leur utilisation fait preuve d'un intérêt en constante progression. Dans les années 1990, cet engouement s'est manifesté par le déploiement des technologies mobiles et la distribution massive de téléphones cellulaires (entre 1990 et 2006, le taux de pénétration de la téléphonie mobile aux USA est passé de 16 à 90 % [34]). Aujourd'hui, les réseaux WLAN (*Wireless Local Area Networks*) et mobiles sont les principales technologies d'accès au réseau Internet pour les utilisateurs de tablettes et autres téléphones intelligents (du terme anglais *smartphones*). Ces deux catégories constituent en effet la majeure partie des terminaux reliés au réseau Internet avec près de 4 millions d'équipements [11]. Avec la convergence des technologies, ces équipements mobiles peuvent maintenant disposer de plusieurs moyens de communications pour accéder aux principaux réseaux qui constituent le paysage sans-fil actuel : les réseaux mobiles, WLAN et WPAN (*Wireless Personal Area Networks*).

Les réseaux mobiles (GSM [83], cdmaOne [121], UMTS [56], cdma2000 [131], LTE [107]) ont évolué à partir des années 1990 d'un fonctionnement analogique dédié à la voix vers un réseau numérique pour le transfert de paquets. La 4^e génération (*e.g.* LTE) de technologies mobiles est actuellement en œuvre et permet un débit maximal descendant théorique de 300 Mbps. Les technologies mobiles fonctionnent sur les bandes de fréquences¹ 850, 900, 1800 et 1900 MHz.

Les réseaux WLAN (IEEE 802.11 [6]) ont été conçus pour remplacer les systèmes filaires dans les réseaux traditionnels (entreprise, domicile, usine, ...) et permettre aux terminaux mobiles ou fixes un accès à l'infrastructure IP. Les réseaux WLAN sont principalement structurés en cellules de quelques dizaines de mètres de diamètres à l'intérieur desquelles les communications entre les nœuds sont possibles. La communication à l'extérieur de la cellule se fait par l'intermédiaire de points d'accès. La principale technologie utilisée dans les WLAN est le standard 802.11 qui utilise les bandes ISM (*Bandes Industrielles, Scientifiques et Médicales*) de 2.4 GHz dans ses versions 802.11b,g et 5 GHz dans la version 802.11a. Aujourd'hui l'usage conjugué des réseaux WLAN à l'intérieur et des réseaux mobiles à l'extérieur permet d'expérimenter l'itinérance globale pressentie à la fin des années 1990, *i.e.* la capacité d'être constamment connecté malgré les déplacements et les voyages.

Les réseaux WPAN (IEEE 802.15.1 [26], IEEE 802.15.4 [7]) sont utilisés comme

1. L'utilisation de ces bandes de fréquences peut-être variable selon les pays et les technologies.

réseaux d'interconnexion pour équipements électroniques portatifs. Leur portée est donc limitée (environ 10 mètres) et concentrée autour du porteur (cellules *picocells*). Le débit des réseaux WPAN est généralement faible (< 24 Mbps) pour économiser l'énergie qui est une contrainte principale des nœuds. Les réseaux WPAN ont été normalisés par le groupe de standards IEEE 802.15 qui définit également les principales technologies utilisées : le Bluetooth IEEE 802.15.1 et IEEE 802.15.4 qui constitue la base de la pile de communication ZigBee [39]. Par conséquent, la plupart des communications WPAN se font sur les bandes ISM de 2.4 GHz où ces réseaux doivent cohabiter avec les réseaux WLAN.

Dans un futur proche, ces technologies sans-fil pourraient constituer les moyens de communication de base des réseaux ubiquitaires et de l'Internet des Objets (traduit de l'anglais *Internet of Things - IoT*) [17].

Les réseaux ubiquitaires organisent les interactions entre le monde physique et les réseaux informatiques. Ils sont constitués de nœuds immergés dans l'environnement ambiant et communiquant entre eux pour accomplir des tâches coopératives. Le résultat de ces tâches peut être invisible aux utilisateurs [130].

Selon Cisco, ces réseaux constitués de centaines de nœuds pourraient générer 3.9 exaoctets de données en 2017 soit 3 % du trafic IP global prévu [40]. Ces réseaux très denses utilisent des technologies sans-fil hétérogènes (802.11b, 802.11g, Bluetooth, ...) qui cohabitent sur les mêmes bandes de fréquences ISM. Ils posent le problème du partage du médium [13] et de la contention spectrale [51]. Le médium est en effet une ressource limitée où les interférences et le bruit sont nocifs pour la qualité des transmissions.

Afin de gérer ces problèmes et ceux du médium d'une manière générale, les réseaux sans-fil du futur devront s'appuyer sur les capacités d'adaptabilité des nœuds. Les travaux actuels traitant des radios intelligentes et adaptatives (radios cognitives [13] et logicielles [99]) ne sont pas suffisants si l'on considère les différents résultats de la littérature : fiabilité discutable [50], modèles radio insuffisants [13], contraintes trop importantes sur le réseau [74]. Concernant la fiabilité des communications, des études sont réalisées pour mieux comprendre les dégradations du signal transmis [90, 75]. En revanche, dans les futurs réseaux IoT, traiter les problèmes de ce point de vue uni-couche (*e.g.* en s'occupant uniquement de la couche radio) ne sera pas suffisant. Afin d'être représentatif, le point de vue de la pile de communication devra se placer à de multiples niveaux (point de vue *inter-couche*, traduction française de *cross-layer*). Les piles protocolaires utilisées dans les réseaux sans-fil définissent en effet les méthodes de transfert de données à la fois localement sur le réseaux sans-fil (couches PHY et MAC) et globalement sur l'infrastructure IP. La dynamique de ces réseaux et leurs besoins sont différents. La conséquence de cette ambivalence est mal connue : le résultat du trafic IP au niveau des couches basses pourrait par exemple interférer avec le médium. De plus, la connaissance exacte de la répercussion du trafic généré par les protocoles de la

couche transport ou de la couche application et de leur forme respective (*e.g.* trafic streaming ou élastique [85]) est faible, ce qui rend l'amélioration de l'utilisation du médium impossible pour l'instant à ces niveaux de la pile. Les besoins en termes d'adaptabilité et de fiabilisation des nœuds évoqués précédemment passeront donc aussi par une meilleure compréhension des liaisons entre couches hautes, moyennes et basses dans les piles de communications sans-fil.

La solution d'adaptabilité proposée et mise en œuvre dans le projet ANR RESCUE² repose sur l'utilisation de la mobilité des nœuds afin de les positionner pour améliorer la qualité des communications. Le projet s'intéresse au déploiement des réseaux de substitution mobiles qui sont des réseaux de secours constitués de nœuds sans-fil ayant des capacités d'infrastructure réseau (*e.g.* des routeurs sans-fil) et rapidement déployables. Ces réseaux sont entre autres utilisés pour augmenter ou rétablir la résilience ou la qualité de service d'un réseau de base en se déployant après ou en prévision d'un accident réseau. Pour parvenir au déploiement de ces nouveaux réseaux, RESCUE définit deux objectifs principaux qui rejoignent les problématiques décrites précédemment dans cette introduction. Le 1^{er} objectif consiste d'une part, à développer de nouveaux algorithmes et protocoles pour gérer les communications entre les nœuds, et d'autre part, à définir de nouvelles procédures pour le déploiement ou le redéploiement du réseau de substitution. Le 2^e objectif consiste à surveiller le trafic et les différents indices de qualité du réseau et du signal. Sur le réseau de base, cette surveillance permettra le déploiement du réseau de substitution si des motifs suspects sont détectés (*e.g.* baisse de trafic, pertes importantes localisées, brouillage, ...). De plus et étant donné que la couverture et la qualité du réseau de substitution sans-fil seront en partie déterminées par le placement des nœuds, l'analyse des tous les indices et notamment des indicateurs localisés dans les couches protocolaires (MAC, PHY et RF) des réseaux sans-fil permettra aux nœuds mobiles de trouver la meilleure position et de s'y déployer afin d'améliorer la qualité de service. Pour ce dernier objectif, la connaissance des liaisons qui existent entre les différents niveaux de la pile de communication sans-fil depuis le niveau physique sera nécessaire.

La transition des réseaux sans-fil vers l'IoT nécessite donc de développer de nouvelles approches pour améliorer la gestion du médium sans-fil.

D'abord en termes de mesure et d'expérimentation, la compréhension du médium et l'analyse de la couche physique nécessitera d'une part de disposer de plateformes offrant un contrôle important de l'environnement et des perturbations. De plus, ces mesures devront descendre jusqu'au niveau physique des communications. Les bancs de type grille comme Senslab [30] et Orbit [100] sont efficaces pour tester de nouveaux protocoles et de nouvelles applications, en revanche ils ne permettent pas un contrôle précis de l'environnement radio. Le même constat peut-être tiré des bancs conçus dans des environnements opérationnels (*e.g.* réseaux d'université,

2. le projet RESCUE du programme VERSO a financé cette thèse.

centres commerciaux, ...) et qui malgré leur réalisme ne permettent pas d'effectuer des mesures de précision [57, 62].

Ensuite, en termes de modélisation : les simulateurs de réseaux sans-fil utilisés pour le test et la mise en œuvre de nouveaux protocoles et applications sont critiqués pour leur manque de réalisme [54, 65]. Afin de les améliorer, des études comparant les résultats des simulations et des expérimentations ont été utilisées [132, 22]. Ces études sont aujourd'hui dépassées par le nombre et la complexité des simulateurs ainsi que par la diversité des équipements sans-fil [59, 18]. Par conséquent, elles ne sont plus réalistes. En conclusion, les résultats de ces études comparatives peuvent dégrader la perception et le réalisme des simulateurs et de leurs modèles en fournissant des résultats contradictoires comme [33] et [76]. De nouvelles méthodologies plus fiables seront donc nécessaires pour permettre des avancées concernant le développement de nouvelles applications et de nouveaux protocoles.

Enfin en termes d'outils pour l'adaptabilité des nœuds, les solutions actuelles ne sont pas toujours réalisables sur les équipements courants (*e.g.* nécessité d'utiliser des radios cognitives [47]). De plus, la fiabilité de ces solutions est souvent mise en doute comme c'est le cas pour les algorithmes de contrôle du mode d'émission utilisés en 802.11 et qui sont parfois jugés défaillants dans les environnements bruités ou à fort niveau d'interférences [117, 82]. Afin de rendre ces outils plus efficaces, de nouvelles solutions d'adaptabilité plus robustes et qui prennent mieux en compte les spécificités du médium sont nécessaires. Les algorithmes d'apprentissage ont fait leur preuve dans de nombreux domaines [20, 14, 28]. En revanche et à notre connaissance, il n'existe pas de travaux qui se sont réellement penchés sur leur application sur du matériel de communication générique (*e.g.* carte radio 802.11) et peu ont réellement étudiés le coût de telles applications (*e.g.* en termes de mémoire ou de calcul). Pourtant, l'application de ces méthodes pourraient améliorer les possibilités d'adaptation des nœuds car elles permettent par l'intermédiaire de leur processus d'apprentissage de découvrir la technique d'adaptation la plus efficace selon les caractéristiques du lien de communication [35]. En tenant compte des récentes études qui caractérisent les relations entre performance et mesures effectuées au niveau des différentes couches de la pile de communications sans-fil et notamment au niveau de la couche physique [24, 108], les futurs outils pourraient être capables de fournir un niveau d'adaptabilité plus important. Par conséquent, afin de rendre l'application de ces méthodes d'apprentissage possible, des études sont nécessaires pour tester la précision, la fiabilité et la faisabilité de ces solutions sur du matériel standard 802.11.

Ces trois points seront l'objet des contributions développées parmi les 6 chapitres de cette thèse.

Le chapitre 1 abordera dans une première partie les difficultés posées par l'utilisation du médium sans-fil. Celui-ci soumet en effet le signal transmis à des pertur-

bations physiques qui peuvent le détériorer et à des interférences qui en perturbent la réception. La deuxième partie du chapitre traitera des différents moyens mis en œuvre par la couche physique pour prendre en charge ces difficultés. Ce chapitre servira ainsi de chapitre introductif pour la compréhension du fonctionnement des équipements de communication numérique utilisés tout au long de cette thèse.

Le chapitre 2 présentera la technologie IEEE 802.11 qui sera prise en exemple dans les différentes contributions de cette thèse. Ce chapitre détaillera entre autres les différents standards de couche physique et MAC ainsi que les mécanismes de gestion comme la sélection automatique du mode de transmission utilisée pour maintenir la qualité des transmissions.

Le chapitre 3 présentera un état de l'art des différentes plates-formes d'instrumentation actuelles pour réseaux sans-fil ainsi que des travaux comparatifs de simulateurs et bancs expérimentaux.

La première contribution de cette thèse sera présentée en chapitre 4. Elle consistera à concevoir et à mettre en œuvre un banc de mesure inter-couche permettant de mesurer l'ensemble de la pile de communication sans-fil et en particulier les couches sans-fil (PHY et MAC). La première partie de cette contribution sera une étude théorique des besoins du banc expérimentale et des différents moyens disponibles (équipements, logiciels, ...) tandis que la seconde partie sera la mise en œuvre de la plate-forme.

Le chapitre 5 présentera la deuxième contribution. Il s'agira de l'amélioration et de la validation des différents modèles WIFI aux niveaux des couches MAC et physique mises en œuvre dans le simulateur ns-3. Malgré les critiques qui sont adressées aux simulateurs concernant leur manque de réalisme, ceux-ci sont encore couramment utilisés par la communauté pour tester de nouveaux protocoles ou de nouvelles applications sans-fil. Par conséquent, afin d'améliorer le réalisme des modèles, des méthodes empiriques de validation des simulateurs sont parfois mises en œuvre pour comparer les résultats des simulations avec ceux des réseaux réels. Dans cette optique, cette contribution proposera un modèle d'analyse par cause racine (RCA) permettant la comparaison de mesures issues de traces respectivement obtenues sur le simulateur ns-3 et sur le banc expérimental décrit au chapitre 4.

Le chapitre 6 détaillera la troisième contribution de la thèse et concernera donc l'utilisation de méthodes d'apprentissage pour l'estimation prédictive du débit IP d'un lien sans-fil à partir de métriques de niveau physique. Celles-ci sont choisies pour être disponibles facilement sur du matériel 802.11 standard à la réception d'une trame. Les données utilisées sont obtenues expérimentalement à partir du banc de mesure. Pour juger de la qualité des estimations, l'erreur quadratique absolue (MSE) ainsi que le pourcentage de bonnes estimations sont utilisés. Additionnellement et étant donné que ces algorithmes sont destinés à être mis en œuvre sur du matériel générique possiblement limité en termes de ressources, les caractéristiques fonctionnelles des algorithmes comme la taille des modèles ou les

temps d'apprentissage et d'estimation sont également étudiées. Enfin, la pertinence de chacune des métriques de niveau physique est analysée en fonction des résultats obtenus avec les différents algorithmes. Pour une question de places, les différentes estimations obtenues seront affichées en annexe A.

Finalement, le manuscrit fera le bilan de la thèse et abordera les différentes perspectives offertes par les résultats obtenus.

1

La couche physique des réseaux sans-fil

Sommaire

1.1	Introduction	7
1.2	Contraintes physiques et spécificités du médium sans-fil	8
1.2.1	Propagation du signal radio	8
1.2.2	Interférences et bruit	8
1.2.3	Impact du médium sur les piles de communication sans-fil	9
1.3	Couche physique pour les réseaux sans-fil	11
1.3.1	Codage canal et binaire des données	11
1.3.2	Modification de la forme d'onde et filtrage du signal	12
1.3.3	Conversion du signal bande-de-base vers le signal RF	14
1.4	Conclusion	14

1.1 Introduction

Le niveau physique des réseaux sans-fil a un impact déterminant sur la qualité des communications. En effet, le médium sans-fil soumet le signal à des phénomènes physiques qui peuvent l'endommager et peut-être affecter la qualité des communications. Pour prendre en compte cette complexité, la couche physique met en œuvre diverses techniques et mécanismes pour rendre le signal plus robuste. L'objectif de ce chapitre est dans un premier temps de présenter de manière succincte les difficultés posées par l'utilisation du médium sans-fil, et dans un deuxième temps d'introduire le modèle de couche physique pour la transmission numérique. L'introduction de celui-ci sera ensuite utile pour la conception du processus de mesure métrologique prenant place à ce niveau de la pile de communication. Ces deux parties seront respectivement détaillées en section 1.2 et 1.3.

1.2 Contraintes physiques et spécificités du médium sans-fil

1.2.1 Propagation du signal radio

La propagation de signaux radio et ses phénomènes sont abondamment décrits dans la littérature [89, 97, 24]. Deux types de facteurs sont essentiellement considérés comme impactant un signal radio : les facteurs environnementaux et les facteurs matériels [125, 24]. Seuls les facteurs environnementaux sont abordés dans cette section. Parmi eux :

- Les *pertes par propagation* sont les résultats de l'atténuation du signal émis due à la distance parcourue.
- Le *multi-trajet* ou l'*affaiblissement* se produisent lorsqu'un signal n'emprunte plus un trajet unique mais plusieurs trajets simultanément. La vitesse de propagation du signal dans l'air ne varie pas, par conséquent, si les trajets sont de tailles différentes, le récepteur reçoit le même signal à des instants différents, les signaux reçus se superposent et des erreurs apparaissent.
- Le *masquage* est l'obstruction des zones de couverture d'un signal RF par des obstacles de dimension importante.

La *réflexion*, la *diffraction* et l'*éparpillement* sont trois phénomènes d'affaiblissement du signal responsables du phénomène de multi-trajet notamment. La *réflexion* se produit quand un signal traverse la frontière entre deux milieux de nature différente. Une partie du signal se reflète alors dans le milieu d'origine tandis que le reste du signal continue à se propager. La valeur de la réflexion dépend de la fréquence du signal et de la nature du milieu. On constate par exemple l'apparition de ce phénomène dans des applications de communications sous-marine [61]. La *diffraction* est la résultante du principe de Huygens qui spécifie que tout point d'un signal peut-être considéré comme la source d'un nouveau signal. Il se produit quand un signal rencontre un objet de taille similaire à sa longueur d'onde et ayant des arrêtes prononcées (angle à 90°). Enfin l'*éparpillement* est la déviation d'un signal dans plusieurs directions après sa rencontre avec la surface d'un obstacle considéré comme un bon réflecteur (les parois d'une armoire en métal par exemple). L'apparition et la nature spatiale de ces phénomènes ainsi que leur impact sont étudiés dans de nombreux articles, parmi eux, on peut citer Watteyne et al. dans [129] et Zhao et Govinda dans [138].

1.2.2 Interférences et bruit

Les interférences et le bruit sont les conséquences du partage du médium sans-fil :

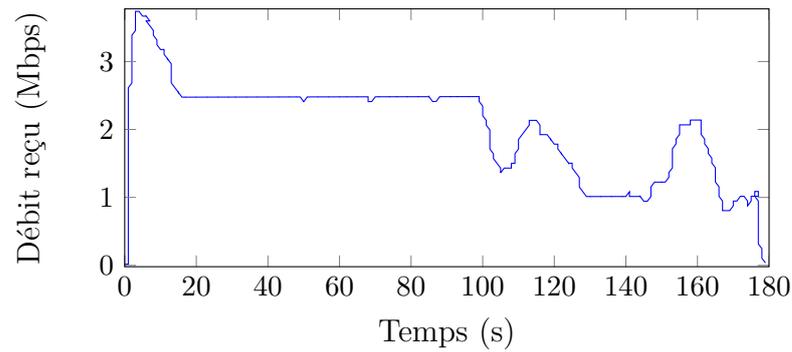
- Les *interférences* sont les signaux d'équipements reçus sur un lien mais non impliqués dans les communications. Les *interférences externes* sont issues des équipements n'appartenant pas au réseau contrairement aux *interférences internes* [133].
- Le *bruit de communication* est le signal résiduel pouvant être émis par des équipements de communication non reconnus ou par des instruments en fonctionnement (micro-onde [139], moteur, ...).
- Le *bruit électromagnétique* est le bruit issu du circuit d'émission ou de réception du lien résultant des imperfections de fabrication et de celles des composants.

Les interférences et le bruit constituent une cause majeure d'erreurs sur les communications sans-fil et font l'objet d'études importantes [133, 77, 139]. MacDonald et Ucci [77] utilisent une métrique appelée température limite d'interférence pour quantifier le niveau d'interférence maximale admis par un réseau 802.11. Ils démontrent ainsi que la limite de résistance réellement mesurée est bien inférieure aux limites théoriques. L'impact des interférences externes est étudié par Villegas et al. dans [126], qui étudient l'effet des émissions sur canaux adjacents des réseaux 802.11 concomitants. Des méthodes de gestion du spectre sont alors proposées. L'effet des interférences et du bruit sur la couche physique des communications sont reconnus, néanmoins certains protocoles de plus haut-niveau peuvent également se retrouver perturbés. C'est ainsi que Scalia et al. dans [106] mettent en évidence l'impact du bruit et des interférences sur le fonctionnement des protocoles CSMA-CA utilisés par 802.11. En effet, en surestimant l'ampleur de ces perturbations, les algorithmes réduisent de manière artificielle l'usage effectif du canal par les nœuds 802.11.

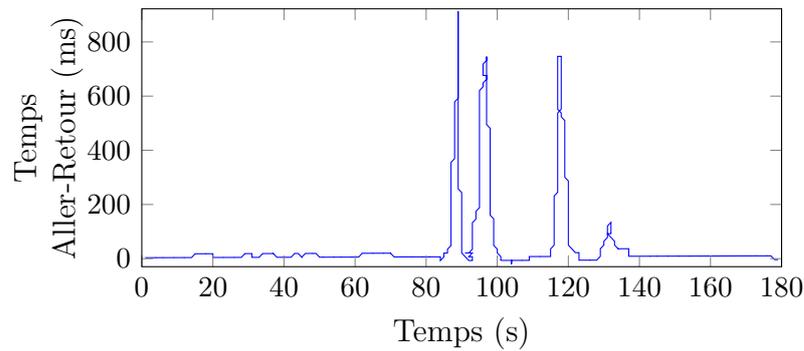
1.2.3 Impact du médium sur les piles de communication sans-fil

Le standard OSI (*Open Systems Interconnection*) divise le fonctionnement du réseau en couches distinctes ayant des fonctionnalités différentes. Chacune de ces couches est alors conçue pour interagir avec les couches disposées immédiatement sur et sous elle. La performance du réseau entier est donc dépendante des performances séparées et des services offerts à chacune des couches, et en particulier aux couches physiques et accès au lien. Celles-ci assurent la transformation des informations numériques en signaux et leur transmission sur l'infrastructure physique.

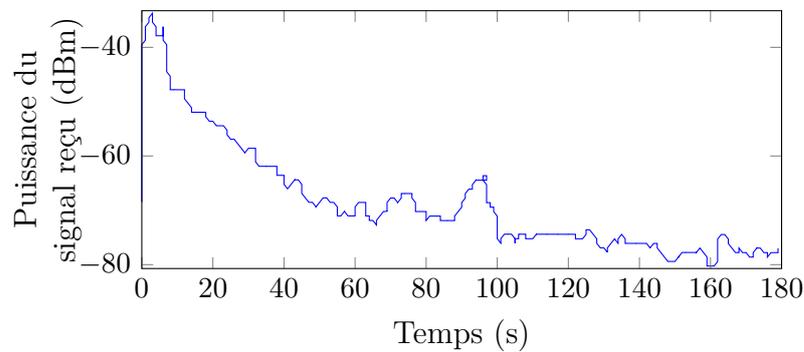
Alors que dans les réseaux filaires, ces deux couches fournissent une intégrité de signal et un niveau de stabilité suffisants, dans les réseaux sans-fil, l'efficacité de ces deux points n'est pas garantie à cause notamment des différentes contraintes du médium. Au final, les couches les plus hautes de la pile, considérées comme fiables dans les réseaux filaires, peuvent se retrouver impactées de manière importante par les imperfections du médium sans-fil.



(a) Débit reçu mesuré au niveau de la couche application.



(b) RTT mesuré au niveau de la couche réseau.



(c) Puissance du signal mesuré au niveau de la couche physique.

FIGURE 1.1 – Mesures de performance des couches application 1.1(a), réseau 1.1(b) et physique 1.1(c) sans-fil lors de la diffusion d’une application vidéo. Graphiques extraits de [72].

Ces phénomènes sont illustrés sur la figure 1.1 qui montre diverses mesures de performance prises à de multiples niveaux d’une même couche de communication lors de la réception d’un flux vidéo sur un lien sans-fil. Sur ces figures, un changement minime au niveau de la couche physique (seconde 90) a pour conséquence

une brusque montée du RTT et une dégradation du niveau de performance aux niveaux des couches applicatives. Cette dégradation se traduit par une chute de plus de 1,6 Mbps du débit reçu [72].

1.3 Couche physique pour les réseaux sans-fil

Les contraintes des réseaux sans-fil réclament donc une attention particulière. La couche physique qui assure la transformation des données numériques en signal lors de l'émission et le processus inverse lors de la réception met en œuvre des mécanismes conçus pour réduire les effets du médium sur les communications. Afin d'illustrer synthétiquement ces processus, cette section présente les étapes d'une chaîne d'émission standard utilisée dans la conception d'équipements de transmission numérique [122]. Les figures présentant respectivement les chaînes de transmission numérique pour l'émission et la réception sont respectivement présentées en figure 1.2(a) et 1.2(b).

1.3.1 Codage canal et binaire des données

La première étape de préparation du signal est le *codage canal*. Sa fonction est de permettre de limiter l'incidence des modifications du signal lors de sa propagation sur le canal. Le codage canal tend à rendre plus robuste les informations transmises en rajoutant de la redondance dans les données. Le récepteur peut ainsi détecter si les données ont été altérées et le cas présent essayer de rétablir les informations d'origine. Le point négatif de ces algorithmes et qu'ils rajoutent des informations 'non utiles' (qui ne convoient pas d'information supplémentaire) au message. La proportion de données utiles dans un flux aussi appelée taux de codage correspond au rapport $C = \frac{k}{n}$ du nombre de bits utiles k sur le nombre de bits de sortie n .

Le *codage binaire* convertit le message discret binaire en un signal discret. Celui-ci est constitué de symboles qui représentent un ou plusieurs bits du message discret initial. La transcription des séquences de bits vers des symboles est effectuée en suivant un schéma de modulation numérique qui conditionne la performance du système et la robustesse du signal. Le schéma de modulation le plus simple et le moins performant est celui d'une modulation d'amplitude à deux états qui convertit un bit en symbole. Par conséquent, un symbole c_k obtenu par une telle transcription représentera deux valeurs possibles 0 et 1 et sera codé en fonction de l'amplitude du signal A selon la relation $c_k \in \{-A, A\}$. La modulation d'amplitude à 4 états permet de coder deux bits par symboles : $c_k \in \{-3A, -A, A, 3A\}$ avec c_k pouvant représenter les couples binaires 00, 11, 01 et 10.

Le schéma de modulation numérique et le taux de codage ont un effet direct sur la robustesse du signal et sur les performances du système de transmission. Un schéma de modulation performant permettra de coder plus d'informations

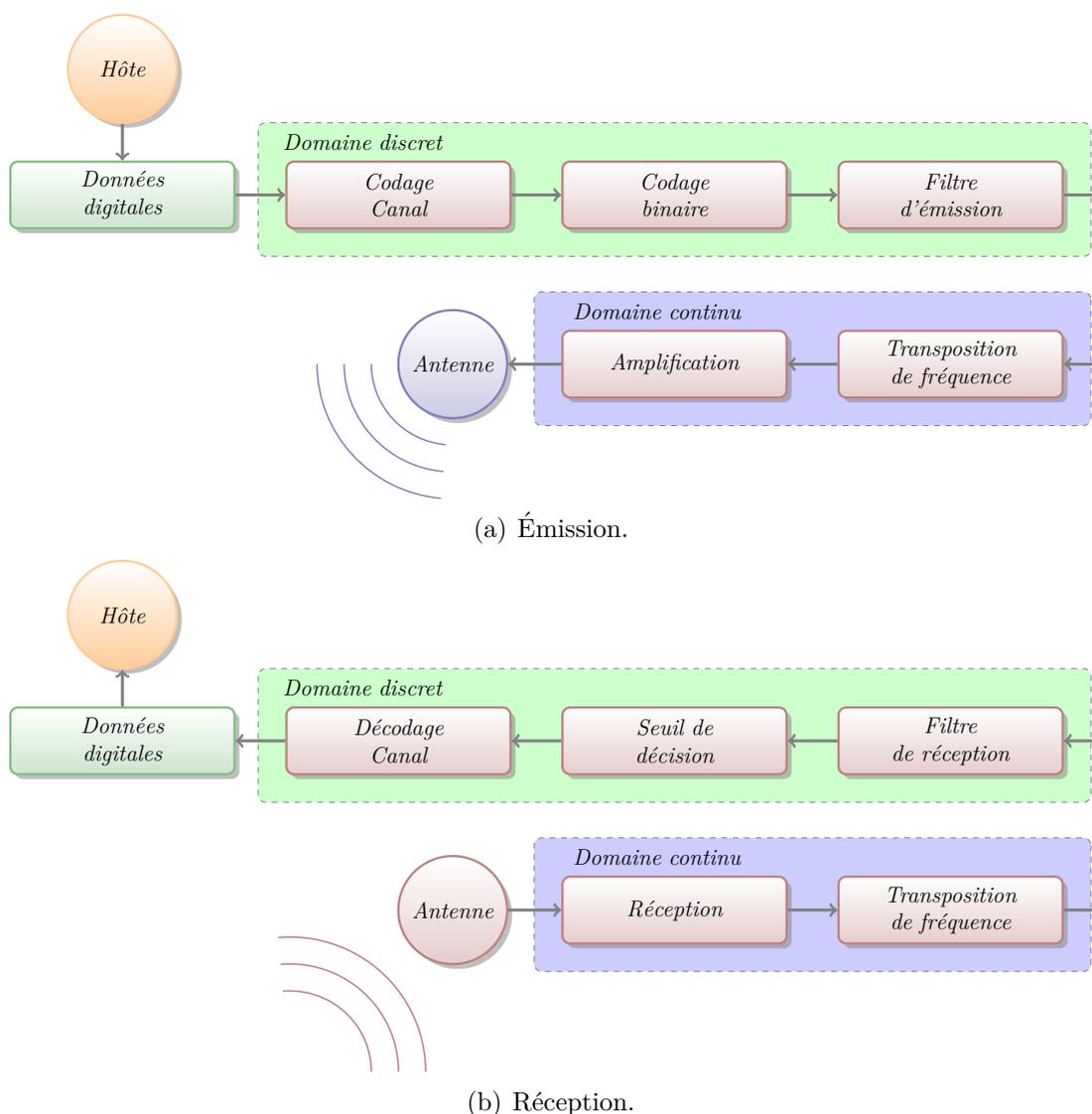
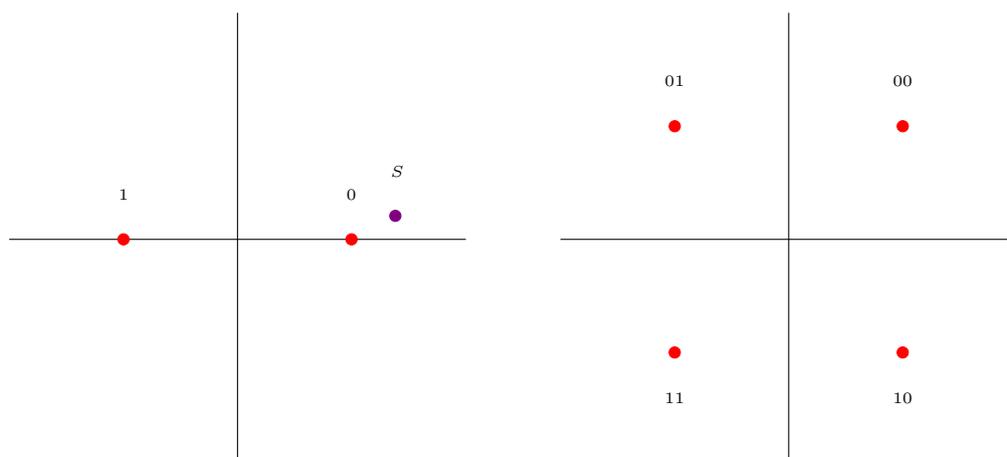


FIGURE 1.2 – Chaînes de transmission analogique/numérique.

dans un même symbole, en revanche, son décodage lors de la réception du signal sera plus difficile comme illustré sur la figure 1.3. Le tableau 1.1 présente ainsi la performance des différents schémas de modulation S_n utilisés dans le standard $802.11a,g$ en fonction de leurs caractéristiques comme le taux de codage ou le nombre de bits codés par symbole.

1.3.2 Modification de la forme d'onde et filtrage du signal

Pour pouvoir émettre les symboles c_k du domaine continu au domaine analogique, il est nécessaire de les interpoler à un signal sinusoïdal continu prenant la



(a) Diagramme de constellation pour la modulation à 2 états.

(b) Diagramme de constellation pour la modulation à 4 états.

FIGURE 1.3 – Une représentation graphique des codages binaires est celle des diagrammes de constellation dans laquelle chaque point représente un symbole c_k . Les diagrammes de constellation théoriques des schémas de modulation numérique à 2 et à 4 états sont respectivement présentés sur les figures 1.3(a) et 1.3(b) (points rouges). Lors de la phase de réception d'un signal, les symboles sont examinés. Ces symboles peuvent avoir été altérés. Le récepteur estime alors le point de la constellation idéale le plus proche du symbole reçu. Un exemple est donné sur la figure 1.3(a) dans lequel un symbole S est reçu déformé. Dans ce cas, le récepteur choisira la valeur 0 lors du décodage du symbol.

Schéma de modulation S_n	Nombre d'états	Bits par symbols	Taux de codage	Débit (Mbps)
<i>BPSK</i>	2	1	1/2	6
<i>BPSK</i>	2	1	3/4	9
<i>QPSK</i>	4	2	1/2	12
<i>QPSK</i>	4	2	3/4	18
<i>16QAM</i>	16	4	1/2	24
<i>16QAM</i>	16	4	3/4	36
<i>64QAM</i>	64	6	2/3	48
<i>64QAM</i>	64	6	3/4	54

TABLE 1.1 – Caractéristiques des schémas de modulation disponibles pour 802.11a,g.

forme de l'équation (1.1) avec A_c l'amplitude de ce signal, f_c la fréquence et ψ la phase. Cette opération est réalisée par le *filtre d'émission*, en appliquant des filtres RF ayant la propriété de contraindre le signal dans sa largeur de bande et sa puissance d'émission.

$$A_c \cos(2\pi f_c t + \phi) \quad (1.1)$$

Pour simplifier son traitement, la fréquence du signal n'est pas encore fixée et se situe entre 0 Hz et sa fréquence de coupure qui correspond à la moitié de la largeur de son canal (voir figure 1.4(a)). Le signal est alors qualifié de bande-de-base.

1.3.3 Conversion du signal bande-de-base vers le signal RF

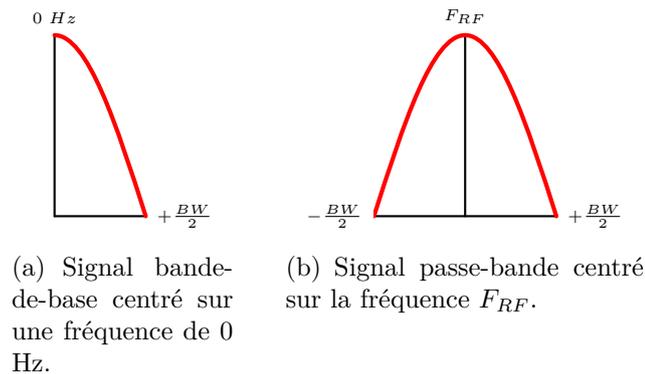


FIGURE 1.4 – Exemples de signaux bande-de-base et passe-bande.

La *transposition de fréquence* translate la fréquence du signal bande-de-base vers les fréquences plus élevées de radio fréquence et qui correspondent aux fréquences de son canal de transmission (figure 1.4(b)). Le signal est alors qualifié de RF ou passe-bande.

L'amplification permet d'élever la puissance du signal au niveau requis. Finalement, le signal est rayonné par une ou plusieurs antennes de transmission suivant les technologies utilisées (respectivement qualifiées de systèmes SISO - *Single Input Single Output* et MIMO - *Multiple Input Multiple Output*).

1.4 Conclusion

Dans un premier temps, ce chapitre a détaillé les différentes contraintes physiques liées à l'usage du médium. Ces contraintes peuvent se diviser en 2 catégories : les phénomènes physiques qui interviennent pendant la propagation du signal et les interférences et le bruit qui en perturbent la réception. Dans un deuxième temps, ce chapitre a décrit l'implémentation d'une couche de transmission numérique et notamment les différents processus qui sont mis en place pour protéger les données et le signal.

Les informations de ce chapitre seront utiles dans la suite de ce manuscrit qui abordera les problématiques de mesure physique et de propagation du signal.

Le prochain chapitre présentera la technologie 802.11 qui est l'une des plus utilisées dans les réseaux sans-fil et qui est prise en exemple dans les contributions de cette thèse.

2

Le standard IEEE 802.11

Sommaire

2.1	Introduction	17
2.2	Couche MAC	18
2.2.1	Architecture d'un réseau 802.11	18
2.2.2	Accès au médium	19
2.2.2.1	Méthode DCF	19
2.2.3	Tramage	23
2.2.3.1	Type de trames	23
2.2.3.2	Format d'une trame de donnée	24
2.3	Couche physique	25
2.3.1	Bandes ISM et interférences	25
2.3.2	Modes de transmission	26
2.3.2.1	Algorithme de choix du mode de transmission	26
2.4	Conclusion	27

2.1 Introduction

Comme le chapitre précédent l'a démontré, le niveau physique des communications sans-fil est complexe. Pour cette raison, les couches de communication physique et MAC responsables de la prise en compte de ces difficultés implémentent des mécanismes supplémentaires par rapport à leurs homologues filaires. Pour des raisons commerciales et de cohabitation technologique, l'ensemble de ces moyens est parfois formalisé sous la forme de *standard technologique*. Un des standards les plus populaires pour les technologies sans-fil est le standard WIFI aussi appelé IEEE 802.11.

IEEE 802.11 est un ensemble de normes pour réseaux sans-fil appartenant à la famille IEEE 802. Ce standard comme l'ensemble des spécifications IEEE 802 définit principalement les 2 couches les plus basses de la couche OSI : la couche physique et la couche d'accès au lien.

- La couche MAC 802.11 utilise la méthode CSMA/CA pour gérer l'accès des nœuds au lien. Celle-ci spécifie que les nœuds doivent s'assurer de la non-utilisation du canal avant d'émettre, ceci afin d'éviter toute interférence. Des mécanismes de gestion d'erreurs, comme l'utilisation d'acquittement positif (traduit de l'anglais *Acknowledgement* - ACK) pour signaler la réception d'une trame sans erreur ou d'une somme de contrôle sur chaque trame, permettent la transmission fiable des informations sur le médium. La couche MAC est aussi responsable de la sécurité des transmissions ou de l'accès au réseau (amendement *802.11i*) et peut incorporer des mécanismes de qualité de service dans son fonctionnement (amendement *802.11e*).
- La couche physique utilise les bandes ISM (*Bandes Industrielles, Scientifiques et Médicales*) de 2.4 GHz et 5 GHz comme fréquences de communication. Les différents modes d'émission disponibles pour l'envoi des trames permettent des débits de l'ordre de 450 Mbps pour les améliorations les plus récentes. Différents standards de communication sont disponibles (OFDM, DSSS, ...).

Les différentes options et mécanismes mis en œuvre par la couche MAC sont décrits plus en détail en section 2.2 tandis que les spécificités de la couche physique 802.11 sont décrites en section 2.3.

2.2 Couche MAC

2.2.1 Architecture d'un réseau 802.11

Le standard 802.11 définit deux types de nœuds :

- les stations : ce sont les points terminaux des communications du réseau.
- les points d'accès : ils constituent la transition entre le réseau local sans-fil constitué des stations et le réseau externe (e.g. Internet). Dans la majeure partie des cas, la transition se fait par l'intermédiaire d'une communication filaire.

Dans le mode *infrastructure*, majoritairement utilisé par les réseaux 802.11, le point d'accès est le coordonnateur principal des communications du réseau. La cellule créée autour d'un point d'accès est appelée ensemble de services de base (*Basic Service Set* - BSS) et regroupe l'ensemble des stations constituant cette cellule. L'ensemble des communications intra et inter-réseaux transite par le point d'accès. Par conséquent, si deux stations situées dans le même réseau local doivent communiquer, leur communication prendra au minimum 2 sauts.

2.2.2 Accès au médium

802.11 définit 3 méthodes d'accès différentes au médium :

- *Distributed Coordination Function (DCF)* : La méthode d'accès DCF implémente la méthode standard CSMA/CA. Elle consiste à vérifier si le médium est déjà utilisé avant d'émettre soit même. Pour éviter les collisions, les émetteurs utilisent des procédures de temporisation aléatoire avant chaque trame, le nœud qui a alors le temps de temporisation le plus faible dispose du médium. Dans certains cas, des techniques de réservations du médium à partir de RTC/CTS permettent de limiter encore plus les risques de collisions. DCF est la méthode de base dans les réseaux 802.11.
- *Point Coordination Function (PCF)* : La méthode PCF offre des méthodes d'accès au médium sans-contention. Il permet à des points d'accès appelés points coordinateurs (PC) un accès privilégié. Pour permettre cet accès prioritaire, les stations PC utilisent des délais réduits par rapport à ceux de la méthode DCF.
- *Hybrid Contention Function (HCF)* : Dans certains cas, un trafic particulier peut nécessiter une qualité de service supérieure à celle offerte par le modèle DCF sans toutefois nécessiter l'utilisation d'un modèle extrême comme PCF. Pour cette raison le modèle HCF permet aux stations d'implémenter des modèles de priorité sous-forme de queue d'envoi et de balancer l'accès au médium en fonction de la qualité du trafic. Ce mode d'accès est rendu disponible par l'amendement 802.11e.

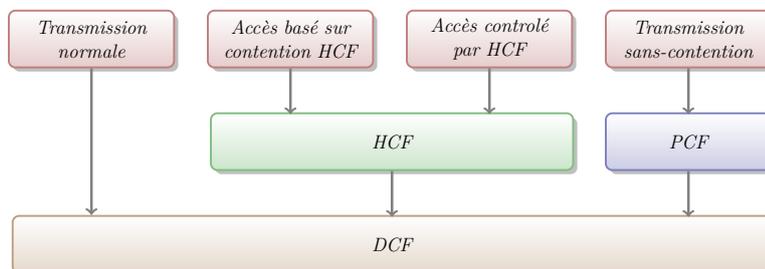


FIGURE 2.1 – Organisation des méthodes d'accès 802.11. Figure reproduite de [42].

La relation entre les différentes méthodes d'accès est présentée sur la figure 2.1. La méthode d'accès DCF est la méthode de base pour l'accès au canal en 802.11 par conséquent, ses principes et mécanismes sont décrits dans la section suivante.

2.2.2.1 Méthode DCF

La méthode DCF débute par une procédure de sondage du canal. Celle-ci permet de s'assurer que le canal est libre avant d'émettre. Le canal est considéré comme libre si aucune transmission n'a été détectée pendant un intervalle de

temps appelé *intervalle intertrame* (*Interframe Space - IFS*). Par conséquent, la procédure de sondage se déroule jusqu'à ce que cette condition soit remplie.

Après la procédure de sondage, la procédure d'émission peut avoir lieu. Plusieurs nœuds peuvent alors être en compétition pour l'accès au canal. Afin d'éviter les collisions, les transmissions effectives des nœuds sont réparties aléatoirement dans le temps selon une procédure appelée procédure de temporisation. La procédure de temporisation consiste pour chacun des nœuds en compétition pour l'accès au médium à tirer aléatoirement un créneau d'émission parmi ceux disponibles. La station qui dispose du créneau le plus court émet la première.

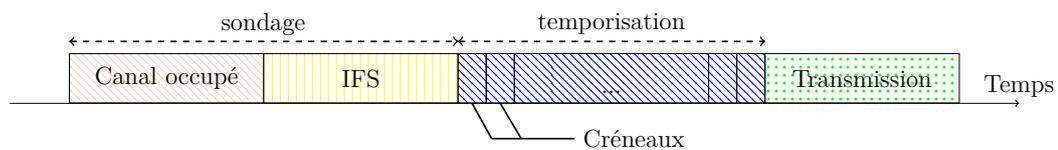


FIGURE 2.2 – Séquence d'émission DCF. Figure extraite de [42].

Entre chaque transmission unicast, l'émetteur de la trame attend la réception de la trame pendant un délai IFS court appelé *SIFS*. Si l'acquittement n'est pas reçu, le nœud recommence la procédure complète d'envoi de la trame. Avant chaque retransmission, le nombre de créneaux disponibles peut-être modifié par l'algorithme de gestion de la temporisation. L'ensemble de la séquence d'envoi est illustré sur la figure 2.2.

Les différents mécanismes mis en œuvre dans la méthode DCF disposent de nombreuses spécificités décrites ci-après.

Sondage du canal

Le sondage du canal permet de s'assurer que le médium n'est pas utilisé. Deux types de sondage sont utilisés :

- le sondage physique consiste à mesurer l'énergie sur le canal à un instant donné. Il dépend du canal et de la modulation utilisés. Si le niveau mesuré dépasse un certains seuil, le canal est considéré comme occupé.
- le sondage virtuel utilise le vecteur d'allocation réseau NAV (*Network Allocation Vector*) et les informations de niveau MAC pour inférer l'occupation du canal.

Le sondage physique peut se révéler compliqué et nécessiter du matériel onéreux. Par conséquent, le sondage virtuel est préféré. Le NAV est un minuteur présent sur chaque nœud et indique la réservation du canal. Si la valeur du compteur NAV est supérieure à 0, alors le canal est considéré occupé, dans le cas contraire, le canal est considéré comme vide. La plupart des trames 802.11 incluent un champs *durée* qui indique le temps nécessaire à leur transmission sur le canal. Ce champ est utilisé pour mettre à jour le compteur NAV.

Délais intertrames (IFS)

Étant donné la contention qui peut exister sur le médium sans-fil, des délais intertrames sont utilisés pour réguler l'accès. Une trame émise avec un délai intertrame réduit disposera d'un accès priorisé au médium. Il existe 4 délais intertrames différents :

- le délai *Short Intreframe Space (SIFS)* : le SIFS est utilisé pour réguler l'accès aux trames de forte priorité comme l'envoi de trames RTS/CTS ou l'envoi d'un acquittement. Son délai est par conséquent le plus court des délais IFS.
- le délai DCF Interframe Space (DIFS) : le délai DIFS est utilisé par le modèle DCF, c'est donc le délai standard pour l'envoi de trames ne nécessitant pas de priorité particulière.
- le délai *PCF Interframe Space (PIFS)* : Le PIFS est utilisé par le modèle PCF. Les trames émises par PCF sont priorisées par rapport aux trames standards (DCF), par conséquent la durée de PIFS se situe entre le SIFS (trames de très grande priorité) et le DIFS.
- le délai *Extended Interframe Space (EIFS)* : Le EIFS a été introduit par l'amendement 802.11e et est défini par l'application en fonction des priorités de la trame.

Gestion des erreurs

Le protocole MAC 802.11 est qualifié de *sans-perte*. Ce terme signifie que son utilisation limite les pertes d'informations liées à la transmission sur le médium. Si une erreur est détectée sur une trame reçue, celle-ci est réémise. Le mécanisme de gestion des erreurs est constitué de plusieurs composants :

- détection des erreurs : une somme de contrôle est calculée et incluse dans la trame avant son émission sur le médium. À la réception de la trame, une somme de contrôle similaire est calculée. Si une différence est constatée entre les deux valeurs, la trame est considérée comme corrompue et ignorée par la pile de réception.
- acquittement de trame reçue et retransmission de trames erronées : si aucune erreur n'est détectée à la réception, un acquittement est envoyé par le récepteur à l'émetteur. Dans le cas contraire, l'émetteur attend un temps SIFS et effectue une retransmission de la trame (traduit en anglais par *retry*).

Le nombre de retransmissions est limité. Par conséquent, quand le nombre maximal de retransmission est atteint, la trame est considérée comme perdue par la couche MAC de l'émetteur qui en informe alors les couches supérieures qui prennent alors une décision en fonction de leur propre logique.

Mécanisme de réservation du canal par RTS/CTS

Afin de garantir la bonne transmission d'une trame de données, les stations peuvent recourir au mécanisme des RTS/CTS. Le mécanisme des RTS/CTS permet de réserver le canal pendant la totalité des échanges liés à l'envoi d'une trame (envoi et réception de la trame et de son acquittement). Les différentes étapes de la procédure de réservation sont les suivantes :

- la station envoie la trame RTS au point d'accès. Cette trame contient une valeur NAV qui correspond au temps nécessaire à l'envoi de la trame de données et la réception de son acquittement. Les autres stations à proximité de la station émettrice reçoivent le RTS et mettent à jour leur horloge NAV. Par conséquent, ces stations n'émettront pas pendant la durée de l'échange.
- le point d'accès retourne une trame CTS à la station émettrice du RTS. La trame CTS contient également un champ NAV qui permet aux stations voisines du point d'accès de mettre à jour leur horloge NAV.
- après réception du CTS par la station, l'échange de la trame de données est protégé et peut avoir lieu selon la procédure habituelle.

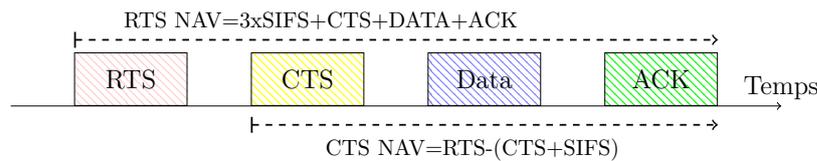


FIGURE 2.3 – Mécanisme de réservation RTS/CTS. Figure extraite de [42].

L'ensemble de la procédure est résumé sur la figure 2.3. L'utilisation des RTS/CTS provoque une charge supplémentaire du réseau. Par conséquent, les standards 802.11 recommandent que son utilisation soit désactivée par défaut [9]. En revanche, l'utilisation de ce mécanisme peut-être bénéfique dans certaines situations. Initialement, le mécanisme de RTS/CTS a ainsi été mis en œuvre pour combattre le problème du terminal caché. Dans ce cas particulier, deux stations appartenant à une même cellule 802.11 ne sont pas situées à portée de transmission radio l'une de l'autre. Par conséquent, la procédure classique de sondage de canal ne peut réguler l'accès au médium des deux stations. Si ces stations émettent en même temps, des collisions peuvent se produire au niveau du point d'accès. L'utilisation du mécanisme permet donc d'inhiber les émissions des autres nœuds de la cellule le temps de l'échange.

Modèle de temporisation exponentielle binaire (BEB)

Le modèle de temporisation exponentielle binaire (traduit de l'anglais *Binary Exponential Backoff - BEB*) gère le nombre de créneaux disponibles lors de chaque

émission. Le fonctionnement de l'algorithme BEB peut-être décrit selon trois variables, CW , CW_{min} et CW_{max} , qui sont respectivement la valeur actuelle du nombre de créneaux, la valeur minimale et la valeur maximale. Le fonctionnement de l'algorithme est le suivant :

1. avant chaque première transmission de trame, CW est initialisé à CW_{min} .
2. à chaque retransmission :
 - si CW est inférieure à CW_{max} , alors la valeur de CW est fixée à $(CW + 1) * 2 - 1$.
 - dans le cas contraire, CW prend pour valeur CW_{max} .

Les valeurs de CW_{min} et CW_{max} sont respectivement fixées à 15 et 1023 pour le standard 802.11g. Par conséquent, les valeurs prises par CW pour les différentes retransmissions sont indiquées sur la figure 2.4.

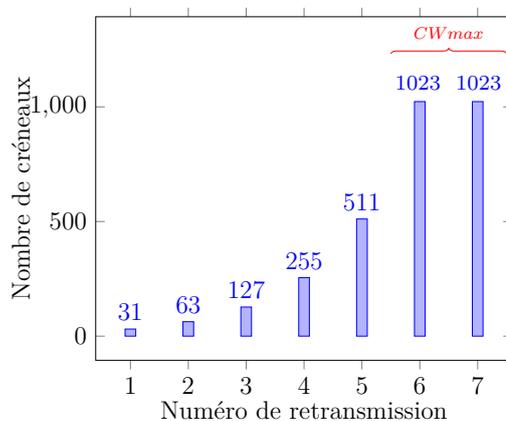


FIGURE 2.4 – Valeurs prises par CW en fonction du numéro de retransmission [9]. À partir de la 7^{ème} retransmission, le nombre de créneaux ne varie plus.

2.2.3 Tramage

Le tramage constitue un point important des protocoles MAC.

2.2.3.1 Type de trames

Trois types de trames sont utilisées par 802.11 :

- les trames de données constituent la majeure partie des trames échangées dans une cellule. Elles contiennent les informations des protocoles de plus haut-niveau.
- les trames de contrôle sont utilisées conjointement aux trames de données pour accomplir les fonctionnalités de niveau MAC : acquittement, réservation et sondage du canal.

- les trames de management sont émises par le point d'accès pour accomplir les tâches de supervision comme l'entrée d'une nouvelle station dans la cellule.

Les trames de données constituent la majorité du trafic MAC dans la plupart des réseaux. Son format est décrit dans la prochaine section.

2.2.3.2 Format d'une trame de donnée

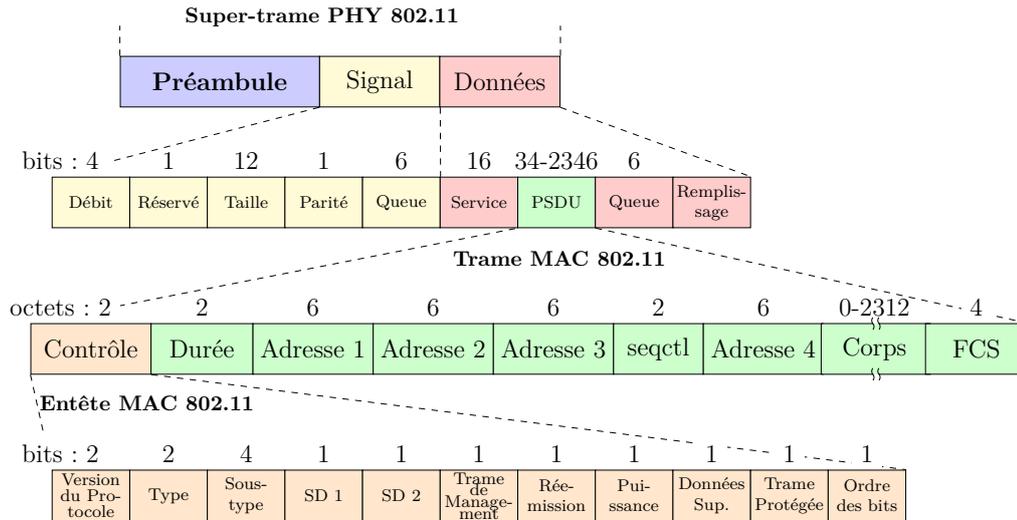


FIGURE 2.5 – Format d'une trame de donnée 802.11.

Le format d'une trame de donnée est présenté sur la figure 2.5. Les trois constituants d'une trame de donnée sont le préambule, le champ *Signal* et le champ *Données*. Seuls le champ *Signal* et le champ *Données* contiennent des données susceptibles d'être décodées par le récepteur. Le préambule est une séquence de bits fixée par le standard et identique à toutes les trames permettant aux récepteurs 802.11 de détecter l'arrivée d'une trame.

Le champ *signal* contient toute information pouvant être utilisée pour le décodage physique et la réception de la trame. Pour des problèmes de compatibilité entre les différentes normes 802.11, les symboles constituant le champ *signal* sont obligatoirement modulés par *BPSK* avec un taux de codage $\frac{1}{2}$.

Le champ *PSDU* du champ *données* contient toutes informations nécessaires au fonctionnement du protocole MAC 802.11 (acheminement local, ...) ainsi que toutes les données passées à la couche MAC pour la couche IP. Le champ *Durée* est utilisé pour le sondage virtuel du canal et la mise à jour de l'horloge NAV. Les champs *Adresse* contiennent les adresses d'émission et de réception nécessaires à l'acheminement à l'intérieur de la cellule. Un numéro de séquence (*seqctl*) est également présent dans les métadonnées et permet de différencier chacune des trames. Le champ *Corps* constitue la charge utile de la trame et est composé des

informations transmises à la couche MAC par les couches supérieures. Le champ *FCS* est la somme de contrôle, calculée sur le champ *Corps* avant l'émission. Le champ *Contrôle* du PSDU incorpore toutes les informations d'ordre protocolaire comme le type de trame (contrôle, donnée ou management), la version du protocole utilisée, si la trame est réémise ou fragmentée ou encore si les données présentes dans la trames sont cryptées.

2.3 Couche physique

Les différentes couches physiques définies pour 802.11 dans la première version du standard sont les suivantes :

- *Frequency-Hopping Spread Spectrum (FHSS)* : les systèmes FHSS sautent d'une fréquence à une autre selon un motif aléatoire. Une partie de l'information est émise dans chacune des fréquences visitées.
- *Direct Sequence Spread Spectrum (DSSS)* : l'énergie du signal est étalée sur une large bande de fréquences.
- *Orthogonal Frequency Division Multiplexing (OFDM)* : les systèmes OFDM divisent les canaux de transmission en sous-canaux où chacune des portions du signal est émise en parallèle.

Ces spécifications ont ensuite été standardisées dans les versions plus récentes du standard :

- *802.11a* permet l'utilisation des fréquences 5 GHz. La technologie d'émission utilisée est OFDM pour des débits atteignables de 54 Mbps.
- *802.11b* utilise la bande de fréquences ISM de 2.4 GHz et la technologie DSSS pour des débits maximaux de 11 Mbps.
- *802.11g* utilise la technologie OFDM similaire à celle utilisée par 802.11a. Le débit maximal est 54 Mbps mais la bande d'émission est la bande ISM 2.4 GHz.
- *802.11n* est le standard MIMO 802.11, il permet l'utilisation des bandes ISM de 2.4 GHz et 5 GHz. Les débits atteignables sont les plus importants parmi les standards 802.11 car ils permettent d'atteindre 450 Mbps.

2.3.1 Bandes ISM et interférences

La technologie 802.11 fonctionne dans les bandes ISM de 2.4 GHz et 5 GHz. Les bandes ISM (figure 2.6) sont des plages de fréquences utilisables pour la communication ou le fonctionnement d'appareils civils et à usage courant (robots d'usine, télécommandes de garage, ...). L'utilisation de ces fréquences ne requiert pas d'enregistrement préalable, *i.e.* les bandes ISM ne sont pas régulées et, par conséquent les interférences y sont nombreuses. L'utilisation et les études de fréquentation des bandes ISM est un sujet de recherche important étant donné la multiplication des équipements personnels (smartphone, ordinateurs portables, ...) [58, 116, 133].

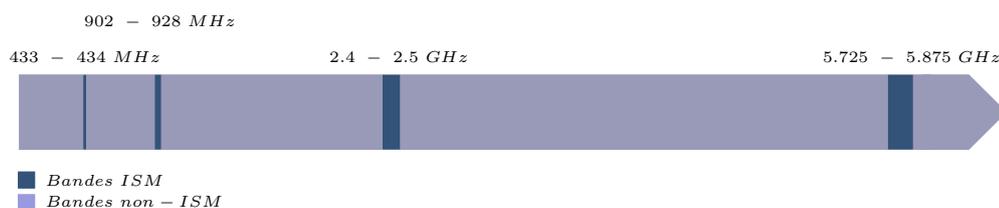


FIGURE 2.6 – Représentation des différentes bandes ISM de fréquence inférieure à 6 GHz.

2.3.2 Modes de transmission

Comme l'explique la section 1.3.1, une grande partie des caractéristiques du signal (robustesse, taux de codage, ...) dépend de la modulation utilisée (aussi appelée mode de transmission). Différents modes de transmission sont disponibles pour 802.11, les modes de transmissions disponibles pour les standards 802.11 a et g sont résumés dans le tableau 1.1.

2.3.2.1 Algorithme de choix du mode de transmission

Le choix de la modulation utilisée dépend de la capacité actuelle du canal et consiste à maximiser le débit atteignable en fonction du taux d'erreur binaire actuel (traduit par *Bit Error Rate* - BER). Le choix idéal du mode de transmission n'est pas encore réglé par la littérature. Par conséquent, différents algorithmes de contrôle existent. Parmi eux, Onoe [82, 118] utilise le ratio de perte de paquet sur un lien. Cette approche est jugée pessimiste par [82] qui propose SimpleRate : un algorithme qui teste l'efficacité des débits en envoyant périodiquement des paquets de données. La méthode des chaînes de retransmissions est actuellement utilisée dans de nombreux algorithmes. Celle-ci module le mode d'émission d'une trame en fonction du nombre de tentative de transmission (ex : 1^{ère} transmission, 1^{ère} retransmission, 2^{ème} retransmission, ...). Cette méthode est utilisée par les algorithmes AMRR [68] et Minstrel [117]. Ce dernier utilise le débit le plus performant pour la 1^{ère} tentative d'envoi, la 1^{ère} retransmission utilisera le 2^{ème} débit en terme de performance. À partir de la 2^{ème} retransmission, le mode d'envoi est le débit qui a la plus grande probabilité de passer (en fonction de l'historique des pertes sur le lien pour chacun des débits). Certaines de ces approches sont jugées optimistes [117] ou au contraire pessimistes dans les environnements à fort niveau d'interférences [82]. Par conséquent, les mesures de puissance du signal RSS ou SNR mesurées par les cartes radio 802.11 sont parfois utilisées, c'est le cas de RBAR [49], CHARM [55], FRAR [94] et SGRA [137].

2.4 Conclusion

Ce chapitre a détaillé le standard IEEE 802.11 qui définit différentes solutions aux niveaux des couches MAC et PHY des réseaux sans-fil. Au niveau MAC, la méthode DCF principalement utilisée consiste à sonder le médium avant de commencer la procédure de transmission. Cette dernière consiste à temporiser aléatoirement l'envoi des trames pour gérer l'accès des multiples nœuds au canal. Des mécanismes de retransmission sont également mis en places.

Au niveau physique, le standard définit l'usage des couches ISM de 2.4 GHz et 5 GHz. Dans la version 802.11g, la technique d'émission est OFDM. Pour gérer les aléas du médium, les trames peuvent être transmises selon différents modes de transmission ayant des capacités différentes en termes de performances et de robustesse de données. Par conséquent, le choix du mode de transmission est important et peut être géré automatiquement ou manuellement pour tenter de maximiser les performances obtenues en fonction de la qualité du canal.

Les informations données dans ce chapitre seront utiles dans la suite du manuscrit où le protocole 802.11g est pris comme exemple. Le chapitre suivant détaillera l'état de l'art des différentes plates-formes de mesure pour réseaux sans-fil et des différents travaux de comparaison entre simulateurs et bancs expérimentaux.

3

État de l'art sur les moyens de mesure de réseaux sans-fil

Sommaire

3.1	Introduction	29
3.2	Plates-formes de mesure pour réseaux sans-fil	30
3.2.1	Plates-formes de type grille	30
3.2.1.1	Utilisation des plates-formes de type grille	31
3.2.2	Plates-formes en environnements de production	32
3.2.2.1	Utilisation des plates-formes de production	32
3.2.3	Plates-formes de laboratoire	32
3.2.3.1	Utilisation des plates-formes de laboratoire	33
3.3	Simulateurs, validation des modèles et comparaison de performances	33
3.4	Conclusion	35

3.1 Introduction

Une part importante de la métrologie des réseaux sans-fil consiste à produire des données sur l'ensemble de la pile de communication et en particulier sur les couches les plus basses. Par conséquent, après les chapitres 1 et 2 qui introduisent entre autres les difficultés rencontrées au niveau de la couche physique et le standard 802.11, ce chapitre fera dans une première partie un état de l'art des différents travaux de conception et de mise en œuvre de bancs expérimentaux pour réseaux sans-fil. Ces travaux sont classifiés en trois grandes catégories définies selon les caractéristiques principales de ces plates-formes : les plates-formes de type grille qui sont constituées d'une grille de nœuds et dont la taille peut être importante, les plates-formes de mesure en environnements de production qui sont conçues sur des

réseaux opérationnels et les plates-formes de laboratoire qui sont de taille réduite et qui isolent certaines propriétés spécifiques du réseau.

La deuxième partie de ce chapitre s'intéresse aux travaux de comparaison effectués entre simulateurs et bancs expérimentaux. En effet, un moyen alternatif aux mesures expérimentales est l'usage des simulateurs de réseau. Ces logiciels sont pourtant largement critiqués pour leur manque de réalisme. Afin de mesurer ce dernier et peut-être l'améliorer, un certain nombre de travaux ont comparé les performances obtenues sur simulateurs et bancs expérimentaux. Dans cette optique, nous pensons que ces travaux peuvent et doivent bénéficier des mesures de métrologie recueillies sur une plate-forme de mesure expérimentale.

Par conséquent, le plan suivi sera le suivant : la partie 3.2 présentera la classification des différentes plates-formes de mesure tandis que la partie 3.3 présentera les différents travaux de comparaison entre simulateurs et bancs expérimentaux.

3.2 Plates-formes de mesure pour réseaux sans-fil

Cette section recense des plates-formes de mesure rencontrées dans la littérature et classifiées selon les trois catégories suivantes :

- les plates-formes de type grille de nœuds [79, 30, 84, 95] : elles disposent d'un grand nombre de nœuds (plusieurs centaines) fixes ou mobiles dans des environnements souvent automatisés. Ce sont des plates-formes principalement conçues pour les tests et la validation grande échelle de protocoles (routage, gradient, consensus, ...). Les plates-formes de ce type sont souvent conçues dans des environnements isolés et dépourvus de toute interférence externe.
- les plates-formes en environnements de production [57, 62, 32, 66, 15] : elles peuvent disposer de plusieurs dizaines voir plusieurs centaines de nœuds et sont souvent conçues dans des environnements de production industrielle ou d'usage courant comme les réseaux domestiques, de bureaux, *etc.* De plus, les équipements utilisés pour ces mesures sont des équipements standards. Par conséquent, les mesures issues de ces réseaux sont les plus réalistes.
- les plates-formes de laboratoire [120, 93, 46, 113] : elles mettent en œuvre de petits réseaux afin de simplifier leurs études. L'environnement utilisé est très isolé et permet un contrôle important des perturbations. En complément de ces particularités, une instrumentation plus complète est utilisée.

3.2.1 Plates-formes de type grille

ORBIT [100] est un banc de test développé pour fournir une infrastructure grande échelle pour le test de réseaux sans-fil. Dans sa version initiale, le banc

était composé d'une grille de 400 nœuds sans-fil (PC) situés dans un environnement totalement isolé. Chacun des nœuds dispose d'une interface Ethernet permettant sa programmation à distance. Récemment des améliorations ont été apportées avec l'intégration de routeurs *OpenFlow* au réseau sans-fil initial. De plus, des boîtiers de radio logicielle URSP2 ont été déployés pour prendre en considération les récentes avancées des réseaux cognitifs et permettre des tests de développement aux couches MAC et physique. ORBIT est le fruit d'un effort de la *National Science Foundation* qui initia en 2002 [27] une réflexion sur le développement de nouveaux bancs de tests expérimentaux pour la mesure sans-fil. Par conséquent il est souvent cité comme référence pour la conception de ce type de bancs.

De manière similaire, la plate-forme Senslab [30] installée à Lille dispose d'un total de 256 nœuds sans-fil répartis en grille, pour l'étude des réseaux 802.15.4. Parmi ces 256 nœuds 32 sont mobiles. Les nœuds sont conçus autour d'un boîtier de communication Texas Instrument CC2420 [8] dont les communications sont gérées par le standard Zigbee [39]. Les nœuds ne disposent en revanche pas de moyens supplémentaires pour la mesure physique. D'autres bancs de mesure Senslab existent permettant des mesures 802.11b ou OpenMAC sur des fréquences de 800 MHz. L'environnement de mesure est lui aussi isolé des interférences externes.

Bien que fournissant un environnement isolé de toute communication externe, la plupart des plates-formes de mesure sans-fil comme ORBIT et Senslab ne permettent pas un contrôle important des perturbations. Par conséquent, les liens de communication peuvent être sujets aux dégradations imprévues du médium. Pour ces raisons, les mesures obtenues sur ces plates-formes peuvent se révéler inexactes. Pour améliorer le contrôle des perturbations, les nœuds de la plate-forme ASSERT [84] sont reliés entre eux via câbles coaxiaux et les effets du médium simulés par des atténuateurs intelligents pilotés par FPGA [110]. Bien qu'avantageux à de nombreux égards, ce type de montage dégrade le réalisme des communications et limite les possibilités de mesure étant donné le nombre maximal de ports sur les matrices d'interconnexion des différents liens du réseau [127]. Le nombre maximal de voisin de chaque nœud est ainsi limité. ASSERT est actuellement un réseau de taille moyenne (40 nœuds) mais a été développé pour supporter plusieurs centaines de nœuds.

3.2.1.1 Utilisation des plates-formes de type grille

L'utilisation typique des plate formes de type grille est la validation d'algorithmes et de protocoles de haut-niveau. Les mises en œuvre se retrouvent ainsi facilitées par le grand contrôle des nœuds, de l'infrastructure et de l'homogénéité des nœuds. Ainsi, les auteurs de [31] mesurent la précision des algorithmes de localisation basés sur les métriques de puissance du signal en utilisant la plate-forme ORBIT. Dans [95], c'est la performance des algorithmes de contrôle de débit qui est étudiée et validée sur des réseaux 802.11 congestionnés. Les auteurs de [109] comparent les performances des trafic VoIP (Voice Over IP), sur des

réseaux expérimentaux (ORBIT) et en simulation. Enfin, dans [88], un mécanisme d'agrégation de données pour WSN est validé sur la plate-forme ASSERT.

3.2.2 Plates-formes en environnements de production

Pour plus de réalisme, de nombreuses plates-formes de mesure sont conçues autour de réseaux opérationnels comme des réseaux de campus, d'universités [57, 62] ou de bureau [32]. Les mesures effectuées sur ces réseaux sont issues du matériel standard déployé sur le réseau initial (point d'accès, nœuds, ...).

Parmi les plates-formes moyenne échelle, DieselNet [29] de l'université du Massachusetts est destiné à étudier les réseaux tolérants aux délais (Delays Tolerant Network - DTN). Le projet équipe ainsi 35 bus d'une brique *Diesel* qui permet de fournir aux passagers ou aux passants une connectivité 802.11g, de scanner d'autres réseaux 802.11 disponibles dans les environs du bus et de se connecter à des boîtes de dépôt lors de leur passage à proximité.

3.2.2.1 Utilisation des plates-formes de production

L'utilisation typique des réseaux moyenne échelle est la caractérisation du trafic des réseaux ainsi instrumentés, les modèles qui en résultent peuvent être utilisés pour l'amélioration de ces réseaux. Ainsi, [48] et [78] instrumentent et analysent des réseaux d'université tandis que [80] et [119] caractérisent du trafic haut-débit résidentiel. Dans chacun de ces cas, les usages des utilisateurs sont étudiés et modélisés.

Plus récemment, des recherches se sont intéressées au trafic mobile et smartphone, comme [38] qui instrumente des terminaux mobile pour étudier les interactions entre le trafic généré et la consommation d'énergie radio. Les auteurs de [96] étudient le trafic mobile de l'application sociale multimédia 'youtube', source dominante du trafic réseau actuel [104].

3.2.3 Plates-formes de laboratoire

Les plate-forme de laboratoire permettent de fournir une isolation et un contrôle de l'environnement important. Ces deux éléments sont fondamentaux pour la mesure du lien sans-fil [84].

Un environnement utilisé pour ce type de plate-forme est celui d'une chambre anéchoïque. Une chambre anéchoïque est un espace fermé dans lequel les murs sont recouverts de matériaux isolants qui permettent la suppression des phénomènes de multi-trajet qui affectent la propagation des signaux sans-fil. Additionnellement, les isolants permettent de prémunir l'intérieur de la chambre de la plupart des interférences externes. L'environnement de mesure situé à l'intérieur d'une chambre anéchoïque est donc neutre électromagnétiquement, ce qui en fait un outil important pour la mesure électromagnétique de précision (tests radars, audio ou mesures

d'antenne). Néanmoins, étant données les qualités de cet environnement pour la mesure RF, son utilisation s'est étendue à la mesure réseau et notamment dans la conception de nombreuses plates-formes de mesure du lien sans-fil comme dans [15, 66].

Dans ce type d'environnement, les perturbations naturelles étant nulles, une partie de la mise au point de ces plates-formes consiste à générer artificiellement des perturbations et ainsi à contrôler précisément leur force et leur apparition. Deux techniques sont essentiellement utilisées [15] : la génération de signal d'interférences par du matériel de communication ou l'injection de bruit artificiel dans l'environnement par utilisation d'un générateur de signal.

Enfin, la dernière facette des plates-formes petite échelle est l'instrumentation et la mesure des phénomènes étudiés. À cet égard, du matériel spécialisé dans la mesure RF est souvent mis en œuvre comme dans [66] où les auteurs utilisent et détaillent la configuration d'un analyseur de spectre.

3.2.3.1 Utilisation des plates-formes de laboratoire

L'environnement isolé et contrôlé des plates-formes de laboratoire les rend idéales pour la caractérisation des relations inter-couche et matérielle des réseaux sans-fil. Dans [15], les auteurs utilisent une chambre anéchoïque où les communications sont perturbées par du bruit généré artificiellement pour caractériser les relations entre la couche physique et les performances au niveau de la couche réseau. Dans [66], des mesures effectuées en chambre anéchoïques sont utilisées pour trouver les caractéristiques statistiques des mesures de puissance du signal obtenues sur du matériel standard. Celles-ci sont ensuite analysées et comparées à celles obtenues avec du matériel RF spécialisé.

3.3 Simulateurs, validation des modèles et comparaison de performances

Les bancs de tests expérimentaux sont essentiels pour la production de traces. Pourtant, des moyens alternatifs d'analyse, de validation et d'estimation de performance sont parfois rencontrés. C'est le cas des simulateurs de réseau qui présentent des avantages en termes de simplicité et de facilité de mise en œuvre malgré des défauts importants de réalisme. En effet, afin de pouvoir effectuer les simulations dans des temps et des ressources raisonnables, les modèles utilisés au niveau des couches physiques et MAC sont souvent simplifiés et abstraient volontairement la réalité de ces couches [65, 71, 63, 64]. De plus, un grand nombre de simulateurs a été conçu selon des hypothèses reconnues comme fausses [64]. Par conséquent, la crédibilité des simulateurs est fréquemment remise en question. Pourtant et notamment grâce à leur facilité d'usage, un nombre important de travaux (11 % des papiers de recherche et jusqu'à 75 % dans certains domaines spécifiques comme les

MANET selon [59]) utilisent toujours les simulateurs. Pour cette raison et malgré leurs limitations, il est important de quantifier le degré de réalisme des simulateurs les plus utilisés.

Une manière empirique pour évaluer les modèles est de comparer les performances obtenues sur simulateur avec celles obtenues sur des bancs de mesure expérimentaux. Pour pouvoir être efficace, cette comparaison doit être effectuée sur des environnements configurés de manière similaire et sur des scénarios identiques. Cette démarche est celle majoritairement suivie par de nombreux travaux dont les efforts se concentrent principalement sur les couches de bas niveaux. En effet, l'implémentation de protocoles, du niveau IP jusqu'aux couches applicatives, est souvent considérée comme verbatim vis à vis des standards par la communauté scientifique [18]. De plus, le fonctionnement des algorithmes complexes comme TCP rajoute un nombre important de variables pouvant perturber les analyses [59].

En revanche, l'impact de la simplification des modèles aux niveaux MAC et physique est activement étudié par la communauté. Ainsi, les performances du simulateur OMNET++ sont évaluées dans [33]. Les résultats montrent que les performances obtenues en termes de débit et de latence sont parfois optimistes par rapport à la réalité. L'article pointe la simplification extrême des modèles de nœuds au niveau MAC. Les simulateurs plus récents disposent de modèles de nœuds plus sophistiqués. Ainsi, [132] et [76] comparent les performances de ns-2 vis-à-vis d'un réseau de capteurs sans-fil expérimental. Les résultats de [132] démontrent une différence constatée moyenne sur le taux de perte de trames se situant entre 8,2 % et 12,3 % selon le choix de paramètres et le modèle de propagation radio. Des résultats similaires sont constatés par [22, 46] qui évaluent les performances respectives des simulateurs Castalia et TOSSIM sur un réseau de capteurs sans-fil. Sur ce dernier les différences en terme de taux de pertes sont comprises entre 5 % et 10 %. Le modèle statistique de pertes sur le lien semble être adapté dans la plupart des cas, en revanche, des anomalies plus importantes sont constatées sur la latence des échanges. Ce résultat est confirmé par [33, 46] et par [53] sur le simulateur ns-2 à partir d'un réseau statique et de taille moyenne. Une explication fournie est le manque de réalisme et de modélisation de l'écosystème propre aux simulateurs. La latence étant une métrique plus complexe que le taux de perte, elle nécessite une granularité de modélisation plus importante. Une des causes ainsi avancée par [53] est la non-prise en compte des délais et temporisation introduits par le système d'exploitation.

Ce constat persiste au niveau physique où un nombre important de modèles peuvent être utilisés pour paramétrer la propagation du signal et notamment les phénomènes de perte [87]. Ainsi, Tan et al. [120] mettent en avant des anomalies dans la puissance du signal mesurée. Selon eux, la non prise en compte de la diversité d'antennes par les simulateurs actuels est responsable de ces divergences. De plus, certains cas de propagation complexes seraient mal gérés par

les couches physiques des simulateurs [60, 71], conduisant à l'absence en simulation de phénomènes de contention ou de difficulté d'accès au médium observables dans les cas expérimentaux. Enfin, tout équipement a ses spécificités et certaines implémentations matérielles peuvent différer des standards comme le démontre [23] qui mesure les délais de temporisation mis en œuvre par des cartes WIFI lors de l'accès au lien. Les résultats obtenus entre l'expérimentation et la simulation peuvent alors être très différents.

Face à la diversité de ces résultats, certains auteurs insistent sur l'importance de définir des méthodologies de comparaison précises et partagées par la communauté [59]. D'autres travaux mettent l'accent sur la nécessité de répéter les tests de manière régulière et automatisée pour prendre en compte les différentes évolutions des simulateurs et les différents réglages [18].

3.4 Conclusion

La 1^{ère} partie de cet état de l'art a concerné la classification et la description des plates-formes de mesure expérimentale. Parmi elles, les plates-formes de laboratoire semblent constituer une bonne solution pour la conception d'un banc de mesure compatible avec la métrologie des réseaux sans-fil. La 2^{ème} partie de l'état de l'art a abordé les problèmes de validation des modèles de simulation et de comparaison de performances entre simulateurs et bancs expérimentaux. Parmi les travaux cités, certains auteurs [59, 18] insistent sur la nécessité de développer de nouvelles méthodes de comparaison plus fiables et plus automatiques que les études comparatives effectuées actuellement.

Ces considérations seront prises en compte dans les prochains chapitres. Le chapitre 4 décrira ainsi la conception d'un banc de test adapté à la métrologie des réseaux sans-fil tandis que le chapitre 5 définira et mettra en pratique un modèle de comparaison entre environnements de mesure. Ce modèle permettra à partir de jeux de données issus de simulations et d'expérimentations de trouver des anomalies d'implémentation, de modélisation ou de configuration entre les deux environnements.

4

Banc de mesure sans-fil inter-couche pour la mesure RF

Sommaire

4.1	Introduction	38
4.2	Contributions	39
4.3	Problématique de la mesure RF	39
4.3.1	Spécificités	39
4.3.2	Environnement de mesure	40
4.3.3	Contrôle des perturbations	40
4.4	Problématiques fonctionnelles associées à la conception d'un banc de test distribué	41
4.4.1	Synchronisation des nœuds	41
4.4.2	Récolte des données et commande du banc de test	41
4.5	Techniques de mesure RF	42
4.5.1	Échantillonnage	42
4.5.2	Filtrage	43
4.5.3	Indicateurs de mesure physique	43
4.5.3.1	RSS	43
4.5.3.2	SNR	44
4.5.3.3	CSI	45
4.5.3.4	EVM	45
4.5.3.5	LQI	46
4.5.3.6	$\frac{E_b}{N_o}$	46
4.5.3.7	ACPR	47
4.5.4	Composants pour la mesure	47
4.5.4.1	Diviseur de puissance	47
4.5.5	Instrumentation	47

4.5.5.1	Oscilloscope LeCroy	47
4.5.5.2	Radio logicielle	49
4.5.5.3	Matériel générique 802.11	51
4.6	Techniques de mesure de trafic et de mesure MAC et réseau	51
4.6.1	Équipements	51
4.6.2	Logiciels pour la mesure et la génération de trafic	52
4.7	Implémentation du banc de test	52
4.7.1	Implémentation de la synchronisation NTP	55
4.7.1.1	Validation	55
4.7.2	Filtrage hors-ligne du signal	57
4.7.3	Impact du diviseur de puissance	57
4.8	Conclusion du chapitre	58

4.1 Introduction

Ce chapitre traite de l'étude et la conception d'un banc expérimental pour réseaux sans-fil capable de mesurer aussi bien les couches réseaux que les couches sans-fil des piles de communication sans-fil. Il s'agira de la première contribution présentée dans cette thèse. Elle permettra entre autres de générer des traces et des mesures adaptées à la métrologie des réseaux sans-fil.

La littérature regorge de bancs de test pour la mesure de réseaux. Néanmoins, peu sont réellement adaptés pour la mesure RF. Parmi les trois classes de bancs décrites dans l'état de l'art dressé en section 3.2, seuls les bancs dits *de laboratoire* sont adaptés à la mesure RF. En effet, de telles mesures nécessitent un contrôle important et une isolation totale de l'environnement. Afin d'obtenir un environnement adéquat, certains travaux situent leurs mesures à l'intérieur d'une chambre anéchoïque [15, 66]. Notre démarche décrite ici reprend celle suivie dans ces travaux tout en améliorant certains points comme la précision des mesures.

Étant donnée l'importance de la mesure RF, une large part de ce chapitre sera consacrée aux techniques de mesure au niveau physique. Y seront notamment détaillées les différentes métriques utilisées dans la littérature et disponibles sur les équipements actuels. Ces derniers seront également évoqués dans la même partie afin d'expliquer les choix effectués pendant la conception du banc expérimental.

Le plan suivi dans ce chapitre sera le suivant. Dans un premier temps, seront abordées les problématiques d'un banc de mesure RF. Nous présenterons donc en section 4.3, le choix de l'environnement de mesure RF et le contrôle des perturbations. La section 4.4 traitera des problèmes fonctionnels comme la synchronisation des nœuds. Les techniques de mesure physique seront abordées en section 4.5. Cette dernière décrira notamment les métriques et équipements de mesure RF testés ou utilisés sur le banc de test. La section 4.6 traitera de la mesure réseau

tant du point de vue matériel que logiciel. Enfin, la section 4.7 sera consacrée à la mise en œuvre du banc ainsi qu'à la caractérisation de ses performances.

4.2 Contributions

Le travail détaillé dans ce chapitre suit les recommandations des travaux désignés dans l'état de l'art sous l'appellation *bancs de laboratoire* [15, 66]. En effet, ces bancs sont les seuls susceptibles de remplir les conditions préalables à une mesure précise du lien sans-fil. En revanche, le banc décrit dans ce chapitre diffère des travaux actuels par plusieurs aspects qui améliorent les mesures recueillies :

- la précision des mesures est ainsi rendue meilleure de part les équipements et les méthodes utilisés : par exemple, les auteurs de [15] utilisent une chambre semi-anéchoïque et [66] utilise plusieurs antennes de réception. Dans notre configuration, la chambre est totalement anéchoïque et une seule antenne de réception est utilisée, son signal étant divisé vers les instruments de mesure et d'enregistrement.
- l'instrumentation et donc la nature des données diverge : dans [66] un analyseur de spectre est utilisé alors que nous utilisons un oscilloscope relié à l'antenne de réception.
- finalement, dans notre proposition, un intérêt particulier est porté au système d'exploitation et aux logiciels libres. Ce point peu abordé dans les travaux cités permet des mesures plus complètes et plus fiables.

4.3 Problématique de la mesure RF

4.3.1 Spécificités

Contrairement à la mesure réseau où l'unité de base est le bit, la mesure RF opère sur des phénomènes physiques. Par conséquent, par rapport aux mesures réseau, la mesure RF se caractérise par :

- un besoin en précision plus important du fait de la nature des observations (ondes, ...)
- une échelle temporelle plus basse : l'ordre de fréquence des technologies de communication est le Gigahertz. Par conséquent l'échelle de mesure des observations est la nanoseconde
- une difficulté de mise-en-œuvre accrue : les phénomènes radio sont aléatoires et difficiles à isoler.

En conséquence, les mesures RF nécessitent des équipements spécialisés qui ont dans certains cas des contraintes particulières (calibration, utilisation, ...). De par leur rareté, ces équipements ont un coût beaucoup plus importants que le matériel 802.11 standard.

De plus, des protocoles expérimentaux spécifiques sont requis pour pouvoir mesurer efficacement les phénomènes sans-fil et doivent être réalisés dans des environnements protégés permettant le contrôle maximal des perturbations (fréquence des apparitions et force) et la répétabilité des mesures.

4.3.2 Environnement de mesure

Dans ce contexte, un environnement intéressant est celui d'une chambre anéchoïque (figure 4.1) car il limite les perturbations causées par le médium. Une chambre anéchoïque est une pièce fermée dont les murs sont tapissés de matériau isolant, ce qui permet de limiter les perturbations du signal liées aux phénomènes de rebond et de multi-trajets. Les isolants prennent la forme de cônes de différentes tailles capables de capturer les ondes qui s'y reflètent. La taille des cônes conditionne la fréquence des ondes capturées. En plus de limiter les perturbations à l'intérieur de la chambre, les caractéristiques des isolants rendent également la chambre en grande partie imperméable aux signaux venant de l'extérieur.

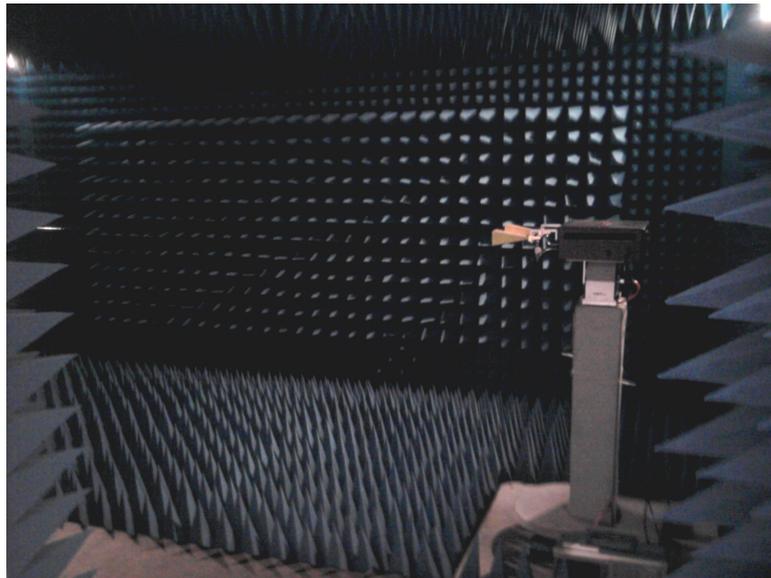


FIGURE 4.1 – La chambre anéchoïque du LAAS-CNRS nous permet d'effectuer des mesures dans un environnement protégé, *i.e.* sans les perturbations physiques naturelles qui existent dans un milieu non-protégé.

4.3.3 Contrôle des perturbations

En tirant partie de la protection apportée par la chambre anéchoïque, le niveau des perturbations peut-être contrôlé :

- (a) En disposant des objets dans la ligne de vue d'un émetteur, la propagation du signal est perturbée, générant des rebonds et favorisant les phénomènes de multi-trajet.
- (b) En multipliant le nombre de nœuds dans le réseau, des interférences sont générées provoquant des collisions ou la mise-en-œuvre de mécanismes de régulation (procédure d'accès au lien par CSMA/CA par exemple).
- (c) Alternativement à la solution (b), du bruit peut-être injecté dans l'environnement pour augmenter le nombre d'erreurs et la difficulté de réception [66].

Cette problématique est évoquée de manière plus détaillée dans la section 4.7.

4.4 Problématiques fonctionnelles associées à la conception d'un banc de test distribué

4.4.1 Synchronisation des nœuds

La mesure simultanée de traces par plusieurs appareils nécessite de les synchroniser, ceci se fait soit par moyens logiciels, soit par moyens matériels. D'un point de vue matériel, la méthode la plus connue est sans doute l'usage d'un signal satellite horodaté comme le GPS (Global Positioning System) qui permet de fournir une horloge commune à tous les nœuds de mesure avec une synchronisation inférieure à la centaine de nanosecondes [21]. Néanmoins pour pouvoir fonctionner, il est nécessaire d'effectuer les mesures à ciel ouvert et éloignées d'édifices élevés ou de tout obstacle qui pourrait limiter la réception du signal satellite. Une autre méthode de synchronisation matérielle est l'utilisation de matériel dédié comme les cartes assistants pour le protocole IEEE 1588 [81] qui permet une synchronisation élevée de l'ordre de quelques centaines de nanosecondes [52]. Le contrepartie est bien sûr l'usage du matériels supplémentaire.

D'un point de vue logiciel, des solutions comme NTP (Network Time Protocol) [81] existent et sont souvent plus faciles à intégrer car elles ne nécessitent pas de modifications matérielles. En revanche, leur précision est limitée à quelques millisecondes [81].

4.4.2 Récolte des données et commande du banc de test

Le rapatriement des données vers un puits pendant les expériences peut-être nécessaire. Ceci se justifie par le manque de mémoire de certains instruments de mesure ou par le besoin de contrôle et de vérification des informations par un nœud maître. Les procédures de récolte peuvent être contraignantes pour le banc. La nécessité de ressources de calcul (compression des données), de mémoire (stockage des données de compression temporaires) et de ressources réseau dues à leur transfert peuvent porter atteinte aux besoins de fidélité scientifique du banc. Par

conséquent, cette problématique doit nécessiter une attention toute particulière ainsi que des réglages importants tant sur l'infrastructure du banc de test (mise en place d'une topologie adéquate) que sur les nœuds en eux même (optimisation du stockage par exemple).

De plus, pendant les mesures, tous les équipements du banc doivent exécuter leurs opérations locales (mesure, rapatriement, synchronisation, ...) en accord avec le protocole expérimental en cours. Ainsi, la nature distribuée du banc nécessite l'usage d'un serveur capable d'orchestrer la conduite du protocole expérimental et de sélectionner les séquences adéquates comme le démarrage des interfaces ou des mesures, mais aussi le réglage des nœuds.

Pour respecter les nécessités de précision et pour ne pas interférer avec les mesures, ces problématiques peuvent nécessiter l'ajout de matériels ou d'équipements supplémentaires. Ainsi un réseau secondaire filaire peut permettre de palier aux nombreux problèmes de dégradation de performances et d'interaction avec les mesures.

4.5 Techniques de mesure RF

4.5.1 Échantillonnage

L'échantillonnage du signal est un processus de conversion du signal analogique vers du signal numérique. En effet, le signal transmis par les circuits de télécommunication est analogique et continu. Après sa réception, le signal reçu est digitalisé par approximation des valeurs de ce signal à intervalles successifs et constants, selon une fréquence appelée fréquence d'échantillonnage. Son unité est exprimée en nombres d'échantillons (sample en anglais) par unité de temps (S/s). Cette fréquence doit respecter la loi de Nyquist-Shannon qui précise qu'un signal continu ne peut-être échantillonné que s'il ne contient pas de composantes fréquentielles au dessus de la moitié de la fréquence d'échantillonnage. Par exemple, un signal échantillonné à 2000 S/s nécessite que le signal analogique ne soit composé que de fréquences inférieures à 1 kHz.

Le signal RF ainsi échantillonné peut prendre plusieurs formes :

- L'*amplitude instantanée* du signal représente la valeur de son amplitude à un instant t . C'est la composante A_c d'un signal sinusoïdal (voir relation 1.1).
- La *représentation complexe IQ* du signal qui est une représentation du signal intégrant sa phase et son amplitude.

La figure 4.2(a) est une représentation polaire du signal où le vecteur complexe intègre l'état instantané du signal (équation (1.1)) et est formé par ses coordonnées de phase et d'amplitude.

Les *données IQ* sont issues de cette représentation après translation des données

de phase et d'amplitude de ce système vers un système cartésien (X,Y) (figure 4.2(b)). La conversion s'effectue par simple utilisation des relations trigonométrique $I(t) = M(t)\cos(\phi(t))$ et $Q(t) = M(t)\sin(\phi(t))$.

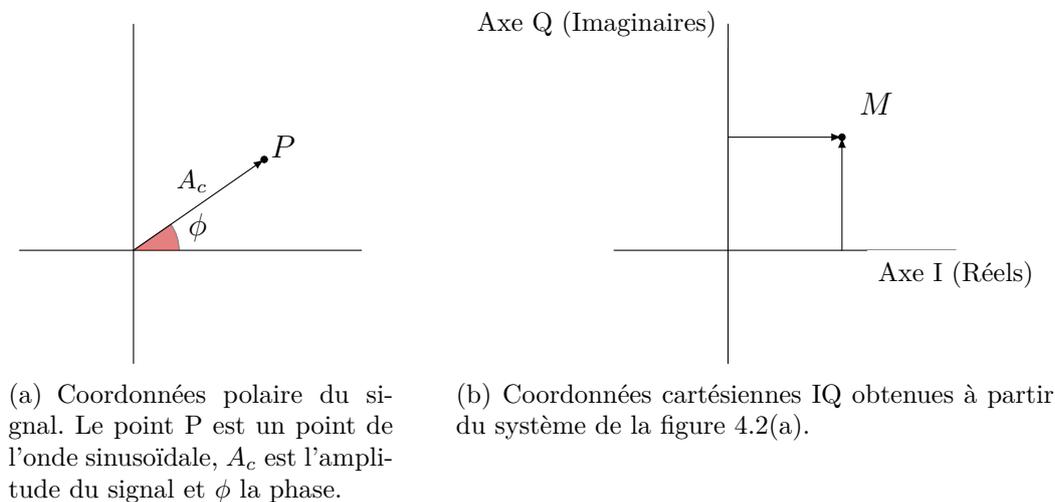


FIGURE 4.2 – Illustration des différentes représentations du signal.

4.5.2 Filtrage

Tout signal capturé contient du bruit, qui peut-être dû à l'imperfection des circuits et des câbles, au mode de mesure mal adapté ou au bruit rayonnant sur le canal. Par conséquent, il peut être nécessaire de filtrer le signal reçu. Des composants matériels existent pour un filtrage du signal réalisé expérimentalement. La contrainte principale de ces filtres est la perte de puissance du signal filtré. Pour pallier à ce problème, des filtres logiciels peuvent-être mis-en-œuvres. En revanche, leur utilisation peut contraindre certains paramètres des mesures comme la fréquence d'échantillonnage (voir section 4.5.1).

4.5.3 Indicateurs de mesure physique

Un des objectifs du banc de test est la mesure physique du lien sans-fil. Par conséquent, la connaissance des principaux indicateurs de mesure physique est nécessaire. Cette partie détaille les indicateurs de mesure les plus rencontrés dans la littérature et initialement envisagés lors de la conception du banc de test.

4.5.3.1 RSS

La puissance du signal reçu (RSS) est la puissance du signal mesurée par l'antenne. C'est la métrique la plus rencontrée dans la littérature. Son unité est le décibel (dBm).

Les variations du RSS sont modélisées par diverses lois physiques. Une loi fondamentale, la loi de Friis [41], indique que la puissance du signal décroît en fonction du carré de la distance vis-à-vis de l'émetteur et est donnée par l'équation 4.1 avec $P_r(d)$ la puissance reçue en fonction de la distance d , P_t la puissance de transmission du signal, G_t et G_r les gains respectifs des antennes de transmission et de réception. Enfin λ est la longueur d'onde du signal et L inclut toutes les pertes liées au matériel (antennes, filtres, lignes de transmission).

$$P_r(d) = P_t \frac{G_t G_r \lambda^2}{(4\pi d)^2 L} \quad (4.1)$$

Comme λ varie en fonction de la vitesse de l'onde c et de sa fréquence f selon $\lambda = \frac{c}{f}$, l'équation 4.1 devient l'équation 4.2.

$$P_r(d) = P_t \frac{G_t G_r c^2}{(4\pi d f)^2 L} \quad (4.2)$$

On remarque donc que la puissance reçue décroît également en fonction de la fréquence de l'onde : plus la fréquence est élevée, plus l'onde subira de pertes due à la distance.

La valeur de perte de propagation (traduite en anglais par *Propagation Path Loss* - PL) est la perte subie par le signal qui se propage en espace libre. Elle est donnée par l'équation 4.3 et peut être calculée sans le gain de l'antenne car elle est la différence entre la puissance transmise ($P_t G_t$) et la puissance reçue en face de l'antenne.

$$P_l = -10 \log_{10} \frac{(\lambda)^2}{(4\pi d)^2} = +20 \log_{10} \frac{(4\pi d)}{\lambda} \quad (4.3)$$

En utilisant ces relations, le RSS peut-être utilisé pour estimer la distance entre le récepteur d'un signal et l'émetteur.

Dans le contexte 802.11, la puissance du signal reçu est un indicateur (Received Signal Strength Indicator - RSSI) renvoyé par la couche Radio et compris entre 0 (pas de signal) et 255 (signal maximal) [6]. La valeur du RSSI est calculée sur le préambule de la super-trame 802.11 (voir figure 2.5). Néanmoins, le standard définit seulement la nature de la valeur mesurée et aucune obligation n'est spécifiée concernant le processus de mesure. Par conséquent, le RSSI mesuré par une carte radio 802.11 est spécifique au constructeur et doit donc être utilisé avec précaution.

4.5.3.2 SNR

Le rapport signal sur bruit (traduit en anglais par *Signal to Noise Ratio* - SNR ou $\frac{S}{N}$) est décrit comme le rapport de l'amplitude d'un signal donné sur l'amplitude des signaux de bruit. Une approximation théorique de mesure du SNR sur un signal numérisé constitué de T échantillons est donnée par l'équation (4.4),

E_t et n_t représentant respectivement l'amplitude et le bruit mesuré du signal pour l'échantillon t .

$$SNR = \frac{\text{Energiedusignal}}{\text{Energiedubruit}} = \frac{\frac{1}{T} \sum_{t=1}^T E_t}{\frac{1}{T} \sum_{t=1}^T n_t} \quad (4.4)$$

Similairement au RSS, l'unité généralement utilisée est le décibel (dBm). Grâce au SNR, la loi de Shannon établit une borne sur la quantité maximale d'informations sans erreur qui peut être transmise sur un canal en présence de bruit et d'interférences et en fonction de la bande passante disponible. Cette valeur aussi appelée capacité d'un canal de transmission numérique est définie par :

$$C = W \cdot \log_2 \left(1 + \frac{S}{N} \right) \quad (4.5)$$

avec W la largeur de bande passante disponible du canal.

Des variantes de cette métrique existent. Par exemple, le rapport signal sur bruit plus interférences (traduit par Signal-to-Noise-plus-Interference ratio SNIR) est défini comme le rapport de l'intensité du signal sur l'intensité du bruit et des interférences.

4.5.3.3 CSI

Les nouvelles technologies MIMO comme la norme 802.11n nécessitent des analyses de signal plus avancées que les technologies SISO. Lors de l'émission, les techniques de formation du faisceau peuvent améliorer la propagation du signal en contrôlant le niveau d'émission et la direction de chacune des antennes. Pour cela, des mesure précises effectuées lors de la réception du signal sont effectuées puis renvoyées à l'émetteur qui après analyse peut alors modifier sa technique ou ses paramètres d'envoi. Dans le contexte du standard 802.11, ces mesures sont appelées CSI et sont constituées entre autre des mesures de phases et d'amplitude pour différentes combinaisons d'antennes de réception et d'émission ainsi que des informations de réception du signal comme la puissance du signal reçue sur les différents canaux OFDM [10].

Le CSI est disponible au même titre que le RSSI pour certains équipements de communication 802.11n.

4.5.3.4 EVM

L'amplitude du vecteur d'erreur (traduite en anglais par *Error Vector Magnitude*) rend compte des différentes distorsions subies par le signal lors de sa propagation sur le médium. L'EVM mesure donc la différence entre le signal reçu et le même signal qui se serait propagé sur un canal idéal et n'aurait pas subi de déformation, ce dernier étant qualifié de référence. Dans la pratique, l'EVM se calcule à partir des diagrammes de constellation des deux signaux (4.3)

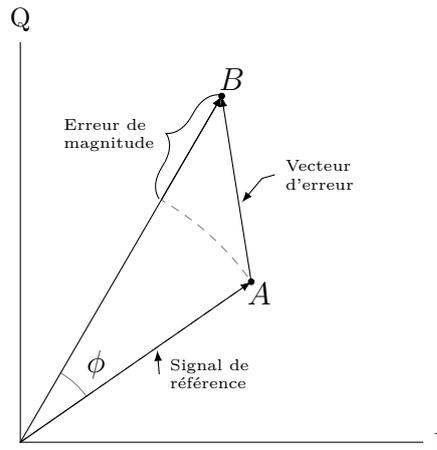


FIGURE 4.3 – Le point A est le signal du diagramme constellation d'origine. Le point B est le signal reçu. Le vecteur d'erreur se calcule comme la différence des coordonnées des deux signaux.

4.5.3.5 LQI

Le LQI est un indicateur qui mesure la qualité du signal reçu en quantifiant les déformations du signal dans les réseaux 802.15.4. Une première manière de calculer le LQI est d'accumuler les erreurs entre la constellation idéale et la constellation reçue sur une partie de la trame reçue. Le LQI est alors similaire à l'EVM. D'autres matériels approximent le LQI en utilisant la relation (4.6) avec $CORR$ la corrélation matérielle. Les coefficients a et b sont des constantes spécifiques au matériel utilisé.

$$LQI = (CORR - a) * b \quad (4.6)$$

Communément à l'EVM, le calcul du LQI dépend de la modulation utilisée. Par conséquent, ces valeurs dépendent en grande partie de facteurs expérimentaux (équipements, protocole expérimental, ...).

4.5.3.6 $\frac{E_b}{N_o}$

Le rapport $\frac{E_b}{N_o}$ correspond à la quantité d'énergie utilisée pour envoyer un bit d'information, *i.e* elle correspond au SNR par bit d'information reçu. Elle est liée au SNR par l'équation (4.7) avec f_b le débit du canal et B la largeur du canal.

$$\frac{S}{N} = \frac{E_b}{N_o} \cdot \frac{f_b}{B} \quad (4.7)$$

Le rapport $\frac{E_b}{N_o}$ est essentiellement une valeur théorique utilisée pour la mise au point ou le développement de matériels de communication.

4.5.3.7 ACPR

Le rapport de l'énergie du canal utile sur l'énergie d'intermodulation (Adjacent Channel Power Ratio) est le rapport de l'énergie émise sur un canal pour communiquer et de l'énergie totale émise sur les canaux voisins. Cette mesure est principalement utile sur les technologies de communication à modulation de fréquence qui divisent leur spectre en canaux symétriques dont les fréquences peuvent se chevaucher et se perturber.

4.5.4 Composants pour la mesure

4.5.4.1 Diviseur de puissance

Un diviseur de puissance est un composant passif qui permet de diviser un signal d'entrée en deux signaux de sortie similaires. La figure 4.4 présente le schéma de conception d'un diviseur de signal à 2 voies utilisé dans le banc de test. Une résistance est appliquée entre les ports B et C. La valeur de cette résistance conditionne certaines propriétés du diviseur de puissance qui ont une influence sur les modifications du signal qui le traverse. En effet, l'utilisation de ce composant peut avoir des conséquences négatives sur la qualité des mesures (voir figure 4.5).

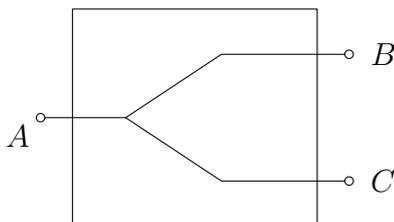
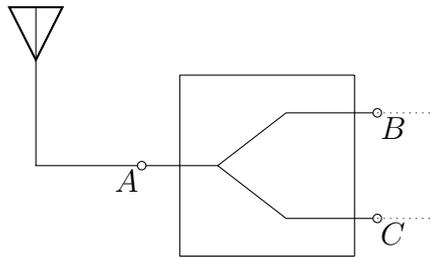


FIGURE 4.4 – Schématique d'un diviseur de signal 2 voies.

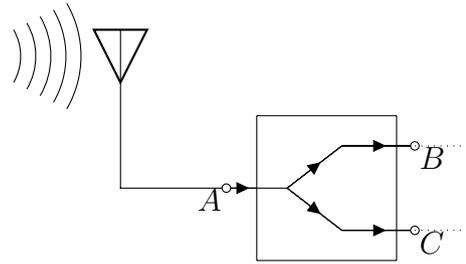
4.5.5 Instrumentation

4.5.5.1 Oscilloscope LeCroy

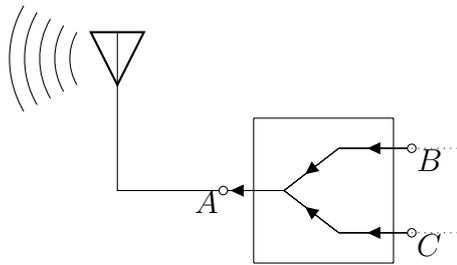
Un oscilloscope est un appareil de mesure du signal dans le domaine temporel. Il prend en entrée un signal analogique et le transcrit dans une représentation numérique échantillonnée (voir section 4.5.1) de son amplitude instantanée. Des analyses simples sur les données mesurées permettent alors d'obtenir facilement les différentes mesures RF. De manière générale, la plupart des oscilloscopes n'effectue aucune transposition de fréquence (voir section 1.3.3). Ils travaillent donc sur des signaux passe-bande. Il existe des oscilloscopes de performances différentes (performance de stockage, fréquence d'échantillonnage maximale, ...). Le choix d'un oscilloscope pour la numérisation du signal dépend de la technologie étudiée et



(a) Configuration habituelle d'un diviseur de signal: le port A est câblé à une antenne et les ports B et C sont reliés à 2 instruments différents (radio, quantificateur de puissance, analyseur, ...).



(b) Cas d'un signal reçu par l'antenne. Le signal reçu par le port A est divisé vers B et C. Comme le composant est passif et ne propose pas d'amplification, l'amplitude reçue en B et C est divisée par 2, *i.e.* le signal perd 3 dB comparé au signal reçu en A. De plus, la conception du diviseur de signal induit une différence de phase entre les sorties de signal de C et B. Celle-ci dépend de la valeur de la résistance et est appelée décalage de phase.



(c) Cas d'un signal émis par l'instrument raccordé par le port B. De manière similaire au cas présenté sur la figure 4.5(a), le signal perd 3 dB quand il arrive au port A. Lors de son passage du port B vers le port C, le signal perd une valeur qui dépend d'une autre caractéristique du diviseur de signal appelé l'isolation de port.

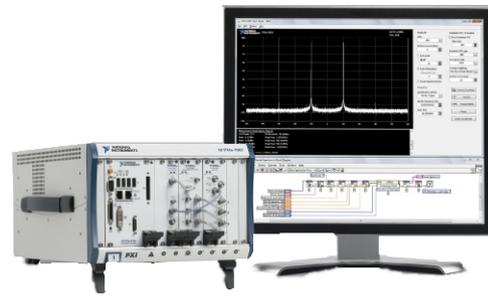
FIGURE 4.5 – Utilisations d'un diviseur de puissance.

notamment de sa fréquence qui impose selon la loi de Nyquist évoquée à la section 4.5.1 une fréquence d'échantillonnage minimale en fonction de la fréquence du signal.

L'oscilloscope dont nous disposons est un LeCroy waveform (figure 4.6(a)) qui permet de travailler sur des fréquences importantes (jusqu'à 40 GHz) et un échantillonnage de capture élevé (jusqu'à 25 GigaSamples/s). Étant donnée la fréquence radio 802.11g se situant entre 2.4 GHz et 2.5 GHz et selon la loi de Nyquist, cette valeur est largement suffisante pour leur capture. Cet oscilloscope dispose de bibliothèques logicielles qui permettent entre autre le filtrage de données en



(a) Oscilloscope LeCroy.



(b) Analyseur de signaux vectoriels National Instruments.



(c) USRP National Instruments avec différentes cartes filles.

FIGURE 4.6 – Équipements utilisés pour la mesure physique.

ligne. Malheureusement, l'exécution de celles-ci peuvent nécessiter des ressources de calcul importantes, ce qui peut perturber les mesures. Par conséquent, ces fonctions n'ont été utilisées que lors de la mise au point et l'expérimentation du banc.

La principale qualité des oscilloscopes est leur coût accessible et leur grande disponibilité. Ils sont également faciles d'usage et permettent donc des mesures fiables. De plus, les bibliothèques logicielles dont disposent certains oscilloscopes apportent des possibilités de filtrage, de calculs ou de traitements supplémentaires. Ceci rend les oscilloscopes idéaux pour l'expérimentation et la mise au point des bancs de test.

4.5.5.2 Radio logicielle

Alors que les matériels d'émission ou de réception radio traditionnels comme les cartes WIFI, réalisent la majeure partie des opérations de décodage de manière matérielle, la radio logicielle a pour objectif d'implémenter une grande partie de ces étapes de manière logicielle offrant par là même des capacités d'inspection du signal accrues.

Parmi les outils utilisant la radio logicielle, les analyseur de signaux vectoriels

sont conçus pour mesurer l'amplitude et la phase du signal d'entrée, *i.e.* les données IQ. Les analyseurs de signaux vectoriels sont donc adaptés pour la mesure de composants avancés sur les signaux comme le LQI, l'EVM ou la mesure de diagramme de constellation.

Un analyseur de signal National Instruments est présenté en figure 4.6(b). Cet appareil nous a été prêté pendant une courte durée par la branche locale de la société National Instruments. L'analyseur est constitué d'un module permettant la translation de fréquence du signal RF reçu en signal bande-de-base. Ce signal est ensuite échantillonné et numérisé à une fréquence maximale de 25 MS/s, *i.e.* supérieure à la valeur requise par la loi de Nyquist pour le signal bande-de-base d'un canal WIFI de largeur de bande 20 MHz. Un troisième module se charge du calcul des métriques et de la commande du banc à l'aide des bibliothèques propriétaires National Instruments. Un court aperçu des données récupérables grâce à l'analyseur National Instruments est présenté dans la liste 4.1.

La principale contrainte de ces analyseurs et leur coût extrêmement élevé et donc leur manque de disponibilité dans les laboratoires de recherche. De plus, leur grande complexité d'utilisation peut rendre leurs mesures moins fiables. Enfin ces équipements s'appuient souvent sur des bibliothèques propriétaires fermées pour fonctionner et nécessitent donc l'usage de logiciels payants.

- Puissance du signal reçu
- Ratio signal sur bruit
- EVM du signal
- Données IQ du signal reçu
- Perte sur canaux adjacents

Liste 4.1 – Aperçu des données récupérables à l'aide de l'analyseur National Instruments.

À contrario, la radio logicielle libre possède des bibliothèques ouvertes. Des boîtiers matériels "grand public" existent à des coûts réduits comme l'Universal Software Radio Peripheral (USRP) (figure 4.6(c)) lui aussi développé et vendu par National Instruments [3] (initialement Ettus Research [2]). Seules les fonctions de traitement du signal analogique sont réalisées de manière matérielle grâce à l'usage de cartes filles interchangeables selon les technologies mises-en-œuvre. l'USRP en lui-même est connecté par câble USB ou Ethernet à une machine hôte qui exécute l'ensemble des fonctions protocolaires de manière logicielle que ce soit en émission ou en réception. l'USRP s'appuie sur la bibliothèque GNU radio [45, 25] qui apporte de nombreuses possibilités technologiques (WIMAX, ...). Malheureusement, la performance des solutions proposées pour ces technologies n'était pas suffisante pour leur réalisation dans notre banc de test.

4.5.5.3 Matériel générique 802.11

Les matériels de communication génériques 802.11 offrent des possibilités de mesure physique. En fonction des standards supportés, la mesure des informations disponibles peut varier :

- selon les standards *802.11a,b,g*, les équipements doivent fournir une mesure de la puissance du signal reçu par trame. Cette mesure est appelée RSSI.
- selon le standard *802.11n*, en plus du RSSI, les équipement mesurent le CSI sur les différents canaux OFDM.

Bien que spécifié comme obligatoire dans les standards, la nature exacte du RSSI n'est pas spécifiée et des différences importantes existent quand à la technique ou la précision des mesures selon les différents équipements [19].

Ces dernières années, le phénomène grandissant des logiciels *libres* (traduction de l'expression anglaise *open source*) permet un accès facilité à une partie ou à la totalité du code du logiciel pilote situé sur le système d'exploitation (appelé *pilote* ou *driver*) et sur le matériel de mesure (appelé *firmware*). Cette plus grande accessibilité permet entre autre d'approfondir la connaissance des informations mesurées ou de recueillir des informations autrement inaccessibles.

Le constructeur Atheros produit du matériel 802.11 dont les logiciels pilotes sont libres. La puce Atheros AR5414 est supportée par une communauté libre importante. Cet équipement est géré par le pilote ath5k [16] (anciennement Madwifi) disponible sur les systèmes Linux. Le driver permet l'accès à de nombreuses informations de débogage et de statistique de la couche physique. Un aperçu des informations renvoyées par la puce et le driver au niveau de la couche physique peut-être trouvé dans la liste 4.3.

- Plafond de bruit courant à la couche physique
- SNR reçu du dernier paquet (couche physique)
- Compteurs d'erreur CRC (couche physique)
- Compteurs d'erreurs de décryptage (couche physique)

Liste 4.2 – Exemples de données de niveau physique disponibles avec le pilote ath5k.

4.6 Techniques de mesure de trafic et de mesure MAC et réseau

4.6.1 Équipements

À partir de la couche IP et jusqu'aux couches applicatives, les protocoles sont identiques sur les réseaux filaires et sans-fil. Par conséquent, les équipements de

mesure au niveau de ces couches ne changent pas et sont abondamment décrits dans la littérature [69]. En revanche, au niveau MAC, certaines fonctionnalités

- Compteur de trames retransmises par la couche MAC
- Compteur de bits envoyés par la couche MAC
- Compteur de bits reçus par la couche MAC
- Compteur de fragment envoyés par la couche MAC
- Compteur de fragments reçus par la couche MAC
- Compteur de trames dupliquées reçues par la couche MAC
- Statistiques de l'algorithme de contrôle de débit Minstrel

Liste 4.3 – Exemples de données de niveau MAC disponibles avec le pilote ath5k.

contraintes par le temps peuvent être implémentées de manière matérielle. C'est notamment le cas des algorithmes de sondage logique du médium ou de réponse par acquittement. Pour cette raison, les équipements en logiciels libres sont à privilégier. Un exemple de données disponibles au niveau de la couche MAC avec le driver ath5k est donné dans le listing 4.3.

4.6.2 Logiciels pour la mesure et la génération de trafic

Les systèmes Linux disposent d'un grand nombre d'outils pour la mesure réseau au niveau de toutes les couches. Ainsi, [36] présente de manière exhaustive les outils open-source de mesure de trafic. Parmi eux, l'application `tcpdump` et la librairie `libpcap` [101] permettent la capture de trafic et l'interprétation des paquets. Parmi les informations renvoyées pour chacun d'eux, certaines mesures RF sont présentes dont la valeur de puissance du signal renvoyée par le driver. Le listing 4.4 présente les différents champs récupérables grâce à `TCPdump`. En complément de ces outils, des logiciels comme `IPerf` [43] ou `Ping` [5] peuvent être utilisés pour générer du trafic UDP ou TCP.

4.7 Implémentation du banc de test

Le banc de test a été conçu dans l'environnement RF d'une chambre anéchoïque mesurant 4,10 m de long et 2,50 m de large. Les boîtiers de communication sont des plate-formes Avila [1] munies des cartes WIFI Atheros Ar5414 pouvant opérer 4 cartes radio simultanément. Ces équipements disposent de 2 ports Ethernet, d'une capacité de stockage de 8 Go sous forme de cartes Flash et d'un système d'exploitation Linux OpenWRT.

Les boîtiers sont situés de part et d'autre de la chambre (figure 4.7). Une cellule 802.11g réglée sur une fréquence de 2,412 GHz (canal 1) est créée entre les deux boîtiers et du trafic unidirectionnel TCP, UDP ou ICMP est transmis par

- Type de trame 802.11 (retransmission, data, accusé de réception, ...)
- SNR mesuré pendant la réception de la trame
- Numéro de sequence MAC
- Technologie de la trame 802.11 a,b,g
- Débit physique de la trame (54 Mbps, 48 Mbps, ...)
- Canal d'émission
- Horodatage de la trame à la réception
- Délai d'émission ou de réception de la trame
- Adresse MAC de destination ou d'émission
- Adresse IP de destination ou d'émission
- Adresse MAC de destination ou d'émission
- Taille de la trame

Liste 4.4 – Champs disponibles dans les trames 802.11 capturées par l'outil TCP-Dump.

l'intermédiaire de l'antenne A à l'antenne B . La génération de trafic peut demander des ressources de calcul conséquentes. Par conséquent, afin de ne pas biaiser les mesures, le trafic est généré sur la machine dédiée $M1$ et envoyé sur ce boîtier $P1$ par l'intermédiaire d'une liaison Ethernet. Le boîtier est configuré comme un pont qui transfère le trafic reçu de cette interface vers son interface sans-fil 802.11. De plus et toujours pour ne pas introduire de biais, les captures de trafic au niveau de la réception et de l'émission peuvent être réalisées par des boîtiers d'écoute passifs connectés aux antennes d'émission et de réception par un diviseur de puissance (nœuds $S1$ et $S2$ respectivement pour l'émission et la réception). De manière similaire, le signal dupliqué par le diviseur de puissance de l'antenne de réception est également redirigé vers l'instrumentation de mesure physique (oscilloscope, analyse de signal, ...). Dans notre implémentation, un oscilloscope LeCroy est utilisé. La section 4.5.2 introduit le filtrage, nécessaire pour nettoyer le signal recueilli du bruit issu de l'imperfection des composants de mesure. Afin de ne pas perturber le processus de mesure, ce filtrage est effectué hors-ligne.

Les perturbations consistent en du bruit généré artificiellement car bien configurée, c'est la solution la plus déterministe : pour un niveau de bruit donné, le taux d'erreurs est plus ou moins constant. De plus, après diverses expérimentations tant dans la chambre anéchoïque que dans un environnement de bureau standard, nous avons remarqué que les interférences et le bruit sont bien plus efficaces que la présence de perturbateurs physiques sur les communications sans-fil. Dans notre implémentation, le générateur de signal est un modèle Anritsu MG37000A qui permet de moduler le signal selon des technologies particulières comme DSSS ou OFDM. Dans notre cas, le modèle de bruit choisi est le Bruit Blanc Gaussien Additif (Additive White Gaussian Noise - AWGN) car la largeur de bande de ce signal est similaire à celui du WIFI étudié. De plus, son amplitude est plus ou

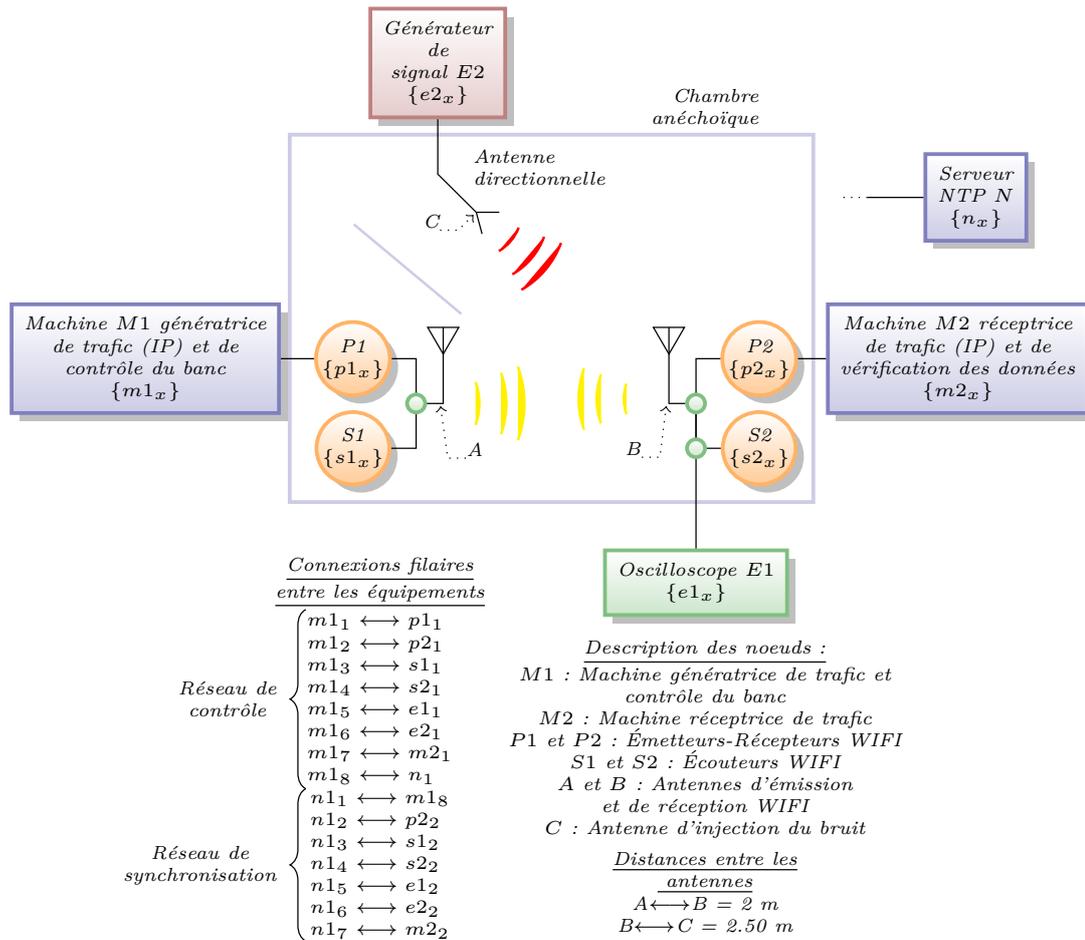


FIGURE 4.7 – Description complète du banc de test et des différents réseaux utilisés. La notation $\{a_x\}$ correspond à l'ensemble des interfaces de l'équipement A. La notation $a \leftrightarrow b$ signifie qu'il existe une connexion entre l'interface a et l'interface b . Pour ne pas surcharger les boîtiers, le trafic est généré sur la machine $M1$ et réceptionné sur la machine $M2$. Les boîtiers $P1$ et $P2$ sont donc paramétrés comme des ponts entre leur connexion filaire et leur connexion sans-fil. Les noeuds $S1$ et $S2$ sont les noeuds écouteurs facultatifs mis en place pour ne pas biaiser les mesures. L'ensemble des équipements est commandé par l'intermédiaire du réseau de contrôle et d'un logiciel situé sur la machine $M1$.

moins constante à 1 dBm près sur l'ensemble de sa largeur de bande.

Le signal de bruit est injecté dans le banc par une antenne directionnelle qui permet d'atteindre avec précision une petite surface du banc et donc de ne perturber qu'un seul boîtier communiquant.

L'ensemble des appareils et équipements est relié au réseau filaire pour répondre aux besoins de synchronisation, de commande et de récolte des données.

4.7.1 Implémentation de la synchronisation NTP

Tous les équipements du banc de test sont synchronisés via NTP en utilisant une connexion filaire. Selon les auteurs de [21], les utilisateurs de NTP peuvent espérer une synchronisation précise à la milliseconde entre le serveur et les clients. Néanmoins dans le cadre de mesures distribuées et pour les besoins de l'analyse comparative, seule la différence de synchronisation d'horloge entre deux boîtiers clients importe et doit être validée.

4.7.1.1 Validation

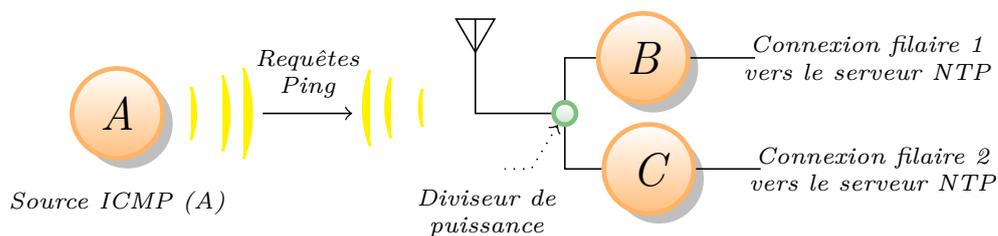


FIGURE 4.8 – Montage expérimental permettant la validation de la synchronisation NTP utilisée sur le banc de test.

Dans cette optique, et en utilisant le montage expérimental décrit sur la figure 4.8, 1000 requêtes de PING sont envoyées du boîtier A au boîtier B. Un troisième boîtier de mesure, le boîtier C, reçoit le même signal que le boîtier B. Le boîtier A est configuré comme un point d'accès tandis que les boîtiers B et C sont configurés comme des clients. Deux ensembles de mesure sont effectués en utilisant différentes synchronisations pour les boîtiers B et C. Dans un premier cas la synchronisation est effectuée une seule fois immédiatement avant le démarrage de l'envoi des requêtes ICMP et dans le deuxième cas toutes les secondes. Des captures de trafic sont effectuées sur les boîtiers B et C à l'aide de TCPDump. En utilisant les numéros de séquence des requêtes ICMP, les différences des horodatages respectifs des paquets reçus sur les 2 nœuds sont calculées.

Les statistiques sur ces mesures sont données dans le tableau 4.5. La figure 4.9(a) présente la différence des horloges en fonction du numéro de séquence ICMP dans le 1er cas de synchronisation et la figure 4.9(b) présente les mêmes données avec synchronisation régulière.

Dans le premier ensemble de test (figure 4.9(a)), on peut constater que la différence entre les horodatages s'accroît avec les séquences de Ping, donc avec le temps. Ceci peut s'expliquer par la dérive respective et naturelle des horloges qui est relative à la température ou à des différences de fabrication. Il semble donc qu'une des deux horloges ait une dérive supérieure à l'autre. La présence d'une exception (séquence 750) s'explique par des défauts d'enregistrement dans

Résultats (synchronisation NTP unique avant le ping)	
Différence d'horodatage minimale	0.000010 s
Différence d'horodatage maximale	0.005653 s
Différence d'horodatage moyenne	0.001541 s
Résultats (une synchronisation NTP par seconde)	
Différence d'horodatage minimale	0.000014 s
Différence d'horodatage maximale	0.002132 s
Différence d'horodatage moyenne	0.000157 s

TABLE 4.5 – Résultats des différences de synchronisation dans les deux cas de test.

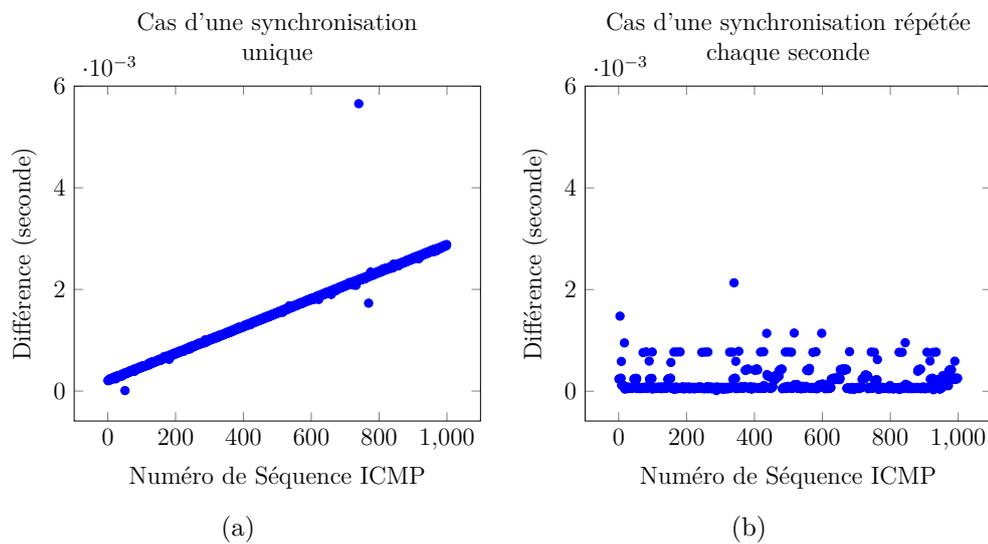


FIGURE 4.9 – Différence entre les estampillages des requêtes ICMP capturées par les boîtiers B et C dans les cas respectifs d'une synchronisation unique au démarrage de l'expérience (a) et d'une synchronisation répétée chaque seconde (b).

la trace. Néanmoins, il est à noter que la différence s'accroît raisonnablement et reste acceptable (inférieure à 3 ms). Dans le second ensemble de test, la synchronisation est effectuée toutes les secondes. La synchronisation est excellente et la différence entre les boîtiers est inférieure à 0.5 ms. Un petit nombre de rafales de valeurs comprises entre 0.5 ms et 0.7 ms sont visibles. Dans ce cas et étant données les valeurs et la figure 4.9, la dérive des horloges ne semble pas en être responsable. Ces valeurs semblent dues à la synchronisations NTP elle même qui garde des statistiques (delta, décalage, gigue) des horloges du serveur et du client et qui applique la synchronisation en fonction de ces valeurs [81]. Ceci permet de prendre en considération les caractéristiques des horloges physiques des boîtiers

et les conditions de synchronisation (température, état du réseau de synchronisation). Par conséquent, il est possible que NTP anticipe des éventuelles décalages d'horloge et applique des mises à niveau plus importantes sur un boîtier que sur l'autre.

D'une manière générale, les valeurs de synchronisation NTP sont meilleures que celles avancées dans l'état de l'art (quelques ms) pour plusieurs raisons :

- la synchronisation étudiée est celle entre les clients eux-mêmes et non pas entre les clients et le serveur.
- le schéma de fonctionnement utilisé ici (un serveur et deux clients sur un même réseau local) est bien plus simple que ceux des cas habituels (mise à jour à travers un serveur NTP internet distant).
- quand un des clients fait une requête de mise à jour, le deuxième client peut bénéficier de la réponse car celle-ci est broadcastée.

4.7.2 Filtrage hors-ligne du signal

Dans notre mise en œuvre et avec l'usage de l'oscilloscope, les données sont filtrées de manière hors-ligne par usage d'une fonction de filtre passe-bande de Butterworth [114]. Les fréquences de coupures $fc1$ et $fc2$ de ce filtre doivent respecter le critère de Nyquist qui spécifie que si fe est la fréquence d'échantillonnage, alors $fc1$ et $fc2$ doivent être comprises entre 0 et $\frac{1}{2}fe$. Par conséquent, pour des fréquences de coupure $fc1$ et $fc2$ égales à 2.4 GHz et 2.5 GHz, la fréquence d'échantillonnage devra être au minimum de 5 GS/s.

4.7.3 Impact du diviseur de puissance

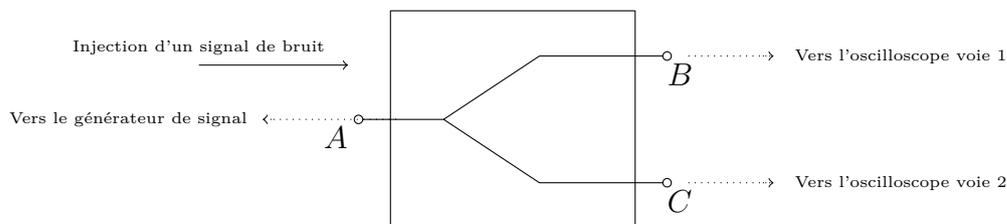


FIGURE 4.10 – Montage expérimental utilisé pour la caractérisation des sorties du diviseur de signal.

Le modèle du diviseur de puissance utilisé dans notre banc de test est un Mini-Circuits ZX10r-14+. Bien que ses caractéristiques soient connues et disponibles dans les documentations techniques [4], il est nécessaire de quantifier son impact réel sur nos mesures en matière de déphasage du signal (figure 4.5). Pour cela, le montage représenté en figure 4.10 est mis en œuvre. Un signal de bruit généré aléatoirement par le générateur de signal est appliqué à l'entrée I du diviseur. Les deux sorties A et B sont enregistrées par un oscilloscope. Une fonction

de corrélation croisée est ensuite utilisée pour mettre en évidence des différences éventuelles entre les deux signaux. Sa formule est la suivante :

$$C_{xy}(k) = \sum_{i=1}^n x(i+k).conj(y(i))$$

avec $C_{xy}(k)$ la valeur de corrélation obtenue pour les séries temporelles de signaux x et y de taille n , pour un décalage k . Sur l'ensemble des tests, la corrélation maximale est obtenue pour des valeurs de décalage comprises entre 0 et 5.8 ns. Par conséquent, les deux signaux étudiés sont identiques ou fortement similaires tant dans leur phase que dans leur amplitude. Il n'y a donc pas d'impact notable du diviseur de signal.

4.8 Conclusion du chapitre

Ce chapitre a détaillé la conception et la mise en œuvre d'un banc de test capable de mesurer l'ensemble de la pile de communication des nœuds et notamment les couches RF, situées sous le niveau 2 des technologies sans-fil. Après étude préliminaire des besoins en termes de contrôle et d'isolation de l'environnement détaillés dans ce chapitre, l'environnement choisi pour implémenter le banc de test est celui d'une chambre anéchoïque. Du bruit AWGN est généré artificiellement et injecté à l'intérieur de l'environnement à l'aide d'une antenne directionnelle afin de perturber la réception des communications sans-fil effectuées dans la chambre. Les équipements et logiciels utilisés sur le banc ont été choisis pour respecter les contraintes de non-intrusivité et d'introspection nécessaires sur un tel banc. Par conséquent, les nœuds de communication sont constitués de matériels WIFI Atheros pilotés par du logiciel libre. De plus, des équipements spécialisés dans la mesure RF comme un oscilloscope LeCroy sont utilisés. Les tests effectués sur le banc permettent de garantir certaines caractéristiques importantes comme la synchronisation précise des nœuds du banc à moins de 0.5 ms et la non-intrusivité des composants utilisés sur les données recueillies.

Dans le chapitre suivant, les données issues de ce banc sont utilisées pour améliorer le réalisme des modèles de simulation ns-3.

5

Validation expérimentale et amélioration des modèles de simulation ns-3 vis à vis d'un banc de test expérimental

Sommaire

5.1	Introduction	60
5.2	Contributions	61
5.3	Description et implémentations du scénario commun pour la comparaison entre traces expérimentales et traces de simulation	62
5.3.1	Mise en œuvre expérimentale	62
5.3.2	Mise en œuvre sous simulateur ns-3	63
5.3.2.1	Injection de bruit dans le modèle YANS	64
5.4	Métriques de comparaison	65
5.4.1	Débit IP	65
5.4.2	Taux de trames erronées au niveau du récepteur	65
5.4.3	Nombre de paquets UDP abandonnés par l'émetteur	66
5.4.4	Nombre de retransmissions nécessaires par datagramme UDP	66
5.4.5	Temps de transmission nécessaire par datagramme UDP	66
5.4.6	Temps d'inter-arrivées des trames 802.11	66
5.5	Méthode d'appairage des traces	67
5.6	Méthodes de comparaison des données	67
5.6.1	Modèle de Gilbert-Elliott pour la caractérisation des motifs d'erreurs	67
5.6.2	Modèle RCA pour la comparaison d'environnements	68

5.6.2.1	Introduction aux modèles d'arbre de comparaison	68
5.6.2.2	Définition d'un arbre pour la comparaison d'environnements	70
5.7	Application de la méthodologie d'association et de comparaison sur les données générées	73
5.7.1	Appairage des traces et validation de la configuration par méthode de Gilbert-Elliott	73
5.7.2	Application du modèle RCA	75
5.8	Analyse des résultats expérimentaux	80
5.8.1	Supplément de diagnostic	80
5.8.2	Modélisation des paramètres BEB	81
5.8.2.1	Facteurs environnementaux	81
5.8.2.2	Régression et calcul des paramètres	82
5.8.2.3	Vérification	84
5.8.3	Améliorations obtenues	84
5.9	Conclusion	85

5.1 Introduction

Les simulateurs de réseau sont régulièrement critiqués pour leur manque de réalisme, en partie à cause des abstractions qu'ils utilisent pour mettre en œuvre ces couches physiques et MAC [54, 65, 63, 64]. Pourtant, leur faible coût d'utilisation, leur capacité à mettre en œuvre des réseaux de grandes tailles et leur facilité d'usage les rendent largement plus attrayants que des bancs de test. Pour ces raisons et malgré les critiques, ils sont donc couramment utilisés dans la littérature et l'industrie [120, 93, 46, 113]. Face à ces faits, de nombreux travaux insistent sur le besoin de mener des études comparatives entre résultats de simulations et expérimentaux [63, 71, 65] dans le but d'une part de valider les modèles et le fonctionnement des simulateurs, et d'autre part, de les améliorer. Une amélioration du réalisme des simulateurs serait en effet bénéfique pour le développement de nouveaux protocoles ou de nouvelles applications pour réseaux sans-fil.

Dans cette optique, après le chapitre 4 qui décrit la mise en œuvre d'un banc de test métrologique, ce chapitre traite de la validation et de l'amélioration des différents modèles WIFI aux niveaux des couches MAC et physique mises en œuvre dans le simulateur ns-3. Ces opérations seront effectuées en tirant partie entre autres des données inter-couche de niveau MAC, PHY et IP recueillies sur le banc.

Par conséquent, ce chapitre propose dans un premier temps, un modèle d'analyse par cause racine (RCA) [112, 111] qui permet d'inférer les causes probables des différences constatées à partir de jeux de données issues des deux environnements. Afin de prouver le fonctionnement du modèle, un scénario de comparaison

identique est mis en œuvre. Celui-ci est volontairement simple pour éviter les biais de mesure et conçu pour que les implémentations dans les deux environnements soient identiques. Il consiste à générer arbitrairement du bruit pour perturber la réception d'une communication UDP unidirectionnelle sur un lien 802.11.

Le plan suivi dans ce chapitre est le suivant. La section 5.3 décrit le scénario expérimental et ses implémentations. La section 5.4 détaille les différentes métriques utilisées pour la comparaison des traces. La section 5.5 décrit le processus d'appariage de trace nécessaire à leur comparaison. La section 5.6 détaille les méthodes de comparaison. La section 5.7 donne et analyse les résultats obtenus. Finalement la section 5.8 apporte un complément de diagnostic, explique la modélisation des données expérimentales de temporisation et présente les améliorations obtenues.

5.2 Contributions

Les trois contributions principales apportées dans ce chapitre sont les suivantes :

- un grand nombre de papiers insiste sur le besoin de tester et de valider le réalisme des modèles de simulations par rapport à des données expérimentales [63, 71, 65]. Ce travail suit leurs recommandations en comparant les résultats issus du simulateur ns-3 et d'un banc expérimental. De plus, et à notre connaissance, peu d'études ont réalisé ce type de travail sur ce simulateur et sur un banc de test avec les mêmes caractéristiques.
- étant donné le grand nombre de simulateurs et de versions différentes une démarche efficace d'amélioration globales des simulateurs nécessite des méthodes permettant l'automatisation des différentes étapes de validation et notamment de comparaison des traces [18, 59]. Pour cette raison, nous proposons un modèle RCA qui permette la comparaison et la détection automatique d'anomalies. Afin de prouver le principe de fonctionnement du modèle, seul le simulateur ns-3 est utilisé et le scénario choisi est unique et volontairement simple.
- l'application du modèle RCA permet de détecter une anomalie dans l'implémentation de l'algorithme de temporisation mis en œuvre sur le matériel utilisé. Pour confirmer le diagnostic et améliorer le simulateur dans le sens du réalisme expérimental, l'algorithme de temporisation expérimental est modélisé et mis en œuvre sur le simulateur ns-3. Les modifications diminuent de manière significative l'écart entre les résultats de simulation et les résultats expérimentaux.

5.3 Description et implémentations du scénario commun pour la comparaison entre traces expérimentales et traces de simulation

Afin de pouvoir comparer les deux environnements le plus efficacement possible et ne pas introduire de biais, le scénario étudié est simple. Il consiste à envoyer du trafic sur un lien 802.11g unidirectionnel tout en perturbant la réception des trames par du bruit. Le trafic ainsi généré est de type UDP. En effet, la complexité du protocole TCP ajoute un nombre important de paramètres supplémentaires qui peuvent introduire des biais dans les comparaisons [59].

Les configurations des deux environnements sont identiques et détaillées dans la table 5.1. Afin de ne pas complexifier les analyses, un maximum de paramètres, comme la taille et le débit UDP, sont fixés. De plus, l'algorithme de contrôle du mode de transmission 802.11 est désactivé et les trames de données sont toutes émises à 54 Mbps. Enfin, les extensions liées à la QoS et aux normes 802.11e, introduisant des priorités d'envoi entre les paquets, ne sont pas utilisées. Dans les deux cas, les valeurs de bruits sont choisies afin d'obtenir une plage de volume d'erreurs complète, *i.e.* pour un taux de trames erronées reçu variant de 0% à 100%.

<i>Paramètre</i>	<i>Nom</i>	<i>Mesures expérimentales</i>	<i>Simulations</i>
Puissance de transmission	P_{ptr}	10 dBm	
Débit d'émission d'UDP	P_{DUDP}	7 Mbps	
Taille des paquets	P_{TP}	1472 B	
Plage de niveaux de bruits	P_{BR}	[-24.00 ; -18.00] dBm	[-67.7 ; -65.5] dBm
Débit binaire trames de données	P_{DT}	fixé à 54 Mbps	
Débit binaire trames de contrôle	P_{DC}	fixé à 24 Mbps	
Normes et configuration 802.11	P_{MAC}	802.11g-DCF-Non-QoS	
Nombre maximum de retransmissions 802.11	P_{RETR}	14	
Distance entre les pairs	P_{DIST}	2 m	
Environnement de propagation	P_{ENV}	Chambre anéchoïque	Espace libre (modèle de Friis)

TABLE 5.1 – Paramètres utilisés dans les scénarios et expérimentations.

5.3.1 Mise en œuvre expérimentale

L'environnement expérimental choisi est celui d'une chambre anéchoïque. En effet, les caractéristiques de cet environnement en termes de rebond, multitrajet, et interférences le rendent très proche d'un simulateur comme ns-3. La mise en œuvre expérimentale se déroule donc sur le banc de test décrit dans le chapitre 4, selon le dispositif décrit sur la figure 5.1. Un générateur de signal placé à l'extérieur de

la chambre produit du bruit AWGN (pour les raisons évoqués dans la section 4.7) qui est injecté dans l'environnement de mesure par une antenne. Cette antenne est directionnelle et pointée vers le récepteur afin de ne pas perturber le nœud émetteur.

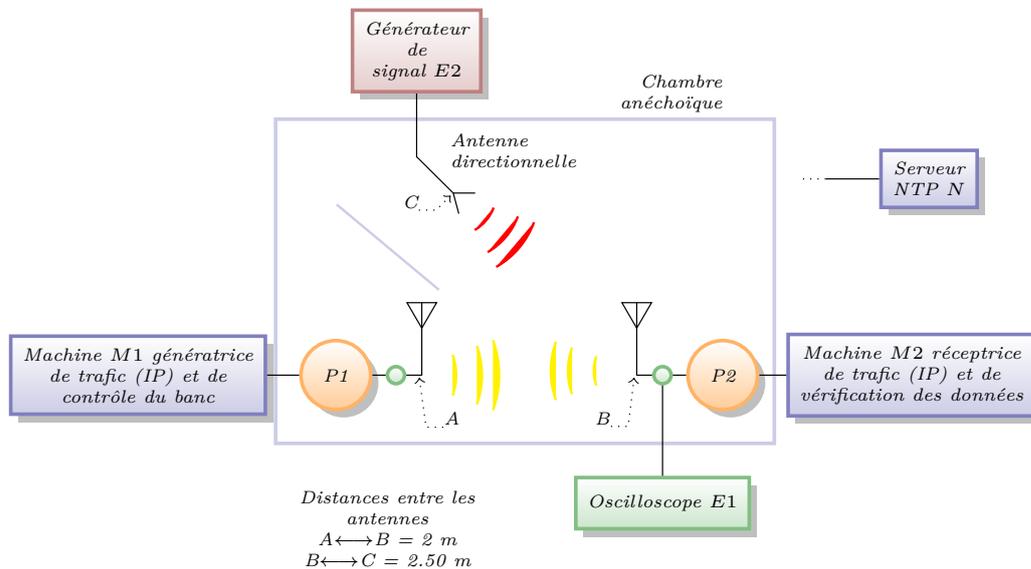


FIGURE 5.1 – Disposition des équipements pour la mesures expérimentales.

Le matériel utilisé pour l'émission et la réception est le matériel utilisé dans le banc de mesure décrit dans le chapitre 4. Les boîtiers WIFI Atheros sont utilisés pour les communications. La configuration du système d'exploitation permet la capture des paquets erronés et normalement détruits par la couche MAC de réception. Les trafics 802.11 et IP des boîtiers de communication sont capturés par l'outil tcpdump qui génère un fichier de trace pcap par niveau de bruit.

5.3.2 Mise en œuvre sous simulateur ns-3

Les modèles WIFI MAC utilisés sont les modèles WIFI standards implémentés par ns-3. Le simulateur est configuré pour capturer le trafic 802.11 reçu et envoyé par chacun des nœuds sous forme de fichiers PCAP.

Le modèle de canal et de couche physique utilisé est le modèle YANS décrit dans [67]. À notre connaissance, il n'existe pas de solution disponible actuellement pour mettre en œuvre le protocole expérimental de perturbation par le bruit sur le simulateur ns-3. Par conséquent et afin de pouvoir injecter du bruit sur le nœud récepteur, le modèle YANS est modifié comme expliqué ci-après.

5.3.2.1 Injection de bruit dans le modèle YANS

Dans la configuration ns-3 actuelle, les premières étapes de réception de trames 802.11 sont effectuées par le module YANS. Ces étapes déterminent si le paquet reçu est sain, i.e. reçu sans erreur, ou erroné. La réception d'une trame p débute par l'évaluation de la puissance du signal reçu, $S(p, t)$, calculée à partir de la puissance du signal émis et de la distance parcourue en utilisant la loi de Friis décrite par l'équation (4.1). Le ratio signal-sur-bruit-plus-interférences $SNIR_A(p, t)$ pour ce même paquet p est alors calculé selon l'équation (5.1). Dans cette relation, N_f représente la valeur du plafond de bruit constant relatif au circuit utilisé et N_i (5.2), la somme de l'énergie de tout autre signal reçu pendant la réception.

$$SNIR_A(p, t) = \frac{S_p t}{N_i(p, t) + N_f} \quad (5.1)$$

$$N_i(p, t) = \sum_{m \neq k} S(k, t) \quad (5.2)$$

L'énergie par bit E_b sur la densité spectrale de bruit N_0 se calcule selon l'équation (5.3), avec B_t la largeur de bande du signal et $R_b(p, t)$ le débit du mode de transmission utilisé par le signal p au temps t .

$$\frac{E_b}{N_0}(p, t) = SNIR_A(k, t) \frac{B_t}{R_b(p, t)} \quad (5.3)$$

La valeur du $BER(p, t)$ pour la modulation QAM utilisée pour le codage des trames émises à 54 Mbps en 802.11g est donnée par les équations (5.4), (5.5) et (5.6).

$$BER(p, t) = 1 - (1 - P_{\sqrt{M}}(p, t))^2 \quad (5.4)$$

$$\text{avec } P_{\sqrt{M}}(p, t) = \left(1 - \frac{1}{\sqrt{M}}\right) X(p, t) \quad (5.5)$$

$$\text{et } X(p, t) = \text{erfc}\left(\sqrt{\frac{1.5}{M-1} \log_2 M \frac{E_b}{N_0}(p, t)}\right) \quad (5.6)$$

La borne supérieure maximisant cette probabilité est donnée par la fonction $P_e(p, i)$, avec i , un intervalle du temps de simulation dans lequel les valeurs de BER et de débit binaire $R_b(p, t)$ sont constantes. $P_e(p, i)$ se définit pour un récepteur 802.11g par l'équation (5.7).

$$P_e(p, i) \leq 1 - (1 - P_u(p, i))^{8L(p, i)} \quad (5.7)$$

avec $L(p, i)$, la taille en bits de l'intervalle i et $P_u(p, i)$, l'inégalité de Boole de la probabilité d'erreur du premier événement (non détaillée ici, plus de détails sont disponibles dans [91] et [67]).

Finalement, la probabilité d'erreur lors de la réception d'une trame p est donnée par l'équation (5.8).

$$P_{err}(p) = 1 - \prod_l (1 - P_e(p, i)) \quad (5.8)$$

Selon l'équation (5.1), la probabilité d'erreur à la réception d'un paquet dépend de la puissance du signal reçu, de la force des interférences et du plafond de bruit constant. Afin de pouvoir implémenter le protocole expérimental et injecter du bruit capable de perturber les communications, la méthode de réception est modifiée en ajoutant une source supplémentaire de bruit, N_g dans le calcul du SNIR selon l'équation (5.9). Cette valeur supplémentaire accroît la probabilité d'erreur sur un paquet reçu.

$$SNIR_B(p, t) = \frac{S_p t}{N_i(p, t) + N_f + N_g(t)} \quad (5.9)$$

Une méthode simple et efficace d'échantillonnage du bruit injecté est la méthode d'échantillonnage naïve. Celle-ci consiste à tirer aléatoirement une valeur selon une fonction de densité gaussienne. Bien que simple, cette méthode présente des résultats suffisamment réalistes en termes d'effet sur les communications [71]. Par conséquent, l'amplitude du bruit N_g est aléatoire selon une densité de probabilité gaussienne de moyenne 0 et de variance N_0 similaire au bruit AWGN généré expérimentalement. La valeur de N_0 est fixée par l'utilisateur et correspond à la valeur du niveau de bruit. Le tirage aléatoire est effectué selon la méthode Box-Muller [70] utilisée dans ns-3.

5.4 Métriques de comparaison

5.4.1 Débit IP

Le débit IP est calculé chaque seconde en utilisant le trafic capturé sur les fichiers PCAP du nœud récepteur. Il est défini à la seconde i et pour la trace x par l'équation $BW_{\{i;x\}} = \sum_{k=1}^n L(p)$ avec $L(p)$ la taille de la charge utile à la couche IP du paquet $p \in P(\text{!erreurs})_{\{i;x\}}$ défini comme l'ensemble des paquets UDP reçus sans erreurs et présents sur la trace x à la seconde i . BW_x est l'ensemble des débits IP calculés pour la trace x .

5.4.2 Taux de trames erronées au niveau du récepteur

Le FER est calculé chaque seconde à partir du trafic 802.11 capturé sur le nœud récepteur. Il correspond au ratio du nombre de trames reçues erronées sur le nombre total de trames de données reçues. Il est défini à la seconde i pour une trace x tel que $FER_{\{i;x\}} = \frac{|P(\text{erreurs})|}{|P|}$ avec $P(\text{erreurs})$ l'ensemble des trames reçues avec erreur et P l'ensemble des trames reçues.

5.4.3 Nombre de paquets UDP abandonnés par l'émetteur

En conséquence d'une congestion sur le lien sans-fil, certaines couches de communication du nœud émetteur peuvent être saturées, les paquets en excès sont alors abandonnés. Le calcul de cette quantité s'effectue en comparant le champ d'identification IP spécifique à chaque datagramme et incrémenté pour chaque nouveau datagramme UDP généré par l'application. Pour une trace x et à la seconde i , le nombre de paquets abandonnés à l'émission est donné par $DROP_{\{i;x\}}$.

5.4.4 Nombre de retransmissions nécessaires par datagramme UDP

Le nombre de retransmissions nécessaires par datagramme UDP est le nombre moyen de retransmissions nécessaires à la couche MAC 802.11 pour transmettre avec succès un datagramme UDP. Cette métrique est calculée sur le trafic IP recueilli sur les émetteurs par comparaison des champs de contrôle de séquence (*seqctl*). Le champ *seqctl* est un identifiant identique pour chaque retransmission 802.11 d'un datagramme UDP. Le nombre de retransmissions nécessaires par datagramme UDP est noté NR .

5.4.5 Temps de transmission nécessaire par datagramme UDP

Le temps de transmission nécessaire est le temps nécessaire pour envoyer une trame UDP avec succès. Ce temps prend en compte les multiples retransmissions nécessaires au niveau de la couche MAC 802.11 et correspond à la différence temporelle entre la première transmission 802.11 et la dernière retransmission pour un même paquet UDP. Le temps de transmission nécessaire par datagramme UDP est noté TT .

5.4.6 Temps d'inter-arrivées des trames 802.11

Le temps d'inter-arrivées des trames est le temps mesuré entre les différentes transmissions et retransmissions 802.11 d'un datagramme UDP. Il est calculé à partir du trafic capturé sur le nœud récepteur. Le temps d'inter-arrivées des trames est noté TI .

La métrique associée TIR est le temps d'inter-arrivées mesuré par tentative de transmission 802.11 r sur l'ensemble des traces (e.g. TIR_2 est l'ensemble des temps d'inter-arrivées mesurés entre les 2^{ème} retransmissions des trames et les 1^{ère} retransmissions associées).

5.5 Méthode d'appairage des traces

Deux ensembles de traces sont générés respectivement dans les deux environnements. Afin de comparer les traces, il est nécessaire de les appairer, i.e. d'associer à chacune des traces expérimentales une des traces obtenues par simulation. Il n'existe pas de correspondance immédiate entre les traces des deux ensembles. Par conséquent, la valeur du taux de trames erronées, FER, respective à chaque trace est utilisée pour réaliser l'appairage selon l'algorithme suivant :

Soient X l'ensemble des traces expérimentales et Y l'ensemble des traces de simulations, pour tout élément $x \in X$, il existe une trace $y \in Y$ qui lui est associée, tel que $|\text{Median}(\text{FER}_x) - \text{Median}(\text{FER}_y)| = z$, avec $\forall t \in Y$, $|\text{Median}(\text{FER}_x) - \text{Median}(\text{FER}_t)| > z$.

5.6 Méthodes de comparaison des données

5.6.1 Modèle de Gilbert-Elliott pour la caractérisation des motifs d'erreurs

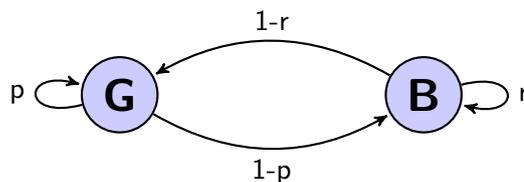


FIGURE 5.2 – Modèle de Gilbert-Elliott.

Le modèle de Gilbert-Elliott [24, 44, 37] présenté sur la figure 5.2, est particulièrement utile pour modéliser les motifs d'erreurs sur les canaux sans-fil.

Le modèle est basé sur une chaîne de Markov cachée (Hidden Markov Model - HMM) constituée des deux états G et B . L'état G correspond à la réception successive de paquets sans-erreur (appelés *intervalles*) tandis que l'état B correspond à la réception successive de paquets avec erreurs, aussi appelées *rafales*. La matrice de transition ainsi associée au modèle est définie par $A = \begin{pmatrix} 1-r & r \\ p & 1-p \end{pmatrix}$ avec $p = P(e_t = B | e_{t-1} = G)$ et $r = P(e_t = G | e_{t-1} = B)$. p et r sont les transitions respectives associées au passage de l'état B vers l'état G et inversement, e_t est l'état du système au temps t . Les probabilités stationnaires π_G et π_B associées aux états G et B sont définies respectivement par (5.10) et (5.11) tandis que la probabilité de pertes π_p est indiquée par l'équation (5.12).

$$\pi_G = \frac{r}{p+r} \quad (5.10)$$

$$\pi_B = \frac{p}{p+r} \quad (5.11)$$

$$\pi_p = p * \pi_G + (1 - r) * \pi_B = \pi_B \quad (5.12)$$

Étant donnée la nature markovienne du canal, il existe une mémoire associée à son comportement [54]. Cette mémoire est définie par le paramètre μ , compris entre -1 et 1, et qui dépend des probabilités de transition entre les états G et B selon la relation (5.13).

$$\mu = 1 - p - r \quad (5.13)$$

Un paramètre μ supérieure à 0 signale un lien dont la mémoire est persistante, i.e. la probabilité de rester dans l'état actuel est la plus importante. Lorsque μ est négatif, la mémoire du canal est dite oscillante, la probabilité de changer d'état est alors plus importante que de rester dans l'état actuel. Une mémoire μ égale à 0 signifie que la probabilité de se trouver dans un état particulier est indépendante de l'état précédent. Les valeurs extrêmes -1 et 1 sont des cas particuliers atteints respectivement pour des liens changeant simultanément d'état ou restant continuellement dans l'état d'origine.

5.6.2 Modèle RCA pour la comparaison d'environnements

Les systèmes d'analyse par cause racine [112, 111] sont conçus pour retrouver les causes originelles d'un événement détecté dans un système de mesure. Dans nos analyses, un modèle RCA est utilisé pour trouver la cause exacte des différences observées sur les traces de mesure issues du banc de test et du simulateur ns-3.

5.6.2.1 Introduction aux modèles d'arbre de comparaison

Pour fonctionner, le système RCA s'appuie sur un modèle constitué d'une base de connaissances où les événements (aussi appelés causes, symptômes ou fermetures) sont associés aux différentes causes possibles les ayant provoquées. Ce système peut-être modélisé sous la forme d'un arbre de comparaison [135, 73] comme celui de la figure 5.3. Dans ce système, chaque nœud est associé à une ou plusieurs métriques, une relation hiérarchique entre deux nœuds de l'arbre signalant une relation entre les métriques respectivement associées aux nœuds dans le système modélisé.

Dans notre modèle, un arbre de déduction est composé de trois types de nœuds : les symptômes, les causes et les fermetures. Les symptômes sont les nœuds non-terminaux de l'arbre. Les causes et les fermetures sont les nœuds terminaux. Chacun de ces nœuds est associé à une expression logique évaluée pendant le parcours

de l'arbre. Il existe une et une seule fermeture issue d'un nœud symptôme. Les fermetures permettent d'inférer la cause possible d'un symptôme même si toutes les valeurs de sa fratrie sont évaluées à *faux*. La valeur associée à une fermeture est ainsi égale à la négation de la disjonction exclusive de l'ensemble des expressions logiques associées aux nœuds frères, i.e. si $V = \{V_1, \dots, V_n\}$ est l'ensemble des valeurs logiques des fils du symptôme S , alors la valeur de la fermeture L associée sera égale à $\neg(V_1 \mid \dots \mid V_n)$.

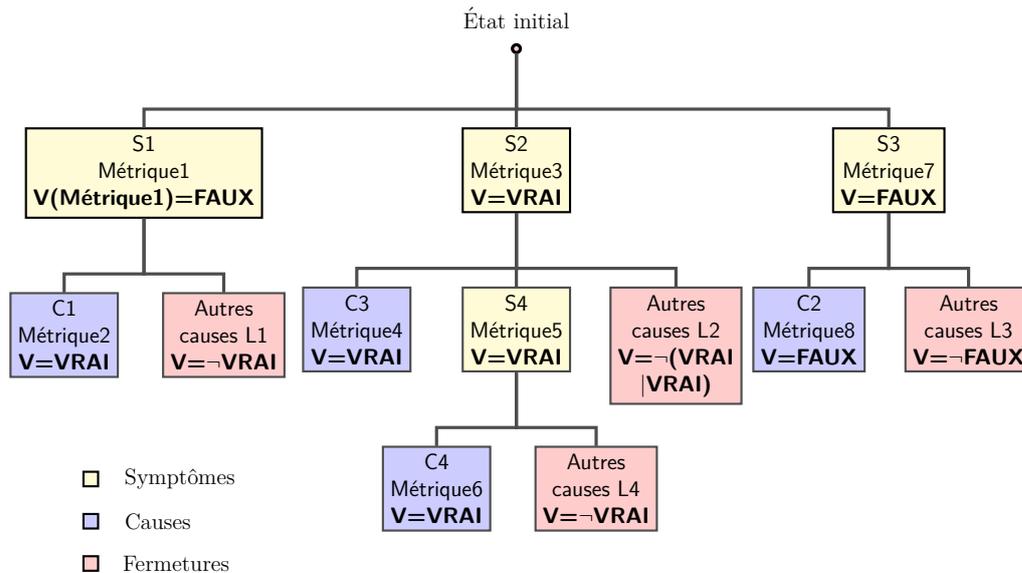


FIGURE 5.3 – Exemple d'arbre de comparaison. Les valeurs V indiquées sur les nœuds sont les valeurs évaluées lors du parcours de l'arbre.

Le parcours de l'arbre débute au nœud initial et se propage par capillarité jusqu'aux nœuds terminaux de l'arbre. Un nœud est évalué seulement si son nœud père est évalué à *vrai*. Ainsi, le parcours de l'arbre décrit par la figure 5.3 est le suivant :

1. Les nœuds de niveau 1, $S1$, $S2$ et $S3$ sont évalués les premiers. Seule la cause $S2$ est vraie
2. Le nœud $S2$ possède 3 causes possibles : $C3$, $S4$ et $L1$, la fermeture qui lui est associée. Les clauses $C3$ et $S4$ sont toutes les deux vraies, par conséquent la cause $L1$ est évaluée à *faux*.
3. Le sous-arbre issu de $S2$ est évalué. La valeur des nœuds $S4$ et $C3$ est *vrai*.
4. Le nœud $C4$ est évalué à *vrai*.

Le parcours de l'arbre a donc permis de détecter des anomalies sur les nœuds $S2$, $C3$, $S4$ et $C4$, il existe alors plusieurs causes racines possible : $C3$ et $C4$.

Dans la pratique, le parcourt de l'arbre est équivalent à la résolution d'une expression logique du 1^{er} ordre composée des valeurs associées aux nœuds de l'arbre. L'expression équivalente au parcourt de l'arbre est donnée par les règles suivantes :

1. La transcription du symptôme S évalué à $V(S)$ et possédant les fils $\{f1, f2, \dots, fn\}$ est donnée par $T(S) \implies V(S) \& (T(f1) | T(f2) | \dots | T(fn) | V(C))$.
2. La transcription de la fermeture C évalué à $V(C)$ sera $T(C) \implies V(C)$

Par conséquent, la transcription de l'arbre exemple est donnée par $T(A_1)$ (5.14).

$$\begin{aligned} T(A_1) \implies & (V(S1) \& (V(C1) | \neg(V(C1)))) | (V(S2) \& (V(C3) | (V(S4) \\ & \& (V(C4) | \neg(V(C4)))) | \neg(V(C3) | V(S4)))) | (V(S3) \\ & \& (V(C2) | \neg(V(C2)))) \end{aligned} \quad (5.14)$$

5.6.2.2 Définition d'un arbre pour la comparaison d'environnements

Les arbres de déduction définis dans la section précédente sont des modèles généralistes applicables sur tout système de mesure. Un modèle d'arbre pouvant être utilisé pour la comparaison d'environnement est présenté en figure 5.4. Sa représentation logique équivalente est donnée par l'équation (5.15). L'utilisation de cet arbre doit permettre, à partir d'un ensemble de mesures issues de deux environnements différents, de trouver les différences existantes et d'en inférer les causes possibles dans la configuration ou l'implémentation des environnements.

$$\begin{aligned} T(A_2) \implies & V(S1) \& (V(S2) \& (V(C1) | V(\neg(V(C3) | V(S3)))) | V(S3) \\ & \& ((V(S6) \& (V(C4) | V(\neg(V(C4)))))) | V(\neg(V(S6) \\ & | V(S4)) | (V(S4) \& (V(C2) | V(C6) | V(\neg(V(C5) | \\ & V(C6)))))) | V(C1) | V(\neg(V(C1) | V(S5) | V(S2))) \\ & | (V(S5) \& (V(C2) | V(\neg(V(C2)))))) \end{aligned} \quad (5.15)$$

Dans ce modèle, chaque nœud est associé à une ou plusieurs caractéristiques présentées dans la section 5.4. Les évaluations logiques des nœuds sont calculées en utilisant les fonctions booléennes D_1 et D_2 qui comparent les valeurs des métriques mesurées sur les traces obtenues en simulation et sur banc de mesure. Les comparaisons se font entre traces appariées. Si $Sim(a)$ est la valeur de la métrique a mesurée en simulation et $XP(a)$, la valeur de la même métrique obtenue expérimentalement, les fonctions D_1 et D_2 sont définies par :

$$1. \quad D_1(a) = \begin{cases} vrai & \text{si } |XP(a) - Sim(a)| > t_a \\ faux & \text{sinon} \end{cases}$$

avec t_a le seuil associé à la métrique a

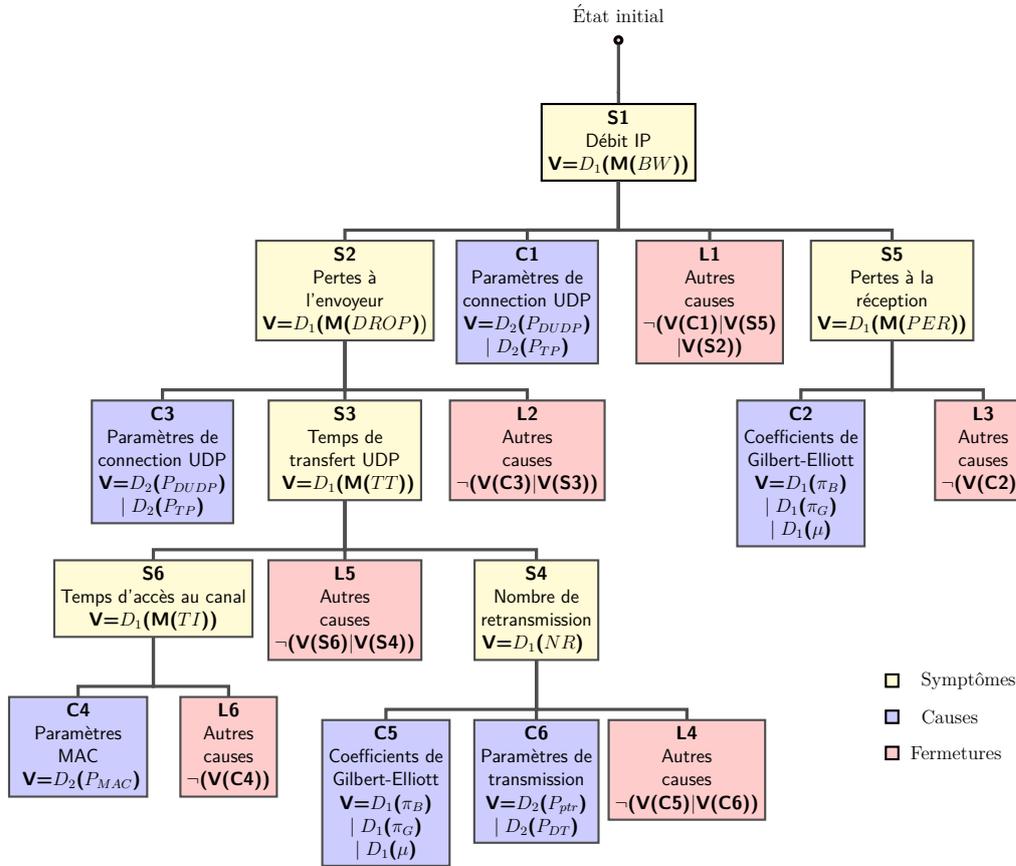


FIGURE 5.4 – Arbre de déduction défini pour la comparaison de traces. La notation $M(a)$ utilisée pour évaluer la valeur des nœuds constitue le médian pour la métrique a (ex : $M(BW)$ est le médian des valeurs de débit mesurées chaque seconde pour chacune des traces).

$$2. \quad D_2(a) = \begin{cases} vrai & \text{si } (Xp(a) \neq Sim(b)) \\ faux & \text{sinon} \end{cases}$$

Description des nœuds

Les nœuds visités en premier lors du parcours de l'arbre conditionnent la visite des nœuds situés aux niveaux inférieurs de l'arbre. Par conséquent, une métrique associée à un nœud situé au niveau n sera jugée plus "globalisante" qu'une métrique associée à un nœud situé au niveau $n + 1$, le but étant de restreindre le champs des anomalies possibles en descendant le long de l'arbre. Pour cette raison, le premier nœud visité utilise Bw , la métrique de mesure de performance du lien. Les variations de débits observés sur un lien sans-fil sont essentiellement dues à des

perdes 802.11 à la réception (métrique FER) et à des pertes UDP à l'émission (métrique $DROP$). Par conséquent le nœud $S1$ est lié aux nœuds $S2$ et $S5$. Le débit reçu est aussi lié aux paramètres des nœuds (débit d'envoi, taille des paquets) et donc la différence de débit entre les deux environnements est liée à ces paramètres (nœud $C1$). Les différences de pertes à la réception (nœud $S5$) sont liées à des différences dans les modèles de bruit généré dans les deux environnements et matérialisé par les coefficients de Gilbert-Elliott (nœud $C2$). Les pertes UDP au niveau de l'émetteur sont principalement la conséquence de congestion du lien sans-fil : le débit généré excède la capacité du lien. Ces congestions sont le résultat de temps de transfert UDP trop importants. Ce temps de transfert est conditionné par deux facteurs : le temps d'accès au canal (nœud $S6$) et le nombre de retransmission 802.11 (nœud $S4$). Le temps d'accès au canal est conditionné par les paramètres MAC (nœud $S6$ et nœud $C4$). Le nombre de retransmissions dépend lui des motifs d'erreurs générées et des paramètres de transmission (nœuds $C5$ et $C6$).

Valeurs de seuil

La fonction $D1$ s'appuie sur l'utilisation d'une valeur seuil différente pour chacun des nœuds, ces valeurs sont choisies arbitrairement selon la métrique mesurée :

- pour le nœud $S1$, le seuil utilisé est noté τ_{bw} . Il correspond à une différence de 500 kbps sur le médian des débits mesurés sur les traces. Cette valeur est choisie arbitrairement et permet d'éviter les faux-positifs dues aux imperfections de mesures.
- le seuil τ_{DROP} utilisé par le nœud $S2$ est fixé à 42 paquets/s et correspond à la perte de 500 kbps choisie pour le nœud $S1$.
- le seuil τ_{FER} fixant le maximum de différence toléré sur les FER est fixé à 1%, correspondant à la valeur de précision maximale atteinte par l'algorithme d'appairage (voir section 5.7.1).
- la valeur du seuil τ_{tt} est une approximation du temps d'envoi de 42 trames de taille 1470 octets envoyées dans le mode de transmission à 54 Mbps. Ce temps correspond à la somme des temps d'accès au médium, de la réception de l'acquiescement et du temps de vol tel que : $42 * DIFS + 42 * SIFS + \frac{42*1470*8}{54*10^6}$. Le temps de temporisation est ici ignoré. Avec les valeurs du DIFS et SIFS respectivement fixée à 28 μs et 10 μs [6], la valeur de τ_{tt} est égale à 0.0091 s.
- τ_{TI} est le seuil fixé pour le temps d'inter-arrivées des paquets. Il correspond à la différence de temps d'inter-arrivées théorique des paquets entre deux flux de 7 Mbps et 6.5 Mbps (selon le seuil de 500 kbps), *i.e.* $\tau_{TI} = \frac{1}{\frac{7*10^6}{1470*8}} - \frac{1}{\frac{7*10^6}{1470*8}} = 0.0002$ s.
- τ_{NR} est le seuil de différence utilisé pour le nombre de retransmissions. Cette valeur est fixée à 1, *i.e.* les valeurs sont considérées comme différentes, si

leur nombre de retransmissions médian est supérieur à 1.

5.7 Application de la méthodologie d'association et de comparaison sur les données générées

5.7.1 Appairage des traces et validation de la configuration par méthode de Gilbert-Elliott

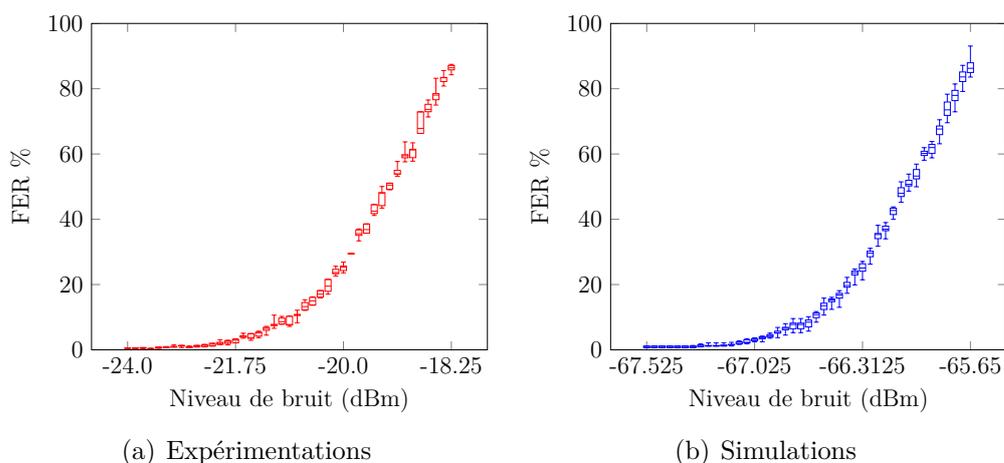


FIGURE 5.5 – Mesures de FER en fonction des différents niveaux de bruits appliqués expérimentalement 5.5(a) et en simulations 5.5(b).

La méthode d'appairage décrite en section 5.5 est utilisée. L'appairage permet d'associer les FER ayant des valeurs médianes distantes de moins de 1% de FER (figure 5.6) à partir des niveaux de FER mesurés expérimentalement (figure 5.5(a)) et en simulation (figure 5.5(b)). Pour une meilleure compréhension dans la suite de ce chapitre, les niveaux de correspondance sont nommés de manière croissante et préfixés par la lettre T (ex : T1, T2, ...). Pour illustrer cette nouvelle notation, un aperçu des associations est affiché dans le tableau 5.2.

Notations après appairage		T1	...	T15	...	T29	...	T43
Niveaux de bruit correspondants (dBm)	expérimentations	-24.0	...	-21.75	...	-20.0	...	-18.25
	simulations	-67.525	...	-67.025	...	-66.3125	...	-65.65

TABLE 5.2 – Notation après le processus d'appairage des traces expérimentales et de simulation. Pour une question de commodité, seule une partie des associations est présentée.

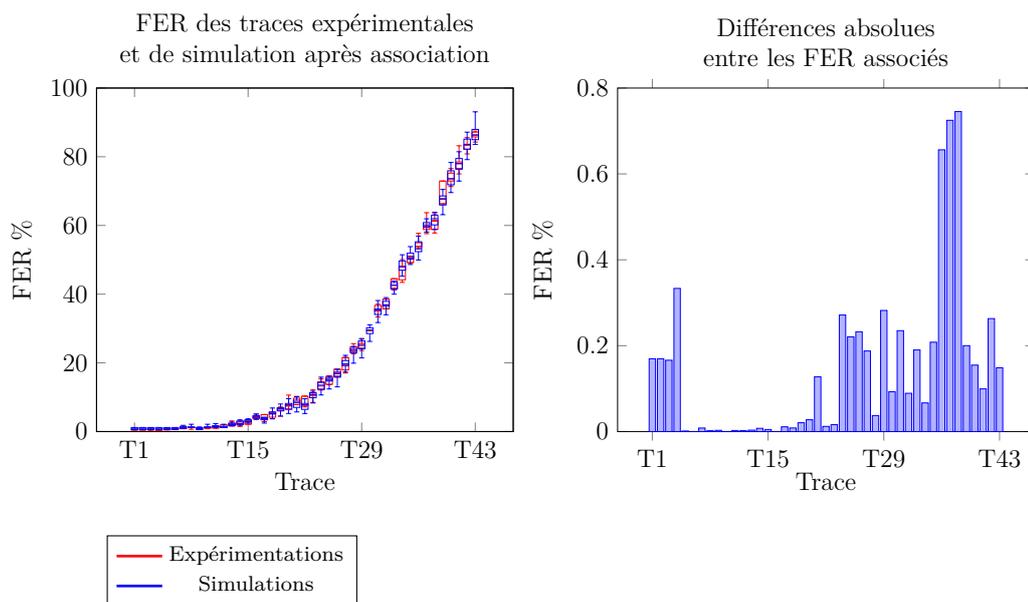


FIGURE 5.6 – Profils des mesures de FER après association et différences absolues mesurées entre les FER des traces associées.

L'appairage associe les traces en fonction du niveau médian de FER. Bien qu'ayant des FER similaires, les motifs d'erreur ou leurs caractéristiques dans les traces associées peuvent être différents. Ces différences ont des impacts sur les communications. Ainsi, la taille des rafales d'erreurs est liée à la valeur du coefficient π_l . Un coefficient important et donc une tendance aux rafales affectera les réseaux sans-perte comme 802.11 car le nombre de retransmissions de trame sera plus important. L'utilisation de l'algorithme BEB qui augmente exponentiellement la taille de la fenêtre de contention entre chaque retransmission successive, aura alors des conséquences sur la capacité du lien. De manière similaire, la valeur du coefficient μ affectera la mémoire et l'indépendance des erreurs du canal. Pour cette raison, un coefficient μ proche de 0 sera privilégié dans les deux environnements.

Par conséquent et pour valider l'appairage des traces au niveau physique, les coefficients de Gilbert-Elliott π_G , π_B et μ sont calculés pour chacune des traces. L'évolution des coefficients est présentée sur la figure 5.7.1.

Comme le montrent les figures, il n'existe pas de différence entre les motifs d'erreurs mesurés expérimentalement et en simulation. De même, les évolutions du coefficient μ calculé sur les traces sont similaires pour les deux environnements et ses valeurs proches de 0, démontrant par là-même l'indépendance des erreurs générées.

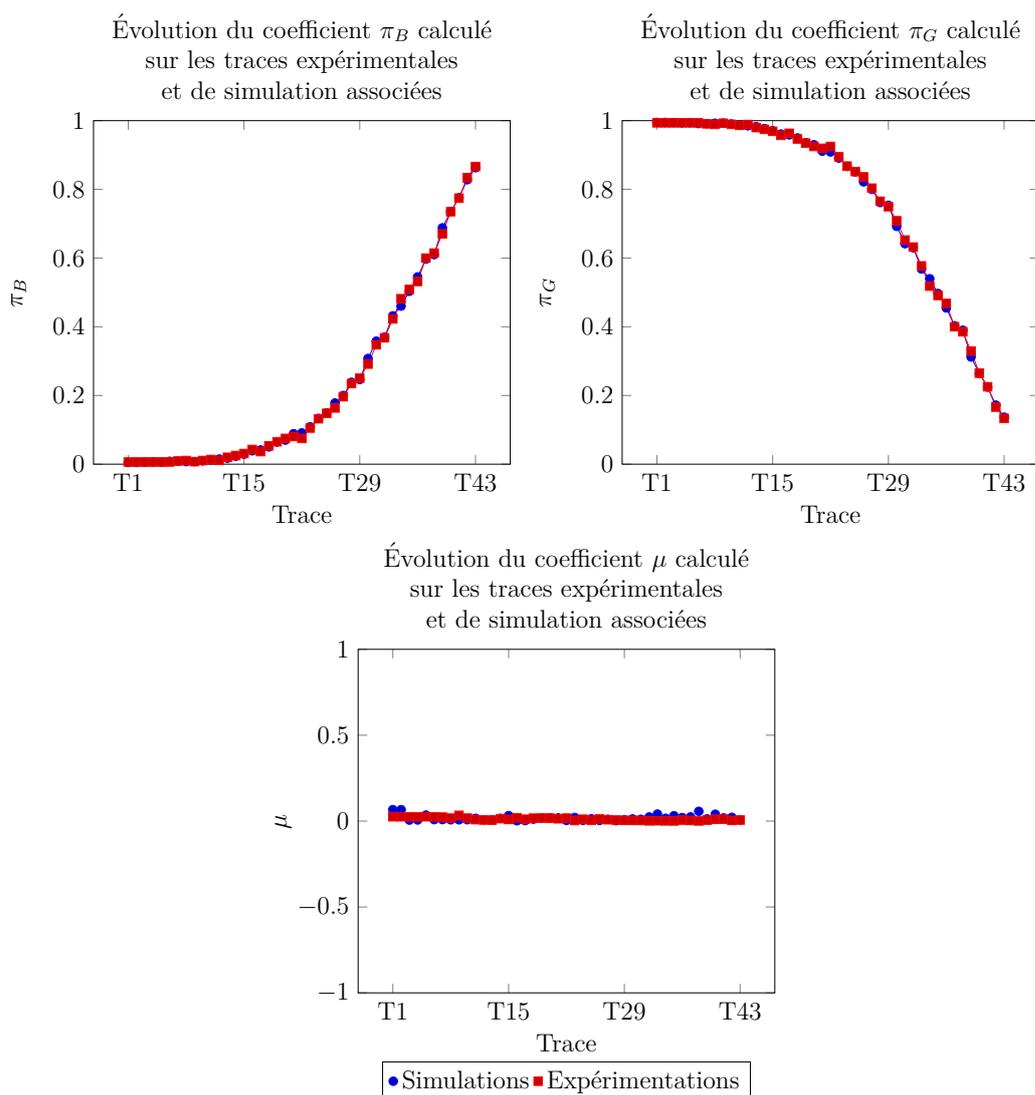


FIGURE 5.7 – Evolutions des différents coefficients de Gilbert-Elliott mesurés en simulation et expérimentalement

5.7.2 Application du modèle RCA

Le modèle présenté sur la figure 5.4 est appliqué sur les différentes traces. Les résultats de cette application sont indiqués dans la table 5.3.

Comme décrit dans la section 5.3, les paramètres appliqués pour les mesures expérimentales et les simulations sont les mêmes. Par conséquent les valeurs des nœuds $C1$, $C3$, $C4$, $C6$ qui comparent les paramètres initiaux sont évaluées à *faux*. De manière similaire, si les nœuds $C2$, $C5$ qui comparent les coefficients GE sont visités, leur valeur sera évaluée à faux étant donnés les résultats illustrés

Nom du nœud	Expression logique associée	Évaluation
S1	$D_1(M(BW))$	<i>vrai</i>
S2	$D_1(M(DROP))$	<i>vrai</i>
C1	$D_2(P_{DUDP}) D_2(P_{TP})$	<i>faux</i>
L1	$\neg(V(C1) V(S5) V(S2))$	<i>faux</i>
S5	$D_1(M(FER))$	<i>faux</i>
C2	$\neg(V(C1) V(S5) V(S2))$	non-évalué
L3	$\neg(V(C2))$	non-évalué
C3	$D_2(P_{DUDP}) D_2(P_{TP})$	<i>faux</i>
L2	$\neg(V(C3) V(S3))$	<i>faux</i>
S3	$D_1(M(TT))$	<i>vrai</i>
S6	$D_1(M(TI))$	<i>vrai</i>
L5	$\neg(V(S6) V(S4))$	<i>faux</i>
S4	$D_1(NR)$	<i>faux</i>
C4	$\neg(V(C1) V(S5) V(S2))$	non-évalué
L4	$\neg(V(C5) V(C6))$	non-évalué
C6	$D_1(\pi_B) D_1(\pi_G) D_1(\mu)$	non-évalué
C4	$D_2(P_{MAC})$	<i>faux</i>
L6	$\neg(V(C4))$	<i>vrai</i>

TABLE 5.3 – Évaluation logique des différents nœuds de l'arbre de déduction.

en section 5.7.1. Enfin la valeur du nœud $S5$ est également évaluée à *faux* étant donnée la procédure d'appairage de traces basée sur le FER qui implique une différence inférieure au seuil τ_{DROP} de 1% d'erreur.

Le premier nœud parcouru $S1$ compare les débits mesurés dans les deux environnements. Ces valeurs sont présentées sur la figure 5.8(a), tandis que la différence des valeurs mesurées entre les traces appairées des deux environnements sont indiquées sur la figure attenante. Cette dernière permet d'évaluer la valeur de $S1$ à *vrai*, étant données les valeurs supérieures au seuil τ_{bw} . Les valeurs de débit restent stables pour les niveaux de bruits inférieurs et correspondant à moins de 2,30% de FER (traces $T1$ à $T17$). A partir de la trace $T18$, le niveau de bruit est assez important pour affecter le débit, une légère différence est mesurée entre les deux traces. Cette différence s'accroît de manière importante après la trace $T33$ (25 % de FER) et atteint un maximum proche de 2 Mbps de débit pour la trace $T35$.

Au deuxième niveau de comparaison, les nœuds $S2$, $C1$, $L1$ et $S5$ sont visités. Le nœud $S2$ est évalué à *vrai* comme démontré sur la figure 5.8(b) : les différences de pertes suivent le même motif que celui des différences de débits entre les ensembles de traces. Les données présentées sur la figure démontrent que la congestion issue de la saturation du médium par l'envoyeur intervient pour des perturbations plus faibles en simulation. Le nombre de pertes par seconde atteint

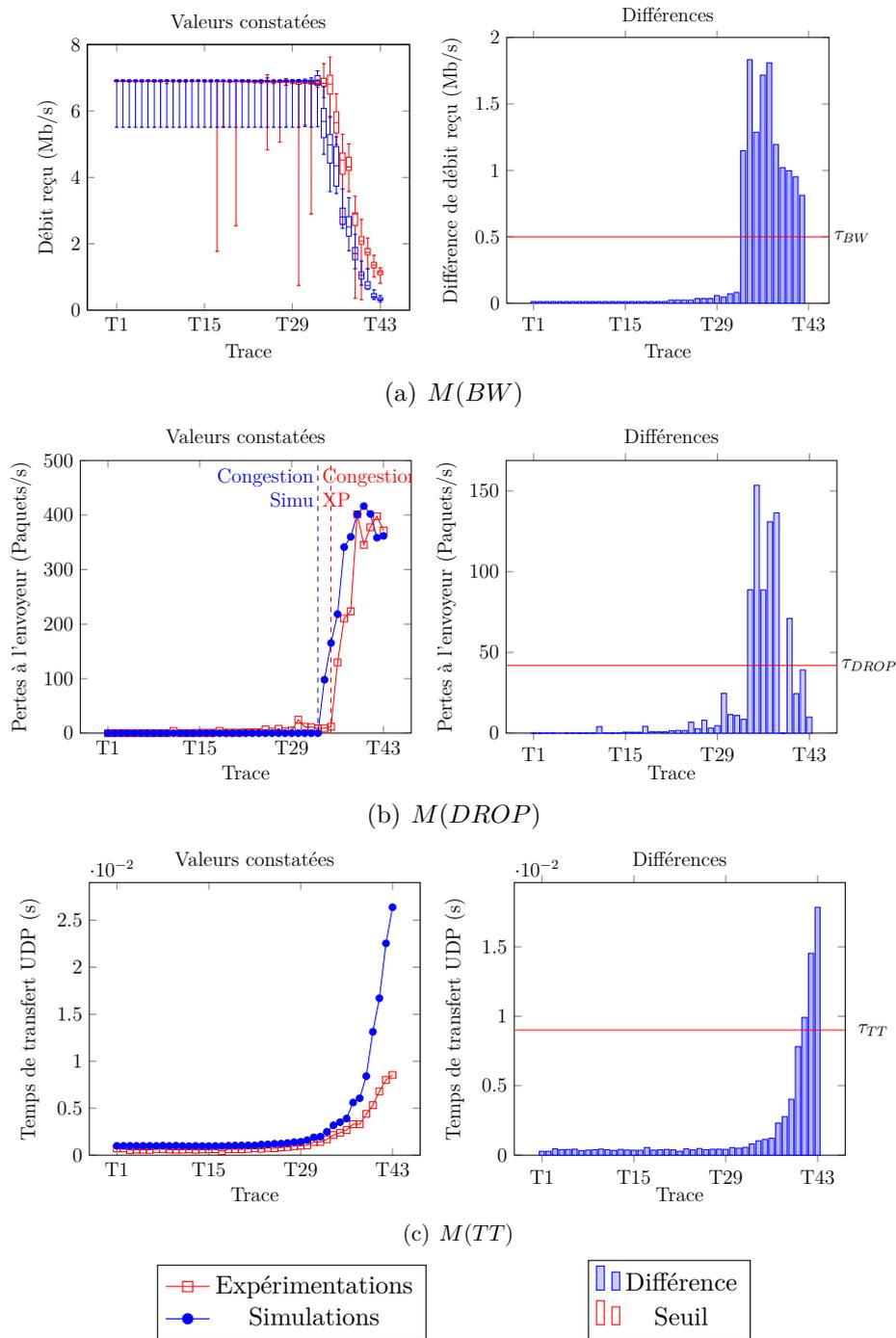


FIGURE 5.8 – Comparaisons des données de simulation et expérimentales pour les métriques BW (a), $DROP$ (b) et TT (c).

finalement un plateau à 400 paquets par secondes pour les deux environnements. La différence de pertes observée sur les traces est mesurée à 150 paquets perdus

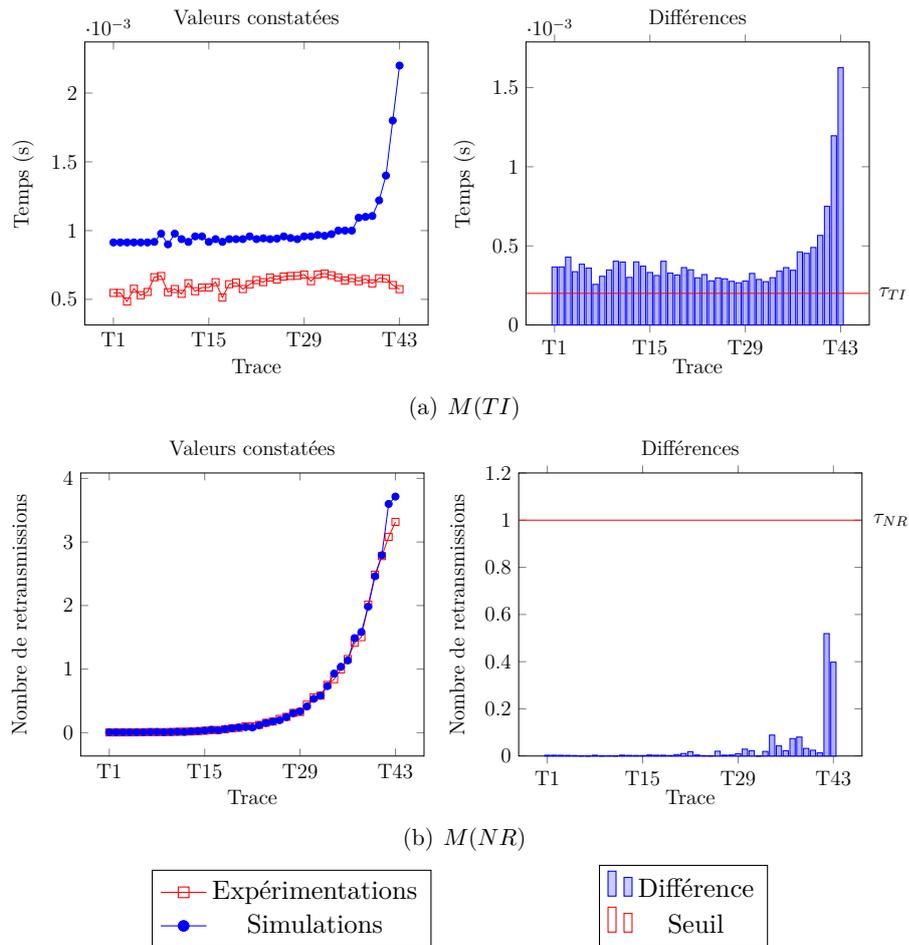


FIGURE 5.9 – Comparaisons des données de simulation et expérimentales pour les métriques TI (a) et NR (b).

par secondes ce qui équivaut à la différence de débit de 2 Mbps observée lors de l'évaluation du nœud $S1$. Le troisième niveau de l'arbre permet de déterminer la cause des dissimilarités sur les pertes de paquets. En effet, le nœud $S3$ qui compare les différences entre les délais de transfert moyen est évalué à *vrai*. Ces temps de transfert plus élevés pourraient être causés par un temps d'accès au lien plus important comme le montre la figure 5.9(a). Sur les données affichées, les temps d'inter-arrivées expérimentaux sont inférieurs au temps de simulation. À partir de la trace $T30$, les temps obtenus en simulation explosent alors que les temps expérimentaux semblent diminuer très légèrement. Finalement le nœud terminal $L6$ est évalué à *vrai*.

L'ensemble du parcours de l'arbre est résumé sur la figure 5.10.

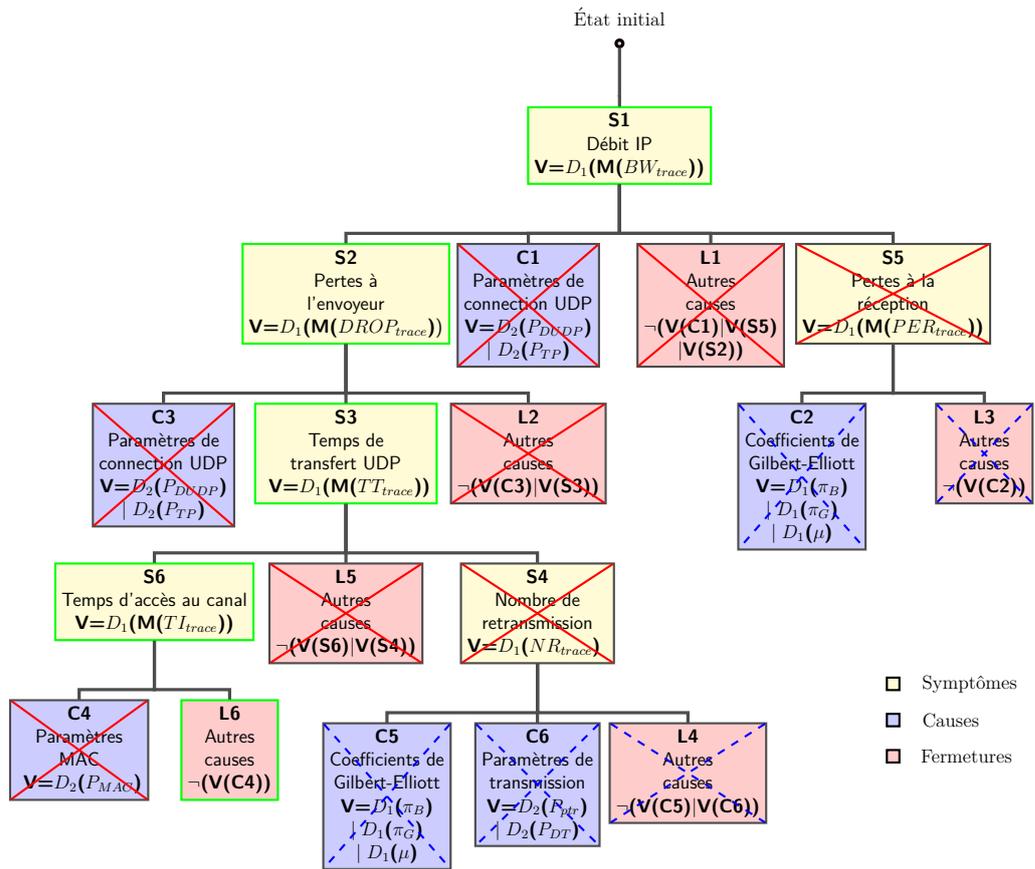


FIGURE 5.10 – Arbre de déduction pour la comparaison de traces avec indication de parcours lors de l'évaluation des nœuds. Les nœuds non barrés sont les nœuds évalués comme vrais, les nœuds barrés par des lignes entières sont évalués comme faux et les nœuds barrés par des tirets ne sont pas évalués pendant le parcours de l'arbre.

5.8 Analyse des résultats expérimentaux

5.8.1 Supplément de diagnostic

La section précédente met en lumière les différences existantes entre les traces de simulation et les traces expérimentales et infère la cause de ces différences aux procédures d'accès au médium. Selon les paramètres affichés dans la table 5.1, les configurations des environnements sont similaires.

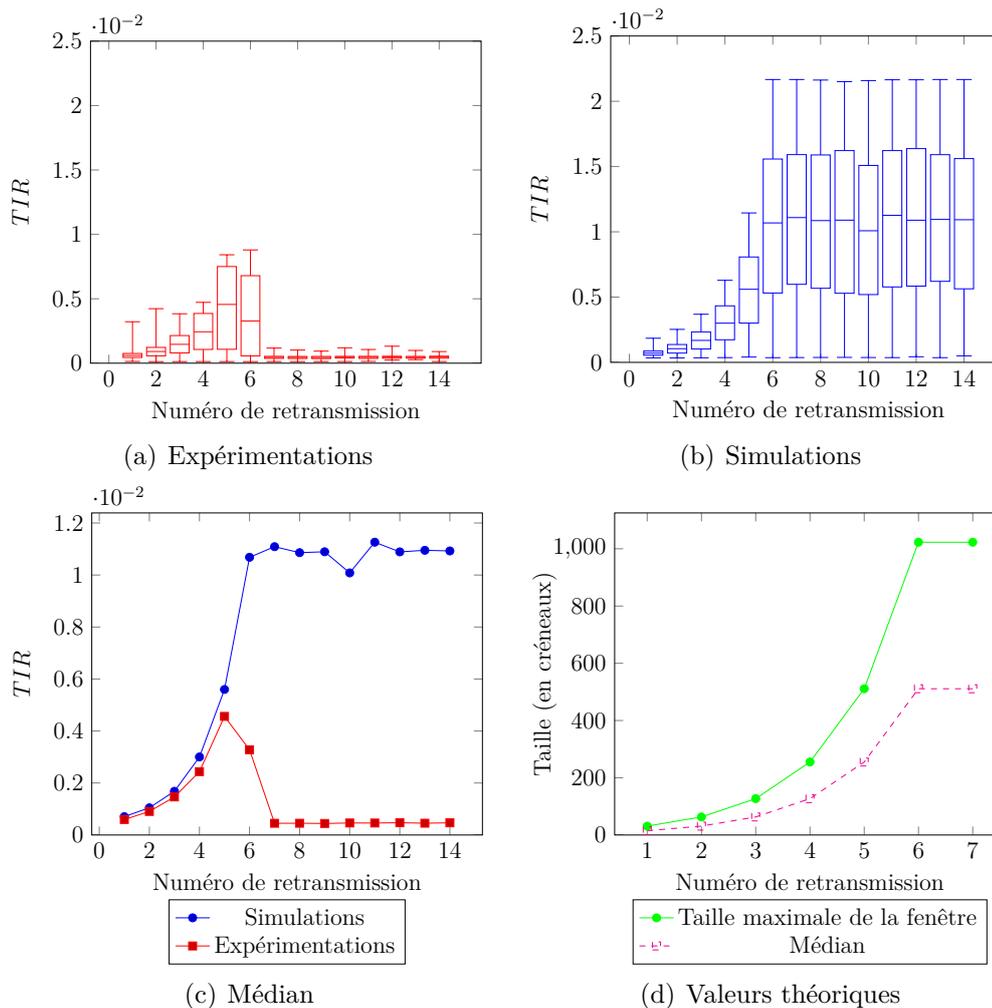


FIGURE 5.11 – Valeurs de TIR mesurées expérimentalement et en simulation.

Les données de la figures 5.11(a) et 5.11(b) complètent celles des figures 5.8 et 5.9. Elles représentent le temps d'inter-arrivées TIR des paquets pour chaque niveau de retransmission 802.11, le médian des valeurs est affiché sur la figure 5.11(c). La figure 5.11(d) présente les valeurs théoriques obtenues par l'algorithme

BEB selon l'équation (5.16). Sur cette dernière figure, les unités sont exprimées en nombre de créneaux.

$$Cw_{n+1} = \max((CW_n + 1) * 2 - 1, 1023) \quad (5.16)$$

avec $Cw_0 = 15$

Les résultats obtenus en simulation semblent proches des résultats théoriques : les valeurs suivent une ascension exponentielle entre chaque retransmission et se stabilisent à la 6^{ème} retransmission. Un plateau est ensuite observé. En revanche, les résultats expérimentaux présentent des anomalies tant vis à vis des résultats en simulation que des résultats théoriques à partir de la 5^{ème} retransmission. Ces résultats confirment ceux obtenus dans [23] qui mettent en évidence les implémentations non standard et déviantes des mécanismes de temporisation de différentes cartes WIFI.

L'apparition de ce phénomène à de hauts-niveaux de retransmission semble concorder avec la différence de débits constatée sur la figure 5.8(a) à partir de la trace T30 correspondant à près de 20% de FER.

5.8.2 Modélisation des paramètres BEB

Un des reproches adressés aux simulateurs et leur manque de réalisme concernant les spécificités des matériels utilisés. Si le simulateur est utilisé pour prototyper un protocole ou un algorithme, les résultats finaux obtenus après mise en œuvre expérimentale peuvent se révéler décevants. Par conséquent, la modélisation de l'algorithme de temporisation expérimental et son implémentation en simulation peuvent permettre d'améliorer son réalisme.

5.8.2.1 Facteurs environnementaux

Pour quantifier l'effet de l'environnement sur le fonctionnement de l'algorithme de temporisation initial, les notions de facteur expérimental et facteur de simulation sont respectivement définis. En effet, il est difficile de convertir les unités utilisées dans le modèle théorique (le créneau) en unités de temps mesurés sur les données expérimentales et de simulation. De plus, sur tout système de mesure, les données mesurées et effectives sont le résultat du modèle théorique mis en œuvre et de l'effet du système et de l'environnement selon la relation $Valeurs_Theoriques * Facteur_environnementale = Valeurs_Mesurees$.

Par conséquent, le facteur environnemental et le facteur de simulation sont respectivement définis par α_n et β_n pour un numéro de retransmission n donné. La relation qui lie la taille de la fenêtre de contention théorique et le temps mesuré pour chaque retransmission peut être notée $\alpha_n * Cw_n = M(XP(TIR_n))$ pour les valeurs expérimentales et $\beta_n * Cw_n = M(Simu(TIR_n))$ pour les simulations.

Dans le cas d'environnements équivalents et en supposant que le standard 802.11 soit identique, l'équation (5.17) est vérifiée et implique la relation (5.18).

$$M(XP(TIR_n)) \equiv M(Simu(TIR_n)) \quad (5.17)$$

$$\text{et donc } \alpha_n * Cw_n \equiv \beta_n * Cw_n \quad (5.18)$$

Dans le cas inverse, la relation (5.17) n'est pas vérifiée. On a donc la relation (5.19). Cette équation correspond au cas actuel de nos données.

$$\alpha_n * Cw_n \not\equiv \beta_n * Cw_n \quad (5.19)$$

L'objectif de la modélisation entreprise et d'obtenir l'équivalence à partir de cette équation. Une première solution consiste à agir sur les différents facteurs environnementaux (environnements de propagation, ...). Cette solution requiert des modifications complexes. Une deuxième solution plus simple consiste à modifier les valeurs Cw_n utilisées dans le modèle de simulation en intégrant les différences environnementales entre les deux environnements. Si $F(Cw_n)$ est la valeur de Cw_n prenant en compte cette différence, la relation (5.20) est alors vérifiée. Par transformations successives, la relation (5.21) permet alors de calculer les valeurs de $F(Cw)$ pour chaque n . Ces valeurs sont données dans la table 5.4 conjointement aux valeurs des facteurs α et β .

$$\beta_n * F(Cw_n) \equiv \alpha_n * Cw_n \quad (5.20)$$

$$\text{d'où } F(Cw_n) = \frac{M(Xp(TIR)) * Cw_n}{M(Simu(TIR))} \quad (5.21)$$

Numéro de retransmission n	α_n	β_n	Cw_n	$F(Cw_n)$
0	4.7e-05	4.0e-05	15	13
1	3.4e-05	2.9e-05	31	27
2	2.7e-05	2.3e-05	63	55
3	2.4e-05	1.9e-05	127	104
4	2.2e-05	1.8e-05	255	208

TABLE 5.4 – Valeurs des coefficients et des fenêtres de contention pour les différents niveaux de retransmission.

En prenant en compte ces modifications, l'algorithme de temporisation permettant l'équivalence se définit alors par la relation (5.22) avec les paramètres k et m .

$$F(Cw_{n+1}, k) = \max((F(CW_n, k) + 1) * k - 1, m). \quad (5.22)$$

5.8.2.2 Régression et calcul des paramètres

À partir des valeurs du tableau 5.4, il est aisé de calculer les paramètres k et m optimums.

Dans un premier temps, le paramètre k est calculé exhaustivement en comparant les différentes valeurs $F(CW_n)$ obtenues par l'équation (5.22) avec les valeurs de la table 5.4. La valeur optimale est celle dont la somme des différences obtenue par la relation $\sum_1^{n=4} |F(Cw_n, k) - F(Cw_n)|$ est la plus faible.

Les paramètres ainsi testés varient de 1.55 à 2.2 par pas de 0.55. Les 5 meilleurs résultats sont présentés dans la table 5.5 et impliquent une valeur optimale de 1.88.

Classement	k	$Cw_{k,1}$	$Cw_{k,2}$	$Cw_{k,3}$	$Cw_{k,4}$	Différence cumulée
1	1.88	29.08	55.55	105.31	198.87	16.57
2	1.94	30.04	59.22	115.82	225.63	40.61
3	1.83	28.28	52.58	97.06	178.44	41.5
4	1.77	27.32	49.13	87.72	156.04	75.73
5	1.99	30.84	62.36	125.09	249.92	78.11

TABLE 5.5 – Meilleurs résultats obtenus pour la recherche du paramètre k optimum.

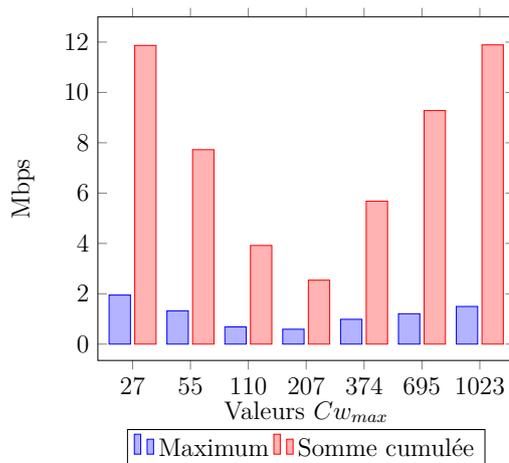


FIGURE 5.12 – Sommes cumulées et maximum des différences entre les débits mesurés en simulation et en expérimentation en fonction de la taille maximale de la fenêtre de contention Cw_{max} .

Dans un deuxième temps, la valeur optimale de m est elle aussi obtenue de manière exhaustive en calculant les différents débits obtenus en simulation pour différents m avec k fixé à 1.88. Les résultats des mesures de différence de débit en termes de somme cumulée et de maximum sont affichés sur la figure 5.12. Les meilleures résultats sont obtenus pour le paramètre m égal à 207.

5.8.2.3 Vérification

Afin de tester l'optimalité des valeurs k et m calculées dans la section précédente et vérifier les résultats de notre méthode de modélisation, l'ensemble des couples constitués des valeurs de k et m sont testés par simulations. Les 5 meilleurs résultats sont donnés sur la table 5.6.

<i>Classement</i>	k	m	<i>Différence cumulée</i>
1	1.83 (3)	207	2.28
2	1.77 (4)	207	2.47
3	1.88 (1)	207	2.54
4	1.61 (9)	374	2.69
5	1.72 (6)	207	2.79

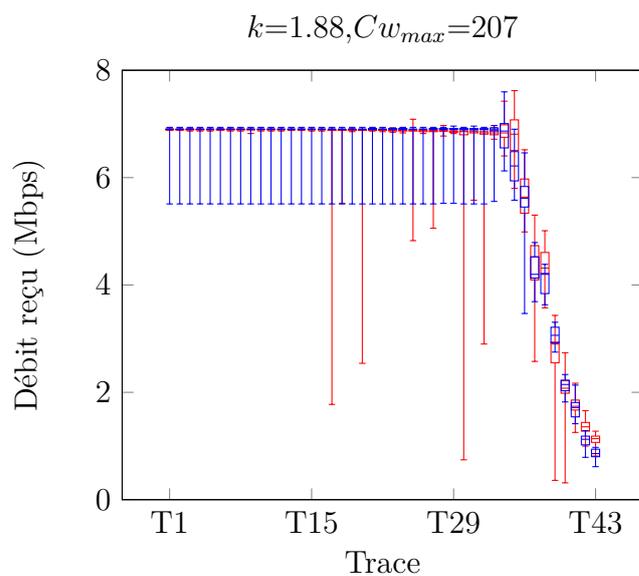
TABLE 5.6 – Valeurs de k et m utilisées pour les 5 meilleurs résultats des tests. La valeur entre parenthèses est le classement du paramètre k dans la section précédente.

Le tableau confirme les résultats obtenus précédemment. La valeur de k classée première est présente en 3^{ème} position dans la table 5.5 tandis que la seconde valeur était classée 4^{ème}. La valeur 1.88 trouvée optimale dans la section précédente est classée 3^{ème}. Enfin, 4 des 5 couples de valeurs utilisent 207 comme taille maximale de fenêtre de contention.

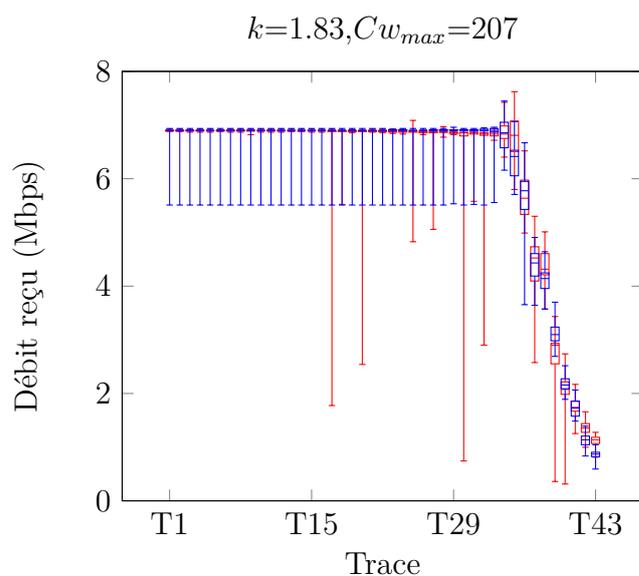
5.8.3 Améliorations obtenues

Les débits obtenus après modification de l'implémentation BEB du simulateur en utilisant les paramètres optimums respectivement obtenus en section 5.8.2.2 et en section 5.8.2.3 sont présentés sur la figure 5.13. La figure 5.14 présente les différences entre les valeurs de simulation et la valeur expérimentale.

Les résultats montrent une amélioration majeure : alors qu'avec la configuration initiale, les différences entre les simulations et les expérimentations s'approchent des 2 Mbps, après modification cette différence maximale n'est que de 500 kbps. Le modèle de simulation MAC est donc plus proche du modèle expérimental. Ces résultats confirment la responsabilité de l'anomalie observée sur l'algorithme de temporisation mesuré expérimentalement dans la différence de débit observée sur la figure 5.8(a).



(a)



(b)

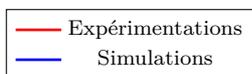


FIGURE 5.13 – Résultats des débits obtenus avec les valeurs optimales.

5.9 Conclusion

Ce chapitre a traité de la validation et de l'amélioration des différents modèles WIFI aux niveaux des couches MAC et physique mises en œuvre dans le simulateur

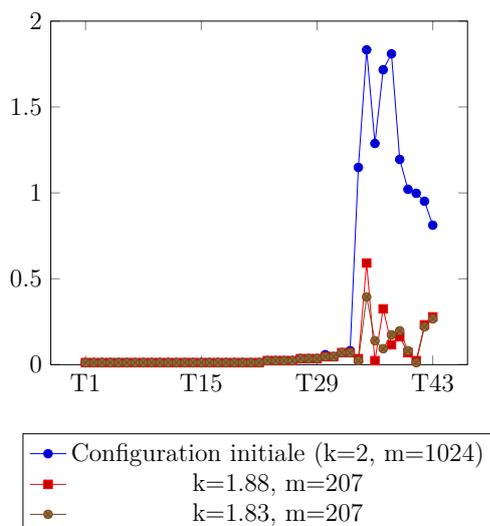


FIGURE 5.14 – Différences observées entre les débits obtenus avec les différentes configurations d’algorithmes de temporisation et le débit expérimental.

ns-3. En effet, malgré les critiques qui sont adressées aux simulateurs concernant leur manque de réalisme, ceux-ci sont encore couramment utilisés par la communauté. Par conséquent, afin d’améliorer le réalisme des modèles, des méthodes empiriques de validation des simulateurs sont parfois mises en œuvre pour comparer les résultats des simulations avec ceux des réseaux réels. Dans cette optique, ce chapitre propose un modèle d’analyse par cause racine (RCA) permettant la comparaison de mesures issues de traces respectivement obtenues sur le simulateur ns-3 et sur le banc expérimental décrit au chapitre 4. Ce modèle a été appliqué sur des données générées dans les deux environnements à partir d’un scénario commun et implémenté de manière similaire. Les résultats ont montré une différence dans la mise en œuvre par les cartes Atheros du mécanisme BEB utilisé pour gérer la durée de temporisation lors de l’accès au lien. Cette différence engendre des écarts de performance entre le simulateur ns-3 et le banc expérimental qui peuvent atteindre 2 Mbps dans le scénario expérimental considéré. Les analyses détaillées dans ce chapitre ont permis de modéliser l’algorithme BEB correspondant au modèle expérimental pour des valeurs de k et m respectivement égales à 1.8 et 207 alors que dans la version théorique et celle du simulateur celles-ci sont égales à 2 et 1024. L’application du modèle expérimental a alors permis d’atteindre des performances similaires expérimentalement et en simulation.

Ces résultats démontrent que les implémentations des standards peuvent différer de manière importante selon les constructeurs et selon les modèles de cartes radio. Par conséquent, les simulateurs et leurs modèles doivent être comparés vis-à-vis de différents matériels par l’intermédiaire de méthodes automatisées et génériques.

En fonction des besoins, plusieurs modèles d'un même standard pourraient être fournis par un simulateur.

Le chapitre suivant utilisera les données expérimentales pour mettre en œuvre et évaluer l'application de méthodes d'apprentissage en vue de la prédiction de performances d'un lien sans-fil bruité.

6

Estimations prédictives de performance d'un lien sans-fil par la mesure RF et l'utilisation d'algorithmes d'apprentissage

Sommaire

6.1	Introduction	90
6.2	Contributions	91
6.3	Algorithmes d'apprentissage automatique pour la régression	92
6.3.1	Méthode SVR	92
6.3.2	Méthode k -nn	93
6.3.3	Méthode DT	94
6.4	Données expérimentales	95
6.4.1	Protocole expérimentale pour la génération de données	95
6.4.1.1	Constitution des jeux de données	96
6.4.2	Caractéristiques de mesure de performance et du canal sans fil	97
6.4.2.1	Caractéristique de performance	97
6.4.2.2	Caractéristiques de mesure physiques du canal	97
6.4.3	Notations alternatives des caractéristiques	98
6.5	Méthodes d'évaluation de performances	98
6.5.1	Méthodes d'évaluation de la précision des estimations	98
6.5.1.1	Erreur quadratique absolue	98
6.5.1.2	Pourcentage de 'bonnes estimations'	99
6.5.2	Méthodes d'évaluation de la pertinence des caractéristiques de mesure physique	99
6.5.3	Méthodes d'évaluation de performance des algorithmes	100

6.5.3.1	Efficacité en temps de calcul	100
6.5.3.2	Efficacité mémoire	100
6.6	Méthodologie de mesure et d'apprentissage	100
6.6.1	Configuration des algorithmes de ML	100
6.6.1.1	SVR	100
6.6.1.2	k -NN	101
6.6.1.3	DT	101
6.6.2	Mesure de performance des algorithmes	101
6.6.2.1	Mesure des délais d'apprentissage et d'estimation	101
6.6.2.2	Mesure de la taille du modèle	101
6.7	Résultats	101
6.7.1	Précision des estimations	101
6.7.2	Pertinence des caractéristiques de mesure RF	105
6.7.3	Fonctionnement des algorithmes	107
6.7.3.1	Efficacité en temps de calcul	107
6.7.3.2	Efficacité mémoire	109
6.8	Conclusion	109

6.1 Introduction

Dans le cadre des communications sans-fil, le médium constitue une des principales sources d'erreurs et de chutes de performance. Celles-ci peuvent apparaître de manière incertaine et aléatoire [89, 97]. Par conséquent, prédire l'évolution du médium même à court terme peut permettre d'améliorer les performances du réseau et d'offrir aux nœuds de nouvelles possibilités d'adaptation. Dans cette optique, ce chapitre propose l'application de méthodes d'apprentissage automatique pour l'estimation de débit reçu au niveau de la couche IP à partir de mesures effectuées sur le nœud récepteur d'un lien sans-fil.

Les trois algorithmes choisis pour cette tâche sont fréquemment rencontrés dans la littérature : la régression par support vecteur SVR [20], la méthode des plus proches voisins k -nn [14] et les arbres de décision DT [28]. Les données utilisées sont générées à partir du banc de test décrit dans le chapitre 4, sur un lien TCP dont la réception est perturbée par du bruit. Pour effectuer les estimations, différentes caractéristiques de niveau physique sont extraites des données enregistrées sur le nœud récepteur. Pour que les résultats de cette étude soient reproductibles sur du matériel standard, les équipements utilisés sont constitués d'équipements WIFI génériques et d'un oscilloscope¹. Ces caractéristiques sont le RSS, le SNR ou encore

1. Bien que les caractéristiques mesurées soient disponibles sur du matériel générique 802.11, la configuration de nos équipements au moment des mesures ne nous permettait pas de mesurer efficacement et simultanément toutes les métriques envisagées. Par conséquent un oscilloscope a été utilisé en complément ou en alternative de l'équipement WIFI.

le bruit.

Afin d'évaluer l'efficacité des différents algorithmes, leurs performances respectives sont comparées selon différents critères. Le premier critère est la précision des estimations mesurée selon deux méthodes : l'erreur quadratique moyenne (MSE) et le pourcentage de bonnes estimations. Le deuxième critère est l'efficacité des algorithmes en termes de calcul, évalué par le ratio temps/précision. Enfin, le troisième critère est l'efficacité mémoire caractérisée par le ratio taille du modèle/précision.

L'efficacité d'une estimation dépend des caractéristiques utilisées. Par conséquent la pertinence de chacune d'entre elles est évaluée selon la performance atteinte, la métrique la plus pertinente permettant les estimations les plus précises.

Par conséquent, le plan suivi dans ce chapitre est le suivant. Dans un premier temps, les aspects théoriques des algorithmes SVR, k -nn et DT sont expliqués en section 6.3. La section 6.4 décrit les protocoles, les méthodes de mesure et les données expérimentales utilisées pour les analyses. Il détaille notamment les caractéristiques utilisées pour les estimations. La 3^{ème} partie définit les méthodes d'évaluation des algorithmes : précision des estimations et performance temps et mémoire, ainsi que la méthode d'évaluation de la pertinence des caractéristiques. Les paramètres et configurations utilisés pour les analyses sont donnés en section 6.6. Enfin, la section 6.7 traite les résultats obtenus. Pour une question d'espace, les graphiques de l'ensemble des estimations obtenues sont contenus en annexe A.

6.2 Contributions

Le travail décrit dans ce chapitre se rapproche des travaux de radio cognitive dont le but est la prédiction de certaines caractéristiques du médium sans-fil [35, 34, 102]. En revanche, nos travaux divergent sur certains points :

- les caractéristiques utilisées dans notre travail sont disponibles sur du matériel générique 802.11 actuel bien configuré contrairement à celles utilisées dans les travaux de radios cognitives et qui nécessitent pour la plupart l'usage de radios logicielles. Les radios logicielles sont des matériels coûteux et peu utilisés par la grande majorité des utilisateurs contrairement au matériel générique 802.11.
- les travaux cités ont pour objectifs de prédire l'occupation du spectre. L'application détaillée ici vise à estimer la performance d'un lien à partir de donnée de niveau physiques, *i.e.* à trouver le meilleur indicateur de la qualité d'un lien en vue de la prédiction. À notre connaissance, seul [128] utilise les méthodes d'apprentissage dans cet objectif en utilisant principalement des informations de niveau MAC. La majorité des travaux de caractérisation comme [129, 138, 61] utilisent quand à eux des méthodes statistiques ne mettant pas en œuvre le principe d'apprentissage.

6.3 Algorithmes d'apprentissage automatique pour la régression

Nous présentons dans cette section les algorithmes d'apprentissage automatique SVR, k -nn et DT utilisés dans l'étude.

6.3.1 Méthode SVR

La méthode SVR [115, 20, 105] est une forme particulière de machine à vecteurs supports pour la génération de modèles de prédiction.

En prenant comme exemple l'application de modélisation des données contenues dans l'ensemble $E = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{X} \times \mathbb{R}$, avec \mathbb{X} l'espace d'entrée, l'objectif de SVR est d'estimer une fonction $f(x)$ avec comme contrainte principale une déviation maximale inférieure à ϵ des valeurs à estimer y_i . On distingue alors deux cas selon la nature des données à modéliser : dans le cas où les données sont de nature linéaire, la régression s'effectue directement dans l'espace de départ (équation (6.1)). Dans le cas inverse, cette régression n'est pas possible. Par conséquent, dans un premier temps, l'algorithme SVR projette les données dans un espace caractéristique \mathbb{F} via une fonction $\phi : \mathbb{X} \rightarrow \mathbb{F}$ (équation (6.2)). Dans un second temps, l'algorithme de régression traditionnel SV est appliqué dans le nouvel espace caractéristique.

$$f(x) = \langle w, x \rangle + b \text{ avec } w \in \mathbb{X} \text{ et } b \in \mathbb{R}. \quad (6.1)$$

$$f(x) = \langle w, \phi(x) \rangle + b \text{ avec } w \in \mathbb{X} \text{ et } b \in \mathbb{R}. \quad (6.2)$$

avec $\langle \cdot, \cdot \rangle$ la notation du produit scalaire

La contrainte secondaire concernant la régression est la maximisation de l'uniformité des poids, ici mesuré par $\|w\|^2$. Ainsi, dans le cas non-linéaire, les coefficients w et b sont estimés par minimisation de la fonction du risque régularisé donnée par la relation (6.4). Dans cette équation, C est une valeur définie par l'utilisateur qui contrôle le compromis entre l'erreur d'apprentissage et l'uniformité du modèle. L_ϵ est la fonction de perte ϵ -insensible définie par l'équation (6.3) et sujet à $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$ avec $\alpha_i^* \in [0, C]$. Grâce à cette fonction, seules les erreurs d'estimation supérieures à ϵ sont pénalisées.

$$L_\epsilon(y_i, f(x(i), w)) = \begin{cases} |y_i - f(x(i), w)| - \epsilon & \text{si } |y_i - f(x(i), w)| \geq \epsilon. \\ 0 & \text{autrement.} \end{cases} \quad (6.3)$$

$$R(f, C) = C \sum_{i=1}^n L_\epsilon(y_i, f(x(i), w)) + \frac{1}{2} \|w\|^2. \quad (6.4)$$

La minimisation de l'égalité (6.4) équivaut à résoudre un problème d'optimisation convexe. Une solution plus accessible consiste à maximiser la forme duale et à introduire les multiplicateurs de Lagrange (α_i, α_j^*) . Le nouveau problème d'optimisation

ainsi que ses contraintes est donné par (6.5).

$$\begin{aligned} \text{Maximizer} \quad & -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \phi(x_i, x_j) \rangle \\ & -\epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y(i)(\alpha_i - \alpha_i^*). \end{aligned} \quad (6.5)$$

Résoudre cette équation aboutit une redéfinition de (6.2) sous la forme de l'équation (6.6).

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b. \quad (6.6)$$

Cette définition montre que la solution peut être trouvée par la seule connaissance des valeurs de $\langle \phi(x_i), \phi(x) \rangle$ sans connaître la fonction ϕ . Une fonction qui correspond à un produit scalaire dans un espace caractéristique quelconque \mathbb{F} répondant à la définition $k(x, x') = \langle \phi(x), \phi(x') \rangle$ est appelée un noyau. Un noyau peut-être n'importe quelle fonction symétrique satisfaisant la condition de Mercer comme le noyau radial gaussien (Gaussian Radial Basis - RBF), définie par (6.7). Le noyau RBF est paramétré par γ ($\gamma > 0$) qui impacte entre autre les capacités de généralisation du régresseur.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2). \quad (6.7)$$

6.3.2 Méthode k -nn

Contrairement à SVR, la méthode d'apprentissage k -NN consiste à mémoriser l'ensemble du jeu d'apprentissage sans effectuer de régression ou de calculs complexes.

En prenant comme exemple l'ensemble d'apprentissage $E = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{X} \times \mathbb{R}$, avec $\mathbb{X} \subseteq \mathbb{R}$, le processus d'estimation de k -nn pour estimer un objet $z = (x', y')$ peut être résumé en trois étapes :

1. l'algorithme calcule les distance $d(x', x)$ entre z et chaque objet $(x_i, y_i) \in D$.
2. l'ensemble F des k voisins les plus proches de z est constitué.
3. l'algorithme k -NN calcule l'estimation \hat{y} avec $\hat{y} = \frac{1}{k} \sum_{i=1}^k x_i, x \in F$.

Des variantes existent et concernent essentiellement la méthode utilisée pour calculer la distance $d(x, x')$. Citons par exemple la distance de Manhattan, la distance euclidienne ou la distance de Minkowski d'ordre p . Celle-ci se définit pour deux ensembles d'objets $F = (x_1, \dots, x_n)$ et $G = (y_1, \dots, y_n) \in \mathbb{R}^n$ par l'équation (6.8).

$$\left(\sum_{i=1}^n |x_i - y_i| \right)^{\frac{1}{p}}. \quad (6.8)$$

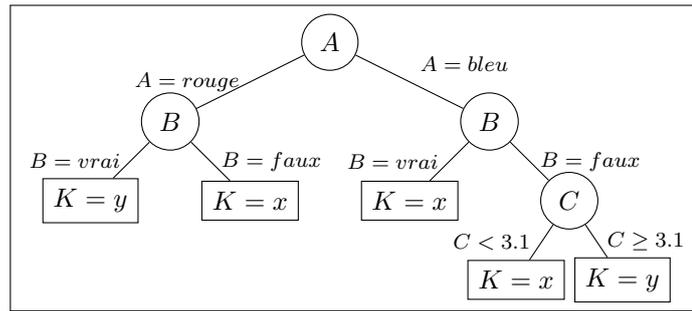


FIGURE 6.1 – Exemple d'arbre de décision.

6.3.3 Méthode DT

L'utilisation d'arbres de décision constitue une troisième forme d'apprentissage.

Un arbre de décision binaire, comme celui de la figure 6.1, représente visuellement une suite de décisions prises à partir du nœud racine vers les nœuds terminaux. L'estimation est obtenue quand un nœud terminal est atteint. Chacun des nœuds terminaux représente alors une valeur d'estimation et chaque séparation constitue une décision prise à partir d'une caractéristique x_i de l'objet à estimer.

Les algorithmes d'apprentissage, comme C4.5 [92] ou CART²[28], sont utilisés pour générer l'arbre de décision à partir d'un jeu de données. Leur différence principale réside dans leur procédure de séparation. La méthode du moindre carré (Least Squares - LS) utilisée par CART [136] se définit pour un nœud t par la relation (6.9) avec $y_i(t)$ la valeur individuelle de la variable au nœud t et $\overline{y_i(t)}$, sa valeur moyenne.

$$SS(t) = \sum_{i=1}^N y_i(t) - \overline{y_i(t)} \quad (6.9)$$

La séparation est alors effectuée en choisissant d'abord la caractéristique dont l'utilité de la séparation calculée par la fonction (6.10) est maximale avec t_r et t_t respectivement les nœuds gauche et droit issus de la séparation.

$$\phi(s, t) = SS(t) - SS(t_r) - SS(t_t) \quad (6.10)$$

L'algorithme général suivi par CART pour construire un arbre à partir des observations X , constituées des caractéristiques (x_1, x_2, \dots, x_i) , est le suivant :

1. À partir du nœud racine et pour chaque attribut x , l'algorithme calcule une mesure d'impureté après séparation, en utilisant la fonction $SS(t)$.
2. L'algorithme sélectionne l'attribut x^* qui maximise $\phi(s, t)$.

2. Le terme CART désigne à la fois la catégorie d'algorithmes d'apprentissage automatique regroupant les arbres de décisions pour la classification et la régression - évoquée dans l'état de l'art - et l'algorithme de construction de l'arbre comme dans le cas présent.

3. Le nœud de séparation correspondant est créé.
4. Une récursion par retour à l'étape 1 est effectuée sur le sous-arbre obtenu après séparation sur x^* .

L'algorithme s'arrête quand les cas suivants sont atteints :

- Aucune des caractéristiques restantes ne permet d'amélioration de la valeur de $\phi(s, t)$.
- Tous les objets restants à apprendre ont les mêmes caractéristiques. Dans ce cas-là l'algorithme crée un seul nœud terminal.

Le processus de construction génère un arbre de grande taille non-optimisé. CART utilise donc une procédure pour réduire sa taille et extraire un modèle plus optimal et plus performant. La taille optimale de l'arbre n'est pas connue avant le processus d'élagage, par conséquent, CART produit une collection d'arbres, éligibles pour l'optimalité, suivant un processus d'élagage de l'arbre initial. *In fine*, l'algorithme sélectionnera le modèle optimal parmi ceux-ci.

Pour générer les différents arbres, CART utilise une fonction de mesure du coût et de la complexité défini par (6.11) avec $R(T)$ le coût de l'arbre d'apprentissage, $|T|$ le cardinal des nœuds terminaux et a la pénalité imposée pour chaque nœud. Ainsi pour $a = 0$, le coût de complexité est minimal et correspond à l'arbre le plus grand possible, *i.e.* il s'agit de l'arbre initial. À contrario, une augmentation progressive de a , permet l'élagage progressif des séparations aux bas de l'arbre. Le processus d'élagage consiste donc à faire varier la valeur de a à partir de 0 et jusqu'à une valeur suffisante pour supprimer toutes les séparations.

$$R_a(T) = R(T) + a |T|. \quad (6.11)$$

Le modèle optimale est celui qui parmi cette collection obtient les meilleures performances sur un jeu de test autre que le jeu d'apprentissage.

6.4 Données expérimentales

6.4.1 Protocole expérimentale pour la génération de données

Le dispositif expérimental et les équipements utilisés pour l'expérience sont indiqués sur la figure 4.7. Deux modules de communication WIFI Atheros sont placés de part et d'autre de la chambre anéchoïque. Le boîtier A génère du trafic TCP vers le boîtier B par l'intermédiaire du logiciel IPERF. Le débit du flot TCP et la taille des paquets sont respectivement fixés à 24 Mbps et 1470 octets. un générateur de signal situé à l'extérieur de la chambre produit un signal AWGN injecté dans la chambre par une antenne directionnelle. L'antenne est dirigée vers le nœud B de manière à ne perturber que la réception de la communication. Afin d'enregistrer le signal reçu par le boîtier B, la sortie de son antenne est dupliquée et

connectée à un oscilloscope qui enregistre l'amplitude instantanée du signal reçu. Enfin, tout le trafic reçu ou envoyé sur les boîtiers est capturé. L'ensemble des équipements est synchronisé par NTP.

Après le processus de mesure, des traitements (filtrage, ...) sont effectués sur les données capturées afin de constituer des jeux de données propres à l'apprentissage automatique.

6.4.1.1 Constitution des jeux de données

Différents échantillons de données sont assemblés en faisant varier l'amplitude du bruit généré et la puissance d'émission. Ces échantillons constituent les ensembles d'apprentissage selon le tableau 6.1(a). Trois jeux d'apprentissage de tailles différentes sont ainsi constitués : le *Dataset1* contient toutes les mesures du jeu d'apprentissage, le *Dataset2* en contient 50% tandis que le *Dataset3* possède seulement 25% des données.

En utilisant les mêmes paramètres d'amplitude du bruit généré et de puissance d'émission, un jeu de test est créé à partir d'une deuxième série d'échantillons. Celui-ci est constitué de l'ensemble des échantillons agencés selon la description de la table 6.1(b).

TABLE 6.1 – Constitutions et caractéristiques des jeux d'apprentissage (a) et jeu de test (b). Chaque vecteur représente 1 seconde de mesure.

(a)

Jeu d'apprentissage	Définition du jeu de données
<i>notation</i>	{Puissance d'émission (dBm); Puissance du bruit (dBm)} _{set de mesures ; {échantillon 2} ; ...}
<i>Dataset1</i> 5323 vecteurs	{10; -20} _A ; {10; -17} _A ; {10; -15} _A ; {10; -13} _A ; {10; -10} _A ; {10; -7} _A ; {10; -5} _A ; {20; -20} _A ; {20; -17} _A ; {20; -15} _A ; {20; -13} _A ; {20; -10} _A ; {20; -7} _A ; {20; -5} _A
<i>Dataset2</i> 2661 vecteurs	{10; -20} _A ; {10; -17} _A ; {10; -15} _A ; {10; -13} _A ; {20; -20} _A ; {20; -17} _A ; {20; -15} _A ; {20; -13} _A ;
<i>Dataset3</i> 1330 vecteurs	{10; -20} _A ; {10; -17} _A ; {10; -15} _A ; {10; -7} _A ; {10; -5} _A ; {20; -20} _A ;

(b)

Jeu de test	Définition du jeu de données
<i>notation</i>	{Puissance d'émission (dBm); Puissance du bruit (dBm)} _{set de mesures ; {échantillon 2} ; ...}
<i>Dataset4</i> 2492 vecteurs	{20; -20} _B ; {20; -15} _B ; {20; -13} _B ; {10; -15} _B ; {10; -17} _B ; {10; -13} _B ; {10; -7} _B ; {20; -7} _B ; {20; -13} _B ; {10; -20} _B ; {20; -5} _B ; {20; -20} _B

6.4.2 Caractéristiques de mesure de performance et du canal sans fil

6.4.2.1 Caractéristique de performance

Dans notre application, la métrique qui constitue la performance d'un lien et qui doit être estimée est le débit moyen par seconde au niveau IP. Elle est calculée sur les traces de trafic PCAP capturées sur le nœud récepteur. Nous définissons BW_i , le débit calculé à la seconde i pour une trace donnée, par $BW_i = \sum_{k=1}^k L(p_k)$ avec $k \in \mathbb{N}$. $L(p)$ est la taille de la charge utile à la couche réseau pour le paquet $p \in P_i$ défini comme l'ensemble des paquets reçus sans erreur FCS durant la seconde i : $P_i = \{p_1, \dots, p_k, \dots, p_n\}$.

6.4.2.2 Caractéristiques de mesure physiques du canal

Mesures radio par la puce Atheros (RSS Atheros)

Le RSS Atheros est une valeur extraite des entêtes RADIOTAP des paquets contenus dans les fichiers PCAP. En effet, les entêtes des paquets reçus 802.11/radiotap disposent d'un champ 'radio'. Ce champ contient une valeur mesurée par la puce lors de la réception du paquet et qui correspond à une mesure de puissance du signal reçu (RSS) ou de ratio puissance du signal sur puissance du bruit (SNR). Cette valeur dépend donc du driver et du matériel utilisé. Dans nos expérimentations, la valeur mesurée est la puissance du signal reçu. Ainsi, $RSS(p)$ est le RSS du paquet p et R_i l'ensemble des paquets capturés pendant la seconde i , nous définissons donc la valeur du RSS Atheros comme : $ATH_RSS_i = \bar{R}_i$ avec $R_i = \{RSS(p_1), \dots, RSS(p_n)\}$.

Bruit mesuré par l'oscilloscope LeCroy

Le bruit mesuré par l'oscilloscope LeCroy est une mesure du niveau de bruit présent sur les captures de l'oscilloscope. Les captures de signal sont constituées de trois parties qui sont le bruit avant la trame (A), la trame elle-même (B) et le bruit après-trame (C). Par conséquent, le calcul du bruit mesuré par l'oscilloscope LeCroy pour une trame p est effectué sur les données de bruits A et C tel que $N_p = \bar{A} \cup \bar{C}$.

Soit p_1, \dots, p_n , l'ensemble des trames capturées par l'oscilloscope LeCroy pendant la seconde i , la caractéristique du plafond de bruit à la même seconde est définie par $LECR_NOISE_i = \{N_{p_1}, \dots, N_{p_n}\}$.

Puissance du signal reçu mesurée par l'oscilloscope LeCroy

Conformément aux standards 802.11g, la puissance du signal reçu est calculée sur les 8 premiers symboles de la trame [6]. Ces points constituent l'ensemble D .

On a donc similairement aux équations précédentes, le RSS pour une trame p égale à $R_p = \bar{D}$ et $LECR_RSS_i = \overline{\{R_{p_1}, \dots, R_{p_n}\}}$ avec $LECR_RSS_i$ la caractéristiques du RSS LeCroy à la seconde i .

SNR mesuré par l'oscilloscope LeCroy

Le SNR S_p pour une trame p est calculé comme la différence entre le plafond de bruit et le RSS de la trame p et donc, de manière similaire aux formules précédentes, le SNR est égal à $S_p = R_p - N_p$ et $LECR_SNR_i = \overline{W_i}$ avec $W_i = \{S_{p_1}, \dots, S_{p_n}\}$.

6.4.3 Notations alternatives des caractéristiques

Pour faciliter la lecture, chacune des caractéristiques de mesure physique est associée à une notation alternative comme présenté dans le tableau 6.2.

<i>Caractéristiques</i>	<i>Notations</i>
<i>ATH_RSS</i>	<i>C1</i>
<i>LECR_RSS</i>	<i>C2</i>
<i>LECR_NOISE</i>	<i>C3</i>
<i>LECR_SNR</i>	<i>C4</i>
<i>ATH_RSS + LECR_NOISE</i>	<i>C5</i>
<i>LECR_RSS + LECR_NOISE</i>	<i>C6</i>

TABLE 6.2 – Notations alternatives des différentes caractéristiques d'apprentissage automatique.

6.5 Méthodes d'évaluation de performances

6.5.1 Méthodes d'évaluation de la précision des estimations

6.5.1.1 Erreur quadratique absolue

La première méthode de calcul de performance pour l'estimation \hat{x} d'un ensemble de données x de taille n , utilise l'erreur quadratique absolue (Mean Squared Error - MSE) de cet ensemble défini par l'équation (6.12).

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2. \quad (6.12)$$

Cette méthode est couramment utilisée pour estimer la performance d'une estimation ou d'une prédiction par apprentissage automatique [12, 103]. En revanche,

c'est une métrique généraliste qui ne prend pas en compte les caractéristiques expérimentales (débit d'émission, ...) et le contexte de l'estimation de bande passante (erreur acceptable, ...).

6.5.1.2 Pourcentage de 'bonnes estimations'

Nous définissons donc une deuxième méthode spécifique à notre application. Celle-ci calcule le pourcentage d'estimations considérées comme bonnes sur l'ensemble du jeu d'estimation. Une estimation est considérée comme bonne si la différence absolue entre cette estimation et la performance réelle est inférieure à une valeur arbitraire. Cette performance est notée $P(e < \xi)$ pour e l'erreur mesurée et ξ la valeur du seuil. Ainsi, la condition de validité d'une estimation notée $V(x, \hat{x}, e < \xi)$ avec \hat{x} une estimation, x sa valeur réelle et ξ le seuil, s'exprime par :

$$V(x, \hat{x}, e < \xi) = \begin{cases} 1 & \text{si } |\hat{x} - x| > \xi. \\ 0 & \text{sinon .} \end{cases} \quad (6.13)$$

La performance pour un ensemble d'estimations \hat{x} et les valeurs correspondantes mesurées x est définie par :

$$P(e < \xi) = \frac{1}{n} \sum_{i=1}^n V(x_i, \bar{x}_i, \xi). \quad (6.14)$$

La valeur de seuil peut-être adaptée aux caractéristiques expérimentales et à la nature du réseau sans-fil. Dans notre application, étant donné le débit d'émission de 24 Mbps et les amplitudes des valeurs comprise entre 24 Mbps et 14 Mbps, le seuil est fixé à 1 Mbps. Ce seuil correspond à une erreur d'estimation de 4 % (89 paquets sur les 2139 transmis en une seconde).

6.5.2 Méthodes d'évaluation de la pertinence des caractéristiques de mesure physique

La pertinence des différentes caractéristiques est évaluée selon trois critères :

- Les deux premiers sont respectivement basés sur les valeurs de MSE et sur le pourcentage de bonnes estimations, la précision des estimations étant jugée comme la qualité principale des algorithmes.
- Le troisième s'appuie sur les mesures d'impureté de l'algorithme de génération d'arbre CART (section 6.3.3). Celle-ci évalue la qualité des caractéristiques à chaque séparation, par conséquent, la somme de chacune de ces valeurs permet d'évaluer leur importance globale [124, 123].

6.5.3 Méthodes d'évaluation de performance des algorithmes

Étant données les contraintes qui peuvent être présentes sur les nœuds d'un réseau sans-fil (mémoire, puissance de calcul, ...), les caractéristiques opérationnelles des algorithmes de ML peuvent s'avérer importantes pour toute mise-en-œuvre future. Par conséquent, nous définissons deux méthodes de mesure de performance des caractéristiques opérationnelles des algorithmes en se basant sur les critères d'efficacité en termes de temps de calcul et de mémoire utilisés par les processus d'estimation et de génération de modèles.

6.5.3.1 Efficacité en temps de calcul

Une méthode d'évaluation de l'efficacité d'un algorithme est le calcul du ratio $\frac{\text{Performance}}{\text{Coûts}}$. En utilisant $P(e < 1Mbps)$ comme indice de performance de l'algorithme, l'efficacité de calcul se définit respectivement pour les temps d'apprentissage et d'estimation par les équations (6.15) et (6.16).

$$R_{T, \text{entraînement}} = \frac{P(e < 1Mbps)}{\text{Temps d'apprentissage}}. \quad (6.15)$$

$$R_{T, \text{estimation}} = \frac{P(e < 1Mbps)}{\text{Temps d'estimation}} \quad (6.16)$$

6.5.3.2 Efficacité mémoire

L'efficacité mémoire d'un algorithme de ML est définie par le facteur multiplicatif M obtenu entre la taille des données et celle du modèle. Ce facteur est défini par la relation (6.17).

$$M_{\frac{|t|}{|M|}} = \frac{\text{Taille des données d'apprentissage}}{\text{Taille du modèle}}. \quad (6.17)$$

6.6 Méthodologie de mesure et d'apprentissage

6.6.1 Configuration des algorithmes de ML

6.6.1.1 SVR

L'algorithme SVR a été utilisé avec un noyau de type RBF. Comme la section 6.3.1 le souligne, dans notre configuration, l'algorithme SVR requiert le choix de trois paramètres par l'utilisateur : C , γ et ϵ . Ces trois paramètres peuvent impacter la performance de l'estimation et par conséquent, doivent être sélectionnés avec soin en fonction de l'application. Pour nos estimations, nous avons utilisé une méthode empirique qui consiste à conduire des séries exhaustives d'apprentissage en utilisant une liste de combinaisons de paramètres générée automatiquement. Cette méthode est connue sous le nom de recherche par quadrillage. Les modèles

généralisés sont ensuite testés en utilisant un jeu de données autre que le jeu d'apprentissage. La combinaison de paramètres C , γ et ϵ sélectionnée est celle qui permet la meilleure précision d'estimation.

6.6.1.2 k -NN

La valeur de k impacte les performances de l'algorithme k -nn. Par conséquent, après plusieurs tests effectués sur des jeux de données différents, nous avons choisi une valeur qui permet un bon compromis entre la précision des estimations et les résultats d'estimation. Ainsi, dans les expériences présentées, la valeur du paramètre k est fixée à 3. La distance utilisée est la distance de Minkowski d'ordre 2 similaire à la distance euclidienne recommandée dans la version initiale de l'algorithme [14].

6.6.1.3 DT

L'algorithme utilisé pour construire l'arbre est CART.

6.6.2 Mesure de performance des algorithmes

6.6.2.1 Mesure des délais d'apprentissage et d'estimation

Une partie des analyses de nos estimations concerne le temps de calcul associé aux processus d'apprentissage et d'estimation. Notre configuration utilise Python et la librairie scikit-learn [86]. Par conséquent la lecture de l'horloge s'effectue par l'intermédiaire de la fonction python 'time', avant et après le processus mesuré. La différence temporelle entre les deux mesures constitue le délai. Pour chacun des processus étudiés, 100 essais indépendants sont réalisés.

6.6.2.2 Mesure de la taille du modèle

La mesure de la taille des modèles générés par les différents algorithmes s'effectue en mesurant la taille des instances de ces modèles dans l'environnement Python.

6.7 Résultats

6.7.1 Précision des estimations

Les figures 6.3 et 6.4 montrent respectivement les résultats en termes de précision des différentes estimations en fonction du MSE et de $P(e < 1Mbps)$. Afin de les illustrer, deux exemples d'estimations sont affichés sur les figures 6.2(a) et 6.2(b). Celles-ci représentent respectivement les estimations obtenues par l'algorithme SVR et le jeu d'entraînement *Dataset1* avec les caractéristiques $C4$ (*LECR_SNR*)

et $C1$ (ATH_RSS). Les caractéristiques $C4$ et $C1$ obtiennent respectivement les meilleures et moins bonnes estimations avec cet algorithme. Le reste des graphiques obtenus pour les différentes estimations est inclus en annexe A.

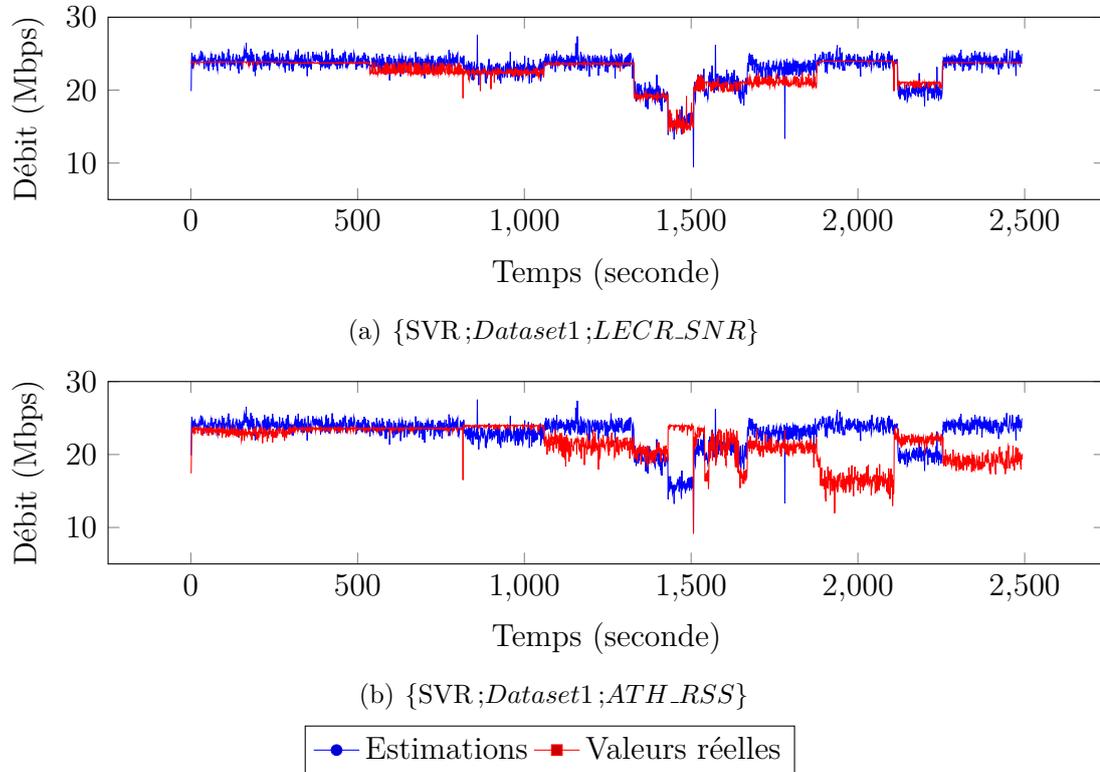


FIGURE 6.2 – Estimations obtenues avec SVR et le jeu d'entraînement *Dataset1* pour les caractéristiques $LECR_SNR$ ($C4$) (figure 6.2(a)) et ATH_RSS ($C1$) (figure 6.2(b)).

L'algorithme SVR permet des estimations précises en termes de MSE (< 6) pour toutes les caractéristiques exceptées pour la caractéristique $C1$ qui présente des résultats plutôt moyens. Les résultats sont similaires pour les 3 jeux d'apprentissage. L'algorithme k -nn présente des résultats identiques sur la plupart des caractéristiques. Les métriques les plus faibles sont également les métriques de puissance du signal avec des estimations dont le MSE qui peuvent-être supérieures à 30. Contrairement aux résultats obtenus avec SVR, on peut observer une amélioration pour certaines caractéristiques entre le *Dataset1* et le *Dataset2*. Les arbres de décision permettent d'excellentes estimations, notamment sur les caractéristiques $C6$, $C4$, $C3$. La métrique $C5$ montrent de bons résultats lors de l'utilisation du *Dataset1* mais subit une forte dégradation lors de l'utilisation du *Dataset3*.

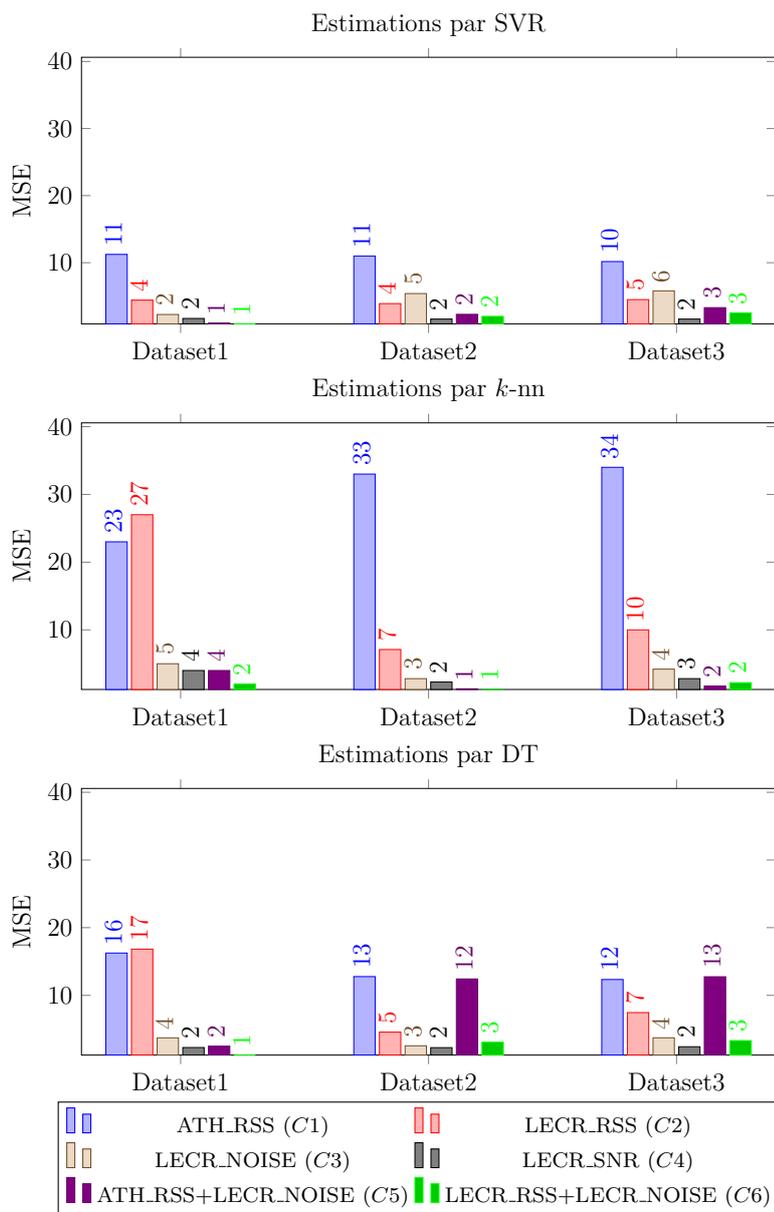


FIGURE 6.3 – Précision moyenne des estimations obtenues pour les différents algorithmes de ML, en fonction des différents jeux d’apprentissage. L’unité de précision est ici le MSE.

Par utilisation de SVR avec le 1^{er} jeu d’apprentissage, les caractéristiques composées C5 et C6 ainsi que la caractéristique de bruit et de SNR ont des résultats prometteurs avec plus de 70% de bonnes estimations. Ceci est confirmé pour le SNR lors de l’utilisation des Dataset2 et Dataset3 avec respectivement 66% et 62% de probabilité de bonnes estimations, tandis que les résultats obtenus par les

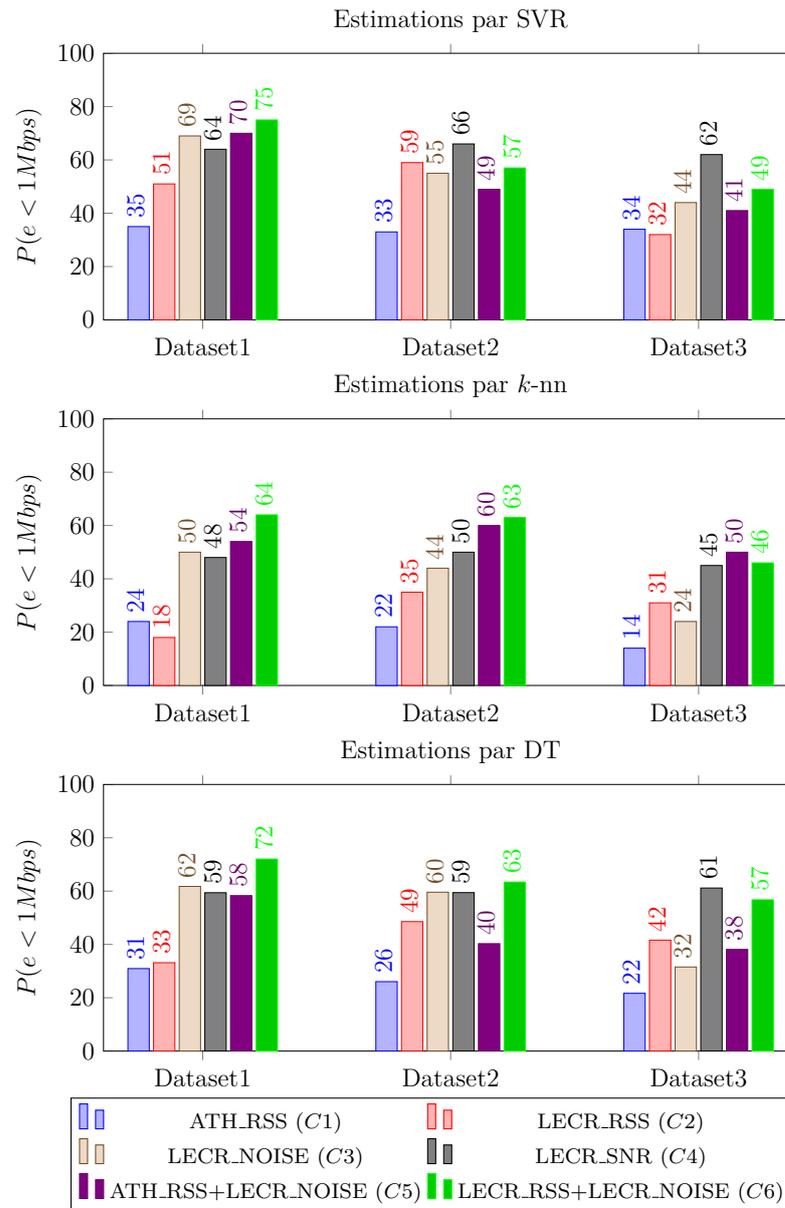


FIGURE 6.4 – Précision moyenne des estimations obtenues pour les différents algorithmes de ML, en fonction des différents jeux d'apprentissage. L'unité de précision est ici le pourcentage de bonnes estimations $P(e < 1Mbps)$.

autres métriques tombent à 50%. L'utilisation de k -nn et de la métrique C6 permet une précision élevée de plus de 60% lors des tests effectués avec les 1^{er} et 2^{ème} jeux de données. Similairement aux observations faites sur le MSE lors de l'usage de k -nn, on peut observer une amélioration sur certaines métriques entre l'utilisation des jeux de test *Dataset1* et *Dataset2*. L'algorithme par arbre de décision se

comporte bien, notamment avec l'utilisation de la métriques de SNR qui obtient plus de 60% de précision sur l'ensemble des jeux de données d'apprentissage. La caractéristique $C6$ permet d'atteindre une précision de 70% avec *Dataset1*, et reste très précise avec les autres modèles d'apprentissage.

À la vue de ces résultats, il apparaît clairement que les métriques composées de bruit et de puissance du signal reçu ($C5$ et $C6$) ainsi que la métrique de SNR permettent une bonne estimation de la performance d'un réseau sans-fil à partir des couches 4 et 5. À l'opposé, les métriques de RSS offrent une précision moyenne. Lors de l'utilisation de l'algorithme k -nn, la précision s'améliore entre le *Dataset1* et le *Dataset2*, ce qui tend à montrer que le 1^{er} jeu de données est inadapté à ces métriques pour l'usage d'un algorithme simple comme k -nn qui peut-être sensible aux données bruitées [134]. En revanche, l'algorithme DT offre des performances égales à celles de SVR en ce qui concernent les métriques composées et le SNR. De plus, il parvient à conserver ces performances malgré la réduction du jeu d'apprentissage et offre donc des capacités de généralisation correctes et similaires à celles de SVR. Les performances de DT pourraient néanmoins être améliorées étant donné que cet algorithme ne requiert normalement pas de traitement de données comme ceux qui ont été effectués ici [134] et permet de travailler sur des données brutes. L'algorithme k -nn ne permet pas une généralisation aussi avancée étant donné la chute de performance entre le *Dataset2* et le *Dataset3*.

6.7.2 Pertinence des caractéristiques de mesure RF

Les figures 6.5 et 6.6 présentent la pertinence des caractéristiques en fonction de la précision de leurs estimations en termes de MSE et de pourcentage de bonnes estimations. L'indice de pertinence par critère de Gini est affiché sur la figure 6.7.

Selon ces données la pertinence des caractéristiques pour l'algorithme SVR ne varie pas en fonction de la taille des jeux d'apprentissage. À contrario, on peut observer le résultat inverse avec l'utilisation de l'algorithme k -nn. Ainsi la mesure du bruit est une mesure pertinente lors de l'utilisation du *Dataset1*, mais plus lors de l'utilisation des *Dataset2* et *Dataset3* où la métrique devient la 5^{ème} plus importante sur les 6 en termes de pourcentage. On observe également ce résultat lors de l'utilisation des arbres de décision où cette métrique passe de la 2^{ème} à la 5^{ème} place entre l'utilisation du *Dataset2* et celle du *Dataset3*. Les métriques composées sont avantageusement utilisées par SVR et k -nn et se classent parmi les 2 premières. Ce résultat peut-aussi être remarqué lors de l'utilisation de la métrique $C6$ et des arbres de décision. En revanche, la caractéristique $C5$ se retrouve 6^{ème} lors de l'utilisation du 3^{ème} modèle d'apprentissage. Les métriques de RSS sont de manière générale assez mal classées. Ces résultats sont cohérents avec ceux de la figure 6.7. Dans ce classement, le SNR obtient un indice de pertinence largement supérieur aux autres métriques avec plus de 80% de bonnes estimations. Le deuxième score

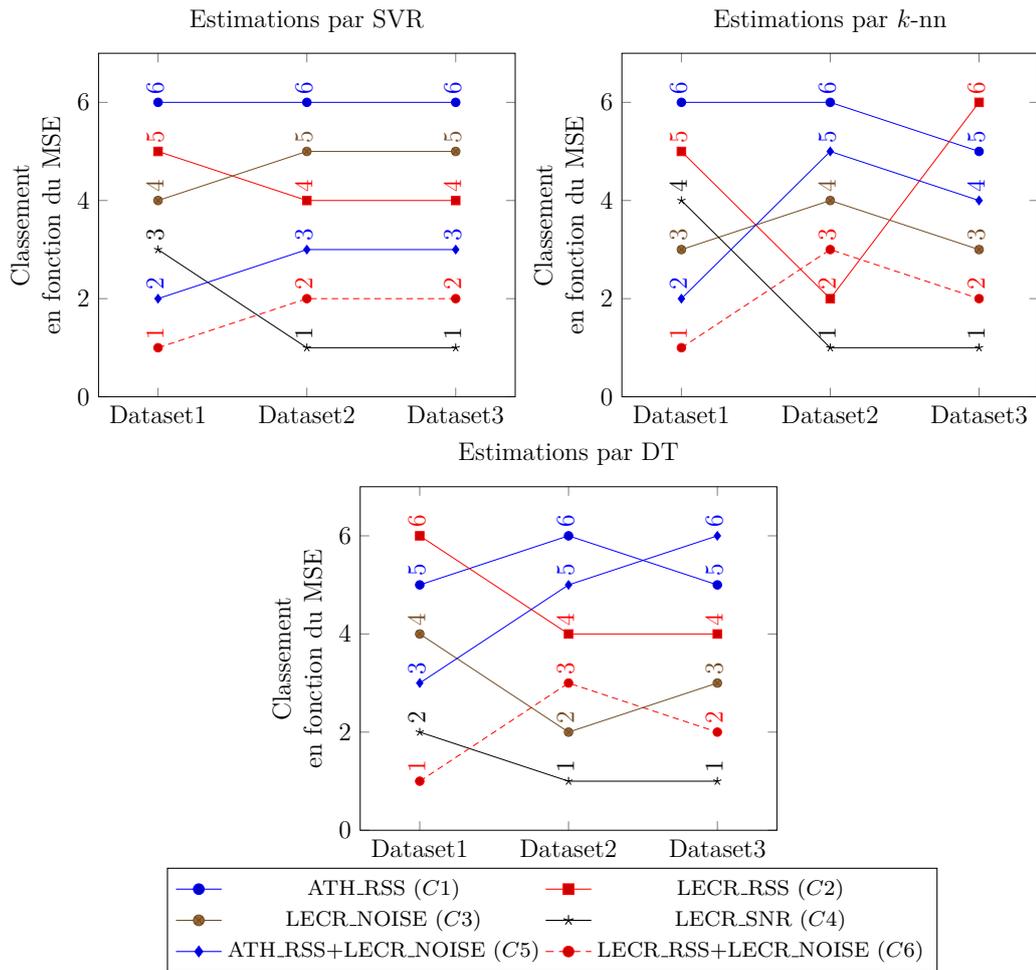


FIGURE 6.5 – Pertinence des différentes métriques pour l'estimation de débit sur un lien sans-fil vis-à-vis des différents algorithmes de ML. Le classement est exprimé en fonction des résultats mesurés en MSE.

est obtenu pour la caractéristique de mesure de bruit avec respectivement 20, 12 et 14% de précision pour les jeux 1, 2 et 3.

Ces résultats montrent que le SNR seul (C4) peut-être considéré sans ambiguïté comme la meilleure caractéristique pour l'estimation d'un lien sans-fil. En revanche, utiliser plusieurs caractéristiques (C5 et C6) permet bien souvent d'obtenir des performances similaires.

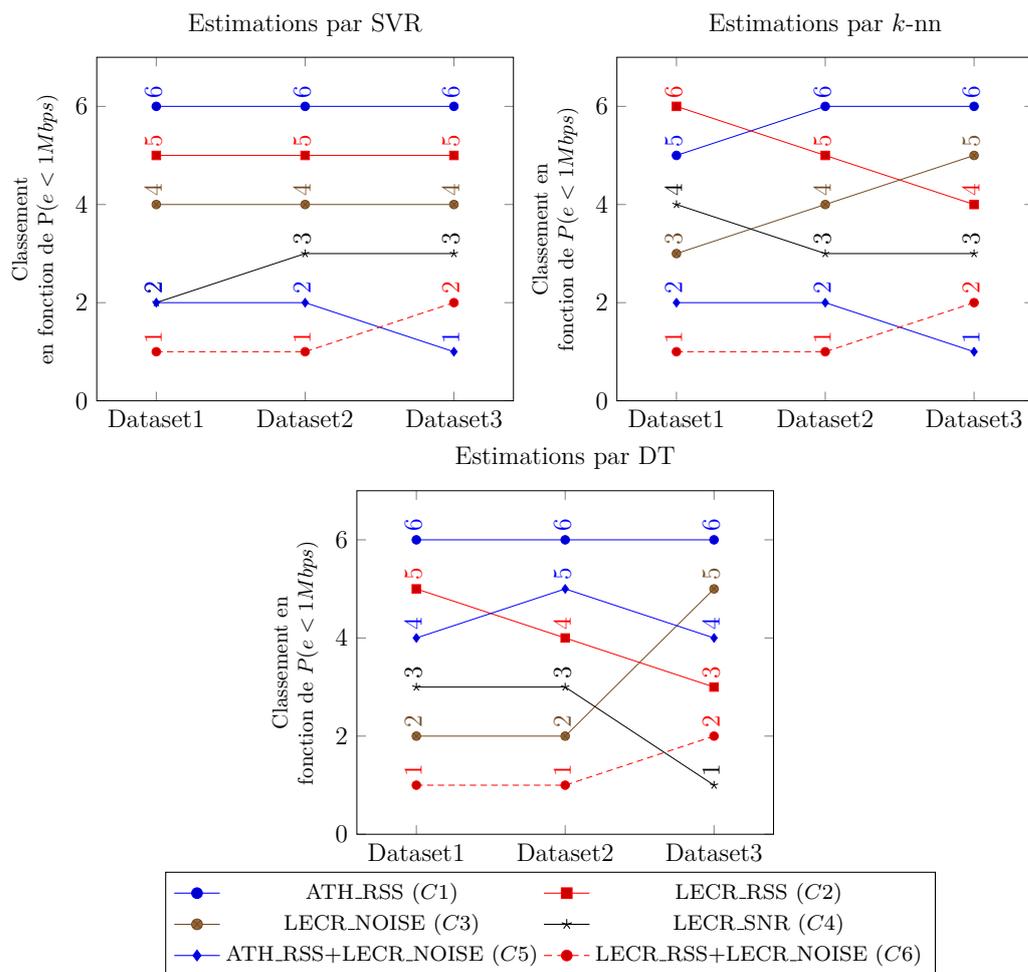


FIGURE 6.6 – Pertinence des différentes métriques pour l’estimation de débit sur un lien sans-fil vis-à-vis des différents algorithmes de ML. Le classement est exprimé en fonction des résultats mesurés en $P(e < 1Mbps)$.

6.7.3 Fonctionnement des algorithmes

6.7.3.1 Efficacité en temps de calcul

Les figures 6.8 et 6.9 présentent respectivement les temps mesurés et l’efficacité correspondante des différents algorithmes pris pour les processus d’apprentissage et d’estimation.

Les résultats montrent que d’une manière générale, ces valeurs peuvent être très importantes dans le cadre de l’utilisation de SVR : ce temps peut même atteindre les 40 secondes pour le RSS mesuré par l’oscilloscope. Les temps d’apprentissage pour les métriques les mieux classées sont correctes et respectivement de 5, 1 et moins de 1 seconde pour les jeux de données 1, 2 et 3. Les temps d’apprentissage

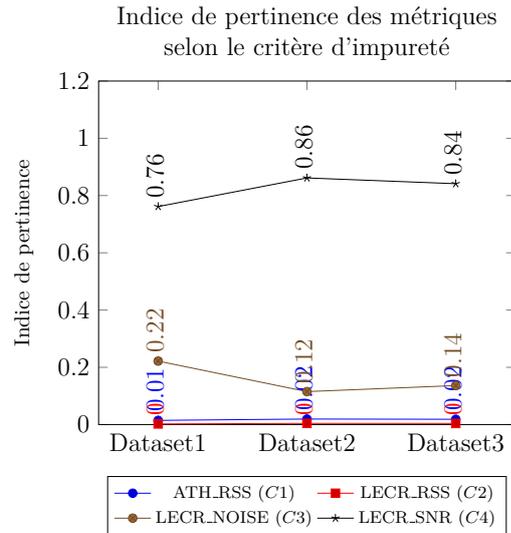


FIGURE 6.7 – Indices de pertinence calculés selon le critère d'impureté par l'algorithme CART. Les indices sont normalisés, *i.e* pour chaque jeu d'apprentissage, la somme des indices est égale à 1.

pour k -nn sont bien meilleurs et proportionnels à la taille du jeu d'apprentissage. De manière similaire, les résultats pour l'algorithme par arbre de décision offrent des temps d'apprentissage inférieurs à 1 seconde. Il est à noter que pour ces deux algorithmes, les valeurs sont toutes similaires ou égales quelque soient les métriques utilisées. Les temps d'estimation sont tous inférieurs à 1 ms peu importe les algorithmes utilisés. Le temps d'estimation moyen est le plus important pour SVR et dépend du jeux de données. En revanche, pour cet algorithme le choix de la métrique utilisée n'a aucune influence. À contrario dans le cas de k -nn, ce temps dépend de la métrique et est systématiquement inférieur à $400 \mu s$. Enfin, les temps d'estimation pour DT sont les plus faibles ($7 \mu s$) pour l'ensemble des données.

L'algorithme le plus efficace en termes d'apprentissage est k -nn qui démontre un ratio pouvant aller jusqu'à 44. Il est suivi de près par DT qui montre des scores similaires. L'algorithme SVR est peu efficace étant donné ses temps d'apprentissage importants. D'une manière globale, l'efficacité augmente avec la diminution du jeu d'apprentissage. En termes d'estimations, les figures 6.9 montrent que les trois algorithmes sont efficaces. Les résultats de k -nn font apparaître deux grappes qui contiennent respectivement dans la première les caractéristiques dites composées et dans la seconde, les caractéristiques simples. On peut voir que l'efficacité de la première est très grande et comparable à celle observée avec SVR. De plus, ces valeurs augmentent avec la diminution des modèles d'apprentissage. Cette tendance est inverse dans le cas de l'algorithme DT qui voit son efficacité diminuer. Encore une fois, cet algorithme démontre une excellente efficacité, ceci pour toutes les métriques.

Les résultats observés sont consistants avec la théorie. En effet, l'algorithme k -nn qui utilise la méthode d'apprentissage la plus simple démontre une efficacité maximale. Ces résultats sont comparables à ceux obtenus avec les arbres de décision qui optimisés permettent de hautes performances d'apprentissage. En revanche, la méthode de régression et d'optimisation utilisée par SVR rend son apprentissage très long comparé aux autres algorithmes, ce qui le rend difficilement utilisable dans un processus en-ligne. En termes d'estimation, les temps observés sont très bas. L'algorithme le plus efficace compte tenu de ses performances et de sa rapidité étant DT.

6.7.3.2 Efficacité mémoire

La taille des modèles obtenus ainsi que les facteurs multiplicatifs sont présentés sur la figure 6.10.

Sans surprise, la taille des modèles générés est inégale selon les algorithmes. L'algorithme DT possède le modèle le plus volumineux. Son facteur multiplicatif est proche de 20. Les algorithmes k -nn et SVR ont des modèles de taille similaire et des facteurs respectifs de 4 et 6. On peut remarquer que les facteurs ne varient pas en fonction de la taille du jeu de données.

Les performances observées s'expliquent par les méthodes de génération de modèles des algorithmes respectifs. En effet, alors que DT doit stocker les arbres et leurs métadonnées, SVR enregistre les vecteurs supports et les coefficients trouvés lors de l'apprentissage. De manière similaire, l'algorithme k -nn qui n'effectue aucun calcul pour la modélisation, observe un facteur 6 sur la taille des données d'entrée. Des études approfondies doivent être menées pour chaque algorithme afin d'étudier l'importance des paramètres de chacun dans ce facteur multiplicatif.

6.8 Conclusion

Ce chapitre a proposé l'application des méthodes d'apprentissage SVR, k -nn et DT pour l'estimation prédictive du débit IP d'un lien sans-fil à partir de mesures de niveau physique. Les données utilisées ont été générées sur le banc expérimental décrit dans le chapitre 4 à partir d'un scénario mettant en œuvre un lien TCP unidirectionnel dont la réception est perturbée par du bruit. Les estimations ont été réalisées à partir de mesures de RSS, SNR et bruit effectuées sur le nœud récepteur. La performance respective des différents algorithmes a alors été évaluée en fonction de la précision des estimations, calculée selon 2 méthodes différentes : l'erreur quadratique moyenne (MSE) et le pourcentage de bonnes estimations. Additionnellement, des mesures du fonctionnement des algorithmes comme la taille des modèles d'apprentissage ou le temps nécessaire à l'estimation ont également été étudiées. Finalement, la pertinence respective de chacune des métriques a été comparée.

Plusieurs conclusions peuvent être tirées des résultats obtenus :

- Les différents algorithmes présentent de très bons résultats d'estimation. Les meilleurs résultats sont obtenus par les algorithmes SVR et DT peu importe la métrique de performance utilisée. L'algorithme k -nn semble avoir des performances plus aléatoires, sûrement à cause de la simplicité de son processus d'apprentissage et d'estimation. SVR et DT supportent aisément la diminution du jeu d'apprentissage et obtiennent des scores élevés : jusqu'à 60% de bonnes estimations avec le jeu le moins complet.
- Les meilleurs résultats sont obtenus avec les caractéristiques dites composées et le SNR. Cette dernière semble la plus efficace étant donné qu'elle supporte facilement la généralisation et semble avoir des performances stables. Les mesures de RSS obtiennent les plus mauvais résultats. Ces résultats sont confirmés par l'utilisation du critère d'impureté qui indique une pertinence de 80% pour le SNR. La seule mesure de bruit obtient 12% de pertinence.
- Le temps d'apprentissage de SVR est extrêmement long comparé à ceux de k -nn et DT. Ce long processus d'apprentissage n'offre pas à SVR une efficacité décisive sur DT. La valeur de $R_{T,entra\grave{a}nement}$ pour cet algorithme est d'environ 30 avec le *Dataset3*. Des résultats similaires sont atteints avec k -nn. Les délais d'estimation les plus courts sont atteints pour DT qui produit une estimation dans des délais inférieurs à la ms et un ratio $R_{T,estimation}$ de l'ordre de 10^5 .
- En termes de mémoire et en observant les facteurs multiplicatifs, SVR est le meilleur candidat quelque soit le jeu d'apprentissage. k -nn obtient des performances proches. L'algorithme DT a en revanche de mauvaises performances avec un facteur multiplicatif proche de 20.

Ces résultats nous permettent d'affirmer qu'il est possible d'estimer le débit IP d'un lien sans-fil à partir de caractéristiques de niveau physique dans les conditions testées. L'algorithme d'apprentissage utilisé a son importance, néanmoins les résultats démontrent qu'il n'est pas forcément nécessaire d'utiliser des algorithmes complexes au temps de calcul important, comme c'est le cas pour SVR, si une caractéristique adaptée est utilisée.

Le prochain chapitre conclura le manuscrit en présentant le bilan des différentes contributions et en évoquant les différentes perspectives.

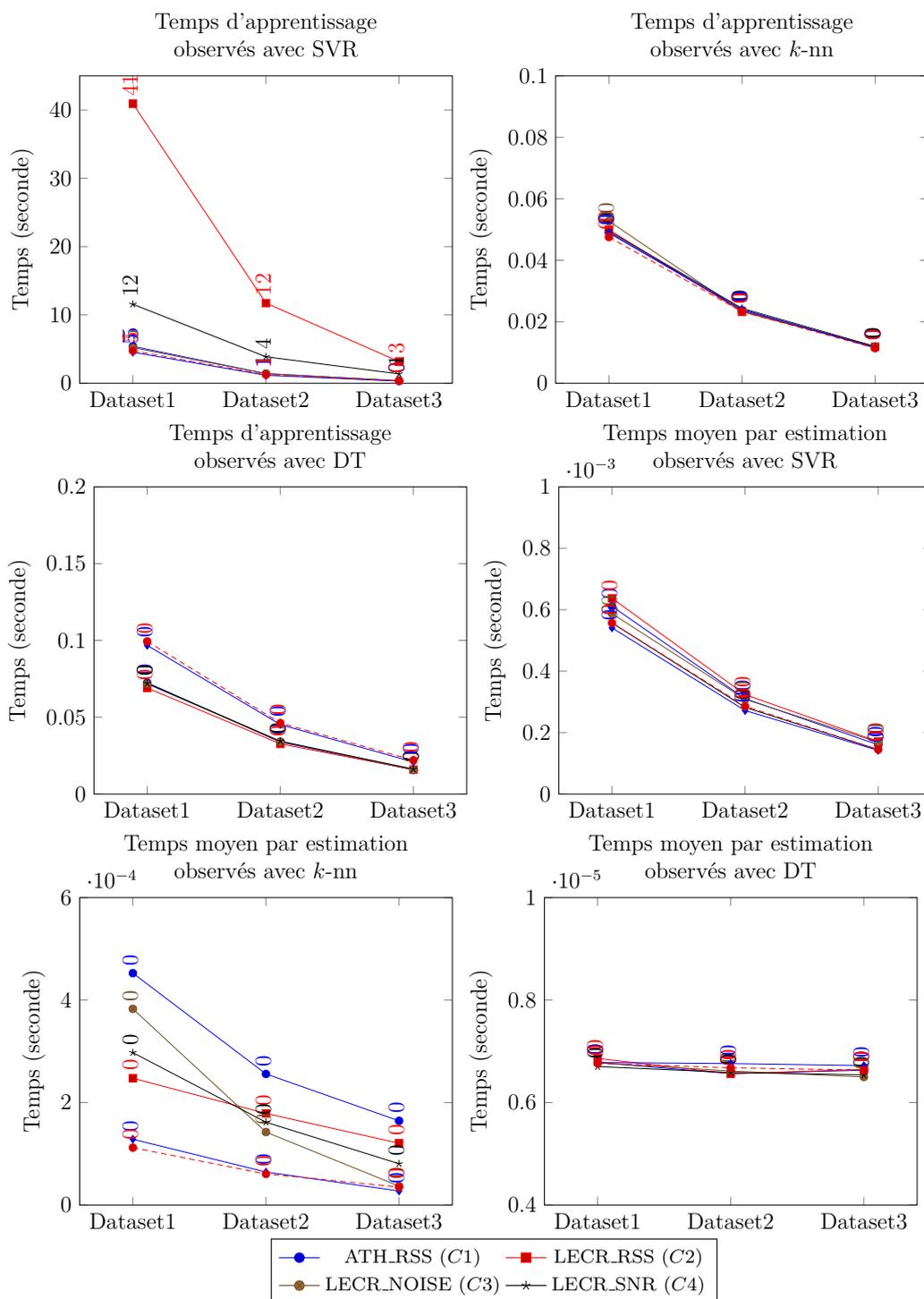


FIGURE 6.8 – Temps d'apprentissage et d'estimation des différents algorithmes SVR, k -nn et DT pour les différentes métriques.

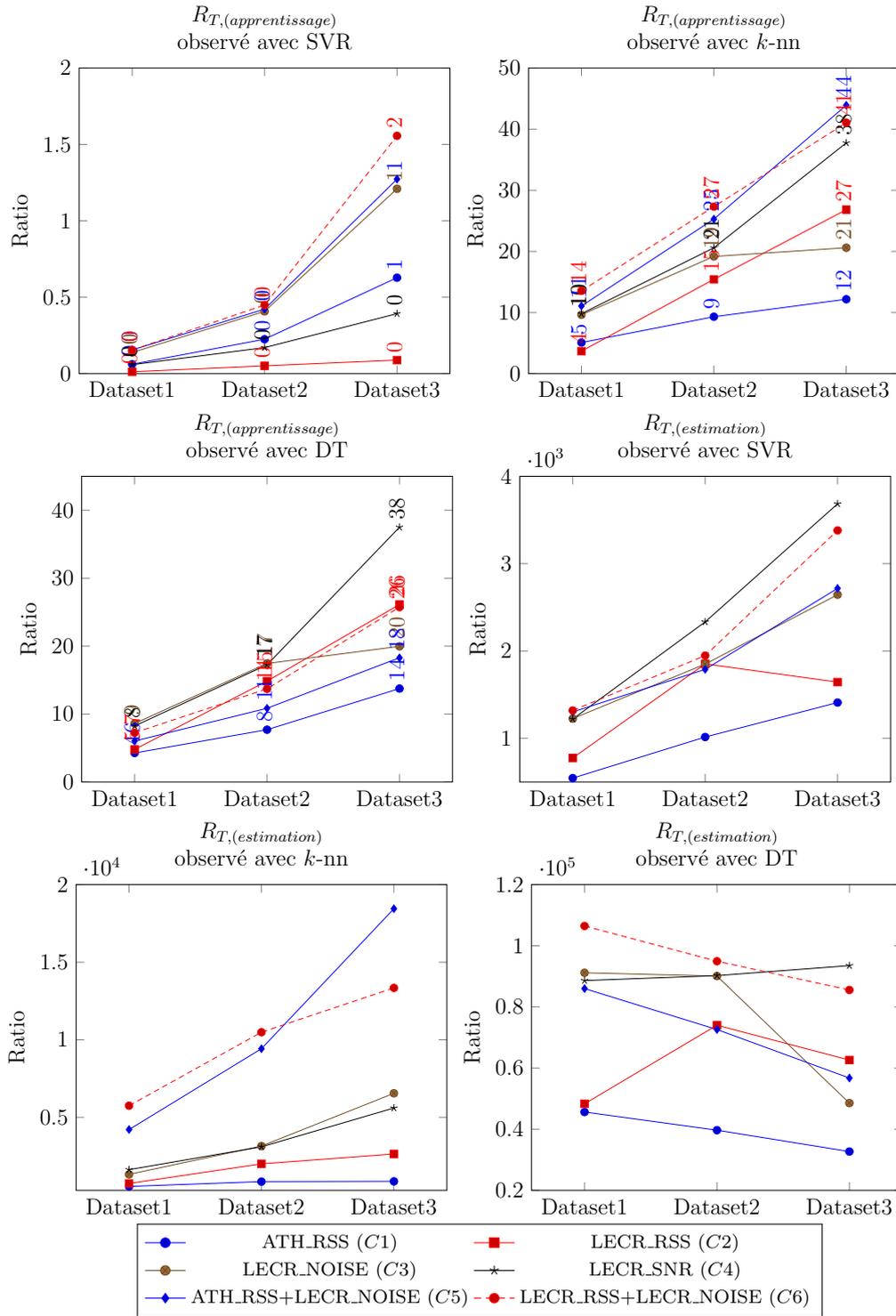


FIGURE 6.9 – Performances de calcul des différents algorithmes SVR, k -nn et DT pour les différentes caractéristiques.

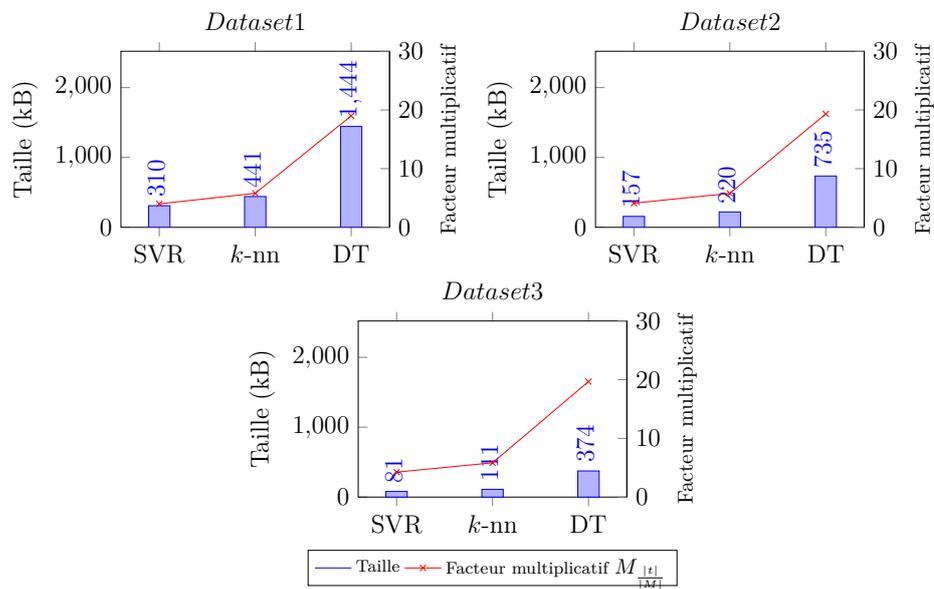


FIGURE 6.10 – Taille moyenne des modèles générés et facteurs multiplicatifs par rapport aux données brutes pour les différents algorithmes et pour chaque jeux d'apprentissage.

Bilan, perspectives et conclusion

L'évolution pressentie des réseaux sans-fil vers les réseaux de l'internet des objets et des réseaux ubiquitaires pourrait aggraver les problèmes d'interférences et dégrader la qualité du médium et des communications. Pour y faire face, les nœuds devront être plus fiables et plus adaptables. Afin d'y parvenir, des études supplémentaires sont nécessaires en termes d'instrumentation, de modélisation ou d'outils pour l'adaptabilité des nœuds. Cette thèse a apporté des contributions sur ces différents points.

Bilan

La 1^e contribution de cette thèse a consisté en l'étude et la mise en œuvre d'un banc expérimental permettant l'étude efficace des communications sans-fil et en particulier au niveau des couches les plus basses (MAC, Phy et RF). Celles-ci ont en effet un impact important sur les communications. Une étude préliminaire des besoins et de l'état de l'art existant sur les bancs de mesure a d'abord permis de choisir l'environnement idéal défini à partir d'une chambre anéchoïque. Cet environnement permet en effet d'effectuer des mesures dans un environnement RF propre de toute perturbation et de pouvoir contrôler le niveau de perturbation du médium.

Des tests ont été effectués sur le banc afin de vérifier la non-intrusivité sur les mesures des composants et le respect de certaines fonctionnalités importantes comme la synchronisation entre les nœuds. Les résultats nous permettent d'affirmer que la synchronisation est très bonne (inférieure à 0.5 ms entre les boîtiers WIFI) et que les composants du banc n'introduisent pas de modifications majeures sur les mesures.

Un des résultats de la mise en œuvre de ce banc de mesure inter-couche est d'offrir des moyens de mesure et d'instrumentation pour mieux comprendre l'impact des couches de moyen et haut-niveau sur le médium, et inversement. Il a donc permis de combler le manque d'outils expérimentaux répondant à ces besoins.

La 2^e contribution a été la validation et l'amélioration des modèles WIFI utilisés dans le simulateur ns-3. Pour ce faire nous avons proposé un modèle de compa-

raison d'environnements par analyse de cause racine (RCA). Le modèle prend en entrée des jeux de mesure obtenus à partir de traces réseaux expérimentales et de simulations. Son application permet de détecter les différences en termes de configuration, de mise en œuvre ou de modélisation entre les environnements à partir des écarts de performance constatés. Afin de valider le réalisme du simulateur, un scénario a été implémenté de manière similaire sur les deux environnements. Ce scénario consiste à perturber la réception d'un lien unidirectionnel en injectant du bruit au niveau du récepteur du lien. L'application des jeux de données dans le modèle RCA a permis de détecter une différence d'implémentation entre les environnements. Cette différence se situait au niveau de l'algorithme de temporisation exponentielle binaire (BEB) qui gère la temporisation de l'accès au lien et pouvait conduire à des écarts de performance entre les deux environnements. Dans les traces comparées, la différence de débit mesurée entre les environnements pouvait atteindre 2 Mbps (soit près de 30 % du débit initial de 7 Mbps). Après modélisation des données expérimentales et mise en œuvre d'un nouveau modèle de temporisation en simulation, cette différence diminue à près de 500 Kbps.

Les résultats de cette contribution permettent de tirer les conclusions suivantes :

- la grande diversité des équipements et des logiciels disponibles pour un même standard rend la modélisation des réseaux sans-fil difficile
- l'utilisation généralisée de méthodologies de comparaison automatiques entre simulateurs et bancs expérimentaux pourrait permettre une amélioration importante des modèles de simulation

Enfin la 3^e contribution avait pour objectif de rendre la prédiction de performances des réseaux sans-fil possible. Le médium sans-fil est sujet à des mécanismes physiques qui peuvent détériorer les signaux transmis. L'apparition et la force de ces mécanismes dépendent de la composition et la nature de l'environnement et peuvent varier dans le temps et l'espace. Ces variations sont mal comprises et sont en partie considérées comme imprévisibles. Pour limiter l'effet de ces perturbations, des mécanismes de gestion déterministes sont mis en œuvre à tous les niveaux de la pile de communication et en particulier au niveau des couches MAC et physiques. Le déterminisme de ces mécanismes laisse présager que la prédiction de performance des réseaux sans-fil est possible.

Pour cette raison, nous avons proposé dans le chapitre 5 l'application des méthodes d'apprentissage à partir de mesures du lien effectuées au niveau des couches RF ou physique des communications sans-fil. Les méthodes d'apprentissage, SVR, k -nn et DT, utilisées ont fait leur preuve dans de nombreux domaines d'estimations prédictives. Afin de constituer les jeux d'entraînement et de test, des données sont générées à partir du banc expérimental. Le scénario considéré est celui du lien unidirectionnel dont la réception est perturbée par du bruit. À la réception du signal, des mesures de RSS, de SNR et de bruit sont effectuées à l'aide

des cartes WIFI Atheros et d'un oscilloscope. Les données ont alors été divisées en jeux d'entraînement de tailles distinctes qui ont permis de tester la précision des différents algorithmes en fonction des caractéristiques de mesure physique et en fonction de la taille du jeu d'entraînement. Afin de tester l'aptitude des algorithmes d'apprentissage à être mis en œuvre sur du matériel générique sans-fil, certaines caractéristiques de fonctionnement des algorithmes ont également été comparées. Les résultats montrent que les estimations les plus précises sont obtenues avec les algorithmes SVR et DT. Ces algorithmes permettent en outre de conserver leur niveau de précision malgré la diminution du jeu d'entraînement. L'algorithme le plus rapide est DT tandis que le plus efficace en mémoire est SVR. Enfin, les meilleurs résultats sont obtenus par usage des métriques composées ou du SNR.

Les résultats obtenus dans cette contribution permettent de tirer le bilan suivant :

- il est possible de prédire une partie des caractéristiques du médium sans-fil dans la situation testée à partir de mesures de niveau physique obtenues sur du matériel 802.11 générique.
- les algorithmes d'apprentissage nécessitent des ressources limitées, il est donc raisonnable de penser que ces méthodes peuvent être appliquées sur du matériel embarqué standard.

Perspectives

Les perspectives futures qui se dessinent dans le domaine des réseaux sans-fil ainsi que les résultats obtenus permettent d'envisager des évolutions ou des extensions quand aux contributions présentées dans cette thèse.

Premièrement, du point de vue de l'instrumentation et des moyens de mesure, seule la technologie *802.11g* est pour le moment mise en œuvre sur le banc expérimental. Actuellement, les réflexions sur le futur des réseaux sans-fil sont en cours au travers de la définition des futurs réseaux 5G tandis que de nouvelles applications sont sur le point d'émerger comme les réseaux intra et inter-véhiculaires. Par conséquent, les réseaux sans-fil se spécialisent et nécessitent le développement des technologies adaptées. Ainsi, la technologie *802.11p* est conçue pour les communications véhicule à véhicule (*vehicle to vehicle - V2V*) et véhicule à infrastructure (*vehicle to infrastructure - V2I*). Pour pouvoir faire face aux contraintes de ces applications, *802.11p* incorpore plusieurs modifications par rapport aux précédentes ratifications *802.11* et notamment au niveau de la couche physique (*e.g.* largeur de bande des canaux qui passe de 20 MHz à 10 MHz). D'autre part, alors que dans le cas des applications généralistes, les antennes sont intégrées dans l'équipement radio (*e.g.* carte radio *802.11* standard) et fournissent un service minimum, les applications spécialisées pourraient utiliser des antennes externes qui répondent

mieux à leurs besoins. Ces antennes viennent en différents types (*e.g.* antennes yagis, dipôles, cornées, ...) et ont des caractéristiques différentes (*e.g.* gain, polarisation, ouverture à mi-puissance, ...). Leur utilisation a un impact important sur la performance des communications.

Par conséquent et pour accompagner ses évolutions, il s'agira dans un premier temps de mettre en œuvre ces nouveaux protocoles. Dans un deuxième temps, il faudra faire évoluer les caractéristiques du banc expérimental comme la génération des perturbations (*e.g.* déplacement de masses métalliques pour 802.11p) ou la mobilité des nœuds. Finalement, l'étude des antennes devra être intégrée au banc de mesure pour pouvoir mieux connaître leur impact sur les communications.

Deuxièmement, en termes de modélisation la mise en pratique du modèle RCA sur les traces générées en simulation et expérimentalement a donné de bons résultats. En revanche, ce modèle connaît plusieurs limitations.

Lors du parcours de l'arbre, les anomalies entre les traces de simulation et expérimentales sont détectées par l'intermédiaire de la fonction $D1$ qui teste le nombre d'occurrences d'une métrique par rapport à une valeur de seuil. À la fin du parcours de l'arbre, une liste de cause racine est proposée à l'utilisateur. Le cas illustré dans l'exemple ci-dessous peut alors se produire :

Exemple (1) Soit le modèle RCA définit par les nœuds A , B , C et D . Le nœud A est le nœud fils de l'état initial de l'arbre et les nœuds B , C et D sont les fils de A . Pendant le parcours de l'arbre, $V(A)$, $V(B)$ et $V(C)$ sont évalués à *vrai* tandis que $V(D)$ est évalué à *faux*. Alors l'ensemble des causes racines de A sera $U = \{B; C\}$ avec $P(B) = P(C)$ les probabilités respectives des causes B et C .

Dans ce cas et étant données les probabilité égale pour chacune des causes racines du nœud A , la détection finale de la cause principale revient à l'utilisateur.

De plus, le modèle RCA utilisé est un modèle qualitatif, qui ne prend pas en compte les temps d'apparition des différents événements ni leur ordre d'apparition. Par exemple, dans le cadre de la comparaison de l'implémentation d'un algorithme de choix de modulation à chaînes de retransmissions (*e.g.* Minstrel, voir partie 2.3.2.1), la séquence d'apparition des différents modes de modulation des trames à son importance. Par conséquent, l'utilisation du modèle peut donc être restreint à certains scénarios.

Pour résoudre ces limitations et appliquer le modèle sur un plus grand nombre de simulateurs et de scénarios, celui-ci doit devenir plus complexe et être complètement automatisé. D'une part, un modèle amélioré pourrait s'appuyer sur un calcul de corrélation entre les valeurs de différence du nœud père et celles des nœuds fils. Par exemple si $C(a, b)$ est la corrélation entre les séries a et b , l'exemple (1) devient :

Exemple (2) Soit le modèle RCA définit par les nœuds A , B , C et D . Le nœud A est le nœud fils de l'état initial de l'arbre et les nœuds B , C et D sont les fils de A . Pendant le parcours de l'arbre, $V(A)$, $V(B)$ et $V(C)$ sont évalués à *vrai* tandis que $V(D)$ est évalué à *faux*. Alors l'ensemble des causes racines de A sera $U = \{B; C\}$ avec $P(B) > P(C)$ si et seulement si $|C(D(A), D(B))| > |C(D(A), D(C))|$.

D'autres part, la deuxième limitation évoquée pourrait être résolue en utilisant de manière conjointe des modèles de nature différente. Par exemple, le cas des algorithmes à chaînes de retransmission pourrait être traité par usage des automates à états finis. Selon ces exemples, la division des différents problèmes de comparaison et leur traitement par des modèles de nature différente pourra conduire aux traitements de scénarios plus complexes.

Troisièmement, en termes de solutions pour l'adaptabilité des nœuds, les résultats obtenus lors de l'application des algorithmes d'apprentissage DT et SVR permettent d'envisager l'intégration de ces algorithmes dans des outils d'adaptation.

Les téléphones intelligents constitueraient une plate-forme d'expérimentation idéale pour ces applications : ils disposent de plusieurs technologies de communication sans-fil et leurs ressources peuvent être importantes. De plus, leur mobilité intrinsèque leur permet d'expérimenter de nombreux environnements sans-fil différents. Par conséquent, l'utilisation des algorithmes d'apprentissage sur ces téléphones intelligents pourrait leur permettre de générer des modèles d'apprentissage plus efficaces. Des études sont en revanche encore nécessaires pour permettre cette implémentation, il s'agira par exemple de trouver comment sélectionner et intégrer les données mobiles dans les modèles d'apprentissage, ou encore de définir des processus en-ligne capables de déterminer les meilleurs paramètres pour l'algorithme SVR.

Conclusion

Les résultats présentés dans cette thèse démontrent que l'adaptabilité et la fiabilisation des nœuds, nécessaires pour les futurs réseaux sans-fil, requièrent :

- une conscience accrue des protocoles de haut-niveau quand aux spécificités du médium et de la pollution spectrale qu'ils génèrent.
- plus de mesures et d'introspection à tous les niveaux de la pile mais surtout aux niveaux des couches physiques et MAC. Les modèles de représentation du réseau utilisés par les nœuds doivent devenir plus complexes.
- plus de réflexion à bas niveau quand à la manière de communiquer. Ces réflexions doivent également être mieux organisées. Les nœuds doivent devenir autonomes et pouvoir apprendre.
- des moyens d'action plus importants à bas niveau pour pouvoir prendre en compte les résultats des analyses, de l'apprentissage et des réflexions.

Ce dernier point nécessitera des moyens matériels plus importants. C'est à ce niveau que se rejoignent nos travaux et ceux traitant des radios logicielles. Les radios logicielles permettent de pouvoir modifier la méthodes et les caractéristiques d'émission-réception à partir d'un même équipement. Aujourd'hui ces solutions sont peu démocratisées car difficiles à intégrer dans les applications mobiles et

embarquées en particulier pour des fréquences et des débits élevés. Il y a peu, un premier pas vers l'usage massif de la radio logicielle a été franchi par le fondateur *Intel* en concevant un dispositif d'émission-réception 802.11 dont la majeure partie des composants sont de nature numérique et implémentés sur une seule et même puce [98]. Ce dispositif va à l'encontre des critiques adressées à la radio logicielle car il permet de moins consommer d'énergie et de baisser le coût des équipements de communication.

À terme, la conjonction de ces solutions matérielles et logicielles (réflexion, autonomie et représentation du réseau) pourraient constituer le point de départ de l'intégration des réseaux sans-fil ubiquitaires et de l'internet des objets.

Bibliographie

- [1] Avila GW2348-4 Network Platform. <http://www.gateworks.com/products/avila/gw2348-4.php>.
- [2] Ettus Research. <http://www.ettus.com/>.
- [3] Gamme USRP National Instruments. <http://www.ni.com/usrp/>.
- [4] Mini-circuits Coaxial Power Splitter / Combiner ZX10R-14+ Datasheet.
- [5] Ping tool website. <http://ftp.arl.army.mil/~mike/ping.html>.
- [6] 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 802.11-1997*, pages i-445, 1997.
- [7] Ieee standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks specific requirements part 15.4 : Wireless medium access control (mac) and physical layer (phy) specifications for low-rate wireless personal area networks (lr-wpans). *IEEE Std 802.15.4-2003*, pages 0-670, 2003.
- [8] Chipcon CC2420 Datasheet : <http://focus.ti.com/lit/ds/symlink/cc2420.pdf>, Texas Instruments, 2007.
- [9] 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 - redline. *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999) - Redline*, pages 1-1238, June 2007.
- [10] Ieee standard for information technology- local and metropolitan area networks- specific requirements- part 11 : Wireless lan medium access control (mac)and physical layer (phy) specifications amendment 5 : Enhancements for higher throughput. *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pages 1-565, 2009.
- [11] Business insider article : The 'internet of things' will be bigger than the smart-phone, tablet, and pc markets combined, 2014.

- [12] P. K. Agrawal, B. B. Gupta, and S. Jain. Svm based scheme for predicting number of zombies in a ddos attack. In *Intelligence and Security Informatics Conference (EISIC), 2011 European*, pages 178–182, 2011.
- [13] Ian F. Akyildiz, Won-Yeol Lee, Mehmet C. Vuran, and Shantidev Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks : A survey. *Comput. Netw.*, 50(13) :2127–2159, September 2006.
- [14] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3) :pp. 175–185, 1992.
- [15] L. Angrisani and M. Vadursi. Cross-layer measurements for a comprehensive characterization of wireless networks in the presence of interference. *Instrumentation and Measurement, IEEE Transactions on*, 56(4) :1148–1156, Aug 2007.
- [16] Ath5k documentation. <http://wireless.kernel.org/en/users/Drivers/ath5k>.
- [17] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things : A survey. *Computer Networks*, 54(15) :2787 – 2805, 2010.
- [18] Nicola Baldo, Manuel Requena-Esteso, José Núñez Martínez, Marc Portolès-Comeras, Jaume Nin-Guerrero, Paolo Dini, and Josep Mangues-Bafalluy. Validation of the ieee 802.11 mac model in the ns3 simulator using the extreme testbed. In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques, SIMUTools '10*, pages 64 :1–64 :9, ICST, Brussels, Belgium, Belgium, 2010. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [19] J. Bardwell. Converting Signal Strength Percentage to dBm Values. http://www.wildpackets.com/elements/whitepapers/Converting_Signal_Strength.pdf, November 2002.
- [20] Debasish Basak, Srimanta Pal, and Dipak C. Patranabis. Support Vector Regression. *Neural Information Processing – Letters and Reviews*, 11, October 2007.
- [21] Ken Behrendt, Ken Fodero, and Schweitzer Engineering Laboratories. The perfect time : An examination of timesynchronization techniques. In *In DistributedTECH*, 2006.
- [22] Lorenzo Bergamini, Carlo Crociani, Andrea Vitaletti, and Michele Nati. Validation of wsn simulators through a comparison with a real testbed. In *Proceedings of the 7th ACM Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, PE-WASUN '10*, pages 103–104, New York, NY, USA, 2010. ACM.
- [23] G. Bianchi, A. Di Stefano, C. Giaconia, L. Scalia, G. Terrazzino, and I. Tinirello. Experimental assessment of the backoff behavior of commercial ieee 802.11b network cards. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 1181–1189, May 2007.

- [24] A. Bildea. *Thèse : Link Quality in Wireless Sensor Networks*. 2013.
- [25] Eric Blossom. Gnu radio : tools for exploring the radio frequency spectrum. *Linux Journal*, 2004(122) :4–, June 2004. <http://dl.acm.org/citation.cfm?id=993247.993251>.
- [26] SIG Bluetooth. Specification of the bluetooth system, version 1.1. <http://www.bluetooth.com>, 2001.
- [27] B. Braden, National Science Foundation (U.S.), National Science Foundation (U.S.). Directorate for Computer, Information Science, Engineering, National Science Foundation (U.S.). Advanced Networking Infrastructure, and Research Division. *Report of NSF Workshop on Network Research Testbeds*. Directorate for Computer and Information Science and Engineering (CISE), Advanced Networking Infrastructure & Research Division, 2002.
- [28] L. Breiman. *Classification and regression trees*. The Wadsworth and Brooks-Cole statistics-probability series. Chapman & Hall, 1984.
- [29] J. Burgess, Brian Gallagher, D. Jensen, and B.N. Levine. Maxprop : Routing for vehicle-based disruption-tolerant networks. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–11, 2006.
- [30] Clément Burin Des Rosiers, Guillaume Chelius, Eric Fleury, Antoine Fraboulet, Antoine Gallais, Nathalie Mitton, and Thomas Noël. SensLAB Very Large Scale Open Wireless Sensor Network Testbed. In *Proc. 7th International ICST Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCOM)*, Shanghai, Chine, April 2011.
- [31] G. Chandrasekaran, M.A. Ergin, Jie Yang, Song Liu, Yingying Chen, M. Gruteser, and R.P. Martin. Empirical evaluation of the limits on localization using signal strength. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. 6th Annual IEEE Communications Society Conference on*, pages 1–9, June 2009.
- [32] Yu-Chung Cheng, John Bellardo, Péter Benkő, Alex C. Snoeren, Geoffrey M. Voelker, and Stefan Savage. Jigsaw : Solving the puzzle of enterprise 802.11 analysis. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '06*, pages 39–50, New York, NY, USA, 2006. ACM.
- [33] Ugo Maria Colesanti, Carlo Crociani, and Andrea Vitaletti. On the accuracy of omnet++ in the wireless sensornetworks domain : Simulation vs. testbed. In *Proceedings of the 4th ACM Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, PE-WASUN '07*, pages 25–31, New York, NY, USA, 2007. ACM.
- [34] Robert C. Daniels. *Thèse : Machine Learning for Link Adaptation in Wireless Networks*. 2011.

- [35] Robert C. Daniels, Ketan Mandke, Steven W. Peters, Scott M. Nettles, and Robert W. Heath, Jr. Machine learning for physical layer link adaptation in multiple-antenna wireless networks. In *Proceedings of the Third ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization*, WiNTECH '08, pages 113–114, New York, NY, USA, 2008. ACM.
- [36] D. Dujovne, T. Turletti, and F. Filali. A taxonomy of ieee 802.11 wireless parameters and open source measurement tools. *Commun. Surveys Tuts.*, 12(2) :249–262, April 2010.
- [37] E. O. Elliott. Estimates of Error Rates for Codes on Burst-Noise Channels. *Bell System Technical Journal*, 42 :1977–1997, September 1963.
- [38] Hossein Falaki, Dimitrios Lymberopoulos, Ratul Mahajan, Srikanth Kandula, and Deborah Estrin. A first look at traffic on smartphones. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, pages 281–287, New York, NY, USA, 2010. ACM.
- [39] Shahin Farahani. *ZigBee Wireless Networks and Transceivers*. Newnes, Newton, MA, USA, 2008.
- [40] Cisco VNI Forecast. Cisco visual networking index : Global mobile data traffic forecast update 2012-2017. *Cisco Public Information*, May, 9, 2013.
- [41] H.T. Friis. A note on a simple transmission formula. *Proceedings of the IRE*, 34(5) :254–256, 1946.
- [42] M. Gast. *802.11 wireless networks : the definitive guide*. Definitive Guide Series. O'Reilly, 2005. ISBN 9780596100520.
- [43] Mark Gates, Ajay Tirumala, Jon Dugan, and Kevin Gibbs. *Iperf version 2.0.0*. NLANR applications support, University of Illinois at Urbana-Champaign, Urbana, IL, USA, May 2004.
- [44] E. N. Gilbert. Capacity of a burst-noise channel. *Bell System Technical Journal*, 39 :1253–1265, September 1960.
- [45] GNU Radio Website, accessed February 2012.
- [46] G. P. Halkes and K. G. Langendoen. Experimental evaluation of simulation abstractions for wireless sensor network mac protocols. *EURASIP J. Wirel. Commun. Netw.*, 2010 :24 :1–24 :2, April 2010.
- [47] Simon Haykin. Cognitive radio : brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2) :201–220, Feb 2005.
- [48] Tristan Henderson, David Kotz, and Ilya Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. Technical Report TR2004-496, Dartmouth College, Computer Science, Hanover, NH, March 2004.

- [49] Gavin Holland, Nitin Vaidya, and Paramvir Bahl. A rate-adaptive mac protocol for multi-hop wireless networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom '01, pages 236–251, New York, NY, USA, 2001. ACM.
- [50] Niels Hoven, Rahul Tandra, and Anant Sahai. Some fundamental limits on cognitive radio. *Wireless Foundations EECS, Univ. of California, Berkeley*, 2005.
- [51] Wendong Hu. On-demand spectrum contention for inter-cell spectrum sharing in cognitive radio networks, January 15 2009. US Patent App. 12/354,606.
- [52] National Instruments. Ethernet instrument controllers ni pci-1588. Technical report, National Instruments, 2006.
- [53] Svilen Ivanov, Andre Herms, and Georg Lukas. Experimental validation of the ns-2 wireless model using simulation, emulation, and real network. In *Communication in Distributed Systems (KiVS), 2007 ITG-GI Conference*, pages 1–12, Feb 2007.
- [54] Glenn Judd and Peter Steenkiste. Repeatable and realistic wireless experimentation through physical emulation. *Computer Communication Review*, 34(1) :63–68, 2004.
- [55] Glenn Judd, Xiaohui Wang, and Peter Steenkiste. Low-overhead channel-aware rate adaptation. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, MobiCom '07, pages 354–357, New York, NY, USA, 2007. ACM.
- [56] Heikki Kaaranen, Ari Ahtiainen, Lauri Laitinen, Siamk Naghian, and Valtteri Niemi. Umts networks. *Architecture, Mobility and Services*. Wiley, 2001.
- [57] Partha Kanuparth, Constantine Dovrolis, Konstantina Papagiannaki, Srinivasan Seshan, and Peter Steenkiste. Can user-level probing detect and diagnose common home-wlan pathologies. *Computer Communication Review*, 42(1) :7–15, 2012.
- [58] P.B. Kenington and D.W. Bennett. Field measurements study into the potential effects of ism emissions on cellular radio equipment. *Science, Measurement and Technology, IEE Proceedings -*, 144(3) :134–140, 1997.
- [59] Shahbaz Khan, Bilal Aziz, Sundas Najeeb, Aziz Ahmed, Muhammad Usman, and Sadiq Ullah. Reliability of network simulators and simulation based research. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pages 180–185, Sept 2013.
- [60] Rabiullah Khattak, Anna Chaltseva, Laurynas Riliskis, Ulf Bodin, and Evgeny Osipov. Comparison of wireless network simulators with multihop wireless network testbed in corridor environment. In *Proceedings of the 9th IFIP TC 6 International Conference on Wired/Wireless Internet Communications, WWIC'11*, pages 80–91, Berlin, Heidelberg, 2011. Springer-Verlag.

- [61] Sea-Moon Kim, Chong-Moo Lee, Young-Chol Choi, Jong-Won Park, and Young-Kon Lim. An experimental model of surface reflection in the underwater anechoic basin at kriso. In *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*, pages 257–260, 2002.
- [62] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. *Wirel. Netw.*, 11(1-2) :115–133, January 2005.
- [63] David Kotz, Calvin Newport, and Chip Elliott. The mistaken axioms of wireless-network research. 2003.
- [64] David Kotz, Calvin Newport, Robert S. Gray, Jason Liu, Yougu Yuan, and Chip Elliott. Experimental evaluation of wireless simulation assumptions. In *Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '04, pages 78–82, New York, NY, USA, 2004. ACM.
- [65] Stuart Kurkowski, Tracy Camp, and Michael Colagrosso. Manet simulation studies : The incredibles. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(4) :50–61, October 2005.
- [66] K.G. Kyriakopoulos, W.G. Whittow, and D.J. Parish. A framework for cross-layer measurements in wireless networks. In *Telecommunications, 2009. AICT '09. Fifth Advanced International Conference on*, pages 237–242, May 2009.
- [67] Mathieu Lacage and Thomas R. Henderson. Yet another network simulator. In *Proceeding from the 2006 Workshop on Ns-2 : The IP Network Simulator*, WNS2 '06, New York, NY, USA, 2006. ACM.
- [68] Mathieu Lacage, Mohammad Hossein Manshaei, and Thierry Turletti. Ieee 802.11 rate adaptation : A practical approach. In *Proceedings of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '04, pages 126–134, New York, NY, USA, 2004. ACM.
- [69] Nicolas Larrieu and Philippe Owezarski. De l'utilisation des mesures de trafic pour l'ingénierie des réseaux de l'internet. *Technique et Science Informatiques*, 23(5-6) :701–731, 2004.
- [70] A.M. Law and W.D. Kelton. *Simulation modeling and analysis*. McGraw-Hill series in industrial engineering and management science. McGraw-Hill, 2000.
- [71] HyungJune Lee, A. Cerpa, and P. Levis. Improving wireless simulation through noise modeling. In *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, pages 21–30, April 2007.
- [72] Feng Li, Mingzhe Li, Rui Lu, Huahui Wu, Claypool Mark, and Kinicki Robert. Tools and techniques for measurement of ieee 802.11 wireless networks. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, pages 1–8, April 2006.

- [73] Wei Li, Min Li, Ruiting Fan, and Lanjun Li. A fault diagnosis method based on decision tree for wireless mesh network. In *Communication Technology (ICCT), 2010 12th IEEE International Conference on*, pages 231–234, Nov 2010.
- [74] Ying-Chang Liang, Yonghong Zeng, E.C.Y. Peh, and Anh Tuan Hoang. Sensing-throughput tradeoff for cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 7(4) :1326–1337, April 2008.
- [75] Xiaobo Long and B. Sikdar. A wavelet based long range signal strength prediction in wireless networks. In *Communications, 2008. ICC '08. IEEE International Conference on*, pages 2043–2047, May 2008.
- [76] Gilberto Flores Lucio, Marcos Paredes-Farrera, Emmanuel Jammeh, Martin Fleury, and Martin J Reed. Opnet modeler and ns-2 : Comparing the accuracy of network simulators for packet-level analysis using a network testbed. *WSEAS Transactions on Computers*, 2(3) :700–707, 2003.
- [77] J.T. MacDonald and D.R. Ucci. Interference temperature limits of ieee 802.11 protocol radio channels. In *Electro/Information Technology, 2007 IEEE International Conference on*, pages 64–69, 2007.
- [78] Ratul Mahajan, Maya Rodrig, David Wetherall, and John Zahorjan. Analyzing the mac-level behavior of wireless networks in the wild. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '06*, pages 75–86, New York, NY, USA, 2006. ACM.
- [79] R. Mahindra, G.D. Bhanage, G. Hadjichristofi, I. Seskar, D. Raychaudhuri, and Y.Y. Zhang. Space versus time separation for wireless virtualization on an indoor grid. In *Next Generation Internet Networks, 2008. NGI 2008*, pages 215–222, April 2008.
- [80] Gregor Maier, Anja Feldmann, Vern Paxson, and Mark Allman. On dominant characteristics of residential broadband internet traffic. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC '09*, pages 90–102, New York, NY, USA, 2009. ACM.
- [81] David L. Mills. Internet time synchronization : the network time protocol. *IEEE Transactions on Communications*, 39(10) :1482–1493, 1991.
- [82] Robert T. Morris, John C. Bicket, and John C. Bicket. Bit-rate selection in wireless networks. Technical report, Master’s thesis, MIT, 2005.
- [83] Michel Mouly and Marie-Bernadette Pautet. *The GSM System for Mobile Communications*. Telecom Publishing, 1992.
- [84] Ehsan Nourbakhsh, Jeff Dix, Paul Johnson, Ryan Burchfield, S. Venkatesan, Neeraj Mittal, and Ravi Prakash. Assert : A wireless networking testbed. In Thomas Magedanz, Anastasius Gavras, NguyenHuu Thanh, and JeffryS.

- Chase, editors, *Testbeds and Research Infrastructures. Development of Networks and Communities*, volume 46 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 209–218. Springer Berlin Heidelberg, 2011.
- [85] Philippe Owezarski and Nicolas Larrieu. Techniques et outils de métrologie pour l'internet et son trafic. *Techniques de l'ingénieur*, 1090 :1090–22, 2006.
- [86] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [87] C. Phillips, D. Sicker, and D. Grunwald. A survey of wireless path loss prediction and coverage mapping methods. *Communications Surveys Tutorials, IEEE*, 15(1) :255–270, First 2013.
- [88] R. Prakash and E. Nourbakhsh. Gradient-based aggregation in forest of sensors (grafs). In *Parallel Processing (ICPP), 2011 International Conference on*, pages 120–129, Sept 2011.
- [89] J.G. Proakis and D.G. Manolakis. *Digital signal processing*. Pearson Prentice Hall, 2007.
- [90] Daniele Puccinelli and Martin Haenggi. Multipath fading in wireless sensor networks : Measurements and interpretation. In *Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing, IWCMC '06*, pages 1039–1044, New York, NY, USA, 2006. ACM.
- [91] M.B. Pursley and D.J. Taipale. Error probabilities for spread-spectrum packet radio with convolutional codes and viterbi decoding. *Communications, IEEE Transactions on*, 35(1) :1–12, Jan 1987.
- [92] J. Ross Quinlan. *C4.5 : Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [93] Abderrezak Rachedi, Stéphane Lohier, Sylvain Cherrier, and Ismail Salhi. Wireless network simulators relevance compared to a real testbed in outdoor and indoor environments. In *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference, IWCMC '10*, pages 346–350, New York, NY, USA, 2010. ACM.
- [94] Hariharan Rahul, Farinaz Edalat, Dina Katabi, and Charles G. Sodini. Frequency-aware rate adaptation and mac protocols. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, MobiCom '09*, pages 193–204, New York, NY, USA, 2009. ACM.
- [95] K. Ramachandran, H. Kremo, M. Gruteser, P. Spasojevic, and I. Seskar. Scalability analysis of rate adaptation techniques in congested ieee 802.11 networks : An orbit testbed comparative study. In *World of Wireless, Mobile and*

- Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pages 1–12, June 2007.
- [96] J.J. Ramos-munoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J.M. Lopez-soler. Characteristics of mobile youtube traffic. *Wireless Communications, IEEE*, 21(1) :18–25, February 2014.
- [97] Theodore Rappaport. *Wireless Communications : Principles and Practice*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 2001.
- [98] Justin Rattner. Connecting the future : It’s a wireless world, the latest research from intel labs. *Intel Developer Forum*, 2012.
- [99] R.D. Raut and K.D. Kulat. Sdr design for cognitive radio. In *Modeling, Simulation and Applied Optimization (ICMSAO), 2011 4th International Conference on*, pages 1–8, April 2011.
- [100] D. Raychaudhuri, M. Ott, and I. Secker. Orbit radio grid testbed for evaluation of next-generation wireless network protocols. In *Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005. Tridentcom 2005. First International Conference on*, pages 308–309, 2005.
- [101] M. Richardson. TCPDUMP public repository, 2004.
- [102] Thomas Warren Rondeau. *Thèse : Application of Artificial Intelligence to Wireless Communications*. 2007.
- [103] A. Ruiz and P.E. Lopez-de Teruel. Nonlinear kernel-based statistical pattern analysis. *Neural Networks, IEEE Transactions on*, 12(1) :16–32, 2001.
- [104] Inc ULC SANDVINE. Global internet phenomena report, 2011.
- [105] N.I. Sapankevych and Ravi Sankar. Time series prediction using support vector machines : A survey. *Computational Intelligence Magazine, IEEE*, 4(2) :24–38, 2009.
- [106] L. Scalia, J. Widmer, and I. Aad. On the side effects of packet detection sensitivity in iee 802.11 interference management. In *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a*, pages 1–7, 2010.
- [107] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE : the UMTS long term evolution*. Wiley Online Library, 2009.
- [108] R.A Shafik, S. Rahman, and R. Islam. On the extended relationships among evm, ber and snr as performance metrics. In *Electrical and Computer Engineering, 2006. ICECE '06. International Conference on*, pages 408–411, Dec 2006.
- [109] Sangho Shin and Henning Schulzrinne. Experimental measurement of the capacity for voip traffic in iee 802.11 wlans. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 2018–2026, May 2007.

- [110] Vikas B. Shivamurthy and Ravi Prakash. Emulation of multipath effects on a wireless testbed and its impact on mac layer performance. *Wirel. Netw.*, 18(4) :457–468, May 2012.
- [111] M. Siekkinen, G. Urvoy-Keller, E. W. Biersack, and T. En-Najjary. Root cause analysis for long-lived tcp connections. In *Proceedings of the 2005 ACM Conference on Emerging Network Experiment and Technology*, CoNEXT '05, pages 200–210, New York, NY, USA, 2005. ACM.
- [112] Matti Siekkinen, Guillaume Urvoy-Keller, Ernst W. Biersack, and Denis Colange. A root cause analysis toolkit for tcp. *Comput. Netw.*, 52(9) :1846–1858, June 2008.
- [113] R. Skehill, P. Scully, and S. McGrath. Characteristics, results and findings of ieee 802.11 in an rf isolated testbed. In *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, pages 1–5, Sept 2007.
- [114] Steven W. Smith. *The scientist and engineer's guide to digital signal processing*. California Technical Publishing, San Diego, CA, USA, 1997.
- [115] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3), August 2004.
- [116] M. Smolnikar, M. Mihelin, G. Berke, G. Kandus, and M. Mohorcic. Ism bands spectrum sensing based on versatile sensor node platform. In *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*, pages 1–5, 2010.
- [117] Spécification de l'algorithme Minstrel. <http://wireless.kernel.org/en/developers/Documentation/mac80211/RateControl/minstrel/>.
- [118] Spécification de l'algorithme Onoe. <http://madwifi-project.org/wiki/UserDocs/RateControl>.
- [119] Srikanth Sundaresan, Walter de Donato, Nick Feamster, Renata Teixeira, Sam Crawford, and Antonio Pescapè. Broadband internet performance : A view from the gateway. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11, pages 134–145, New York, NY, USA, 2011. ACM.
- [120] Kefeng Tan, D. Wu, An Chan, and P. Mohapatra. Comparing simulation tools and experimental testbeds for wireless mesh networks. In *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a*, pages 1–9, June 2010.
- [121] D. Terasawa and Jr. Tiedemann, E.G. cdmaone(r) (is-95) technology overview and evolution. In *Radio Frequency Integrated Circuits (RFIC) Symposium, 1999 IEEE*, pages 213–216, 1999.
- [122] S. Tertois. *Thèse : Réduction des effets de non-linéarités dans une modulation multiporteuse à l'aide de réseaux de neurones*. 2003.

- [123] E. Test, V. Kecman, R. Strack, Qi Li, and R. Salman. Feature ranking for pattern recognition : A comparison of filter methods. In *Southeastcon, 2012 Proceedings of IEEE*, pages 1–5, 2012.
- [124] E. Test, L. Zigic, and V. Kecman. Feature ranking using gini index, scatter ratios, and nonlinear svm rfe. In *Southeastcon, 2013 Proceedings of IEEE*, pages 1–5, 2013.
- [125] M. Valkama, A. Springer, and G. Hueber. Digital signal processing for reducing the effects of rf imperfections in radio devices 2014; an overview. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 813–816, 2010.
- [126] Eduard Garcia Villegas, Elena Lopez-Aguilera, R. Vidal, and J. Paradells. Effect of adjacent-channel interference in ieee 802.11 wlans. In *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*, pages 118–125, 2007.
- [127] B. Walker and J. Seastrom. Addressing scalability in a laboratory-based multihop wireless testbed. In *Testbeds and Research Infrastructures for the Development of Networks Communities and Workshops, 2009. TridentCom 2009. 5th International Conference on*, pages 1–10, April 2009.
- [128] Yong Wang, Margaret Martonosi, and Li-Shiuan Peh. Predicting link quality using supervised learning in wireless sensor networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 11(3) :71–83, July 2007.
- [129] T. Watteyne, S. Lanzisera, A. Mehta, and K.S.J. Pister. Mitigating multipath fading through channel hopping in wireless sensor networks. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5, 2010.
- [130] Mark Weiser. The computer for the 21st century. *Scientific american*, 265(3) :94–104, 1991.
- [131] Serge Willenegger. cdma2000 physical layer : An overview. *Communications and Networks, Journal of*, 2(1) :5–17, March 2000.
- [132] Georg Wittenburg and Jochen Schiller. A quantitative evaluation of the simulation accuracy of wireless sensor networks. *Proceedings of 6. Fachgespräch “Drahtlose Sensornetze” der GI/ITG-Fachgruppe “Kommunikation und Verteilte Systeme*, pages 23–26, 2007.
- [133] M. Woehrle, M.C. Bor, and K.G. Langendoen. 868 mhz : a noiseless environment, but no free lunch for protocol design. In *9th int. conf. on Networked Sensing Systems*, INSS, pages 1–8, jun 2012.
- [134] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1), December 2007.

- [135] Xu Xiang-Hua, Zhou Biao, and Wan Jian. Tree topology based fault diagnosis in wireless sensor networks. In *Wireless Networks and Information Systems, 2009. WNIS '09. International Conference on*, pages 65–69, Dec 2009.
- [136] Y. Yohannes and P. Webb. *Classification and Regression Trees, CART : A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. Microcomputers in policy research. International Food Policy Research Institute, 1999.
- [137] Jiansong Zhang, K. Tan, Jun Zhao, Haitao Wu, and Yongguang Zhang. A practical snr-guided rate adaptation. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages –, 2008.
- [138] Jerry Zhao and Ramesh Govindan. Understanding packet delivery performance in dense wireless sensor networks. In *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, SenSys '03*, pages 1–13, New York, NY, USA, 2003. ACM.
- [139] Youping Zhao, B.G. Agee, and J.H. Reed. Simulation and measurement of microwave oven leakage for 802.11 wlan interference management. In *Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, 2005. MAPE 2005. IEEE International Symposium on*, volume 2, pages 1580–1583 Vol. 2, 2005.

Annexe A

Graphiques des estimations par apprentissage automatique

Sommaire

A.1 Estimations par SVR	135
A.1.1 Estimation de SVR pour la métrique <i>ATH_RSS</i>	135
A.1.2 Estimation de SVR pour la métrique <i>LECR_RSS</i>	136
A.1.3 Estimation de SVR pour la métrique <i>LECR_NOISE</i>	137
A.1.4 Estimation de SVR pour la métrique <i>LECR_SNR</i>	138
A.1.5 Estimation de SVR pour la métrique <i>ATH_RSS+LECR_NOISE</i>	139
A.1.6 Estimation de SVR pour la métrique <i>LECR_RSS + LECR_NOISE</i>	140
A.2 Estimations par <i>k</i>-nn	141
A.2.1 Estimation de <i>k</i> -nn pour la métrique <i>ATH_RSS</i>	141
A.2.2 Estimation de <i>k</i> -nn pour la métrique <i>LECR_RSS</i>	142
A.2.3 Estimation de <i>k</i> -nn pour la métrique <i>LECR_NOISE</i>	143
A.2.4 Estimation de <i>k</i> -nn pour la métrique <i>LECR_SNR</i>	144
A.2.5 Estimation de <i>k</i> -nn pour la métrique <i>ATH_RSS+LECR_NOISE</i>	145
A.2.6 Estimation de <i>k</i> -nn pour la métrique <i>LECR_RSS + LECR_NOISE</i>	146
A.3 Estimations par DT	147
A.3.1 Estimation de DT pour la métrique <i>ATH_RSS</i>	147
A.3.2 Estimation de DT pour la métrique <i>LECR_RSS</i>	148
A.3.3 Estimation de DT pour la métrique <i>LECR_NOISE</i>	149
A.3.4 Estimation de DT pour la métrique <i>LECR_SNR</i>	150
A.3.5 Estimation de DT pour la métrique <i>ATH_RSS+LECR_NOISE</i>	151
A.3.6 Estimation de DT pour la métrique <i>LECR_RSS+LECR_NOISE</i>	152

Cette annexe présente les graphiques issus des estimations par apprentissage automatique présenté en chapitre 6. Chaque graphique présente les estimations pour une combinaison d'algorithme, de jeux d'entraînement et de métrique particulière utilisées pour obtenir les estimations. Cette combinaison est exprimée dans le titre des figures sous le format A.1.

$$\{\text{algorithme, jeux d'entraînement, métrique}\}. \quad (\text{A.1})$$

A.1 Estimations par SVR

A.1.1 Estimation de SVR pour la métrique ATH_RSS .

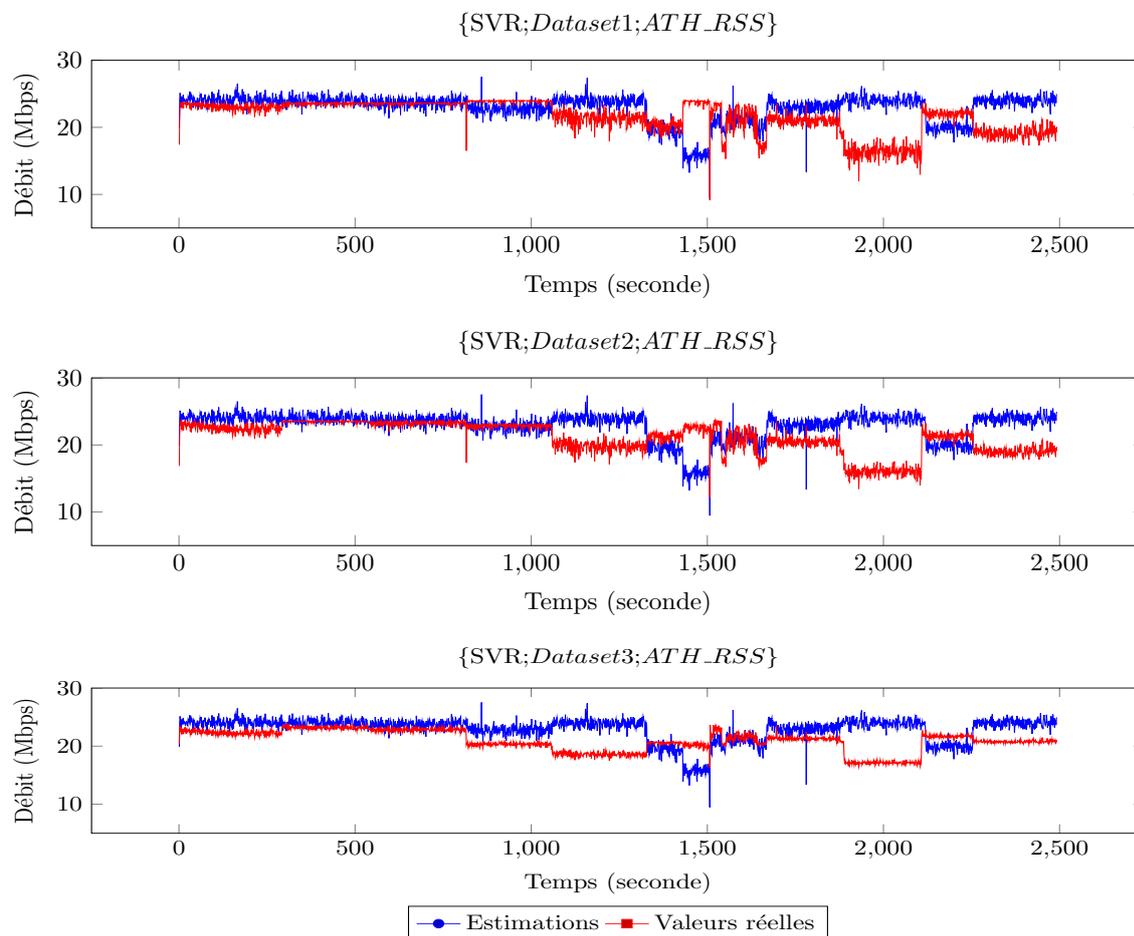


FIGURE A.1 – Graphiques des estimations de SVR pour la métrique ATH_RSS .

A.1.2 Estimation de SVR pour la métrique $LECR_RSS$.

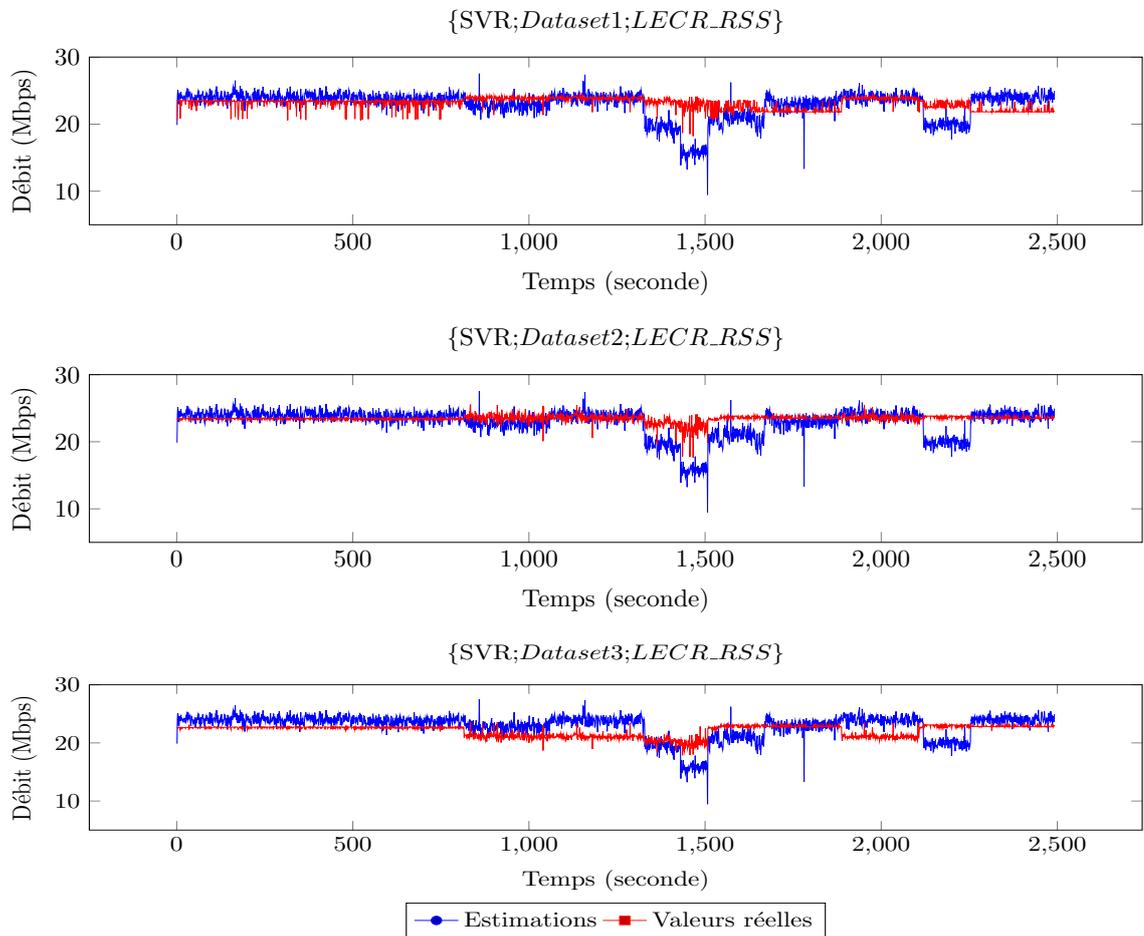


FIGURE A.2 – Graphiques des estimations de SVR pour la métrique $LECR_RSS$.

A.1.3 Estimation de SVR pour la métrique *LECR_NOISE*.

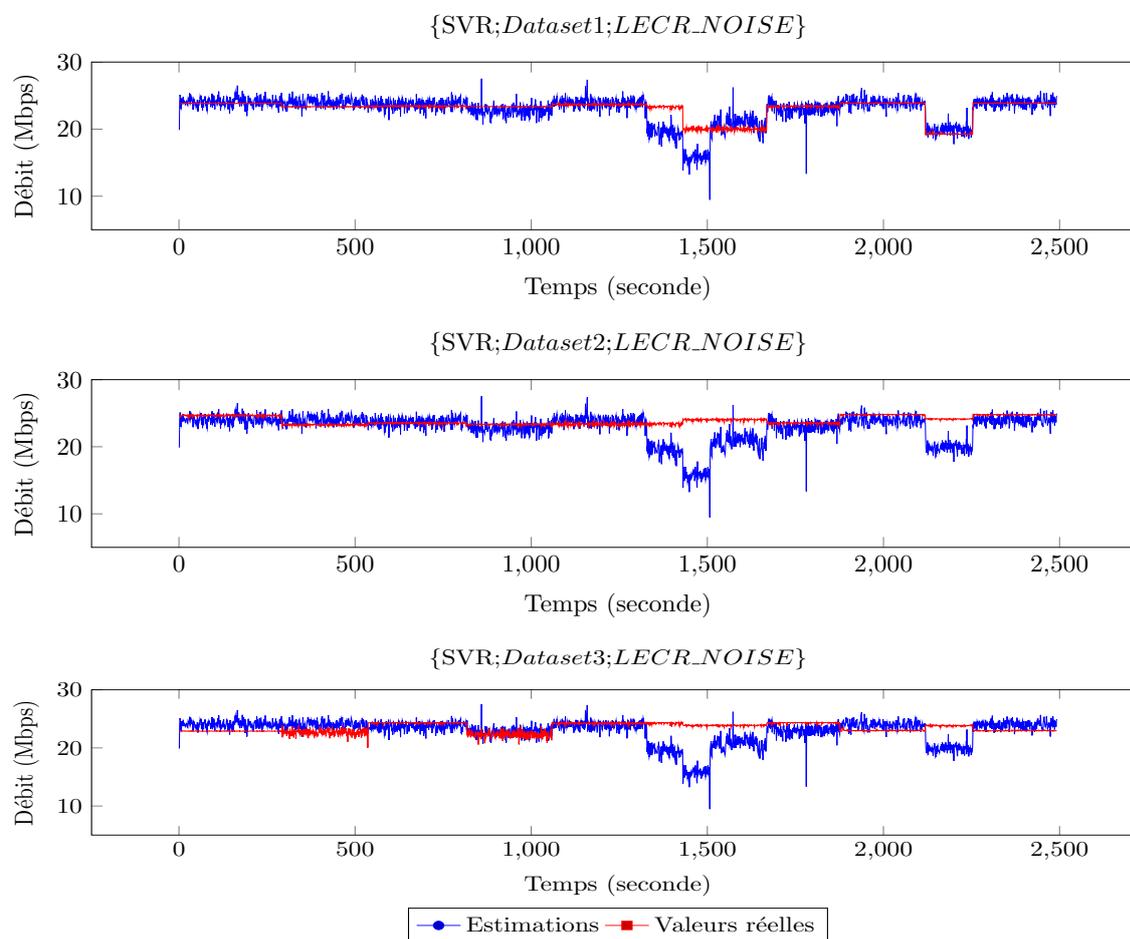


FIGURE A.3 – Graphiques des estimations de SVR pour la métrique *LECR_NOISE*.

A.1.4 Estimation de SVR pour la métrique $LECR_SNR$.

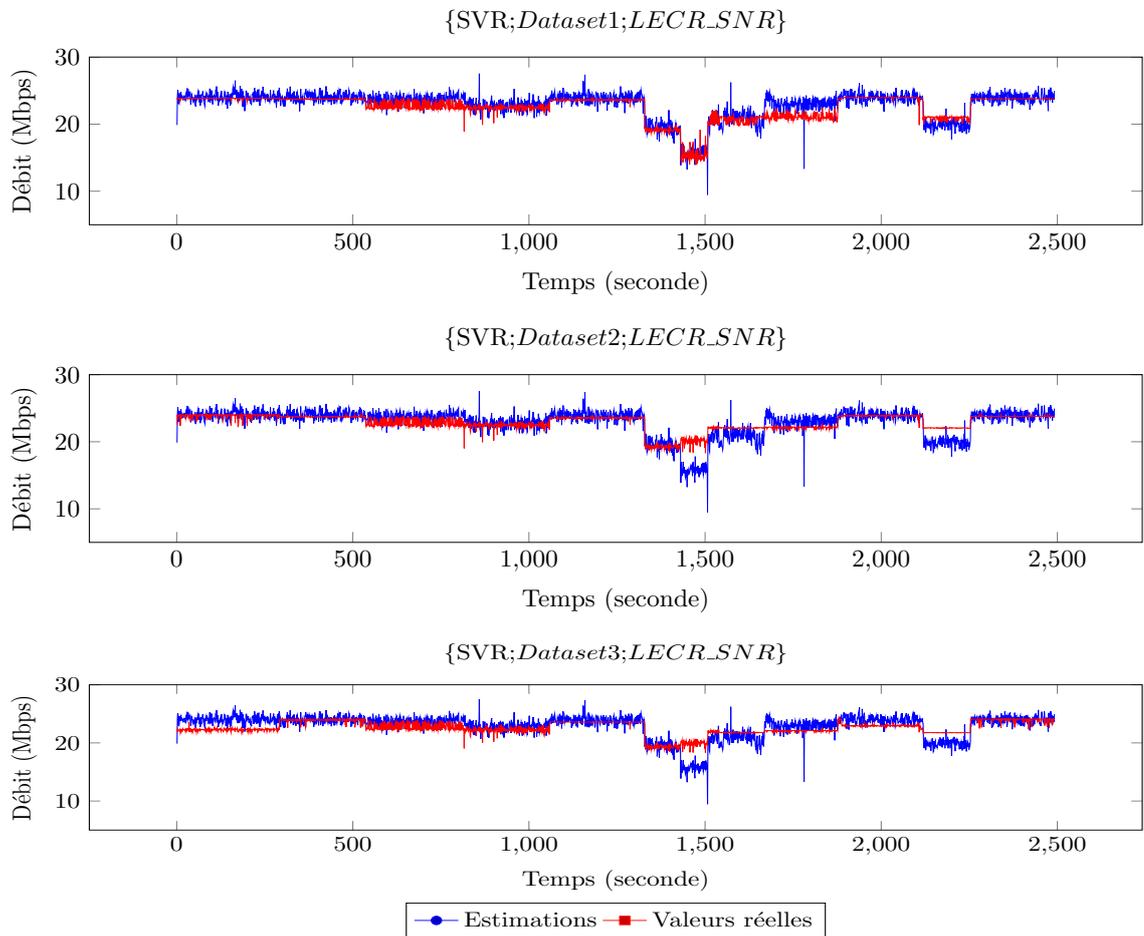


FIGURE A.4 – Graphiques des estimations de SVR pour la métrique $LECR_SNR$.

A.1.5 Estimation de SVR pour la métrique $ATH_RSS + LEQR_NOISE$.

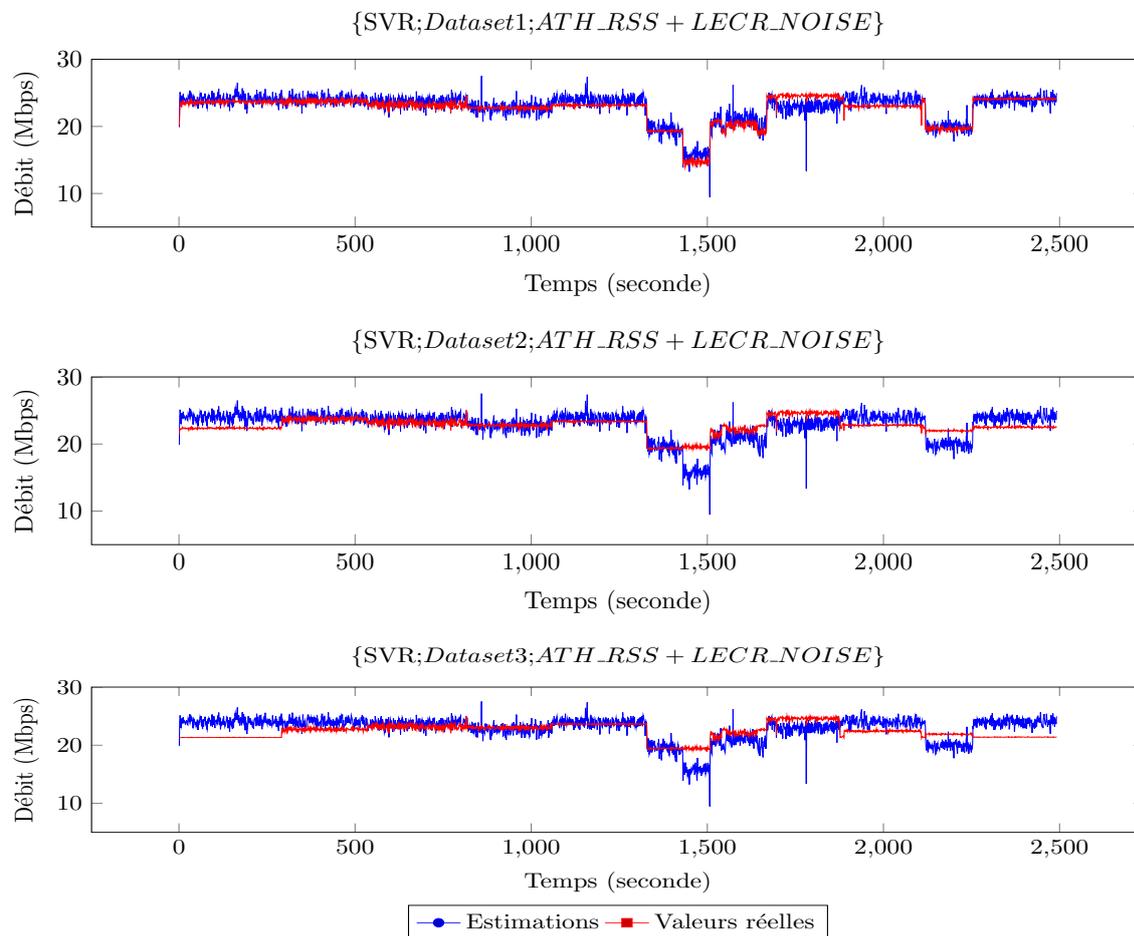


FIGURE A.5 – Graphiques des estimations de SVR pour la métrique $ATH_RSS + LEQR_NOISE$.

A.1.6 Estimation de SVR pour la métrique $LECR_RSS + LECR_NOISE$.

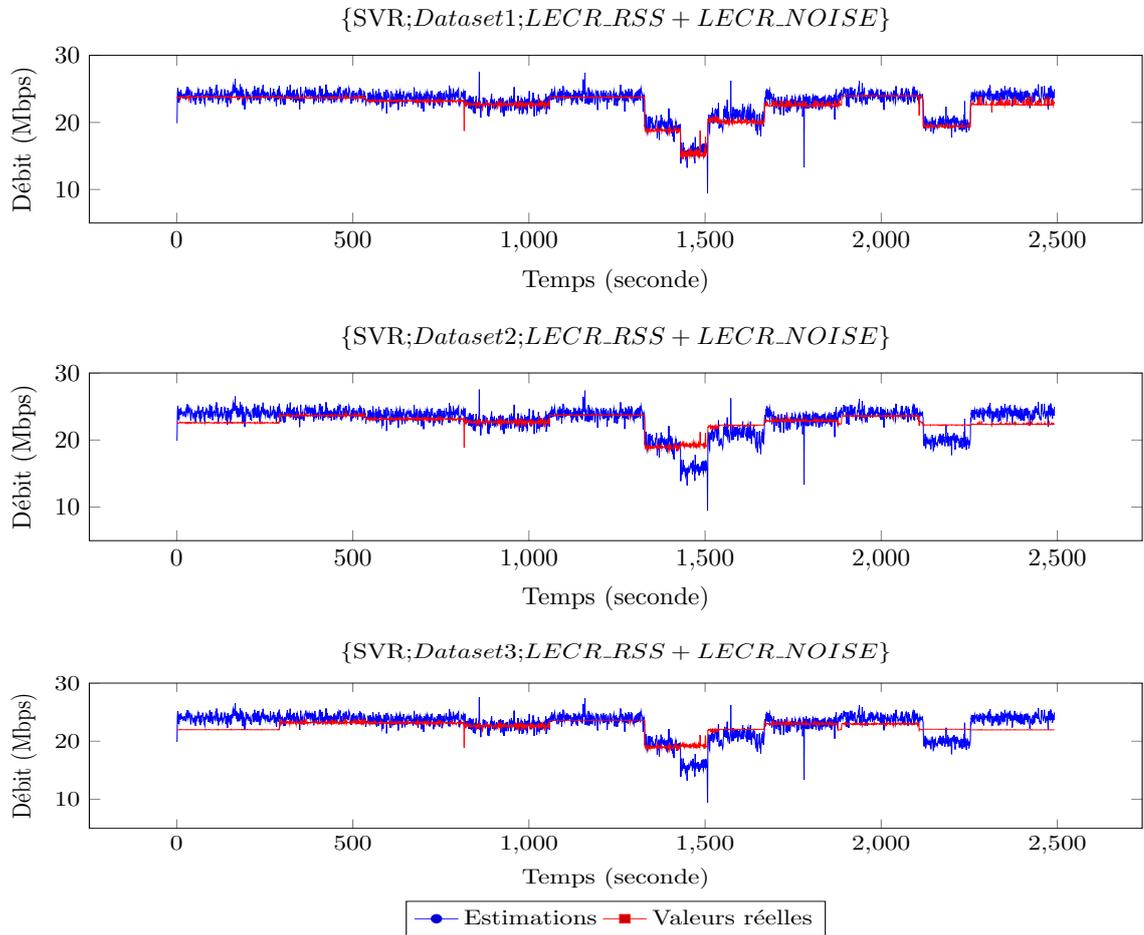


FIGURE A.6 – Graphiques des estimations de SVR pour la métrique $LECR_RSS + LECR_NOISE$.

A.2 Estimations par k -nn

A.2.1 Estimation de k -nn pour la métrique ATH_RSS .

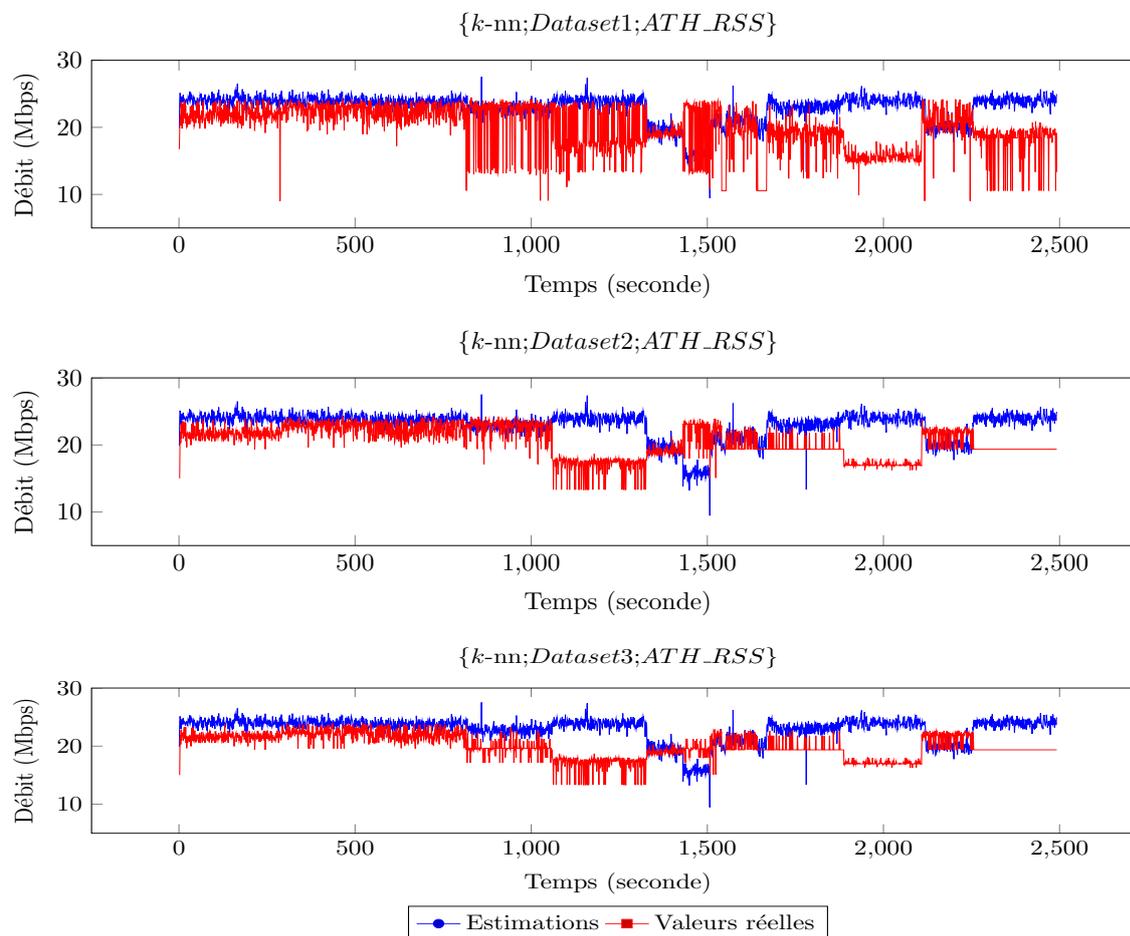


FIGURE A.7 – Graphiques des estimations de k -nn pour la métrique ATH_RSS .

A.2.2 Estimation de k -nn pour la métrique $LECR_RSS$.

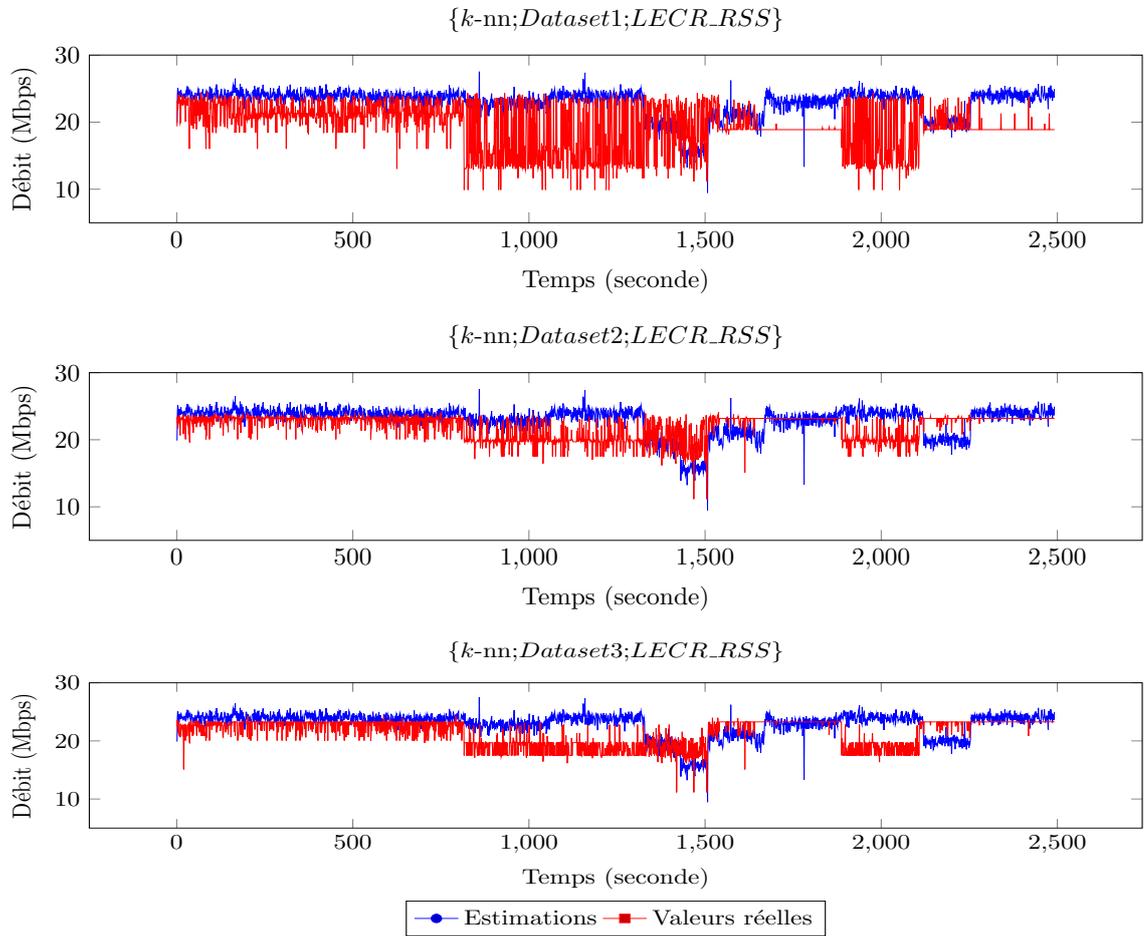


FIGURE A.8 – Graphiques des estimations de k -nn pour la métrique $LECR_RSS$.

A.2.3 Estimation de k -nn pour la métrique $LECR_NOISE$.

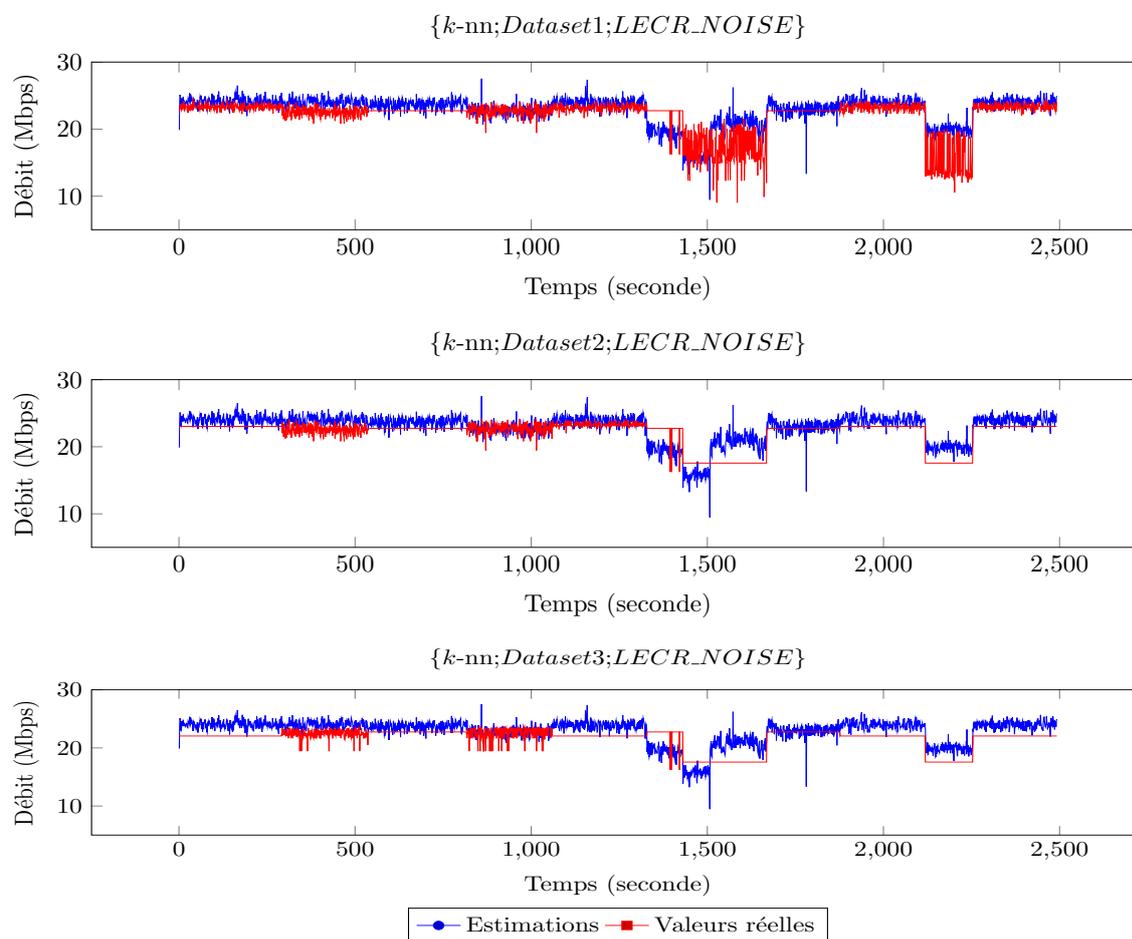


FIGURE A.9 – Graphiques des estimations de k -nn pour la métrique $LECR_NOISE$.

A.2.4 Estimation de k -nn pour la métrique $LECR_SNR$.

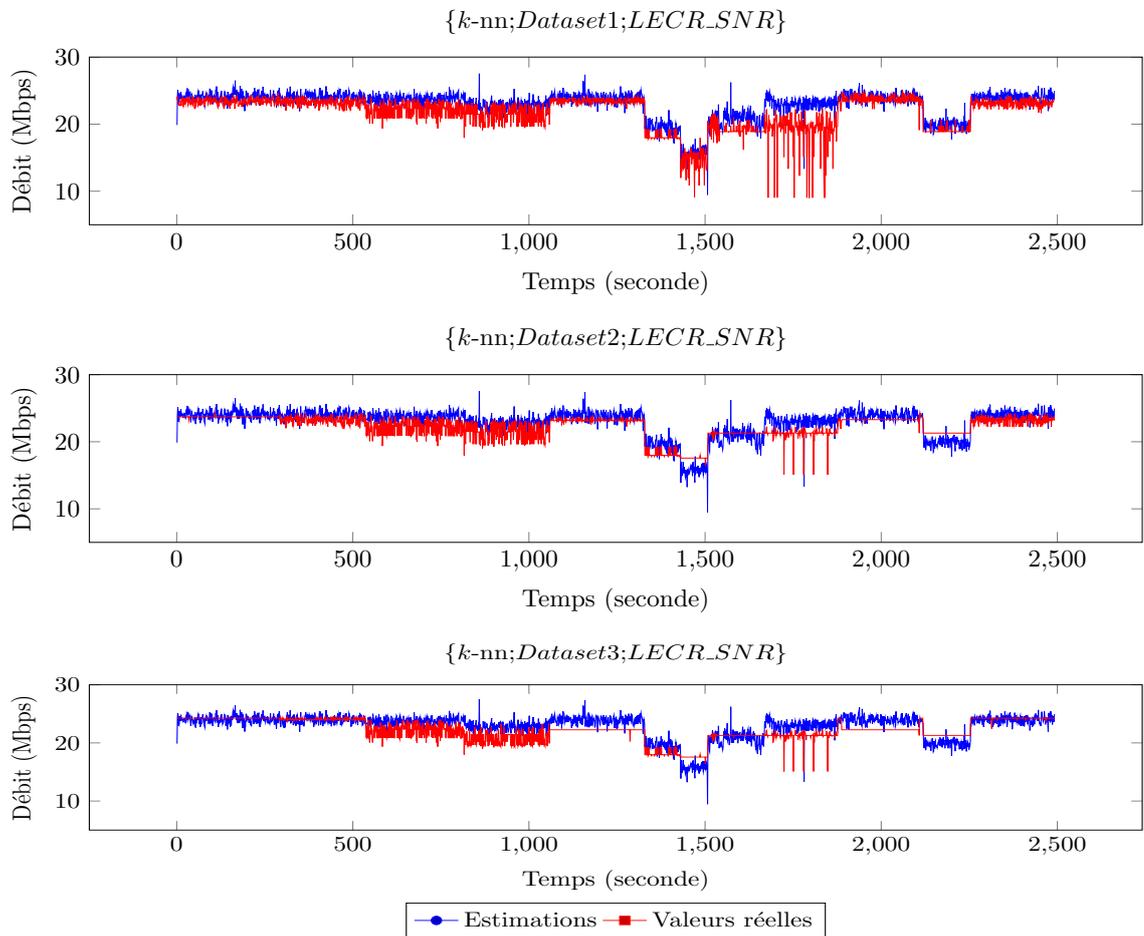


FIGURE A.10 – Graphiques des estimations de k -nn pour la métrique $LECR_SNR$.

A.2.5 Estimation de k -nn pour la métrique $ATH_RSS + LEQR_NOISE$.

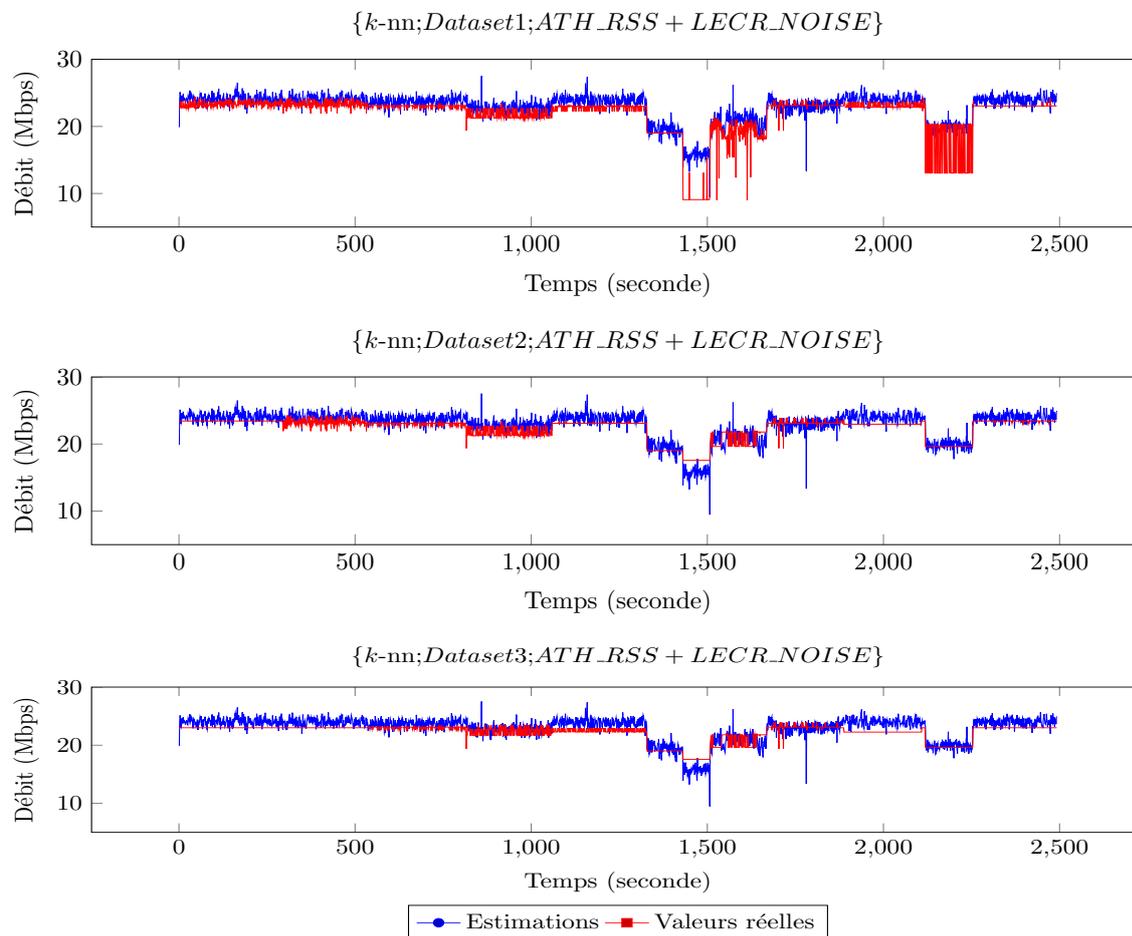


FIGURE A.11 – Graphiques des estimations de k -nn pour la métrique $ATH_RSS + LEQR_NOISE$.

A.2.6 Estimation de k -nn pour la métrique $LECR_RSS + LECR_NOISE$.

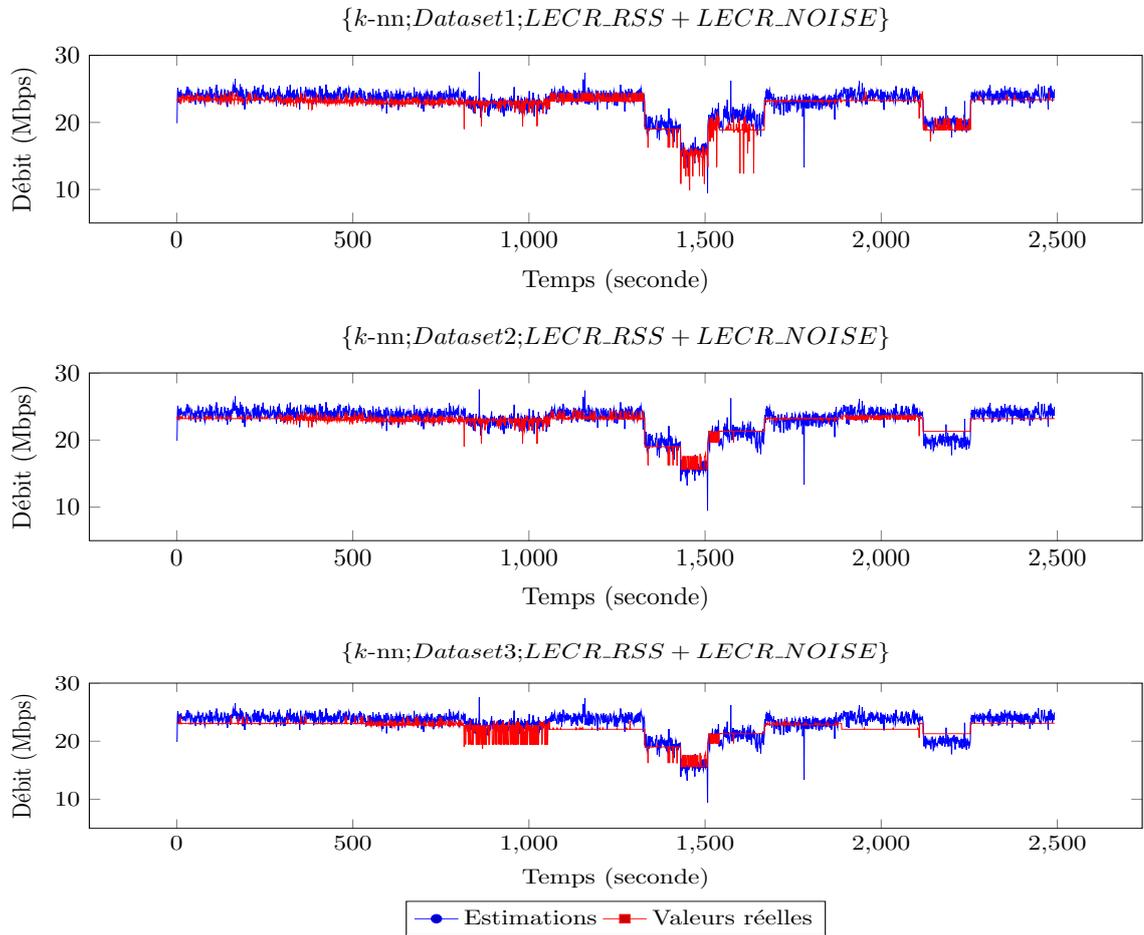


FIGURE A.12 – Graphiques des estimations de k -nn pour la métrique $LECR_RSS + LECR_NOISE$.

A.3 Estimations par DT

A.3.1 Estimation de DT pour la métrique ATH_RSS .

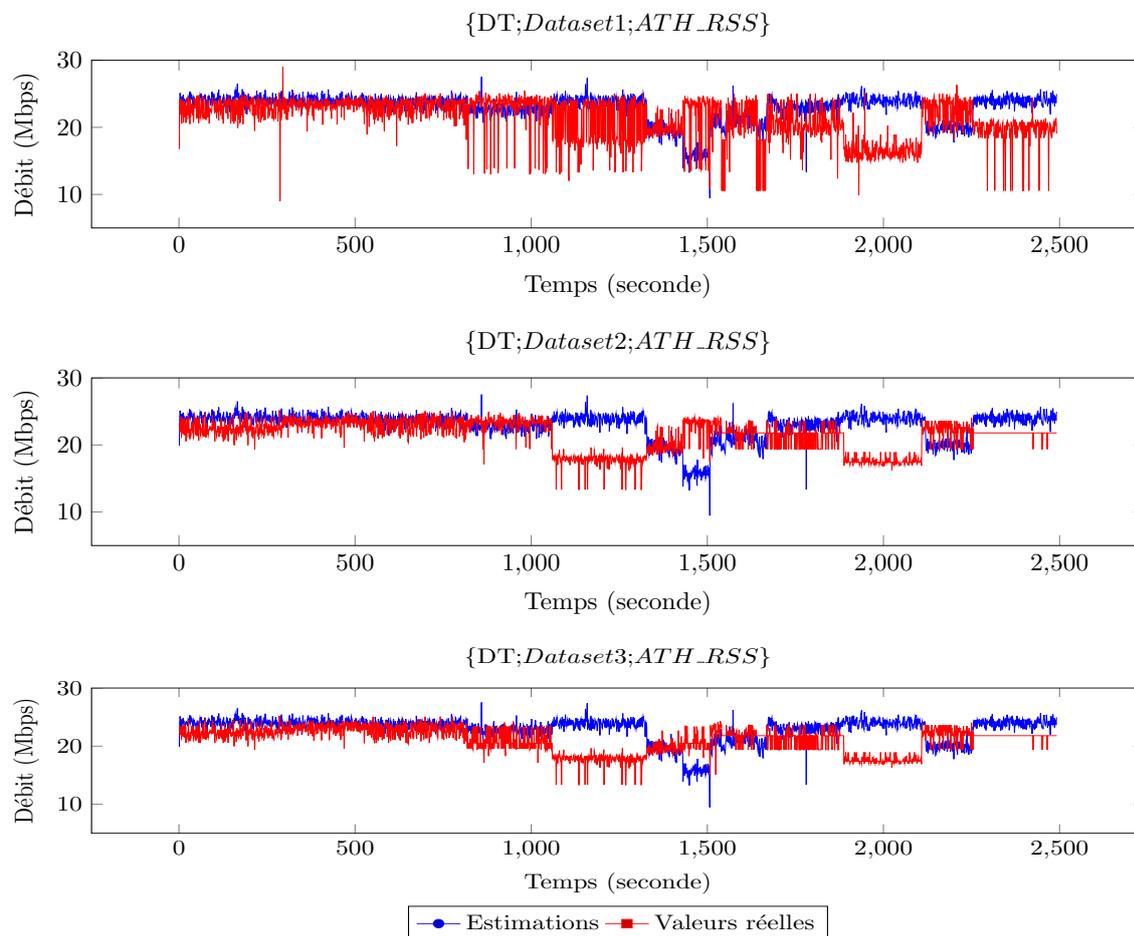


FIGURE A.13 – Graphiques des estimations de DT pour la métrique ATH_RSS .

A.3.2 Estimation de DT pour la métrique $LECR_RSS$.

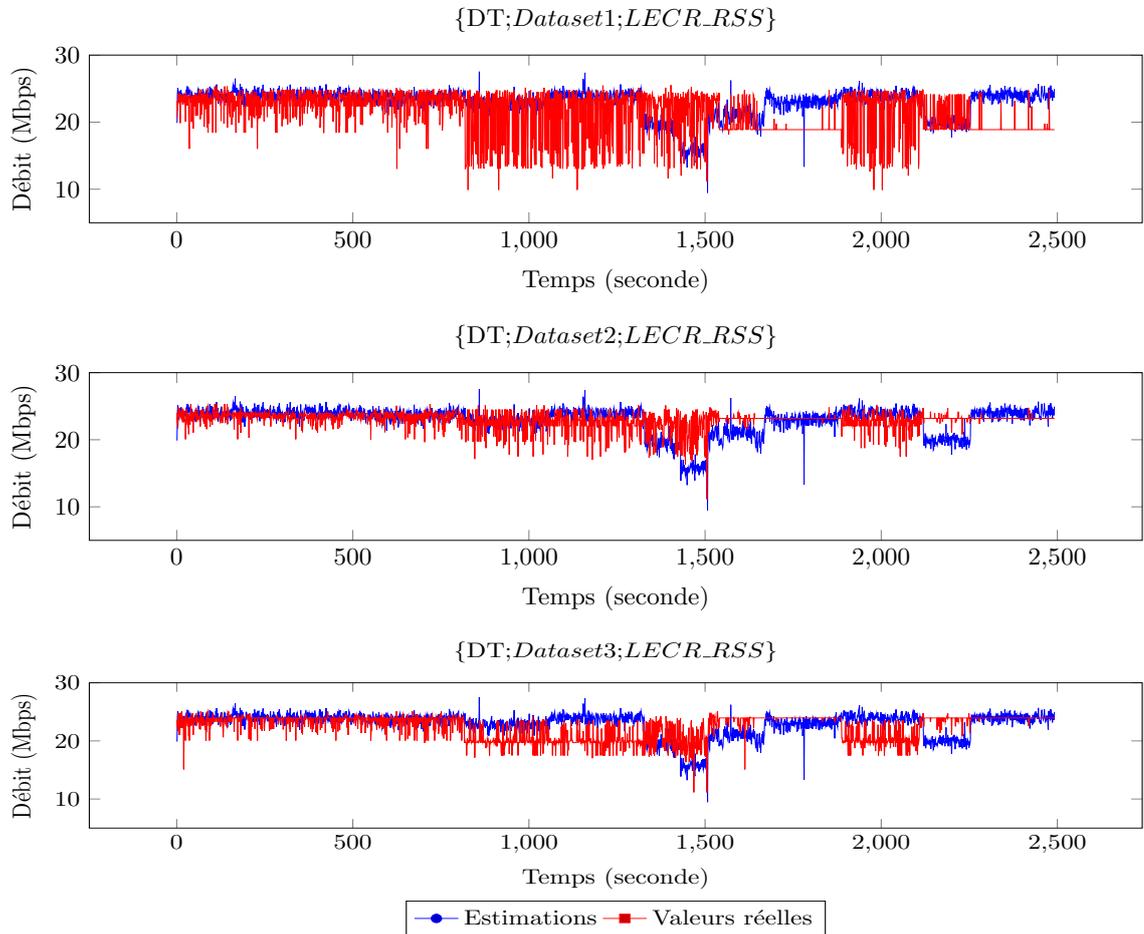


FIGURE A.14 – Graphiques des estimations de DT pour la métrique $LECR_RSS$.

A.3.3 Estimation de DT pour la métrique *LECR_NOISE*.

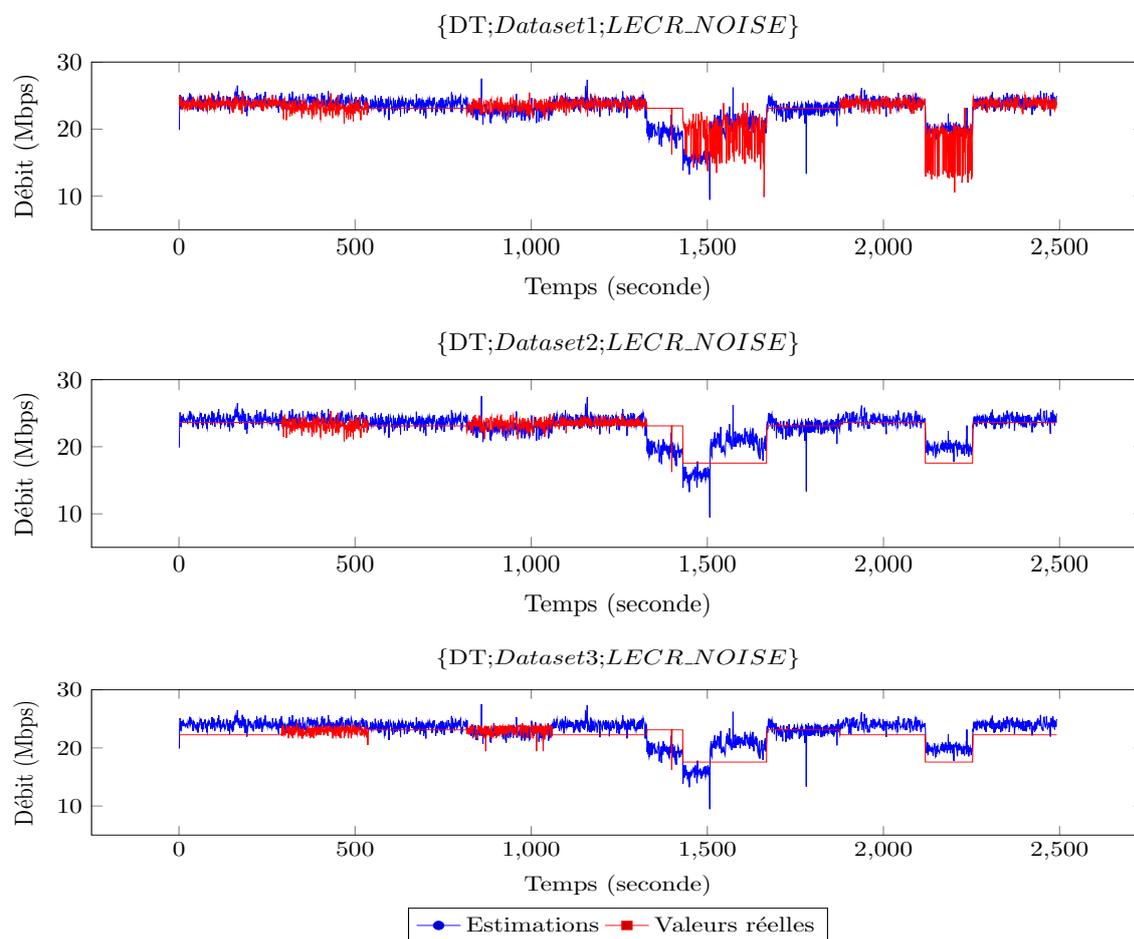


FIGURE A.15 – Graphiques des estimations de DT pour la métrique *LECR_NOISE*.

A.3.4 Estimation de DT pour la métrique $LECR_SNR$.

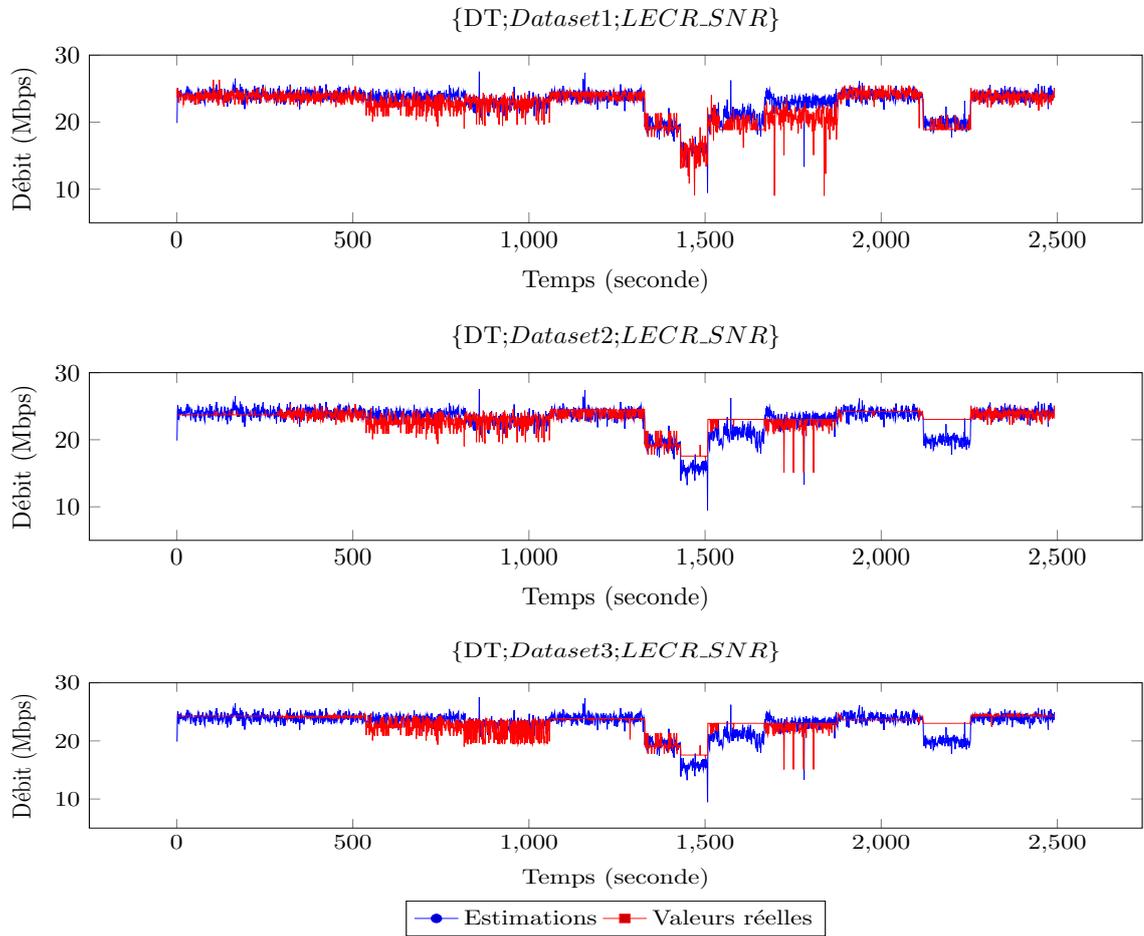


FIGURE A.16 – Graphiques des estimations de DT pour la métrique $LECR_SNR$.

A.3.5 Estimation de DT pour la métrique $ATH_RSS+LECR_NOISE$.

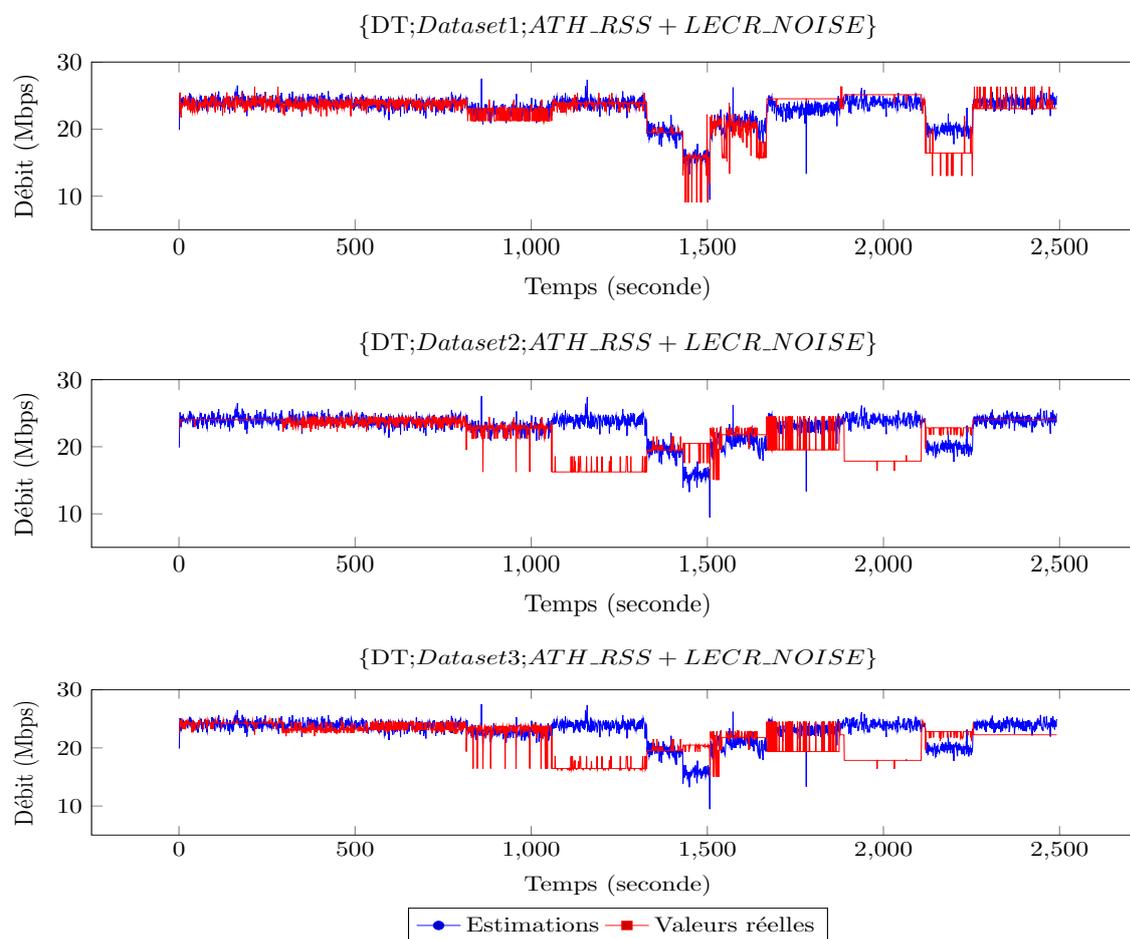


FIGURE A.17 – Graphiques des estimations de DT pour la métrique $ATH_RSS+LECR_NOISE$.

A.3.6 Estimation de DT pour la métrique $LECR_RSS + LECR_NOISE$.

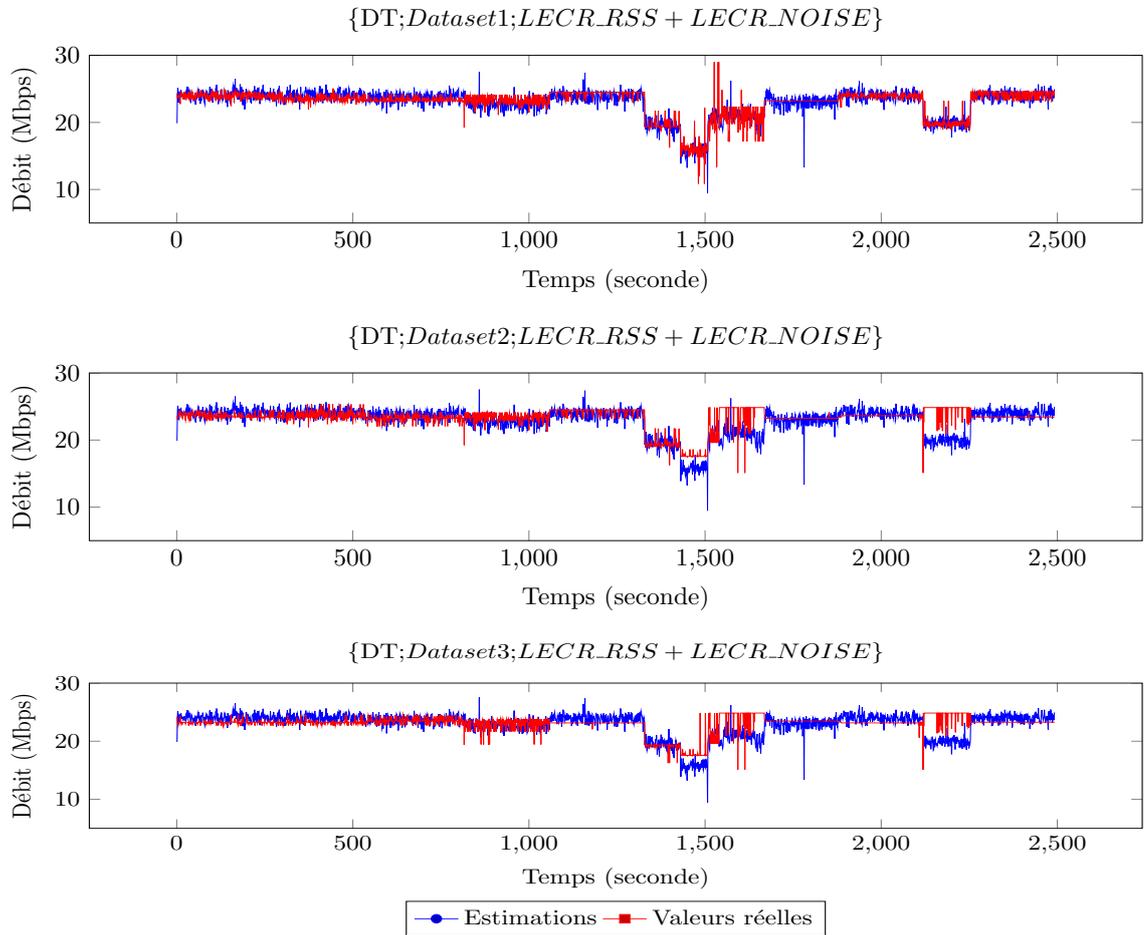


FIGURE A.18 – Graphiques des estimations de DT pour la métrique $LECR_RSS + LECR_NOISE$.

