SCIENCES

RÉSEAUX ET COMMUNICATIONS

Gestion et contrôle des réseaux

Gestion et contrôle intelligents des réseaux

sécurité intelligente, optimisation multicritères, Cloud Computing, Internet of Vehicles, radio intelligente

sous la direction de
Badr Benmammar

ISTE editions

Gestion et contrôle intelligents des réseaux

First published 2020 in Great Britain by ISTE Editions Ltd.

ISTE Editions Ltd
27-37 St George's Road
London SW19 4EU
UK

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-78948-008-5 (print)
ISBN: 978-1-78949-008-4 (e-book)

ERC code:
PE7 Systems and Communication Engineering
PE7_1 Control engineering
PE7_8 Networks (communication networks, sensor networks, networks of robots, etc.)

Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY, March 2020

Encyclopédie SCIENCES

Réseaux et communications, domaine dirigé par Guy Pujolle

Gestion et contrôle des réseaux, thème dirigé par Francine Krief

Gestion et contrôle intelligents des réseaux

sécurité intelligente,
optimisation multicritères,
Cloud Computing, Internet of Vehicles
et radio intelligente

sous la direction de
Badr Benmammar

iSTE editions

Table des matières

Introduction

Badr BENMAMMAR
Université Abou Bekr Belkaid, Tlemcen, Algérie

La gestion et le contrôle d'un réseau informatique sont un domaine qui, auparavant, se concentrait principalement sur des tâches purement techniques de maintenance des équipements constituant le réseau. Cette maintenance vise à assurer son bon fonctionnement et à le faire évoluer.

À l'heure actuelle, avec l'émergence des réseaux informatiques et le développement de toujours plus d'applications susceptibles d'opérer sur un réseau (et plus généralement sur Internet), la gestion et le contrôle d'un réseau informatique ne peuvent plus être envisagés sans l'introduction de l'intelligence artificielle dans l'ensemble de ses étapes.

C'est ce que tente de montrer cet ouvrage d'introduction à la gestion et aux contrôles intelligents des réseaux informatiques. Notre travail vise principalement la présentation de l'utilisation de l'intelligence artificielle dans les réseaux, à travers le contrôle intelligent de ces derniers.

L'objectif principal de l'intelligence artificielle est de concevoir des systèmes capables de reproduire le comportement de l'humain dans ses activités de raisonnement. Toutefois, définir l'intelligence artificielle n'est pas chose simple. Le domaine est si élargi qu'il est impossible de le limiter à un domaine de recherche particulier.

L'intelligence artificielle est définie par l'un de ses créateurs, Marvin Lee Minsky, comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR. © ISTE Editions 2020.

êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ».

Dans les réseaux informatiques, l'application de l'intelligence artificielle peut être liée à plusieurs domaines, comme les communications radio intelligentes, les nouvelles architectures de communication, le *Cloud Computing*, l'optimisation des réseaux et la sécurité.

L'objectif visé par cet ouvrage est de traiter des thèmes d'actualité qui sont liés principalement à la sécurité intelligente des réseaux informatiques, au déploiement de services de sécurité dans les réseaux SDN (*Software-Defined Networking*), à l'optimisation des réseaux à l'aide des techniques de l'intelligence artificielle et aux méthodes d'optimisation multicritères pour la sélection des réseaux dans un environnement hétérogène. L'ouvrage s'intéresse également à la sélection des services *Cloud Computing*, au déchargement intelligent des calculs dans le contexte du *Mobile Cloud Computing*, à la gestion intelligente des ressources dans un système *Smart Grid-Cloud* pour une meilleure efficacité énergétique, à l'*Internet of Vehicles* (IoV), en se basant sur ses nouvelles architectures, à l'application de l'intelligence artificielle dans les réseaux de radio cognitive et enfin à l'apport de la radio intelligente pour répondre aux besoins de communication sur route des véhicules autonomes.

Les différents thèmes traités dans cet ouvrage sont regroupés par parties, contenant deux chapitres chacune. L'idée est de faciliter au lecteur la compréhension de l'apport de l'intelligence artificielle dans chaque domaine particulier.

PARTIE 1

L’IA et la sécurité des réseaux

1

La sécurité intelligente des réseaux informatiques

Abderrazaq SEMMOUD et Badr BENMAMMAR
Université Abou Bekr Belkaid, Tlemcen, Algérie

1.1. Introduction

L'intelligence artificielle (IA) (*Artificial Intelligence* en anglais) et l'apprentissage automatique (*Machine Learning*) ont progressé rapidement ces dernières années, ce qui a permis le développement d'une vaste gamme d'applications. Par exemple, l'intelligence artificielle est un composant essentiel des technologies largement utilisées telles que la reconnaissance automatique de la parole, la traduction automatique, les filtres antispams et la reconnaissance faciale. Parmi les autres technologies prometteuses faisant actuellement l'objet de recherches ou de projets pilotes à petite échelle, citons les voitures sans conducteur, les assistants numériques et les drones activés par l'intelligence artificielle. Encore plus loin dans le futur, l'intelligence artificielle avancée pourrait réduire le besoin de main-d'œuvre non désirée et améliorer la qualité de la gouvernance.

L'intelligence artificielle est utilisée afin d'automatiser un large éventail de tâches. Parmi les tâches habituellement étudiées par les chercheurs de l'intelligence artificielle, nous pouvons évoquer les jeux, la conduite de véhicules et la classification des images. L'ensemble des tâches pouvant être transformées par l'intelligence artificielle est vaste. Au minimum, toute tâche pour laquelle les humains utilisent leur intelligence peut être une cible d'innovation de l'intelligence artificielle. Tandis que le domaine de l'intelligence artificielle remonte aux années 1950, plusieurs années de progrès et de croissance rapides ont récemment conduit à une plus grande

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

fiabilité. Les chercheurs ont réalisé des gains de performance soudains dans un certain nombre de domaines. La figure 1.1 illustre cette tendance dans le cas de la reconnaissance d'images où, au cours des dernières années, les performances des systèmes d'intelligence artificielle sont passées d'une classification correcte d'environ 70 % à une classification presque parfaite (98 %) et supérieure à la référence humaine (95 %) (Brundage *et al*. 2018).

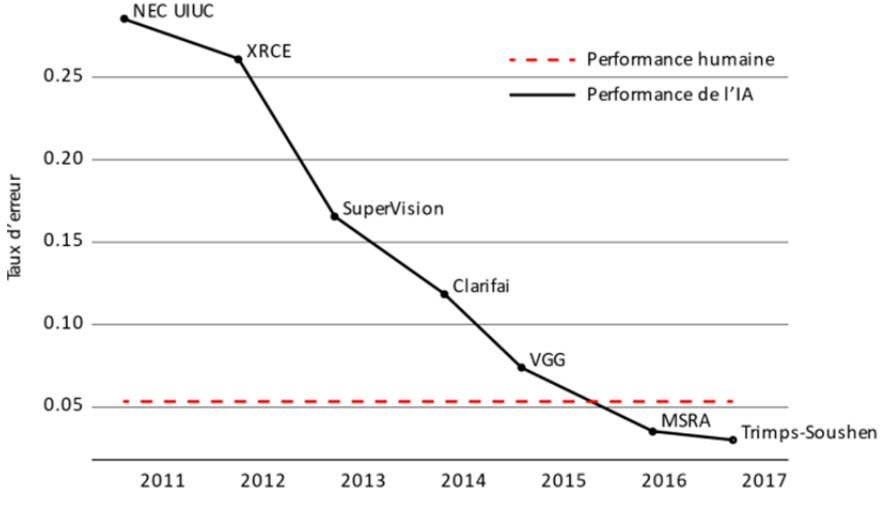


Figure 1.1. *Progrès de la reconnaissance des images (benchmark ImageNet) « Electronic Frontier Foundation's AI Progress Measurement » (août 2017)*

D'un point de vue sécuritaire, un certain nombre des évolutions de l'intelligence artificielle méritent d'être relevées. Par exemple, la capacité de reconnaître le visage d'une cible et de naviguer dans l'espace peut s'appliquer aux systèmes d'armes autonomes. De même, la possibilité de générer des images, du texte et de la voix pourrait être utilisée pour imiter d'autres personnes en ligne ou pour influencer l'opinion publique en diffusant le contenu généré par l'intelligence artificielle *via* les réseaux sociaux. Ces développements techniques peuvent également être considérés comme des indicateurs précoces du potentiel de l'intelligence artificielle. Il ne serait pas étonnant que les systèmes de l'intelligence artificielle deviennent bientôt compétents pour une gamme encore plus large de tâches liées à la sécurité.

La sécurité de l'information est définie comme la protection des systèmes informatiques contre tout accès, utilisation, perturbation, modification ou destruction non autorisés afin d'assurer la confidentialité, l'intégrité et la disponibilité (Peltier 2010). La sécurité de l'information ne définit aucune technologie de sécurité

particulière, mais plutôt une stratégie composée des personnes, des processus, des règles et des outils nécessaires pour détecter, prévenir, documenter et atténuer les menaces auxquelles nous sommes confrontés aujourd'hui. Les réseaux étant de plus en plus interconnectés, il devient de plus en plus important de fournir des services de sécurité. Dans le monde commercial, la connectivité n'est plus une option et les risques éventuels de la connectivité ne l'emportent pas sur ses avantages. Par conséquent, les services de cybersécurité doivent fournir une protection adéquate aux entreprises qui exercent leurs activités dans un environnement relativement ouvert. Par rapport aux approches classiques de la sécurité informatique, plusieurs nouvelles hypothèses doivent être formulées sur les réseaux informatiques actuels :

– les réseaux modernes sont très grands et davantage interconnectés et ils sont plus accessibles ; par conséquent, un attaquant potentiel peut facilement se connecter et accéder à distance à ces réseaux ;

– l'interconnexion de réseaux augmente la probabilité que des attaques soient menées sur les réseaux de grande taille tels qu'Internet, à l'aide d'un ensemble de protocoles largement connus et ouverts.

Les systèmes et les applications informatiques deviennent de plus en plus complexes. Par conséquent, il est devenu plus difficile d'analyser, de sécuriser et de tester correctement la sécurité des systèmes informatiques. Lorsque ces systèmes et leurs applications sont connectés à de grands réseaux, le risque des menaces augmente considérablement. Pour assurer une protection adéquate des réseaux informatiques, les procédures et les technologies que vous déployez doivent garantir (Khidzir *et al*. 2018) :

– la confidentialité : la confidentialité des données garantit que seuls les utilisateurs autorisés peuvent afficher des informations sensibles ;

– l'intégrité : l'intégrité des données garantit que seuls les utilisateurs autorisés peuvent modifier des informations sensibles ; l'intégrité pourrait également garantir l'authenticité des données ;

– la disponibilité : la disponibilité d'un système et des données garantit aux utilisateurs autorisés un accès ininterrompu aux ressources et aux données importantes.

La triade confidentialité, intégrité et disponibilité (*CIA triad*, en anglais) est un concept fondamental de la sécurité de l'information. Ce sont ces trois éléments du système d'information que chaque organisation tente d'assurer. La confidentialité empêche la divulgation non autorisée des informations sensibles (Kumar *et al*. 2018). L'intégrité empêche toute modification non autorisée des informations, garantissant ainsi l'exactitude des informations. Les fonctions de hachage cryptographique

(telles que SHA-1 ou SHA-2) peuvent être utilisées pour assurer l'intégrité des données. La disponibilité est la prévention de la perte d'accès aux ressources et aux informations (Kumar *et al*. 2018).

1.2. L'intelligence artificielle au service de la cybersécurité

Les systèmes de l'intelligence artificielle sont généralement efficaces, car ils peuvent réaliser une tâche donnée plus rapidement et à moindre coût qu'un humain. Les systèmes de l'intelligence artificielle sont également évolutifs car, au fur et à mesure du temps, leur puissance de calcul permet de terminer beaucoup plus de tâche. Par exemple, un système de reconnaissance faciale typique est à la fois efficace et évolutif ; une fois développé, il peut être appliqué à de nombreux flux de caméras pour un coût bien inférieur au coût de l'embauche d'analystes humains pour effectuer un travail équivalent. Pour cela, il est facile de comprendre pourquoi les experts en cybersécurité se penchent sérieusement sur l'intelligence artificielle et en quoi cela pourrait contribuer à atténuer certains de ces problèmes. À titre d'exemple, l'apprentissage automatique utilisé par de nombreux algorithmes de l'intelligence artificielle peut aider à détecter les logiciels malveillants, de plus en plus difficiles à identifier et à isoler, car ces derniers deviennent de plus en plus capables de s'adapter aux solutions de sécurité traditionnelles (Veiga 2018).

L'institut de recherche Capgemini a interrogé 850 dirigeants de sept grandes sociétés industrielles : parmi les membres de la haute direction interrogés, 20 % sont des directeurs des systèmes d'information et 10 % des responsables de la sécurité des systèmes d'information. Les entreprises dont le siège est en France, en Allemagne, au Royaume-Uni, aux États-Unis, en Australie, en Inde et en Italie figurent dans le rapport (Capgemini Research Institute 2019). Capgemini a constaté que, à mesure que les entreprises digitales se développent, leur risque de cyberattaques augmente de manière exponentielle. 21 % ont déclaré que leur organisation avait subi une violation de la cybersécurité menant à un accès non autorisé en 2018. Les entreprises paient un lourd tribut aux violations de la cybersécurité (20 % ont déclaré des pertes supérieures à 50 millions de dollars). Ce sondage a relevé que 69 % des entreprises estiment que l'intelligence artificielle sera nécessaire pour réagir aux cyberattaques. La majorité des entreprises de télécommunications (80 %) déclarent compter sur l'intelligence artificielle pour identifier les menaces et contrecarrer les attaques. Capgemini a constaté que le secteur des télécommunications affichait le plus grand nombre de pertes déclarées dépassant 50 millions de dollars, faisant de l'intelligence artificielle une priorité pour contrecarrer les infractions coûteuses dans ce secteur. Il est compréhensible que les vendeurs de produits pour consommateurs (78 %) et les banques (75 %) se classent respectivement en deuxième et en troisième positions,

compte tenu du fait que chacun de ces secteurs repose de plus en plus sur des modèles numériques. Les entreprises basées aux États-Unis accordent la plus haute priorité aux applications et plateformes de cybersécurité basées sur l'intelligence artificielle.

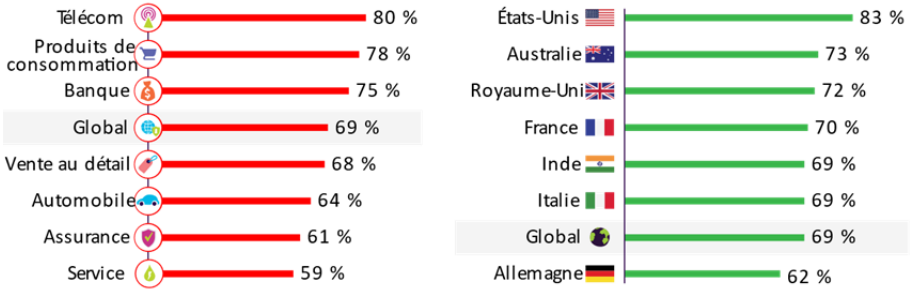


Figure 1.2. *Les organisations et les pays comptant sur l'intelligence artificielle pour identifier les menaces et contrecarrer les attaques*

Chaque jour, de nouvelles vulnérabilités sont découvertes dans les programmes courants, qui pourraient infecter et diriger tout le réseau d'une entreprise. En contraste aux vulnérabilités logicielles traditionnelles (par exemple, les débordements de mémoire tampon), les systèmes intelligents actuels souffrent d'un certain nombre de vulnérabilités. Il s'agit notamment de l'introduction des données provoquant des erreurs dans un système d'apprentissage (Biggio *et al*. 2012), de l'exploitation des défauts de la conception des objectifs des systèmes autonomes (Amodei *et al*. 2016) et de l'utilisation des entrées conçues pour fausser la classification des systèmes d'apprentissage automatique (Szegedy *et al*. 2013). Ces vulnérabilités démontrent que, même si les systèmes intelligents peuvent battre les performances humaines, ils peuvent également échouer de manière inimitable.

Une cyberdéfense idéale offrirait une protection complète aux utilisateurs, tout en préservant les performances du système. Nous sommes très loin de cette situation, mais nous pourrions nous en approcher en rendant la cyberdéfense plus intelligente. L'idée d'utiliser les techniques de l'intelligence artificielle dans la cybersécurité ne date pas d'aujourd'hui. Dans (Landwehr 2008), Carl Landwehr a affirmé que, à ses débuts, la sécurité informatique et l'intelligence artificielle ne semblaient pas avoir grand-chose à se dire. Les chercheurs en intelligence artificielle souhaitaient que les ordinateurs fassent seuls ce que les humains étaient capables de faire, alors que les chercheurs en sécurité cherchaient à réparer les fuites dans les systèmes informatiques qu'ils jugeaient vulnérables. Selon Bruce Schneier (Schneier 2008), « Internet peut être considéré comme la machine la plus complexe jamais

construite par l'homme. Nous comprenons à peine comment cela fonctionne, sans parler de la façon de le sécuriser ». Vu la multiplication rapide des nouvelles applications web et l'utilisation croissante des réseaux sans-fil (Barth et Mitchell 2008) et de l'Internet des objets, la cybersécurité est devenue la menace la plus complexe pour la société.

La sécurité des applications web contre les attaques (telles que les scripts XSS – *Cross Site Scripting* –, le CSRF – *Cross Site Request Forg*ery – et l'injection du code) devient de plus en plus un besoin évident et urgent. Au fil du temps, les scripts XSS et CSRF ont été utilisés pour réaliser diverses attaques. Certaines d'entre elles peuvent être interprétées comme un contournement direct de la même politique de sécurité d'origine. La même politique de sécurité ressemblait à une protection simple et efficace, mais il s'est avéré qu'elle pouvait être contournée assez facilement et bloquer certaines fonctionnalités des sites web modernes. Selon D. Crockford (2015), les politiques de sécurité adoptées par la plupart des navigateurs « empêchent des contenus utiles et autorisent des contenus dangereuses ». Aujourd'hui, ces politiques sont en train d'être revues. Cependant, la détection des attaques telles que XSS, CSRF ou l'injection des codes nécessite plus qu'une simple règle, mais la capacité de raisonner en fonction du contexte.

En général, l'utilisation de l'intelligence artificielle en cybersécurité consiste à utiliser certains outils intelligents et à les appliquer à la détection des intrusions (Ahmad *et al*. 2016 ; Kalaivani *et al*. 2019) ou à d'autres aspects de la cybersécurité (Ahlan *et al*. 2015). Cette approche consiste à utiliser d'autres techniques de l'intelligence artificielle développées pour des problèmes totalement différents à la cybersécurité ; ce qui peut fonctionner dans certains cas, mais cela présente des limitations inhérentes et sévères. La cybersécurité a des besoins spécifiques et, pour y répondre, de nouvelles techniques de l'intelligence artificielle devront être spécifiquement développées. De toute évidence, l'intelligence artificielle a fait beaucoup de progrès dans certains domaines, mais il y a toujours un besoin à apprendre et à développer de nouvelles techniques intelligentes adaptées à la cybersécurité. Dans ce contexte, C. Landwehr (2008) constate qu'une « branche de l'IA qui est liée à la sécurité informatique depuis le plus jeune âge est le raisonnement automatisé, en particulier lorsqu'il est appliqué aux programmes et aux systèmes. Bien que le programme SATAN de Dan Farmer et Wietse Venema, lancé en 1995, n'ait pas encore été identifié comme étant une IA, il a automatisé un processus de recherche de vulnérabilités dans des configurations de système qui nécessitaient beaucoup plus d'efforts humains ». S. Forrest (Ingham *et al*. 2007) a proposé un système de raisonnement inductif pour la protection des applications web. Les travaux de Vigna *et al.* (Mutz *et al*. 2007 ; Cova *et al*. 2007 ; Kirdaa *et al*. 2009 ; Cova *et al*. 2010) et de (Robertson *et al*. 2010) ont traité aussi de la protection

des applications web contre les cyberattaques. Les pare-feu utilisant des inspections de paquets en profondeur peuvent être considérés comme une sorte d'instanciation de l'intelligence artificielle dans la cybersécurité. Les pare-feu font partie de l'arsenal de la cyberdéfense depuis de nombreuses années. Bien que des techniques plus sophistiquées (Mishra *et al*. 2011 ; Valentín et Malý 2014 ; Tekerek et Bay 2019) soient également utilisées dans la plupart des cas, le filtrage est basé sur les numéros de port. Les pare-feu ne peuvent pas compter sur le numéro de port, car la plupart des applications web utilisent le même port que le reste du trafic web. L'inspection approfondie des paquets est la seule option permettant de distinguer un code malveillant dans une application légitime. L'idée de filtrer au niveau de la couche Application du modèle TCP/IP a déjà été introduite dans la troisième génération de pare-feu dans les années 1990. Le succès modeste de ces technologies reflète la nécessité de travailler beaucoup plus sur l'intelligence artificielle, afin qu'elle puisse faire une différence significative en matière de cybersécurité. Cependant, il est important de noter que l'utilisation de l'intelligence artificielle en cybersécurité n'est pas nécessairement une solution miracle. Par exemple, les attaques sans logiciel malveillant, qui ne nécessitent aucun téléchargement de logiciel et dissimulent les activités malveillantes au sein de services du *Cloud Computing* légitimes, sont en augmentation, et l'intelligence artificielle n'est pas encore capable de contrecarrer ces types de violations du réseau.

1.3. L'intelligence artificielle appliquée à la détection d'intrusion

La détection d'intrusion est définie comme le processus de surveillance intelligente des événements se produisant dans un système informatique ou un réseau et de leur analyse pour rechercher des signes de violation de la politique de sécurité (Bace 2000). Le principal objectif des systèmes de détection d'intrusion (*Intrusion Detection Systems*) est de protéger la disponibilité, la confidentialité et l'intégrité des réseaux. Les systèmes de détection d'intrusion sont définis à la fois par la méthode utilisée pour détecter les attaques et par leur emplacement sur le réseau. Le système de détection d'intrusion peut être déployé en tant que système basé sur le réseau ou sur l'hôte, afin de détecter les anomalies. La détection des utilisations abusives repose sur la correspondance des modèles des activités hostiles connus avec les bases de données d'attaques antérieures. Ces modèles sont très efficaces pour identifier les attaques et les vulnérabilités connues, mais moins pertinents pour identifier les nouvelles menaces de sécurité. La détection d'anomalies cherchera quelque chose de rare ou d'inhabituel, en appliquant des mesures statistiques ou intelligentes pour comparer l'activité actuelle aux connaissances préalables. Les systèmes de détection d'intrusion reposent sur le fait qu'ils ont souvent besoin de nombreuses données pour les algorithmes d'apprentissage artificiel. Ils nécessitent

généralement davantage de ressources informatiques, car plusieurs métriques sont souvent conservées et doivent être mises à jour pour chaque activité du système (Ahmad *et al*. 2016). Le système expert de détection d'intrusion (IDES) (Lunt 1993) de SRI formule les connaissances d'un expert sur les modèles d'attaque connus et les vulnérabilités du système sous forme de règles *if-then*. La machine inductive temporelle (Teng et Chen 1990) apprend des modèles séquentiels, afin d'assurer la détection d'anomalies dans un réseau. Plusieurs approches utilisant les réseaux de neurones artificiels pour les systèmes de détection d'intrusion ont été proposées (Kang et Kang 2016 ; Kim *et al*. 2016 ; Vinayakumar *et al*. 2017 ; Hajimirzaei et Navimipour 2019). Les techniques basées sur l'intelligence artificielle sont catégorisées en différentes classes (Mukkamala et Sung 2003a ; Novikov *et al*. 2006).

1.3.1. *Les techniques basées sur les arbres de décision*

Les arbres de décision (*Decision Tree*) sont des outils puissants et populaires d'apprentissage non paramétrique utilisés pour des problèmes de classification et de prédiction. Le but est de créer un modèle qui prédit les valeurs de la variable cible, en se basant sur un ensemble de séquences de règles de décision déduites à partir des données d'apprentissage. Rai a développé un algorithme à base de l'approche de l'arbre de décision C4.5 (Rai *et al*. 2016). Les caractéristiques les plus pertinentes sont sélectionnées à l'aide du gain d'information et la valeur fractionnée est sélectionnée de manière à rendre le classificateur non biaisé par rapport aux valeurs les plus fréquentes. Dans le travail de Sahu et Babu (2015), une base de données appelée « Kyoto 2006+ » est utilisée pour les expérimentations. Dans Kyoto 2006+, chaque instance est étiquetée comme « normal » (pas d'attaque), « attaque » (attaque connue) et « attaque inconnue ». L'algorithme Decision Tree (J48) est utilisé pour classifier les paquets. Les expérimentations confirment que les règles générées fonctionnent avec une exactitude de 97,2 %. Moon *et al.* (2017) ont proposé un système de détection d'intrusion basé sur les arbres de décision utilisant l'analyse du comportement des paquets pour détecter les attaques. Peng *et al.* (2018) ont proposé une technique qui consiste à effectuer un prétraitement pour numériser les données, puis pour les normaliser, afin d'améliorer l'efficacité de la détection. Ensuite, une méthode basée sur les arbres de décision est utilisée.

1.3.2. *Les techniques basées sur l'exploration de données*

L'exploration de données vise à éliminer les éléments manuels utilisés pour la conception des systèmes de détection d'intrusion. Diverses techniques d'exploration

de données ont été développées et largement utilisées. Les principales techniques d'exploration de données sont abordées dans les sections suivantes.

1.3.2.1. *La logique floue*

La logique floue (*Fuzzy Logic*) a été utilisée dans le domaine de la sécurité des réseaux informatiques, en particulier dans le domaine de la détection d'intrusion (Idris et Shanmugam 2005 ; Shanmugavadivu et Nagarajan 2011 ; Balan *et al*. 2015 ; Kudłacik *et al*. 2016 ; Sai Satyanarayana Reddy *et al*. 2019), pour deux raisons principales. Tout d'abord, plusieurs paramètres quantitatifs utilisés dans le contexte de la détection d'intrusion, par exemple le temps d'utilisation du processeur, l'intervalle de connexion, etc., peuvent potentiellement être considérés comme des variables floues. Deuxièmement, le concept de sécurité lui-même est flou. En d'autres termes, le concept de flou aide à éviter la séparation abrupte des comportements normal et anormal. Kudłacik *et al.* (2016) ont appliqué la logique floue pour la détection des intrusions. La solution proposée analyse l'activité de l'utilisateur sur une période relativement courte, en créant un profil local de l'utilisateur. Une analyse plus poussée implique la création d'une structure plus générale basée sur un nombre défini de profils locaux d'un utilisateur, appelé « profil flou ». Le profil flou représente le comportement de l'utilisateur du système informatique. Les profils flous sont utilisés directement pour détecter les anomalies dans le comportement des utilisateurs, et donc les éventuelles intrusions. Idris et Shanmugam (2005) ont proposé le système FIRE modifié. Il s'agit d'un mécanisme pour automatiser le processus de génération de règles floues et réduire l'intervention humaine en utilisant les techniques de l'intelligence artificielle.

1.3.2.2. *Les algorithmes génétiques*

Les algorithmes génétiques (*Genetic Algorithm*) sont des techniques dérivées de la génétique et de l'évolution naturelle, utilisés pour trouver des solutions approximatives à des problèmes d'optimisation et de recherche. Les principaux avantages des algorithmes génétiques sont la flexibilité et la robustesse en tant que méthode de recherche globale. Leur inconvénient est qu'ils sont coûteux en temps de calcul, puisqu'ils manipulent plusieurs solutions simultanément. Dans le domaine de la détection d'intrusion, les algorithmes génétiques ont été utilisés de différentes manières (Hoque *et al*. 2012 ; Aslahi-Shahri *et al*. 2016 ; Hamamoto *et al*. 2018). Hoque *et al.* (2012) ont présenté un système de détection d'intrusion utilisant un algorithme génétique pour détecter efficacement les anomalies sur le réseau. Aslahi-Shahri *et al.* (20016) ont proposé une méthode hybride qui utilise les machines à vecteurs de support et les algorithmes génétiques pour le problème de détection d'intrusion. Les résultats révèlent que cet algorithme est capable d'atteindre un taux de 97,3 % pour les vrais positifs et de 1,7 % pour les faux positifs.

1.3.3. *Les techniques basées sur les règles*

Les techniques basées sur les règles (Li *et al*. 2010 ; Yang *et al*. 2013) impliquent généralement l'application d'un ensemble de règles d'association pour classifier les données. Dans ce contexte, si une règle stipule que *si l'événement X se produit, alors l'événement Y est susceptible de se produire*, les événements X et Y peuvent être décrits comme des ensembles de paires (*variable, valeur*). L'avantage d'utiliser des règles est qu'elles ont tendance à être simples et intuitives, non structurées et moins rigides. Cependant, l'inconvénient est que les règles sont difficiles à maintenir et, dans certains cas, inadéquates pour représenter différents types d'informations.

Turner *et al.* (2016) ont développé un algorithme pour surveiller l'état activé/désactivé des règles d'un système de détection d'intrusion basé sur les signatures. L'algorithme est implémenté en Python et est exécuté sur Snort (Roesch 1999). Agarwal et Joshi (2000) ont proposé un cadre général en deux étapes pour l'apprentissage d'un modèle à base de règles (PNrule) afin d'apprendre des modèles de classifieur sur un ensemble de données. Ils ont largement utilisé différentes distributions de classes dans les données d'apprentissage. La base de données KDD Cups a été utilisée pour l'apprentissage et le test de leur système.

1.3.4. *Les techniques basées sur l'apprentissage automatique*

L'apprentissage automatique peut être défini comme la capacité d'un programme à apprendre et à améliorer les performances d'une série de tâches au fil du temps. Les techniques d'apprentissage automatique se concentrent sur la création d'un modèle de système qui améliore ses performances en fonction des résultats précédents. Ou encore, on peut dire que les systèmes basés sur l'apprentissage automatique ont la capacité de manipuler la stratégie d'exécution en fonction de nouvelles entrées. Les principales techniques d'apprentissage automatique sont développées dans les sections suivantes.

1.3.4.1. *Les réseaux de neurones artificiels*

Les réseaux de neurones artificiels (*Artificial Neural Networks*) apprennent à prédire le comportement des différents utilisateurs du système. S'ils sont correctement conçus et mis en œuvre, les réseaux de neurones peuvent potentiellement résoudre plusieurs problèmes rencontrés par les approches basées sur les règles. Le principal avantage des réseaux de neurones est leur tolérance aux données imprécises et aux informations incertaines et leur capacité à déduire des solutions sans avoir une connaissance préalable des régularités des données. Cunningham et

Lippmann (2000) du MIT Lincoln Laboratory ont effectué un certain nombre de tests en utilisant des réseaux de neurones (Lippmann et Cunningham 2000). Le système recherchait des mots-clés spécifiques aux attaques dans le trafic réseau. Dans (Ponkarthika et Saraswathy 2018), un modèle de système de détection d'intrusion est exploré en fonction d'un apprentissage profond. L'architecture LSTM (*Long Short Term Memory*) est appliquée à un réseau de neurones récurrents (*Recurrent Neural Network*) pour l'apprentissage du système de détection d'intrusion à l'aide du jeu de données KDD Cup 1999.

1.3.4.2. *Les réseaux bayésiens*

Un réseau bayésien (*Bayesian Network*) est un modèle graphique probabiliste représentant un ensemble de variables aléatoires sous la forme d'un graphe orienté acyclique. Cette technique est généralement utilisée pour la détection d'intrusion en combinaison avec des schémas statistiques. Elle présente plusieurs avantages, notamment la capacité de coder les interdépendances entre les variables et de prédire les événements, ainsi que la possibilité d'intégrer à la fois les connaissances et les données antérieures (Heckerman 2008). L'inconvénient majeur est que les résultats sont comparables aux techniques statistiques, mais que cela nécessite des efforts de calcul supplémentaires. Kruegel *et al.* (2003) ont proposé une approche de fusion multisensorielle utilisant un classifieur à base de réseaux bayésiens pour la classification et la suppression des fausses alarmes, selon laquelle les sorties de différents capteurs du système de détection d'intrusion ont été agrégées pour produire une seule alarme. Han *et al.* (2015) ont proposé un algorithme de détection d'intrusion basé sur les réseaux bayésiens reposant sur l'analyse en composantes principales. Les auteurs calculent la valeur caractéristique des données d'attaque du réseau d'origine, puis extraient les propriétés principales par l'analyse en composantes principales.

1.3.4.3. *Les chaînes de Markov*

Une chaîne de Markov (*Markov Chain*) est un processus aléatoire portant sur un nombre fini d'états, avec des probabilités de transition sans mémoire. Pendant la phase d'apprentissage, les probabilités associées aux transitions sont estimées à partir du comportement normal du système cible. La détection des anomalies est ensuite réalisée en comparant le score de l'anomalie obtenu pour les séquences observées à un seuil fixé. Dans le cas d'un modèle de Markov caché (Hu *et al*. 2009 ; Zegeye *et al*. 2018 ; Liang *et al*. 2019), le système qui nous intéresse est supposé être un processus de Markov dans lequel des états et des transitions sont masqués. Dans la littérature, plusieurs méthodes ont été présentées pour résoudre le problème de la détection d'intrusion en inspectant les en-têtes des paquets. Mahoney et Chan (2001) ont expérimenté la détection d'anomalies sur les données du réseau

DARPA en comparant les champs d'en-tête de paquet réseau. Plusieurs systèmes utilisent le modèle de Markov pour la détection d'intrusion : PHAD (*Packet Header Anomaly Detector*) (Mahoney et Chan 2001), LERAD (*Learning Rules for Anomaly Detection*) (Mahoney et Chan 2002a) et ALAD (*Application Layer Anomaly Detector*) (Mahoney et Chan 2002b). Dans le travail de Zegeye *et al.* (2018), un système de détection d'intrusion à l'aide du modèle de Markov caché (*Hidden Markov Model*) est proposé. La phase d'analyse du trafic réseau comprend des techniques d'extraction de caractéristiques, de réduction de dimensions et de quantification vectorielle qui jouent un rôle important dans les grands ensembles de données, car le nombre de données transmises augmente chaque jour. Les performances du modèle par rapport au jeu de données KDD 99 démontrent une précision supérieure à 99 %.

1.3.4.4. *Les machines à vecteurs de support*

La machine à vecteurs de support (*Support-Vector Machine*) est une technique utilisée pour résoudre divers problèmes d'apprentissage, de classification et de prédiction. La machine à vecteurs de support est à l'origine d'une implémentation du principe de minimisation du risque structurel (SRM) de Vapnik, qui minimise l'erreur de généralisation, c'est-à-dire l'erreur vraie sur des exemples non vus (Vapnik 1998). La machine à vecteurs de support de base traite des problèmes à deux classes, dans lesquels les données sont séparées par un hyperplan défini par un certain nombre de vecteurs de support. Les vecteurs de support sont un sous-ensemble de données d'apprentissage servant à définir la limite entre les deux classes. Dans les situations où la machine à vecteurs de support ne peut pas séparer deux classes, il résout ce problème en mappant les données d'entrée dans des espaces de fonctions de grande dimension à l'aide d'une fonction du noyau. Dans un espace de grande dimension, il est possible de créer un hyperplan permettant une séparation linéaire (ce qui correspond à une surface incurvée dans l'espace d'entrée inférieur). En conséquence, la fonction du noyau joue un rôle important dans la machine à vecteurs de support. En pratique, diverses fonctions du noyau peuvent être utilisées, telles que linéaire, polynomiale ou gaussienne. Une propriété remarquable de la machine à vecteurs de support est sa capacité d'apprentissage indépendante de la dimensionnalité de l'espace caractéristique. Cela signifie que la machine à vecteurs de support peut bien généraliser en présence de nombreuses fonctionnalités. Mukkamala et Sung (2003b) ont démontré de nombreux avantages de la machine à vecteurs de support par rapport aux autres techniques. Les machines à vecteurs de support dépassent les réseaux de neurones en ce qui concerne l'évolutivité, le temps d'apprentissage, le temps d'exécution et la précision des prévisions. Mukkamala et Sung (2003a) ont également appliqué les machines à vecteurs de support à l'extraction des caractéristiques pour la détection d'intrusion de fichiers KDD. Ils ont prouvé empiriquement que les

fonctionnalités sélectionnées à l'aide de la machine à vecteurs de support conduisent à des résultats similaires à ceux de l'utilisation de l'ensemble complet des fonctionnalités. Cette réduction du nombre de fonctionnalités améliore les efforts de calcul. Chen *et al.* (2005) ont également prouvé que les machines à vecteurs de support étaient meilleures que les réseaux de neurones.

1.3.5. *Les techniques basées sur le* clustering

Les techniques de *clustering* fonctionnent en regroupant les données observées en groupes, en fonction d'une similarité donnée ou d'une mesure de distance. La similarité peut être mesurée en utilisant la formule cosinus, la formule cosinus pondérée binaire proposée par Rawat (2005) ou encore d'autres formules. La procédure la plus couramment utilisée pour le *clustering* implique la sélection d'un point représentatif pour chaque *cluster*. Ensuite, chaque nouveau point de données est classé comme appartenant à un groupe donné en fonction de la proximité au point représentatif correspondant. Il existe au moins deux approches pour la détection des anomalies basée sur la classification. Dans la première approche, le modèle de détection d'anomalie est formé à l'aide de données non étiquetées, comprenant à la fois le trafic normal et le trafic d'attaque. Dans la seconde approche, le modèle est formé en utilisant uniquement des données normales et un profil d'activité normale est créé. L'idée sous-jacente à la première approche est que les données anormales ou d'attaque constituent un faible pourcentage du total des données. Si cette hypothèse est vérifiée, des anomalies et des attaques peuvent être détectées en fonction de la taille des grappes : les grappes volumineuses correspondent à des données normales et les autres points de données à des attaques. Liao et Vemuri (2002) ont utilisé l'approche k-NN (*k-Nearest Neighbor*), basée sur la distance euclidienne, pour définir l'appartenance des points de données à un *cluster* donné. Le système de détection d'intrusion de Minnesota est une approche de détection d'anomalie basée sur le réseau qui utilise des techniques d'exploration de données et de mise en *cluster* (Levent *et al*. 2004). Leung et Leckie (2005) ont proposé une approche de détection d'anomalie non supervisée pour la détection d'intrusion sur un réseau. L'algorithme proposé, appelé « fpMAFIA », est un algorithme de *clustering* basé sur la densité et sur la grille pour les grands ensembles de données. L'avantage majeur de cet algorithme est qu'il peut produire des de formes quelconques et couvrir plus de 95 % de l'ensemble des données avec des valeurs de paramètres appropriées. Les auteurs ont prouvé que l'algorithme évolue linéairement par rapport au nombre d'enregistrements dans l'ensemble des données. Ils ont évalué la précision du nouvel algorithme proposé et ont montré qu'il permet d'atteindre un taux de détection raisonnable.

1.3.6. *Les techniques hybrides*

De nombreux chercheurs ont suggéré que la capacité de surveillance des systèmes IDS actuels pouvait être améliorée en adoptant une approche hybride comprenant à la fois des techniques de détection des anomalies et des signatures (Lunt *et al*. 1992 ; Anderson *et al*. 1995 ; Fortuna *et al*. 2002 ; Hwang *et al*. 2007). Sabhnani et Serpen (2003) ont prouvé qu'aucune technique de classification unique ne permettait de détecter toutes les classes d'attaques à un taux de fausse alarme acceptable et à une bonne précision de détection. Les auteurs ont utilisé différentes techniques pour classifier les intrusions à l'aide du jeu de données KDD 1998. De nombreux chercheurs ont prouvé que la technique de classification hybride ou par ensemble peut améliorer la précision de la détection (Mukkamala *et al*. 2005 ; Chen *et al*. 2005 ; Aslahi-Shahri *et al*. 2016 ; Hamamoto *et al*. 2018 ; Hajimirzaei et Navimipour 2019 ; Sai Satyanarayana Reddy *et al*. 2019). Une approche hybride implique l'intégration de différents modèles d'apprentissage ou de prise de décision. Chaque modèle d'apprentissage fonctionne de manière différente et exploite un ensemble de fonctionnalités différent. L'intégration de différents modèles d'apprentissage donne de meilleurs résultats que les modèles individuels d'apprentissage ou de prise de décision et réduit leurs limitations individuelles. Un avantage significatif de la combinaison de techniques de classification redondantes et complémentaires consiste à augmenter la robustesse et la précision dans la plupart des applications.

Différentes méthodes combinant différentes techniques de classification ont été proposées dans la littérature (Menahem *et al*. 2009 ; Witten *et al*. 2016). L'objectif commun des méthodes d'ensemble est de construire une combinaison de certains modèles, au lieu d'utiliser un seul modèle pour améliorer les résultats. Mukkamala et ses collaborateurs (2005) ont prouvé qu'en utilisant des classificateurs d'ensemble on pouvait obtenir la meilleure précision possible pour chaque catégorie de modèles d'attaque. Chebrolu *et al.* (2005) ont utilisé l'approche CART-BN pour la détection d'intrusion. Zainal *et al.* (2009) ont proposé l'hybridation de la programmation génétique linéaire (*Linear Genetic Programming*), du système d'inférence neuro-floue (*Adaptive Neural Fuzzy Inference System*) et des forêts d'arbres décisionnels (*Random Forest*) pour la détection d'intrusion. Ils ont prouvé empiriquement qu'en attribuant des poids appropriés aux classificateurs dans une approche hybride la précision de détection de toutes les classes de trafic réseau est améliorée par rapport à un classificateur individuel. Menahem *et al.* (2009) ont utilisé différents classificateurs et ont essayé d'exploiter leurs points forts. Hwang *et al*. (2007) ont proposé une approche hybride à trois niveaux pour détecter les intrusions. Le premier niveau du système est une approche basée sur la signature pour filtrer les attaques connues en utilisant le concept de la liste noire. Le deuxième niveau du

système est un détecteur d'anomalies qui utilise le concept de liste blanche pour distinguer le trafic normal et le trafic d'attaque dépassé par le premier niveau. Le troisième niveau du système utilise les machines à vecteurs de support pour classifier le trafic d'attaque inconnu. Le succès d'une méthode hybride dépend de nombreux facteurs, notamment la taille de l'échantillon d'apprentissage, le choix d'un classifieur de base, la manière exacte dont l'ensemble de formation est modifié, le choix de la méthode de combinaison et enfin la distribution des données et la capacité potentielle du classifieur de base choisie pour résoudre le problème (Rokach 2010).

1.4. L'utilisation malveillante de l'intelligence artificielle

L'intelligence artificielle est un domaine technologique à double usage. Les systèmes d'intelligence artificielle et la manière de les concevoir peuvent être utilisés à des fins tant civiles que militaires, et plus largement à des fins bénéfiques ou néfastes. Étant donné que certaines tâches requérant une intelligence sont bénignes et que d'autres ne le sont pas, l'intelligence artificielle est à double tranchant, au même titre que l'intelligence humaine. Les chercheurs en intelligence artificielle ne pourront pas éviter de produire des systèmes pouvant servir à des fins préjudiciables. Par exemple, la différence entre les capacités d'un drone autonome utilisé pour livrer des packages et les capacités d'un drone autonome utilisé pour livrer des explosifs n'est pas forcément très grande. En outre, les recherches fondamentales visant à améliorer notre compréhension de l'intelligence artificielle, de ses capacités et de son contrôle semblent être par nature à double usage.

L'intelligence artificielle et l'apprentissage automatique ont un impact de plus en plus important sur la sécurité des citoyens, des organisations et des États. L'utilisation malveillante de l'intelligence artificielle aura une incidence sur la manière dont nous construisons et gérons notre infrastructure numérique, ainsi que sur la conception et la distribution de systèmes d'intelligence artificielle, de ce fait elle nécessitera probablement une politique institutionnelle. À noter ici, que les menaces causées par l'utilisation malveillante de l'intelligence artificielle ont été mises en évidence dans des contextes très médiatisés (comme lors d'une audience du Congrès (Moore et Anderson 2012), d'un atelier organisé par la Maison Blanche et d'un rapport du Département de la sécurité intérieure des États-Unis).

L'utilisation croissante de l'intelligence artificielle pour le développement des techniques de cyberattaques et l'absence de développement de défenses adéquates conduit à trois conséquences majeures.

1.4.1. *Élargissement des menaces existantes*

Pour de nombreuses attaques connues, nous nous attendons à ce que les progrès de l'intelligence artificielle élargissent l'ensemble des acteurs capables de mener l'attaque, la vitesse à laquelle ces acteurs peuvent la mener et l'ensemble des cibles possibles. Cette affirmation découle de l'efficacité, de l'évolutivité et de la facilité de diffusion des systèmes de l'intelligence artificielle. En particulier, la diffusion des systèmes intelligents et efficaces peut augmenter le nombre d'acteurs pouvant se permettre de mener des attaques particulières. Si les systèmes intelligents pertinents sont également évolutifs, alors même les acteurs qui possèdent déjà les ressources nécessaires pour mener ces attaques peuvent acquérir la capacité de les exécuter à un rythme beaucoup plus rapide.

Un exemple de menace susceptible de se développer de cette manière est la menace d'attaques par hameçonnage. Ces attaques utilisent des messages personnalisés pour extraire des informations sensibles ou de l'argent à leurs victimes. L'attaquant se présente souvent comme l'un des amis, des collègues ou des contacts professionnels de la cible. Les attaques de *phishing* les plus avancées nécessitent une main-d'œuvre qualifiée importante, car l'attaquant doit identifier les cibles de grande valeur, rechercher les réseaux sociaux et professionnels de ces cibles, puis générer des messages acceptables par la cible.

1.4.2. *Introduction de nouvelles menaces*

Les progrès de l'intelligence artificielle permettront de nouvelles variétés d'attaques. Ces attaques peuvent utiliser des systèmes de l'intelligence artificielle pour effectuer certaines tâches avec plus de succès que tout être humain.

Le fait d'être illimités par rapport aux capacités humaines implique que les systèmes intelligents pourraient permettre aux acteurs de mener des attaques qui seraient autrement impossibles. Par exemple, la plupart des personnes ne sont pas capables d'imiter la voix des autres de manière efficace. Par conséquent, la création de fichiers audio ressemblant à des enregistrements de la parole humaine devient primordiale dans ce cas. Cependant, des progrès importants ont récemment été réalisés dans le développement de systèmes de synthèse de la parole, qui apprennent à imiter la voix des individus. De tels systèmes ouvriraient à leur tour de nouvelles méthodes pour répandre la désinformation et imiter les autres.

En outre, les systèmes d'intelligence artificielle pourraient également être utilisés pour contrôler certains aspects du comportement des logiciels malveillants qu'il serait impossible de contrôler manuellement. Par exemple, un virus conçu pour

modifier le comportement d'ordinateurs ventilés, comme dans le cas du programme Stuxnet utilisé pour perturber le programme nucléaire iranien, ne peut pas recevoir des commandes une fois que ces ordinateurs sont infectés. Des problèmes de communication limités apparaissent également sous l'eau et en présence de brouilleurs de signaux.

1.4.3. *Modification du caractère typique des menaces*

L'efficacité, l'évolutivité et le dépassement des capacités humaines sont des propriétés de l'intelligence artificielle menant à des attaques très pertinentes. Les attaquants sont souvent confrontés à un compromis entre la fréquence, l'ampleur de leurs attaques et leur efficacité. Par exemple, le *spear phishing* est plus efficace que le *phishing* classique, qui ne consiste pas à adapter les messages aux individus, mais il est relativement coûteux et ne peut être effectué en masse. Les attaques de *phishing* plus génériques parviennent à être rentables malgré des taux de réussite très bas, simplement en raison de leur ampleur. En améliorant la fréquence et l'évolutivité de certaines attaques, y compris le *spear phishing*, les systèmes d'intelligence artificielle peuvent atténuer ces compromis. De plus, les propriétés d'efficacité et d'évolutivité, en particulier dans le contexte de l'identification et de l'analyse de cibles, conduisent également à des attaques finement ciblées. Les attaquants ont souvent intérêt à adapter leurs attaques aux propriétés de leurs cibles et à viser les cibles ayant certaines propriétés, telles que des avoirs élevés ou une association avec certains groupes politiques. Cependant, les attaquants doivent souvent faire la part des choses entre l'efficacité, l'extensibilité de leurs attaques et la précision de leur cible. Un autre exemple pourrait être l'utilisation d'essaims de drones qui déploient une technologie de reconnaissance faciale pour tuer des membres spécifiques de la foule, à la place de formes de violence moins ciblées.

Les cyberattaques deviennent de plus en plus alarmantes en complexité et en quantité, à cause du manque de conscience et de compréhension des besoins réels. Ce manque de soutien est à l'origine du manque de dynamisme, d'attention et de volonté d'engager des fonds et des ressources pour la cybersécurité dans de nombreuses organisations. Pour limiter l'impact des cyberattaques, les recommandations suivantes sont suggérées (Brundage *et al*. 2018) :

– les décideurs devraient collaborer étroitement avec les chercheurs techniques pour étudier, prévenir et limiter les utilisations malveillantes potentielles de l'intelligence artificielle ;

– les chercheurs et les ingénieurs en intelligence artificielle doivent prendre au sérieux la nature à double usage de leur travail, en permettant à des considérations liées à une utilisation abusive d'influencer les priorités et les normes de recherche et

en s'adressant de manière proactive aux acteurs concernés lorsque des applications nuisibles sont prévisibles ;

– les pouvoirs publics doivent chercher activement à élargir l'éventail des parties prenantes et des experts du domaine impliqués dans les discussions sur ces défis.

1.5. Conclusion

L'intelligence artificielle est un vaste domaine que les chercheurs et les experts en cybersécurité doivent explorer. À mesure que les systèmes intelligents gagneront en capacité, ils vont d'abord atteindre, puis dépasser les capacités humaines dans de nombreux domaines. Dans la cybersécurité, l'intelligence artificielle peut être utilisée pour renforcer les défenses de l'infrastructure informatique. À noter que, au fur et à mesure que l'intelligence artificielle s'étend à des domaines considérés comme étant réservés aux humains, les menaces de sécurité vont devenir variées, différentes et intelligentes par rapport aux techniques actuellement existantes. Il est très difficile de se défendre contre ces menaces, car même les experts en cybersécurité peuvent être la proie des attaques de *spear phishing*. Par conséquent, se préparer aux utilisations malveillantes potentielles de l'intelligence artificielle associées à cette transition est une tâche urgente. L'utilisation des techniques intelligentes vise à identifier les attaques en temps réel, avec peu ou pas d'interaction humaine, et à les arrêter avant qu'elles ne causent des dégâts. Nous pouvons en conclure que l'intelligence artificielle est considérée comme un outil puissant pour résoudre les problèmes de cybersécurité.

1.6. Bibliographie

Agarwal, R., Joshi, M.V. (2000). A new framework for learning classifier models in data mining [En ligne]. Disponible à l'adresse : https://pdfs.semanticscholar.org/db6e/1d67f7912efa65f94807dc81b24dea2de158.pdf.

Ahlan, A.R., Lubis, M., Lubis, A.R. (2015). Information security awareness at the knowledge-based institution: its antecedents and measures. *Procedia Computer Science*, 72, 361–373.

Ahmad, J. *et al*. (2016). A deep learning approach for network intrusion detection system. Dans *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*. ICST, New York.

Amodei, D. *et al*. (2016). Concrete Problems in AI Safety [En ligne]. Cornell University. Disponible à l'adresse : https://arxiv.org/abs/1606.06565.

Anderson, D., Frivold, T., Valdes, A. (1995). Next-generation intrusion detection expert system (NIDES). Rapport, Department of the Navy, Space and Naval Warfare Systems Command, San Diego.

Aslahi-Shahri, B.M. *et al*. (2016). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural computing and applications*, 27(6), 1669–1676.

Bace, R.G. (2000). *Intrusion detection*. Sams Publishing, Indianapolis.

Balan, E.V. *et al*. (2015). Fuzzy based intrusion detection systems in MANET. *Procedia Computer Science*, 50, 109–114.

Barth, C.J., Mitchell, J.C. (2008). Robust Defenses for Cross-Site Request Forgery. Dans *Proceedings of 15th ACM Conference*. CCS, Alexandria.

Biggio, B., Nelson, B., Laskov, P. (2012). Poisoning attacks against support vector machines. Dans *29th International Conference on Machine Learning*. ICML, Édimbourg, 1467–1474.

Brundage, M. *et al*. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation [En ligne]. Disponible à l'adresse : https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.

Capgemini Research Institute (2019). Reinventing Cybersecurity with Artificial Intelligence – The new frontier in digital security [En ligne]. Disponible à l'adresse : https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf.

Chebrolu, S. *et al*. (2005). Feature deduction and ensemble design of intrusion detection systems. *Computers & security*, 24(4), 295–307.

Chen, W.-H., Hsu, S.-H., Shen, H.-P. (2005). Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 32(10), 2617–2634.

Cova, M., Balzarotti, D., Felmetsger, V., Vigna, G. (2007). Swaddler: An Approach for the Anomaly-based Detection of State Violations in Web Applications. Dans *10th Proceedings of the International Symposium on Recent Advances in Intrusion Detection*. RAID, Gold Coast.

Cova, M., Kruegel, C., Vigna, G. (2010). Detection and Analysis of Drive-by-Download Attacks and Malicious JavaScript Code. Dans *Proceedings of the 19th international conference on World wide web*. WWW, Raleigh.

Crockford, D. (2015). Json [En ligne]. Disponible à l'adresse : https://github.com/douglascrockford/JSON-js/blob/master/README [Consulté en mars 2018].

Cunningham, R., Lippmann, R. (2000). Detecting Computer Attackers: recognizing patterns of malicious stealthy behavior. Présentation, CERIAS, Anderlecht.

Fortuna, C., Fortuna, B., Mohorčič, M. (2002). Anomaly detection in computer networks using linear SVMs [En ligne]. Disponible à l'adresse : http://ailab.ijs.si/dunja/SiKDD2007/Papers/Fortuna_Anomaly.pdf.

Hajimirzaei, B., Navimipour, N.J. (2019). Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *ICT Express*, 5(1), 56–59.

Hamamoto, A.H. *et al*. (2018). Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Systems with Applications*, 92, 390–402.

Han, X. *et al*. (2015). A Naive Bayesian network intrusion detection algorithm based on Principal Component Analysis. Dans *7th International Conference on Information Technology in Medicine and Education*. IEEE. Huangshan.

Heckerman, D. (2008). A tutorial on learning with Bayesian networks. Dans *Innovations in Bayesian networks*, Holmes, D.E., Jain, L.C. (dir). Springer, Berlin, 33–82.

Hoque, M.S. *et al*. (2012). An implementation of intrusion detection system using genetic algorithm. *International Journal of Network Security & Its Applications*, 4(2).

Hu, J. *et al*. (2009). A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection. *IEEE network*, 23(1), 42–47.

Hwang, T.S., Lee, T.-J., Lee, Y.-J. (2007). A three-tier IDS via data mining approach. Dans *Proceedings of the 3rd annual ACM workshop on Mining network data*. ACM, San Diego.

Idris, N.B., Shanmugam, B. (2005). Artificial intelligence techniques applied to intrusion detection. Dans *Annual IEEE India Conference-Indicon*. IEEE, Chennai.

Ingham, K., Somayaji, A., Burge, J., Forrest, S. (2007). Learning DFA representations of HTTP for protecting web applications. *Journal of Computer Networks*, 51(5), 1239–1255.

Kalaivani, S., Vikram, A., Gopinath, G. (2019). An Effective Swarm Optimization Based Intrusion Detection Classifier System for Cloud Computing. Dans *5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, Coimbatore.

Kang, M.-J., Kang, J.-W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. *PLOS ONE*, 11(6).

Khidzir, N.Z. *et al*. (2018). Information Security Requirement: The Relationship Between Cybersecurity Risk Confidentiality, Integrity and Availability in Digital Social Media. Dans *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, Yacob, N.A., Mohd Noor, N.A., Mohd Yunus, N.Y., Lob Yussof, R., Zakaria, S.A.K.Y. (dir.). Springer, Berlin.

Kim, J. *et al*. (2016). Long short term memory recurrent neural network classifier for intrusion detection. Dans *International Conference on Platform Technology and Service (PlatCon)*. IEEE, Jeju.

Kirdaa, E., Jovanovicb, N., Kruegelc, C., Vigna, G. (2009). Client-side cross-site scripting protection. *Computers & Security*, 28(7).

Kruegel, C. *et al*. (2003). Bayesian event classification for intrusion detection. Dans *Proceedings of the 19th Annual Computer Security Applications Conference*. IEEE, Las Vegas.

Kudłacik, P., Porwik, P., Wesołowski, T. (2016). Fuzzy approach for intrusion detection based on user's commands. *Soft Computing*, 20(7), 2705–2719.

Kumar, S., Krishna, C.R., Solanki, A.K. (2018). A Technique to Resolve Data Integrity and Confidentiality Issues in a Wireless Sensor Network. Dans *8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, Noida.

Landwehr, C. (2008). Cybersecurity and Artificial Intelligence: From Fixing the Plumbing to Smart Water. *IEEE, Security and privacy*, 6(5), 3–4.

Leung, K., Leckie, C. (2005). Unsupervised anomaly detection in network intrusion detection using clusters. Dans *Proceedings of the 28th Australasian conference on Computer Science*. Australian Computer Society Inc., Darlinghurst, 333–342.

Levent, E. *et al*. (2004). Minds-minnesota intrusion detection system. *Next generation data mining*, août, 199–218.

Li, L., De-Zhang, Y., Chen, F.-S. (2010). A novel rule-based Intrusion Detection System using data mining. Dans *3rd International Conference on Computer Science and Information Technology*. IEEE, Chengdu.

Liang, J. *et al*. (2019). A filter model for intrusion detection system in Vehicle Ad Hoc Networks: A hidden Markov methodology. *Knowledge-Based Systems*, 163, 611–623.

Liao, Y., Vemuri, V.R. (2002). Use of k-nearest neighbor classifier for intrusion detection. *Computers & security*, 21(5), 439–448.

Lippmann, R.P., Cunningham, R.K. (2000). Improving intrusion detection performance using keyword selection and neural networks. *Computer networks*, 34(4), 597–603.

Lunt, T. (1993). Detecting intruders in computer systems. Dans *Proceedings of the 1993 conference on auditing and computer technology*. Baltimore Convention Center, Baltimore.

Lunt, T., Tamaru, A., Gillham, F. (1992). A real-time intrusion-detection expert system (IDES). *Computer Science Laboratory*.

Mahoney, M.V., Chan, P.K. (2001). PHAD: Packet header anomaly detection for identifying hostile network traffic [En ligne]. Disponible à l'adresse : https://pdfs.semanticscholar.org/1505/f3658f5af7dff88e88d6a2b381de12e03036.pdf.

Mahoney, M.V., Chan, P.K. (2002a). Learning models of network traffic for detecting novel attacks. Rapport technique, Institut technologique de Floride, Melbourne.

Mahoney, M.V., Chan, P.K. (2002b). Learning nonstationary models of normal network traffic for detecting novel attacks. Dans *Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, Edmonton.

Menahem, E. *et al*. (2009). Improving malware detection by applying multi-inducer ensemble. *Computational Statistics & Data Analysis*, 53(4), 1483–1494.

Mishra, A., Agrawal, A., Ranjan, R. (2011). Artificial intelligent firewall. Dans *Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*. ACM, Rajpura/Punjab.

Moon, D. *et al*. (2017). DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. *The Journal of supercomputing*, 73(7), 2881–2895.

Moore, T., Anderson, R. (2012). *Internet security. The Oxford Handbook of the Digital Economy*. Oxford University Press, Oxford.

Mukkamala, S., Sung, A.H. (2003a). Artificial intelligent techniques for intrusion detection. Dans *International Conference on Systems, Man and Cybernetics*. IEEE, Washington.

Mukkamala, S., Sung, A.H. (2003b). A comparative study of techniques for intrusion detection. Dans *Proceedings of the 15th IEEE international conference on tools with artificial intelligence (ICTAI'03)*. IEEE, Washington.

Mukkamala, S., Sung, A.H., Abraham, A. (2005). Intrusion detection using an ensemble of intelligent paradigms. *Journal of Network and Computer Applications*, 28, 167–182.

Mutz, D., Robertson, W., Vigna, G., Kemmerer, R. (2007). Exploiting Execution Context for the Detection of Anomalous System Calls. Dans *Proceedings of the International Symposium on Recent Advances in Intrusion Detection*. RAID, Gold Coast.

Novikov, D., Yampolskiy, R.V., Reznik, L. (2006). Artificial intelligence approaches for intrusion detection. Dans *IEEE Long Island Systems, Applications and Technology Conference*. IEEE, Long Island.

Peltier, T.R. (2010). *Information security risk analysis*. CRC Press, Boca Raton.

Peng, K. *et al*. (2018). Intrusion detection system based on decision tree over big data in fog environment [En ligne]. Disponible à l'adresse : https://www.hindawi.com/journals/wcmc/2018/4680867/.

Ponkarthika, M., Saraswathy, V.R. (2018). Network intrusion detection using deep neural networks. *Asian Journal of Applied Sciences*, 2(2), 665–673.

Rai, K., Devi, M.S., Guleria, A. (2016). Decision tree based algorithm for intrusion detection. *International Journal of Advanced Networking and Applications*, 7(4), 2828.

Rawat, S. (2005). Efficient data mining algorithms for intrusion detection. Dans *Proceedings of the 4th Conference on Engineering of Intelligent Systems (EIS 2004)*. EIS, Madère.

Robertson, W., Maggi, F., Kruegel, C., Vigna, G. (2010). Effective Anomaly Detection with Scarce Training Data. Dans *Proceedings of the Network and Distributed System*. Security Symposium (NDSS), San Diego.

Roesch, M. (1999). Snort: Lightweight intrusion detection for networks. *Lisa*, 99(1).

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1/2), 1–39.

Sabhnani, M., Serpen, G. (2003). Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context. Dans *International Conference on Machine Learning; Models, Technologies and Applications*. MLMTA, Las Vegas.

Sahu, S., Mehtre, B.M. (2015). Network intrusion detection system using J48 Decision Tree. Dans *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, Kochi.

Sai Satyanarayana Reddy, S., Chatterjee, P., Mamatha, C. (2019). Intrusion Detection in Wireless Network Using Fuzzy Logic Implemented with Genetic Algorithm. Dans *Computing and Network Sustainability*, Peng, S.-L, Dey, N., Bundele, M. (dir.). Springer, Berlin, 425–432.

Scharre, P. (2015). Counter-Swarm: A Guide to Defeating Robotic Swarms [En ligne]. Disponible à l'adresse : https://warontherocks.com/2015/03/counter-swarm-a-guide-todefeating-robotic-swarms/.

Schneier, B. (2008). The psychology of security. Dans *International Conference on Cryptology in Africa.* AFRICACRYPT, Casablanca.

Shanmugavadivu, R., Nagarajan, N. (2011). Network intrusion detection system using fuzzy logic. *Indian Journal of Computer Science and Engineering*, 2(1), 101–111.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R. (2013). Intriguing properties of neural networks [En ligne]. Disponible à l'adresse : https://arxiv.org/ abs/1312.6199.

Tekerek, A., Bay, O.F. (2019). Design and implementation of an artificial intelligence-based web application firewall model. *Neural Network World*, 189, 206.

Teng, H.S., Chen, K. (1990). Adaptive real-time anomaly detection using inductively generated sequential patterns. Dans *Proceedings. 1990 IEEE Computer Society Symposium on Research in Security and Privacy*. IEEE, Oakland.

Turner, C. *et al*. (2016). A rule status monitoring algorithm for rule-based intrusion detection and prevention systems. *Procedia Computer Science*, 95, 361–368.

Valentín, K., Malý, M. (2014). Network firewall using artificial neural networks. *Computing and Informatics*, 32(6), 1312–1327.

Vapnik, V. (1998). *Statistical learning theory*. Wiley, Hoboken.

Veiga, A.P. (2018). Applications of artificial intelligence to network security [En ligne]. Disponible à l'adresse : https://arxiv.org/abs/1803.09992.

Vinayakumar, R., Soman, K.P., Poornachandran, P. (2017). Applying convolutional neural network for network intrusion detection. Dans *6th International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Manipal University, Karnataka.

Witten, I.H. *et al*. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, Burlington.

Yang, Y. *et al.* (2013). Rule-based intrusion detection system for SCADA networks. Dans *2nd IET Renewable Power Generation Conference (RPG 2013)*. RPG, Beijing.

Zainal, A., Maarof, M.A., Shamsuddin, S.M. (2009). Ensemble classifiers for network intrusion detection system. *Journal of Information Assurance and Security*, 4(3), 217–225.

Zegeye, W.K., Moazzami, F., Dean, R. (2018). Hidden Markov Model (HMM) based Intrusion Detection System (IDS). *International Telemetering Conference Proceedings*, 1–10.

2

Un plan de contrôle intelligent pour le déploiement de services de sécurité dans les réseaux SDN

Maïssa MBAYE[1], Omessaad HAMDI[2] et Francine KRIEF[3]

[1] *Université Gaston Berger, Saint-Louis, Sénégal*

[2] *IMT Atlantique, Rennes, France*

[3] *ENSEIRB-MATMECA, Bordeaux, France*

2.1. Introduction

L'approche SDN (*Software-Defined Networking*) consiste à piloter une infrastructure réseau par des applications logicielles (Fortes 2013 ; ONF 2015). Ces applications peuvent être hébergées sur un ou plusieurs serveurs et permettent de *contrôler* les composants réseaux (physiques ou virtuels) de l'infrastructure. Ce nouveau modèle d'architecture des réseaux, fondé sur la séparation entre le *plan de contrôle* et le *plan de données*, attire de plus en plus l'attention des grands acteurs des réseaux et télécoms (ONF, IETF, ITU-T, ETSI) (Kreutz *et al.* 2015).

Le plan de données correspond aux équipements réseaux qui sont responsables du transfert optimisé (au mieux) des données jusqu'au destinataire. La tâche principale dans ce plan consiste à consulter une table de transfert/flux afin d'effectuer la retransmission correcte des données. Pour le cas des réseaux SDN/OpenFlow, cette table contient des règles de transfert des flux qui vérifient la valeur des champs des différents en-têtes (Ethernet, IPv4/v6, MPLS, TCP/UDP, etc.) et en

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

déduit l'action à mener (retransmettre, supprimer, modifier l'entête, etc.). À partir de l'année 2015, les principaux constructeurs (IBM, Hewlett-Packard, Huawei, Juniper, NEC, etc.) avaient déjà mis sur le marché des équipements réseaux compatibles SDN/OpenFlow (Kreutz *et al*. 2015).

Le plan de contrôle représente toute la logique qui permet de piloter le contenu des tables de transfert. Il a une vue globale du réseau et inclut des protocoles permettant la programmation du plan de données en fonction des besoins des applications qui sont déployées dans le réseau (routage, contrôle d'accès, QoS, répartition de charge, ingénierie de trafic, sécurité, etc.). Ce plan de contrôle est implanté sur une plateforme logicielle centralisée (logiquement) dans un *Cloud* et qui est appelé *contrôleur*. Ce plan offre des fonctionnalités innovantes, comme la virtualisation des réseaux.

Les principaux bénéfices des réseaux SDN et de la virtualisation réseau sont que les réseaux de nouvelle génération seront plus flexibles, agiles, adaptables et hautement automatisés. En effet, la transition IPv4/IPv6 montre que le fait que les équipementiers soient responsables du déploiement de nouveaux protocoles et applications réseaux dans leurs produits rallonge la mise sur le marché. Avec cette nouvelle approche, il devient possible pour chaque structure de programmer de nouvelles applications et protocoles réseaux sur les équipements sans être obligés d'attendre une nouvelle version de l'équipement. Toutefois, cette possibilité pose aussi un certain nombre de défis pour la sécurité et le passage à grande échelle.

Les défis de sécurité se présentent sous deux aspects principalement : la sécurisation du réseau SDN et le déploiement de services de sécurité. Les travaux sur la sécurité des réseaux SDN se focalisent essentiellement sur celle du contrôleur, qui est un point de décision centralisé. En effet, si le contrôleur est compromis, hors service ou déconnecté du plan de données, la cohérence du comportement du réseau n'est plus garantie. Le contrôleur peut être l'objet d'attaques, telles que le DDoS, les accès non autorisés, l'injection de politiques de sécurité, etc. (Shu *et al*. 2016). Différentes solutions ont été proposées allant des pare-feu (Wang *et al.* 2013) aux systèmes de détection/prévention d'intrusion (IDS/IPS) (Gowtham *et al.* 2018), afin de faire face aux menaces visant la sécurité des réseaux SDN.

Le déploiement d'applications de sécurité (par exemple IDS/IPS) dans les réseaux SDN consiste, quant à lui, à fournir la sécurité comme service du réseau SDN pour les clients. Les deux défis majeurs du déploiement des services de sécurité sont : le caractère dynamique du contrôle des réseaux SDN et la grande quantité de données à traiter par ces solutions. Il s'agit ici de déployer des services de sécurité qui offriront des temps de réponses et de traitement acceptables.

Au niveau des réseaux classiques, l'apprentissage machine (*Machine Learning*) et l'intelligence artificielle (IA) ont montré, de manière générale, leur efficacité pour

la sécurité (Das et Nene 2017). Aujourd'hui, l'IA est de plus en plus populaire pour la résolution de tout type de problème, grâce en partie à la quantité de données produites à travers Internet. La sécurité des réseaux SDN n'échappe pas à cette tendance et plusieurs travaux abordent déjà le problème de sécurité dans les réseaux SDN, avec des solutions s'appuyant sur des outils de l'IA (Abubakar et Pranggono 2018 ; Dey *et al.* 2018 ; Xie *et al*. 2019). La finalité ultime serait d'avoir des réseaux SDN intelligents capables de s'autoprotéger et de s'auto-optimiser.

L'objectif de ce chapitre est d'aborder des techniques de contrôle intelligent, basées sur l'IA, afin de permettre un pilotage intelligent du déploiement de la sécurité.

Nous présentons tout d'abord les réseaux SDN dans la section 2.2, puis nous abordons le problème de la sécurité, avant de décrire une architecture pour le déploiement de services de sécurité pour les réseaux SDN dans la section 2.3. Par la suite, section 2.4, nous présentons différentes initiatives en vue de rendre les réseaux SDN « intelligents ». La section 2.5 est dédiée à l'apport de l'intelligence artificielle pour la sécurité puis, section 2.6, pour la sécurité dans les réseaux SDN. Nous décrivons notre proposition de déploiement d'un service intelligent de protection contre les intrusions dans la section 2.7. La section 2.8 est dédiée aux principaux défis à relever pour utiliser largement les outils de l'IA pour la sécurité dans les réseaux SDN. Finalement, la section 2.9 conclut ce chapitre.

2.2. Les réseaux SDN (*Software-Defined Networking*)

Le concept de réseaux SDN est une évolution importante des architectures réseaux traditionnels. Dans cette section, nous allons en présenter les principaux concepts et éléments architecturaux.

2.2.1. *Architecture générale*

L'architecture SDN est composée principalement de trois plans (figure 2.1) : *le plan application*, *le plan de données*, *le plan de contrôle*. Le cœur de cette architecture consiste en la séparation entre le plan de contrôle et le plan de données (ONF 2015).

Le plan application contient aussi bien les applications SDN que des applications réseaux. Le contrôleur SDN utilise une API, appelée *Northbound API*, pour interagir avec le plan application. Dans ce plan, on peut trouver des applications permettant d'implanter des fonctionnalités purement réseaux, comme la QoS, le routage, etc. Il peut y avoir aussi d'autres applications SDN dont le rôle est de contrôler la logique du réseau SDN (Bannour *et al.* 2018).

Le plan de données est constitué des équipements physiques ou virtuels d'interconnexion dont la principale tâche est l'acheminement des données (*Forwarding*). En effet, dans les réseaux traditionnels, les équipements réseaux contiennent le plan de contrôle et le plan de données. Avec cette nouvelle approche, le plan de contrôle est externalisé pour rendre l'équipement réseau plus efficace. Ces éléments réseaux (principalement des *switchs*) contiennent des tables de transfert (*Forwarding Tables*) et sont contrôlables par logiciel à distance *via* des API.

Le plan de contrôle est implanté au niveau du contrôleur SDN, généralement sur un serveur physique ou dans le Cloud. Le contrôleur gère toute la logique « intelligente » du réseau SDN en programmant (manipulation à la volée) le contenu des tables de transfert au niveau du plan de données. Le contrôleur SDN gère l'infrastructure dans sa globalité et est capable de s'informer en temps réel sur l'état et l'activité des équipements (physiques ou virtuels) qu'il pilote.

Dans le contexte du routage, le plan de contrôle correspondrait au composant chargé de trouver les meilleurs chemins pour que l'équipement de transfert identifie déjà sa table prête à l'emploi. Il est logiquement centralisé mais peut être physiquement distribué entre plusieurs éléments (Bannour *et al.* 2018). Lorsque les contrôleurs sont physiquement distribués, ils peuvent communiquer en utilisant les *Application Programming Interface* (API)-Est et Ouest.

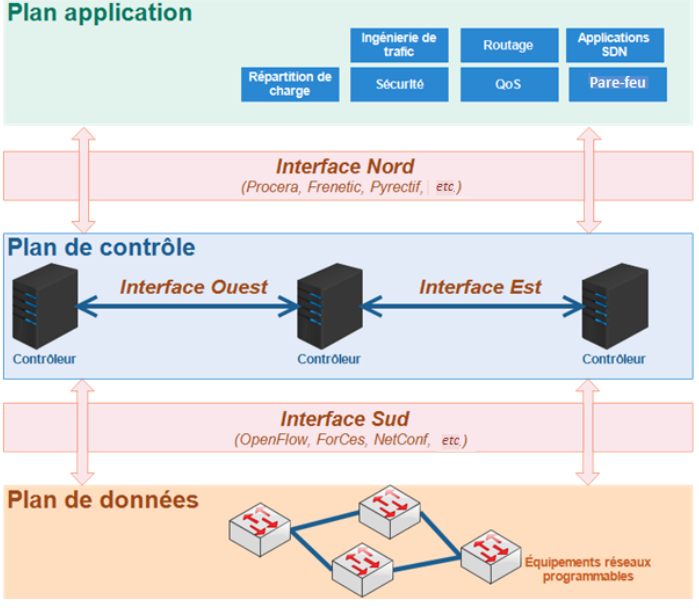


Figure 2.1. *Architecture simplifiée de SDN (Zhang et al. 2017)*

La communication entre le plan de contrôle et le plan de données se fait *via* une classe de protocoles appelée « API-Sud » (*SouthBound Interface*). En 2019, l'API-Sud la plus avancée et disponible dans le commerce pour les réseaux SDN est OpenFlow[1]. Toutefois, il existe des alternatives telles que ForCES, SNMP et NetConf de l'IETF. Il y a aussi d'autres approches de la programmation des équipements réseaux comme P4 (Bosshart *et al.* 2014 ; Cordeiro *et al.* 2017), qui sont en train de se développer.

OpenFlow était à l'origine un projet de l'université de Stanford. C'est aujourd'hui un protocole réseau standard, publié par l'Open Networking Foundation (ONF) et qui sert de lien entre le plan de contrôle et le plan de données. Ce protocole est constitué d'instructions (règles) qui permettent de programmer les tables de transfert d'un équipement réseau, appelées ici « tables de flux » (*flow tables*). Les instructions définissent une action sur le trafic (transmettre le paquet, rejeter le paquet, etc.). Différentes versions du protocole OpenFlow ont été introduites pour ajouter plus de flexibilité et de fiabilité en incluant plusieurs tables de flux, des capacités de correspondance/action améliorées, des ports optiques, des tables de groupe, etc.[2]. En outre, il existe de nombreux contrôleurs OpenFlow disponibles, tels que POX (Kaur *et al.* 2014), Beacon (Erickson 2013), OpenDayLight[3].

2.2.2. *Distribution logique du contrôle SDN*

La centralisation physique du plan de contrôle dans un seul composant logiciel programmable, appelé « contrôleur », pose un certain nombre de problèmes et en particulier de passage à l'échelle, de disponibilité et de fiabilité (Nkosi *et al*. 2016 ; Karakus et Durresi 2017). Ainsi, plus le nombre de commutateurs augmente, plus le contrôleur SDN est sollicité et donc surchargé en termes de bande passante, de puissance de traitement et de capacité mémoire. Les délais de communication entre le contrôleur SDN et les commutateurs augmentent également avec la taille et l'étendue du réseau, ce qui a pour conséquence d'affecter la latence des flux de données (Bannour *et al*. 2018). Une façon de résoudre le problème de passage à l'échelle consiste à étendre les responsabilités du plan de données afin d'alléger la charge du contrôleur, mais cela impose de modifier la conception des commutateurs (Rebecchi *et al.* 2017).

Il semble donc nécessaire de considérer le plan de contrôle comme un système distribué dans lequel plusieurs contrôleurs SDN sont chargés de gérer l'ensemble du

1. Disponible à l'adresse : www.opennetworking.org/software-defined-standards/specifications/.

2. Plus de détails concernant les différentes versions de la spécification OpenFlow sont disponibles à l'adresse : www.opennetworking.org/software-defined-standards/specifications/.

3. Disponible à l'adresse : http://www.opendaylight.org/.

réseau, tout en maintenant une vue réseau logiquement centralisée. Cette solution assure en effet un meilleur passage à l'échelle du plan de contrôle réseau, tout en réduisant les latences du plan de contrôle. L'utilisation de plusieurs contrôleurs améliore également la fiabilité en éliminant le problème du point de défaillance unique.

Plusieurs travaux se sont intéressés à la manière de fournir un contrôle distribué. Dans (Oktian *et al.* 2017), les auteurs ont classé les différentes approches sur la manière de fournir une vue logiquement centralisée à plusieurs instances de contrôleurs distribués en fonction du choix de conception réalisé. Ils ont ainsi identifié deux architectures distribuées : l'architecture plate et l'architecture hiérarchique (figure 2.2). Dans l'architecture plate, chaque contrôleur gère un sous-réseau/domaine du réseau global. Dans l'architecture hiérarchique, des contrôleurs locaux gèrent les besoins des applications locales à leur domaine, tandis que le contrôleur principal, généralement appelé « racine », traite les besoins des applications demandant une vue réseau globale.

Dans (Bannour *et al.* 2018), les auteurs ont classé les architectures de contrôle en fonction de la manière dont les connaissances sont diffusées entre les instances du contrôleur. Deux architectures sont mentionnées, la première centraliseé et la seconde distribuée. Cette classification leur a ensuite permis de comparer les différentes plateformes de contrôleur SDN en termes de critères d'évolutivité, de fiabilité et de performance. Les auteurs privilégient l'organisation hiérarchique du plan de contrôle pour une meilleure évolutivité et de meilleures performances. Chaque contrôleur peut ainsi avoir différentes responsabilités et prendre des décisions basées sur une vue partielle du réseau. Le niveau le plus haut agit comme contrôleur centralisé, avec le problème du point de défaillance unique.

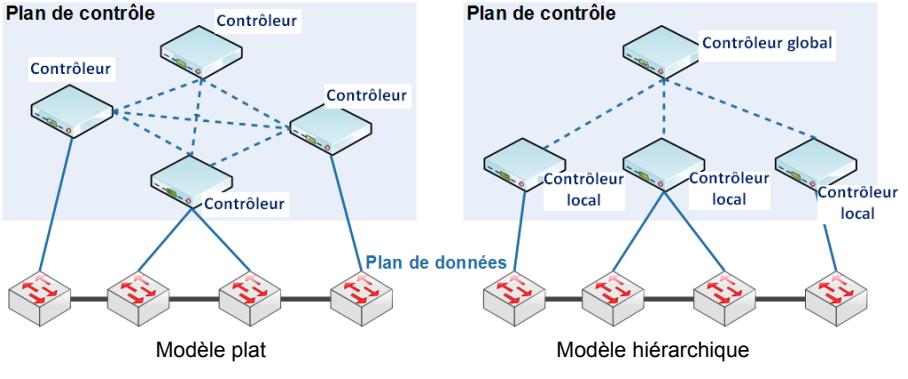


Figure 2.2. *Modèles de distribution du plan de contrôle*

Le tableau 2.1 présente une classification des informations réseau devant être échangées entre contrôleurs distribués, afin de maintenir la vue globale.

Catégorie	Exemples
État du réseau local (statique)	Atteignabilité Topologie Capacité Paramètres QoS
État du réseau local (dynamique)	Tables des flux Règles de flux Utilisation en temps réel de la bande passante Chemins des flux
Événement contrôleur	État du contrôleur Information propre à l'application SDN
Inventaire	Liste des applications SDN installées Liste des commutateurs connectés

Tableau 2.1. *Classification des informations réseau partagées (Oktian et al. 2017)*

La distribution du contrôle SDN pose également un certain nombre de défis, parmi lesquels l'identification du nombre requis de contrôleurs ainsi que leurs emplacements appropriés par rapport aux objectifs de performance et de fiabilité visés. Ainsi, l'organisation hiérarchique du plan de contrôle nécessite un schéma de répartition du contrôle pertinent prenant en compte à la fois l'organisation du plan de contrôle SDN et le placement physique des contrôleurs SDN (Bannour *et al*. 2018). Enfin, pour conserver la vue logiquement centralisée, la nécessité de partager les connaissances entre contrôleurs peut introduire de nouveaux problèmes de passage à l'échelle. En effet, la propagation fréquente des mises à jour d'état peut rendre le réseau indisponible, ce qui aura pour effet d'augmenter la latence entre contrôleur-commutateurs.

Un autre défi important concerne la tolérance aux pannes de l'architecture de contrôle distribuée. Ainsi, des stratégies de coordination des contrôleurs doivent être appliquées pour parvenir à des accords et résoudre également les problèmes de concurrence de mises à jour et de cohérence des états (Bannour *et al*. 2018). La cohérence d'état entre contrôleurs SDN logiquement centralisés est d'ailleurs un défi

de conception majeur des réseaux SDN, qui implique des compromis entre l'application des règles et les performances du réseau. Il est en effet très difficile d'obtenir une cohérence élevée dans un environnement SDN sujet aux défaillances du réseau sans compromettre sa disponibilité et sans ajouter de la complexité à la gestion de l'état du réseau. Les approches récentes ont introduit le concept de cohérence adaptative, où les contrôleurs peuvent ajuster leur niveau de cohérence pour atteindre le niveau de performances souhaité, en fonction de métriques spécifiques. Cette cohérence devrait être prise en compte lors de la recherche du placement optimal des contrôleurs. En effet, la minimisation des distances entre contrôleurs est importante pour les performances du système ; elle facilite les communications entre contrôleurs et améliore la cohérence de l'état du réseau. Dans (Canini *et al.* 2014), les auteurs se sont intéressés à ce problème et ont proposé un plan de contrôle SDN distribué robuste, appelé « STN » (*Software Transactional Networking*).

La sécurité du réseau SDN distribué constitue un autre défi crucial. La décentralisation du contrôle SDN réduit le risque associé à un seul point de défaillance et à des attaques (DDoS par exemple). Cependant, l'intégrité des flux de données entre les contrôleurs SDN et les commutateurs n'est toujours pas sûre. On pourrait imaginer qu'un attaquant puisse corrompre le réseau en agissant en tant que contrôleur SDN.

Enfin, il est également difficile d'assurer l'interopérabilité entre contrôleurs SDN distribués appartenant à différents domaines SDN et utilisant différentes technologies de contrôleur.

Plusieurs travaux de recherche ont proposé d'intégrer des approches automatiques et adaptatives dans le plan de contrôle distribué afin de relever certains défis cités précédemment. Ainsi, dans (Ma *et al*. 2018), les auteurs utilisent l'apprentissage par renforcement pour automatiser le processus de gestion et d'allocation des ressources dans un réseau SDN distribué.

2.3. La sécurité dans les réseaux SDN

La sécurité est l'un des facteurs freinant le déploiement des architectures SDN, qui sont aujourd'hui principalement centralisées. L'aspect centralisé du plan de contrôle présente en effet beaucoup d'avantages, mais il soulève également un problème de sécurité très important, car tous les plans de contrôle des switchs sont placés en un seul point. Typiquement, les attaques de déni de service (DoS) deviennent très impactantes, car le plan de contrôle n'est plus distribué, ce qui risque de compromettre tout le réseau.

Nous présenterons dans cette section les différentes surfaces d'attaques des réseaux SDN, avant de décrire une architecture pour le déploiement de services de sécurité pour les réseaux SDN.

2.3.1. *Les surfaces d'attaques*

En termes d'identification et de réponse aux attaques, SDN présente deux avantages essentiels par rapport aux réseaux traditionnels (Fortes 2013 ; Tang *et al.* 2016) :

– le plan de contrôle permet à un administrateur de séparer et de bloquer les attaques sur tout matériel hétérogène (pas besoin de reconfigurer individuellement chaque composant) ;

– au lieu d'investir dans un système de détection d'intrusion coûteux, SDN peut en faire une tâche répartie entre les nœuds ; de plus, chaque nœud peut devenir un pare-feu, un proxy, etc.

En inconvénients, SDN offre une opportunité aux attaquants lorsqu'il expose de nouvelles interfaces, c'est-à-dire une communication entre le plan de contrôle et le plan de données. En compromettant le contrôleur SDN, l'ensemble du réseau peut être compromis. Par conséquent, lors de l'utilisation de SDN pour fournir des services IDS (Systèmes de détection d'intrusion), il faut garder à l'esprit la sécurité de ce dernier. Une stratégie de contrôle d'accès étanche pour le contrôleur SDN doit donc être conçue et mise en œuvre.

Sécurité du plan d'application : le plan d'application comprend différents types d'applications. Certaines jouent un rôle très important dans l'élaboration des règles de flux. Attaquer ces applications peut causer un dysfonctionnement du réseau SDN. Un attaquant peut injecter un code malicieux dans l'application ou accéder illégalement au réseau SDN. Développer un modèle de contrôle d'accès ou un mode de vérification du code est efficace contre ce genre d'attaques (Klaedtke *et al*. 2014).

Sécurité des contrôleurs : le contrôleur SDN est l'élément le plus important. Un attaquant qui arrive à attaquer le contrôleur peut avoir facilement le contrôle total du réseau. Les vulnérabilités au sein du plan de contrôle peuvent aussi mener à contrôler le contrôleur d'une manière illégale. De plus, un attaquant peut lancer une *flooding attack* en utilisant les vulnérabilités du switch (envoyer plusieurs paquets d'un switch compromis à un contrôleur peut rendre ce contrôleur hors service) ou du protocole OpenFlow (un attaquant peut envoyer des paquets qui ne correspondent pas à la table et donc le switch va tous les transmettre au contrôleur, ce qui peut

engendrer un DoS/DDoS). Le contrôleur est la cible principale des attaques de type déni de service :

– sécurité du plan de données : le switch est la partie cruciale du plan de données. Il transfère les paquets, lit la table des adresses MAC et les requêtes ARP, etc. Plusieurs attaques sont possibles dans le plan de données, telles que les DoS/DDoS. Une autre attaque peut se faire *via* la taille des tables de flux, qui peut être très importante et peut donc ralentir la vitesse de commutation, surtout quand l'interface Sud est compromise ;

– sécurité des protocoles : les protocoles SDN sont principalement les protocoles des interfaces *SouthBound* et *NorthBound*. OpenFlow est typiquement un protocole de l'interface *SouthBound*. Bien qu'il soit largement utilisé, il présente encore des vulnérabilités puisqu'il n'y a pas d'identification et de contrôle d'accès pour les communications entre le *switch* et le contrôleur SDN. La sécurité des protocoles peut être assurée en ajoutant un contrôle d'accès aux ressources du réseau dans le plan de données, à partir du plan de contrôle.

2.3.2. *Exemple de déploiement de services de sécurité dans les réseaux SDN : le service IPSec*

Les travaux dans (Coly et Mbaye 2019) proposent un *framework* pour le déploiement de services de sécurité sur un réseau SDN. Les tunnels IPSec sont pris comme exemple de service de sécurité pour illustrer le fonctionnement de ce *framework*. IPsec (*Internet Protocol Security*) (Kent et Seo 2005) est une suite de protocoles assurant la sécurité au niveau de la couche IP du modèle TCP/IP. Ce *framework* peut être utilisé pour fournir un réseau privé virtuel (VPN) ou établir des tunnels sécurisés entre deux sites. Ce protocole utilise le protocole *Internet Key Exchange* (IKE) pour la négociation et la gestion des clés.

Les principaux points forts de cette proposition sont : la proposition d'une architecture pour le déploiement de services de sécurité dans les réseaux SDN, une nouvelle extension du protocole OpenFlow pour la gestion des tunnels sécurisés et, enfin, l'intégration d'un mécanisme de tunnel basé sur IPSec dans les réseaux SDN comme cas d'utilisation.

2.3.2.1. *Architecture générale*

L'architecture proposée est présentée figure 2.3. Les réseaux (clients) sont reliés par un réseau de cœur SDN. Le cœur du réseau SDN est composé de commutateurs BGS (*Border Gateway Switches*) et de commutateurs CIS (*Core Internal Switches*).

Ces commutateurs communiquent avec le contrôleur SDN à travers l'API-Sud OpenFlow. Le contrôleur SDN coordonne le déploiement des services de sécurité dans le réseau SDN. Dans le cas de tunnels déployés entre des réseaux clients différents, le service de sécurité est uniquement déployé au niveau des BGS *ingress* et *egress*. Dans les autres cas de figure, le service est déployé dans tous les systèmes CIS du chemin.

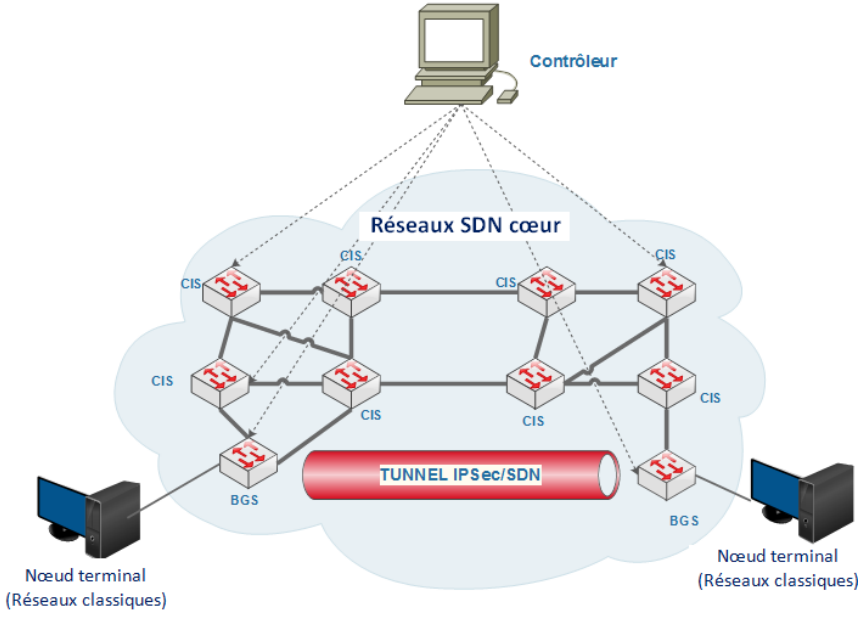


Figure 2.3. *Architecture de déploiement d'un service de tunnels IPSec dans un réseau SDN*

Lorsqu'un client souscrit à un SLA (*Service Level Agreement*) comprenant un service de sécurité, le contrôleur SDN est configuré conformément à ce contrat. Les flux du client vont désormais entraîner le déploiement de ce service en envoyant un message « SecTrans » avec les politiques de sécurité aux commutateurs BGS impliqués (*ingress/egress*) dans cette communication. La transaction restante est gérée par *SouthBound Extension*. En tant que cas d'utilisation illustrant le fonctionnement de notre proposition, nous déployons un service de *tunneling* sécurisé basé sur IPsec, en tant que service de sécurité SDN. Pour ce faire, nous avons étendu le protocole OpenFlow pour l'établissement du tunnel IPsec et défini une structure étendue pour les tables de flux.

Le contrôleur SDN est chargé de la génération et de la transmission des informations d'identification IKE. Il est également responsable du contrôle et de l'application des SPD IPsec. Il dispose donc d'une vue centralisée du réseau et des politiques de sécurité. Le composant IKE implémenté dans la ressource réseau s'exécute pour créer les associations de sécurité IPsec à l'aide de ces stratégies et informations d'identification. La figure 2.4 illustre le processus de déploiement de tunnel.

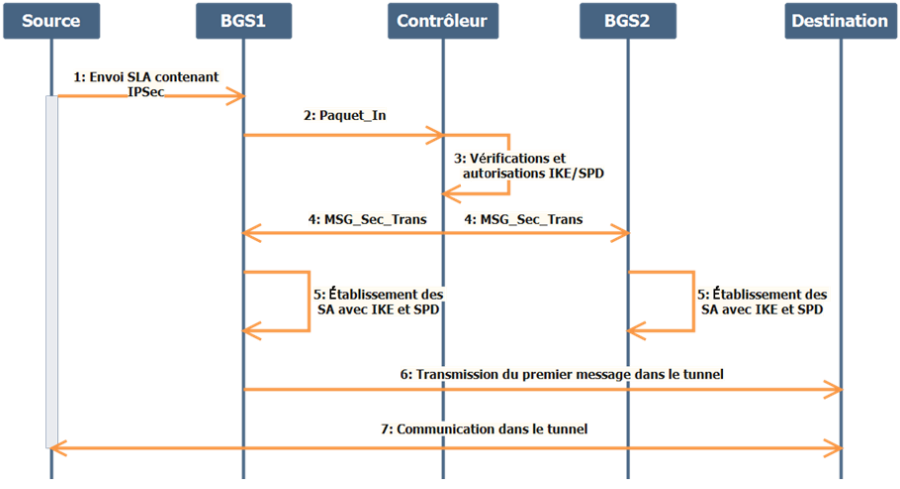


Figure 2.4. *Processus de déploiement de tunnel*

Si un nœud terminal envoie un trafic vers un autre client en utilisant un tunnel IPsec, la procédure suivante aura lieu :

1) le BGS connecté à la source enverra un message « Packet_In » (un message OpenFlow) au contrôleur pour lui demander le traitement à appliquer à ces paquets ;

2) s'il y a correspondance avec le trafic d'un abonné de service, le contrôleur génère les informations d'identification IKE et les politiques SPD puis les envoie aux deux commutateurs (entrée et sortie) BGS impliqués en plus du message SecTrans, pour permettre la transmission dans le tunnel IPsec ; le message ajoutera/modifiera un flux avec « yes » sur le champ IPsec ;

3) les BGS utilisent les informations d'identification IKE et les stratégies SPD pour établir des associations de sécurité avant de commencer la transmission.

Une fois le tunnel établi, tous les messages entre ces deux points de terminaison sont transmis en utilisant ce tunnel.

2.3.2.2. *Évaluation de performance du déploiement*

La proposition a été évaluée en utilisant Mininet[4], OpenSwitch et Floodlight[5] comme contrôleur SDN. Le tableau 2.2 résume les configurations des nœuds de notre banc de test.

Nœud	Système d'exploitation	Composants (software)	CPU	RAM
Contrôleur	Debian 8.4	Floodlight master	(4) @ 3,2 GHz	4 GB
Machine A	Debian 8.4	Mininet 2.2.2, Racoon, ipsec-tools	(4) @ 3,2 GHz	4 GB
Machine B	Ubuntu 18.04	Mininet 2.2.2, Racoon, ipsec-tools	(4) @ 2,4 GHz	4 GB
Nœud terminal A	Ubuntu 18.04	Racoon, ipsec-tools, iperf, top	(8) @ 4,0 GHz	8 GB
Nœud terminal B	Ubuntu 14.04	Racoon, ipsec-tools, iperf, top	(4) @ 3,2 GHz	8 GB

Tableau 2.2. *Configuration système des réseaux SDN*

Le cœur des réseaux SDN de notre banc de test contient dix commutateurs OpenFlow, cinq sur chaque réseau (figure 2.5). Ces commutateurs sont hébergés sur des ordinateurs physiques interconnectés comme illustré par la figure 2.6.

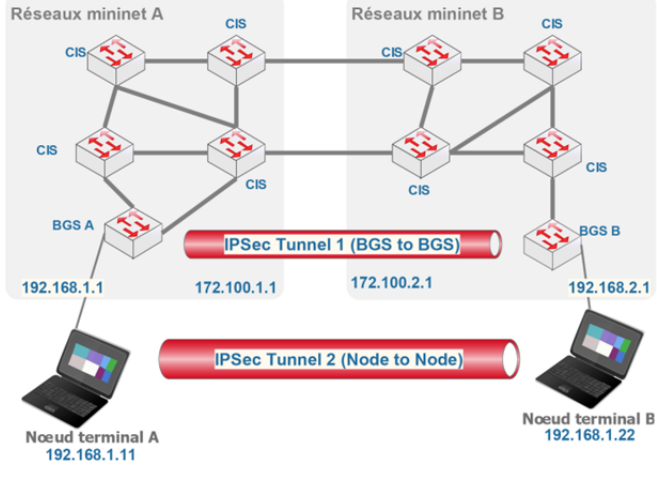


Figure 2.5. *Topologie logique du banc de test*

4. Disponible à l'adresse : http://www.mininet.org.

5. Floodlight OpenFlow Controller – Project Floodlight, Big switch network, disponible à l'adresse : http://www.projectfloodlight.org/floodlight/.

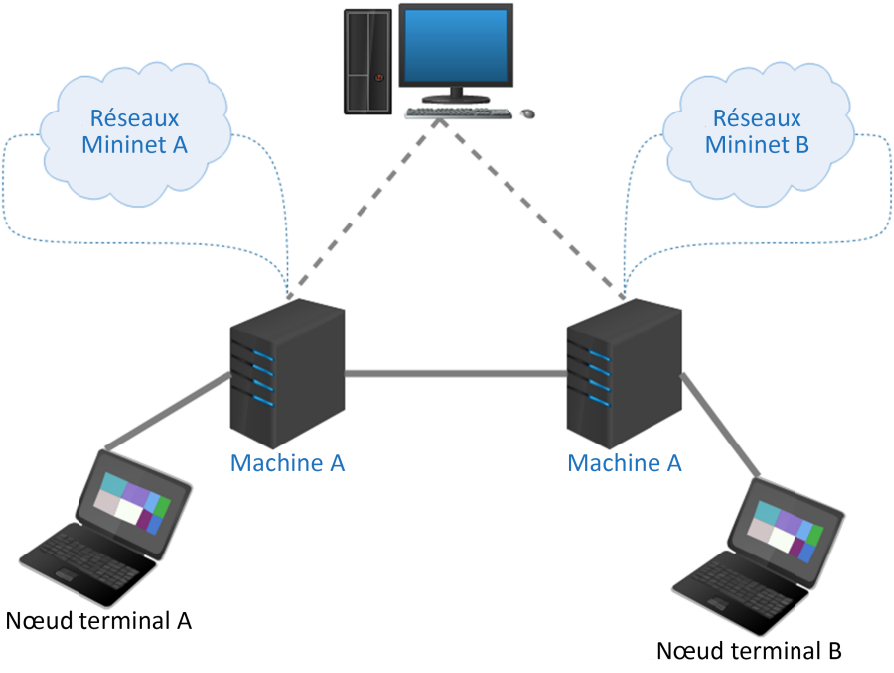
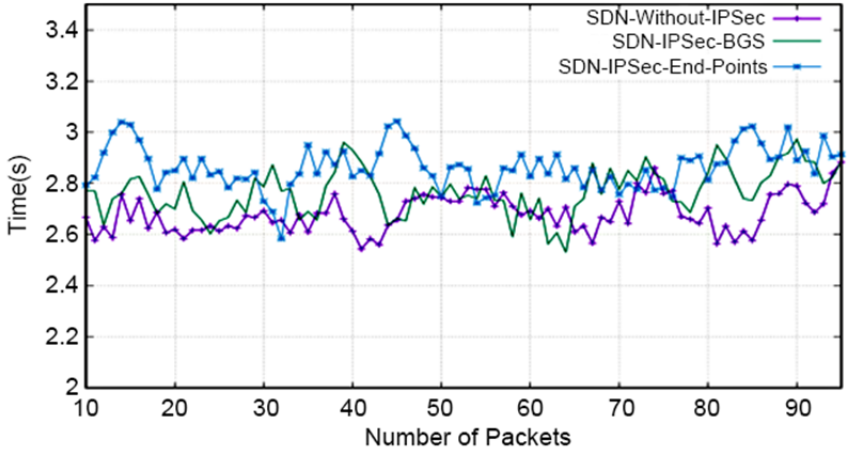


Figure 2.6. *Topologie physique du banc de test*

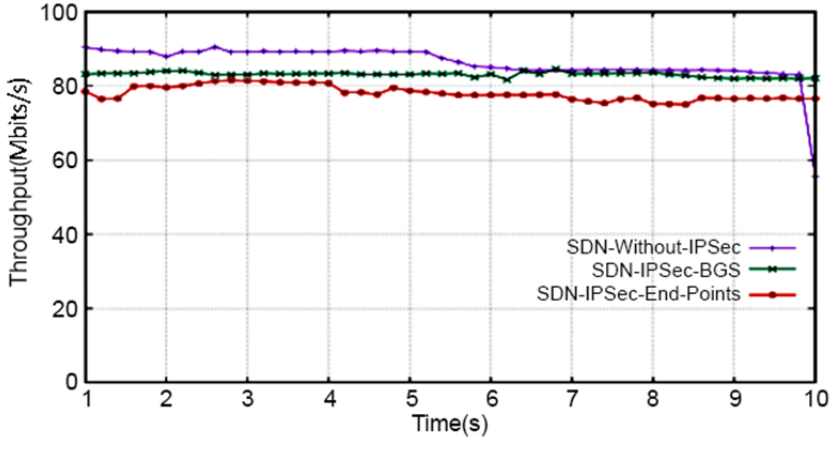
L'évaluation des performances est faite en termes de délai de transmission des données, débit, gigue et charge du processeur. Lors des tests, nous comparons les performances d'un réseau SDN sans IPsec avec un tunnel IPsec aux extrémités, au niveau des clients, et enfin dans le cas d'un réseau SDN avec un tunnel IPSec entre des commutateurs BGS.

Sur ces résultats, nous pouvons noter que, pour cette configuration, le délai est plus important lorsque les extrémités du tunnel IPsec sont au niveau des nœuds terminaux qu'entre les passerelles de sécurité que sont les BGS (figure 2.7A(a)). Nous avons un résultat similaire si nous regardons le *throughput* (figure 2.7A(b)), la gigue (figure 2.7B(c)) et la charge du processus (figure 2.7B(d)).

Bien que la portée de ces résultats soit à confirmer, ils peuvent être un argument en faveur du déploiement de services de sécurité au niveau des passerelles (en opposition aux nœuds terminaux). Avec ce type de résultat, il est envisageable de gérer la sécurité comme service au niveau du contrôleur.

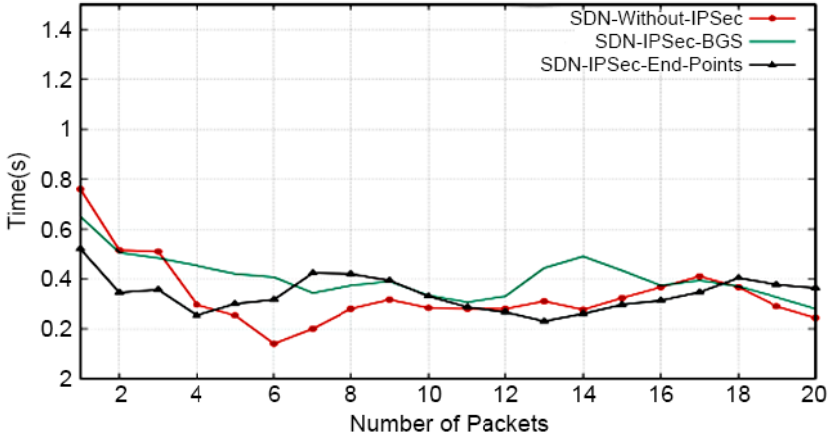


a) Performance en termes de délai

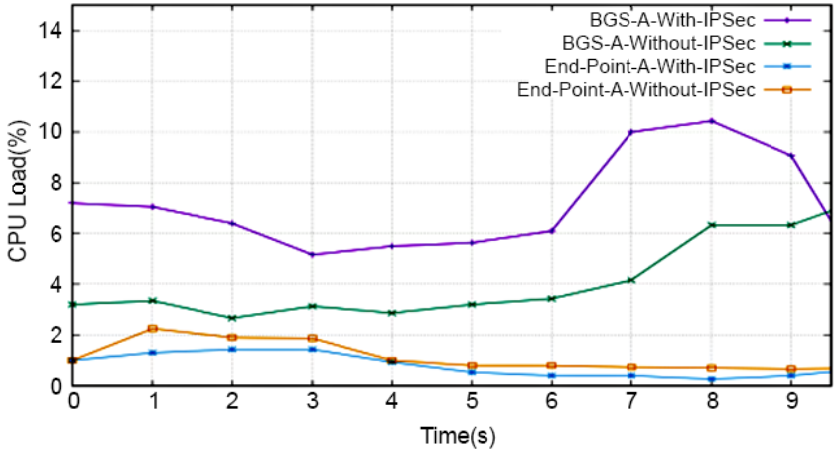


b) Performance du service de sécurité SDN en termes de débit

Figure 2.7A. *Performance du déploiement du service de sécurité IPSec sur SDN*



c) Performance du service de sécurité SDN en termes de gigue



d) Performance du service de sécurité SDN en termes de charge CPU

Figure 2.7B. *Performance du déploiement du service de sécurité IPSec sur SDN (suite)*

2.4. L'intelligence dans les réseaux SDN

L'ajout de l'intelligence dans les réseaux n'est pas nouveau mais est un défi d'un enjeu important. Dans cette section, nous commencerons par présenter le plan de connaissance qui a été proposé dans les années 2000 pour permettre aux réseaux de

prendre des décisions, en particulier de gestion, de manière « autonomique » (Krief 2010). Nous introduirons ensuite la notion de réseaux KDN (*Knowledge-Defined Networking*), qui doit permettre un contrôle et une exploitation autonomes des réseaux SDN. Nous terminerons avec la notion, très similaire, de réseaux IDN (*Intelligence-Defined Networking*) définis par l'intelligence ou réseaux fondés sur l'intention, nouveau paradigme porté par Huawei, qui se positionne en leader sur ce marché.

2.4.1. *Plan de connaissance*

En 2003, Clark *et al.* (2003) ont proposé un nouveau plan permettant de gérer la connaissance dans un réseau : le plan de connaissance (*Knowledge Plane* en anglais). L'objectif principal de ce nouveau plan était de résoudre les limites spécifiques à Internet, en particulier le manque de fiabilité et d'adaptation face à une situation nouvelle telle qu'une attaque de sécurité (Mbaye et Krief 2009). Ce travail a été intégré, par la suite, dans les architectures des réseaux autonomes. Par autonomie, on entend la capacité des processus de gestion et de l'infrastructure sous-jacente à se déployer, s'organiser et opérer sans aide extérieure. Le rôle de l'administrateur se limite au guidage de ces processus, en leur fixant des objectifs de haut niveau (Krief 2010). Le plan de connaissance joue un rôle essentiel dans les réseaux autonomes en permettant à chaque entité autonome de fermer la boucle de contrôle réseau sans l'aide de l'administrateur. Les tâches de gestion peuvent ainsi être effectuées par le réseau lui-même, chaque entité autonome étant capable de s'autoconfigurer, s'auto-optimiser, s'autoprotéger et s'autoréparer. Z. Movahedi *et al.* (2012) ont comparé différentes architectures de réseaux autonomes et ont souligné l'intérêt de l'apprentissage pour permettre une adaptation intelligente et converger vers un fonctionnement optimal du réseau.

Dans le contexte des SDN, la séparation claire des plans de contrôle et de données et la centralisation de l'intelligence peut influencer positivement le développement du plan de connaissance avec de nouvelle fonctionnalités.

2.4.2. *Réseaux KDN (*Knowlegde-Defined Networking*)*

Mestres *et al.* (2017) ont étudié les raisons du manque d'adoption des techniques d'intelligence artificielle en pratique et ont affirmé que la montée en puissance de deux paradigmes récents, à savoir le réseau défini par logiciel (SDN) et l'analyse de réseau pour *Network Analysis* (NA) faciliteront l'adoption de ces techniques pour le contrôle et l'exploitation des réseaux. Ils ont ainsi proposé un nouveau paradigme, appelé « réseau défini par la connaissance » (KDN pour *Knowlegde-Defined Networking*), qui associe SDN, NA et apprentissage automatique pour fournir un contrôle de réseau

automatisé. Dans (Mestres *et al.* 2018), ils ont exploré la possibilité d'appliquer différents modèles et techniques de *Machine Learning* (ML) pour modéliser des éléments de réseaux complexes, tels que les fonctions de réseau virtuel (VNF pour *Virtual Network Function*). Ils ont ainsi démontré que le comportement de différentes fonctions VNF pouvait être appris à l'aide de techniques de *Maching Learning* (ML), telles que leur consommation de CPU en fonction du trafic entrant.

Le projet ALLIANCE (Careglio *et al.* 2018) s'appuie sur le nouveau paradigme de KDN. Il a pour objectif de concevoir et mettre en place une infrastructure réseau 5G capable de prendre en charge des services omniprésents, tout en répondant aux performances et aux exigences commerciales de multiples parties prenantes. Le résultat attendu de ce projet est de proposer, comparer et interconnecter trois prototypes différents d'architecture complète de réseau (allant du réseau d'accès jusqu'au réseau cœur), autonomes et orientés 5G. Plusieurs solutions réseau seront étudiées, telles que SDN/NFV, les réseaux *overlays* programmables et l'architecture récursive interréseaux RINA. Chacune des solutions retenues s'appuiera sur un orchestrateur KDN, qui tirera parti des techniques d'apprentissage automatique pour déployer, exploiter, surveiller et dépanner les réseaux automatiquement.

Dans (Hyun et Hong 2017), les auteurs présentent une architecture réseau autonome combinant télémétrie réseau, KDN, SDN et P4 INT. P4 INT (INT pour *In-band Network Telemetry*) permet de collecter la télémétrie réseau. Ces données seront ensuite utilisées par KDN pour apporter de l'intelligence à la gestion du réseau. SDN sera ensuite chargé de gérer et contrôler le réseau en fonction de la décision prise par le plan de connaissance. Dans (Hyun *et al.* 2018), les auteurs présentent une première mise en œuvre du système de surveillance réseau.

Dans (Lu *et al.* 2019), les auteurs s'intéressent à l'orchestration au niveau KDN, dans le but de faire fonctionner un DCN hybride optique/électronique de manière performante et économe en énergie. Cette orchestration s'appuie sur trois modules d'intelligence artificielle basés sur l'apprentissage en profondeur (*deep learning*), à savoir le module de prédiction de trafic, le module de prédiction de la demande de machine virtuelle et le module de reconfiguration du réseau. Les résultats expérimentaux montrent que cette approche permet d'améliorer non seulement les performances de la fourniture de services, mais également l'efficacité énergétique du système.

2.4.3. *Réseaux IDN (*Intelligence-Defined Networks*)*

Les approches de type IDN (*Intelligence-Defined Networks*) se présentent comme une évolution de SDN en ajoutant une couche cognitive au-dessus de la couche contrôle. Cela permet ainsi d'introduire de l'intelligence artificielle dans la gestion et le contrôle

du réseau. Il s'agit donc d'une approche assez similaire au KDN mais portée par des industriels, et en particulier par Huawei (Jiang 2016 ; Huawei 2018).

Ainsi, ce dernier a proposé au Mobile World Congress 2018 une nouvelle offre appelée *Intent-Driven Network* (IDN), visant à faire évoluer les réseaux SDN vers des réseaux « fondés sur l'intention », grâce à l'ajout de technologies d'IA, de Big Data et du Cloud. L'accent est mis sur la capacité du plan de contrôle à connaître – voire à prédire – le type d'application/service qu'un client souhaitera utiliser et le contexte dans lequel il désirera le faire, pour lui réserver automatiquement et de bout en bout la bande passante dont il a besoin, afin de lui procurer la meilleure expérience utilisateur. Le réseau, à présent centré sur l'utilisateur, est ainsi capable d'identifier avec précision l'intention de celui-ci et de la réaliser. Il est capable également de détecter en temps réel la qualité de l'expérience utilisateur et de réaliser une analyse prédictive, afin d'optimiser les performances de manière proactive (Huang *et al.* 2018).

2.5. L'apport de l'IA pour la sécurité

Dans cette section, nous commençons par présenter les techniques d'apprentissage (*Machine Learning*) les plus utilisées pour la sécurité, puis nous prendrons comme exemple les systèmes de détection d'intrusion pour illustrer l'apport de l'IA dans le domaine de la sécurité des réseaux SDN.

2.5.1. *Techniques d'apprentissage machine*

Le *Machine Learning* (ML) est un sous-domaine de l'IA (intelligence artificielle) ; il peut se subdiviser en 4 groupes :

1) l'apprentissage supervisé où, en disposant d'un ensemble d'objets et pour chaque objet d'une valeur cible associée, le système doit apprendre un modèle capable de prédire la bonne valeur cible d'un objet nouveau ;

2) l'apprentissage non supervisé où, en disposant d'un ensemble d'objets sans aucune valeur cible associée, le système doit apprendre un modèle capable d'extraire les régularités au sein des objets pour comprendre la structure des données ;

3) l'apprentissage semi-supervisé où, en disposant d'un petit ensemble d'objets avec pour chacun une valeur cible associée et d'un plus grand ensemble d'objets sans valeur cible, le système doit être capable de résoudre des problèmes supervisés et/ou non supervisés ;

4) l'apprentissage par renforcement où, en disposant d'un ensemble de séquences de décisions dans un environnement dynamique et pour chaque action d'une séquence d'une valeur de récompense, le système doit apprendre un modèle capable de prédire la meilleure décision.

Les travaux dans le domaine de la sécurité utilisent principalement les techniques d'apprentissage supervisé et non supervisé. Ces techniques seront présentées dans les sections suivantes.

2.5.1.1. *Techniques d'apprentissage supervisé*

L'apprentissage supervisé est composé d'un ensemble de techniques qui sont bien établies et utilisées dans différents domaines.

Les plus utilisées dans le domaine de la recherche en sécurité sont :

1) l'arbre de décision : c'est un outil d'aide à la décision représentant un ensemble de choix sous la forme d'un arbre. L'arbre contient des nœuds et des branches. Les nœuds sont étiquetés par des attributs et les branches sont étiquetées par un prédicat qui s'applique au nœud parent. Les arbres de décision réussissent très bien à résoudre les problèmes de classification. L'arbre de décision peut décrire un ensemble de données par une structure arborescente (Negnevitsky 2005). Les données d'entrée et de sortie peuvent être discrètes ou continues. Les arbres de décision peuvent représenter toutes les fonctions booléennes. Un arbre de décision effectue une séquence de tests, chaque nœud interne de l'arbre correspondant à un test de l'un des attributs d'entrée (Russell et Norvig 1995). L'apprentissage consiste à construire un arbre récursivement en choisissant l'attribut qui affecte les exemples dans leurs propres classes, aux nœuds enfants pour chaque valeur de l'attribut choisi. L'arbre de décision est un outil utilisé dans plusieurs domaines, tels que la sécurité et la fouille de données ;

2) le réseau bayésien : le réseau bayésien est un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphe. Ce graphe représente les relations, qui ne sont pas déterministes mais probabilistes, causales, entre les variables. Les réseaux bayésiens consistent à tenir compte simultanément de connaissances *a priori* (dans le graphe) et de l'expérience contenue dans les données ;

3) les algorithmes génétiques : c'est l'une des techniques les plus connues et les plus utilisées (Holland 1975). Ici, la solution d'un problème d'optimisation est représentée par un chromosome. Un ensemble de chromosomes forme la population.

Les deux opérations fondamentales mais très importantes de cette technique sont : le croisement et la mutation. L'opération de croisement combine des individus précédemment sélectionnés en échangeant certaines de leurs pièces. La mutation, en revanche, introduit un peu d'aléas dans la recherche pour éviter le problème des *optima* locaux. Les facteurs importants pour la mise en œuvre de tout algorithme génétique sont : la stratégie de sélection et le type d'opérateurs de croisement et de mutation (Boussaid *et al.* 2013) ;

4) la machine à vecteurs de support : les machines à vecteurs de support (en anglais *Support Vector Machine*, SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM ont été développées dans les années 1990 et ont été adoptées pour leur capacité à travailler avec des données de grandes dimensions. Les SVM se basent sur des séparateurs linéaires qui maximisent la marge entre deux classes différentes afin de fournir une meilleure classification. Les méthodes d'apprentissage peuvent être utilisées pour transformer les données d'entrée en un espace de grande dimension afin de traiter des cas linéairement non séparables (Nguyen et Armitage 2008) ;

5) les réseaux de neurones : les réseaux de neurones artificiels sont principalement inspirés des neurones biologiques du cerveau humain (Negnevitsky 2005). Les réseaux neuronaux sont très utilisés grâce à leur capacité à traiter de grandes quantités d'information et à leur stabilité face au bruit. Les réseaux neuronaux sont un ensemble de neurones (petits processeurs) travaillant en parallèle. Ces neurones reçoivent des données, appliquent une fonction dite « fonction d'activation » et les renvoient à d'autres neurones ou à une source externe. Les réseaux de neurones présentent de nombreux avantages. Tout d'abord, ils peuvent s'ajuster aux données sans spécifier explicitement de fonction ou de distribution pour représenter le modèle sous-jacent (Zhang 2000). Deuxièmement, les réseaux de neurones forment un approximateur fonctionnel universel, qui peut approcher n'importe quelle fonction (Zhang 2000). Troisièmement, les réseaux de neurones sont des modèles non linéaires, ce qui leur donne la souplesse nécessaire pour représenter et modéliser des relations complexes (Zhang 2000). Les réseaux multicouches (MLP pour *MultiLayer Perceptron*) sont les systèmes de classification de réseau neuronal les plus couramment utilisés. Les MLP sont principalement formés avec des algorithmes d'entraînement supervisés. Les réseaux de neurones sont sujets à des surajustements lorsque nous utilisons trop de paramètres dans le modèle (Russell et Norvig 1995) ;

6) les *Random Forest* (arbres aléatoires) : cette technique appartient à la famille des agrégations de modèles. C'est un cas particulier de *bagging* (*Bootstrap*

Aggregation) appliqué aux arbres de décision. Le principe est de faire la moyenne des prévisions de plusieurs modèles indépendants pour réduire la variance, et donc l'erreur de prévision. Pour construire ces différents modèles, on sélectionne plusieurs échantillons *bootstrap*, c'est-à-dire des tirages avec remises. Les forêts aléatoires ajoutent de l'aléa au niveau des variables en plus du principe de *bagging*. Pour chaque arbre, il faut sélectionner un échantillon *bootstrap* d'individus et la construction d'un nœud de l'arbre se fait sur un sous-ensemble de variables tirées aléatoirement ;

7) K-NN (*K-Nearest Neighbors*) : c'est un algorithme de classification. L'idée est de faire voter les plus proches voisins d'une observation x. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x. La détermination du plus proche voisin est basée sur une fonction distance.

2.5.1.2. *Techniques d'apprentissage non supervisé*

L'approche supervisée est aussi composée d'un ensemble d'algorithmes bien établis :

1) *K-means* est l'une des approches de *clustering* les plus connues. Sa mise en œuvre nécessite la connaissance préalable du paramètre k, qui indique le nombre de *clusters* résultants. Chaque point de données sera assigné au centroïde le plus proche de chaque groupe. K-means minimise une fonction d'objectif, qui représente la distance entre les points de données et leurs centroïdes correspondants (Khan et Ahmad 2004). Le processus de mise à jour des centroïdes, en fonction des points de données qui leur sont attribués, sera répété jusqu'à ce que les centroïdes restent les mêmes ou qu'aucun point ne change. K-means dépend principalement de l'ensemble initial de classes. Par conséquent, un choix inapproprié de k peut entraîner des résultats médiocres (Zhang et Xia 2009). De plus, la classification fuzzy-C-Means (FCM) (Pal et Bezdek 1995), également appelée « soft K-means », permet à chaque point de données d'appartenir à plusieurs classes. En d'autres termes, un point de données peut appartenir à toutes les classes avec un degré d'adhésion différent ;

2) SOM : utilisé en *DoS attack*, l'algorithme SOM (*Self-Organizing Map*) est connu aussi sous le nom d'algorithme de Kohonen. C'est un algorithme stochastique de classification qui intègre une notion de voisinage entre les classes.

Ces différentes techniques d'apprentissage sont résumées dans le tableau 2.3.

Algorithme ML	Problème	Avantages	Inconvénients
Arbre de décision	Classification Régression	Facile à implémenter Décisions simples à comprendre Classification des données sans trop de calcul Traitement des données continues et discrètes	Coûteux en calcul et en espace mémoire Instabilité : un petit changement dans le jeu de données peut entraîner des modifications importantes dans le modèle
Réseau bayésien	Classification	Facile à implémenter Phase d'apprentissage rapide	Difficulté à gérer des données continues
Algorithmes génétiques	Classification	Conclure de bonnes règles de classification	Coût de calcul
Machine à vecteurs de support	Classification Régression	Traitement des données de grande dimension	Moins efficace sur un jeu de données bruité Coûteux en calcul
Réseaux de neurones	Classification Régression	La phase de prédiction est rapide Travaille bien sur un grand volume de données d'entraînement	Coût de calcul Difficulté d'interprétation du modèle d'apprentissage par un humain
Random Forest	Classification Régression	Travaille bien sur des données d'entraînement de grande dimension Réduit l'instabilité du modèle comparé aux arbres de décision	Faible vitesse d'entraînement
K- NN	Classification Régression	Facile à implémenter	Coûteux en calcul et en espace mémoire
K-means	*Clustering*	Facile à implémenter Résultats faciles à interpréter	Coût de calcul linéaire avec la taille des données d'entraînement
SOM	*Clustering*	Manipulation de données très volumineuses	Coûteux en calcul

Tableau 2.3. *Techniques d'apprentissage les plus utilisées pour la sécurité*

2.5.2. *L'apport de l'IA pour un service de sécurité : la détection d'intrusion*

Les techniques d'IA citées plus haut ont été utilisées pour résoudre beaucoup de problèmes de sécurité en général, et plus particulièrement pour la détection d'intrusion (Soheily-Khah *et al.* 2018).

Dans (Barapatre *et al*. 2008), une approche a été proposée avec un réseau de neurones MLP-BP (MLP pour *Multilayer Perceptron* et BP pour *BackPropagation*). Le système proposé a en entrée les caractéristiques de l'ensemble de données KDD (*Knowlege Discovery from Database*)[6] et en sortie la classification des paquets normaux et des paquets suspects présents dans le jeu de données. Il a été démontré que le réseau de neurones MLP-BP détectait les attaques DoS et « Probe » avec plus de précision que les attaques U2R (*User to Root*). Dans (Lu *et al*. 2015), les auteurs utilisent les réseaux de neurones *Radial Basis Function* (RBF), qui sont très pratiques dans les systèmes de détection d'intrusions. Ils ont comparé RBF et MLP-BP en utilisant un jeu de données KDD traité en convertissant toutes les chaînes en chiffres, réduisant ainsi la dimension du jeu de données. Les résultats de la simulation ont montré que le réseau de neurones RBF est meilleur que MLP-BP en termes de temps d'entraînement, de précision et de détection des attaques.

Les travaux présentés dans (Canbay et Sagiroglu 2015) proposent d'utiliser une approche hybride pour détecter les attaques. Les méthodes de l'algorithme génétique (AG) et des *k* plus proches voisins (KNN) ont été combinées pour modéliser et détecter les attaques. KNN a été utilisé pour classifier les attaques et AG pour sélectionner *k* voisins d'un échantillon d'attaques. Ce système hybride a été appliqué pour la première fois dans le domaine de la détection d'intrusions. Les résultats ont montré que le système proposé fournit de meilleurs résultats qu'un système unique en termes de précision de détection.

La plupart des chercheurs ont utilisé l'ensemble de données KDD, qui a été largement critiqué pour ne pas représenter fidèlement le réseau. Dans (Sahu et Mehtre 2015), les auteurs ont utilisé un nouvel ensemble de données réseau étiqueté, appelé « ensemble de données Kyoto 2006+ ». Dans Kyoto 2006+, chaque instant est étiqueté comme « normal » (pas d'attaque), « attaque » (attaque connue) et « attaque inconnue ». Les utilisateurs ont utilisé l'algorithme *Decision Tree* (J48) pour classifier le paquet réseau pouvant être utilisé pour un NIDS (*Network IDS*). Les résultats ont montré que l'arbre de décision présente une très bonne précision de classement et qu'il permet également de classer les attaques inconnues.

6. Disponible à l'adresse : http://nsl.cs.unb.ca/NSL-KDD/.

Ces quelques échantillons de travaux montrent à quel point le potentiel de l'IA n'a pas encore été épuisé dans le domaine de la sécurité, car la combinaison de technique peut permettre d'avoir de meilleurs résultats.

2.6. L'apport de l'IA pour la sécurité dans les réseaux SDN

Nous prenons dans cette section les systèmes de détection d'intrusion comme exemples pour illustrer le potentiel de l'IA dans le domaine de la sécurité des réseaux SDN.

Les administrateurs réseau implémentent des IDS pour éviter les attaques par intrusion et appliquer la politique de sécurité du réseau. Un IDS surveille le trafic et envoie des alertes d'intrusion sur la console de l'administrateur lorsqu'un message suspect est détecté.

Dans (Tang *et al.* 2016), les auteurs proposent une approche d'apprentissage profond (*Deep Learning*) pour la détection d'anomalies dans un environnement SDN. Ils ont construit un modèle de réseau de neurones profonds (DNN) pour un système de détection d'intrusions. Ils ont prouvé que l'apprentissage en profondeur a un fort potentiel d'utilisation pour la détection d'anomalies dans des environnements SDN. Dans (Tang *et al.* 2018), les auteurs améliorent l'approche présentée dans (Tang *et al.* 2016). Ils proposent un système de détection d'intrusions utilisant le réseau de neurones récurrents (GRUN-RNN pour *Gated Recurrent Unit Recurrent Neural Network*) pour les réseaux SDN. L'approche proposée est testée à l'aide du jeu de données NSL-KDD[7]. Les résultats d'expérimentation montrent que l'approche proposée présente également un fort potentiel de détection d'intrusions dans les environnements SDN. Thaseen et Kumar (2013) visent à évaluer différents algorithmes de classification basés sur des arbres qui classent les événements du réseau dans les systèmes de détection d'intrusions. Les expériences sont menées sur le jeu de données NSL-KDD. Les résultats montrent que le modèle *RandomTree* détient la meilleure précision en minimisant les faux positifs.

Dans (Chen et Yu 2016), les auteurs proposent une nouvelle architecture collaborative de prévention des intrusions (CIPA pour *Collaborative Intrusion Prevention Architecture*). CIPA est déployé en tant que réseau virtuel d'un réseau de neurones sur le sous-réseau. Tirant parti de la manipulation mathématique parallèle et simple des neurones dans un réseau de neurones, CIPA peut disperser sa puissance de calcul légère vers les commutateurs programmables du sous-réseau. Chaque commutateur programmable virtualise un à plusieurs neurones. L'ensemble du réseau neuronal fonctionne comme un IDS/IPS intégré. Cela permet à CIPA de

7. Disponible à l'adresse : http://nsl.cs.unb.ca/NSL-KDD/.

détecter les attaques distribuées dans une vue globale, ce qui ne nécessite pas de temps de communication et de calcul élevé.

L'attaque par déni de service distribué (DDoS) est l'un des problèmes les plus récurrents en matière de sécurité réseau. Récemment, bien que les mécanismes d'attaque DDoS soient largement compris, les problèmes deviennent de plus en plus fréquents en raison de la similitude entre une attaque DDoS et le trafic normal. Dans (Nam *et al.* 2018), les auteurs proposent des approches de détection d'attaques DDoS. Les algorithmes proposés dans l'architecture de détection sont implémentés dans un environnement SDN. Le contrôleur SDN permet de compiler rapidement un algorithme complexe de classification et de détection. Les résultats expérimentaux montrent que ces algorithmes présentent un temps de traitement relativement faible, tout en maintenant une bonne précision.

Dans (Mihai-Gabriel et Victor-Valeriu 2014), les auteurs présentent un moyen d'atténuer les attaques DDoS dans un environnement SDN en évaluant les risques par le biais d'un système de cyberdéfense basé sur les réseaux de neurones.

Dans (Niyaz *et al.* 2016), les auteurs proposent un système de détection DDoS multivecteur basé sur l'apprentissage en profondeur dans un environnement SDN. Ils ont implémenté le système en tant qu'application réseau au-dessus d'un contrôleur SDN. Le *deep learning* a été utilisé pour réduire un grand nombre de fonctionnalités dérivées des en-têtes de réseau trafic. Les auteurs ont obtenu, grâce aux traces du trafic collectées à partir de différents scénarios, une grande précision avec très peu de faux positifs pour la détection d'attaque.

Braga *et al.* (2010) présentent une méthode légère de détection d'attaque DDoS basée sur les caractéristiques de flux de trafic, dans laquelle l'extraction de telles informations est réalisée avec un temps relativement court comparé aux approches traditionnelles. Cela est possible grâce à l'utilisation de la plate-forme NOX, qui fournit une interface de programmation facilitant le traitement des informations de commutation. L'approche présente une bonne performance de détection et un taux de faux positifs très faible.

Dans (Mehdi *et al.* 2011), les auteurs développent des algorithmes de détection d'anomalies de trafics importants qui peuvent être implémentés dans un contexte SDN en utilisant des commutateurs compatibles Openflow et NOX en tant que contrôleur. Ils prouvent que ces algorithmes sont précis dans l'identification des activités malveillantes sur les réseaux domestiques par rapport aux ISP (*Internet Service Provider*). De plus, l'analyse de l'efficacité des implémentations SDN sur un routeur de réseau domestique programmable indique que les détecteurs d'anomalies peuvent fonctionner sans introduire de pénalité de performances pour le trafic du réseau domestique.

2.7. Le déploiement d'un service de protection contre les intrusions

Dans le contexte des IDS basés sur les signatures, lorsqu'une attaque réseau se produit, une des tâches les plus difficiles pour l'administrateur ou l'expert réseaux consiste à écrire une signature précise décrivant cette nouvelle attaque.

Dans cette section, nous allons présenter les travaux de (Hamdi *et al.* 2015), qui utilisent une autre technique de l'IA : la programmation logique inductive (Muggleton 1991) pour la génération automatique de signatures.

Ensuite, nous allons montrer une proposition d'évolution de ces travaux, afin de permettre le déploiement « intelligent » d'un tel service de sécurité pour les réseaux SDN.

2.7.1. *Service d'apprentissage de signatures d'attaques comme service du* Cloud

La proposition dans (Hamdi *et al.* 2015) est de fournir l'apprentissage de signatures d'attaques comme un service dans le Cloud. Le service est déployé dans le Cloud avec une architecture client/serveur. L'accès des clients au service se fait *via* un réseau privé virtuel sécurisé.

Le service d'apprentissage est fourni par des nœuds d'apprentissage (LN pour *Learning Node*) contenant des modules ILP (programmation logique inductive). En entrée, les LN reçoivent des trafics malveillants et normaux aux formats standards (par exemple, PCAP). Le système renvoie une règle de Prolog (signature) qui peut être traduite en toute grammaire cible spécifique.

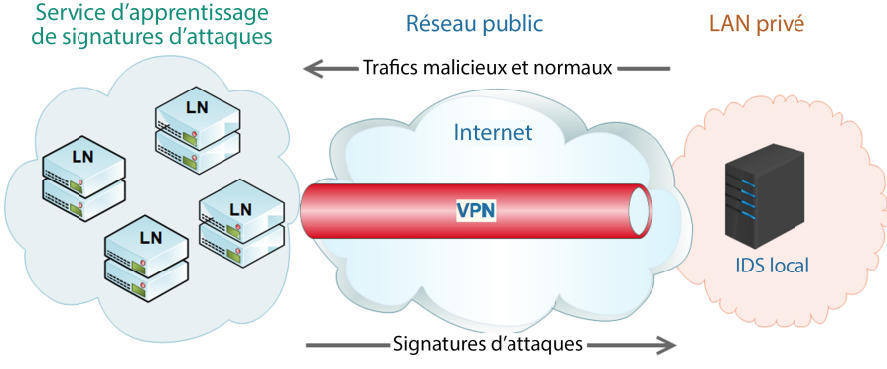


Figure 2.8. *Architecture du service d'apprentissage de signatures*

À l'intérieur d'un LN, il y a un moteur ILP, une base de faits et un GTP (*Grammar Translator Proxy*). Ce proxy est un composant de médiation de données entre le moteur ILP et le client. Il traduit les données d'entrée provenant du client en prédicats correspondant aux paramètres du moteur ILP :

– trafic malicieux → exemples positifs (E^+) ;

– trafic normal → exemples négatifs (E^-).

Le moteur de programmation logique inductive apprend des règles correspondant aux signatures des attaques qui lui sont présentées. Ce moteur a besoin d'une base de faits composée de connaissances de base (BK pour *Background Knowledge*) et d'exemples qui correspondent aux trafics classés comme malveillants.

Une fois que le moteur ILP a appris les règles, GTP transforme ces règles de la grammaire Prolog vers la grammaire spécifique à l'IDS ciblé.

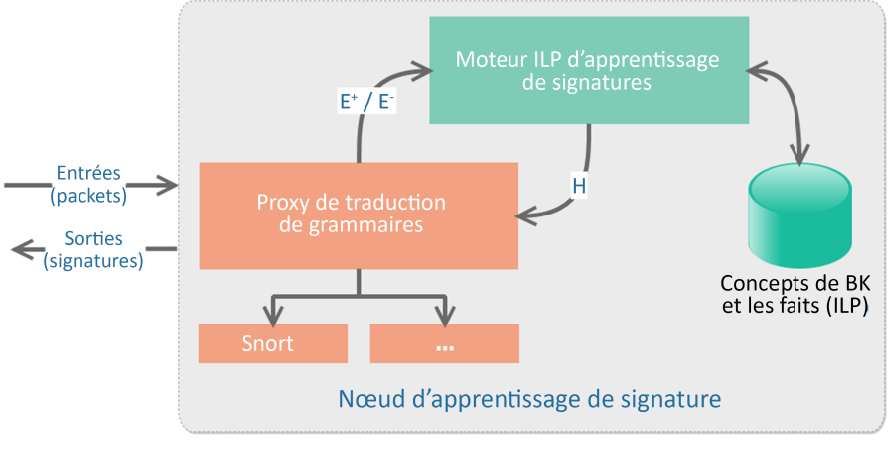


Figure 2.9. *Architecture d'un LN (Learning Node)*

Pour le moment, seule la grammaire des règles de l'IDS SNORT est prise en charge par la proposition. Soit la signature apprise $C \models P_1 \ldots P_n$, où C est le concept se trouvant à la tête de la clause et $P_1 \ldots P_n$ sont des prédicats de BK. La transformation de la règle de signature SNORT s'effectue en deux étapes.

Étape 1 : construire l'en-tête de règles à partir de la signature Prolog en extrayant les adresses IP source et destination, les numéros de port dans la signature apprise. Les prédicats de BK qui peuvent être dans le corps de la règle de signature sont : `Pi=ip_src(packet)`, `Pj=ip_dst(packet)`, `Pk=src_port(packet)`,

`dst_port(packet)`, P_1=`proto(X,tcp)`. Le tableau 2.4 donne une idée de la transformation effectuée.

Signature (Prolog)	Signature SNORT
ip_src(192.168.0.1) ; ip_dst(10.0.0.1),proto(P,tcp) ; dst_port(X),src_port(Y)	alert tcp 192.168.0.1:ANY > 10.0.0.1:ANY
ip_src(X),dst_ip(10.0.0.1), proto(P,icmp)	alert icmp ANY > 10.0.0.1

Tableau 2.4. *Construction de l'en-tête de règles*

Étape 2 : construire le corps des règles. Dans le corps SNORT, C est transformé en (`msg: some_text_id`). Les prédicats du corps $P_1 \ldots P_n$ sont transformés en filtre SNORT le plus proche F_1, F_2, F_3, etc. Par exemple, `syn_tcp_actif` (paquet) pourrait être traduit dans le flux du filtre SNORT : (…, Flow:Etablished,…). Le corps final de la règle SNORT est : (F1, F2,…). Enfin, la signature est envoyée au client pour alimenter sa base de signatures.

Aleph (*A Learning Engine for Proposing Hypotheses*) (Srinivasan 2000) a été utilisé comme moteur ILP pour réaliser cette solution. Il a été montré que cette approche peut être très efficace sur des attaques de type DoS. Le principal avantage de cette solution est que la tâche complexe d'écriture de la signature à partir des journaux du réseau peut maintenant être faite par le service de nuage.

Cette approche présente un grand intérêt avec la conception des réseaux SDN qui intègre déjà le principe de la délocalisation de l'intelligence des réseaux dans un contrôleur logiquement centralisé. La section suivante montre les propositions sur le plan architectural pour permettre le déploiement de tels services dans les réseaux SDN.

2.7.2. *Déploiement d'un service de protection contre les intrusions dans les réseaux SDN*

Un des premiers défis, afin de déployer une architecture IDS qui apprend à la volée les signatures et les déploient, est le temps de réponse acceptable. Classiquement, les IDS ont des systèmes de langages de filtres beaucoup plus riches que l'API-Sud dans les réseaux SDN. Il devient donc un défi de mettre en place un système de prévention d'intrusions avec un temps de réponse acceptable.

L'architecture proposée a pour objectif de fournir un système IPS (système de prévention d'intrusions) autogéré pour les réseaux SDN. L'IPS permet d'étendre les fonctions de sécurité d'un contrôleur SDN à travers une interface P4 (Bosshart *et al.* 2014) et grâce à l'intelligence artificielle. En effet, le contrôleur est chargé de transférer des règles de retransmission ou ACL (*Access Lists*) aux switchs supportant OpenFLow à l'intérieur du réseau. Cependant, il n'est pas conçu pour analyser et détecter les attaques à l'intérieur du réseau. Les IDS et les IPS sont les outils dédiés à cette tâche. L'implémentation de ces outils est un vrai défi au vu de la grande quantité de données qui circule dans le domaine SDN et des contraintes de performance liées à la virtualisation. Cette architecture aborde ce défi à l'aide de trois blocs fonctionnels : un système de détection et prévention d'intrusions à base de signatures, un système pour l'apprentissage à la volée des signatures d'attaques basé sur la programmation logique inductive et le *Deep Learning*, et une interface P4 permettant le déploiement de nouvelles règles de détection dans le réseau.

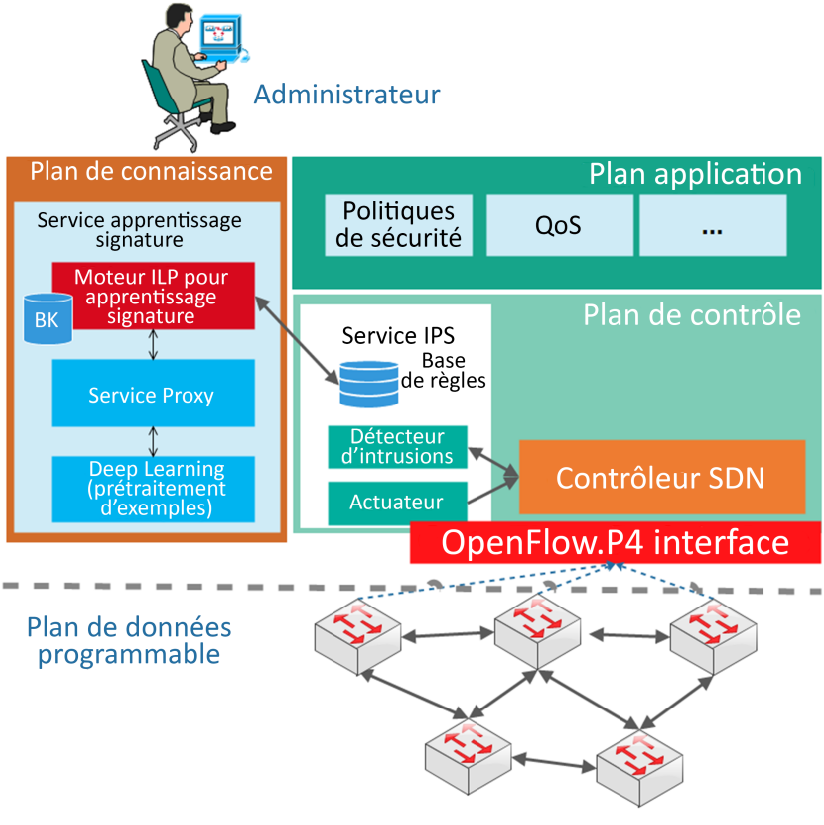


Figure 2.10. *Architecture d'un réseau SDN intelligent avec IDS*

Le système (figure 2.10) fonctionne comme une chaîne de traitement qui s'auto-optimise avec l'intervention occasionnelle de l'administrateur en tant qu'oracle lorsque c'est nécessaire. Le plan de données joue un rôle classique dans cette architecture ; en revanche, le plan de contrôle a une activité plus intelligente à travers le plan de connaissance.

Le plan de connaissance au niveau des réseaux SDN gère toute la connaissance afin d'améliorer/optimiser le fonctionnement du contrôleur pour assurer les fonctions de sécurité (optimisation de politique, prévention d'intrusion, détection de fonctionnement malicieux d'un switch, etc.), de routage (optimisation des routes, autoréparation des routes disparues, prédiction de la répartition de la charge, etc.) et de QoS.

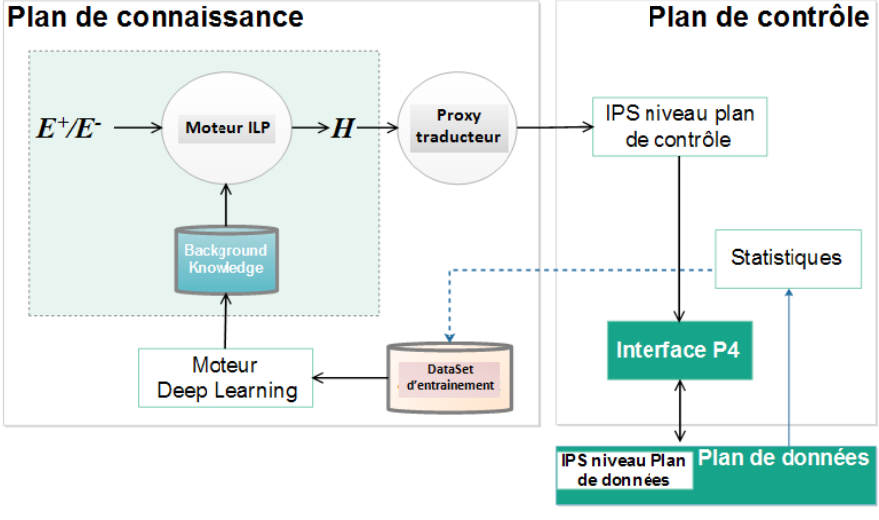


Figure 2.11. *Interactions entre les composants de l'architecture des données*

Le plan de connaissance permet d'avoir une boucle locale intelligente pour la gestion d'un service IPS à base de signatures. La boucle est décrite sur la figure 2.11. Sur cette boucle, il y a deux outils d'apprentissage qui ont des rôles liés :

– l'outil d'apprentissage principal est basé sur la programmation logique inductive. L'ILP permet d'apprendre une hypothèse à partir des exemples de paquets malicieux comme indiqué dans (Hamdi *et al.* 2015) ;

– le module de *Deep Learning* qui, à partir des exemples d'un *dataset* de référence comme NSL-KDD, construit un *background knowledge* en permettant de ne prendre que les concepts qui sont pertinents pour apprendre une attaque :

$$\Delta(D_{set}, E) = \{p \in P \,/\, \Pi(p, D_{set}) > 0\} \quad [2.1]$$

où Δ est la fonction de *Deep Learning* qui prend en paramètre un *dataset* (comme NLS_KDD) et des exemples positifs de l'attaque. En sortie, cette fonction doit donner l'ensemble des p propriétés pertinentes pour le BK. La pertinence est représentée par la fonction $\Pi(p, D_{set}) > 0$. Cette formule exprime le fait que le système va utiliser une base d'entraînement pour déterminer les champs des paramètres qui sont potentiellement pertinents pour figurer dans la signature de l'attaque. En effet, le *background knowledge* du module ILP contient en général les prédicats qui sont utilisables pour décrire l'attaque. Le choix du contenu BK est crucial pour avoir les règles les plus précises possibles. Pour ajouter un biais dans la conception de BK, nous utilisons le *Deep Learning* afin de déterminer pour chaque attaque les propriétés qui sont les plus susceptibles de figurer dans BK. Le module de *Deep Learning* va utiliser les bases de traces d'entraînement, pour notre cas NSL-KDD, afin d'optimiser le contenu de BK. Cette optimisation consiste principalement à réduire la dimension qui permet de filtrer en amont les paramètres qui seront les plus pertinents pour décrire une attaque, et donc réduire l'espace de recherche du moteur ILP.

Une interface P4 permet de transformer les filtres et les actions du système de détection d'intrusion en programmes déployables au niveau des *switchs* programmables.

Le service de détection d'intrusions présenté dans cette section est déployé à l'aide d'un unique contrôleur SDN. Pour permettre le passage à l'échelle, nous préconisons l'existence de contrôleurs distribués physiquement mais logiquement centralisés, ainsi qu'un modèle plat, solution qui semble faire consensus aujourd'hui. Pour les échanges entre contrôleurs afin de maintenir la vue globale, et en l'absence d'API-Est/Ouest standardisé, l'approche Pub/Sub nous semble également à privilégier, car elle permet de tenir un contrôleur informé uniquement lors d'un changement d'état.

2.8. Enjeux

Les enjeux de l'utilisation des outils de l'IA et de l'apprentissage automatique pour la gestion de la sécurité dans les réseaux SDN sont très importants. Ils le sont non seulement sur le plan technique mais aussi sur le plan sociétal. En effet, aujourd'hui la sécurité informatique a un impact sur la vie de tous les jours, surtout avec l'avènement de la 5G. Cette technlogie aura un impact significatif sur plusieurs domaines tels que la médecine, la santé, la sécurité routière, etc.

Tout d'abord, avec l'avènement du Big Data et de la Big Analytics, les dispositifs qui permettent d'étudier les vulnérabilités des systèmes sont de plus en plus efficaces (Terzi *et al.* 2017 ; Hajizadeh *et al.* 2018 ; Hoon *et al.* 2018). Si de tels dispositifs se mettent à découvrir des vulnérabilités de sécurité qui sont intrinsèques à la conception des contrôleurs SDN, et qu'il n'est pas possible de se protéger, les conséquences peuvent être l'abandon pur et simple de cette technologie. Il est donc important que les contrôleurs SDN aient une intelligence et une autonomie leur permettant de se reprogrammer et de reprogrammer le plan de données de manière adaptée.

D'un autre côté, les systèmes d'attaques intelligents pourraient faire apparaître des vulnérabilités à la limite de la normalité pour les systèmes de détection d'intrusion à base d'anomalies traditionnelless de telle sorte à générer beaucoup de faux positifs et négatifs. Or, un grand nombre de faux positifs pourrait tenter les administrateurs à relâcher un peu la sécurité ou à intégrer des exceptions pour de vraies attaques. En effet, face à des attaques intelligentes, le moyen de défense le plus adapté est une autre intelligence artificielle capable d'aider les administrateurs à prendre de bonnes décisions.

L'arrivée de la 5G met encore plus l'accent sur l'usage de l'IA dans les outils de sécurité. En effet, la grande masse de données à analyser ne peut pas l'être avec les technologies et outils classiques. D'un côté les outils de sécurité pourraient ne pas fournir une réponse dans un temps acceptable : une attaque qui réussit le temps que le système de détection fournisse une réponse. De l'autre, ces outils pourraient créer des goulots d'étranglement si le système fonctionne en mode paranoïaque en regardant tous les paquets. Les outils de l'IA/ML pourraient permettre de réduire l'analyse aux trafics qui sont pertinents à analyser. Si cette quantité de données pouvait avoir un comportement linéaire par rapport à l'augmentation de trafic, les services de sécurité pourraient être déployés tout en garantissant le passage à l'échelle.

2.9. Conclusion

Le concept de réseau SDN a été créé pour faire face à des défis liés à l'évolution des réseaux. Les principes de base des réseaux SDN sont : l'introduction de la programmabilité des équipements de transmission *via* des API-Sud standard et ouvertes (*SouthBound Interface*) ; le découplage du plan de contrôle et de celui des données et la vue globale du réseau par une centralisation logique de « l'intelligence du réseau ».

Avec cette nouvelle approche, les éléments du plan de données sont programmables et très efficaces dans la transmission des paquets, car ils effectuent des

tâches moins complexes. Les composants du plan de contrôle sont implantés sur un contrôleur qui concentre les activités de pilotage du réseau grâce à sa vue globale.

Toutefois, la question de la sécurité est un enjeu important pour le développement des réseaux SDN. En effet, le point de faiblesse que représentent le contrôleur et la sécurisation d'équipements réseaux programmables dans un environnement qui produit une grande quantité de données sont les principaux défis à relever.

2.10. Bibliographie

Abubakar, A., Pranggono, B. (2018). Machine learning based intrusion detection system for software defined networks. Dans *7th International Conference on Emerging Security Technologies (EST)*, 138–143.

Bannour, F., Souihi, S., Mellouk, A. (2018). Distributed SDN Control: Survey, Taxonomy, and Challenges. *IEEE Communications Surveys & Tutorials*, 20(1).

Barapatre, P., Tarapore, N.Z., Pukale, S.G., Dhore, M.L. (2008). Training MLP neural network to reduce false alerts in IDS. Dans *International Conference on Communication and Networking*, 1–7.

Bi, J., Zhang, K., Cheng, X. (2009). Intrusion detection based on RBF neural network. Dans *International Symposium on Information Engineering and Electronic Commerce*, 357–360.

Bosshart, P., Daly, D., Gibb, G. *et al.* (2014). P4: Programming protocol-independent packet processors. *ACM SIGCOMM Computer Communication Review*, 44(3), 87–95.

Boussaid, I., Lepagnot, J., Siarry, P. (2013). A survey on optimization meta-heuristics. *Information Sciences*, 237, 82–117.

Braga, R., Mota, E., Passito, A. (2010). Lightweight DDoS flooding attack detection using NOX/OpenFlow. Dans *35th Conference on Local Computer Networks (LCN)*, 408–415.

Canbay, Y., Sagiroglu, S. (2015). A hybrid method for intrusion detection. Dans *14th International Conference on Machine Learning and Applications (ICMLA)*, 156–161.

Canini, M., De Cicco, D., Kuznetsov, P., Levin, D., Schmid, S., Vissicchio, S. (2014). STN: A Robust and Distributed SDN Control Plane. Dans *Proceedings of Open Networking Summit (ONS'14)*.

Careglio, D., Spadaro, S., Cabellos, A., Lazaro, J.A., Perelló, J., Barlet, P., Gené, J.M., Paillissé, J. (2018). ALLIANCE Project: Architecting a Knowledge-Defined 5G-Enabled Network Infrastructure. Dans *20th International Conference on Transparent Optical Networks (ICTON)*.

Chen, X.F., Yu, S.Z. (2016). CIPA: A collaborative intrusion prevention architecture for programmable network and SDN. *Computers & Security*, 58, 1–19.

Chung, S.P., Mok, A.K. (2006). Allergy attack against automatic signature generation. Dans *9th International Conference on Recent Advances in Intrusion Detection*.

Clark, D.D., Partridge, C., Ramming, J.C., Wroclawski, J.T. (2003). A Knowledge plane for Internet. Dans *Conference on Applications, technologies, architectures, and protocols for computer communications*, 3–10.

Coly, A., Mbaye, M. (2019). S-SDS: A Framework for Security Deployment as Service in Software Defined Networks. Dans *Third EAI International Conference*.

Cordeiro, W.L.D.C., Marques, J.A., Gaspary, L.P. (2017). Data plane programmability beyond openflow: Opportunities and challenges for network and service operations and management. *Journal of Network and Systems Management*, 25(4), 784–818.

Das, S., Nene, M.J. (2017). A survey on types of machine learning techniques in intrusion prevention systems. Dans *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2296–2299.

Dey, S.K., Rahman, M.M., Uddin, M.R. (2018). Detection of Flow Based Anomaly in OpenFlow Controller: Machine Learning Approach in Software Defined Networking. Dans *4th International Conference on Electrical Engineering and Information & Communication Technology*.

Erickson, D. (2013). The beacon openflow controller. Dans *Proceedings of the second ACM SIGCOMM workshop on hot topics in software defined networking*.

Fortes, J. (2013). Cloud computing security: What changes with software-defined networking?. Dans *ARO Workshop on Cloud Security*.

Gowtham, V.N., Baratheraja, R.N., Jayabarathi, G., Vetriselvi, V. (2018). Collaborative Intrusion Detection System in SDN Using Game Theory. Dans *Proceedings of the International Conference on Computing and Communication Systems*, Mandal, J., Saha, G., Kandar, D., Maji, A. (dir.). Springer, Singapour.

Hajizadeh, M., Phan, T.V., Bauschert, T. (2018). Probability Analysis of Successful Cyber Attacks in SDN-based Networks. Dans *Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 1–6.

Hamdi, O., Mbaye, M., Krief, F. (2015). A Cloud-based Architecture for Network Attack Signature Learning. Dans *Conference on New Technologies Network and Security*.

Holland, J.H. (1975). *Adaption in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

Hoon, K.S., Yeo, K.C., Azam, S., Shunmugam, B., De Boer, F. (2018). Critical review of machine learning approaches to apply big data analytics in DDoS forensics. Dans *International Conference on Computer Communication and Informatics*, 1–5.

Huang, C.H., Lee, T.H., Chang, L.H., Lin, J.R., Horng, G. (2018). Adversarial Attacks on SDN-Based Deep Learning IDS System. Dans *International Conference on Mobile and Wireless Technology*, 181–191.

Huawei (2018). Roads to a better future. Dans *Mobile World Congress*.

Hyun, J., Hong, J.W.K. (2017). Knowledge-Defined Networking using In-band Network Telemetry. Dans *19th Asia-Pacific Network Operations and Management Symposium*.

Hyun, J., Van Tu, N., Hong, J.W.K. (2018). Towards Knowledge-Defined Networking using In-band Network Telemetry. Dans *IEEE/IFIP Network Operations and Management Symposium*.

Jiang, S. (2016). Intelligence Defined Network IDN. Dans *IETF 97*.

Karakus, M., Durresi, A. (2017). A survey: Control plane scalability issues and approaches in Software-Defined Networking (SDN). *Computer Networks*, 112, 279–293.

Kaur, S., Singh, J., Singh Ghumman, N. (2014). Network Programmability Using POX Controller. Dans *International Conference on Communication, Computing and Systems*.

Kent, S., Seo, K. (2005). Security Architecture for the Internet Protocol. Mémo, Network Working Group, Obsoletes 2401.

Khan, S.S., Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. *Pattern recognition letters*, 25(11), 1293–1302.

Klaedtke, F., Karame, G.O., Bifulco, R., Cui, H. (2014). Access control for SDN controllers. Dans *Proceedings of the 3rd workshop on Hot topics in software defined networking*, 219–220.

Kreutz, D., Ramos, F.M.V., Veríssimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. *Proceedings of the IEEE*, 103(1), 14–76.

Krief, F. (dir.) (2010). *Communicating embedded Networks: Network Applications*. ISTE Ltd, Londres et Wiley, New York.

Lu, J., Hongping, H., Yanping, B. (2015). Generalized radial basis function neural network based on an improved dynamic particle swarm optimization and AdaBoost algorithm. *Neurocomputing*, 152, 305–315.

Lu, W., Liang, L., Kong, B., Li, B., Zhu, Z. (2019). AI-Assisted Knowledge-Defined Network Orchestration for Energy-Efficient Datacenter Networks. *IEEE Communications Magazine*.

Ma, L., Zhang, Z., Ko, B., Srivatsa, M., Leung, K.K. (2018). Resource management in distributed SDN using reinforcement learning. Dans *SPIE Defense + Security*. 15–19 avril.

MacQueen, J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Dans *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.

Mbaye, M., Krief, F. (2009). A Collaborative Knowledge Plane for Autonomic Networks. Dans *Autonomic Communication*, Vasilakos, A.V., Parashar, M., Karnouskos, S., Pedrycz, W. (dir.). Springer, Boston, 60–90.

Mehdi, S.A., Khalid, J., Khayam, S.A. (2011). Revisiting traffic anomaly detection using software defined networking. Dans *International workshop on recent advances in intrusion detection*, 161–180.

Mestres, A., Rodriguez-Natal, A., Carner, J. *et al.* (2017). Knowledge-defined networking. *ACM SIGCOMM Computer Communication Review*, 47(3), 2–10.

Mestres, A., Alarcón, E., Cabellos, A. (2018). A Machine Learning-Based Approach for Virtual Network Function Modeling. Dans *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*.

Mihai-Gabriel, I., Victor-Valeriu, P. (2014). Achieving DDoS resiliency in a software defined network by intelligent risk assessment based on neural networks and danger theory. Dans *15th International Symposium on Computational Intelligence and Informatics (CINTI)*, 319–324.

Movahedi, Z., Ayari, M., Langar, R., Pujolle, G. (2012). Survey of Autonomic Network Architectures and Evaluation Criteria. *IEEE Communications Surveys & Tutorials*, 14(2).

Muggleton, S. (1991). Inductive logic programming. *New Generation Computing*, 8, 295–318.

Nam, T.M., Phong, P.H., Khoa, T.D. *et al.* (2018). Self-organizing map-based approaches in DDoS flooding detection using SDN. Dans *International Conference on Information Networking (ICOIN)*, 249–254.

Negnevitsky, M. (2005). *Artificial: A Guide to Intelligent Systems*, 2e édition. Addison-Wesley, Boston.

Newsome, J., Karp, B., Song, D. (2006). Paragraph: Thwarting signature learning by training maliciously. Dans *Proceedings of the 9th International Conference on Recent Advances in Intrusion Detection*, 81–105.

Nguyen, T.T., Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys & Tutorials*, 10(4).

Niyaz, Q., Sun, W., Javaid, A.Y. (2016). A deep learning-based DDoS detection system in software-defined networking (SDN) [En ligne]. Disponible à l'adresse : arXiv:1611.07400.

Nkosi, M., Lysko, A., Ravhuanzwo, L., Nandeni, L., Engelberencht, A. (2016). Classification of SDN distributed controller approaches: a brief overview. Dans *International Conference on Advances in Computing and Communication Engineering (ICACCE)*.

Oktian, Y.E., Lee, S., Lee, H., Lam, J. (2017). Distributed SDN controller system: A survey on design choice. *Computer Networks*, 121, 100–111.

ONF (2015). Software-Defined Networking: The new norm for networks [En ligne]. Livre blanc. Disponible à l'adresse : https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf.

Pal, N.R, Bezdek, J.C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3), 370–379.

Rebecchi, F., Boite, J., Nardin, P.-A., Bouet, M., Conan, V. (2017). Traffic monitoring and DDoS detection using stateful SDN. Dans *Conference on Network Softwarization (NetSoft)*, 1–2.

Russell, S., Norvig, P. (1995). *Artificial Intelligence (A Modern Approach)*, 3e édition. Prentice Hall, Upper Saddle River.

Sahu, S., Mehtre, B.M. (2015). Network intrusion detection system using J48 Decision Tree. Dans *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2023–2026.

Shu, Z., Wan, J., Li, D., Lin, J., Vasilakos, A.V., Imran, M. (2016). Security in Software-Defined Networking: Threats and Countermeasures. *Mobile Networks and Applications*, 21(5), 764–776.

Soheily-Khah, S., Marteau, P.F., Béchet, N. (2018). Intrusion Detection in Network Systems Through Hybrid Supervised and Unsupervised Machine Learning Process: A Case Study on the ISCX Dataset. Dans *1st International Conference on Data Intelligence and Security (ICDIS)*, 219–226.

Srinivasan, A. (2000). The aleph manual. Rapport technique, Oxford University, Oxford.

Tang, T.A., Mhamdi, L., McLernon, D., Zaidi, S.A.R., Ghogho, M. (2016). Deep learning approach for network intrusion detection in software defined networking. Dans *International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 258–263.

Tang, T.A., Mhamdi, L., McLernon, D., Zaidi, S.A.R., Ghogho, M. (2018). Deep recurrent neural network for intrusion detection in sdn-based networks. Dans *4th Conference on Network Softwarization and Workshops (NetSoft)*, 202–206.

Terzi, D.S., Terzi, R., Sagiroglu, S. (2017). Big data analytics for network anomaly detection from netflow data. Dans *International Conference on Computer Science and Engineering (UBMK)*, 592–597.

Thaseen, S., Kumar, C.A. (2013). An analysis of supervised tree-based classifiers for intrusion detection system. Dans *International Conference on Pattern recognition, informatics and mobile engineering (PRIME)*, 294–299.

Wang, J., Wang, Y., Hu, H., Sun, Q., Shi, H., Zeng, L. (2013). Towards a Security-Enhanced Firewall Application for OpenFlow Networks. Dans *Cyberspace Safety and Security. Lecture Notes in Computer Science*, Wang, G., Ray, I., Feng, D., Rajarajan, M. (dir.). Springer, Bâle.

Xie, J., Yu, F.R., Huang, T., Xie, R. *et al.* (2019). A Survey of Machine Learning Techniques Applied to Software Defined Networking (SDN): Research Issues and Challenges. *IEEE Communications Surveys and Tutorials*, 21(1), 393–430.

Zhang, G.P. (2000). Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451–462.

Zhang, C., Xia, S. (2009). K-means clustering algorithm with improved initial center. Dans *Second International Workshop on Knowledge Discovery and Data Mining*, 790–792.

Zhang, Y., Cui, L., Wang, W., Zhang, Y. (2017). A survey on software defined networking with multiple controllers. *Journal of Network and Computer Applications*, 103, 101–118.

PARTIE 2

L'IA et l'optimisation des réseaux

3

Optimisation des réseaux à l'aide des techniques de l'intelligence artificielle

Asma AMRAOUI et Badr BENMAMMAR
Université Abou Bekr Belkaid, Tlemcen, Algérie

3.1. Introduction

Les télécommunications ont beaucoup évolué au cours des dernières années avec l'explosion des marchés des mobiles et d'Internet, le déploiement des réseaux à hauts débits et de réseaux intelligents. Ces évolutions rendent les environnements réseau plus complexes, car ils traitent en permanence une énorme quantité d'informations ; ce qui rend la gestion des réseaux plus difficile.

Les fournisseurs de services de communications actuels doivent faire face aux demandes croissantes des clients pour des services de meilleure qualité et une meilleure expérience client. Les entreprises de télécommunication saisissent ces opportunités en exploitant les vastes quantités de données collectées au fil des années auprès de leur vaste clientèle. Ces données sont extraites d'appareils, de réseaux, d'applications mobiles, de géolocalisations, de profils clients détaillés.

Pour traiter et analyser ces énormes volumes de données et obtenir des informations exploitables, les télécoms profitent de la puissance de l'intelligence artificielle (IA) pour offrir une meilleure expérience client, améliorer les opérations et augmenter les revenus des entreprises grâce à de nouveaux produits et services.

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

En effet, l'IA peut aider à identifier les anomalies et résoudre les problèmes de manière proactive avant que les clients ne soient affectés, et ainsi optimiser le réseau. L'optimisation du réseau, la maintenance prédictive, les assistants virtuels sont des exemples de cas où l'IA a eu un impact sur le secteur des télécommunications.

Dans ce chapitre, nous allons parler de l'IA de façon générale et définir les différentes techniques intelligentes utilisées habituellement dans le secteur des télécommunications, en passant par les systèmes experts, l'apprentissage automatique, les systèmes multiagents, mais aussi l'Internet des objets et les Big Data, qui sont très tendance et ont beaucoup de succès auprès des entreprises de télécommunications.

Ce chapitre se focalise sur quatre aspects de l'optimisation réseau : les performances du réseau, la qualité de service, la sécurité et la consommation d'énergie. Pour chacun de ces critères, nous allons expliquer en quoi consiste leur optimisation et comment l'IA contribue à une meilleure utilisation.

3.2. Intelligence artificielle

3.2.1. *Définition*

L'intelligence humaine s'oppose à l'instinct, qui correspondrait davantage à un réflexe qu'à une pensée élaborée. L'intelligence artificielle est une science qui est apparue après la Seconde Guerre mondiale avec l'invention des premiers ordinateurs électroniques. Le but de cette science était double : simuler les capacités humaines pour mieux comprendre l'intelligence humaine et remplacer l'homme dans certaines tâches automatiques et répétitives. L'intelligence est souvent associée à la capacité de raisonnement et de réflexion d'une personne.

L'IA est un terme créé par John Mc Carthy et Marvin Lee Minsky :

> « Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisantes par des êtres humains car elles demandent des processus mentaux de haut niveau tels que l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. »

On pourrait dire que l'IA sert à concevoir des systèmes capables de reproduire le comportement de l'humain dans ses activités de raisonnement.

Depuis quelques années, on associe presque toujours l'intelligence aux capacités d'apprentissage. C'est grâce à l'apprentissage qu'un système intelligent est capable d'exécuter une tâche et peut améliorer ses performances avec l'expérience.

3.2.2. *Techniques de l'intelligence artificielle*

3.2.2.1. *Systèmes experts*

Un système expert (SE) est un outil capable de reproduire les mécanismes cognitifs d'un expert humain, dans un domaine particulier. Il s'agit de l'une des voies tentant d'aboutir à l'IA. Plus précisément, un SE est un logiciel capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connus.

Les systèmes experts sont généralement constitués de :

– une base de connaissance (*knowledge base*) ;

– une interface (*interface*) ;

– un moteur d'inférence (*inference engine*).

La base de connaissance est un ensemble de données qui sont utilisées par le moteur d'inférence. C'est là qu'est stocké le savoir du système propre au domaine de connaissance. Elle rassemble toutes les connaissances d'un expert du domaine considéré.

La base de connaissance contient :

– les standards d'engagement (connaissances de l'expert) : l'information de base et de configuration du système, mesures, lois, paramètres, données contractuelles ;

– les règles d'inférence (savoir-faire) : ensemble de règles logiques de déduction utilisées par le moteur d'inférence ;

– la base de faits (expérience) : ensemble des données de départ sur lequel le système va commencer à travailler. Cette base s'enrichit au fur et à mesure des déductions faites par le système. Cet espace de travail est un peu la mémoire à court terme, le système y stocke aussi les règles en attente, les sous-problèmes, etc.

Les interfaces servent au dialogue entre l'expert, chargé de la création de la base de connaissances, et la machine.

Le moteur d'inférence est le mécanisme qui permet d'inférer des connaissances nouvelles à partir de la base de connaissances du système. C'est le cerveau du système et il sert à déclencher les règles et à les enchaîner les unes après les autres.

Les deux mécanismes les plus employés pour déclencher les règles sont :

– le chaînage avant ;

– le chaînage arrière.

Un SE est différent d'un logiciel classique. En effet, un logiciel classique est développé autour d'un ensemble de procédures algorithmiques. La résolution d'un problème suit une séquence d'étapes bien définie par le programmeur. Un SE quant à lui peut intégrer la capacité à déterminer lui-même les traitements adaptés à un état donné des paramètres d'entrée, c'est-à-dire une séquence d'étapes non prédéfinie par le programmeur pour cet état. Cette différence entre un logiciel classique et un système expert est essentiellement due à la méthode d'organisation et d'utilisation des connaissances spécialisées.

L'avantage principal d'un SE est qu'il est très performant pour résoudre les problèmes rencontrés durant la période d'expertise pour lesquels des règles ont été formulées. Cependant, pour un domaine de grande dimension, le nombre de règles augmente de manière considérable et il devient très difficile d'assurer leur maintenance. En effet, il faut pouvoir poursuivre l'expertise du domaine étudié, formuler de nouvelles règles et les corréler manuellement avec toutes les règles existantes.

Un système expert est donc très adapté pour des domaines très peu changeants. Par contre, si le domaine est très dynamique, certaines règles expertes peuvent devenir très rapidement obsolètes et peuvent fragiliser le système, qui deviendra incapable de résoudre certains problèmes. Ces faiblesses des systèmes experts à base de règles ont conduit au développement d'une nouvelle approche de représentation des connaissances expertes.

Il y a deux types de systèmes experts :

– les systèmes experts classiques à base de règles, qui formulent des règles pour décrire et comprendre le phénomène de propagation des pannes et des alarmes sur un réseau de télécommunication ;

– les systèmes experts évolutifs à base de modèles, qui s'inspirent des sciences de l'artificiel et qui considèrent qu'un phénomène n'est compris que lorsqu'il peut être reproduit ou simulé. Dans cette catégorie se rangent les méthodes de diagnostic à base de modèle qui développent des raisonnements sur une représentation explicite de la structure et du fonctionnement du réseau et les méthodes qui essayent d'apprendre artificiellement le comportement du réseau sans le modéliser.

3.2.2.2. *Raisonnement à partir de cas*

Le raisonnement à partir de cas (RàPC) est un paradigme de l'IA qui consiste à résoudre un nouveau problème, appelé « problème cible », en utilisant un ensemble de problèmes déjà résolus. Le RàPC est un raisonnement analogique qui satisfait globalement ce que l'on appelle « le carré d'analogie » tel qu'illustré dans la figure 3.1.

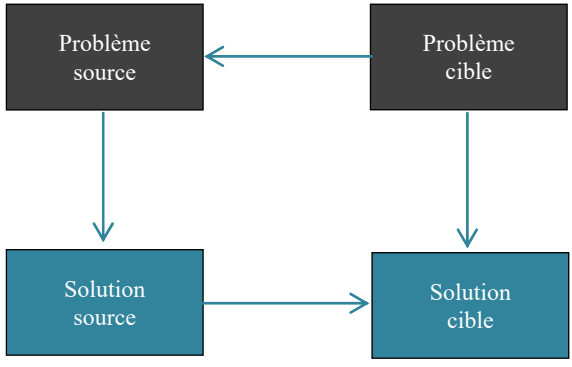


Figure 3.1. *Carré analogique des SE*

La recherche de cas sources similaires est naturellement essentielle dans le cycle. Nous rappelons que le cas source qui sera choisi sera normalement le cas ayant la description de problème la plus proche possible de la description du problème cible.

La réutilisation consiste à réutiliser un cas similaire pour avoir une trace de raisonnement du cas cible. Alors que la révision permet de corriger, de façon à ce que le cas soit un cas avec une solution correcte.

L'apprentissage du nouveau cas résolu est l'occasion d'enrichir la base de connaissances.

Le principe du RàPC consiste à récupérer, adapter et exécuter les solutions des précédents problèmes pour évaluer les problèmes actuels. Les solutions de diagnostic passé sont stockées sous forme de cas dans une base de connaissances. Les cas contiennent les caractéristiques les plus pertinentes des solutions de diagnostics passés ; ils sont adaptés et utilisés pour résoudre les nouveaux problèmes.

L'expérience acquise en diagnostiquant ces nouveaux problèmes constitue de nouveaux cas stockés pour une prochaine utilisation. Ce système intègre la capacité d'apprendre non seulement de ses précédentes solutions correctes de diagnostic, mais aussi de ses échecs. En effet, lorsqu'une tentative de diagnostic d'une situation échoue, le système identifie et journalise la raison de cet échec, afin de s'en souvenir lors de futurs diagnostics de situations.

Un système RàPC dispose d'une base de cas. Chaque cas possède une description détaillée du problème et une solution. Pour utiliser ces informations, un moteur est aussi nécessaire. Celui-ci va retrouver les cas similaires au nouveau

problème posé. Après analyse, le moteur fournit une solution adaptée qui doit être validée. Enfin, le moteur ajoute le problème et sa solution dans la base de cas.

3.2.2.3. *Apprentissage automatique*

Il s'agit de techniques issues de l'IA, permettant à des machines d'apprendre, d'une manière plus ou moins autonome, à réaliser des tâches sans être explicitement programmées.

L'apprentissage automatique fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques.

Il existe trois grands types d'apprentissage automatique :

– apprentissage supervisé : l'algorithme cherche à prédire un phénomène ou une mesure en se basant sur l'historique des réalisations de cette dernière. La base de données est formée avec des données étiquetées ;

– apprentissage non supervisé : il ne s'agit pas de prédire une mesure particulière ; l'algorithme cherche plutôt de lui-même à déceler des structures ou regroupements caractéristiques sur un ensemble d'observations qu'on lui fournit. Les données ne sont pas étiquetées ; l'objectif est alors de trouver une relation entre les données ;

– apprentissage par renforcement : l'agent intelligent observe les effets de ses actions et déduit la qualité de ses actions pour améliorer ses actions futures. L'action de l'algorithme sur l'environnement produit une valeur de retour (une récompense ou une pénalité) qui guide l'algorithme d'apprentissage.

Pour conclure, on peut affirmer que l'objectif principal de l'apprentissage automatique est d'extraire et d'exploiter automatiquement l'information présente dans un jeu de données. Mais le véritable potentiel de l'apprentissage automatique réside dans le traitement de données jamais vues auparavant, tout en apportant les bonnes réponses. Pour cette raison, le cœur de l'apprentissage automatique est la quantité et la qualité des données, ainsi que le choix du meilleur algorithme d'apprentissage automatique qui s'intègre dans nos données.

Réseaux de neurones

Le cerveau humain se compose d'un ensemble de neurones interconnectés transmettant des modèles élaborés de signaux électriques. Les dendrites reçoivent les signaux d'entrée et, sur la base de ces entrées, un neurone produit un signal de sortie *via* un axone (Shiffman 2012).

Les réseaux de neurones artificiels s'inspirent du fonctionnement biologique du cerveau humain et donc, par analogie avec un neurone biologique, un neurone artificiel est perçu comme un processeur autonome avec des canaux unidirectionnels pour la communication avec les autres neurones qui lui sont connectés.

Un neurone artificiel a plusieurs canaux d'entrées fonctionnant comme des dendrites, et un seul canal de sortie fonctionnant comme un axone. Les points de connexions entre les neurones sont appelés « synapses ». L'opération typique d'un neurone artificiel est de calculer une somme pondérée des signaux d'entrée et de générer un signal de sortie si cette somme dépasse un certain seuil. La somme pondérée des signaux d'entrée est effectuée par la fonction de combinaison, qui multiplie le vecteur des entrées par une matrice de transformation. Le signal de sortie est généré par la fonction.

La figure 3.2 représente la structure d'un réseau de neurones artificiel.

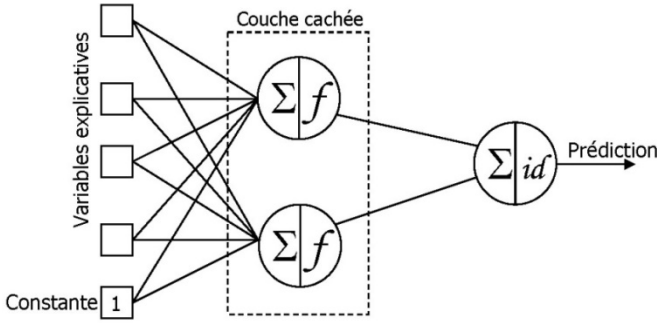


Figure 3.2. *Réseau de neurones artificiel (Decourt 2018)*

3.2.2.4. *Systèmes multiagents*

Un système multiagents (SMA) est un regroupement d'agents où chaque agent possède une ou plusieurs compétences élémentaires. Le but est de faire travailler ensemble les agents pour résoudre un problème ou effectuer une tâche spécifique. En quelque sorte, on distribue l'intelligence, chaque agent autonome n'ayant qu'une vision locale du problème ou une tâche élémentaire d'un travail à effectuer.

Ferber et Perrot (1995) définissent un SMA de la manière suivante :

> « Un système multiagents est un système composé des éléments suivants :

– un environnement qui est un espace disposant généralement d'une métrique ;

– un ensemble d'objets situés dans l'espace ; ils sont passifs ; ils peuvent être perçus, détruits, créés et modifiés par les agents ;

– un ensemble d'agents, qui sont les entités actives du système ;

– un ensemble de relations, qui unissent les objets entre eux ;

– un ensemble d'opérations permettant aux agents de percevoir, de détruire, de créer, de transformer et de manipuler les objets ;

– un ensemble d'opérateurs chargés de représenter l'application de ces opérations et la réaction du monde à cette tentative de modification (les lois de l'univers). »

Les SMA sont utilisés en général lorsque le problème est trop complexe pour être résolu par un seul système, à cause de quelques limitations logicielles ou matérielles. En particulier, si les composantes entretiennent des relations multiples entre elles. Les SMA représentent un excellent outil pour assurer un contrôle autonome dans un système largement distribué et dont les caractéristiques sont très dynamiques.

Quand on a besoin d'un système qui doit s'adapter dynamiquement lorsque l'on ajoute ou que l'on retire de nouvelles composantes et que ces dernières doivent s'adapter facilement lorsque l'environnement subit des modifications, les SMA sont sûrement la solution idéale pour ce genre de scénarios. Il ne faut pas oublier que l'un des avantages les plus importants des SMA est leur modularité, qui permet de rendre la programmation plus simple, c'est-à-dire que l'ajout de nouveaux agents à un SMA ne pose aucun problème important ; ce qui explique leur extensibilité (Amraoui 2015).

L'intérêt de la solution à base d'agents réside dans l'absence totale d'entité centrale régissant le fonctionnement des agents, ce qui garantit une grande résistance et une grande fiabilité (car si un agent tombe en panne, le système continue de fonctionner).

3.2.2.5. *Internet des objets*

Le terme « Internet des objets » (*Internet of Things*, IoT) est utilisé généralement pour décrire un système où les objets physiques sont connectés à Internet, appelé maintenant « écosystème des objets connectés ».

L'IoT commence ainsi dans le monde physique avec les capteurs qui recueillent les informations ; celles-ci sont ensuite transmises grâce à la connexion et à l'intégration des systèmes entre eux ; les données sont enfin traitées et stockées pour être analysées et exploitées.

Une caractéristique essentielle est que l'IoT peut transformer des objets ordinaires en dispositifs. Ils peuvent être identifiés par une adresse IP, enregistrer des états *via* des capteurs et disposer d'une capacité de mémoire *via* des puces. Les mini-ordinateurs intégrés leur permettent de se contrôler, de gérer leur environnement et d'échanger des données automatiquement. Parfois, grâce à l'apprentissage automatique, ils sont même capables de reconnaître et de généraliser des modèles et de tirer des conclusions pour s'adapter aux situations et s'optimiser continuellement.

3.2.2.6. Cloud Computing

La définition officielle du *Cloud Computing* est donnée par (Mell *et al.* 2011) : « Le *Cloud Computing* est un modèle permettant un accès réseau omniprésent et pratique à la demande à un pool partagé de ressources informatiques configurables (par exemple, réseaux, serveurs, stockage, applications et services). Ces ressources peuvent être rapidement mises en service et libérées avec un effort de gestion minimal ou avec une interaction des fournisseurs de services. »

Le *Cloud Computing* désigne la fourniture de diverses solutions matérielles et logicielles *via* Internet. Les performances des processeurs, l'espace de stockage et les environnements logiciels peuvent être loués au besoin par les utilisateurs pour étendre ou remplacer leur propre infrastructure. Le *Cloud Computing* offre aux utilisateurs la possibilité de stocker une quantité énorme de données ainsi qu'un accès depuis n'importe quel endroit et et à n'importe quel moment. Pour utiliser les informations stockées, les utilisateurs doivent disposer d'une connexion internet de base.

3.2.2.7. *Big Data*

Big Data englobe tout terme pour décrire toute collection de données tellement volumineuse et complexe qu'il devient difficile de la traiter en utilisant des outils classiques de traitement d'applications. C'est un terme générique employé pour désigner les stratégies et technologies mises en œuvre pour rassembler, organiser et analyser de vastes ensembles de données.

3.3. Optimisation des réseaux

Optimiser un réseau signifie améliorer son fonctionnement en termes de sécurité et fiabilité, de performance et rapidité, de qualité de service et, bien sûr, en termes de consommation d'énergie.

Dans cette section, nous allons donner un aperçu sur la façon dont l'IA optimise les réseaux.

3.3.1. *Intelligence artificielle et optimisation des performances du réseau*

Avec la demande Wi-Fi qui augmente quotidiennement et l'intégration des objets connectés dans nos vies, le manque de visibilité et le contrôle sur le réseau sont des facteurs majeurs de la gestion informatique, et donc, assurer continuellement la même qualité de service (QoS) et la même performance est un défi qui doit être relevé. Une solution prometteuse est la mise en place d'un réseau cognitif dit intelligent.

Les réseaux cognitifs sont un type particulier de réseaux capables d'apprendre, de prévoir et d'améliorer les performances, la sécurité et l'expérience utilisateur. Ces solutions exploitent le *Cloud*, l'analyse de données, l'apprentissage automatique et l'intelligence artificielle pour déterminer les performances de base, suivre l'activité et identifier les problèmes.

La grande quantité de données entrant au réseau à partir de différents nœuds nécessite forcément une puissance de calcul très élevée. Avec l'utilisation de l'IA, toutes ces données doivent être étudiées dans le cadre d'un apprentissage automatique. Ainsi, le réseau pourra comprendre, par exemple, le moment où les applications atteignent des performances non optimales ou bien comparer un flux constant de données historiques et actuelles.

3.3.2. *Intelligence artificielle et optimisation de la qualité de service*

Les progrès technologiques récents ont permis la fabrication de petits éléments à faible coût, comme les petits capteurs sans-fil qui sont utilisés pour mesurer les conditions ambiantes de l'environnement où ils sont installés. Ces derniers sont utilisés pour leur faible consommation d'énergie, leur faible portée radio, leur faible mémoire et leur faible coût.

Notons également que le trafic multimédia a considérablement augmenté ces dernières années, car les technologies récentes (IoT par exemple) ajoutent de nouveaux types de trafic et surtout les flux vidéo. D'après le rapport effectué par Cisco[1], le trafic vidéo en 2021 sera de trois fois celui du trafic vidéo de 2016. De plus, ce trafic représente 82 % du trafic internet total en 2021.

Tout cela pour dire que la QoS réseau sera toujours nécessaire pour la transmission de données imprévisibles en temps réel. Par conséquent, nous pouvons affirmer que les tâches lourdes devront toujours être effectuées dans le Cloud.

1. Cisco Visual Networking Index, Forecast and Methodology, 2016-2021.

Ceci dit, les FAI n'ont aucun contrôle et généralement aucune connaissance sur les points d'accès Wi-Fi qu'utilisent les utilisateurs des appareils mobiles et ils ne peuvent pas assurer que la QoS soit livrée comme promis.

Lorsque l'on parle de QoS, cela ne concerne pas uniquement le débit, la perte de paquets, la latence ou la gigue. C'est aussi une question de disponibilité. Pour pouvoir optimiser encore plus les réseaux et permettre une gestion efficace de l'énergie, une fiabilité et une disponibilité élevée, il est important d'assurer une sécurité des communications et pour cela l'intégration de l'IA est nécessaire pour une meilleure gestion dynamique du trafic réseau.

Avec l'utilisation des techniques de l'IA, nous pouvons découvrir les différents types de flux qui sont en cours de transmission sur le réseau. Ainsi, des modèles de trafic peuvent être obtenus, qui peuvent aider à la prise de décision.

Dans (Nowicki et Uhl 2017), les auteurs pensent que le trafic multimédia peut être géré le plus efficacement avec les techniques de l'IA. Leur papier propose un système intelligent pour garantir la QoS et la QoE dans la vidéosurveillance du trafic généré par les appareils de l'IoT.

La qualité dans le réseau ne peut pas être rétablie une fois que celle-ci est perdue. Mais il y a la possibilité d'intégrer qualité et intelligence cognitive à chaque extrémité de la connexion.

L'apprentissage automatique peut également être utile pour atténuer les risques liés à l'indisponibilité du réseau ou à des exploits de sécurité. Avec la radio cognitive (Benmammar *et al.* 2012) par exemple, l'application sait que vous êtes sur le point de traverser une zone noire de couverture et elle peut agir en conséquence à l'avance.

D'autres travaux existent dans la littérature et se basent sur les métaheuristiques comme le papier de (Benmammar 2017). Ce dernier utilise la métaheuristique SFLA (*Shuffled Frog Leaping Algorithm*) pour améliorer la QoS dans un réseau de radio cognitive. L'objectif des auteurs est de maximiser le débit, de minimiser le taux d'erreur et de minimiser la consommation d'énergie dans ce type de réseau.

3.3.3. *Intelligence artificielle et sécurité*

D'après l'étude effectuée par (Cigref 2018), plus d'une entreprise sur deux a été touchée par des cyberattaques, ce qui représente un chiffre énorme en termes de coûts.

En France, 29 % des sociétés seulement perçoivent la cybersécurité comme un enjeu prioritaire[2]. Plus gênant encore, seule une entreprise sur deux a mis en place une stratégie dédiée pour lutter contre les cyberrisques.

Pour faire face à ce genre d'attaques, il faut sensibiliser les employés aux mesures de base de sécurité et à la nécessité d'utiliser les antivirus et les *firewalls*. Mais ceci n'est bien sûr pas suffisant pour assurer la sécurité totale et, pour cela, il existe d'autres solutions plus évoluées et plus performantes à base d'IA, et plus précisément à base d'apprentissage automatique. Ces nouvelles méthodes permettent de détecter plus aisément les anomalies et génèrent des alertes assez rapidement afin de prévenir les administrateurs du système.

Contrairement aux méthodes classiques telles que les antivirus, qui s'assurent que la machine ne contient pas une signature spécifique (qui indique que tel programme est considéré comme malveillant ou pas), les systèmes à base d'apprentissage automatique apprennent à rechercher les différentes caractéristiques des *malwares*, afin d'apprendre leur comportement pour pouvoir les détecter plus rapidement ; ces systèmes sont donc plus flexibles.

Les techniques dites intelligentes sont très utilisées dans la lutte contre le spam et le *phishing* et ont donné de bons résultats. Ces dernières peuvent aussi être utilisées autrement pour protéger le système des attaques intérieures (qui viennent des employés malintentionnés). Pour cela, l'IA peut faire une analyse comportementale : elle étudie le comportement d'un utilisateur sur un ordinateur et permet d'alerter les responsables de la sécurité du système lors d'un comportement déviant de l'ordinateur (tentative de fuites de données par exemple).

Ceci dit, le fonctionnement des techniques intelligentes leur permet d'analyser les situations et les comportements, mais pour cela elles ont besoin d'une grande quantité de données pour pouvoir donner des résultats efficaces et satisfaisants. Ce qui implique qu'un manque de données peut aboutir à de faux résultats et donc à de fausses alertes ; ce qui rend impossible l'automatisation entière d'un système de sécurité, car l'intervention humaine reste indispensable dans certains cas.

Les moteurs d'IA actuels utilisent les données statistiques pour effectuer la classification (malveillante ou honnête), mais la capacité de ces derniers peut aussi devenir une faiblesse. En effet, le moteur d'apprentissage automatique équivaut à la

2. Selon le document de référence IPSOS de 2017 et « Les entreprises face aux enjeux de la cybersécurité » de PWC, disponible à l'adresse : https://www.pwc.fr/fr/assets/files/pdf/2018/10/pwc-barometre-cybersecurite-septembre-2018.pdf.

capacité d'apprentissage des êtres humains, mais à une échelle et avec une rapidité bien plus grande.

Pour avoir une IA efficace, on soumet à son moteur d'apprentissage de très nombreuses informations. Au fil des informations reçues, le moteur élabore un modèle statistique qui lui permet de déterminer de façon autonome lorsque le phénomène recherché se manifeste.

De fait, puisque la force des algorithmes statistiques tient à la reconnaissance de modèles, de schémas, les attaquants sont susceptibles d'adapter progressivement leurs comportements pour qu'ils paraissent normaux, ou de réaliser leurs actions de sorte à induire une confusion.

D'ailleurs, de nombreux systèmes peuvent détecter des anomalies au départ, mais au bout d'un moment ils s'entraînent à les accepter comme des comportements normaux. Ce qui donne un avantage aux attaquants, car ils peuvent masquer leurs activités en observant les comportements normaux, comme l'utilisation du protocole « https » pour l'envoi des données à un serveur.

Les attaquants peuvent ajouter des étapes d'exécution inutiles en rapport avec le but poursuivi, mais pensées pour donner une apparence de normalité au processus. À cela peuvent également s'ajouter des signaux faibles d'apparence anodine pour l'analyste humain, mais efficaces pour tromper les algorithmes d'apprentissage automatique.

Pour finir, on peut dire que l'intelligence artificielle apporte une couche de sécurité supplémentaire et qu'elle peut ralentir considérablement des pirates informatiques. En effet, même si l'humain est capable de traiter plusieurs menaces par heure, il peut vite être dépassé devant un flux important de menaces. L'IA aide l'homme dans ses actions de traitement des incidents de sécurité et va même pouvoir suggérer ou encore appliquer des actions de remédiations rapidement.

Les systèmes intelligents sont faiblement consommateurs en ressources CPU et Ram par rapport aux antivirus traditionnels et ne nécessitent pas forcément de connexion internet. Ils n'ont plus besoin de connaître la menace pour la bloquer ni de faire des mises à jour en permanence, car le modèle repose sur une approche statistique. Le système va analyser une grande quantité de données comportant les différentes caractéristiques d'un fichier, son éventuelle signature, sa taille, son code, toutes ces suites de bits qui se répètent. À partir de là, un score lui sera attribué pour savoir si un fichier peut s'exécuter ou non.

3.3.4. *Intelligence artificielle et consommation d'énergie*

L'intelligence artificielle, lorsqu'elle est utilisée dans la production ou la consommation d'énergie, fonctionne grâce à des capteurs, qui sont installés dans les systèmes de contrôle. Cela permet de faire un traitement en temps réel des données. Grâce à cela, les anomalies ou les dysfonctionnements du système sont détectés et pris en charge beaucoup plus rapidement. Une fois les problèmes mis en lumière, le système ou les équipements défectueux peuvent être remplacés ; cela permet d'optimiser au maximum l'efficacité énergétique.

Le développement d'objets connectés associé à l'utilisation de technologies de l'IA permet de déployer des outils d'assistance à la consommation et de gestion intelligente de l'énergie. Il permet aussi de déployer des systèmes de prédiction et de gestion de la consommation en temps réel, basés sur le recours au stockage et à l'autoconsommation.

La prédiction de la consommation et de la production d'énergie pour la gestion énergétique en temps réel peut être effectuée à l'aide d'algorithmes de régression du type forêt d'arbres décisionnels ou machine de Boltzmann restreinte. En ce qui concerne l'amélioration de l'efficacité énergétique, des méthodes de *clustering* de type k-means peuvent être utilisées.

3.4. Application de l'intelligence artificielle dans les réseaux

3.4.1. *Systèmes experts et réseaux*

3.4.1.1. *Système expert pour la maintenance des machines*

Un système expert de diagnostic est une suite d'applications informatiques qui intègrent une grande base de connaissances ou de raisonnements d'experts sur des tâches de diagnostic bien précises et qui infèrent de façon automatique les causes racines des anomalies observées.

C'est un système informatique destiné à déterminer une cause de défaillance d'un équipement à partir d'une analyse et d'une représentation des connaissances et du raisonnement d'un ou de plusieurs spécialistes de maintenance. Il permet de donner la possibilité à un opérateur possédant des connaissances techniques moyennes ou même faibles de chercher la cause probable d'une défaillance, en communiquant au système le minimum d'informations telles que le type de machine défaillante et le mode de défaillance observé.

La modélisation des connaissances est la partie la plus importante dans la conception d'un système d'aide au diagnostic, donc il faut bien définir les éléments à étudier et les relations entre eux.

La base de faits est l'espace de travail du système d'aide au diagnostic qui s'enrichit au fur et à mesure du déroulement du système du point de vue de la sélection de la cause de défaillance la plus probable et du remède qui lui convient.

Pour constituer une base de raisonnements valides, les connaissances d'experts en matière de diagnostic peuvent être formalisées sous forme de règles, d'arbre de décisions, de logique propositionnelle, etc.

Dans le cas du SE qui aide à la maintenance des machines (Kaushik *et al.* 2011 ; Raja'a et Jassim 2014), la base de connaissances contient les connaissances propres à la machine qui sont fournies par les experts de la maintenance. Ces connaissances sont exprimées sous forme de faits et de règles.

L'apprentissage se fait dans ce cas à partir des données générées par les équipements du réseau à diagnostiquer et consiste essentiellement à résoudre par interpolation ou induction le problème inverse de la propagation de pannes et d'alarmes sur un réseau de télécommunication.

3.4.1.2. *Système expert pour le diagnostic des réseaux de multiplexeurs*

Dans (Lor 1993), les auteurs ont développé un système de diagnostic des réseaux de multiplexeurs. Les connaissances expertes de diagnostic sont classifiées en deux catégories : les connaissances expertes génériques et les connaissances expertes spécialisées sur des tâches de diagnostic bien précises. Le système expert utilise une base de données contenant des informations statiques et des informations dynamiques nécessaires durant le processus de diagnostic. Ces informations concernent les relations entre les entités logiques (groupes de canaux) et les entités physiques (équipements et liens), telles que les informations de routage, les attributs des entités physiques, les relations d'incidences entre les nœuds du réseau et les liens, etc.

Le diagnostic d'une ligne se fait en deux étapes. On commence d'abord par collecter les données disponibles de cette ligne, telles que les niveaux de puissances transmises et reçues par les équipements de cette ligne, les tensions d'alimentation, les courants de polarisation, les températures des équipements de cette ligne, les compteurs d'erreurs de transmission, les alarmes observées, etc. Chaque donnée est stockée dans une KPI (*Key Performances Identifier*). Ensuite, des règles expertes prédéfinies utilisent ces KPI pour produire une indication ou une décision finale de diagnostic appelée « conclusion ».

DELC (Diagnostic expert de la ligne cliente) est un système expert à base de règles développé par Orange Labs France pour le diagnostic automatique des réseaux d'accès cuivre xDSL (*Digital Subscriber Line*) et optique GPON (*Gigabit Passive Optical Network*) de type FTTH (*Fiber To The Home*).

3.4.2. *Raisonnement à partir de cas et réseaux de télécommunications*

Le plus souvent, et notamment dans un système RàPC complexe, tel que celui du diagnostic d'un réseau de télécommunication avec une très grande diversité de signatures d'anomalies, une adaptation des solutions préexistantes est presque toujours requise.

Un système RàPC de diagnostic des pannes d'un réseau appelé DUMBO a été proposé dans (Melchiors et Tarouco 1999). Ce système utilise les connaissances des cas de diagnostic stockés dans un système à tickets d'incidences pour proposer des solutions de diagnostic aux nouvelles anomalies survenues. Ce système vise à faciliter les étapes de diagnostic et de résolution des problèmes de gestion de réseau.

L'unité de connaissance d'un système CBR est le cas et non pas la règle. Il est plus facile d'articuler, d'examiner et d'évaluer un cas qu'une règle (Hounkonnou 2013). Un système RàPC est également capable d'apprendre de ses propres erreurs/échecs et d'auto-améliorer ses performances. Il faut faire attention à la phase d'évaluation des solutions de nouveaux problèmes. En effet, une mauvaise évaluation pourrait entraîner l'intégration de cas erronés dans la base des connaissances et faire ainsi dériver le système dans son ensemble.

3.4.3. *Apprentissage automatique et réseaux de télécommunications*

L'apprentissage automatique est une méthode qui peut diagnostiquer un plus grand nombre de pannes que les systèmes experts à base de règles ; il peut diagnostiquer les problèmes hors de son expertise, même si ses performances se dégradent dans ces cas-là.

Effectuer le diagnostic d'un réseau de télécommunication requiert la compréhension du phénomène de propagation de pannes et d'alarmes dans ce réseau. Cette compréhension permet d'acquérir des connaissances pertinentes, afin de résoudre automatiquement le problème inverse de la propagation de pannes et d'alarmes.

Pour pouvoir diagnostiquer les anomalies qui pourraient survenir sur un réseau de télécommunications, le système de diagnostic doit être un système apprenant, c'est-à-dire

doté de capacités d'induction qui lui permettront d'utiliser sa base de connaissances pour trouver les causes racines de nouvelles anomalies qui lui étaient jusqu'alors inconnues.

Dans cette méthode, on n'utilise plus une base de raisonnements spécialisés sur des tâches précises de diagnostic, comme c'est le cas dans les SE et les RàPC, mais des connaissances sur le comportement ou le fonctionnement du réseau de télécommunications. Ces connaissances sont utilisées pour construire une représentation structurée et explicite du fonctionnement du réseau. La complexité de développement d'un système de diagnostic à base de modèle est due au fait qu'un réseau de télécommunications de grande échelle est très souvent hétérogène et dynamique avec un grand nombre d'équipements de différents types.

Dans (Łgorzata Steinder et Sethi 2004), les auteurs expliquent que la construction du modèle n'est que la première étape du développement d'une approche de diagnostic d'un réseau à base d'un modèle de ce réseau. La seconde étape consiste à développer ou à appliquer un algorithme sur le modèle. L'algorithme commence par les entités ayant déclenché des alarmes et explore les relations entre les entités du réseau formalisées par le modèle. Leur algorithme est ainsi capable de déterminer quelles alarmes sont corrélées et de localiser ainsi les entités incriminées du réseau.

Dans (Yu *et al*. 2009 ; Fan *et al.* 2012), les auteurs expliquent comment utiliser l'apprentissage automatique, et spécialement les réseaux de neurones artificiels pour la détection d'intrusion.

L'approche à base de modèle est facile à déployer et à modifier et est appropriée pour un réseau à grande échelle, si les informations relatives aux ressources du réseau sont disponibles.

3.4.4. *Big Data et réseaux de télécommunications*

3.4.4.1. *Big Data et amélioration du service client*

Les entreprises de télécommunication collectent d'énormes quantités de données à partir des enregistrements d'appels, de l'utilisation du téléphone mobile, des équipements réseau, des journaux de serveur, de la facturation et des réseaux sociaux, fournissant ainsi de nombreuses informations sur leurs clients et leur réseau. Avec la technologie Big Data, les entreprises de télécommunication vont utiliser ces données pour améliorer leur activité, grâce à l'utilisation d'analyses avancées.

Avec l'expansion rapide des téléphones intelligents et d'autres appareils mobiles connectés, les fournisseurs de services de communication doivent rapidement traiter,

stocker et tirer des informations du volume diversifié de données transitant sur leurs réseaux. Les analyses de données volumineuses peuvent :

– aider à améliorer la rentabilité en optimisant l'utilisation du réseau, en améliorant l'expérience client et en renforçant la sécurité ;

– prédire les périodes d'utilisation intensive du réseau et cibler les étapes pour réduire la congestion ;

– identifier les clients les plus susceptibles de faire défaut et cibler les étapes permettant d'éviter le roulement ;

– identifier les clients les plus susceptibles d'avoir des problèmes pour payer leurs factures et cibler les étapes pour améliorer le recouvrement des paiements.

En raison du volume considérable des données, il est important de les traiter près de la source, puis de les transférer efficacement vers divers centres de données pour une utilisation ultérieure.

L'analyse des événements en temps réel est la clé d'une analyse opportune des services réseau afin d'améliorer la satisfaction de la clientèle. On peut par exemple analyser les appels abandonnés, les endroits où la qualité de la couverture réseau est médiocre, un mauvais temps de téléchargement, un temps d'attente inacceptable, etc.

Dans les applications réseau, la clé d'une exploitation réussie du Big Data consiste à se concentrer sur les problèmes, et non sur les points de données.

En matière d'administration réseau, les données Big Data sont collectées à partir de sondes déployées en différents points, ainsi qu'au moyen d'un logiciel de couche réseau installé sur des équipements client et serveur. Lorsqu'elles sont présentées au sein d'une infrastructure système à administration standard, une partie de ces informations peut correspondre aux pratiques courantes de gestion.

3.4.4.2. *Big Data et sécurité*

Dans le Big Data, les données ont des volumes plus importants, mais surtout exponentiels, variables, et provenant de sources différentes.

Puisque l'entreprise peut avoir une vision sur l'ensemble de ces volumes de données qui transitent chaque jour sur son système d'information, elle peut, au lieu de se contenter d'attendre que les problèmes surgissent pour les traiter, tâcher de repérer tous les événements éventuellement annonceurs.

Une fois le risque identifié, il s'agira alors de mettre en place une protection afin de prévenir sa propagation. On pourra alors en déduire une vision proactive de la

sécurité informatique, d'autant que la précision des informations ainsi remontées permettra de mieux identifier les menaces en remontant directement à leur source.

3.4.5. *Systèmes multiagents et réseaux de télécommunications*

Le développement d'une ingénierie des connaissances permettant de ramener le traitement de l'information à un raisonnement sur des connaissances est la tâche principale des techniques agent. Ces dernières permettent aussi le développement de techniques de génie logiciel adaptées à la fourniture de services.

Le domaine des télécommunications offre des perspectives d'environnements ouverts, que ce soit au niveau du Web ou des futurs services réseau. Il permet en outre d'explorer les différentes techniques d'agent : les agents mobiles, les assistants web et les agents raisonnant sur des connaissances.

Il y a plusieurs décennies, lorsque des sociétés souhaitaient disposer d'un réseau de télécommunication privé, elles utilisaient une infrastructure de télécommunication qui leur était propre. Par la suite, ces demandes ont été satisfaites par des réseaux privés constitués de liaisons louées à un opérateur, l'exploitation de ces liaisons pouvant être sous-traitées à l'opérateur. Ces liaisons, qui sont en dehors des réseaux publics, garantissent une qualité de service, par exemple un débit requis entre plusieurs points donnés et une complète confidentialité des données échangées.

Les réseaux privés virtuels (VPN) sont des offres de réseaux privés mises en œuvre sur les réseaux publics. L'offre VPN permet de répondre à une demande croissante de connexions temporaires, la bande passante inutilisée par une entreprise à un instant donné étant potentiellement disponible pour une autre utilisation. Les SMA peuvent être utilisés pour automatiser l'approvisionnement de VPN nécessitant plusieurs fournisseurs de services réseau et pour automatiser la négociation des ressources réseau dans ce contexte.

Les agents désignent des composants logiciels contrôlant ou supervisant de façon décentralisée des ressources réseau. Ils sont utilisés pour développer des stratégies de coopération permettant de coordonner l'allocation ou la supervision de ressources dépendant de différentes autorités, ainsi que pour développer des stratégies de contrôle de la surcharge du réseau qui pourrait être engendrée par la signalisation liée aux nouveaux services.

3.4.5.1. *SMA et radio cognitive*

Dans (Mir 2011), l'auteur propose une coopération entre les PU et les SU, et entre les SU seulement. Des agents sont déployés sur les terminaux des utilisateurs

pour coopérer et aboutir à des contrats régissant l'allocation du spectre. Les agents SU coexistent et coopèrent avec les agents PU dans un environnement RC *ad hoc* en utilisant des messages et des mécanismes de prise de décision. Vu que les comportements internes des agents sont coopératifs et désintéressés, cela leur permet de maximiser la fonction d'utilité des autres agents sans ajouter de coût conséquent en termes de messages échangés.

Cependant, l'allocation des ressources est un enjeu important dans les systèmes de RC. Elle peut être faite en effectuant la négociation parmi les utilisateurs secondaires (Li 2009 ; Qian *et al.* 2011). Dans (Qian 2011), les auteurs proposent un modèle basé sur les agents pour la négociation du spectre dans un réseau RC. Dans le modèle proposé par les auteurs, au lieu de négocier le spectre directement entre des PU et des SU, un agent courtier est inclus. Ce qui veut dire que l'équipement du PU ou du SU ne nécessite pas une grande intelligence, vu qu'il n'a pas besoin d'effectuer la détection du spectre ou d'autres tâches plus compliquées de la RC. L'objectif de cette négociation est de maximiser les bénéfices et les profits des agents pour satisfaire le SU. Les auteurs ont proposé deux situations, la première utilise un seul agent qui va exploiter et dominer le réseau et, dans la deuxième, il va y avoir plusieurs agents en concurrence.

Une étude a été faite par (Xie *et al*. 2007) sur la RC dans les réseaux WLAN (*Wireless Local Area Network*) quant à la possibilité d'introduire la technologie d'agents ; en d'autres termes, ils essayent de résoudre le problème de l'allocation des ressources radio en associant la gestion des ressources WLAN dans un environnement décentralisé, ceci en utilisant les SMA. Pour cela, ils proposent une approche basée sur les agents pour le partage d'information et la distribution des décisions parmi de multiples WLAN, d'une manière distribuée.

Dans (Amraoui 2015), une architecture multiagents est proposée, composée de trois niveaux : le premier est le niveau physique, où les auteurs ont donné quelques remarques sur le type de terminal utilisé, vient ensuite le niveau cognitif, où ils ont proposé un cycle de cognition modifié sur la base des SMA, et enfin le niveau comportemental, où ils ont étudié les différents comportements que peuvent avoir les agents au moment de la négociation du spectre.

3.4.5.2. *SMA et réseaux de transport*

La transposition des notions orientées agent au domaine de transport est en accord avec les caractéristiques propres aux deux domaines. On y retrouve en effet des caractéristiques d'autonomie, de comportements distribués et d'environnement partiellement observable. Les approches existantes s'intéressent aux propriétés des systèmes

multiagents : émergence, auto-organisation, coopération. De plus, l'évolution d'un grand nombre de véhicules sur un réseau routier partagé correspond parfaitement aux problématiques de conflit de ressources étudiées *via* les SMA (Guériau 2016).

Les systèmes de transports intelligents offrent un ensemble d'outils s'appuyant sur les dernières avancées en termes de puissance de calculs, de communication et de perception, afin de produire un système supervisé, intégré, universel et abordable. Les SMA semblent constituer la meilleure voie d'amélioration, à la fois en termes d'interaction et de calcul distribué.

La gestion des intersections peut être aussi améliorée avec l'utilisation des SMA. En effet, les agents peuvent être déployés au niveau de l'intersection ou pour chaque feu, et leur coopération permet d'optimiser les cycles face à la demande.

Les SMA peuvent être également pratiques pour la régulation du trafic et la gestion de la congestion, où les agents peuvent coopérer ou négocier afin d'assurer une meilleure gestion du trafic routier et un réseau de transport plus intelligent.

3.4.6. *Internet des objets et réseaux*

Pour s'adapter en temps réel à une situation donnée, les appareils connectés doivent comprendre la valeur de l'information qu'ils recueillent et apprendre les uns des autres. Grâce à l'IA et à l'analytique, ils prendront alors des décisions adaptées au sein de systèmes autonomes.

Aujourd'hui, les objets que nous utilisons dans nos maisons, nos bureaux, nos hôpitaux et nos usines sont en phase de connexion. En leur donnant des capacités d'apprentissage autonome et de personnalisation, l'IA les fera entrer dans une phase de disruption.

La combinaison (IoT + IA) provoque de véritables changements de perceptives. Les secteurs de la sécurité, de la santé, de l'industrie et de l'énergie peuvent bénéficier des avantages qu'offre cette combinaison.

Un exemple d'application de la combinaison IA + IoT dans le domaine de la sécurité est par exemple les logiciels de sécurité incrustés dans les caméras connectées dotées de *computer vision* capables de repérer une personne dans la foule et de prévenir les autorités compétentes grâce aux techniques de reconnaissance de forme**.** Le logiciel permet notamment de compter le nombre de personnes dans une

pièce, de repérer un criminel fiché ou une personne recherchée, d'autoriser l'entrée d'une personne dans des zones confidentielles, etc.

Dans le domaine de la santé, les objets connectés combinés à l'IA ont leur place dans la plupart des situations. Ils permettent, par exemple, grâce à une caméra connectée ou à une paire de lunettes connectée, de repérer les symptômes d'une maladie. Les capteurs connectés peuvent transmettre les données vitales des patients à une plateforme dotée d'IA, qui fera appel à un infirmier ou à un aide-soignant en cas d'alerte et pourra lui fournir tous les détails de l'événement.

Dans (Srinidhi *et al*. 2018), les auteurs proposent plusieurs algorithmes issus de l'intelligence artificielle pour optimiser les réseaux avec IoT. En effet, ils abordent plusieurs types d'algorithmes, tels que les algorithmes génétiques où l'on utilise des critères multi-objectifs pour sélectionner les meilleurs capteurs ayant un espace de stockage maximal.

3.5. Conclusion

Avec l'accélération et la transformation des réseaux de télécommunication grâce aux nouvelles technologies, les opérateurs doivent améliorer l'efficacité de leurs services tout en réduisant les coûts.

Nous pensons que l'utilisation de l'intelligence artificielle et de la science des données peut améliorer les performances, la fiabilité et la sécurité des réseaux.

En effet, avec l'IA, le réseau sera en mesure de réagir automatiquement à toute surcharge importante susceptible de survenir. Il deviendra possible pour le réseau de détecter une surcharge, de créer automatiquement le nombre de machines virtuelles nécessaires pour gérer la quantité de trafic entrant.

Le diagnostic de pannes dans un réseau de télécommunications de grande échelle est un problème complexe qui intéresse aussi bien les opérateurs de télécommunications que la communauté de l'intelligence artificielle. Ce problème a fait l'objet de nombreuses recherches et différentes approches ont été proposées à base de systèmes experts, de systèmes de raisonnement, à partir de cas et d'apprentissage automatique.

Nous pensons qu'à l'avenir les réseaux de télécommunications vont devenir complètements autonomes et ne dépendront plus de l'intervention humaine, grâce à l'IA et surtout grâce aux technologies Big Data.

3.6. Bibliographie

Amraoui, A. (2015). Vers une architecture multiagents pour la radio cognitive opportuniste. Thèse de doctorat, Université de Tlemcen, Algérie.

Benmammar, B. (2017). Optimisation de la QoS dans un réseau de radio cognitive en utilisant la métaheuristique SFLA (Shuffled Frog Leaping Algorithm) [En ligne]. Disponible à l'adresse : arXiv:1703.07565.

Benmammar, B., Amraoui, A., Baghli, W. (2012). Performance improvement of wireless link reliability in the context of cognitive radio. *IJCSNS International Journal of Computer Science and Network Security*, 12(1), 15–22.

Cigref (2018). Cybersécurité : visualiser, comprendre, décider [En ligne]. Rapport. Disponible à l'adresse : https://www.cigref.fr/wp/wp-content/uploads/2018/10/Cigref-Rapport-Cybersecurite-Visualiser-Comprendre-Decider-Octobre-2018. pdf.

Decourt, O. (2018). Les réseaux de neurones expliqués à ma fille [En ligne]. Disponible à l'adresse : https://od-datamining.com/knwbase/les-reseaux-de-neurones-expliques-a-ma-fille/.

Fan, W., Bouguila, N., Ziou, D. (2012). Variational learning for finite Dirichlet mixture models and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 23(5).

Ferber, J., Perrot, J.-F. (1995). *Les systèmes multiagents : vers une intelligence collective*. InterEditions, Paris.

Guériau, M. (2016). Systèmes multiagents, auto-organisation et contrôle par apprentissage constructiviste pour la modélisation et la régulation dans les systèmes coopératifs de trafic. Thèse de doctorat, Université Claude Bernard Lyon 1.

Hounkonnou, C. (2013). Active self-diagnosis in telecommunication networks. Thèse de doctorat, Université européenne de Bretagne et Université de Rennes 1.

Kaushik, A., Barnela, M., Khanna, S. *et al*. (2011). A Novel Expert System for PC Network Troubleshooting and Maintenance. *International Journal of Advanced Research in Computer Science*, 2(3).

Łgorzata Steinder, M., Sethi, A.S. (2004). A survey of fault localization techniques in computer networks. *Science of computer programming*, 53(2), 165–194.

Li, H. (2009). Multiagent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case. Dans *International Conference on Systems, Man and Cybernetics*.

Lor, K.W.E. (1993). A network diagnostic expert system for Acculink multiplexers based on a general network diagnostic scheme. Dans *Proceedings of the 3rd International Symposium on Integrated Network Management with participation of the IEEE Communications Society CNOM and with support from the Institute for Educational Services*, 659–669.

Melchiors, C., Tarouco, L.M.R. (1999). Fault management in computer networks using case-based reasoning: DUMBO system. Dans *International Conference on Case-Based Reasoning*, 510–524.

Mell, P., Grance, T. *et al*. (2011). The NIST definition of cloud computing [En ligne]. Disponible à l'adresse : https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

Mir, U. (2011). Utilization of Cooperative Multiagent Systems for Spectrum Sharing in Cognitive Radio Networks. Thèse de doctorat, Université de technologie de Troyes, Troyes.

Nowicki, K., Uhl, T. (2017). QoS/QoE in the Heterogeneous Internet of Things (IoT). Dans *Beyond the Internet of Things*, Batalla, J.M., Mastorakis, G., Mavromoustakis, C.X., Pallis, E. (dir.). Springer, Bâle, 165–196.

Qian, L., Ye, F., Gao, L. *et al*. (2011). Spectrum trading in cognitive radio networks: an agent-based model under demand uncertainty. *IEEE Transactions on Communications*, 59(11), 3192–3203.

Raja'a, A.K., Jassim, R.O. (2014). Expert system to troubleshoot the wireless connection problems. *International Journal of Science, Engineering and Computer Technology*, 4(8), 238.

Shiffman, D. (2012). *The Nature of Code: Simulating Natural Systems with Processing*. Daniel Shiffman.

Srinidhi, N.N., Dilip Kumar, S.M., Venugopal, K.R. (2018). Network optimizations in the Internet of Things: A review. *Engineering Science and Technology*, 22(1), 1–21.

Xie, J., Howitt, I., Raja, A. (2007). Cognitive radio resource management using multiagent systems. Dans *4th IEEE Consumer Communications and Networking Conference*.

Yu, W., He, H., Zhang, N. (2009). Advances in Neural Networks. Dans *6th International Symposium on Neural Networks*.

4

Méthodes d'optimisation multicritères pour la sélection de réseaux dans un environnement hétérogène

Fayssal BENDAOUD
ESI-SBA, Sidi Bel Abbès, Algérie

4.1. Introduction

Le secteur des télécommunications a connu une évolution explosive au cours de la dernière décennie, ce qui lui a permis d'avoir une place importante et cruciale dans les sociétés actuelles. L'innovation la plus importante dans le domaine des télécommunications est sans aucun doute les technologies « sans-fil ». La possibilité de se connecter avec des ondes hertziennes a donné naissance à plusieurs types de réseaux sans-fil. Ces réseaux sont en concurrence pour attirer l'attention de l'utilisateur en offrant des performances plus élevées, surtout en termes de qualité de service QoS. Du point de vue du modèle TCP/IP, la couche 3 est la plus importante, car elle est fusionnée avec les couches 1 et 2 pour devenir une couche appelée « couche IP ». Par conséquent, IP doit prendre en charge les fonctionnalités-clés associées aux réseaux sans-fil, notamment la gestion du transfert inter/intra cellulaire *handover* et celle de la localisation GPS et de la mobilité. En outre, en raison de la diversité des choix de technologies d'accès radio proposés par de nombreux fournisseurs de réseaux cellulaires/sans-fil, et prenant en considération les exigences de qualité de service des applications moderne, choisir un type de réseau s'avère une tâche très compliquée. De ce fait, une approche unique et unifiée dans le

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

choix de réseau s'impose, et on parle donc de la sélection de réseau dans un environnement hétérogène. Un environnement sans-fil hétérogène est un ensemble d'équipements sans-fil utilisant différentes technologies d'accès radio. Dans un environnement hétérogène, un utilisateur n'est pas lié à un seul réseau, mais il peut plutôt profiter des avantages de tous les réseaux existant dans son entourage. Un tel environnement présente plusieurs avantages par rapport aux réseaux sans-fil homogènes traditionnels, notamment :

– une meilleure QoS, puisque à chaque instant une application lance l'algorithme de choix de réseau et connecte l'utilisateur à ce réseau ;

– une disponibilité du service puisque, si une technologie d'accès radio disparaît, l'application connecte l'utilisateur à un autre réseau ;

– un équilibrage de charge des réseaux et une utilisation efficace du spectre : à chaque instant, l'application met l'utilisateur dans le réseau le moins chargé pour espérer avoir une bonne QoS.

Aucune technologie d'accès radio ne peut seule assurer ces avantages. Cependant, il reste plusieurs problèmes à résoudre dans un environnement sans-fil hétérogènes, tels que :

– l'interopérabilité entre les réseaux et la sélection de réseaux ;

– la gestion de la mobilité ;

– le passage inter/intra cellulaire (le *handover*) ;

– la qualité de service/d'expérience ;

– l'interférence entre les réseaux.

Dans ce chapitre, nous allons mettre l'accent principalement sur la sélection de réseau dans un environnement hétérogène et la qualité de service. L'objectif principal pour un environnement hétérogène est de concrétiser le concept « Always Best Connected », en offrant aux utilisateurs mobiles la possibilité de tirer profit des réseaux ayant des performances différentes. De nos jours, nous avons une variété de technologies d'accès radio, les WLAN (essentiellement IEEE802.11) et les réseaux cellulaires (UMTS, HSPA et le LTE). Cette variété constitue l'environnement hétérogène. La procédure de sélection de réseau consiste à sélectionner le meilleur réseau parmi ceux qui sont disponibles. Cependant, le grand nombre de paramètres impliqués dans le processus de sélection, tels que le coût d'utilisation d'un tel réseau, la qualité de service, l'énergie consommée, rend la décision très difficile et complexe. Cela nous permet de dire que la définition de base du problème de sélection de réseau est le choix dynamique et automatique du meilleur réseau d'accès sans-fil, en tenant compte des paramètres que nous avons cités. Dans les

systèmes cellulaires classiques, l'utilisateur mobile est lié par un contrat avec un seul réseau, la sélection du réseau n'existe même pas. La nouvelle vision des réseaux stipule que l'utilisateur peut choisir à tout moment le réseau idéal pour lui et l'utiliser pour sa session sans qu'un contrat l'oblige à rester connecté à un seul réseau (figure 4.1).

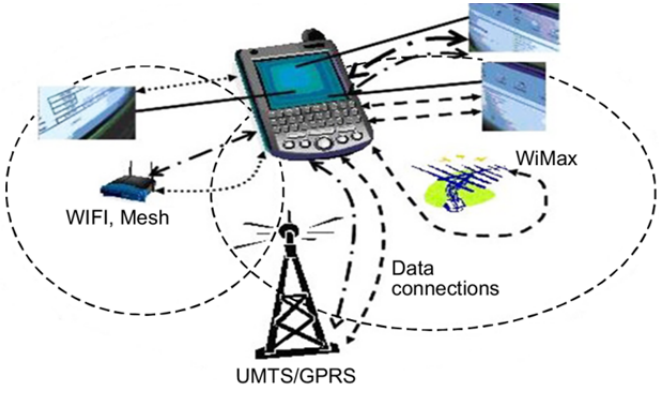


Figure 4.1. *Un environnement hétérogène (Bendaoud 2018)*

Dans la section 4.2, nous allons parler des méthodes d'optimisation multicritères ; la section 4.3 proposera une modification d'une des méthodes d'optimisation multicritères (accompagnée d'une motivation de ce choix et d'une comparaison entre la méthode modifiée et celle sans modification) ; la section 4.4 sera consacrée à une conclusion ouvrant sur un ensemble de perspectives.

4.2. Optimisation multicritères et sélection de réseaux

Jusqu'aux réseaux de troisième génération, la notion de sélection de réseau n'existait même pas, puisque l'utilisateur était lié par un contrat avec un seul réseau. Par la suite, le développement rapide des technologies de réseaux a entraîné une croissance impressionnante des services et des applications qui utilisent Internet et, en même temps, il a induit une augmentation des utilisateurs, surtout mobiles. Actuellement, les utilisateurs équipés de périphériques intelligents cherchent à avoir le concept ABC. Il est évident qu'aucune technologie d'accès radio ne peut offrir l'ABC seule ; il est donc primordial de changer les idées, c'est-à-dire de passer des systèmes homogènes (utilisateurs liés par un contrat avec un seul réseau) aux systèmes hétérogènes. L'objectif d'un environnement hétérogène est de concrétiser le concept ABC en offrant aux utilisateurs mobiles la possibilité de tirer profit des réseaux qui ont des performances différentes.

La sélection de réseau dans un environnement hétérogène peut être comparée aux problèmes d'optimisation multicritères en raison du nombre de paramètres impliqués et des critères à respecter. Cette approche mathématique a été largement utilisée pour résoudre le problème de sélection de réseau (Kovvali *et al.* 2015 ; Scherzer et Scherzer 2015 ; Wu *et al.* 2015 ; Bendaoud *et al.* 2018b). D'autres méthodes, telles que la logique floue et la théorie des jeux, ont été proposées pour traiter ce problème (Watanabe *et al.* 2008 ; Alkhawlani et Ayesh 2008). Dans cette section, nous allons nous focaliser surtout sur l'utilisation des méthodes d'optimisation multicritères pour la résolution du problème de la sélection du meilleur réseau.

4.2.1. *Le processus de sélection de réseaux*

Comme nous l'avons déjà dit, le processus de sélection de réseau consiste à basculer entre les technologies d'accès radio pour être toujours mieux servi, à n'importe quel moment. Ainsi, lorsqu'un utilisateur avec un appareil multimode découvre l'existence de plusieurs réseaux dans sa zone, il devrait pouvoir sélectionner le meilleur réseau pour lui parmi ceux présents et l'utiliser pour la durée de sa session. Les différents réseaux offrent plusieurs caractéristiques en termes de délai, de gigue, de débit et de taux de paquets perdus. Pour cette raison, de nos jours, il est primordial de bien sélectionner le meilleur réseau pour l'efficacité du système (figure 4.2).

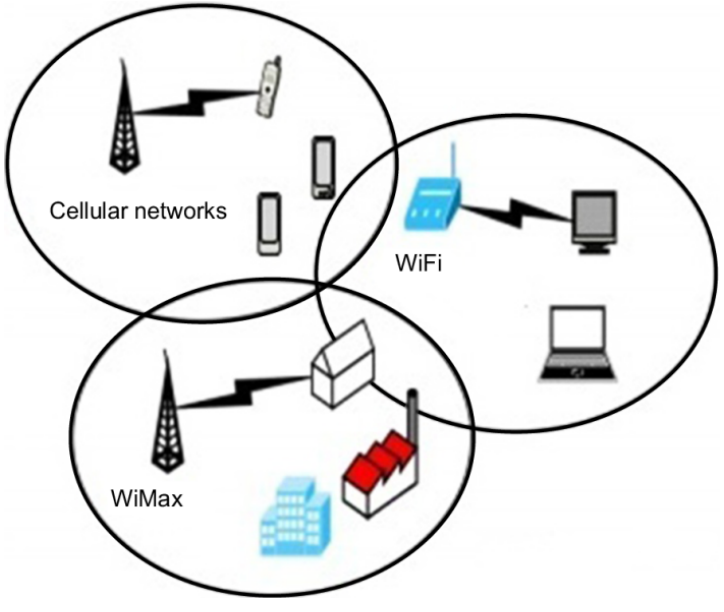


Figure 4.2. *Environnement hétérogène (Bendaoud 2018)*

Il est connu que la procédure de sélection de réseau est le cas général du processus de transfert intercellulaire *handover* ; celui-ci peut être centralisé, c'est-à-dire effectué par l'opérateur, ou décentralisé, c'est-à-dire axé sur l'utilisateur.

Pour l'approche centrée sur le réseau, l'opérateur contrôle la totalité du processus et prend des décisions. Les utilisateurs obéissent à ces décisions et les exécutent. Cela peut être considéré comme une bonne stratégie pour éviter des problèmes tels que le comportement égoïste des utilisateurs, qui essaient tous d'obtenir le meilleur réseau en même temps, ce qui entraîne une situation de congestion du réseau sollicité. De plus, cette approche suppose une situation avec un opérateur unique et plusieurs réseaux ; cette approche ne peut pas être utilisée dans le cas de plusieurs opérateurs. Pour l'approche centrée sur l'utilisateur, les utilisateurs prennent les décisions eux-mêmes ; cette approche est décentralisée et peut facilement générer une situation de congestion, en raison de la nature égoïste des utilisateurs. De nos jours, presque tous les opérateurs offrent un accès radio 3G et 4G, en plus des connexions Wi-Fi ; il est donc préférable d'utiliser la première approche (centralisée).

Nombreux sont les paramètres qui influencent le processus de décision du meilleur réseau, comme l'état de la batterie, l'énergie nécessaire pour obtenir les services demandés, le RSS reçu (*Received Signal Strength*), le coût à payer pour l'utilisation d'un réseau, la bande passante allouée, les préférences de l'utilisateur (l'utilisateur préfère que la qualité soit excellente ou bien il cherche l'équilibre entre la qualité et le prix à payer), etc. Ces paramètres sont classés en différentes classes et différents ensembles (figure 4.3).

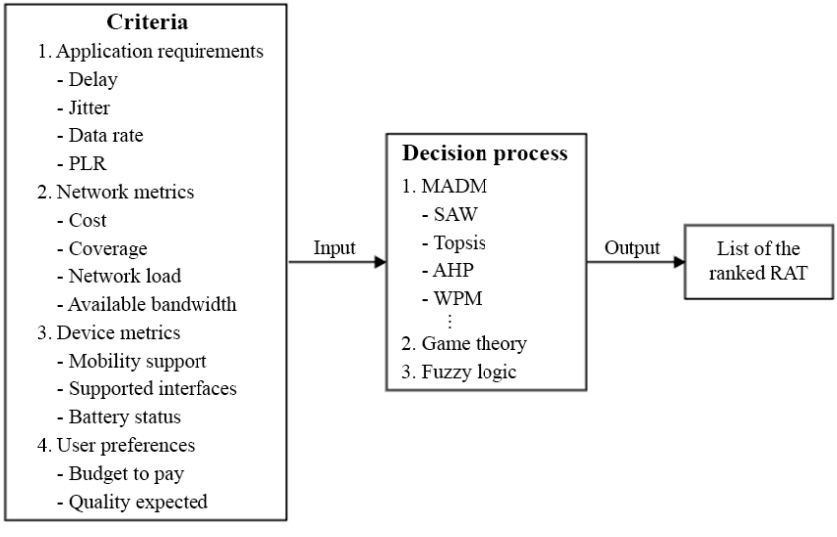


Figure 4.3. *Processus de sélection de réseau (Bendaoud et al. 2018b)*

Les paramètres impliqués dans le processus de sélection de réseau sont détaillés comme suit :

– paramètres liés aux conditions de réseau : ils regroupent des informations sur les conditions de réseau, telles que la charge du réseau, la zone de couverture (radius), le temps de connexion au réseau, la bande passante disponible, etc. ;

– paramètres relatifs aux exigences de l'application : ils correspondent aux informations sur le seuil nécessaire à l'application pour être à l'état normal. On peut citer entre autres : le débit requis, le délai de transmission des paquets, la gigue, le taux de perte de paquets et l'énergie nécessaire à l'application. Ce sont les paramètres de la qualité de service QoS ;

– paramètres liés aux préférences utilisateur : ils représentent les informations nécessaires aux utilisateurs de l'application ; ils incluent entre autres le budget que l'utilisateur est prêt à payer, la préférence entre le coût et la qualité de service ;

– paramètres de l'équipement mobile : ils regroupent les informations sur le périphérique de l'utilisateur, comme l'état de la batterie du mobile à l'instant t et la gestion de mobilité.

Compte tenu de tous ces paramètres impliqués et de leurs classifications, on peut dire que le problème de la sélection de réseau est une tâche très complexe. Dans la littérature, plusieurs travaux ont traité le problème de la sélection de réseau ; ces études se concentrent principalement sur l'optimisation de la décision de sélection de réseau pour les utilisateurs, afin que les applications soient servies avec la meilleure qualité de service possible à l'instant t et afin de maintenir les utilisateurs connectés à ce meilleur réseau aussi longtemps que possible. Plusieurs auteurs ont traité le problème de la sélection du meilleur réseau dans le but de trouver la solution la plus appropriée. Nous allons présenter les méthodes d'optimisation multicritères les plus connues et les plus utilisées pour la résolution de ce problème.

4.2.2. *Méthodes d'optimisation multicritères pour la sélection de réseaux*

L'optimisation multicritères est une branche des mathématiques traitant spécifiquement des problèmes d'optimisation ayant plusieurs fonctions objectifs, plusieurs critères, plusieurs alternatives, mais les objectifs à optimiser portent sur un seul problème. Tout le monde utilise cette approche dans la vie ordinaire, par exemple pour acheter une voiture qui a plusieurs caractéristiques. Dans la matrice de décision, les lignes représentent les critères et les colonnes sont les alternatives. Les valeurs a_{ij} décrivent les performances des alternatives A_j par rapport aux critères C_i. Dans la table de décision, les poids $w_1, \dots, w_m$ sont affectés aux critères, la

pondération w_i reflète l'importance relative du critère C_i dans le processus de la décision. Les poids des critères sont généralement déterminés sur une base subjective : ils représentent l'opinion d'un seul ou de plusieurs experts. Par contre, il existe des méthodes objectives permettant de déterminer les poids des critères. Ils seront abordés dans la suite du chapitre. Cette approche est très adaptée au problème de sélection de réseau, en raison de la nature multicritère de ce dernier (Sgora *et al*. 2010). La base de cette approche est la suivante :

– alternatives : c'est l'ensemble des acteurs qui seront classés. Dans le cas de la sélection des réseaux, les alternatives sont les listes des réseaux ;

– attributs : ils représentent les paramètres et les critères utilisés dans le processus de prise de décision. Dans le cas de la sélection des réseaux, les paramètres sont le débit, la gigue et le délai ;

– pondération : ceci traduit l'importance du paramètre dans le processus de décision.

Plusieurs méthodes ont été proposées dans ce contexte, telles que *Simple Additive Weight* (SAW), *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS), *Weighted Product Model* (WPM), *Analytical Hierarchy Process* (AHP), etc. (Salih *et al*. 2015). Les méthodes SAW, TOPSIS et WPM classifient seulement les alternatives ; elles nécessitent donc d'autres méthodes pour la pondération des critères. En revanche, la méthode AHP est à la fois une méthode de classement et de pondération ; elle contient un processus qui effectue les pondérations des critères. Dans le tableau 4.1, nous présentons quelques méthodes de pondération parmi les plus connues.

Méthode	Description
Entropy	$w_j = 1 - \frac{1}{N} * \sum_{i=1}^{N} x_{ij} * \ln x_{ij}$
Variance	$w_j = \frac{1}{N} * \sum_{i=1}^{N} x_{ij}$
Eigenvector	$(\mathrm{B} - \lambda \times \mathrm{I}) \times \mathrm{w} = 0$

Tableau 4.1. *Méthodes de pondération*

4.2.2.1. Simple Additive Weight *(SAW)*

La méthode SAW, également connue sous le nom de « méthodes mathématiques d'analyse multicritère », est une technique de décision multicritères simple et parmi

les plus utilisées, basée sur le score moyen pondéré. Cette méthode suppose que les données traitées ont la même unité. Par conséquent, pour être homogène, il est obligatoire de normaliser les données pour chaque paramètre (Savitha et Chandrasekar 2011a ; Abdullah et Adawiyah 2014). Enfin, l'alternative ayant la valeur la plus élevée ou la plus basse (suivant la fonction objectif) est sélectionnée. La formulation mathématique de SAW est la suivante :

$$R_{SAW} = \sum_{i=1}^{n} w_j * r_{ij} \qquad [4.1]$$

avec :

– R_{SAW} : valeur de chaque alternative ;

– w_j : valeur de pondération du critère j ;

– r_{ij} : valeur normalisée du critère j et de l'alternative i.

Une fois le processus achevé, nous choisissons la valeur maximale/minimale en fonction de la tendance de la fonction objectif (à maximiser ou à minimiser).

Depuis sa création, SAW a été très utilisée dans le contexte du problème de sélection de réseau (Nguyen-Vuong *et al*. 2008 ; Savitha et Chandrasekar 2011a). La méthode est utilisée comme mécanisme de sélection et de classement des réseaux dans (Salih *et al*. 2015) ; dans (Nguyen-Vuong *et al*. 2008 ; Savitha et Chandrasekar 2011b), elle est utilisée avec la théorie de jeu pour classer les réseaux, puisque les auteurs ont compris que SAW toute seule ne permet pas de donner de meilleures performances pour tous les cas. Nous reviendrons en fin de section sur les problèmes des méthodes d'optimisation multicritères, et notamment sur ceux de la méthode SAW.

4.2.2.2. Technique for Order Preference by Similarity to Ideal Solution *(TOPSIS)*

La technique de l'ordre de préférence par similarité avec la solution idéale (TOPSIS) est une méthode d'analyse décisionnelle multicritères ; elle a été proposée à l'origine par Hwang et Yoon en 1981, puis développée par Yoon en 1987 et par Hwang, Lai et Liu en 1993. TOPSIS est une méthode d'agrégation et de compensation basée sur le concept selon lequel la solution choisie doit avoir la distance géométrique la plus courte de la solution idéale positive et la plus grande distance géométrique de la solution idéale négative (Olson 2004). Le processus TOPSIS s'effectue comme suit :

– création d'une matrice d'évaluation composée de *m* alternatives et de *n* critères ; l'intersection des alternatives avec les critères est représentée sous la forme x_{ij}, nous avons donc une matrice $(x_{ij})_{m*n}$;

– la matrice $(x_{ij})_{m*n}$ est normalisée pour obtenir les valeurs $(r_{ij})_{m*n}$ en utilisant une des méthodes de normalisation ;

– calcul de la matrice de décision normalisée pondérée, avec :

$$t_{ij} = w_j * (r_{ij})_{m*n}$$

$w_j = \frac{W_j}{\sum_{k=1}^{n} W_k}$ avec $\sum_{i=1}^{n} w_i$ et W_k est le vecteur poids des critères

– détermination des meilleures et des mauvaises alternatives :

$$A_b = \left\{ b_j^p = \max t_{ij} \,\middle|\, j \in J_+ \right\}$$

$$A_{wo} = \left\{ wo_j^N = \min t_{ij} \,\middle|\, j \in J_- \right\}$$

avec J_+ et J_- : deux ensembles contenant des critères avec impact positif et négatif respectivement ;

– calcul de la mesure de séparation pour chaque alternative :

$$D_p = \sqrt{\sum_{i=1}^{n} (w_j^2 * (r_{ij} - b_j^p)^2)}$$

$$D_N = \sqrt{\sum_{i=1}^{n} (w_j^2 * (r_{ij} - wo_j^N)^2)}$$

– calcul de solution relative idéale :

$$R_{TOPSIS} = D_p D_p + D_N$$

TOPSIS a été utilisée dans plusieurs travaux en littérature, notamment en (Bakmaz *et al*. 2007 ; Sgora *et al*. 2010 ; Savitha et Chandrasekar 2011a). Dans ce dernier article, les auteurs comparent les performances du *handover* vertical avec deux méthodes, TOPSIS et SAW ; ils ont constaté que TOPSIS est meilleure que SAW. Dans (Bakmaz *et al.* 2007), les auteurs réalisent la sélection de réseau *via* la méthode TOPSIS et ils disent que TOPSIS est sensible aux préférences des utilisateurs et aux valeurs des paramètres en entrées.

4.2.2.3. Weighted Product Model *(WPM)*

La méthode WPM, appelée également « méthode de pondération exponentielle multiplicative » est similaire à SAW (Lahby *et al*. 2012). La différence réside dans le remplacement de l'opération d'addition dans la méthode SAW par l'opération de

multiplication dans la méthode WPM, en plus du fait que chaque alternative est comparée aux autres critères de décision. Chaque rapport est élevé à une puissance équivalente au poids relatif du critère correspondant. La description mathématique de cette méthode est la suivante :

$$R_{WPM} = \prod_{i=1}^{n} (r_{ij})^{w_j} \quad [4.2]$$

En fonction du choix de r_{ij} , deux variantes de WPM se présentent :

$$r_{ij} = x_{ij} \; ou \; r_{ij} = \frac{x_{ik}}{x_{jk}} \quad [4.3]$$

Les auteurs de (Savitha et Chandrasekar 2011b) ont comparé les méthodes SAW et WPM dans le contexte du *handover* vertical : ils utilisent l'écart type relatif comme métrique de comparaison entre les deux méthodes et ils concluent que WPM est meilleur que la méthode SAW. Dans (TalebiFard et Leung 2011), les auteurs utilisent la méthode WPM pour la sélection de réseau ; la conclusion du travail est que la méthode WPM constitue une approche plus robuste pour la prise de décision dynamique.

4.2.2.4. Analytical Hierarchy Process *(AHP) et* Grey Relational Analysis *(GRA)*

AHP considère la décomposition d'un problème compliqué en plusieurs sous-problèmes simples. Les étapes AHP sont les suivantes :

– la décomposition du problème en sous-problème hiérarchique, où le nœud supérieur est l'objectif final et, pour chaque critère, nous listons les alternatives ;

– la comparaison paire à paire des attributs et leur transformation en valeurs numériques de 1 à 9 ;

– le calcul des poids (pondération) pour chaque niveau de la hiérarchie ;

– la synthèse des poids et l'obtention du poids global.

En ce qui concerne la méthode GRA, elle est utilisée pour classer les réseaux candidats. Son principe est le suivant :

– la normalisation des données est effectuée en tenant compte de trois situations : la plus élevée est la meilleure, la faible est la mauvaise et la nominale (modérée) est la demandée ;

– la définition de la séquence idéale dans les trois situations considérées : la séquence idéale contient la limite supérieure, la limite inférieure et la limite modérée ;

– le calcul du coefficient relationnel (GRC) : la séquence la plus favorable est celle avec le GRC le plus grand.

La méthode AHP est généralement associée à la méthode *Gray Relational Analysis* (GRA) ; souvent AHP est utilisée pour la pondération (poids des alternatives) et la méthode GRA est utilisée pour le classement de ses alternatives. Les auteurs de (Shen *et al*. 2010) ont utilisé une version modifiée de AHP et l'ont comparée à la AHP traditionnelle en utilisant un critère de qualité d'expérience (QoE) ; enfin, leurs résultats numériques montrent que la AHP modifiée proposée est meilleure par rapport à la AHP conventionnelle, ce qui résulte d'un bon équilibre de charge pour les réseaux. Dans (Lin et Hsu 2003), les auteurs utilisent AHP pour classer différents critères pour les réseaux de publicité en ligne ; le travail proposé fournit un modèle de décision objectif et efficace que les agences publicitaires doivent utiliser pour choisir un réseau de publicité sur Internet.

4.2.2.5. *Discussion*

En résumé, les méthodes d'optimisation multicritères sont largement utilisées pour résoudre le problème de sélection de réseau, puisque cette dernière a le même schéma et les mêmes caractéristiques que les problèmes résolus par ce type de méthodes. De plus, ces méthodes sont connues pour leur facilité d'utilisation, leur clarté et leur faible complexité de calcul. Néanmoins, elles ont quelques inconvénients, que nous allons citer dans ce qui suit :

– ces méthodes n'ont pas les mêmes performances vis-à-vis des différents services (VoIP, appels vidéo et service *best effort*). Par exemple, une méthode peut produire de bonnes performances avec le service VoIP et une mauvaise performance pour des services vidéo, ce qui n'est pas idéal ;

– ces méthodes souffrent du problème de renversement du rang : il s'agit d'un phénomène qui se produit dans les méthodes d'optimisation multicritères lorsqu'une réplique exacte ou une copie d'une alternative a été ajoutée ou éliminée. Les auteurs de (Wang et Luo 2009) ont montré que le problème d'inversement de rang est fréquent dans la plupart des méthodes d'optimisation multicritères connues. Ce problème a été abordé dans d'autres travaux (Huszak et Imre 2010 ; Shin *et al*. 2013) en proposant des modifications de ces méthodes, mais les versions originales des méthodes souffrent du phénomène d'inversion de rang ;

– la méthode AHP est très compliquée et nécessite un calcul très complexe lors du calcul du vecteur de pondération (calcul des valeurs propres).

Pour toutes ces raisons, on peut dire que les méthodes d'optimisation multicritères sont une bonne solution, mais l'absence d'une méthode qui pallie aux aspects évoqués précédemment pose problème.

Avantages	Inconvénients
Facile à comprendre	Phénomènes d'inversion de rang
Facile à implanter	Complexité élevée pour certaines méthodes telles que AHP et ELECREE
Très bons résultats dans certains cas	Bonne performance avec certaines applications et mauvais résultats pour d'autres

Tableau 4.2. *Avantages et inconvénients des méthodes d'optimisation multicritères*

4.3. « Modified-SAW » pour la sélection de réseaux dans un environnement hétérogène

Dans cette section, nous présentons une méthode appelée « modified-SAW» (Bendaoud *et al*. 2017) pour faire face aux problèmes des solutions existantes du problème de sélection de réseaux. En effet, les méthodes existantes d'optimisation multicritères souffrent toutes principalement du fameux problème d'inversement du rang une fois qu'une alternative est ajoutée ou supprimée ; d'autres problèmes se posent pour ces méthodes, telles que la sensibilité à la préférence de l'utilisateur (notamment dans la méthode TOPSIS), la pénalisation des alternatives avec de mauvaises valeurs pour certains attributs WPM. La méthode proposée permet d'éviter les problèmes cités auparavant et, en même temps, elle offre l'opportunité de surpasser et de donner de bonnes performances par rapport aux autres méthodes.

On rappelle que, dans la vision des réseaux de nouvelle génération, l'accès sans-fil hétérogène est une caractéristique prometteuse dans laquelle les utilisateurs ont la capacité « ils sont suffisamment flexibles » pour sélectionner le réseau le plus approprié en fonction de leurs besoins. Donc, la sélection du réseau a une tâche importante pour le bon fonctionnement de tout le système de communication hétérogène. En effet, le processus de sélection de réseau consiste à basculer entre les RAT pour servir l'utilisateur du meilleur réseau. Ainsi, lorsqu'un utilisateur possédant un terminal multimode découvre l'existence de plusieurs RAT dans la même zone, il doit pouvoir sélectionner le meilleur réseau pour obtenir le service souhaité (figure 4.2).

4.3.1. *Méthode proposée « modified-SAW »*

La procédure de sélection de réseau proposée, « modified-SAW » (Bendaoud *et al*. 2017), a pour but de garantir à l'utilisateur une bonne qualité de service

pendant la session d'appel et, en même temps, de garantir la répartition correcte des utilisateurs sur chaque réseau. La solution proposée assigne à l'utilisateur le meilleur réseau parmi les réseaux disponibles à l'instant actuel et que le réseau sélectionné est accessible, c'est-à-dire non surchargé. Ce processus est répété à plusieurs reprises jusqu'à la fin de la session d'appel de l'utilisateur.

Ainsi, lorsqu'un utilisateur veut utiliser un service particulier, il envoie une demande à l'opérateur. Cette demande contient des informations telles que le service requis et le niveau de la batterie de l'utilisateur. Les autres paramètres nécessaires à la procédure de sélection du réseau sont recueillis par l'opérateur. Ensuite, ce dernier déclenche le processus de classement des réseaux disponibles. Les résultats seront transmis à l'utilisateur, qui sélectionne le meilleur réseau disponible. Évidemment, il choisira le réseau ayant le meilleur rang puisqu'un réseau surchargé ne peut normalement pas obtenir le meilleur rang, simplement parce qu'un réseau chargé a un délai plus long et un débit inférieur ; il a donc une mauvaise performance en termes de qualité de service (figure 4.4).

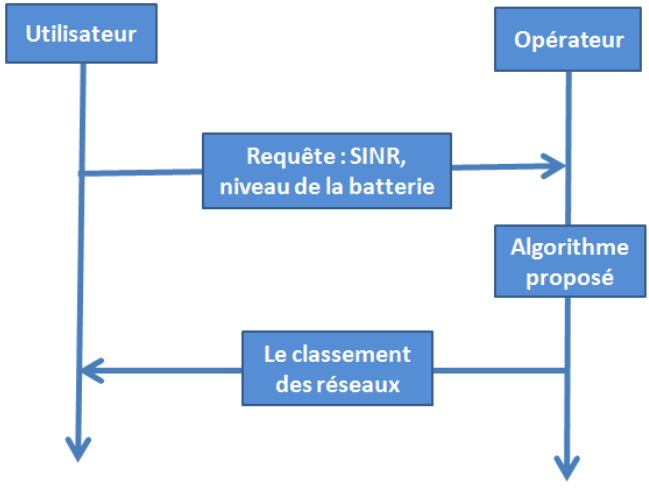


Figure 4.4. *Le processus de sélection de réseau*

Dans notre description basée sur la figure 4.4, nous avons deux agents : l'utilisateur mobile, qui recherche les meilleurs RAT, et l'opérateur, qui déclenche le processus de classement des réseaux disponibles. L'idée derrière la fonction objectif est simple. Nous formulons le système comme une fonction de minimisation dans

laquelle la valeur la plus basse pour chaque critère donne l'ordre le plus élevé pour le réseau. Par conséquent, cela confère au réseau le gain local le plus élevé. Ce processus est répété jusqu'à ce que tous les critères soient évalués. Sa représentation est la suivante, en tenant compte que pour chaque réseau « i », le calcul suivant est effectué :

$$R_i = \sum_{j=1}^{m} income_{ij} \quad [4.4]$$

$$income_{ij} = (\alpha - k_{ij}) * w[j] \quad [4.5]$$

$$k_{ij} = \min(Vect_{ij}) \quad [4.6]$$

avec :

– α : nombre entier fixe égal au nombre d'alternatives ;

– k_{ij} : ordre de classement du réseau *i* pour le critère *j* ;

– $Vect_{ij}$: vecteur colonne de la matrice mat où *j* est fixe ;

– *w[j]* : vecteur pondération associé au critère « *j* » pour la variante « *i* » ;

– *i* : alternative ;

– *j* : critère ;

– mat [n] [m] : matrice d'entrée, représentée par le tableau 4.2.

Nous commençons par diviser la matrice d'entrée en un ensemble de vecteurs colonne afin d'obtenir un groupe de vecteurs égal au nombre de critères. Pour chaque vecteur, les réseaux (alternatives) sont classés en fonction de leurs valeurs de données. Ensuite, chaque réseau reçoit un revenu local égal à la multiplication mathématique entre « α », moins la valeur de rang pour ce réseau, et la valeur de pondération du critère. « α » est une valeur fixe égale au nombre de critères. Ce processus est répété pour les autres critères. La valeur du revenu global est égale à la somme de tous les revenus locaux pour chaque réseau (voir l'équation [4.4]). Nous utilisons la valeur du poids de chaque critère pour différencier les critères, comme le font toutes les méthodes d'optimisation multicritères. Il est bien connu que les critères de délai et de taux de perte de paquets (PLR) sont les plus importants par rapport aux autres critères. Ils représentent les paramètres les plus importants de QoS.

```
Data: The matrix mat [m][n] and the weight vector
        W[], α=m
Result: Income vector for each network
for (j=0; j<m; j++) do
    for (i=0;i<n;i++) do
        tab[i]=mat[i][j]
    end
    tabind=Sort(tab)
    for (k=0;k<tabind.length;k++) do
        income[tabind[k]] = income[tabind[k]]+(α − k)*w[k]
    end
    return income[]
end
```

Algorithme 4.1. *La méthode SAW modifiée*

Le système de revenu local basé sur le rang du réseau pour chaque critère nous permet de contrôler le revenu de chaque alternative en fonction de son rang pour tous les critères. Par exemple, un réseau proposant une valeur de délai égale à 30 ms et un autre proposant une valeur de 100 ms n'ont pas le même revenu local.

Dans cet exemple, une faible valeur de retard est bonne et offre un gain plus élevé au réseau. D'autre part, une bande passante plus élevée se traduira par des revenus locaux plus élevés. Le processus d'évaluation des critères est répété jusqu'à épuisement des critères.

Pour chaque réseau « i », le meilleur des cas est d'avoir une valeur minimale pour le critère « j », donc, le classement local le plus élevé, c'est-à-dire $k_{ij} = 0$, et le revenu est $R_{ij} = \alpha * w[j]$, sachant que dans cette étude α = 5 (le nombre de critères).

Le cas le plus défavorable survient lorsque le réseau a la valeur maximale pour le critère « j », c'est-à-dire que $k_{ij} = \alpha - 1$, donc, le revenu devient $R_{ij} = w[j]$.

L'utilisation du concept de pondération nous permet de donner plus de sens à notre fonction objectif, qui classe les réseaux en fonction de leurs valeurs pour chaque critère. Le vecteur de pondération nous permet de distinguer les critères significatifs des autres moins importants, en fonction des exigences de l'application et des attributs, ce qui donne des revenus plus intéressants aux réseaux avec les valeurs les plus élevées pour les critères importants.

```
Data: vector tab[]
Result: Sort function
int t[][], double min;
int ind, cpt, deb=0;
int a, nb, val=0:
while (deb<t.length) do
        while (tab[a]<0) do
                a++;
        end
        min=tab[a]; ind=a;
for(int i=0;i<tab.length;i++) do
        if (min>tab[i] and tab[i]>=0) then
                        min=tab[i]; ind=i; nb=0;
        end
        else if (min==tab[i] and a!=i) then
                        nb++
        end
end
for(int p=deb; p<deb+nb+1;p++) do
        for(int h=0;h<2;h++) do
                if(h==0) then
                                t[p][h]=val;
                end
                else
                                t[p][h]=ind+cpt; cpt++;
                end
        end
end
```

Algorithme 4.2. *La fonction de tri de l'algorithme M-SAW*

Dans cette étude, nous modifions l'utilisation du vecteur de poids (la pondération) en l'associant à la valeur « α ». Cette modification a principalement deux avantages :

– éviter la situation dans laquelle un réseau qui a une bonne valeur pour un critère non important aura le même revenu qu'un autre réseau avec une bonne valeur pour un critère important. Cette situation existe dans la méthode SAW en raison de l'utilisation du vecteur de poids ordinaire (sans la modification proposée dans ce travail) ;

– le concept de la pondération désigne la représentation des exigences des applications dans le système. C'est ainsi que nous distinguons les applications, car chacune a des exigences spécifiques. La VoIP (Voix sur IP) nécessite un délai minimum à respecter ; pour le service vidéo, le taux de paquets perdus ne doit pas dépasser un certain seuil. Ces informations sont transformées en valeurs numériques pour obtenir le vecteur de pondération ; ceci est fait avec la méthode des valeurs propres.

Dans ce travail, nous avons considéré de nombreux paramètres dans la matrice d'entrée, tels que le coût, la consommation d'énergie, le débit moyen atteint, le délai moyen, le taux de paquet perdus et la charge du réseau. Un exemple de notre matrice est présenté dans le tableau 4.3 ; les valeurs sont tirées de simulations sur NS3.

	Bande passante	Délai	Paquets perdus	Coût	Énergie
Wi-Fi	1-11	100-150	0,2-8	1	-
3G	1-14	25-50	0,2-8	5	-
LTE	1-100	60-100	0,2-8	2	-

Tableau 4.3. *Les valeurs des simulations*

Le paramètre de consommation d'énergie est lié à la batterie du périphérique mobile ainsi qu'à la durée de la session de l'application, c'est-à-dire qu'avec un niveau de batterie élevé la consommation d'énergie ne sera pas aussi importante dans le système, car l'équipement mobile peut assurer que la session ne sera pas interrompue. Dans le cas contraire, si le niveau de charge de la batterie est faible et la durée de la session longue, cela signifie que la session peut être interrompue en raison de l'épuisement de la batterie. Donc, dans le cas où la batterie est faible, ce paramètre est très important et aura un poids plus important dans le système pour éviter l'épuisement de la batterie, et donc une interruption de la session. La valeur de consommation d'énergie est définie à l'aide de l'équation suivante (Huang *et al*. 2012) :

$$P[mJ/s] = \alpha_u * th_u + \alpha_d * th_d + \beta \qquad [4.7]$$

avec :

– α_u, α_d et β : paramètres avec des valeurs différentes d'une RAT à une autre suivant (Huang *et al*. 2012) ;

– th_u et th_d : débit de la liaison montante et débit de la liaison descendante.

La puissance en mJ/s signifie que l'énergie dépend de la durée de la session de l'utilisateur.

4.3.2. *Évaluation des performances*

Dans cette section, nous évaluons les performances du modèle proposé et nous le comparons avec les méthodes d'optimisation multicritères décrites ci-dessus à partir des données d'entrée du tableau 4.3. Trois types de services sont donc considérés : la VoIP, le service vidéo et le service *best effort*. Ces services sont représentés dans le système par les vecteurs de pondération associés, puisque chaque type d'application a des exigences de QoS particulières. De nombreux utilisateurs pensent que plus de bande passante (débit) résoudra le problème, mais avoir seulement plus de bande passante n'est pas la solution idéale pour toutes les applications. Dans les réseaux à commutation de paquets, la qualité de service est affectée par divers facteurs, puisque les paquets peuvent subir de nombreuses altérations lorsqu'ils transitent de la source à la destination. Les facteurs qui déterminent la QoS sont : la bande passante, le délai de réception des paquets, le taux de paquets perdus, la variation du temps de réception des paquets, connue sous l'appellation « gigue ».

Nous présentons ci-dessous les facteurs techniques importants :

– faible débit : en raison du nombre élevé des différents utilisateurs partageant les mêmes ressources du réseau, le débit effectif pouvant être fourni à un flux de données peut être trop faible et insuffisant pour les services multimédias en temps réel (Bendaoud 2018) ;

– latence : pour chaque paquet, arriver à destination peut prendre beaucoup de temps, car il peut être bloqué dans de longues files d'attente ou il peut aussi prendre un itinéraire différent pour éviter la situation de congestion (Bendaoud 2018). Dans certains cas, une latence excessive peut rendre une application telle que la VoIP ou les jeux en ligne inutilisable. Donc, nous pouvons affirmer que la latence est un paramètre très important pour un comportement idéal du réseau ;

– gigue : c'est la variation du temps nécessaire pour recevoir les paquets de leurs sources à leurs destinations. Le délai de réception d'un paquet varie suivant sa position dans la file d'attente des routeurs et de la longueur du chemin entre la source et la destination (Bendaoud 2018). Cette variation de délai peut sérieusement affecter la qualité du streaming audio et vidéo ;

– paquets perdus : il est possible que les routeurs ne livrent pas (abandonnent) certains paquets si ces paquets de données sont corrompus, ou si les paquets arrivent lorsque les mémoires tampons du routeur sont déjà pleines (Bendaoud 2018). L'application réceptrice peut demander la retransmission de ces informations (paquets) à nouveau, ce qui peut entraîner de graves retards dans la transmission globale.

Ce sont les facteurs les plus importants qui influencent une application dans un réseau sans-fil. Nous utilisons donc le vecteur de pondération pour représenter ces exigences dans le système. L'évaluation des performances est composée de deux parties :

– dans la première partie, nous comparons notre proposition avec les méthodes d'optimisation multicritères existantes, dans le cas normal où aucune RAT ne disparaît au milieu du processus de sélection ;

– dans la deuxième partie, un réseau disparaît de la liste des réseaux disponibles ; c'est la preuve que les méthodes d'optimisation multicritères souffrent du phénomène d'inversion de rang et que ce problème ne se produit pas dans la méthode que nous proposons. Le tableau 4.4 représente la matrice d'entrée utilisée dans nos comparaisons ; cette matrice est basée sur les valeurs présentées dans le tableau 4.3.

	Bande passante	Délai	Paquets perdus	Coût	Énergie
N(0)	1,730	105,85	7,94	0,2	1,00
N(1)	5,076	134,88	6,70	0,2	2,6
N(2)	6,849	43,98	2,84	1	6,26
N(3)	6,329	32,15	3,05	1	5,86
N(4)	66,66	95,15	6,32	0,4	12,78
N(5)	62,5	99,73	5,80	0,4	10,28

Tableau 4.4. *La matrice d'entrée*

Les valeurs de pondération de chaque type d'application VoIP, service vidéo et service *best effort*, sont générées à l'aide de la méthode Eigenvector (voir équation [4.8]). Nous avons décidé d'utiliser la méthode Eigenvector, car elle est déjà utilisée dans la méthode AHP. Donc, pour être cohérent et juste, nous avons décidé d'utiliser la même méthode pour avoir les mêmes vecteurs de pondération pour toutes les méthodes :

$$(\text{mat} - \lambda \times \text{I}) \times \text{w} = 0 \qquad [4.8]$$

avec :

– mat : matrice d'entrée ;

– λ : valeur propre ;

– w : vecteur propre associé contenant les valeurs de pondération ;

– I : matrice d'identité.

Le tableau 4.5 contient le vecteur de pondération pour chaque type d'application.

	B. passante	Délai	Paquets perdus	Coût	Énergie
VoIP	0,047	0,486	0,371	0,047	0,047
Vidéo	0,458	0,101	0,302	0,074	0,063
Best effort	0,299	0,146	0,146	0,108	0,299

Tableau 4.5. *Les vecteurs de pondération*

Nous commençons par la première partie de cette étude, le cas ordinaire, c'est-à-dire tous les réseaux sont disponibles.

4.3.2.1. *Simulation 1 : disponibilité de tous les réseaux*

Dans ce cas, tous les réseaux sont disponibles ; cela signifie que les réseaux situés dans la zone de couverte de l'utilisateur ne disparaissent pas. Dans cette situation, nous considérons les algorithmes de sélection de réseaux déjà présentés et nous les comparons avec l'algorithme M-SAW proposé. Les simulations concernent trois types d'applications, à savoir : la VoIP, le service vidéo et les services *best effort*.

Le cas VoIP

Le premier cas de cette simulation concerne l'application VoIP. Cette dernière est représentée *via* un vecteur de pondération qui donne plus d'importance aux délais et à la perte de paquets. Le débit dans le cas de la VoIP n'est pas trop important, car les paquets sont petits et peuvent être transmis avec un débit relativement faible (Lewis et Pickavanc 2006). Les résultats présentés dans le tableau 4.6 concernent le premier cas, qui est VoIP. L'analyse du tableau 4.6 est effectuée en considérant le tableau 4.4 comme entrée de données.

Méthode	Classement
SAW	N(1) N(0) N(2) N(3) N(5) N(4)
TOPSIS	N(3) N(2) N(0) N(1) N(5) N(4)
WPM	N(3) N(2) N(0) N(1) N(5) N(4)
AHP	N(3) N(2) N(5) N(4) N(1) N(0)
M-SAW	N(3) N(2) N(4) N(5) N(0) N(1)

Tableau 4.6. *Classement pour VoIP*

Dans cette simulation, nous nous intéressons à l'ordre total des réseaux, et pas seulement au réseau classé comme le meilleur, car le réseau classé premier est susceptible de devenir chargé rapidement après un laps de temps donné, puis il sera

indisponible, c'est-à-dire surchargé. De ce fait, il est primordial d'avoir un classement total optimal des réseaux. D'après les résultats du tableau 4.6, les méthodes TOPSIS, WPM et M-SAW donnent le même ordre pour les deux premiers réseaux N(0) et N(1). Pour la troisième position, notre méthode M-SAW choisit N(4), mais TOPSIS et WPM choisissent N(0), tandis que AHP choisit N(5). Nous allons maintenant comparer les performances de N(4), N(0) et N(5) pour voir quelle méthode a fait le bon choix. N(0) a un délai de 105,85 et un taux de paquets perdus de 7,94. Le N(4) propose un délai de 95,15 et un taux de paquets perdus de 6,32.

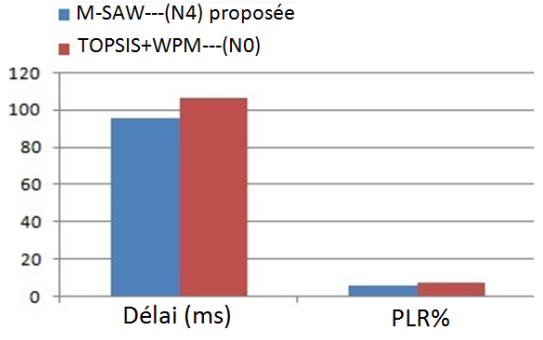


Figure 4.5. *Comparaison du délai et paquets perdus pour N(0) et N(4)*

Donc, N(4) est meilleur que N(0) et cela montre que notre méthode est plus performante que TOPSIS, WPM et AHP (figures 4.5 et 4.6). Dans le cas de la VoIP, nous avons prouvé que la méthode proposée M-SAW donne le meilleur ordre de classement par rapport aux méthodes d'optimisation multicritères.

Le cas service vidéo

Le deuxième cas de nos simulations est dédié aux applications de type vidéo. Le vecteur de poids (pondération) associé donne plus d'importance au débit qu'aux délais de transmission et aux pertes des paquets. Le débit est plus important, car les paquets sont volumineux et doivent être transmis sans perte (Szigeti *et al*. 2013).

Méthode	Classement
SAW	N(1) N(0) N(2) N(3) N(5) N(4)
TOPSIS	N(5) N(4) N(2) N(3) N(5) N(0)
WPM	N(4) N(5) N(3) N(2) N(1) N(0)
AHP	N(2) N(3) N(5) N(4) N(1) N(0)
M-SAW	N(4) N(2) N(3) N(5) N(1) N(0)

Tableau 4.7. *Classement pour service vidéo*

Dans le tableau 4.7, les M-SAW et WPM sélectionnent le N(4) et TOPSIS choisit le N(5). N(4) a un grand débit et un délai faible. Le N(5) a un meilleur taux de paquets perdus. Mais, dans ce cas « service vidéo », l'importance est donnée au débit et au délai. Donc, le meilleur choix est le N(4) (figure 4.6).

De plus, pour le second ordre du classement, la méthode WPM sélectionne le N(5) et M-SAW sélectionne N(2). N(5) a une bande passante de 62,5, un délai de 99,73 ms et un taux de paquets perdus de 5,80. N(2) propose un débit de 6,85, un délai de 43,98 ms et un taux de paquets perdus de 2,84.

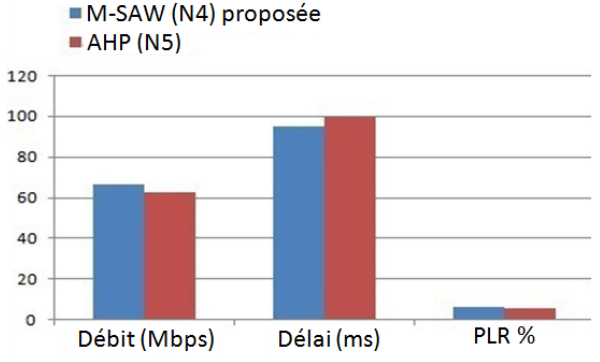


Figure 4.6. *Comparaison du débit et délais pour N(5) et N(4)*

À partir de ces valeurs, on voit que N(5) propose un débit plus grand, mais celui de N(2) est également bon et peut largement servir le service interactif (service vidéo).

Dans ce cas, nous utilisons la propriété de donner à l'utilisateur la valeur minimale qui satisfait aux exigences de l'application. Cela signifie que si les exigences de l'application sont satisfaites, le réseau est considéré comme acceptable et l'utilisateur peut le choisir.

Pour les autres paramètres (délai et paquets perdus), N(2) est très bon par rapport à N(5) (figure 4.7). Nous pouvons voir ici que le paramètre de débit monopolise la décision de classement dans TOPSIS, c'est-à-dire que le réseau ayant le meilleur débit oblige l'algorithme à négliger le délai énorme et le taux élevé des paquets perdus.

Ainsi, basée sur la figure 4.7, notre méthode M-SAW apporte les meilleurs choix pour ce type de service (service vidéo).

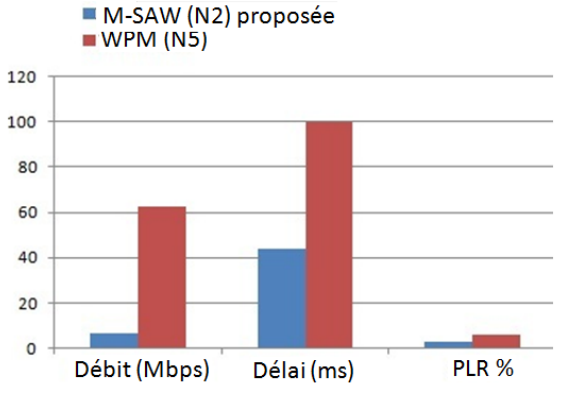


Figure 4.7. *Comparaison entre N(2) et N(4)*

Scénario services *best effort*

Le troisième type d'application utilisé concerne les applications de type *best effort*, comme le service mail et la navigation web. Le vecteur de poids associé ne donne aucune importance particulière à aucun critère : il donne des valeurs égales à tous les critères de manière équitable avec l'espoir d'obtenir un débit élevé. Les paquets sont transmis et l'utilisateur n'a aucune idée sur le temps nécessaire à leurs transmissions, ni sur la qualité des paquets reçus.

Le tableau 4.8 représente les résultats du troisième cas (téléchargement de fichiers). WPM et M-SAW ont les mêmes ordres de premier et deuxième rang ; pour la troisième position, WPM choisit N(1) et M-SAW choisit N(2). Dans la figure 4.8, nous voyons clairement que N(2) est meilleur que N(1).

Méthode	Classement
SAW	N(1) N(0) N(2) N(3) N(5) N(4)
TOPSIS	N(5) N(4) N(2) N(3) N(1) N(0))
WPM	N(4) N(5) N(1) N(2) N(3) N(0)
AHP	N(5) N(1) N(4) N(2) N(3) N(0)
M-SAW	N(4) N(5) N(2) N(3) N(1) N(0)

Tableau 4.8. *Classement pour service best effort*

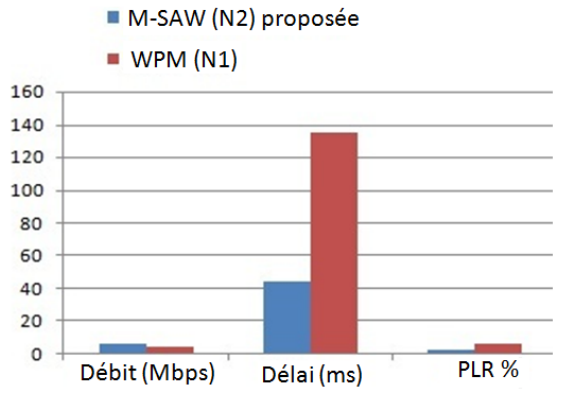


Figure 4.8. *Comparaison entre N(2) et N(1)*

Dans cette première partie de simulation, nous avons effectué une étude comparative entre notre algorithme M-SAW et les algorithmes d'optimisation multicritères présentés dans les premières sections. Nous avons simulé trois cas : VoIP, service vidéo et les applications de type *best effort*. Dans les trois cas, l'algorithme proposé fournit des résultats plus précis et donne à chaque fois le rang total adapté pour le système. Les figures présentées dans cette section montrent que notre algorithme donne dans tous les cas l'ordre exact total de toutes les alternatives.

4.3.2.2. *Simulation 2 : le cas d'inversement de rang*

Dans cette section, nous étudions le problème de l'inversion de rang appelé aussi « anomalie de classement ». Dans cette deuxième simulation, deux objectifs sont à atteindre : le premier, on vise à confirmer que ce problème (inversement de rang) se produit dans les méthodes d'optimisation multicritères lorsque nous supprimons une ligne de la matrice (une alternative disparaît) ; le deuxième objectif est de montrer que notre méthode proposée M-SAW évite ce problème. Si le deuxième objectif est atteint, nous disons que notre méthode proposée est très bien adaptée aux problèmes des multicritères en général et pour la sélection de réseau en particulier.

Maintenant, supposons le cas où un réseau disparaît ; ce cas va nous permettre d'étudier le phénomène d'inversion de rang. Nous allons donc éliminer un réseau (une alternative de la matrice d'entrée), par exemple le réseau N(4) du tableau 4.4, et appliquer les algorithmes de sélection sur les réseaux restants dans le cas du service VoIP (ce sera la même procédure pour les autres types de services). Les résultats sont présentés dans le tableau 4.9.

Méthode	Classement
SAW	N(1) N(3) N(2) N(0) N(5)
TOPSIS	N(5) N(3) N(2) N(0) N(1)
WPM	N(5) N(3) N(2) N(0) N(1)
AHP	N(0) N(1) N(5) N(3) N(2)
M-SAW	N(3) N(2) N(5) N(0) N(1)

Tableau 4.9. *Classement dans le cas de disparition d'une alternative*

Les résultats du tableau 4.9 montrent deux choses : premièrement, toutes les méthodes d'optimisation multicritères (TOPSIS, SAW, WPM et AHP) souffrent du phénomène d'inversion de rang, ce résultat correspond à l'affirmation des auteurs dans (Savitha et Chandrasekar 2011b), selon laquelle toutes les méthodes d'optimisation multicritères souffrent du phénomène d'inversion de rang. Ceci confirme que l'utilisation directe de ce type de méthode n'est pas le bon choix pour le problème de sélection de réseau.

La deuxième remarque est que ce problème (inversement de rang) ne se présente pas dans notre méthode M-SAW lorsqu'un réseau disparaît. Ce résultat nous permet de dire que le M-SAW est plus adapté que les méthodes d'optimisation multicritères pour la résolution du réseau. Nous ajoutons également que le M-SAW est peu complexe (facile à comprendre, à implanter et rapide en temps d'exécution), puisque nous avons apporté simplement quelques modifications à la méthode SAW, connue pour sa faible complexité.

4.3.2.3. *Discussion*

Pour résumer, les méthodes d'optimisation multicritères sont utilisées pour avoir l'ordre de classement des alternatives dans un problème à critères multiples ; cet ordre de classement n'est pas toujours optimal. Il est donc possible qu'une méthode classe un réseau comme le meilleur, l'utilisateur voulant se connecter à ce réseau ; il découvre qu'il est chargé, compte tenu du nombre d'utilisateurs qui sélectionnent le meilleur réseau. Ainsi, la sélection automatique du meilleur réseau chargera facilement ce meilleur réseau et, après un certain moment, ce dernier n'acceptera plus les demandes de connexion des utilisateurs. Pour cette raison, il est important de disposer de la liste de classement correcte et optimale des réseaux disponibles. Par conséquent, si le meilleur réseau est chargé, l'utilisateur passe au deuxième meilleur réseau ; donc, l'algorithme de classement doit ordonner les réseaux d'une manière optimale et précise. En même temps, l'algorithme doit éviter le problème d'inversement de rang car, si un réseau est chargé, il sera inaccessible et supprimé

de la liste des réseaux candidats. Toutes ces exigences ne sont pas respectées dans les méthodes d'optimisation multicritères traditionnelles et sont fournies par la méthode M-SAW. Le processus de sélection est effectué du côté opérateur pour bénéficier de la capacité de traitement de l'opérateur et de l'alimentation (l'énergie) en permanence, ce qui nous donne à la fois efficacité et rapidité ; de plus, l'opérateur dispose de toutes les informations concernant les réseaux et les utilisateurs. Donc, l'algorithme cherche à trouver le classement total le plus optimal des réseaux et non pas seulement le meilleur réseau parmi une liste de réseaux. Avoir l'ordre optimal des réseaux nous permet d'être sûrs qu'à chaque instant l'utilisateur se connecte au meilleur réseau disponible parmi les réseaux existants. Un deuxième avantage de cet algorithme est qu'il fonctionne bien dans le cas normal et aussi dans le cas où un réseau disparaîtrait. Dans ce dernier cas, les méthodes multicritères traditionnelles présentent le problème d'inversement de rang.

4.4. Conclusion

Dans le but de trouver le meilleur réseau à chaque instant, l'idée était de classer les réseaux existants et d'obtenir le classement optimal, et que l'opérateur fournisse aux utilisateurs le meilleur réseau disponible dans la liste des réseaux classés. Dans ce travail, nous avons présenté une méthode appelée « M-SAW », dans laquelle la fonction objectif est basée sur l'ordre relatif de chaque alternative pour chaque critère à chaque itération du processus. Cette méthode nous a permis d'obtenir un algorithme de type « glouton », qui donne de bons résultats. En effet, les simulations montrent que notre méthode surpasse celles existantes dans la littérature et utilisées auparavant. À travers les tests effectués, nous avons montré que toutes les méthodes traditionnelles d'optimisation multicritères présentent le phénomène d'inversion de rang. Or, notre algorithme évite ce problème et reste cohérent en apportant le même ordre de classement, tout en éliminant le réseau qui a disparu.

4.5. Bibliographie

Abdullah, L., Adawiyah, C.W. (2014). Simple additive weighting methods of multi criteria decision making and applications: A decade review. *International Journal of Information Processing and Management*, 39.

Alkhawlani, M., Ayesh, A. (2008). Access network selection based on fuzzy logic and genetic algorithms. *Advances in Artificial Intelligence*, 8(1), 1.

Bakmaz, B., Bojkovic, Z., Bakmaz, M. (2007). Network selection algorithm for heterogeneous wireless environment. Dans *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, 1–4.

Bendaoud, F. (2018). Management of joint radio resources in heterogeneous networks Beyond 3G. Thèse de doctorat, École supérieure en informatique, Sidi Bel Abbès.

Bendaoud, F., Didi, F., Abdennebi, M. (2017). A modified-SAW for network selection in heterogeneous wireless networks. *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 15(2), 8–17.

Bendaoud, F., Abdennebi, M., Didi, F. (2018). Network Selection in Wireless Heterogeneous Networks : a Survey. *Journal of Telecommunications and Information Technology*, 4, 64.

Huang, J., Qian, F., Gerber, A., Mao, Z.M., Sen, S., Spatscheck, O. (2012). A close examination of performance and power characteristics of 4G LTE networks. Dans *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 225–238.

Huszak, A., Imre, S. (2010). Eliminating rank reversal phenomenon in GRA-based network selection method. Dans *2010 IEEE International Conference on Communications*, 1–6.

Kovvali, S.K. *et al*. (2015). Content and ran aware network selection in multiple wireless access and small-cell overlay wireless access networks. Brevet.

Lahby, M., Cherkaoui, L., Adib, A. (2012). An intelligent network selection strategy based on MADM methods in heterogeneous networks. *International Journal on Wireless & Mobile Networks*, 4(1), 83–96.

Lewis, C.S., Pickavanc, S. (2006). *Selecting MPLS VPN Services*. Cisco Press, Indianapolis.

Lin, C.T., Hsu, P.F. (2003). Adopting an analytic hierarchy process to select Internet advertising networks. *Marketing Intelligence & Planning*, 21(3), 183–191.

Nguyen-Vuong, Q.T., Ghamri-Doudane, Y., Agoulmine, N. (2008). On utility models for access network selection in wireless heterogeneous networks. Dans *IEEE Network Operations and Management Symposium*, 144–151.

Olson, D.L. (2004). Comparison of weights in TOPSIS models. *Mathematical and Computer Modelling*, 40(7–8), 721–727.

Salih, Y.K., See, O.H., Ibrahim, R.W., Yussof, S., Iqbal, A. (2015). A user-centric game selection model based on user preferences for the selection of the best heterogeneous wireless network. *Annales des télécommunications*, 70(5–6), 239–248.

Savitha, K., Chandrasekar, C. (2011a). Trusted network selection using SAW and TOPSIS algorithms for heterogeneous wireless networks. *International Journal of Computer Applications*, 26, 8.

Savitha, K., Chandrasekar, C. (2011b). Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks. *Global Journal of Computer Science and Technology*, 11.

Scherzer, S., Scherzer, T. (2015). Method and system for selecting a wireless network for offloading. Brevet.

Sgora, A., Vergados, D.D., Chatzimisios, P. (2010). An access network selection algorithm for heterogeneous wireless environments. Dans *The IEEE symposium on Computers and Communications*, 890–892.

Shen, D.M., Tian, H., Sun, L. (2010). The QoE-oriented heterogeneous network selection based on fuzzy AHP methodology. Dans *The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 275–280.

Shin, Y.B., Lee, S., Chun, S.G., Chung, D. (2013). A critical review of popular multi-criteria decision making methodologies. *Issues in Information Systems*, 14(1), 358–365.

Szigeti, T. *et al*. (2013). *End-to-End QoS Network Design: Quality of Service for Rich-Media & Cloud Networks*. Cisco Press, Indianapolis.

TalebiFard, P., Leung, V.C. (2011). Context-Aware Mobility Management in Heterogeneous Network Environments. *JoWUA*, 2(2), 19–32.

Wang, Y.M., Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling*, 49(5–6).

Watanabe, E.H., Menasché, D.S., de Souza e Silva, E., Leao, R.M.M. (2008). Modeling resource sharing dynamics of VoIP users over a WLAN using a game-theoretic approach. Dans *IEEE INFOCOM 2008 – The 27th Conference on Computer Communications*. IEEE, Phoenix, 915–923.

Wu, J., Cheng, B., Yuen, C., Shang, Y., Chen, J. (2015). Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks. *IEEE Transactions on Mobile Computing*, 14(4), 688–701.

PARTIE 3

L'IA et l'approche *Cloud*

5

Sélection des services *Cloud Computing* : apport des méthodes intelligentes

Ahmed Khalid Yassine SETTOUTI
Université Abou Bekr Belkaid, Tlemcen, Algérie

5.1. Introduction

De nos jours, chaque objet informatique génère des données. Ces dernières sont relatives à son utilisation, à son environnement ou à son état. Une donnée générée reste et restera toujours importante pour son constructeur ou son utilisateur.

Tout de suite après l'explosion du Web, l'utilisation des dispositifs informatiques a augmenté soudainement. Une conséquence directe est la croissance des données générées par ces dispositifs. Cet essor a engendré de nouveaux besoins et exigences quant aux ressources capables de gérer les informations créées. Ainsi, le *Cloud Computing* a vu le jour.

Un peu plus tard, les fournisseurs de Cloud Computing n'ont cessé de croître. Ceci a eu un impact sur la complexité du choix entre eux lorsqu'il s'agit d'avoir des exigences peu conventionnelles. En plus de cela, chaque fournisseur a ses propres tarifs, ses modèles de facturation, ses avantages comme ses inconvénients. Cela a rendu le choix entre les fournisseurs ou entre leurs services encore plus difficile.

Notre chapitre est un état de l'art qui se concentre sur la sélection, la composition, l'évaluation ou la recommandation des services Cloud Computing. Les travaux

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

surveillés doivent avoir une problématique, suivre une approche ou utiliser un outil d'intelligence artificielle. Beaucoup d'essais ont été constatés dans la littérature comme Ahmed *et al.* (2019), qui a préféré se concentrer sur la sélection des services Cloud Computing dans le cas de fédération de nuages hétérogènes, tandis que Whaiduzzaman *et al.* (2014) s'est focalisé sur les approches utilisant les méthodes multidécisionnelles à choix multiples. La même chose a été faite par Le *et al.* (2014) en réduisant l'ensemble des travaux surveillés à ceux qui utilisent la logique floue. Et pour finir, Papathanasiou *et al.* (2015) a repris le travail de Whaiduzzaman *et al.* (2014) et y a ajouté une comparaison analytique.

L'objectif de notre chapitre est de donner aux lecteurs une idée globale sur l'actualité des travaux sélectionnant les services Cloud Computing en utilisant l'intelligence artificielle. L'intérêt est de réduire le temps de la recherche bibliographique pour les chercheurs dans le domaine et d'initier plus facilement les chercheurs non éclairés.

La suite de ce chapitre est structurée comme suit :

– la section 5.2 présente quelques notions générales nécessaires à la compréhension du chapitre ;

– la section 5.3 énumère quelques travaux similaires dans le domaine de la sélection des services Cloud Computing ;

– la section 5.4 expose les travaux surveillés, les critique et leur suggère quelques améliorations ;

– la section 5.5 récapitule l'avancée de la recherche dans la sélection des services Cloud Computing en utilisant l'intelligence artificielle et propose des travaux pouvant être réalisés dans le futur proche.

5.2. Prérequis scientifique et technique

Dans cette section, nous allons décrire quelques notions scientifiques ou techniques nécessaires à la compréhension du chapitre. Nous allons commencer par une définition simpliste, mais assez complète du Cloud Computing. Enfin, nous terminerons par le concept de l'intelligence artificielle, tout en expliquant ce qui l'est et ce qui ne l'est pas.

5.2.1. Cloud Computing

Dans cette section, nous allons présenter brièvement le concept du Cloud Computing (sa définition, ses caractéristiques, ses modèles de déploiement et ses niveaux de services).

Le Cloud Computing (ou « informatique en nuage ») est un modèle permettant un accès à des ressources (physiques ou virtuelles) configurables à distance (Mell et Grance 2011).

L'accès doit d'un côté être ubiquitaire et à la demande des clients. D'un autre côté, il doit nécessiter un minimum d'efforts de la part du client et ne doit en aucun cas solliciter l'interaction avec le fournisseur (Mell et Grance 2011).

Le nuage informatisé doit assurer une certaine qualité pour ses clients, telle qu'un taux de disponibilité minimum (généralement de 99,95 %) (Hayes 2008 ; Mell et Grance 2011).

5.2.1.1. *Caractéristiques du* Cloud Computing

Bien que la définition précédente (voir la section 5.2.1) liste un ensemble de caractéristiques permettant de distinguer le modèle du Cloud Computing des autres modèles et architectures (comme la virtualisation, l'informatique à la demande et l'informatique ubiquitaire), il est important de décrire chaque critère à lui seul. Dans les sections suivantes, nous allons justement détailler les propriétés du Cloud Computing.

5.2.1.1.1. Service à la demande

Le client peut s'allouer autant de ressources que nécessaire à sa demande. Ceci à n'importe quel moment et sans intervention des services du fournisseur (Ahmad *et al.* 2017).

5.2.1.1.2. Accès *via* Internet

Toutes les ressources du nuage informatique sont accessibles par le biais d'Internet (Armbrust *et al.* 2010). Ces moyens sont à disposition de tout type d'internaute, qu'ils soient lourds (applications bureau) ou légers (applications web) (Mell et Grance 2011).

5.2.1.1.3. Groupement des ressources

Les différentes ressources sont groupées pour un ensemble de clients (Syntec Numérique 2012). Selon la charge des processus et leur nombre, les ressources sont allouées et libérées (Hayes 2008).

Les consommateurs n'ont aucune connaissance sur la position exacte de leurs traitements (Mell et Grance 2011). Ceci dit, ils peuvent choisir dans quel centre de données dérouler leurs services, mais pas dans quel serveur exactement (Armbrust *et al.* 2010).

5.2.1.1.4. Flexibilité rapide

Les ressources doivent être allouées, supprimées, étendues ou réduites de la manière la plus facile et rapide possible (Mell et Grance 2011). Pour les clients, ces dernières paraissent comme illimitées, accessibles et modifiables à n'importe quel moment et à partir d'un quelconque emplacement connecté, comme si elles étaient locales (Hayes 2008).

5.2.1.1.5. Service mesuré

Les utilisateurs payent ce qu'ils utilisent uniquement. Le service est donc mesuré par son utilisation (Syntec Numérique 2012). Ceci peut varier selon le type de service, mais le principe reste toujours le même. Par exemple, un service de stockage (Google Drive, qui est gratuit pour 10 Go et payant pour plus) peut mesurer l'espace alloué, sans pour autant facturer les autres ressources comme le processeur et la bande passante, alors qu'un service de base de données (Heroku par exemple, qui est gratuit pour une certaine taille de bases de données, mais payant pour plus) peut mesurer la taille des tables, sans pour autant inclure le coût de leur création et/ou de leur traitement.

5.2.1.2. *Modèles de déploiement*

Déployer un nuage informatisé peut se faire de quatre façons différentes. Le principal critère les différenciant est le choix de l'audience.

5.2.1.2.1. Modèle public

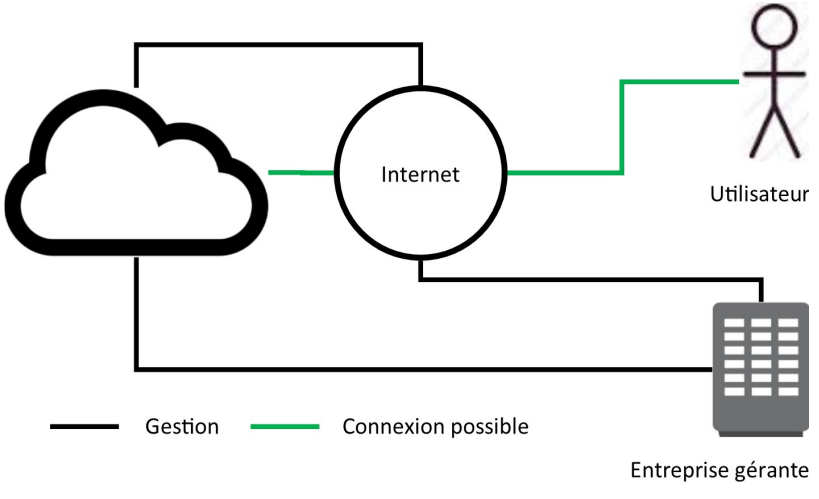


Figure 5.1. *Modèle de déploiement public du Cloud Computing*

Bien qu'il soit géré par une entreprise particulière (comme la compagnie Microsoft, qui gère son service public Azure), le modèle public du Cloud Computing reste un modèle accessible par tout internaute connecté (Mell et Grance 2011). Par conséquent, il est le moins sécurisé, mais pas pour autant le moins utilisé (Hayes 2008).

La figure 5.1 illustre un cas d'exemple pour le modèle public du Cloud Computing. Un service en nuage public reste toujours accessible au grand public. La seule condition pour un client est qu'il doit être connecté à Internet.

Comme tout matériel, il doit être géré par un organisme, une entreprise, etc., mais ceci pourrait être fait d'une manière locale ou *via* Internet.

5.2.1.2.2. Modèle privé

L'infrastructure du nuage ou d'une partie du nuage est dédiée à une entreprise spécifique (Ahmad *et al.* 2017). Un internaute non concerné par l'entreprise ne pourra pas accéder aux installations dites *privées* (Mell et Grance 2011). Ceci dit, l'infrastructure pourrait être gérée par l'entreprise en question (cliente) ou par une tierce personne (gérante) (Syntec Numérique 2012).

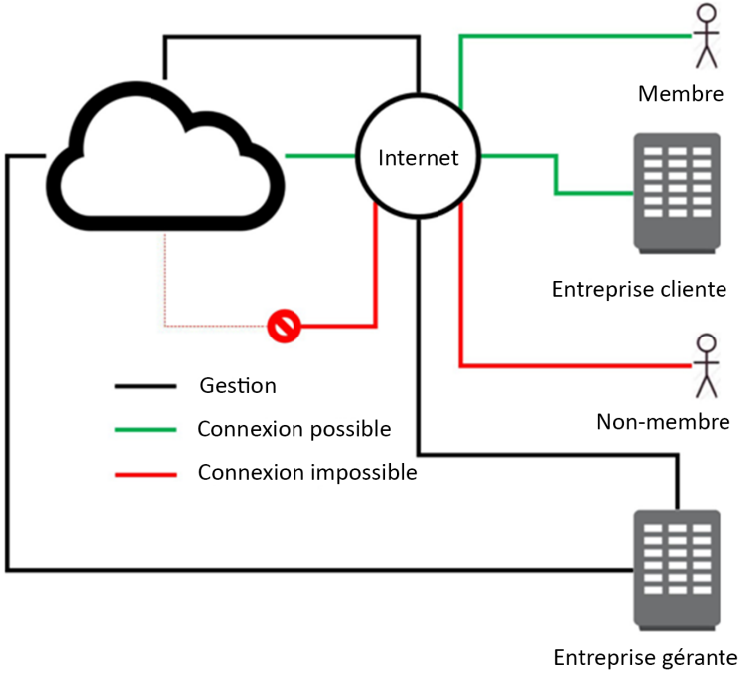


Figure 5.2. *Modèle de déploiement privé du Cloud Computing*

La figure 5.2 illustre un exemple d'un modèle de déploiement privé du Cloud Computing. Un membre de l'entreprise cliente peut accéder aux installations qui lui sont dédiées alors qu'un autre ne pourra pas. D'un autre côté, le matériel peut être géré par un organisme tiers (entreprise gérante) ou par l'entreprise elle-même (cliente).

5.2.1.2.3. Modèle communautaire

L'infrastructure du nuage ou d'une partie du nuage est dédiée à une communauté d'entreprises partageant les mêmes requis et préférences spécifiés dans leurs contrats de type *Service Level Agreement* (Mell et Grance 2011).

L'infrastructure en question peut être gérée par la communauté en totalité, ou par une partie de cette dernière, une entreprise membre ou un organisme tiers (Syntec Numérique 2012).

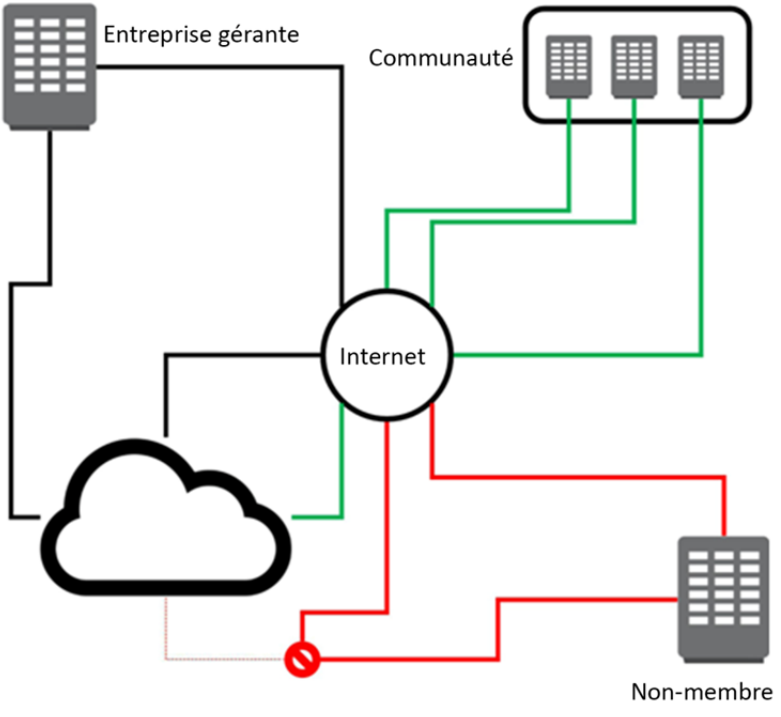


Figure 5.3. *Modèle de déploiement communautaire pour le Cloud Computing*

La figure 5.3 illustre à titre d'exemple un modèle de déploiement communautaire. La gestion de l'infrastructure peut être exercée par l'entreprise en dehors de la communauté (entreprise gérante), mais peut aussi être faite par la communauté elle-même ou par une partie de la communauté. Ceci dit, un groupe d'entreprises doit avoir un minimum de requis et/ou de principes en commun pour tirer un maximum de bénéfices d'un tel modèle.

5.2.1.2.4. Modèle hybride

L'infrastructure est une composition d'au moins deux modèles précédemment cités (Ahmad *et al.* 2017).

5.2.1.3. *Niveaux de services* Cloud Computing

5.2.1.3.1. *Infrastructure-as-a-Service* (IaaS)

On donne aux utilisateurs des ressources accessibles et partagées, se présentant sous forme d'unités de calcul, de supports de stockage, de moyens de communication (réseau), etc. (Mell et Grance 2011). Avec ces moyens-là, les consommateurs pourront dérouler leurs propres logiciels et outils à distance (Hayes 2008). Ceci veut dire que le client a le contrôle sur le système d'exploitation, l'aspect applicatif et le stockage (Syntec Numérique 2012), mais il n'a pas la gestion des couches basses comme l'infrastructure et le réseau matériel (Hayes 2008). Ceci dit, il peut parfois avoir la gestion logicielle et partielle du réseau (si le fournisseur autorise cela), comme le choix d'un pare-feu (Armbrust *et al.* 2010).

5.2.1.3.2. *Software-as-a-Service* (SaaS)

L'objectif visé est d'utiliser des applications web s'exécutant sur les serveurs du nuage (Hayes 2008). Les logiciels sont accessibles par des clients différents, répartis sur le globe (Mell et Grance 2011). Ces derniers n'ont de contrôle que sur les paramètres laissés ouverts explicitement par le fournisseur du service (Ahmad *et al.* 2017), mais n'ont aucun contrôle sur la gestion de l'infrastructure, le système d'exploitation, le réseau, l'application elle-même (Syntec Numérique 2012).

5.2.1.3.3. *Platform-as-a-Service* (PaaS)

Un niveau intermédiaire entre les deux niveaux précédemment cités est le PaaS (Ahmad *et al.* 2017). Le client peut dérouler des applications de sa création sur les outils offerts ou proposés par le fournisseur (Mell et Grance 2011). Ces derniers peuvent regrouper les bases de données et les langages de programmation (Hayes 2008). Le consommateur ne gère pas les couches basses comme les serveurs et les systèmes d'exploitation (Syntec Numérique 2012), mais il a tout de même le

contrôle sur les applications déployées ainsi que la possibilité de configurer l'environnement les hébergeant (Armbrust *et al.* 2010).

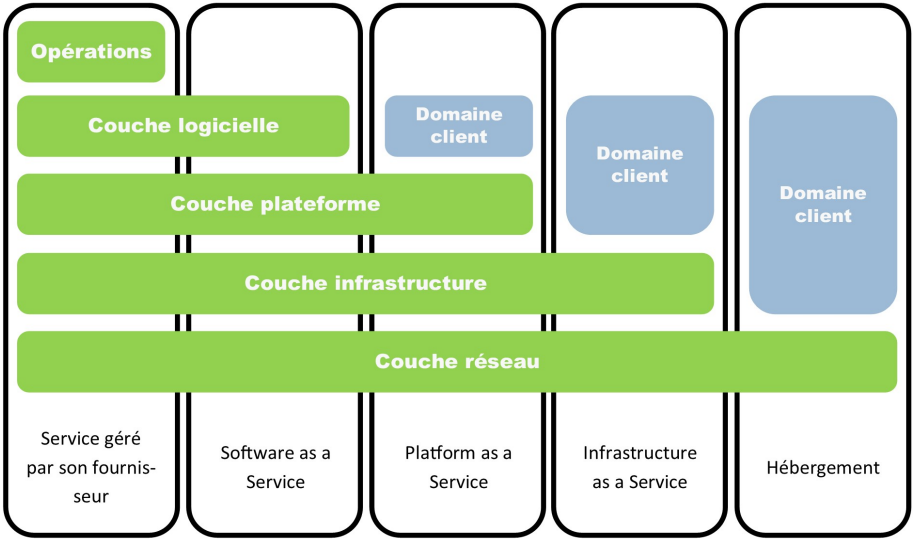


Figure 5.4. *Répartition des gestions client/fournisseur dans un environnement Cloud Computing selon le niveau de service (Syntec Numérique 2012)*

La figure 5.4 montre la façon dont Syntec Numérique (2012) fait la différence entre les niveaux de services Cloud Computing. Nous pouvons citer Facebook et la plupart des produits Google comme représentation des plus célèbres services du modèle SaaS.

Le PaaS est un niveau assez difficile à comprendre, vu l'ambiguïté dans le départage du contrôle entre le client et le fournisseur. Par exemple, le client peut avoir uniquement le contrôle sur des données logiques comme le service de Heroku, ou il peut avoir le contrôle sur un SGBD réel mais dans un environnement cloud comme MySQL Cloud. Et pour finir, le client peut avoir le contrôle sur tout un environnement de développement intégré comme Salesforce (le service donne à ses clients la possibilité de créer un site web complexe en quelques étapes).

Le niveau IaaS offre le contrôle au client sur une machine virtuelle distante. Par conséquent, le client a un ordinateur distant qu'il ne possède pas physiquement, mais qui est toujours disponible pour lui.

5.2.2. *Intelligence artificielle*

L'intelligence artificielle fait référence à l'intelligence démontrée par les machines, contrairement à l'intelligence naturelle qui est exercée par les êtres humains ou les animaux (Russel et Norvig 2016).

L'intelligence artificielle (I.A. ou IA) est une science étudiant des *agents intelligents* (Legg 2007), qui sont des entités capables à la fois de percevoir l'environnement dans lequel ils existent et de prendre des décisions, afin de maximiser les chances d'atteindre leurs objectifs (Kaplan et Haenlein 2019).

C'est la définition la plus simple et la plus récurrente dans le domaine, car beaucoup de chercheurs estiment que les gens confondent tout au sein de l'intelligence artificielle (McCorduck 2009). D'ailleurs, ce phénomène a un nom qui est « l'effet IA » (McCorduck 2009). À titre d'exemple, beaucoup de chercheurs en intelligence artificielle ne considèrent pas la reconnaissance optique des caractères comme faisant partie de l'IA. Ceci à cause des algorithmes déterministes et optimaux résolvant ce genre de problèmes (Schank 1991).

5.2.2.1. *Problématiques*

Les problématiques traitées dans le domaine de l'intelligence artificielle sont diverses et variées. La plus notoire reste la résolution de problèmes (par raisonnement) (Poole *et al.* 1998). Puisque le raisonnement est une des facettes de l'intelligence naturelle, les chercheurs en intelligence artificielle ont essayé depuis les années 1980 d'implémenter le raisonnement pas à pas dans des machines (Nilsson 1998).

Le concept a évolué jusqu'à présent, mais les algorithmes proposés de nos jours subissent une montée en temps d'exécution (exponentielle) dès que le problème dépasse un certain niveau de complexité (Russel et Norvig 2016). Avec une intelligence naturelle, il est plus facile de résoudre ces problèmes d'une complexité importante, parce que les humains utilisent la majeure partie du temps des raccourcis (comme l'intuition et la mémoire) (Foss et Dodwell 1966).

Puisque l'intelligence artificielle étudie les agents intelligents capables de percevoir leurs environnements, il est donc évident que la perception en elle-même est une problématique dans le domaine (Nilsson 1998 ; Russel et Norvig 2016). Nous pouvons citer la reconnaissance vocale (Russel et Norvig 2016), la reconnaissance faciale et la reconnaissance des objets (Russel et Norvig 2016), qui sont des sous-problématiques liées à l'intégration des capteurs aux agents intelligents.

L'intelligence artificielle traite beaucoup de problématiques. Il faut noter que celles expliquées dans cette section ne représentent en aucun cas l'ensemble global des problématiques.

En effet, certaines problématiques sont considérées davantage comme des méthodes (comme l'apprentissage machine) ou des applications (comme la reconnaissance du langage).

5.2.2.2. *Approches*

Afin de faire face aux problématiques précédemment citées, l'intelligence artificielle a à sa guise toute une panoplie d'approches. Alors que quelques chercheurs pensaient que les machines devaient ressentir des émotions, d'autres préféraient qu'elles trouvent l'essence même des résolutions de problèmes logiques et variés (Crevier 1993 ; McCorduck 2009) (comme la représentation des connaissances, la planification et l'apprentissage).

D'autres approches peuvent être mentionnées, comme l'informatique intelligente. Elle étudie l'intégration des réseaux artificiels de neurones en informatique (Crevier 1993 ; McCorduck 2009) et elle est utilisée en vue de résoudre les problèmes avec peu de certitude et ne nécessitant pas forcément une solution optimale (globale).

L'intelligence artificielle suit beaucoup d'approches qui ne peuvent pas être citées dans un chapitre ne traitant pas spécifiquement le domaine. Les approches brièvement expliquées dans cette section ne représentent pas toutes les approches possibles en IA.

En plus de cela, certaines approches sont davantage considérées comme étant des domaines (comme la cybernétique) ou des sous-domaines tout simplement (systèmes d'agents intelligents).

5.2.2.3. *Outils*

Chaque domaine a ses problématiques, chaque problématique peut être résolue par une approche, et chaque approche a besoin d'outils.

Parmi les outils les plus utilisés en IA nous pouvons citer les réseaux de neurones artificiels. Ces derniers ont été largement inspirés de l'architecture et du comportement du cerveau humain. Ils sont connus sous la forme d'un ensemble de nœuds interconnectés. Chaque nœud est associé à un poids, afin de voter pour une décision qu'il est censé activer (Domingos 2015). Le seul défaut d'un réseau de neurones artificiels est qu'il a besoin d'un certain temps afin d'apprendre à partir

d'exemples. En plus de cela, même si le cas d'apprendre une décision fausse n'est pas toujours probable, ce cas ne reste pas impossible.

D'autre part, certains chercheurs utilisent les méthodes probabilistes pour des raisonnements incertains. En d'autres termes, des algorithmes n'ayant pas la totalité de leurs entrées. Parmi ces méthodes, nous pouvons trouver les réseaux bayésiens (Nilsson 1998), les modèles de Markov (Russel et Norvig 2016), les filtres de Kalman (Russel et Norvig 2016), etc.

Nous ne trouvons pas judicieux de décrire toutes les méthodes utilisées dans la sélection des services Cloud Computing dans cette section. Nous trouvons plus intéressant pour le lecteur de trouver un rappel (une brève description) de la méthode utilisée au moment de parler des travaux et des contributions sélectionnant les services Cloud Computing. Par conséquent, nous allons expliquer plus en détails les outils des contributions dans la section 5.4.

Nous aurions pu détailler davantage les définitions des problèmes récurrents, les procédés courants, les outils ainsi que les applications de l'intelligence artificielle. Mais nous avons préféré nous concentrer sur les spécificités du chapitre, c'est-à-dire les services Cloud Computing, ainsi que sur leurs approches de sélection.

5.3. Travaux similaires

Dans cette section, nous allons présenter brièvement quelques travaux dans le même domaine. Ce qui veut dire des travaux surveillant des contributions dans la sélection des services Cloud Computing.

Beloglazov *et al.* (2012) ont surveillé les heuristiques utilisées afin de découvrir et d'allouer les ressources d'un centre de données dans un environnement Cloud Computing, tout en réduisant la consommation énergétique et en assurant la qualité de service négociée. Afin d'y parvenir, les chercheurs ont divisé leurs travaux en trois axes distincts. Le premier se concentre sur les architectures dans le nuage informatique permettant une réduction ou un contrôle sur la consommation d'énergie. Le deuxième axe s'intéresse aux politiques et aux algorithmes de planification capables d'assurer une certaine qualité de service préalablement déterminée, tout en réduisant la consommation d'énergie. Le troisième et dernier axe s'intéresse plus aux problématiques qu'aux solutions. Il présente les problématiques encore ouvertes de l'époque tout en cherchant à allouer les ressources d'un nuage informatisé, à réduire la consommation d'énergie et à assurer la qualité de service requise (négociée).

Le travail que nous venons de citer est plus que recommandé pour tout chercheur voulant contribuer dans l'allocation de ressources des nuages informatisées (que le chercheur soit non-initié, débutant ou expert).

Certains travaux ont appliqué les méthodes de décision à choix multiples afin de sélectionner les services Cloud Computing IaaS. Dans un premier temps, Rehman *et al.* (2012) se sont contentés de dérouler un ensemble de méthodes sur le même ensemble de services Cloud Computing IaaS. Puis, ils ont comparé les résultats trouvés. Dans un autre papier, les mêmes auteurs ont rajouté la qualité de services à leurs critères de recherche (Rehman *et al*. 2014).

Les deux surveillances précédemment citées sont d'excellents travaux pour tout chercheur qui veut contribuer dans la sélection de services Cloud Computing en utilisant les réseaux de neurones. Parce que la majorité des approches nécessitant un apprentissage se reposent généralement sur des méthodes de décision à choix multiple.

Durant la même année, Mandal *et al.* (2013) ont comparé les approches afin de découvrir, évaluer et sélectionner les services Cloud Computing pour des applications à grand flux de données. Les auteurs ont d'abord présenté une architecture, puis ils ont comparé les résultats de leurs travaux avec ceux en actualité. Ceci dit, les contributions sont toutes basées sur les arbres.

L'article en question est assez dépassé certes, mais reste un excellent moyen de s'initier à la découverte de services Cloud Computing en utilisant des agents intelligents.

Le *et al.* (2014) ont proposé une approche de décision à critères multiples, basée sur la logique floue, afin de sélectionner les meilleurs services Cloud Computing. Pour ce faire, les chercheurs ont proposé une approche basée à la fois sur l'*Interpretive Structure Modeling* (ISM), ainsi que sur l'*Analytic Network Process* (ANP). Le premier a servi à la modélisation des relations interactives entre les critères d'évaluation, alors que le second a servi à la gestion des données incertaines. Vu que l'objectif était de maximiser les résultats utiles, un résultat optimal a été perçu comme un service maximisant la fonction utilité. Ceci dit, les auteurs ont négligé le temps nécessaire pour arriver à de tels résultats.

Certaines applications de l'intelligence artificielle se focalisent sur la prédiction de données incertaines. Pour un chercheur non initié cherchant une contribution dans cette optique, cette surveillance est plus que recommandée.

Sun *et al.* (2014) ont résumé les avancées scientifiques achevées en 2014 dans le domaine de sélection des services Cloud Computing. Puis, ils ont proposé quelques

directions pour le futur proche. Les auteurs ont certes négligé le côté expérimental des travaux étudiés, mais ils ont néanmoins su comment résumer l'évolution des problématiques durant les années 2009-2014.

(Sun *et al.* 2014) est un excellent moyen de s'initier dans le domaine de sélection de services Cloud Computing, si le chercheur n'a pas encore cerné une problématique claire.

Whaiduzzaman *et al.* (2014) ont proposé une surveillance des travaux sélectionnant les services Cloud Computing utilisant les méthodes multicritères à choix multiple (MCCM). Afin d'aboutir à leur objectif, ils ont d'abord synthétisé les techniques MCCM (parmi elles, il y avait des techniques reposant sur l'intelligence artificielle). Ensuite, ils ont analysé les approches dans leur domaine étudié. Et enfin, ils ont mis en évidence certains aspects des méthodes MCCM dans la sélection des services Cloud Computing. Parmi ces évidences démontrées, beaucoup reposent sur l'intelligence artificielle.

L'écrit que nous venons de citer est certes orienté vers les méthodes multicritères à choix multiple, mais ce genre de démarches repose souvent et grandement sur l'intelligence artificielle. Par conséquent, l'écrit de Whaiduzzaman *et al.* (2014) est recommandé pour les chercheurs non initiés au domaine de l'IA voulant faire une contribution dans la sélection des services Cloud Computing.

Papathansiou *et al.* (2015) ont repris le travail de Whaiduzzaman *et al.* (2014) afin d'y ajouter une comparaison analytique en termes d'objectifs et de performances. Pour y arriver, les chercheurs devaient implémenter les procédés surveillés, les évaluer et les comparer. Puisque l'article de Whaiduzzaman (2014) est recommandé, (Papathanasiou *et al.* 2015) l'est encore plus.

En dernier lieu, Ahmed *et al.* (2019) ont proposé une surveillance ainsi qu'une analyse des exigences pour les approches sélectionnant les services Cloud Computing les plus en confiance à des fins de fédération de nuages (ou fédération de services Cloud Computing). Afin d'y parvenir, les chercheurs ont commencé par définir les caractéristiques des nuages informatiques fédérées, ainsi que les services en nuage fédérés. Ensuite, ils ont proposé leur propre définition de niveau de confiance d'un service Cloud Computing. De ce fait, ils ont comparé leurs mesures de confiance avec les autres proposées dans la littérature. Enfin, ils ont mis en évidence les avantages de leur proposition par rapport aux propositions dans le domaine de fédération des services Cloud Computing.

Bien que le travail (Ahmed *et al.* 2019) semble être plus une contribution qu'une surveillance, nous avons préféré le considérer comme étant un état de l'art pour la

simple raison que les chercheurs le mentionnent explicitement. Même si un chercheur n'est pas forcément intéressé par la fédération des nuages informatiques, cet écrit lui est recommandé pour la composition de services Cloud Computing (cas général).

À la fin, il faut distinguer une certaine implicité dans le domaine. Certains travaux se positionnent du côté du client, tandis que d'autres prennent le côté du fournisseur. Ces derniers contribuent dans les schémas de facturation (Aishwarya et Muzammil 2017), l'efficacité et la consommation d'énergie (Sharma *et al*. 2016b), la planification des tâches (Singh *et al*. 2017), la maximisation des profits (Das *et al.* 2014) et la gestion des ressources (Liaqat 2017) (ceci n'est qu'une liste non exhaustive des objectifs).

Par contre, les travaux du côté client se concentrent plus sur l'optimisation du coût (Pandey *et al.* 2011), la migration des applications vers le Cloud Computing (Andrikopoulos *et al.* 2013), la gestion des risques (Gupta *et al.* 2015) et la confidentialité des données (Xu *et al*. 2016) (ceci n'est qu'une liste non exhaustive des objectifs).

5.4. Travaux surveillés

Dans cette section, nous allons énumérer un certain nombre de travaux dans la sélection des services Cloud Computing utilisant l'intelligence artificielle (problématiques, outils, approches, domaines d'application, etc.).

Afin de lire le chapitre aisément, nous avons divisé l'ensemble des travaux à surveiller en familles selon les procédés (sous-domaines de l'intelligence artificielle) utilisés. Ceci a pour but de réduire le nombre de contributions par section, et de pouvoir ainsi discuter plus en détails des problèmes spécifiques au sous-domaine.

5.4.1. *Apprentissage machine*

Dans cette section, nous allons surveiller les travaux dans la sélection des services Cloud Computing utilisant l'apprentissage machine à leurs fins. Sachant que l'intelligence artificielle est une implémentation de l'intelligence humaine par des machines et que l'apprentissage par les exemples est une des formes de l'intelligence naturelle. L'apprentissage machine permet aux agents intelligents d'apprendre à partir des exemples avant d'être confrontés à des cas réels.

Kang *et al.* (2013) ont proposé une approche pour sélectionner une machine virtuelle dans un environnement en nuage afin de planifier et de répartir les tâches des processus utilisateurs. L'objectif était de maximiser la prédiction de la meilleure machine virtuelle ainsi que ses performances, avant de voir le prix ou la qualité, tout en faisant un apprentissage sur un ensemble d'exemples. Malgré l'intérêt de la contribution, un apprentissage nécessite un certain temps afin d'arriver à un stade utile ; ce côté (temps nécessaire à l'apprentissage) a été d'ailleurs négligé de la part des auteurs.

Xiaogang *et al.* (2015) ont proposé une approche dynamique afin de sélectionner le meilleur service Cloud Computing. Pour ce faire, ils implémentent une stratégie nommée *Dynamic Cloud Service* (DCS) dans chaque intermédiaire (*broker*). La stratégie est basée sur l'apprentissage machine, et son but est de retourner le meilleur service en temps réel, quels que soit les changements qui ont eu lieu auparavant. Les chercheurs ont testé la performance de leur contribution en termes de pourcentage de succès et de probabilité d'échec. Bien que le travail soit intéressant, bien expliqué et bien présenté, il serait plus judicieux d'évaluer selon le pourcentage d'échec (ou de succès) ainsi qu'en fonction du temps nécessaire pour prendre en compte le changement de prix d'un service Cloud Computing en retournant le meilleur service (car une réponse en temps réel n'est pas forcément qu'une réponse très rapide).

Sharma *et al.* (2016b) ont proposé une approche pour prédire les paramètres aidant (attirant ou incitant) à adopter le Cloud Computing. Afin d'y parvenir, les chercheurs ont d'abord amélioré deux techniques. Ces dernières s'appuient sur *la régression linéaire multiple* et sur *les réseaux de neurones*. Même si les auteurs ont testé *via* des cas d'études les deux techniques améliorées, ils n'ont pas trouvé le même ordre d'importance des paramètres (pour le même ensemble de critères au départ). En plus de cela, les auteurs ne se sont pas penchés sur la raison d'une telle différence.

Li *et al.* (2017) ont proposé une méthode de décision conjonctive à choix multiple afin de composer les services Cloud Computing à des fins préalablement déterminées (*Cloud Manufacturing*). Pour y parvenir, les chercheurs ont d'abord déterminé leurs critères de sélection. En d'autres termes, ils ont dit comment ils allaient juger un service unique. Ensuite, les auteurs ont utilisé les réseaux de neurones, la logique floue, le processus d'analyse hiérarchique, TOPSIS, etc., afin de sélectionner, composer et évaluer les services Cloud Computing. Bien que le travail soit bien expliqué, illustré et présenté, les chercheurs ont néanmoins négligé une évaluation des performances de la contribution proposée, ainsi que la comparaison avec d'autres approches similaires. En plus de cela, nous ne recommandons

pas cet article pour un chercheur non expert dans les méthodes utilisées, car une certaine implicité a été détectée quant aux combinaisons de ces dernières.

Alipoufard *et al.* (2017) ont proposé un système complet sélectionnant les instances Cloud Computing de type IaaS à des fins d'analyses Big Data. Pour y parvenir, les chercheurs ont utilisé l'apprentissage statistique et l'optimisation bayésienne. Le travail est bien expliqué, joliment illustré et rigoureusement présenté. En plus de cela, les auteurs ont pu expérimenter sur des exemples réels de services Cloud Computing de type IaaS et ont pu aussi évaluer les performances de leur approche, mais ils n'ont pas jugé nécessaire de comparer leur contribution avec des travaux similaires. Ceci dit, l'article reste toujours recommandé à tout chercheur voulant contribuer dans la sélection des services Cloud Computing en utilisant l'apprentissage statistique ou les algorithmes bayésiens.

Wassim *et al.* (2018) ont proposé une analyse de facteurs en utilisant l'apprentissage machine non supervisé afin d'évaluer les services Cloud Computing. Pour ce faire, les chercheurs ont considéré deux sortes de métriques de qualité, la première est le retour des clients et la deuxième est le retour des serveurs. Ensuite, ils ont analysé les critères des services Cloud Computing, grâce à l'apprentissage machine. De ce fait, ils ont utilisé les retours des clients afin d'avoir des mesures de qualité objectives et subjectives. À la fin, ils ont pu calculer un score pour les services Cloud Computing de manière rigoureuse. C'est une contribution très bien expliquée, et nous remarquons les efforts des auteurs pour l'évaluation de leur approche. Ceci dit, les chercheurs n'ont pas pris la peine de comparer leur travail avec d'autres similaires.

Les travaux cités dans cette section forment une liste non exhaustive des articles scientifiques dans la sélection des services Cloud Computing utilisant l'apprentissage machine. Nous ne pouvions pas les citer tous, vu l'ampleur qu'a prise l'intelligence artificielle dans le monde de l'informatique, mais nous avons pu constater une négligence commune à propos du temps nécessaire pour apprendre. Effectivement, les chercheurs se désintéressent souvent d'évaluer le temps nécessaire pour leurs approches en vue d'apprendre. En d'autres termes, les chercheurs ont tendance à évaluer leurs approches proposées selon le taux de succès (très souvent), la précision (assez souvent), le taux d'échec (peu fréquent), mais ils négligent continuellement d'évaluer le temps nécessaire à l'apprentissage de leurs approches.

5.4.2. *Heuristiques*

Afin de résoudre des problèmes d'optimisation, de recherche ou les deux, certains chercheurs font appel aux heuristiques et métaheuristiques. Dans la vie de

tous les jours, nous remarquons que les êtres humains éliminent certains cas durant leurs recherches dans le but de gagner du temps. Sachant que les solutions éliminées peuvent être justes (valides) et/ou optimales, nous éliminons ce genre de situations généralement par manque de temps. C'est le principe même des heuristiques.

Gao *et al.* (2012) ont proposé une technique afin d'équilibrer les charges des processus entre les clients mobiles et leurs services dans un nuage informatisé. Pour ce faire, les chercheurs ont d'abord fait le point sur quelques concepts théoriques nécessaires (comme *Cloudlet*, *Coordinated Mobile Device*, etc.). Ensuite, ils ont présenté le modèle sur lequel ils se sont reposés. Ce dernier regroupe les caractéristiques des processus, des réseaux, etc. Après, ils ont proposé un algorithme avec une heuristique cherchant à équilibrer les charges des processus pour les clients et les services. L'algorithme produit des plans de décharges vérifiant certaines fonctions d'utilité préalablement établies. L'heuristique de son côté distingue entre une solution optimale et une autre peu optimale durant la recherche. En jetant un coup d'œil sur les tests de performance de la contribution faite par les auteurs eux-mêmes, nous pouvons voir que l'approche n'a pas été comparée avec d'autres similaires. En plus de cela, les chercheurs ont négligé la consommation énergétique des deux côtés (client et serveur), le coût du service en nuage informatisé, etc. Ceci dit, le travail reste de loin et de près intéressant. Par conséquent, nous le recommandons à tout type de chercheurs voulant contribuer dans l'équilibrage de charges (que ce soit dans le domaine de Cloud Computing, de *Grid Computing*, etc.).

Nacsimento *et al.* (2016) ont proposé une amélioration et une facilitation de la mise en échelle des instances virtuelles dans les nuages informatisés. Afin d'y parvenir, ils ont d'abord investigué sur l'efficacité des techniques d'apprentissage machine pour passer à l'échelle. Ensuite, ils ont proposé un ensemble d'heuristiques afin d'améliorer les algorithmes d'apprentissages. Ceci dit, l'objectif à optimiser n'est pas très clair (comme temps de démarrage de l'instance virtuelle, maximisation du nombre de machines virtuelles pour chaque machine physique, etc.). À noter également que le travail n'a été ni comparé, ni évalué, mais les auteurs ont pris la peine de bien expliquer la problématique et de dérouler un cas d'étude.

Hoang *et al.* (2016) ont proposé une approche afin de gérer les requêtes clients dans un environnement Cloud Computing, en se basant sur l'heuristique de colonies de fourmis et le *Particle Swarm Optimisation* (PSO). Pour ce faire, les chercheurs se sont d'abord fixé quelques objectifs (ils peuvent faire office de fonctions d'utilité) qui sont : la minimisation du coût global du système, la satisfaction des mesures de qualité de service et la maximisation du profit pour les fournisseurs de services Cloud Computing. Le travail est bien expliqué, joliment illustré, évalué, mais pas comparé à d'autres approches similaires. Ceci dit, l'article demeure intéressant et nous le recommandons vivement aux

chercheurs voulant contribuer dans la planification, l'ordonnancement et les gestions des requêtes clients dans un environnement Cloud Computing.

De même pour Xue *et al.* (2016), qui ont proposé une approche afin d'ordonnancer les processus dans un nuage informatisé. Se basant sur le *Particle Swarm Optimisation* et une heuristique, les chercheurs s'étaient fixé un seul objectif, qui est la maximisation de la qualité de service. Les auteurs ont pris la peine d'expliquer leur travail, d'évaluer leur approche, de comparer leur contribution par rapport à d'autres similaires. Ceci dit, les chercheurs ont exprimé leur objectif (maximisation de la qualité de service) en termes de minimisation du temps total d'exécution des requêtes clients, minimisation du temps total nécessaire à la terminaison des tâches clients et minimisation du coût total pour l'exécution des requêtes clients, tandis que la qualité de service dans ce sens s'exprime plus par la minimisation du temps moyen de terminaison de tâches pour les clients.

De leur côté, Samieifar et Mardukhi (2017) ont proposé une approche afin d'allouer les ressources d'un nuage informatisé de façon dynamique. Considérant le problème comme étant *NP-Hard*, les chercheurs ont proposé une métaheuristique combinant les algorithmes génétiques et la compétition coloniale. Les auteurs ont très bien expliqué les concepts utilisés, très bien illustré l'architecture et le contexte de la recherche, et ont comparé leur contribution par rapport à d'autres similaires. Même si les auteurs ont pris en compte le temps d'exécution et le coût des services, un chercheur peut évaluer son approche en termes de ressources consommées par les services. Par conséquent, nous recommandons le travail pour tout chercheur voulant contribuer dans l'allocation des ressources des nuages informatisés.

À la fin, Hajlaoui *et al.* (2017) ont proposé un système découvrant et sélectionnant les services Cloud Computing de type IaaS. Pour ce faire, ils ont introduit deux heuristiques qui sont *Hungarian* et *Volgenant-Jonker*. Ensuite, ils ont utilisé la théorie des graphes, la programmation linéaire, ainsi que la transformation linéaire symétrique, afin de calculer la qualité des services découverts. À la fin et pour démontrer l'efficacité de leur approche, les auteurs ont comparé entre les concepts utilisés, et non entre les approches proposées. Ceci dit, l'article demeure intéressant et recommandé à lire pour tout chercheur voulant contribuer dans la découverte et/ou la sélection des services Cloud Computing.

Dans cette section, nous avons présenté quelques travaux utilisant les heuristiques ou métaheuristiques afin de sélectionner, évaluer, composer ou découvrir les services Cloud Computing. Cette liste de contributions est loin d'être exhaustive, mais vous pourrez avoir un semblant d'idées pour vos futures recherches. En plus de cela, nous avons remarqué une négligence répandue dans les travaux proposant une

ou plusieurs heuristiques. Effectivement, les chercheurs ne comparent en général pas leurs approches utilisant leurs heuristiques proposées avec leurs approches utilisant des heuristiques standards (déjà proposées dans la littérature). Ceci a pour but d'évaluer l'heuristique et l'algorithme séparément.

5.4.3. *Les systèmes multiagents intelligents*

Les systèmes multiagents intelligents sont la première alternative à laquelle les chercheurs en IA font appel lorsqu'il s'agit d'étudier l'impact d'un agent sur les autres. Prenons un exemple, dans lequel nous considérons N agents intelligents clients et M agents intelligents fournisseurs de services Cloud Computing. Supposons que, parmi les services proposés, un fournisseur se distingue par rapport aux autres par sa qualité optimale et son prix réduit (qu'on note m_i). Nous constatons une augmentation des demandes des N agents clients à ce service du fournisseur m_i. Par conséquent, ce dernier sera obligé de prendre des précautions (ou mesures en général) afin d'éviter une surcharge de ses serveurs. Il peut par exemple ajouter des serveurs matériels (ce qui a tendance à augmenter le prix), ou augmenter le nombre de serveurs virtuels pour chaque machine physique (ce qui a tendance à diminuer la qualité de service), etc. Dans tous les cas, les mesures prises pourront avoir un impact sur les clients, qui pourront changer de fournisseur.

Ceci n'est qu'un simple exemple parmi d'autres des applications des systèmes multiagents intelligents dans la sélection des services Cloud Computing. Dans cette section, nous allons présenter quelques contributions utilisant l'intelligence artificielle et les systèmes multiagents, afin de découvrir, sélectionner, évaluer ou composer les services Cloud Computing.

Rabbani *et al.* (2014) ont proposé une méthode de sélection des services Cloud Computing, en faisant appel à des agents intelligents. Les chercheurs ont remarqué que, si nous entraînions les agents suffisamment, ils pouvaient à la fin remarquer les similarités et les différences entre les services pour pouvoir les sélectionner. Vu que l'objectif était de répondre au mieux aux requis des clients, il serait intéressant de voir l'évaluation d'une telle approche selon le temps d'apprentissage, la pertinence des résultats, etc., mais les auteurs n'ont intégré ni évaluation, ni comparaison à leur travail.

Chichin *et al.* (2014) ont proposé un marché intelligent pour les services Cloud Computing, en se basant sur les agents intelligents. Leur principe repose sur la simulation de cas par les agents intelligents (logiciels) selon plusieurs politiques, diverses situations, etc. Si un client veut utiliser la plateforme proposée, les auteurs lui recommandent l'agent le plus adéquat à ses besoins et à sa situation.

C'est un travail assez remarquable, vu que les chercheurs ont évalué leur approche et comparé avec plusieurs autres contributions similaires. En plus de cela, les critères de comparaison sont divers et variés. Nous recommandons vivement cet article à tout chercheur voulant contribuer dans la sélection des services Cloud Computing en utilisant les agents intelligents.

Lacheheub *et al.* (2016) ont proposé une approche afin de sélectionner les services Cloud Computing, en se basant sur les agents intelligents. Tout d'abord, les agents découvrent au fur et à mesure les services Cloud Computing. Ensuite, chaque agent les regroupe (en hiérarchie). De ce fait, les clients sélectionnent les services qui sont les plus proches de leurs anciens services (cas de changement de fournisseurs) ou de leurs applications en local (cas de migration aux nuages informatisés).

Parmi toutes nos contributions surveillées, (Lacheheub et Maamrin 2016) est le meilleur en termes d'explication, d'illustrations et d'évaluation. Les auteurs n'ont pas seulement expliqué et illustré leur approche, mais étudié un exemple, déroulé la démarche à suivre (mathématiquement, graphiquement et linguistiquement), évalué leur approche et comparé avec d'autres contributions. Par conséquent, nous recommandons (scientifiquement et méthodologiquement) cet article à tout chercheur voulant contribuer dans la sélection des services Cloud Computing en utilisant les agents intelligents.

Jahani *et al.* (2017) ont proposé un système nommé ARank, sélectionnant les services Cloud Computing et reposant sur les systèmes multiagents intelligents. Tout d'abord, chaque agent prend quelques services Cloud Computing. Ensuite, chacun évalue ses candidats selon la qualité de service. De ce fait, les agents incluent dans leur évaluation l'historique du taux de satisfaction des clients précédents de chaque service Cloud Computing. Le principe est simple à comprendre, vu que les auteurs ont bien expliqué le concept, mais nous avons remarqué une certaine incohérence entre l'objectif et l'évaluation (comparaison entre l'approche proposée, AHP, SVD). Les chercheurs affirment optimiser le temps d'attente des clients, alors qu'ils comparent leur contribution avec d'autres en termes de temps d'évaluation par rapport au nombre total de services Cloud Computing. Dans ce cas, le lecteur peut comprendre que les services ne sont pas évalués au moment de la sélection, alors que c'est le cas.

5.4.4. *La théorie des jeux*

Souvent, les problèmes mathématiques et/ou informatiques sont exprimés sous forme de jeux en suivant un modèle mathématique afin d'étudier des preneurs de décision dans

leurs environnements. Le modèle proie/prédateur est un bon début pour comprendre l'intérêt des théories des jeux. Prenons l'exemple d'une surface déterminée contenant un ensemble de prédateurs et un ensemble de proies. Nous remarquons tout de suite que les deux groupes de joueurs sont en compétition. Par exemple, les lions chassent les buffles et ils sont en compétition. Les lions ne peuvent pas se passer de la chasse en vue de survivre. Et les buffles ne peuvent pas s'empêcher de se protéger, de peur d'une extinction ou de la faiblesse du troupeau. Ceci dit, le prédateur a besoin de la résistance de la proie et il est bénéfique à sa façon pour elle. Si les buffles arrêtent de se défendre, alors les lions attaqueront plus facilement et plus fréquemment le troupeau. Ceci a pour conséquence d'affaiblir le troupeau ou de l'éteindre. Dans ces cas-là, si les buffles ne sont pas morts, ils seront obligés de migrer à une place plus sécurisée. Ce qui a pour conséquence de les rendre vulnérables durant le voyage et d'affaiblir la troupe prédatrice des lions (en effet, ces derniers auront moins de sources de nutrition). Par contre, si les lions arrêtent de chasser les buffles, ils seront éteints à coup sûr et les buffles auront plus de chance de se multiplier et d'accroître la force de leur troupeau, mais les buffles sont aussi les prédateurs d'autres êtres vivants dans la zone étudiée. Si leurs sources de nutrition ne suffisent pas, ils seront à leur tour éteints.

La théorie des jeux ne se limite pas à la simple étude de preneur de décisions (que ce soit dans un jeu ou dans la réalité). Elle étudie aussi l'impact d'une décision sur les joueurs (les autres preneurs de décision) dans le même environnement (cas de partage de ressources matérielles dans un nuage informatisé).

Hassan *et al.* (2014) ont proposé une approche allouant les ressources Cloud Computing dans un contexte de fédération horizontale dynamique de nuages informatisés. Afin d'y arriver, les chercheurs ont d'abord présenté une architecture d'un système de fédération horizontale et dynamique de services Cloud Computing. Ensuite, les auteurs ont étudié deux cas de maximisation pour la fonction d'utilité pour l'allocation des ressources suivant le contexte présenté. Ensuite, ils ont proposé un algorithme qui sélectionne les ressources, en se basant sur les prix de ces derniers. Le travail est bien présenté, illustré et expliqué, mais les auteurs n'ont pas jugé nécessaire de mettre une comparaison de leur approche par rapport aux contributions déjà existantes. En plus de cela, leurs mesures d'évaluation (maximisation de la fonction d'utilité, maximisation du bien-être social, minimisation du nombre de machines virtuelles par fournisseurs Cloud Computing, etc.) ne concordent pas avec les objectifs de leur approche (maximisation de la robustesse et réduction du temps).

Do *et al.* (2015) ont présenté un modèle de coopération interclients afin de sélectionner le meilleur service Cloud Computing. Pour ce faire, ils ont d'abord divisé l'ensemble des clients en groupes. Ensuite, ils ont fait en sorte que chaque

client choisisse un fournisseur au hasard pour un début. Puis, chaque client communique les caractéristiques de son fournisseur aux autres clients qui sont avec lui dans le même groupe. À ce moment-là, la phase découverte est achevée et les clients vont devoir changer de fournisseur si les caractéristiques d'un meilleur fournisseur leur sont communiquées. Vu que l'objectif de l'étude était de présenter une définition compréhensive des marchés des services Cloud Computing hétérogènes, nous pensons que l'objectif a été largement rempli, car les chercheurs ont proposé une modélisation mathématique du problème ainsi qu'une évaluation numérique. Ceci dit, les auteurs n'ont pas jugé nécessaire de comparer leur approche avec d'autres similaires (si elles existent encore).

Ardagna *et al.* (2015) ont proposé une approche afin de maximiser les revenus et de minimiser les coûts des fournisseurs de services Cloud Computing SaaS, eux-mêmes clients de services Cloud Computing IaaS. Pour ce faire, ils ont proposé un modèle de jeu de *Nash* dans lequel les fournisseurs de services Cloud Computing SaaS parient sur les ressources allouées dans le niveau IaaS. Les chercheurs ont pu prouver l'existence d'un certain point d'*équilibre de Nash*, et que la convergence se passe dans un nombre d'itérations limité. Ceci dit, les auteurs ont mentionné avoir comparé leur approche avec d'autres techniques similaires dans l'état de l'art, mais nous ne trouvons pas ce genre d'évaluation. Les chercheurs se sont contentés d'évaluer la qualité de service au cours d'un certain laps de temps.

Liu *et al.* (2016) ont proposé une étude de l'impact de la communication des modèles de facturation sur la vitesse de convergence vers l'état d'équilibre dans un environnement fournisseur/client au sein d'un nuage informatisé. Afin d'y parvenir, les chercheurs ont d'abord formalisé le problème et présenté le modèle de jeu de *Stackelberg*. Ensuite, ils ont présenté les stratégies optimales pour les clients, comme pour les fournisseurs, afin de maximiser leurs utilités (les auteurs ont démontré aussi l'optimalité de telles stratégies). De ce fait, ils ont pu voir l'arrivée au point d'équilibre de Stackelberg en déroulant les stratégies optimales présentées dans un environnement en nuage informatisé. À la fin, les chercheurs ont modifié quelques paramètres de la simulation, afin de voir leurs impacts sur l'équilibre et le temps nécessaire pour l'atteindre. Malgré cela, une comparaison avec les approches similaires, ou au moins une approche basée sur un modèle de jeu de Nash, serait grandement appréciée.

Wu *et al.* (2016) ont proposé un mécanisme d'allocation de ressources des nuages informatisés auto-organisé. Pour ce faire, les chercheurs ont proposé deux nouvelles stratégies économiques allouant les ressources en se basant sur l'architecture et les prix. Plus spécifiquement, ils ont utilisé un système d'*Enchères de Vickrey Modifié* quand les ressources sont suffisantes, et le système de *Double*

Enchères Continues quand les ressources sont insuffisantes. Malgré l'explication remarquable des auteurs, ils n'ont pas jugé nécessaire de comparer leur approche avec d'autres techniques similaires. En plus de cela, les chercheurs ont évalué leur contribution en termes de coût de procuration des ressources par rapport au nombre de fournisseurs et à l'efficacité de l'exécution. Ceci dit, il existe beaucoup d'autres paramètres qui peuvent entrer en jeu, comme le délai de procuration, le nombre de machines virtuelles par rapport au nombre de machines physiques, etc.

Cette section vient de présenter un ensemble de travaux dans la sélection de services Cloud Computing en se basant sur les techniques d'intelligence artificielle dans le domaine de la théorie des jeux. Nous remarquons bien sûr qu'il s'agit d'une intelligence collective, puisque la majorité des contributions considèrent les clients et/ou les fournisseurs comme étant des joueurs. Chacun a un objectif à réaliser et les joueurs doivent partager les mêmes ressources matérielles. Techniquement parlant, les chercheurs ont essayé d'abord de prouver l'existence d'un certain point d'équilibre (explicitement ou implicitement). Ensuite, ils ont essayé de trouver un moyen d'y arriver. À la fin, les auteurs avaient juste à prouver que le point d'équilibre pouvait être atteint en un nombre fini d'itérations (un temps plus ou moins limité). Ceci dit, nous avons remarqué certaines opportunités de recherche non exploitées, comme la possibilité qu'un point d'équilibre ne se produise pas même s'il existe auparavant, l'hétérogénéité des joueurs (clients et/ou fournisseurs), leurs objectifs finaux, leurs stratégies, etc.

Nous pouvons citer d'autres techniques dans l'intelligence artificielle utilisées pour sélectionner les services Cloud Computing, comme les algorithmes génétiques. Ils regroupent par exemple l'algorithme de colonies fourmis (Gao 2014), l'algorithme d'essaims d'abeilles (Tian *et al*. 2013 ; Seghir *et al*. 2016 ; Xu *et al*. 2017 ; Zhou et Yao 2017), l'algorithme de recherche des corbeaux (Satpathy *et al*. 2017), etc.

L'intelligence artificielle possède toute une panoplie de méthodes, de techniques et de procédés, afin de répondre au mieux à ses problématiques. Par conséquent, nous n'avons pas pu citer toutes les contributions, ni même une contribution de chaque sous-domaine de l'I.A.

5.5. Conclusion

Ce chapitre a présenté un état de l'art sur les travaux de sélection, de composition, d'évaluation ou de recommandation des services Cloud Computing utilisant l'intelligence artificielle. Cet écrit donne un aperçu général sur les

problématiques considérées, les approches suivies et les outils utilisés dans les deux domaines, sauf que ce genre d'études de surveillance est rarement précis et/ou détaillé.

Dans un futur proche, nous voudrions orienter ce travail vers une comparaison, en plus de la surveillance. En d'autres termes, le lecteur aura les résultats de tests de performance des approches surveillées (en plus des critiques).

Nous voulons aussi intégrer les limites de l'intelligence artificielle dans le cadre de la sélection des services Cloud Computing. En d'autres termes, nous voulons intégrer l'impact des limites de l'intelligence artificielle sur les études surveillées.

De plus, nous voulons approfondir la surveillance des travaux basés sur la théorie des jeux, vu qu'ils sont divers et variés. À titre d'exemple, la technique de pari bidirectionnel est un procédé souvent utilisé dans le domaine de la théorie des jeux en utilisant l'intelligence artificielle.

5.6. Bibliographie

Ahmad, I., Bakht, H., Mohan, U. (2017). Cloud Computing – A Comprehensive Definition. *Journal of Computing and Management Studies*, 1(1), 8.

Ahmed, U., Raza, I., Hussain, S.A. (2019). Trust Evaluation in Cross-Cloud Federation: Survey and Requirement Analysis. *ACM Computing Surveys (CSUR)*, 52(1), 19.

Aishwarya, S., Muzammil, H. (2017). Pricing schemes in Cloud Computing: a review. *International Journal of Advanced Computer Research*, 7(29), 60.

Alipoufard, O. *et al.* (2017). Cherrypick: Adaptively unearthing the best cloud configurations for big data analytics. Dans *Symposium on Networked Systems Design and Implementation*. NSDI, Boston, 469–482.

Andrikopoulos, V., Song, Z., Leymann, F. (2013). Supporting the Migration of Applications to the Cloud through a Decision Support System. Dans *IEEE Sixth International Conference on Cloud Computing*. IEEE, Santa Clara, 565–572.

Ardagna, D., Ciavotta, M., Passacantando, M. (2015). Generalized nash equilibria for the service provisioning problem in multi-cloud systems. *IEEE Transactions on Services Computing*, 10(3), 381–395.

Armbrust, M. *et al.* (2010). A View of Cloud Computing. *Communications of the ACM*, 53(4), 50–58.

Beloglazov, A., Abawajy, J., Buyya, R. (2012). Energy-aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing. *Future generation computer systems*, 28(5), 755–768.

Chichin, S. *et al.* (2014). Smart Cloud Marketplace-Agent-Based Platform for Trading Cloud Services. Dans *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE/WIC/ACM, Varsovie.

Crevier, D. (1993). *AI: the Tumultuous History of the Search for Artificial Intelligence*. Basic Books, New York.

Das, A.K. *et al.* (2014). A QoS and profit aware cloud confederation model for IaaS service providers. Dans *8th International Conference on Ubiquitous Information Management and Communication*. ACM, Siem Reap.

Do, C.T. *et al.* (2015). Dynamics of service selection and provider pricing game in heterogeneous cloud market. *Journal of Network and Computer Applications*, 69, 152–165.

Domingos, P. (dir.) (2015). How does your Brain Learn?. Dans *The Master Algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books, New York.

Foss, B., Dodwell, P.C. (1966). *New Horizons in Psychology*. Penguin Books, Londres.

Gao, Z. (2014). The Allocation of Cloud Computing Resources Based on the Improved Ant Colony Algorithm. Dans *Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*. IHMSC, Hangzhou.

Gao, B. *et al.* (2012). From Mobiles to Clouds: Developing Energy-Aware Offloading Strategies for Workflows. Dans *ACM/IEEE 13th International Conference on Grid Computing*. ACM/IEEE, Beijing.

Gupta, S. *et al*. (2015). Risk-driven Framework for Decision Support in Cloud Service Selection. Dans *15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*. IEEE/ACM, Shenzhen.

Hajlaoui, J.E. *et al*. (2017). Qos Based Framework for Configurable Iaas Cloud Services Discovery. Dans *IEEE International Conference on Web Services (ICWS)*. IEEE, Hawaï.

Hassan, M.M. *et al.* (2014). Cooperative Game-based Distributed Resource Allocation in Horizontal Dynamic Cloud Federation Platform. *Information Systems Frontiers*, 16(4), 523–542.

Hayes, B. (2008). Cloud Computing. *Communications of the ACM*, 51, 9–11.

Hoang, H.N. *et al.* (2016). Admission Control and Scheduling Algorithms Based on ACO and PSO Heuristic for Optimizing Cost in Cloud Computing. Dans *Studies in Computational Intelligence*, Hoang, H.N., Le, V.S. *et al*. (dir.). Springer, Berlin.

Jahani, A. *et al* (2017). ARank: A Multiagent Based Approach for Ranking of Cloud Computing Service. *Scalable Computing: Practice and Experience*, 18(2), 105–116.

Kang, D.S. *et al*. (2013). Adaptive process execution in a service cloud: service selection and scheduling based on machine learning. Dans *IEEE 20th International Conference on Web Services*. IEEE, Santa Clara.

Kaplan, A., Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.

Lacheheub, M.N., Maamrin, R. (2016). Towards a Construction of an Intelligent Business Process Based on Cloud Services and Driven by Degree of Similarity and QoS. *Information Systems Frontiers*, 18(6), 1085–1102.

Le, S. *et al*. (2014). Multicriteria Decision Making with Fuzziness and Criteria Interdependence in Cloud Service Selection. Dans *International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, Beijing.

Legg, S. *et al*. (2007). A Collection of Definitions of Intelligence. *Frontiers in Artificial Intelligence and applications*, 157, 17.

Li, L. *et al*. (2017). A Conjunctive Multiple-Criteria Decision-Making Approach for Cloud Service Supplier Selection of Manufacturing Enterprise. *Advances in Mechanical Engineering*, 9(3).

Liaqat, M. *et al*. (2017). Federated Cloud Resource Management: Review and Discussion. *Journal of Network and Computer Applications*, 77, 87–105.

Liu, C. *et al*. (2016). Stackelberg Game Based Optimal Workload Allocation and Pricing Mechanism in Crowdsourcing. Dans *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*. IEEE, Atlanta.

Mandal, A.K., Changder, S., Sarkar, A. (2013). Selection of Services for Data-Centric Cloud Applications: A QoS Based Approach. Dans *International Conference on Advanced Computing, Networking and Security*. IEEE, Mangalore, 102–107.

McCorduck, P. (2009). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. AK Peters/CRC Press, Natick.

Mell, P., Grance, T. (2011). The NIST Definition of Cloud Computing. *Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology*, 800(145), 1–7.

Nascimento, D.C. *et al*. (2016). Applying Machine Learning Techniques for Scaling out Data Quality Algorithms in Cloud Computing Environments. *Applied Intelligence*, 45(2), 530–548.

Nilsson, N.J. (1998). *Artificial intelligence: a new synthesis*. Morgan Kaufmann, Burlington.

Pandey, G. *et al.* (2011). Current Cloud Scenario Review and Cost Optimization by Efficient Resource Provisioning. Dans *The Fourth Annual ACM Bangalore Conference*. ACM, Bangalore.

Papathanasiou, J. *et al*. (2015). A Comparative Analysis of Cloud Computing Services Using Multicriteria Decision Analysis Methodologies. *International Journal of Information and Decision Sciences*, 7(1), 51–70.

Poole, D. *et al*. (1998). *Computational intelligence: a logical approach.* Oxford University Press, New York.

Rabbani, I.M. *et al.* (2014). Intelligent Cloud Service Selection Using Agents. Dans *9th International Conference on Computing and Information Technology (IC2IT2013).* IC2IT, Bangkok.

Rehman, Z.U., Hussain, O.K., Hussain F.K. (2012). Iaas Cloud Selection Using MCDM Methods. Dans *IEEE 9th International Conference on e-business Engineering*. IEEE, Hangzhou.

Rehman, Z.U., Hussain, O.K., Hussain, F.K. (2014). Time series QoS forecasting for management of cloud services. Dans *Ninth International Conference on Broadband and Wireless Computing, Communication and Applications.* IEEE, Guangdong.

Russel, S.J., Norvig, P. (2016). *Artificial intelligence: a modern approach.* Pearson Education Limited, Harlow.

Samieifar, S., Mardukhi, F. (2017). Dynamic Resource Allocation in Cloud Computing Using a Combination of Meta-heuristic Algorithms. *International Journal of Computer Science and Network Security (IJCSNS)*, 17, 332.

Satpathy, A. *et al*. (2017). A Resource Aware VM Placement Strategy in Cloud Data Centers Based on Crow Search Algorithm. Dans *4th International Conference on Advanced Computing and Communication Systems.* ICACCS, Coimbatore.

Schank, R.C. (1991). Where's the AI?. *AI magazine*, 12(4), 38–38.

Seghir, F. *et al*. (2016). A new discrete imperialist competitive algorithm for QoS-aware service composition in cloud computing. Dans *The International Symposium on Intelligent Systems Technologies and Applications*. ISTA, Jaipur.

Sharma, S.K. *et al*. (2016a). Predicting motivators of cloud computing adoption: A developing country perspective. *Computers in Human Behavior*, 62, 61–69.

Sharma, Y. *et al*. (2016b). Reliability and energy efficiency in cloud computing systems: Survey and taxonomy. *Journal of Network and Computer Applications*, 74, 66–85.

Singh, P. *et al*. (2017). A review of task scheduling based on meta-heuristics approach in cloud computing. *Knowledge and Information Systems*, 52(1), 1–51.

Sun, L. *et al*. (2014). Cloud service selection: State-of-the-art and future research directions. *Journal of Network and Computer Applications*, 45, 134–150.

Syntec Numérique (2012). Cloud Computing : nouveaux modèles !. Livre blanc, Syntec Numérique, Paris.

Tian, S. *et al*. (2013). A Discrete Hybrid Bees Algorithm for Service Aggregation Optimal Selection in Cloud Manufacturing. Dans *14th International Conference on Intelligent Data Engineering and Automated Learning*. IDEAL, Hefei.

Wassim, M.U. *et al*. (2018). Cloud Service Providers Optimized Ranking Algorithm Based on Machine Learning and Multi-Criteria Decision Analysis. *Preprints*.

Whaiduzzaman, M. *et al*. (2014). Cloud service selection using multicriteria decision analysis. *The Scientific World Journal*, 2014.

Wu, X. *et al*. (2016). A scalable and automatic mechanism for resource allocation in self-organizing cloud. *Peer-to-peer networking and applications*, 9(1), 28–41.

Xiaogang, W. *et al*. (2015). Dynamic cloud service selection using an adaptive learning mechanism in multi-cloud computing. *Journal of Systems and Software*, 100, 195–210.

Xu, Y. *et al*. (2016). Survey on Privacy Preserving for Intelligent Business Recommendation in Cloud. Dans *Wireless Communications, Networking and Applications*, Shi, G., Ming, Y. (dir.). Springer India, New Delhi.

Xu, X. *et al*. (2017). S-ABC: A paradigm of service domain-oriented artificial bee colony algorithms for service selection and composition. *Future Generation Computer Systems*, 68, 304–319.

Xue, S. *et al*. (2016). A heuristic scheduling algorithm based on PSO in the cloud computing environment. *International Journal of u- and e-Service, Science and Technology*, 9, 349–362.

Zhou, J., Yao, X. (2017). A hybrid artificial bee colony algorithm for optimal selection of QoS-based cloud manufacturing service composition. *The International Journal of Advanced Manufacturing Technology*, 88(9/12), 3371–3387.

6

Le déchargement intelligent des calculs dans le contexte du *Mobile Cloud Computing*

Zeinab MOVAHEDI
Iran University of Science and Technology, Téhéran, Iran

6.1. Introduction

Aujourd'hui, avec le progrès des technologies mobiles aussi bien sur l'aspect logiciel que matériel, les appareils mobiles sont utilisés de plus en plus largement, devenant une partie inséparable de notre vie quotidienne (Li *et al*. 2014). En raison de ces avancées, le rôle des téléphones mobiles a changé, passant d'un simple appareil de communication à un outil indispensable pour de nombreuses d'autres applications répondant à nos divers besoins de tous les jours. Parmi ces applications, nous pouvons citer les simulations, la compression/décompression, le traitement d'images, la réalité virtuelle, les jeux vidéo, etc. De plus, et suite aux fortes attentes des utilisateurs mobiles, d'autres applications de plus en plus sophistiquées seront aussi à utiliser sur les appareils mobiles dans l'avenir. Cependant, le développement de ces nouvelles applications est restreint par les limitations des appareils mobiles en termes d'espace de stockage, de puissance de calcul et de durée de vie de leur batterie (Chen *et al*. 2015).

D'autre part, les avancées récentes des réseaux de télécommunication, en termes de débit offert pour la transmission de données ainsi que le nombre d'utilisateurs pris en

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR. © ISTE Editions 2020.

charge, ouvrent l'horizon pour pouvoir décharger[1] les calculs et les données gourmandes vers le centre de calcul et de stockage du *Cloud*. Ce dernier paradigme, appelé *Mobile Cloud Computing* (MCC), pourrait donner naissance à un nombre important d'applications émergentes, qui ne sont soit pas envisageables avec la performance d'aujourd'hui des appareils mobiles, soit faisables mais avec un temps de traitement important et une consommation d'énergie de batterie significative, demandant un rechargement fréquent de l'appareil mobile (Abolfazli *et al*. 2014 ; Khan *et al*. 2014).

REMARQUE. *Mobile Cloud Computing* est un nouveau paradigme de la technologie mobile dans lequel les capacités des appareils mobiles sont étendues en utilisant les ressources des centres de données et de calculs du Cloud.

Cependant, l'efficacité de déchargement de données et de calculs vers le Cloud dépend fortement de la qualité du lien radio, qui est essentiellement variable dans l'espace et dans le temps en termes de qualité de signal, d'interférence, de débit de transmission, etc. Ce contexte dynamique nécessite l'emploi d'un mécanisme de décision de déchargement en charge de déterminer si un déchargement de données et de calculs sera bénéfique. Le bénéfice de déchargement est en général évalué en termes de temps d'exécution achevé et de montant d'énergie consommé, tout en considérant le statut du réseau sous-jacent (Zhang *et al*. 2016). De plus, étant donné le chevauchement des réseaux mobiles tels que WIMAX et LTE avec les réseaux sans-fil locaux (WAN[2]) comme le Wi-Fi et *femtocell*, ainsi que la propriété multi homing des appareils mobiles d'aujourd'hui, le mécanisme de décision de déchargement pourrait aussi contribuer à la sélection du réseau d'accès approprié pour la transmission des données de la tâche déchargée (Magurawalage *et al.* 2015).

En outre, la compétition des fournisseurs du service Cloud pour attirer plus de clients pourrait les amener à jouer un rôle sur le marché du MCC, menant à un environnement multiCloud pour les utilisateurs des appareils mobiles. Dans un tel environnement, le mécanisme de décision de déchargement devrait ainsi déterminer le fournisseur de Cloud approprié en considérant la qualité de service offerte et le coût des ressources de calcul et de stockage dédiées à une requête d'utilisateur. Cependant, le prix du marché peut varier suivant les conditions de l'offre et de la demande et la compétition inter fournisseurs (Hong et Kim 2019).

Ce contexte dynamique multiréseaux d'accès et multiClouds nécessite d'enrichir les mécanismes de décision de déchargement par des outils d'intelligence artificielle (IA), permettant une décision multicritères répondant au mieux à des besoins des utilisateurs

1. *Offload*, en anglais.
2. *Wireless Local Area Network*.

mobiles en monde réel. Dans un tel contexte, ce chapitre traite les différentes applications d'IA pour optimiser l'efficacité de déchargement du point de vue d'utilisateurs mobiles, de fournisseurs du Cloud et des fournisseurs des réseaux d'accès.

Le reste de ce chapitre est organisé comme suit. La section 6.2 définit les notions de base liées au problème de déchargement. Dans la section 6.3, nous présentons l'architecture de MCC dans le contexte des réseaux d'accès classique aussi bien que dans celui des réseaux d'accès radio à base de Cloud[3]. Ensuite, nous détaillons le problème de décision de déchargement dans la section 6.4. La section 6.5 traite les solutions proposées à base d'IA pour résoudre le problème de déchargement. Enfin, la section 6.6 conclut ce chapitre et propose quelques orientations de recherche pour l'avenir.

6.2. Définitions de base

Comme décrit précédemment, le déchargement des calculs consiste à confier la tâche de traitement du calcul aux serveurs du Cloud, afin d'optimiser le temps d'exécution de l'application mobile et l'énergie consommée pour exécuter la tâche.

Étant donné la mobilité de l'utilisateur mobile, la nature dynamique du canal radio et le coût variable des ressources de stockage et de traitement du Cloud, le déchargement de la totalité des calculs pourrait ne pas être toujours le meilleur choix. Par exemple, lorsque la bande passante[4] du réseau d'accès est faible, il est plus intéressant de décharger seulement une sous-partie de la tâche complète qui correspond aux calculs gourmands mais ayant des données d'entrées très légères.

REMARQUE. Le déchargement de la totalité du calcul pourrait, dans certaines situations, ne pas mener au temps d'exécution et à la consommation d'énergie optimale.

Par conséquent, le déchargement du calcul pourrait être fait *à grain fin*[5] ou *à grain grossier*[6], en considérant les caractéristiques de la tâche en question et les conditions du contexte environnant (Khan *et al*. 2015 ; Wu 2018).

Dans les sections suivantes, nous décrivons ces deux types de déchargement en détail.

3. *Cloud-Radio Access Networks*, en anglais.
4. *Bandwidth,* en anglais.
5. *Fine-grain Offloading*, en anglais.
6. *Coarse-grain Offloading*, en anglais.

6.2.1. *Déchargement à grain fin*

DÉFINITION.– *Le déchargement à grain fin consiste à externaliser seulement une sous-partie de la tâche originale pour être exécutée sur le Cloud.*

Pour permettre de déterminer la sous-partie de tâche dont l'externalisation optimisera le temps d'exécution de l'application et l'énergie consommée de l'appareil mobile, les composants de la tâche devront tout d'abord être extraits. Cette étape peut être effectuée en se basant sur un niveau de granularité, tel que classe, objet, *thread*, etc. Indépendamment du niveau de granularité choisi, l'extraction des composants de l'application pourrait être effectuée avec une approche d'analyse de code *statique*[7] ou *dynamique*[8]. Dans l'analyse de code statique, les composants de l'application sont extraits sans exécuter le code, alors que l'analyse de code dynamique est basée sur l'extraction des composants de l'application lors de l'exécution du code.

L'analyse de code statique permet d'extraire les composants et de créer le graphe de relation pondérée de l'application (WRG[9]) (décrit ci-après) avant que l'utilisateur ait besoin de prendre une décision de déchargement. Cela permettra de réduire le temps d'obtention de résultat de décision de déchargement. Cependant, l'extraction des composants de l'application basée sur l'analyse de code statique pourrait conduire à un résultat moins précis, en particulier pour certains types de granularité pour lesquels les composants du programme pourront varier d'une exécution à l'autre.

EXEMPLE. Si la granularité de déchargement est par objet, l'extraction des composants de l'application basée sur l'analyse de code dynamique est plus appropriée, car les objets d'un code pourront être créés et détruits suivant les conditions produites lors de l'exécution de l'application.

À partir des composants de l'application extraits par l'analyse statique ou dynamique, le graphe de relation pondéré de l'application peut être construit. Ce dernier consiste en un graphe $WRG = (V, E, W_V, W_E)$, dans lequel chaque sommet $v \in V$ représente un composant de l'application et chaque arête $e \in E$ décrit l'invocation entre les deux composants d'extrémités. Chaque $w_v \in W_V$ et $w_e \in W_E$ représente respectivement le poids d'un sommet v et d'une arête e du

7. *Static Code Analysis*, en anglais.
8. *Dynamic Code Analysis*, en anglais.
9. *Weighted Relation Graph*.

graphe. Chaque sommet et arête de ce graphe est respectivement pondéré selon le nombre d'instructions compris dans le composant correspondant et la quantité de données transmises entre les deux composants adjacents.

Un exemple de code d'une application et le graphe de relation pondéré correspondant sont illustrés dans la figure 6.1.

```
Public Class M {
Public static void main (String args[]){
                A a1 = new A();B b1 = new B();B b2 = new B();
                byte [] dataSize = new type [100];
                inti=0;

  while (i< 10){
                        var1 = a1.getdata(datasize);
                        var2 = b1.process(datasize/2);
                        i++;
  }
                for ( ; i<4; i++){
                        var3 = b2.process(datasize/2);
                        }
                }
  }
Class A {
        Public void getdata(byte datasize){
                ---
        }
  }
Class B {
        Public void process (byte [] datasize){
                type block = new type [10];
                c.arrangeBlock(block);
                }
  }
Class C {
        Public static void arrangeBlock (byte[] block) {
        ---
      }
        }
```

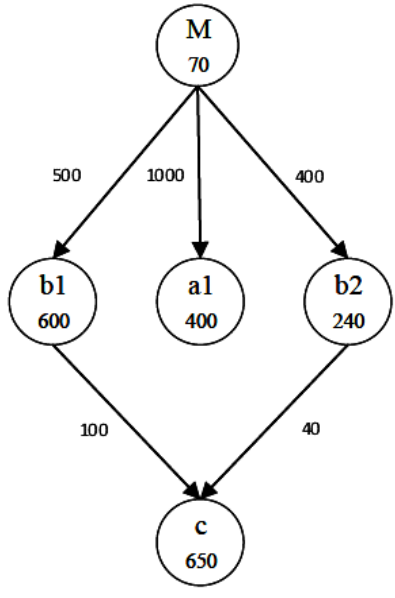


Figure 6.1. *Conversion du code d'une application à un graphe de relation pondéré (Movahedi 2018)*

À partir du graphe de relation pondéré, l'algorithme de décision de déchargement à grain fin devrait déterminer les composants de l'application dont leur déchargement aboutira au temps d'exécution et à la consommation d'énergie optimale. Cette décision est multicritères et son efficacité dépend de sa conscience de contexte environnant qui comprend l'environnement radio et Cloud sous-jacents.

La figure 6.2 illustre le modèle conceptuel décrivant les différentes étapes nécessaires pour prendre la décision de déchargement à grain fin.

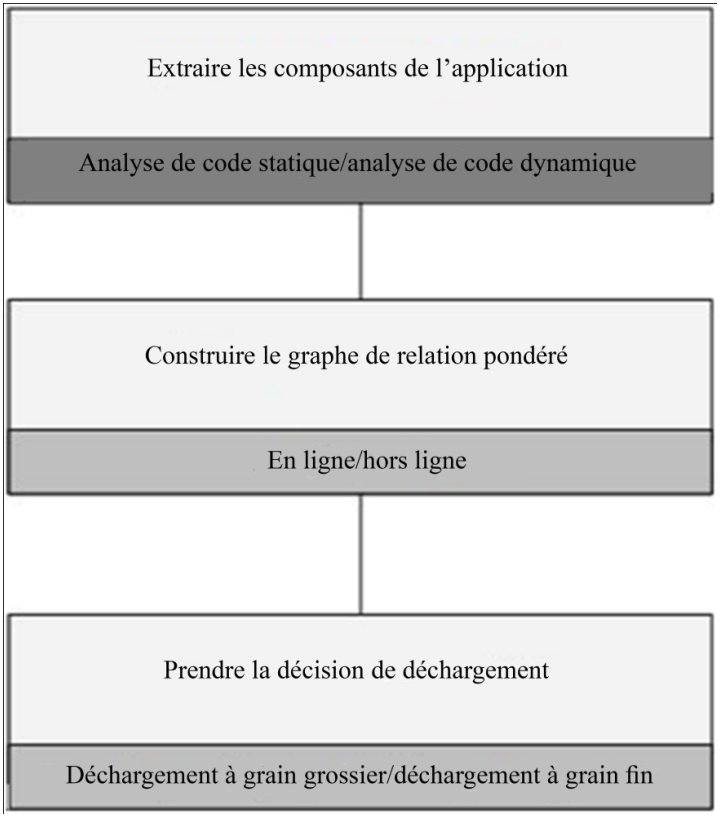


Figure 6.2. *Les étapes de la décision de déchargement*

6.2.2. *Déchargement à grain grossier*

Malgré les efficacités offertes par le déchargement à grain fin, certains travaux préfèrent soit décharger la totalité d'une application ou ne pas décharger du tout un programme. Cette dernière approche appelée « le déchargement à grain grossier » est motivée par la simplicité de la décision comparée au déchargement à grain fin.

DÉFINITION.– *Le déchargement à grain grossier consiste à externaliser la totalité d'un programme pour être exécutée sur le Cloud.*

REMARQUE. Le déchargement à grain grossier est avantageux en particulier quand un programme ne pourra pas être divisé en plusieurs composants.

Malgré la rapidité de décision offerte par le déchargement à grain grossier, cette approche perd les avantages liés au déchargement des sous-parties de l'application. Ces avantages concernent l'optimisation du temps de calcul de l'application ainsi que la quantité de batterie consommée de l'appareil mobile.

Afin de remédier à ce problème, certains travaux divisent une application en plusieurs sous-tâches non liées les unes aux autres. Cela permet de prendre une décision appropriée pour chaque composant de l'application sans avoir à gérer l'impact des invocations entre les composants. Considérant ces travaux, le déchargement à grain grossier pourrait être redéfini comme suit.

DÉFINITION.– *Dans une définition plus générique, le déchargement à grain grossier est défini par le déchargement individuel des tâches d'un programme. Il en résulte que l'application est soit considérée en totalité ou en quelques composants non liés les uns aux autres.*

Le déchargement à grain grossier de plusieurs composants d'une application permet de garder l'avantage de simplicité de décision de déchargement en même temps que de profiter de la possibilité de prendre une décision individuelle pour chaque composant du programme et d'améliorer ainsi l'efficacité résultante. Cependant, la division de programme en plusieurs composants non liés est compliquée et dépend de la granularité de déchargement, des fonctionnalités des composants extraits à partir de cette granularité et des caractéristiques et liaisons de ces composants. Pour ce faire, l'agrégation de plusieurs composants liés pour les représenter par un seul composant combiné serait parfois nécessaire.

Une autre possibilité est d'extraire les composants d'une application normalement, en se basant sur la granularité choisie, mais de représenter la liaison entre deux composants *via* l'intermédiaire du son programme principale (*main*), qui est évidemment implémenté dans l'appareil mobile. Autrement dit, chaque composant prend ses paramètres d'entrées à partir de l'appareil mobile et retourne ses sorties à l'appareil mobile, qui jouera après le rôle d'intermédiaire pour invoquer le composant relié.

Le traitement consiste à convertir le graphe de relations pondéré d'une application en un graphe étoile[10] bidirectionnel. Pour ce faire, le *main* du programme jouera le rôle du sommet central du graphe étoile et les composants de l'application joueront le rôle des sommets environnants, connectés chacun au sommet central par une arête entrante et une arête sortante. L'arrête entrante à un composant désigne la ou les invocation(s) entrée(s) d'un ou plusieurs autre(s) composant(s) pour laquelle/lesquelles

10. *Star Graph*, en anglais.

le composant *main* a joué le rôle d'intermédiaire. Elle est pondérée par la somme des poids des composants invoquant ce composant.

De manière similaire, l'arrête sortante d'un composant représente les résultats à la sortie de ce composant, qui est transmis au *main* pour être retransmis par la suite au(x) composant(s) lié(s). Le poids de l'arrête sortante est la somme des poids des arrêtes de WRG recevant une entrée à partir de ce composant.

RAPPEL. Un graphe étoile est un graphe connexe dont tous les sommets, sauf un, sont de degré 1. On peut aussi le voir comme un arbre avec un nœud et k feuilles, du moins lorsque k > 1.

La figure 6.3 illustre le graphe étoile pondéré correspondant au graphe de relation de figure 6.1.

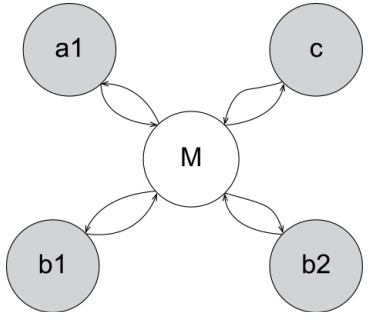


Figure 6.3. *Le graphe étoile*

6.3. Architecture du MCC

Afin de clarifier l'écosystème de décision du déchargement, nous décrivons dans cette section l'architecture générique de MCC. De plus, nous présentons l'architecture du MCC fondée sur les réseaux C-RAN.

6.3.1. *Architecture générique du MCC*

Comme illustré dans la figure 6.4, l'architecture générique de MCC est composée de cinq éléments de base, à savoir l'appareil mobile, le réseau d'accès, le réseau *backhaul* et *backbone*, le Cloud et le *middleware* de décision du déchargement. Dans ce qui suit, chacun de ces éléments de base est décrit en détail.

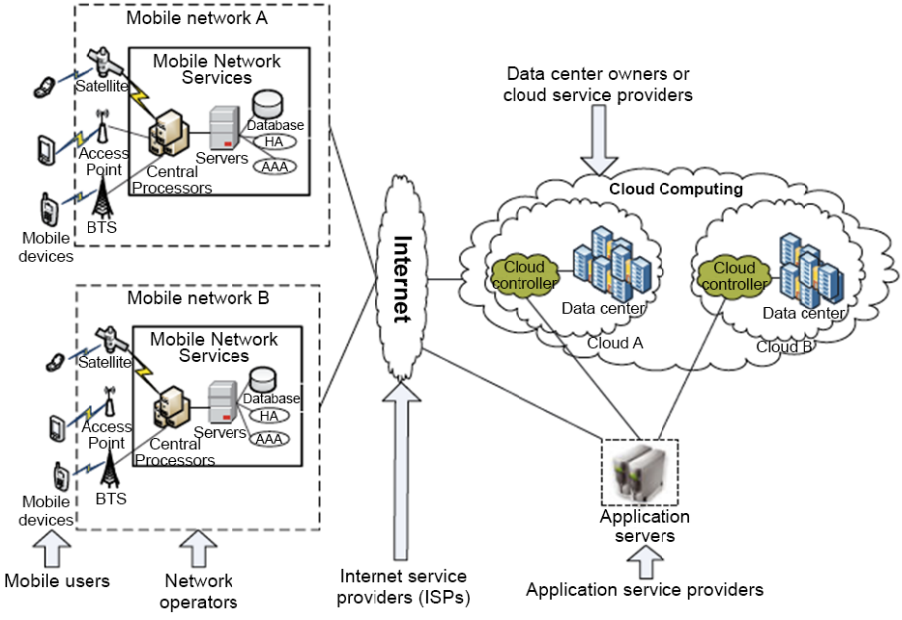


Figure 6.4. *Architecture générique du MCC (Gupta et Gupta 2012)*

6.3.1.1. *Appareil mobile*

L'appareil mobile consiste en un appareil portable sans-fil qui a un calcul intensif à effectuer. Les téléphones mobiles intelligents[11], les ordinateurs portables[12], les assistants électroniques de poche[13], les appareils mettables[14], les capteurs[15], les systèmes embarqués[16] (comme les lecteurs de radio-identification (RFID)[17] et les lecteurs biométriques[18]) sont des exemples d'appareils mobiles qui pourront avoir des calculs intensifs à effectuer. Nous soulignons ici que les appareils mobiles sont, de nature, caractérisés par la limitation de la durée de vie de leur batterie ainsi que par leur puissance de calcul restreinte.

11. *Smartphone*, en anglais.
12. *Laptop*, en anglais.
13. *Personal Digital Assistant* (PDA), en anglais.
14. *Wearable Devices*, en anglais.
15. *Sensor*, en anglais.
16. *Embedded Systems*, en anglais.
17. *RFID Reader*, en anglais.
18. *Biometric Readers*, en anglais.

6.3.1.2. *Réseau d'accès*

Le réseau d'accès radio entre en jeu pour transmettre les données d'entrées des composants de l'application que l'on a décidé d'externaliser. Il est également utilisé pour retourner le résultat final des calculs effectués dans le Cloud. La transmission des données d'entrée vers le Cloud utilise la liaison montante, tandis que la transmission du résultat final de traitement du composant vers l'appareil mobile utilise la liaison descendante du réseau d'accès. Les caractéristiques du lien radio en termes de débit, d'interférence et de stabilité du canal influent considérablement sur le temps du traitement et sur l'énergie de transmission. Cependant, il est difficile de déterminer ses caractéristiques à l'avance et pour toute la durée de la transmission, vu qu'ils sont fortement variables dans le temps et dans l'espace. Cela est dû à la dépendance de ces paramètres et à d'autres facteurs, tels que le positionnement de l'utilisateur mobile par rapport au point d'accès, les sources d'interférence environnantes, le nombre de clients dans la cellule ainsi que la puissance de transmission des utilisateurs, etc.

6.3.1.3. *Réseau* backhaul *et* backbone

Le réseau *backhaul* est le réseau qui connecte la station de base ou le point d'accès au réseau cœur[19] (ou le réseau *backbone*), sur lequel on pourrait accéder au Cloud. Le réseau *backhaul* pourrait être filaire ou sans-fil alors que le réseau cœur est généralement filaire et basé sur Internet. Vu que les données d'entrée et de sortie, après avoir atteint le point d'accès, devront être transportées aussi sur le réseau *backhaul* et *backbone*, la qualité de service offerte par ces deux réseaux est parmi les paramètres importants influant sur l'efficacité de déchargement.

6.3.1.4. Cloud

Le Cloud est un centre de calcul et de stockage puissant situé à distance, accessible par l'intermédiaire d'Internet et exploitable pour stocker des données ou traiter des calculs. Bien que l'utilisation du Cloud des grands fournisseurs situé sur Internet, appelé ici le « Cloud distant », est avantageuse du point de vue de la puissance de calcul et de l'espace de stockage gigantesque offerts, le transport des données sur les liens de *backhaul* et *backbone* introduit un temps de transmission supplémentaire augmentant le temps total d'accomplissement d'une tâche à distance.

Afin d'alléger ce problème, le concept d'*Edge Cloud* a émergé, qui consiste à amener les ressources de traitement et de stockage à proximité de l'utilisateur mobile (Magurawalage *et al*. 2015 ; Mach et Becvar 2017). Bien que l'*Edge Cloud* pourrait ne pas être aussi riche en termes de ressources par rapport au Cloud distant, il permet de réduire le délai de déchargement et d'apporter plus d'agilité aux

19. *Core Network*, en anglais.

traitements en augmentant la bande passante de bout-en-bout[20] et en mettant en commun des ressources locales (Zhang *et al*. 2016 ; Jiang *et al*. 2019).

Afin de profiter des ressources extérieures lorsque la connexion au point d'accès est inappropriée ou inaccessible, une forme mobile de l'*Edge Cloud*, composée des appareils mobiles à proximité, est aussi envisageable (Zhang *et al*. 2015). Ce Cloud mobile, appelé aussi « ad hoc Cloud » ou « D2D[21] Cloud », est avantageux non seulement en raison de son indépendance face à la disponibilité et à la qualité de la connexion sans-fil au point d'accès, mais aussi en raison du débit important de transmission accessible entre des appareils de proximité ainsi que par l'évitement du coût relatif à l'utilisation des ressources radio. Bien évidemment, les ressources mises à disposition par le *D2D Cloud* sont bien inférieures par rapport aux ressources offertes par le Cloud distant ou l'*Edge Cloud.*

6.3.1.5. Middleware *de la décision du déchargement*

Le *middleware* de la décision du déchargement est un élément logiciel de l'architecture du MCC chargé de prendre la décision de déchargement. Du point de vue de la position physique, cet élément pourrait être situé sur l'appareil mobile, être lui-même externalisé, situé sur le point d'accès, sur le Cloud ou sur un point tiers dédié pour prendre la décision de déchargement.

REMARQUE. Le *middleware* de décision peut être implémenté à l'intérieur ou à l'extérieur de l'appareil mobile.

6.3.2. *Architecture à base de C-RAN*

Dans cette approche, le réseau radio classique est remplacé par le C-RAN, permettant entre autres d'enrichir la décision de déchargement par des connaissances accessibles *via* le Cloud du réseau radio.

RAPPEL. L'architecture du réseau C-RAN est une nouvelle architecture du réseau dans laquelle la bande de base[22] et le traitement du canal sont externalisés vers l'unité de bande de base centrale[23] (BBU) située dans le Cloud. Par conséquent,

20. *End-to-End*, en anglais.
21. *Device-to-Device*, en anglais.
22. *Baseband*, en anglais.
23. *Centralized Base Band Unit*, en anglais.

l'antenne sans-fil, appelée RRH[24], joue seulement le rôle d'un relai qui compresse et relie les signaux reçus de l'équipement de l'utilisateur au pool de BBU à travers des liens sans-fil du *fronthaul*.

Étant donnée le Cloud dédié pour les communications radio, la décision de déchargement pourrait aussi être externalisée pour être exécutée dans le C-RAN. Cela permettrait d'exécuter la décision par les serveurs de proximité, qui sont riches en termes de ressources. De plus, la décision pourrait tirer profit des informations collectées par le C-RAN, concernant le réseau d'accès de l'utilisateur tout au long de son mouvement d'un point à l'autre (caractéristiques des réseaux visités, etc.). Cette connaissance du contexte permettrait de prendre une décision plus réaliste et d'optimiser ainsi le temps d'exécution et l'énergie consommée pour exécuter l'application.

6.4. Décision de déchargement

Dans cette section, nous présentons d'abord les différents modèles de placement de *middleware* de la décision du déchargement. Par la suite, nous décrivons les variables de décision et la modélisation du problème de déchargement.

6.4.1. *Placement de* middleware *de décision de déchargement*

Dans cette section, nous décrivons les différentes architectures de MCC du point de vue de l'endroit où la décision de déchargement s'exécute. Le placement du *middleware* de décision est important, étant donné son impact sur la complexité admise lors du développement des algorithmes de décision de déchargement.

6.4.1.1. *Le* middleware *intégré dans l'appareil mobile*

Dans cette architecture, qui est représentée similaire de celle illustrée dans la figure 6.4, le *middleware* de décision du déchargement est intégré dans l'appareil mobile. L'avantage principal de cette approche est lié à l'indépendance totale de déchargement par rapport à un tiers quelconque tant que le réseau d'accès et le Cloud sont accessibles. Cependant, cette approche architecturale nécessite l'utilisation d'algorithmes de décision légers exécutables en temps acceptable sur les appareils mobiles ayant une puissance de calcul limitée.

6.4.1.2. *Le* middleware *externalisé*

Dans cette deuxième approche architecturale, le *middleware* de décision du déchargement est implémenté en dehors de l'appareil mobile, plus souvent dans l'*Edge*

24. *Remote Radio Head*, en anglais.

Cloud ou dans un serveur près de l'utilisateur mobile. Dans l'architecture à base de C-RAN, le *middleware* de décision pourrait aussi être implémenté dans le *BBU pool* (Cai *et al*. 2016). Les paramètres importants de décision tels que les caractéristiques du réseau radio et du Cloud devront donc être transmis au *middleware* de décision.

Nous pourrons différencier deux modèles de l'architecture à base de décision externalisée. Le premier consiste en un modèle dans lequel seule la décision de déchargement est confiée au *middleware* distant ; toutefois, cette décision est retournée à l'appareil mobile pour être exécutée sur l'application. Les composants de l'application choisis pour être déchargés sont alors transmis au Cloud à partir de l'appareil mobile. Dans le deuxième modèle, non seulement le graphe de relation pondéré de l'application mais aussi le code du programme sont transmis au *middlleware* sous la forme d'une machine virtuelle[25] (VM) ou *container*, afin de permettre de lancer la décision prise à partir du *middleware* de décision.

REMARQUE. Dans l'architecture à base de décision externalisée, les composants de l'application pourront aussi être transmis au *middleware* de décision. Celui-ci permet d'appliquer la décision à partir du *middleware* sans avoir besoin de retourner la décision à l'appareil mobile avant de pouvoir commencer l'exécution de l'application selon la décision prise.

Peu importe le modèle d'application de la décision du déchargement : l'avantage principal de cette architecture réside dans la puissance de calcul du *middleware* extérieur comparée à celle de l'architecture de l'appareil mobile. Cette propriété permet d'utiliser des algorithmes de décision plus sophistiqués sans perte de l'agilité de la décision. Cependant, l'externalisation de la décision dans un serveur distant introduit un délai supplémentaire pour la transmission des paramètres de décision vers le *middleware*.

6.4.2. *Formulation générale*

La formulation du problème de déchargement dépend du contexte unique ou multiple du Cloud et du réseau d'accès sous lequel le déchargement est effectué. Dans un environnement à Clouds multiples[26], plusieurs Clouds de type *Cloud distant*, *Edge Cloud*, *Cloud D2D* ou une combinaison de ces types de Cloud, sont exploitables. En outre, dans le contexte à réseau d'accès multiples[27], plusieurs réseaux d'accès de type macro ou micro sont accessibles. Basé sur ces terminologies,

25. *Virtual Machine*, en anglais.
26. *Multi-site Offloading*, en anglais.
27. *Multi-access Network*, en anglais.

nous pouvons distinguer entre le contexte à réseaux d'accès et Cloud unique[28], à réseau d'accès unique et Clouds multiples[29], à réseaux d'accès multiples et Cloud unique[30] et à réseau d'accès et Cloud multiples[31].

Dans la suite de cette section, nous considérons le cas plus général d'un contexte à réseau d'accès et Clouds multiples et proposons une formulation générique applicable à l'ensemble des quatre environnements du déchargement.

6.4.2.1. *Variables de décision du problème de déchargement*

Considérant le contexte à réseau d'accès et Cloud multiples, les variables de décision du problème de déchargement sont définies en considérant les problèmes auxquels la décision tente de trouver une solution. Les problèmes sont les suivants :

– pour chaque composant de l'application, déterminer s'il devrait être exécuté localement ou à distance, afin d'aboutir à un résultat optimal (le « comment exécuter »). Soit $C=\{c_1,c_2,...,c_N\}$ (l'ensemble de composants de l'application où $N=|C|$), la solution à ce problème est décrite par un ensemble $P=\{p_1,p_2,...,p_N\}$ de taille N, où chaque élément p_i est 0 (respectivement 1) si le composant correspondant c_i est décidé d'être exécuté localement (respectivement d'être déchargé) ($p_i \in \{0,1\}$). Il est à noter que la représentation de l'ensemble P est la même pour le déchargement à grain grossier, où $N=|C|=|P|=1$;

– pour chaque composant à décharger, déterminer le Cloud où ce composant devrait être exécuté (le « où exécuter »). Soit $M=\{m_1,m_2,...,m_K\}$ l'ensemble de sites du Cloud accessibles pour le déchargement des calculs, où $K=|M|$, la solution à ce problème est un ensemble $Q=\{q_1,q_2,...,q_N\}$, où chaque élément q_i décrit le Cloud d'exécution choisi pour le composant c_i de l'application ($q_i \in M$). Comme pour le problème précédant, la représentation de la solution Q est la même pour le déchargement à grain grossier, or $|Q|=1$;

– pour les composants à décharger, déterminer le réseau d'accès par lequel les données d'entrée de ces composants devraient être transmises (le « par qui

28. *Single-access Single-site*, en anglais.
29. *Single-access Multi-site*, en anglais.
30. *Multi-access Single-site*, en anglais.
31. *Multi-access Multi-site*, en anglais.

transmettre »). Soit $F = \{f_1, f_2, ..., f_K\}$ l'ensemble des points d'accès accessibles au moment de la transmission des données, la solution à ce problème est une variable $f \in F$ qui décrit le réseau d'accès pour le déchargement des composants choisis pour être exécutés sur le Cloud.

Afin de minimiser le nombre de variables de décision, les ensembles P et Q pourront être fusionnés en un seul ensemble $X = \{x_1, x_2, ..., x_N\}$, où chaque élément x_i représente le site d'exécution de chaque composant de l'application quel qu'il soit sur l'appareil mobile ou sur l'un des Clouds accessibles ($x_i \in \{local, m_1, m_2, ..., m_K\}$). La solution du déchargement S pourrait donc être décrite par un pair (X, f), où chaque variable de décision x_i représente le site d'exécution de i-ième composant de l'application et où la variable f signifie le point d'accès à partir duquel les données de déchargement devront être transmises.

6.4.2.2. *Fonction d'objectif du problème de déchargement*

Indépendamment de la granularité du déchargement, l'objectif de la décision de déchargement est de trouver la solution de déchargement permettant d'optimiser le coût d'exécution de l'application. Le coût est en général exprimé en termes de temps d'exécution de l'application et de la consommation d'énergie dévouée à l'exécution de l'application. Cependant, d'autres critères de sélection, tels que le prix et la sécurité, sont aussi envisageables. La fonction d'objectif générale du déchargement est donc modélisée comme décrit dans l'équation [6.1] :

$$\arg\min_{S} \cos t(S)$$

avec :

- $S = (X, f)$: solution du déchargement détaillée dans la section 6.4.2.1 ;
- $\cos t(S)$: fonction du coût de solution du déchargement S.

RAPPEL. L'argument du minimum, noté « arg min », est l'ensemble des points en lesquels une expression atteint sa valeur minimale. En notation mathématique, pour une fonction f : X → Y, avec Y un ensemble totalement ordonné, arg min est défini par :

$$\arg\min_{x} f(x) := \{x \mid \forall y : f(y) \geq f(x)\}$$

La fonction de coût est définie par la somme pondérée du temps d'exécution de l'application d'un coté et de la consommation d'énergie de l'autre coté, comme décrit dans l'équation [6.2] :

$$\cos t(S) = w_t \times \frac{T(S)}{T(S_{local})} + (1 - w_t) \times \frac{E(S)}{E(S_{local})} \quad [6.2]$$

avec :

– $T(S)$ et $E(S)$: respectivement le temps d'exécution et la consommation d'énergie résultant de la solution S ;

– $T(S_{local})$ et $E(S_{local})$: respectivement le temps et la consommation d'énergie de l'exécution complètement locale de l'application.

Ces deux derniers termes sont utilisés pour normaliser le temps et l'énergie, afin de les rendre sommables dans l'équation 6.2. De l'autre coté, w_t est un coefficient précisant l'importance relative du temps d'exécution de l'application par rapport à la consommation d'énergie dans le calcul du coût total d'une solution du déchargement S. La valeur de ce paramètre pourra être fixée en fonction des préférences de l'utilisateur et des besoins de l'application.

6.4.3. *Modélisation du coût de déchargement*

6.4.3.1. *Coût de déchargement à grain fin*

Afin de modéliser le coût de déchargement à grain fin, nous différencions entre deux modes d'exécution des composants d'un programme : le mode *d'exécution en série* et le mode *d'exécution en parallèle*. Dans le mode d'exécution en série, les composants de l'application sont exécutés les uns après les autres dans un ordre approprié, en considérant le niveau de granularité et leurs données d'entrées. Dans le mode d'exécution parallèle, les composants ayant des données d'entrées ne dépendant pas l'une de l'autre peuvent s'exécuter en parallèle. Dans la suite de ce chapitre, pour des raisons de simplicité, nous nous focalisons sur le mode d'exécution en série.

Basé sur ce principe, le coût du déchargement à grain fin en termes de temps d'exécution est la somme du temps dépensé pour (i) l'exécution des composants de l'application suivant le site d'exécution choisi et pour (ii) la transmission des données d'entrée entre deux composants liés. Le coût du déchargement en termes de

temps d'exécution pourrait donc être modélisé comme indiqué dans l'équation [6.3] :

$$T(S) = \sum_{c=1}^{N} \sum_{m=0}^{K} y_c^m \times t_{c,m}^{exec} + \sum_{c=1}^{N} \sum_{c'=1}^{N} z_{c,c'}^{m,m'} \times t_{c,c'}^{m,m'} \quad [6.3]$$

avec :

– $t_{c,m}^{exec}$: temps d'exécution du composant c dans le site m ;

– $t_{c,c'}^{m,m'}$: temps d'invocation entre deux composants c et c' exécutés dans le site m et m' respectivement.

Les variables binaires y_c^m et $z_{c,c'}^{m,m'}$ sont définies selon les conditions suivantes :

$$y_c^m = \begin{cases} 1, x_c = m \\ 0, x_c \neq m \end{cases} \quad [6.4]$$

$$z_{c,c'}^{m,m'} = \begin{cases} 0, x_c = x_{c'} \\ 1, x_c \neq x_{c'} \end{cases} \quad [6.5]$$

Dans l'équation [6.3], le temps d'exécution du composant c dans le site m est définit par :

$$t_{c,m}^{exec} = \frac{wl_c}{ps_m} \quad [6.6]$$

avec :

– wl_c : charge de travail[32] du composant c ;

– ps_m : vitesse de traitement de CPU offerte par le site m .

Le temps d'invocation entre deux composants c et c' *exécuté* dans le site m et m' est représenté par l'équation [6.7] :

32. *Workload*, en anglais.

$$t_{c,c'}^{m,m'} = \frac{d_{c,c'}}{bw_{m,m'}} \quad [6.7]$$

avec :

– $d_{c,c'}$: quantité de données transmise entre les composants c et c' (représentée par $w_{cc'}$ en WRG) ;

– $bw_{m,m'}$: bande passante du réseau connectant les sites m' et m'.

L'énergie dévouée au déchargement est calculée du point de vue de l'appareil mobile. Par conséquent, l'énergie dévouée pour l'exécution d'un composant dans le Cloud n'est pas prise en compte dans la modélisation du coût énergétique. L'énergie dévouée au déchargement est modélisée par l'équation [6.8] :

$$E(S) = \sum_{c=1}^{N}\sum_{m=0}^{K} y_c^{local} \times e_{c,local}^{exec} + \sum_{c=1}^{N}\sum_{c'=1}^{N} z_{c,c'}^{m,m'} \times e_{c,c'}^{m,m',f} \quad [6.8]$$

avec :

– $e_{c,local}^{exec}$: énergie dévouée à l'exécution du composant c sur l'appareil mobile ;

– $e_{c,c'}^{m,m',f}$: énergie dévouée pour les transmissions inter composants.

Ces deux paramètres sont définis comme suit :

$$e_{c,m}^{exec} = t_{c,local}^{exec} \times p_{local}^{cpu} \quad [6.9]$$

$$e_{c,c'}^{m,m',f} = \begin{cases} t_{c,c'}^{m,m'} \times p_{local}^{transmit}, m = local \\ t_{c,c'}^{m,m'} \times p_{local}^{recv}, m' = local \end{cases} \quad [6.10]$$

avec :

– p_{local}^{cpu} : consommation d'énergie de CPU de l'appareil mobile par l'unité de traitement ;

– $p_{local}^{trqnsmit}$ et p_{local}^{recv} : puissance de transmission et puissance de réception de l'antenne de l'appareil mobile.

6.4.3.2. *Coût de déchargement à grain grossier*

Le temps de déchargement à grain grossier pourrait être modélisé comme un cas spécifique de déchargement à grain fin où il y aurait un seul composant. Celui-ci est représenté ci-dessous :

$$T(S) = \sum_{m=0}^{K} y_c^m \times t_{c,m}^{exec} + z_{main,c}^{local,m} \times t_{main,c}^{local,m} \quad [6.11]$$

De manière similaire, le coût énergétique de déchargement est modélisé comme suit :

$$T(S) = y_c^{local} \times e_{c,local}^{exec} + z_{main,c}^{local,m} \times e_{main,c}^{local,m} \quad [6.12]$$

6.5. Solutions à base d'intelligence artificielle

Le problème du déchargement à grain fin décrit précédemment est un problème NP-difficile[33] (Wang *et al*. 2015 ; Wu *et al*. 2016). La solution à ce problème n'est donc pas trouvable dans un temps de calcul raisonnable (polynomial). Par conséquent, les solutions optimales et leurs variantes à base d'intelligence artificielle sont applicables seulement pour les petits scénarios, avec un nombre limité de variables de décision. Pour les scénarios plus larges, les solutions non optimales à base d'intelligence artificielle, telles que les heuristiques ou les métaheuristiques, sont utilisées. Parmi les métaheuristiques utilisées, nous pouvons citer le recuit simulé[34], la méthode de recherche avec tabous[35], les algorithmes évolutionnaires, les algorithmes basés sur l'éthologie, etc. Dans la suite de cette section, nous présentons quelques algorithmes d'optimisation à base d'intelligence artificielle pour résoudre ce problème.

6.5.1. *Algorithme de séparation et évaluation (B&B)*[36]

L'algorithme de séparation et évaluation est un algorithme optimal de résolution des problèmes d'optimisation combinatoire. Afin d'optimiser les performances de

33. *NP-hard*, en anglais.
34. *Simulated Annealing*, en anglais.
35. *Tabu Search*, en anglais.
36. *Branch-and-Bound* (B&B), en anglais.

cet algorithme en termes de temps de calcul pour trouver la solution optimale, quelques techniques à base d'intelligence artificielle ont été développées.

Dans les solutions de déchargement basées sur cet algorithme, le graphe de relation pondéré de l'application est tout d'abord transformé en un arbre représentant toutes les combinaisons possibles d'exécution locale ou à distance de chaque sommet de WRG. À cette fin, l'arbre est construit à partir d'une racine vide. À cette racine, le premier composant de l'application est rajouté en autant de copies que de sites d'exécution possibles. Chaque nœud de ce niveau représente donc le premier composant étiqueté avec un des sites d'exécution. De manière similaire, à chaque enfant ainsi produit, le deuxième composant de WRG est rajouté en autant de copies que de sites d'exécution possibles. Ce processus est répété jusqu'à ce que tous les sommets de WRG soient ajoutés à l'arbre. Par conséquent, chaque branche de l'arbre ainsi produite représente une solution potentielle à ce problème de déchargement. Le poids de chaque sommet dans WRG est copié sur chaque copie de ce composant dans l'arbre. Cependant, le lien entre un sommet et son parent dans l'arbre prend le poids de lien correspondant dans le graphe de relation pondéré, si le lien entre ces deux composants existe. Sinon, il prend le poids zéro. La figure 6.5 illustre l'exemple de l'arbre correspondant à un WRG ayant trois sommets dans un contexte avec trois sites d'exécution possibles (le site local, le Cloud m1 et le Cloud m2).

Dans la phase d'évaluation, chaque branche est explorée en utilisant la stratégie de *parcours en profondeur*[37]. Quand une branche est ainsi explorée, son coût est progressivement calculé en se basant sur la fonction de coût définie en équation 6.2. Afin d'accélérer l'exploration de l'arbre, une branche pourrait être séparée (coupée) à partir du moment où son coût partiel dépasse la *valeur borne*[38]. Dans notre problème de décision de déchargement, la borne est initiée au minimum du coût de déchargement local ou à distance de tous les composants dans un des sites de Cloud. Lorsqu'une branche est complètement explorée sans être coupée, la valeur de borne est remplacée par le coût de la solution donnée par cette branche.

Afin d'accélérer la résolution du problème de déchargement, certaines solutions existantes à base d'algorithme B&B proposent des optimisations, en vue de permettre de couper les branches inappropriées plus tôt dans le temps. Par exemple, Goudarzi *et al.* (2016) proposent de mettre plus haut dans l'arbre les sommets avec un poids plus important, permettant de couper plus tôt dans le temps les branches inappropriées. D'un autre côté, il est possible d'initier la valeur de la borne au coût de la solution trouvée par une heuristique ou une métaheuristique.

37. *Depth-First Search* (DFS), en anglais.

38. *Bounding Value*, en anglais.

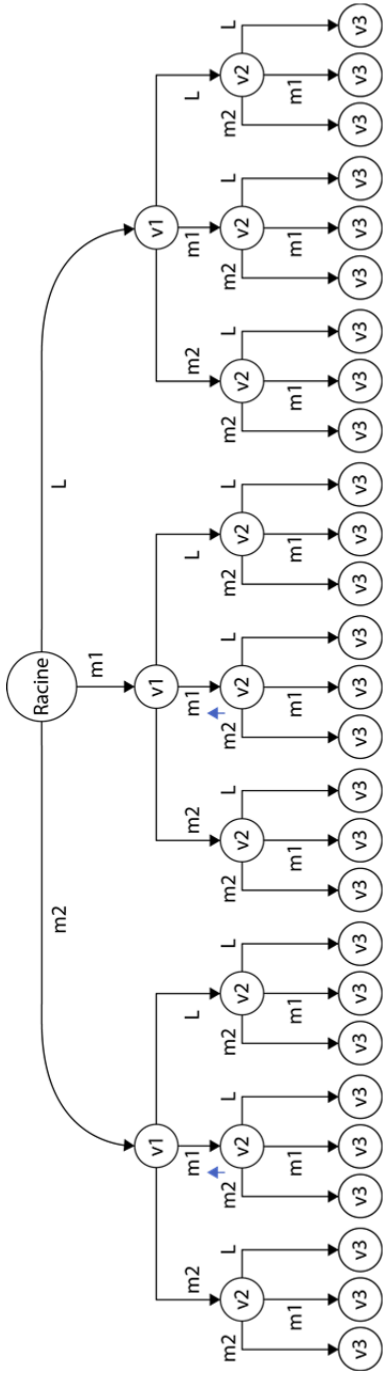


Figure 6.5. *L'arbre B&B*

6.5.2. *Algorithmes métaheuristiques bio-inspirés*

Afin de résoudre le problème de décision de déchargement en un temps de calcul raisonnable, certains travaux se basent sur les algorithmes métaheuristiques évolutionnaires tels que l'algorithme génétique. Dans la suite, nous décrivons l'idée générale des travaux se basant sur ce type de solution.

Tout d'abord, la *population* initiale est créée à partir d'un certains nombres de solutions potentielles. Chaque solution est décrite par un chromosome composé d'un ensemble de gènes. Chaque gène représente un composant de l'application étiqueté par un site d'exécution. La population initiale est générée en général par une approche aléatoire. Cependant, dans certaines optimisations proposées, la population initiale contient des chromosomes avec tous les gènes labélisés pour être exécutés dans un même site. La figure 6.6 illustre l'exemple d'un chromosome.

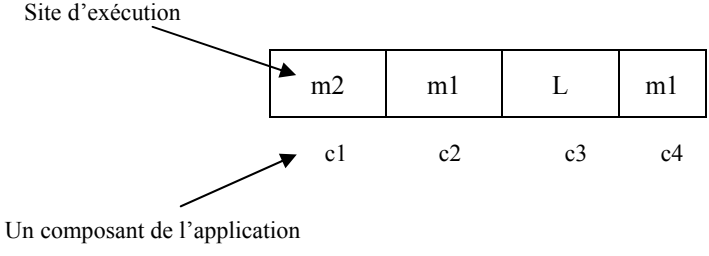


Figure 6.6. *Un chromosome*

À partir d'une population donnée, l'algorithme tente de produire une nouvelle génération en utilisant les processus de sélection et de reproduction. Ces deux processus permettront d'optimiser la population dans chaque itération. Le processus de sélection est en général basé sur la fonction d'adaptation[39], qui mesure le degré de pertinence d'une solution à l'objectif visé. Pour notre problème de décision de déchargement, la fonction d'adaptation correspond à la fonction de coût donnée en équation [6.2] que l'on vise à minimiser. Par conséquent, l'étape de sélection est basée sur une probabilité inversement proportionnelle au coût calculé pour chaque chromosome à partir d'équation [6.2].

Après la phase de sélection, les fonctions de mutation et de croissement sont ensuite appliquées sur les chromosomes sélectionnés pour reproduire une nouvelle génération. Dans la plupart des travaux basés sur les algorithmes génétiques, le

39. *Fitness Function*, en anglais.

croissement des parents est effectué en se basant sur des méthodes telles que coupure et échange. D'un autre côté, les gènes sont choisis pour être mutés en utilisant la fonction *roulette wheel.* Cette dernière consiste à attribuer un coût (une *fitness*) à chaque gène d'un chromosome. La *fitness* d'un gène est calculée à partir du coût du gène divisé par le coût de son chromosome. Par conséquent, le gène qui a le coût le plus élevé a le plus de probabilité d'être choisi pour être muté. La mutation du gène sélectionné se fait par le changement aléatoire de son site attribué.

Ces étapes de sélection, mutation et croissement continuent jusqu'à ce que (i) le nombre maximal d'itération soit atteint ou (ii) qu'aucune amélioration significative ne soit observée lors du passage d'une génération à l'autre.

6.5.3. *Algorithmes métaheuristiques à base d'éthologie*

Certains travaux se focalisent sur une décision de déchargement basée sur les métaheuristiques à base d'éthologie, tels que l'optimisation par colonies de fourmis (ACO)[40], l'optimisation par essaims particulaires (PSO)[41], l'optimisation par colonies d'abeilles (ABC)[42], etc. Dans la suite de cette section, nous nous focalisons sur la décision de déchargement à base d'ACO, comme l'exemple typique des métaheuristiques à base d'éthologie.

L'algorithme ACO est un algorithme métaheuristique inspiré par le comportement collectif des fourmis pour trouver le meilleur chemin entre leur nid et une source de nourriture. Dans la nature, une fourmi commence par chercher de la nourriture d'une façon aléatoire. Une fois qu'une fourmi trouve une nourriture, elle revient au nid en laissant une substance chimique appelée *phéromone*. Grâce à cette trace, d'autres fourmis peuvent trouver la nourriture. Au fil du temps, la traînée de phéromone commence à s'évaporer. Plus une fourmi met longtemps à parcourir le chemin, plus les phéromones doivent s'évaporer. En comparaison, un chemin court est parcouru plus souvent par les fourmis et la densité de phéromones devient plus élevée sur les chemins les plus courts que sur les plus longs.

En se basant sur ce principe, certains travaux représentent chaque solution potentielle du problème de déchargement par le chemin parcouru par une fourmi. Initialement, un nombre de fourmis est créé aléatoirement. À chaque itération d'algorithme, chacune des fourmis met à jour son chemin en fonction de la pertinence de site d'exécution choisi pour chaque composant dans son chemin précédent. Pour notre problème, cette dernière

40. *Ant-colony Optimization*, en anglais.
41. *Particle Swarm Optimization*, en anglais.
42. *Artificial Bee Colony Optimization*, en anglais.

est définie par une fonction d'adaptation basée sur l'équation [6.2]. Ces étapes continueront jusqu'à ce qu'un nombre maximal d'itération soit atteint.

6.6. Conclusion

Dans ce chapitre, nous avons présenté le nouveau paradigme de *Mobile Cloud Computing* qui permet d'étendre les capacités des appareils mobiles en utilisant les ressources abondantes du Cloud. Dans ce contexte, nous avons défini la notion de déchargement de calcul, qui consiste à externaliser les calculs compliqués afin qu'ils soient effectués dans le Cloud. Nous avons introduit deux types de déchargement, à grain fin et à grain grossier, selon le déchargement d'application en entier ou en sous-parties. Nous avons également souligné la nécessité d'une décision de déchargement, étant donné les conditions variables du canal radio et du site Cloud. Le mécanisme de la décision de déchargement dans ce contexte devra déterminer vers quel Cloud et par quel réseau d'accès il est préférable de décharger les calculs. Cet environnement multiCloud multiaccès conduit à un problème de décision NP-difficile, auquel il n'est pas possible de trouver une solution en temps polynomial. Pour résoudre ce problème, les solutions à base d'intelligence artificielle, y compris les algorithmes heuristiques et métaheuristiques, sont proposées dans la littérature. Pour éclaircir les approches prises pour résoudre ce problème, nous avons présenté les différentes classes d'algorithmes que nous avons appliquées.

Bien que le problème général de déchargement décrit dans ce chapitre soit bien étudié dans la littérature, certains problèmes liés au contexte dynamique et mobile, multiCloud*s* et multiutilisateurs ne sont pas encore traités profondément. En particulier, le contexte dynamique et mobile nécessite la proposition d'algorithmes de décision très légers, capables d'être exécutés d'une façon réactive suite aux changements dans le contexte. Certains travaux visant cet axe se basent sur la prédiction de la mobilité de l'utilisateur et traitent aussi la question de déchargement décalé. D'un autre côté, la question de la détermination du prix des ressources radio et du Cloud est un domaine très intéressant pour lequel les solutions à base de théories des jeux[43] et de théories des enchères[44] devraient être proposées. Une autre orientation de recherche consiste à proposer des algorithmes d'allocation de ressources de l'*Edge Cloud* et du réseau d'accès en considérant les besoins de tous les utilisateurs. L'allocation jointe des ressources radio et du Cloud permettra d'optimiser considérablement à la fois la qualité de service, le taux d'admission des utilisateurs et l'utilisation des ressources.

43. *Game Theory*, en anglais.

44. *Auction Theory*, en anglais.

6.7. Bibliographie

Abolfazli, S., Sanaei, Z., Ahmed, E., Gani, A., Buyya, R. (2014). Cloud-Based Augmentation for Mobile Devices: Motivation, Taxonomies, and Open Challenges. *IEEE Communications Surveys & Tutorials*, 16(1), 337–368.

Cai, Y., Yu, F.R., Bu, S. (2016). Dynamic operations of cloud radio access networks (c-ran) for mobile cloud computing systems. *IEEE Transactions on Vehicular Technology*, 65(3), 1536–1548.

Chen, M., Hao, Y., Li, Y., Lai, C.F., Wu, D. (2015). On the computation offloading at ad hoc cloudlet: architecture and service modes. *IEEE Communications Magazine*, 53(6), 18–24.

Goudarzi, M., Movahedi, Z., Pujolle, G. (2016). A Priority-based Fast Optimal Computation Offloading Planner for Mobile Cloud Computing. *International Journal of Information & Communication Technology Research*, 8(1), 43–49.

Gupta, P., Gupta, S. (2012). Mobile Cloud Computing: The Future of Cloud. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 1(3), 134–145.

Hong, S., Kim, H. (2019). QoE-Aware Computation Offloading to Capture Energy-Latency-Pricing Tradeoff in Mobile Clouds. *IEEE Transactions on Mobile Computing*, 18(9), 2174–2189.

Jiang, C., Cheng, X., Gao, H., Zhou, X., Wan, J. (2019). Toward Computation Offloading in Edge Computing: A Survey. *IEEE Access*, 7(1), 131543–131558.

Khan, A.U.R., Othman, M., Madani, S.A., Khan, S.U. (2014). A Survey of Mobile Cloud Computing Application Models. *IEEE Communications Surveys & Tutorials*, 16(1), 393–413.

Khan, A.U.R., Othman, M., Xia, F., Khan, A.N. (2015). Context-Aware Mobile Cloud Computing and Its Challenges. *IEEE Cloud Computing*, 2(3), 42–49.

Li, B., Liu, Z., Pei, Y., Wu, H. (2014). Mobility prediction based opportunistic computational offloading for mobile device cloud. Dans *IEEE 17th International Conference on Computational Science and Engineering (CSE)*. IEEE, Chengdu, 786–792.

Mach, P., Becvar, Z. (2017). Mobile Edge Computing: A Survey on Architecture and Computation Offloading. *IEEE Communications Surveys & Tutorials*, 19(3), 1628–1656.

Magurawalage, C.S., Yang, K., Wang, K. (2015). Aqua computing: Coupling computing and communications [En ligne]. Disponible à l'adresse : https://arxiv.org/abs/1510.07250.

Movahedi, Z. (2018). Green, trust and computation offloading perspectives to optimize network management and mobile services. Thèse de doctorat, Sorbonne Universités, Paris.

Wang, X., Wang, J., Wang, X., Chen, X. (2015). Energy and delay tradeoff for application offloading in mobile cloud computing. *IEEE Systems Journal*, 11(2), 858–867.

Wu, H. (2018). Multi-objective decision-making for mobile cloud offloading: A survey. *IEEE Access*, 6(1), 3962–3976.

Wu, H., Knottenbelt, Z., Wolter, K., Sun, Y. (2016). An optimal offloading partitioning algorithm in mobile cloud computing. Dans *Quantitative Evaluation of Systems*, Agha, G., Van Houdt, B. (dir.). Springer, Berlin, 311–328.

Zhang, Y., Niyato, D., Wang, P. (2015). Offloading in Mobile Cloudlet Systems with Intermittent Connectivity. *IEEE Transactions on Mobile Computing*, 14(12), 2516–2529.

Zhang, K., Mao, Y., Leng, S., Zhao, Q., Li, L., Peng, X., Pan, L., Maharjan, S., Zhang, Y. (2016). Energy-efficient offloading for mobile edge computing in 5g heterogeneous networks. *IEEE Access*, 4(1), 5896–5907.

PARTIE 4

L'IA et les nouvelles architectures de communication

7

Gestion intelligente des ressources dans un système *Smart Grid-Cloud* pour une meilleure efficacité énergétique

Mohammed Anis BENBLIDIA[1], Leila MERGHEM-BOULAHIA[1], Moez ESSEGHIR[1] et Bouziane BRIK[2]

[1] *Université de technologie de Troyes, Troyes, France*

[2] *CESI, Rouen, France*

7.1. Introduction

Avec la prolifération de l'Internet des objets (*Internet of Things*, IoT), notre monde d'aujourd'hui devient de plus en plus connecté. D'après une étude menée par Gartner et l'IDATE (Institut de l'audiovisuel et des télécommunications en Europe), le nombre d'objets connectés dans le monde s'élèvera à 50 milliards en 2030[1]. Vu que les objets connectés sollicitent généralement les services du *Cloud Computing* en termes de calcul et de stockage, les *Data Centers* du *Cloud* vont devoir traiter un nombre considérable de requêtes issues des utilisateurs.

Par conséquent, le rôle majeur que jouent les Data Centers du Cloud dans le stockage, le calcul et la gestion des données issues de l'IoT croît de manière

1. Disponible à l'adresse : https://www.juniperresearch.com/press/press-releases.

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

exponentielle avec le temps. L'un des plus gros problèmes du développement des Data Centers du Cloud est leur grande consommation d'énergie. En outre, on estime que le secteur des Data Centers représente 1,4 % de la consommation mondiale d'électricité (Avgerinou *et al.* 2017).

D'un autre côté, le *Smart Grid* – qui est le futur réseau électrique intelligent – permet l'échange bidirectionnel des données et de l'énergie entre le producteur et le consommateur. Grâce à cet échange de données, le gestionnaire du Smart Grid possède plus d'informations sur ses clients et sera capable de leur proposer des services plus personnalisés et répondant davantage à leurs besoins.

Dans ce travail, nous examinons de plus près l'efficacité énergétique des infrastructures de l'information et de la communication dans un système Smart Grid-Cloud. Nous nous intéressons plus particulièrement aux réseaux de communication et aux Data Centers du Cloud. Nous nous sommes focalisés sur ces derniers à cause de leur grande consommation en énergie, qui les positionne comme des éléments importants dans le réseau. Le reste du chapitre est organisé comme suit. Nous donnons d'abord, dans la section 7.2, un aperçu général sur les Smart Grids ainsi que sur leurs interactions avec les Data Centers du Cloud. La section 7.3 met en avant un état de l'art sur les différentes techniques d'efficacité énergétique utilisées dans les Data Centers. La section 7.4 introduit les techniques d'aide à la décision dans un système Smart Grid-Cloud. Et la section 7.5 conclut le chapitre.

7.2. *Smart Grid* et *Data Center* du *Cloud* : concepts fondamentaux et architecture

Presque inchangée depuis près d'un siècle, l'infrastructure du réseau électrique telle que nous la connaissons aujourd'hui a réussi à répondre à nos besoins avec succès. Cependant, cette infrastructure, vieillissant avec le temps, devient de moins en moins efficace, se heurte sans cesse à ses limites et s'efforce en permanence de faire face à nos exigences. De plus, ce réseau électrique produit l'électricité d'une manière centralisée et l'achemine à un grand nombre d'utilisateurs où la communication est unidirectionnelle : des producteurs vers les consommateurs. Les besoins en matière de fiabilité, de facilité de gestion de l'énergie et de production d'énergie renouvelable soulignent la nécessité d'un réseau modernisé et intelligent pour demain. Dans ce contexte, le réseau électrique intelligent ou Smart Grid ouvre la voie au futur réseau électrique, qui vise à créer un système propre, sûr, sécurisé et fiable (Markovic *et al.* 2013).

Afin de gérer le réseau électrique, il est nécessaire de procéder à des changements importants sans en perturber le fonctionnement. Cela se traduit, dans le

cadre du réseau électrique, par la mise en place d'un réseau de capteurs et compteurs intelligents pour acheminer les données de production et de consommation vers le gestionnaire du Smart Grid. Ce dernier devient alors communicant et interactif.

7.2.1. *Architecture réseaux pour les* Smart Grids

Le Smart Grid peut être considéré comme un réseau de nombreux systèmes et sous-systèmes interconnectés intelligemment pour fournir de l'énergie d'une manière rentable et fiable. L'intelligence apportée au Smart Grid est réalisée par l'ajout des Technologies de l'information et de la communication (TIC) au réseau électrique (Gungor *et al.* 2011). Cette infrastructure de communication permet la collecte de données sur la production et le transport ainsi que sur les réseaux de distribution. En termes d'infrastructure de communication, le Smart Grid peut être divisé en trois principales entités : le réseau domestique de communication (HAN pour *Home Area Network*), le réseau voisin de communication (NAN pour *Neighborhood Area Network*) et le réseau de communication à grande distance (WAN pour *Wide Area Network*) (Hossain *et al.* 2012).

7.2.1.1. Home Area Network *(HAN)*

Le HAN est le plus petit sous-système de la chaîne hiérarchique du Smart Grid. Il représente une unité résidentielle unique dotée d'appareils intelligents, d'outils de contrôle de la consommation d'énergie, de solutions de stockage, de panneaux solaires, de petites éoliennes, de véhicules électriques et de compteurs intelligents. Ces composants communiquent entre eux en utilisant des technologies de communication filaire comme les courants porteurs en ligne (CPL) et des technologies sans-fil comme le Wi-Fi et ZigBe. Les technologies sans-fil telles que ZigBee sont en train de devenir un choix populaire, contrairement aux technologies filaires, en raison de leur faible coût d'installation et de leur flexibilité (Yan *et al.* 2013).

7.2.1.2. Neighborhood Area Network (*NAN*)

Le réseau NAN est un ensemble de réseaux HAN, qui correspond à un groupe de maisons éventuellement alimentées par le même transformateur. Le NAN met en œuvre la connexion qui permet d'offrir plusieurs applications Smart Grid telles que le comptage intelligent (*Smart Metering*), la gestion de la charge, l'automatisation de la distribution d'énergie et la gestion des pannes. L'infrastructure avancée de comptage (AMI pour *Advanced Metering Infrastructure*) collecte les données des équipements intelligents d'un réseau NAN et les regroupe avant qu'elles ne soient envoyées au WAN. Ce dernier connecte les réseaux NAN à l'opérateur du réseau électrique.

7.2.1.3. Wide Area Network *(WAN)*

Le WAN relie plusieurs systèmes de distribution et sert de pont entre les réseaux NAN et HAN et le réseau d'opérateur. Comme le montre la figure 7.1, le réseau WAN fournit l'infrastructure de communication nécessaire pour connecter les équipements des clients du Smart Grid à l'opérateur du réseau électrique. Cela peut être réalisé en adoptant plusieurs technologies de communication (Ethernet, réseaux cellulaires, etc.) pour transférer les données issues des réseaux NAN à l'opérateur réseau.

Parmi les principales applications qu'offre le réseau WAN, on peut citer la surveillance, le contrôle et la protection à grande distance. Ces applications représentent des solutions efficaces pour améliorer la planification, l'exploitation et la protection du réseau électrique dans le Smart Grid (Terzija *et al.* 2011). Les applications de surveillance, de contrôle et de protection dans un réseau WAN offrent une résolution de données supérieure et un temps de réponse plus court que les systèmes classiques de contrôle, de supervision et d'acquisition de données (SCADA pour *Supervisory Control And Data Acquisition*) et de gestion de l'énergie (EMS pour *Energy Management System*). Ces derniers fournissent un intervalle de mise à jour des mesures de plusieurs secondes, voire de quelques minutes, alors que les applications de surveillance, de contrôle et de protection sur le réseau WAN fournissent des données haute résolution, à savoir 60 échantillons par seconde (Khan *et al.* 2016).

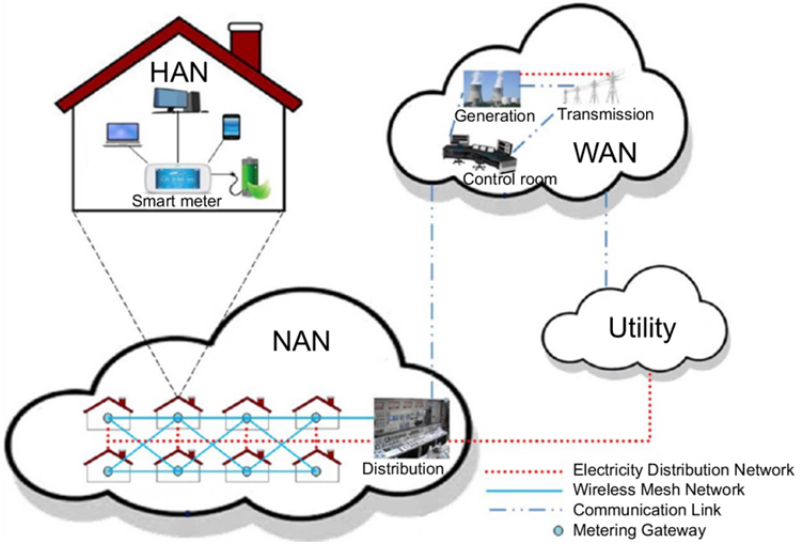


Figure 7.1. *Architecture réseau du Smart Grid*[2]

2. Voir (Bera *et al.* 2015).

7.2.2. *Principales caractéristiques des* Smart Grids

Afin d'approvisionner ses clients en énergie, le Smart Grid gère principalement trois pôles techniques : le pôle génération, le pôle transport et le pôle distribution. Le pôle génération est constitué de centrales traditionnelles produisant de l'électricité. Le pôle transport veille à livrer la quantité d'électricité produite au pôle distribution. Ce dernier se charge de la distribuer vers les clients du Smart Grid. Ces trois principaux composants sont illustrés dans la figure 7.2.

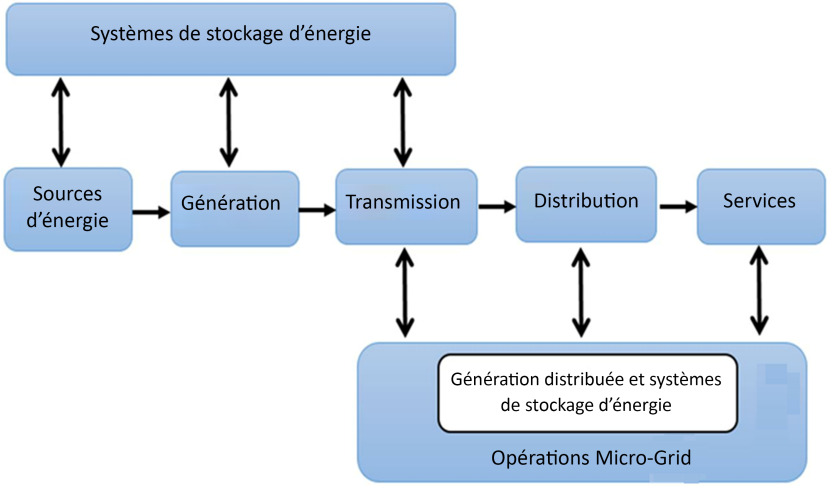


Figure 7.2. *Les systèmes de génération, transport et distribution dans le Smart Grid*

L'un des objectifs majeurs du Smart Grid est le contrôle de la consommation d'électricité chez les clients, en mettant en place différentes méthodes d'optimisation. Pour atteindre cet objectif, le Smart Grid utilise principalement : le comptage intelligent (communication bidirectionnelle) et les Micro-Grids (ressources d'énergie distribuées), les mécanismes de la réponse à la demande (*Demand Response*) et la tarification intelligente (*Smart Pricing*) (Bera *et al.* 2015).

7.2.2.1. *La communication bidirectionnelle*

Afin de créer un réseau de distribution automatisé et largement distribué, le Smart Grid ajoute au réseau électrique traditionnel les avantages des technologies de la communication, notamment, des capteurs, des objets connectés, ainsi que des compteurs intelligents (*Smart Meter* ou SM) pour fournir des informations en temps réel et permettre un équilibre quasi instantané de la gestion de l'offre et de la

demande. À cet égard, il est important de mettre en place une infrastructure de communication fiable permettant d'établir un transport de données en temps réel robuste *via* des réseaux étendus (WAN) jusqu'au niveau du client (Bera *et al.* 2015).

7.2.2.2. *Les ressources d'énergie distribuées*

Le Smart Grid permet d'intégrer des ressources d'énergie distribuées qui comprennent plusieurs technologies : les piles à combustible, les panneaux photovoltaïques, les éoliennes, etc. Le contrôle coordonné du DER ainsi que les charges contrôlables des dispositifs de stockage, comme les volants d'inertie et les batteries, sont au cœur du concept de Micro-Grid (MG).

Le MG est un réseau électrique composé : de sources de consommation énergétiques, de systèmes de production d'électricité (renouvelables ou thermiques), de systèmes de stockage de l'énergie et de systèmes de contrôle et de gestion des données (Guerassimof 2017). Il constitue une solution efficace pour surmonter les problèmes dus aux infrastructures traditionnelles. La capacité du MG à intégrer des générateurs locaux, comme ressources de production d'énergie, et des systèmes de stockage offre de nombreux avantages par rapport aux systèmes existants. Les MG peuvent opérer tout en étant connectés au réseau principal de distribution, où leurs systèmes de stockage permettent une optimisation locale de l'offre et de la demande. De plus, ils peuvent fonctionner en mode îlot d'une manière isolée, en étant déconnectés du réseau pendant une période de temps[3]. Les deux modes de fonctionnement sont gérés par l'opérateur en fonction de la situation technique ou économique du système électrique.

L'un des avantages de l'intégration des MG dans l'architecture du Smart Grid consiste à accroître la fiabilité. Cela se fait en assurant une alimentation de secours en cas de panne du réseau ou lorsque le prix de l'énergie augmente. En outre, les MG permettent de réduire les émissions de CO_2 en alimentant les clients par des sources d'énergie renouvelables (Asmus 2017).

7.2.2.3. *La gestion de l'équilibre du système électrique à travers les mécanismes de la réponse à la demande*

La réponse à la demande (ou DR pour *Demand Response*) est une approche utilisée pour réduire la charge sur le réseau électrique et améliorer la fiabilité du système. La DR est une approche dans laquelle les utilisateurs finaux modifient leurs modèles de consommation d'électricité en fonction des variations des prix. Du point de vue Smart Grid, la réponse à la demande est un moyen efficace de réordonnancer la consommation d'énergie des utilisateurs afin de réduire les dépenses d'exploitation générées par des

3. Disponible à l'adresse : http://www.smartgrids-cre.fr.

générateurs coûteux. Aussi, cette technologie rend le système électrique plus fiable, améliore la transparence et l'efficacité du marché de l'électricité et engendre des avantages financiers mutuels pour l'opérateur Smart Grid et ses clients.

Le mécanisme de la réponse à la demande peut être réalisé en utilisant trois méthodes (Deng *et al.* 2015) :

– réduire la consommation d'énergie de pointe, afin d'empêcher la charge électrique de dépasser la capacité d'alimentation des sous-stations de distribution. La satisfaction des utilisateurs serait réduite, car cette méthode atténue une partie de leur demande ;

– promouvoir la consommation d'énergie en période creuse par le biais de dispositifs de stockage d'énergie, tels que les batteries rechargeables et les véhicules électriques (EV) ;

– équilibrer la consommation d'énergie sur un horizon temporel. Par exemple, transférer une partie de la charge électrique des heures de pointe aux heures creuses.

Le concept de la réponse à la demande a été utilisé dans le secteur commercial et industriel pendant un certain temps pour améliorer la stabilité du réseau. Avec l'émergence de réseaux électriques intelligents, la DR a maintenant le potentiel d'être étendue à grande échelle aux marchés de l'électricité résidentielle (Yan *et al.* 2013).

7.2.2.4. *Tarification intelligente*

La tarification intelligente se présente comme une alternative au prix fixe de l'électricité. Les programmes basés sur les prix sont principalement intégrés aux programmes de la réponse à la demande dans le Smart Grid. Ils offrent aux utilisateurs différents prix de l'électricité à des moments différents. Sur la base de ces informations, les utilisateurs consommeront naturellement moins d'électricité lorsque les prix de l'électricité seront élevés et réduiront donc la demande aux heures de pointe. En d'autres termes, ces programmes incitent indirectement les utilisateurs à modifier de manière dynamique leurs schémas d'utilisation de l'énergie en fonction de la variation des prix de l'électricité, au lieu de contrôler directement leurs charges. On peut trouver différents programmes de tarifications basés sur l'échelle de temps (Deng *et al.* 2015) :

– *Time of Use* (ToU) : représente la forme de tarification dynamique la plus utilisée. Ce mode de tarification décompose la journée en blocs de temps auxquels est associé un prix spécifique. Lorsque les utilisateurs consomment de l'énergie à différents moments de la journée ou à différentes saisons de l'année, ils sont facturés à des prix différents. Afin d'inciter les utilisateurs à déplacer leurs charges durant la journée, le prix de l'électricité annoncé aux périodes de pointes est plus élevé par

rapport à celui fixé aux périodes creuses. Dans ce type de tarification, les prix sont souvent fixés et annoncés à l'avance et ils restent inchangés pendant une longue période (Vardakas *et al.* 2015) ;

– *Critical Peak Pricing* (CPP) : principalement basé sur la tarification ToU, sauf certains jours. La tarification CPP permet d'envoyer un jour à l'avance un signal d'urgence pour prévenir de la période d'extrême pointe (Bergaentzlé et Clastres 2013). Durant cette période de pointe, le prix augmente de manière significative par rapport au tarif habituel, de manière à inciter fortement les utilisateurs à ne pas consommer à ce moment. CPP n'est employé que pendant un nombre limité d'heures ou de jours par an, pour assurer la fiabilité du système et garantir l'équilibre entre l'offre et la demande ;

– *Real Time Pricing* (RTP) : se réfère à la tarification dynamique, où le prix de l'électricité varie généralement à des intervalles de temps différents dans la journée (toutes les 15 minutes ou toutes les heures) (Allcott 2009). RTP est la tarification la plus efficace, car elle permet de transmettre heure par heure au consommateur les variations du prix du kilowattheure sur le marché de gros. Elle donne le coût réel de l'électricité au moment où elle est consommée ;

– *Inclining Block Rate* (IBR) : est une tarification progressive qui vise à réduire la consommation globale de l'électricité. Cette tarification est conçue avec des structures tarifaires à deux niveaux (blocs inférieur et supérieur), de sorte que le prix unitaire de l'électricité, que le consommateur devra payer, augmente avec la quantité d'énergie consommée. En d'autres termes, le prix de l'électricité par consommation d'énergie atteindra une valeur supérieure si la consommation d'énergie horaire, journalière ou mensuelle de l'utilisateur dépasse un certain seuil. IBR incite les utilisateurs à répartir leurs charges entre différents moments de la journée afin d'éviter des taux plus élevés, ce qui contribue à réduire le ratio pic/moyenne du réseau (Gabr *et al.* 2018).

7.2.3. *Interaction des* Data Centers *du* Cloud *avec le* Smart Grid

Victimes de leur développement, le Cloud Computing et le Smart Grid sont confrontés à plusieurs défis. En particulier, les grands fournisseurs de services de Cloud Computing voient leur facture annuelle d'électricité augmenter d'une année à l'autre. À titre d'exemple, cette facture a atteint les 67 millions de dollars fin 2014 (Deng *et al.* 2014), chiffre qui continue d'augmenter avec l'essor des services du Cloud Computing et la hausse du prix de l'électricité. En parallèle, le Smart Grid intègre un grand nombre de ressources d'énergie distribuées, telles que les panneaux solaires et les éoliennes, et doit également garantir une stabilité opérationnelle élevée. Cela peut générer des problèmes économiques, vu la nature intermittente de la production d'énergie décentralisée.

Les préoccupations susmentionnées du Cloud Computing et du Smart Grid peuvent être atténuées par une coopération appropriée entre les deux parties. La réponse à la demande des Data Centers peut être un grand atout pour le Smart Grid, car elle est motivée par les demandes des utilisateurs pouvant être réparties sur des Data Centers géographiquement distribués et desservis par plusieurs sources d'énergie. Ainsi, l'électricité consommée par un Data Center à partir du Smart Grid peut être ajustée de manière flexible en équilibrant la charge de travail, ou en modulant la génération de l'électricité sur le site. Dans ce contexte, l'adoption de moyens de stockage par les Data Centers dans le Smart Grid permet de résoudre plusieurs problèmes et présente de multiples intérêts aux différents acteurs concernés : producteurs, consommateurs et gestionnaires de réseaux. Les systèmes de stockage d'énergie pourront permettre de garantir l'équilibre et la stabilité du réseau tout en améliorant l'efficacité des installations de transmission et de distribution. Ils représentent des solutions ayant un rôle primordial à jouer, aussi bien au niveau économique qu'environnemental, en garantissant une fourniture d'énergie électrique efficace et durable (Guerassimof et Maizi 2013).

L'interaction des Data Centers du Cloud avec le gestionnaire du Smart Grid se fait d'une manière directe, en échangeant les données *via* l'infrastructure de communication du Smart Grid. Mais aussi, d'une manière indirecte *via* les requêtes issues des utilisateurs vers les Data Centers du Cloud. La figure 7.3 illustre l'interaction des Data Centers du Cloud avec le Smart Grid et ses utilisateurs.

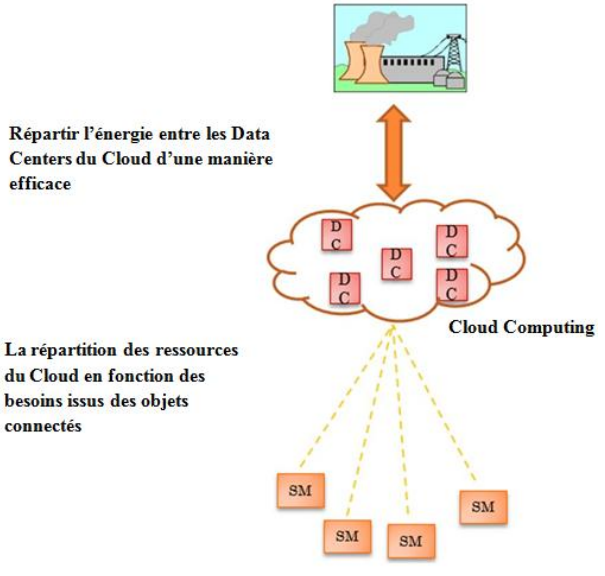


Figure 7.3. *Interaction des Data Centers avec le Smart Grid et ses utilisateurs*

Dans un système Smart Grid-Cloud, les requêtes des utilisateurs sont regroupées par un agrégateur Cloud. Ce dernier se charge de répartir les requêtes vers les Data Centers d'un Cloud suivant une politique prédéfinie. Exécuter ces requêtes tout en garantissant un niveau de qualité de service satisfaisant reste parmi les plus grands challenges à relever. Dans ce contexte, les différentes requêtes issues des objets connectés (IoT) exigent un traitement intensif et une faible latence, en particulier pour les applications temps réel. Il est très difficile pour le Cloud Computing de faire face à cette communication de plus en plus contraignante et de satisfaire toutes les demandes. Le *Fog Computing* est une extension émergente de l'architecture du Cloud Computing, qui permet de traiter les données au plus proche de leur source et met à disposition les connexions réseau nécessaires pour acheminer les données du nœud Fog vers le terminal de l'utilisateur. Chaque nœud Fog a des capacités de traitement qui lui permettent de répondre dans les plus brefs délais aux requêtes des objets qui sont connectés au réseau. Ainsi, en présence de plusieurs nœuds Fog proches des extrémités du terminal, il est possible d'obtenir de meilleures performances en termes de délai qu'avec le Cloud Computing (Chiang et Zhang 2016 ; Chiang *et al.* 2017). De ce point de vue, le *Fog Computing* ne vise pas à remplacer le Cloud Computing, mais à le compléter dans un nouveau paradigme *Fog-Cloud Computing*, qui vise à satisfaire les applications des utilisateurs (Elmroth *et al.* 2017). La figure 7.4 illustre un système *Fog-Cloud Computing*.

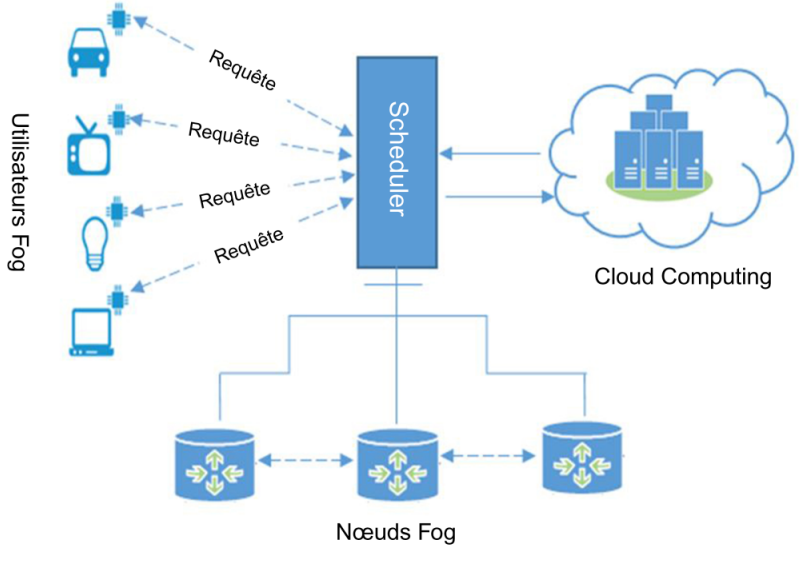


Figure 7.4. *Un système Fog-Cloud Computing*

7.3. État de l'art sur les techniques d'efficacité énergétique des *Data Centers* du *Cloud*

L'amélioration de l'efficacité énergique des Data Centers en termes de coût, de consommation ou bien d'impact sur l'environnement devient un sujet qui attire beaucoup de chercheurs. De nombreux travaux de recherche considèrent le problème de la grosse consommation d'énergie des Data Centers et proposent des solutions mettant en œuvre différentes techniques. Dans cette section, nous regroupons les efforts qui ont été menés à ce sujet en trois sous-sections : les techniques d'efficacité énergétique des installations hors équipements de technologie de l'information (IT pour *Information Technology*), les techniques d'efficacité énergétique des serveurs d'un Data Center et enfin les techniques d'efficacité énergétiques d'un ensemble de Data Centers.

7.3.1. *Techniques d'efficacité énergétique des équipements non-IT d'un* Data Center

Un Data Center regroupe plusieurs installations : des serveurs, des onduleurs, des équipements de refroidissement, une salle d'opération, etc. La majeure partie de l'énergie consommée par un Data Center est utilisée pour refroidir les serveurs et les onduleurs. La règle de base pour la gestion des installations des Data Centers est d'organiser de manière appropriée des îlots d'air chaud et des îlots d'air froid. Un îlot d'air froid se forme de la prise d'air froid des serveurs se faisant face, tandis que l'îlot d'air chaud se forme de la sortie d'air chaud des rangées de serveurs. L'élévation des sols ainsi que l'utilisation d'équipements de refroidissement à des températures plus élevées font partie des techniques permettant d'économiser de l'énergie. Afin de réduire la consommation d'énergie des conteneurs des Data Centers, Endo *et al.* (2013) ont combiné le refroidissement direct à l'air frais au refroidissement par évaporation pour atténuer la chaleur produite par les équipements IT du Data Center. Somani *et al.* (2009) ont présenté un algorithme appelé « gestion de la charge basée sur l'intelligence ambiante », qui améliore la capacité de dissipation de la chaleur du Data Center. L'algorithme utilise la température d'entrée des racks de serveurs et répartit les charges de travail en fonction de l'environnement thermique du Data Center. Pour analyser l'efficacité des techniques mises en œuvre, la dynamique des fluides numérique (CFD pour *Computational Fluid Dynamics*) est utilisée pour modéliser le flux d'air et le transfert de chaleur dans les Data Centers. Joshi et Kumar (2012) ont classifié les modèles CFD des Data Centers en deux catégories : modélisation du flux d'air permettant de prévoir les débits des tuiles perforées et modélisation des effets thermiques de la disposition du rack et de la distribution de l'énergie. Ils ont aussi

présenté des schémas alternatifs d'approvisionnement et de retour et quelques métriques de performances thermiques. Généralement, l'efficacité énergétique des installations d'un Data Center est indépendante des concepts du Smart Grid.

7.3.2. *Techniques d'efficacité énergétique des serveurs d'un* Data Center

Considérant la disponibilité d'un ensemble de Data Centers, cette catégorie améliore la consommation énergétique de chaque Data Center individuellement, par le biais de plusieurs méthodes comme, par exemple, minimiser l'utilisation de leurs équipements informatiques tels que les processeurs. L'efficacité énergétique des équipements informatiques au sein d'un Data Center est mise en œuvre *via* un matériel économe en énergie, ainsi que par la virtualisation et la consolidation dynamique des charges de travail (Cavdar et Alagoz 2012). Bien que les Data Centers contiennent des commutateurs réseaux intra-Data Centers, la consommation totale des commutateurs centraux, d'agrégation et d'accès représente environ la moitié de la consommation électrique des équipements IT (Kliazovich *et al.* 2010). Par conséquent, la plupart des efforts réalisés à ce sujet dans la littérature portent sur l'efficacité énergétique des serveurs d'un Data Center.

Gandhi *et al.* (2009) proposent un algorithme qui permet de trouver une allocation optimale de l'électricité aux serveurs d'un Data Center, afin de minimiser leur temps moyen de réponse. Pour ce faire, ils exploitent deux mécanismes de redimensionnement de la tension et de la fréquence du processeur : le dimensionnement dynamique de la fréquence (DFS pour *Dynamic Frequency Scaling*) et le dimensionnement dynamique de la tension et de la fréquence (DVFS pour *Dynamic Voltage and Frequency Scaling*). Ces derniers comptent parmi les techniques les plus exploitées pour réduire la consommation d'énergie des serveurs. Wang *et al.* (2013) proposent une application qui permet de virtualiser la distribution d'électricité dans un Data Center. Leur proposition permet aux serveurs de définir leurs besoins en énergie et, par conséquent, l'application répartit l'électricité d'une manière équitable. Liu *et al.* (2018) considèrent diverses sources d'énergie et proposent une architecture de gestion d'énergie nommée DiPSN. Cette dernière connecte efficacement différentes sources d'énergie à des serveurs appropriés.

L'allocation de ressources et la migration de machines virtuelles (VM) considérant l'optimisation d'énergie ont été largement prises en compte. Dai *et al.* (2016) ont modélisé le placement des machines virtuelles dans les serveurs d'un Data Center en tant que problème d'optimisation linéaire. Les auteurs ont développé deux algorithmes d'approximation appelés « MinES » et « MinCS », afin d'obtenir la solution optimale

qui minimise la consommation d'énergie et garantit le niveau de service des utilisateurs. Sharma *et al.* (2019) proposent une approche hybride qui combine algorithmes génétiques (AG) et optimisation de l'essaimage de particules (PSO pour *Particale Swarm Optimization*) appelée « HGAPSO ». Cet algorithme permet la migration des machines virtuelles tout en réduisant la consommation d'énergie et en évitant la violation du niveau de service prédéfini. Dans un autre travail, Duong-Ba *et al.* (2018) traitent le problème du placement et de la migration optimale des machines virtuelles pour minimiser l'utilisation des ressources et la consommation d'énergie des Data Centers du Cloud. Ils ont modélisé le problème en tant que fonction multi-objectif et ont résolu le problème en proposant deux algorithmes multi niveaux, qui combinent placement et migration des machines virtuelles.

7.3.3. *Techniques d'efficacité énergétique d'un ensemble de* Data Centers

Nous considérons dans cette section les techniques conçues pour un ensemble de Data Centers géographiquement distribués tels que ceux détenus et exploités par Google, Yahoo, Microsoft, Amazon, eBay, etc. L'idée générale est de diriger les requêtes des utilisateurs vers un ou plusieurs Data Centers afin de minimiser leur coût, leur consommation ou leur émission en CO_2.

Cette catégorie tire parti des fonctionnalités offertes par le Smart Grid pour améliorer l'efficacité énergétique des Data Centers dans le Cloud. En outre, nous examinerons l'adoption des Data Centers pour l'hébergement de services de Smart Grid, afin de brosser un tableau complet des interactions entre le Smart Grid et les Data Centers du Cloud. Nous nous concentrerons sur les techniques d'efficacité énergétique pour plusieurs Data Centers et leurs réseaux de transport correspondants. Ces techniques sont appliquées sous deux architectures : une architecture Smart Grid-Cloud et une architecture Smart Micro-Grid-Cloud.

7.3.3.1. *Architecture* Smart Grid-Cloud

Plusieurs travaux de recherches ont considéré l'interaction des Data Centers du Cloud avec le Smart Grid pour les fonctionnalités qu'offre ce dernier, comme la tarification dynamique et la réponse à la demande (DR). Gu *et al.* (2016) considèrent un système Smart Grid-Cloud où les Data Centers du Cloud sont alimentés par le Smart Grid et par l'énergie renouvelable qu'il autoproduit. Ils proposent un algorithme de répartition de requêtes des utilisateurs qui minimise les coûts énergétiques des Data Centers du Cloud. Ce travail a été étendu dans (Gu *et al.* 2018) en proposant un ordonnanceur vert qui minimise à la fois le coût énergétique et les émissions de carbone.

Dans un autre travail, Wang *et al.* (2016a) ont modélisé l'interaction entre les réseaux électriques et les Data Centers du Cloud sous forme de problème en deux étapes. Ils utilisent d'abord la tarification dynamique pour réaliser l'équilibre de charge puis, dans un second temps, les Data Centers réagissent aux prix de l'énergie et gèrent la répartition des requêtes des utilisateurs pour minimiser leur coût total en énergie. Wang et Ye (2016) proposent un algorithme de réponse à la demande qui considère les énergies renouvelables, afin de minimiser les coûts énergétiques. Leur proposition optimise conjointement l'approvisionnement en énergie et l'allocation de la charge de travail issue des utilisateurs dans les Data Centers du Cloud. Kiani *et al.* (2018) visent à maximiser les profits associés à l'exploitation des énergies renouvelables dans les Data Centers répartis géographiquement. Ils considèrent dans leur système à la fois la production d'énergie renouvelable et la tarification dynamique des marchés de l'électricité. Ding *et al.* (2018) proposent un algorithme stochastique de répartition des ressources, afin d'optimiser les coûts énergétiques et l'émission du CO_2 des Data Centers.

7.3.3.2. *Architecture* Smart Micro-Grid-Cloud

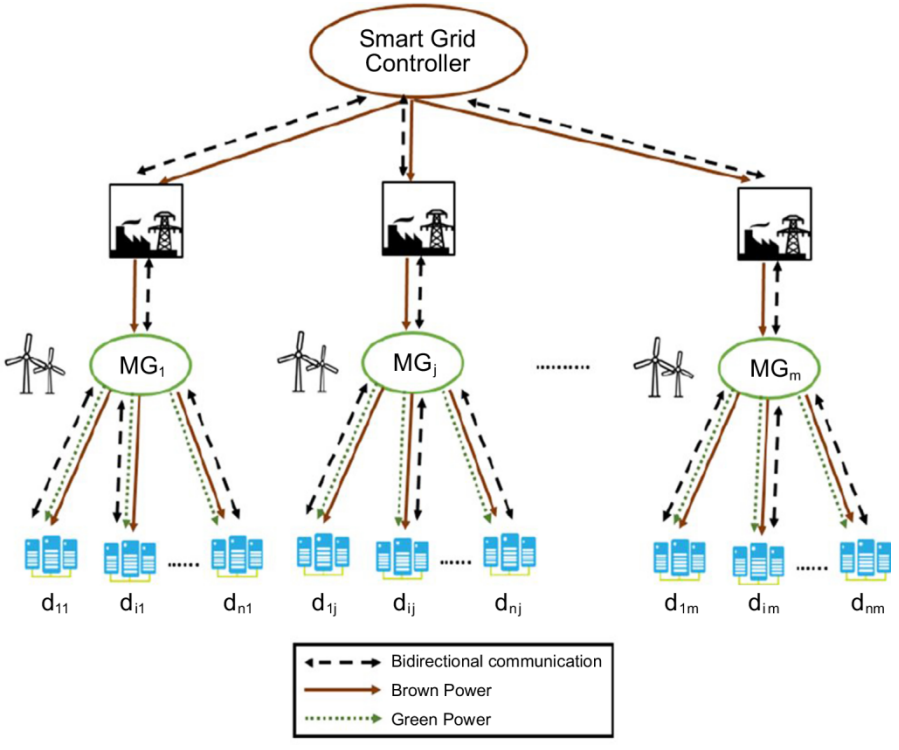


Figure 7.5. *Exemple d'une architecture Smart Micro-Grid-Cloud*

Considérer l'interaction des Data Centers du Cloud avec un opérateur Smart Grid a conduit à des résultats significatifs en termes d'efficacité énergétique. Cependant, certaines études estiment que faire tourner des Data Centers dans des Micro-Grids peut être plus intéressant en termes d'efficacité énergétique, étant donné que les Micro-Grids peuvent améliorer la durabilité et la fiabilité des services électriques. La figure 7.5 illustre un exemple d'une architecture Smart Micro-Grid-Cloud. Yu *et al.* (2016) étudient le problème de la gestion de l'énergie des Data Centers du Cloud alimentés par des Micro-Grids. Ils minimisent le coût énergétique en considérant les paramètres d'efficacité de chargement et de déchargement des batteries. Ils étudient également dans (Yu *et al.* 2014) le problème de minimisation des coûts de l'énergie et des émissions de carbone des Data Centers alimentés par des Micro-Grids. Pour ce faire, ils modélisent un problème d'optimisation stochastique et le résolvent en utilisant la technique d'optimisation de Lyapunov. Yu *et al.* (2015) ont modélisé de la même manière le problème de minimisation du coût énergétique des Data Centers en prenant en considération les pannes de courant.

7.3.4. *Discussion*

Grâce à cette étude de l'état de l'art, on peut constater qu'il existe peu de travaux de recherche qui ont étudié l'efficacité énergétique des équipements IT dans un Data Center du point de vue du Smart Grid. La majorité des travaux réalisés à ce sujet se concentre sur l'efficacité énergétique d'un seul élément IT, sans tirer profit des fonctionnalités qu'apporte le Smart Grid.

Dans ce travail, nous portons notre intérêt sur les techniques prenant en considération un ensemble de Data Centers interagissant avec le Smart Grid. Le tableau 7.1 présente une étude comparative entre les travaux susmentionnés dans cette catégorie. La comparaison est effectuée selon quatre critères : réduction des coûts énergétiques, réduction des émissions de gaz à effet de serre, réduction de la charge du réseau électrique et intégration des énergies renouvelables. Nous constatons que la plupart des solutions proposées dans un système Smart Grid-Cloud sont appliquées du côté des fournisseurs de services Cloud et nous notons que la prise en considération des prix et l'utilisation des énergies renouvelables demeurent deux facteurs principaux de l'efficacité énergétique dans les Data Centers. Cependant, peu de travaux ont examiné le rôle du Smart Grid dans la gestion de l'allocation d'énergie aux Data Centers du Cloud. En outre, les systèmes proposés ne traitent pas l'impact que l'énorme consommation d'énergie des Data Centers peut avoir sur le Smart Grid.

Généralement, les travaux de recherche utilisant des architectures à base de Micro-Grids ont pour objectif de réduire le coût énergétique et l'émission de

carbone des Data Centers du Cloud. Cependant, ce type de travaux ne présente pas une analyse détaillée de la manière dont leurs approches peuvent affecter le réseau électrique, car ils ne considèrent pas le fait qu'un Smart Grid ait une quantité d'énergie limitée à fournir.

Travaux de recherche	Architecture		Métriques considérées			
	Smart Grid-Cloud	*Smart Mico-Grid-Cloud*	Coût d'énergie	Émissions de carbone	Charge électrique sur le réseau	Ressources d'énergie renouvelables
(Gu *et al.* 2016, 2018)		X	X	X		X
(Wang *et al.* 2016)		X	X		X	
(Wang et Ye 2016)		X	X			X
(Kiani et Ansari 2018)		X	X	X		X
(Ding *et al.* 2018)		X	X	X		X
(Yu *et al.* 2016)	X		X			X

Tableau 7.1. *Tableau comparatif des travaux*

7.4. État de l'art sur les techniques d'aide à la décision dans un système *Smart Grid-Cloud*

Les techniques d'efficacité énergétique dans un système Smart Grid-Cloud nécessitent l'interaction des Data Centers avec l'opérateur du Smart Grid. Chacune

de ces deux entités a un comportement vis-à-vis de l'autre. Ces comportements peuvent être modélisés en utilisant les fonctions d'utilité et les fonctions de coûts.

Fonction d'utilité : les comportements des différents Data Centers sont modélisés par différents choix de fonctions d'utilité. Plus formellement, l'utilité représente le niveau de confort/ satisfaction obtenu par le Data Center en fonction de sa consommation d'énergie, non décroissante et concave. Généralement, les fonctions d'utilité les plus considérées sont les fonctions d'utilité quadratiques (Samadi *et al.* 2012).

Fonction de coût : le coût de la production et de la fourniture d'électricité par l'opérateur Smart Grid est modélisé par la fonction de coût, qui est croissante et strictement convexe. Deux alternatives peuvent être adoptées : la fonction de coût linéaire par morceaux (correspondant à IBR pour *Inclining Block Rate*) et la fonction de coût quadratique (Mohsenian-Rad *et al.* 2010).

Les problèmes d'efficacité énergétiques dans un système Smart Grid-Cloud sont souvent formulés en problèmes d'optimisation. Ces derniers peuvent être résolus par plusieurs approches. Dans cette section, nous présentons les principales approches de modélisation et de résolution des problèmes d'efficacité énergétique dans les Data Centers du Cloud.

7.4.1. *Théorie des jeux*

La théorie des jeux est un modèle de processus décisionnels interactifs qui étudie le comportement égoïste ou rationnel des individus. On appelle ces derniers « les joueurs » et ils représentent l'élément le plus important dans un jeu. Un joueur peut représenter un seul individu ou un groupe d'individus prenant une décision. Après avoir défini l'ensemble des joueurs, on peut distinguer deux types de modèles : les jeux dont les éléments de base sont des actions des joueurs individuels ou ceux basés sur les actions jointes d'un groupe de joueurs. Les modèles du premier type sont appelés les « jeux non coopératifs » et ceux de second type, les « jeux coopératifs » (Yildizoglu 2011). Un jeu G se compose de trois éléments fondamentaux : les joueurs N ; les stratégies $\{S_i\}_{i\in N}$ et la fonction de gain $\{P_i\}_{i\in N}$. Chaque joueur choisit une stratégie $s_i \in S_i$ pour maximiser sa fonction de gain $P_i(s_i,s_{-i})$, qui dépend non seulement de sa stratégie s_i, mais aussi des stratégies des autres joueurs s_{-i}.

L'un des plus importants concepts dans la théorie des jeux est *l'équilibre de Nash*. Il représente la stratégie stable et statique dans laquelle aucun joueur n'a envie de dévier unilatéralement, étant donné les stratégies jouées par les autres joueurs.

Les approches basées sur la théorie des jeux pour l'optimisation d'énergie des Data Centers convergent généralement vers un équilibre de Nash ; cela nous conduit à avoir une solution optimale à la fois pour l'offre et la demande. Ainsi, la théorie des jeux représente une approche efficace pour faciliter la prise de décision d'une manière intelligente dans un système Smart Grid-Cloud.

Dans ce contexte, Wang *et al.* (2014a) proposent un jeu Stackelberg à deux niveaux entre le contrôleur du Smart Grid et le contrôleur du Cloud Computing. L'objectif du contrôleur du Smart Grid est d'effectuer l'équilibre de charge dans les bus électriques en se basant sur une politique de tarification définie. Tandis que l'objectif du contrôleur du Cloud est de maximiser le prix total obtenu en répondant aux demandes des utilisateurs. Une extension de ce travail est présentée dans (Wang *et al.* 2014b), où les auteurs considèrent une structure de réseau électrique existante (la structure IEEE 24 bus) et l'intégration des sources d'énergies renouvelables dans le système. Par conséquent, la charge sur chaque bus d'alimentation dépend à la fois de la consommation d'énergie et de la production d'énergie renouvelable. Benblidia *et al.* (2018) modélisent l'interaction des Data Centers du Cloud avec le Smart Grid en un jeu non coopératif pour répartir efficacement l'énergie entre les Data Centers. La fonction de gain du jeu est modélisée en un problème d'optimisation linéaire et résolue *via* les multiplicateurs de Lagrange.

7.4.2. *Optimisation convexe*

Dans l'optimisation convexe, on considère des fonctions objectifs qui sont convexes. Généralement, les méthodes d'efficacité énergétiques dans les Data Centers visent à maximiser la fonction d'utilité ou à minimiser la fonction de coût. Il faut noter que la fonction d'utilité est concave, alors que la fonction de coût est convexe. En outre, le problème de maximiser une fonction concave f peut être reformulé de la même manière lorsque l'on veut minimiser la fonction $-f$, qui est convexe (Deng *et al.* 2015).

Wang *et al.* (2016) utilisent l'optimisation convexe pour modéliser le problème de réduction du coût énergétique des Data Centers dans le Cloud. Ils proposent une approche qui permet à un opérateur Cloud d'optimiser conjointement l'approvisionnement en énergie et l'allocation de la charge de travail des utilisateurs dans des Data Centers distribués. Chen *et al.* (2014) considèrent un Data Center doté de dispositifs de stockage d'énergie. Leurs travaux visent à minimiser la combinaison linéaire ente le coût de l'électricité et le temps moyen de réponse des requêtes des utilisateurs. Des solutions basées sur des techniques d'optimisation convexe sont proposées pour résoudre le problème.

7.4.3. *Processus de décision markovien*

Le processus de décision markovien fait référence à la prise de décision séquentielle basée sur l'observation périodique ou continue des systèmes dynamiques de Markov. Les actions de ces derniers sont aléatoires, mais les probabilités de transition d'état présentent des propriétés de Markov.

Rao *et al.* (2010) modélisent l'incertitude du prix d'énergie et la charge de travail traitée par les Data Centers du Cloud en un processus de décision markovien. Ils visent à réduire le coût de l'électricité tout en garantissant la qualité de service des utilisateurs. Yang *et al.* (2017) utilisent un processus de décision markovien pour modéliser le problème d'approvisionnement de ressources des serveurs d'un Data Center. Ils proposent un algorithme d'autogestion basé sur l'apprentissage automatique. Leur proposition fournit les ressources aux serveurs d'une manière autonome et dynamique pour réduire le coût de l'électricité des Data Centers.

7.4.4. *La logique floue*

La logique floue est un sous-ensemble de la logique conventionnelle booléenne, qui a été étendue pour prendre en charge le concept de vérité partielle : les valeurs de vérité qui sont entre « complètement vrai » et « complètement faux ». La logique floue (Zadeh 1965) est un outil qui traite des informations incertaines, imprécises ou qualitatives dans des systèmes qui ne sont pas définis avec des modèles mathématiques formels. En théorie, l'ensemble flou indique l'appartenance progressive d'un élément à un ensemble.

Soit A un ensemble flou, défini sur un univers X par une fonction d'appartenance $\mu_A: X \rightarrow [0,1]$. μ_A indique le degré d'appartenance de x à A. $\mu_A = 1$ indique l'appartenance complète, $\mu_A = 0$ la non-appartenance absolue et $0 < \mu_A < 1$ l'appartenance partielle (plus μ_A s'approche de 1, plus x appartient à A).

La logique floue couvre un large éventail de conditions de fonctionnement. En outre, les algorithmes à base de logique floue peuvent traiter les incertitudes et sont robustes aux environnements à forte mutation (Zadeh 1965). De plus, le processus d'inférence est simple, comparé à d'autre systèmes (Marrouchi et Ben Saber 2014), ce qui permettra d'économiser de l'énergie. Dans notre cas, cette fonctionnalité est très intéressante, car nous considérons un système temps réel où les requêtes des utilisateurs doivent être exécutées dans de brefs délais tout en minimisant la consommation d'énergie.

Chopra *et al.* (2017) ont mis en œuvre un système de surveillance et de contrôle de l'électricité en utilisant la logique floue et le Cloud Computing. Ce système est responsable de l'ajustement automatique du temps de travail des appareils électriques. Le système prend l'humidité et la température comme entrées et calcule le temps de travail des appareils en sortie. Les résultats ont montré que le système proposé permet d'économiser de l'énergie et aide à avoir une utilisation énergétiquement efficace. Benblidia *et al.* (2019) traitent la nécessité d'une méthode efficace de répartition des tâches dans une architecture *Fog-Cloud*. Leur proposition vise à envoyer les requêtes des utilisateurs au meilleur nœud Fog, tout en satisfaisant les préférences des utilisateurs et les contraintes des nœuds Fog. Pour cela, ils ont utilisé des quantificateurs linguistiques et des propositions quantifiées floues regroupant efficacement les préférences des utilisateurs et les contraintes des nœuds Fog, afin de classer les nœuds Fog du plus satisfaisant au moins satisfaisant. Les résultats des simulations ont montré que le système répartit les tâches tout en satisfaisant les préférences des utilisateurs. Il fournit également un compromis entre la satisfaction moyenne des utilisateurs, le délai d'exécution et la consommation électrique.

7.5. Conclusion

Les chercheurs et les industriels portent un grand intérêt à la réduction des coûts d'exploitation et de consommation d'énergie ainsi qu'aux émissions de carbone des Data Centers. Avec le développement du Smart Grid, ces efforts ont pris une autre dimension en intégrant des concepts visant à améliorer l'efficacité énergétique des Data Centers. La réponse à la demande, la tarification dynamique et l'intégration des ressources d'énergie distribuées sont les approches les plus adoptées.

Dans ce chapitre, nous nous sommes intéressés à l'efficacité énergétique des Data Centers dans un système Smart Grid-Cloud. Nous avons tout d'abord introduit le Smart Grid, son architecture intégrant les technologies de l'information et de la communication, ainsi que ses principales fonctionnalités. Après avoir détaillé l'interaction du Smart Grid avec les Data Centers du Cloud, nous avons résumé les principales techniques d'efficacité énergétiques utilisées. Aussi, nous avons fait une étude comparative des travaux réalisés à ce sujet. Ensuite, nous avons donné un aperçu des techniques d'aide à la décision dans un système Smart Grid-Cloud.

Par rapport aux travaux que nous avons étudiés, nous constatons que la prise en considération des prix et l'utilisation des énergies renouvelables demeurent deux facteurs principaux de l'efficacité énergétique dans les Data Centers. Actuellement, deux principales thématiques commencent à être exploitées pour une interaction

Smart Grid-Cloud plus efficace énergétiquement. La première vise à faire tourner les Data Centers dans des Micro-Grids pour assurer une fiabilité des systèmes électriques, tandis que la deuxième vise à répartir la charge de travail entre les Data Centers de manière à minimiser l'émission de gaz à effet de serre.

7.6. Bibliographie

Allcott, H. (2009). *Real Time Pricing And Electricity Markets*. Harvard University, Cambridge.

Asmus, P. (2017). Data centers and advanced microgrids. Libre blanc, NAVIGANT Research.

Avgerinou, M., Bertoldi, P., Castellazzi, L. (2017). Trends in data center energy consumption under the european code of conduct for data centre energy efficiency. Commission européenne.

Benblidia, M.A., Brik, B., Esseghir, M., Merghem-Boulahia, L. (2018). A Game Based Power Allocation in Cloud Computing Data Centers. Dans *14th International Conference on Wireless and Mobile Computing, Networking and Communications*. WiMob, Limassol, 1–7.

Benblidia, M.A., Brik, B., Esseghir, M., Merghem-Boulahia, L. (2019). Ranking Fog nodes for Tasks Scheduling in Fog-Cloud Environments: A Fuzzy Logic Approach. Dans *14th International Wireless Communications and Mobile Computing*. IWCMC, Tanger, 1–7.

Bera, S., Misra, S., Rodrigues, J.J.P.C. (2015). Cloud Computing Applications for Smart Grid: A Survey. *IEEE Transactions on Parallel and Distributed Systems*, 26(5), 1477–1494.

Bergaentzlé, C., Clastres, C. (2013). Tarifications dynamiques et efficacité énergétique : l'apport des Smart Grids. *Presses de l'ISMEA*, XLVII(2), 348–363.

Cavdar, D., Alagoz, F. (2012). A survey of research on greening data centers. *IEEE GLOBECOM*, 3237–3242.

Chen, S., Wang, Y., Pedram, M. (2014). Resource allocation optimization in a data center with energy storage devices. Dans *IECON 2014 – 40th Annual Conference of the IEEE Industrial Electronics Society*. IECON, Dallas, 2604–2610.

Chiang, M., Zhang, T. (2016). Fog and iot: An overview of research opportunities. *IEEE Internet of Things Journal*, 3(6), 854–864.

Chiang, M., Ha, S., Chih-Lin, I., Risso, F., Zhang, T. (2017). Clarifying fog computing and networking: 10 questions and answers. *IEEE Communications Magazine*, 55(4), 18–20.

Chopra, P., Bedi, R.P.S. (2017). Application of Fuzzy logic in Cloud Computing: A Review. *International Journal of Scientific Research Engineering & Technology (IJSRET)*, 6(11).

Dai, X., Wang, J.M., Bensaou, B. (2016). Energy-efficient virtual machines scheduling in multi-tenant data centers. *IEEE Transactions on Cloud Computing*, 4(2), 210–221.

Deng, W., Liu, F., Jin, H., Li, B., Li, D. (2014). Harnessing Renewable Energy in Cloud Datacenters: Opportunities and Challenges. *IEEE Network Magazine*, 28(1), 48–55.

Deng, R., Yang, Z., Chow, M., Chen, J. (2015). A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. *IEEE Transactions on Industrial Informatics*, 11(3), 570–582.

Ding, Z., Xie, L., Lu, Y., Wang, P., Xia, S. (2018). Emission-aware stochastic resource planning scheme for data center microgrid considering batch workload scheduling and risk management. *IEEE Transactions on Industry Applications*, 54(6), 5599–5608.

Duong-Ba, T.H., Nguyen, T., Bose, B., Tran, T.T. (2018). A dynamic virtual machine placement and migration scheme for data centers. *IEEE Transactions on Services Computing*.

Elmroth, E., Leitner, P., Schulte, S., Venugopal, S. (2017). Connecting fog and cloud computing. *IEEE Cloud Computing*, 4(2), 22–25.

Endo, H., Kodama, H., Fukuda, H., Sugimoto, T., Horie, T., Kondo, M. (2013). Effect of climatic conditions on energy consumption in direct fresh-air container data centers. Dans *International Green Computing Conference Proceedings*, 1–10.

Gabr, A.Z., Helal, A.A., Abbasy, N.H. (2018). Dynamic pricing; different schemes, related research survey and evaluation. Dans *9th International Renewable Energy Congress*. Hammamet, 1–7.

Gandhi, A., Harchol-Balter, M., Das, R., Lefurgy, C. (2009). Optimal power allocation in server farms. *SIGMETRICS Performance Evaluation Review*, 37(1), 157–168.

Gu, C. *et al.* (2016). Lowering down the cost for green cloud data centers by using ESDs and energy trading. Dans *IEEE Trustcom/BigDataSE/ISPA 2016*. IEEE, Tianjin.

Gu, C., Fan, L., Wu, W., Huang, H., Jia, X. (2018). Greening cloud data centers in an economical way by energy trading with power grid. *Future Generation Computer Systems*, 78, 89–101.

Guerassimof, G. (2017). *Microgrids : pourquoi, pour qui ?*. Presses des Mines, Paris.

Guerassimof, G., Maizi, N. (2013). *Smart Grids : au-dela du concept, comment rendre les réseaux plus intelligents.* Presses des Mines, Paris.

Gungor, V.C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., Hancke, G.P. (2011). Smart Grid Technologies: Communication Technologies and Standards. *IEEE Transactions on Industrial Informatics*, 7(4), 529–539.

Hossain, E., Han, Z., Poor, H.V. (2012). *Communication architectures and models for Smart Grid*. Cambridge University Press, Cambridge.

Joshi, Y., Kumar, P. (2012). *Energy Efficient Thermal Management of Data Centers*. Springer-Verlag, New York.

Khan, A.A., Rehmani, M.H., Reisslein, M. (2016). Cognitive Radio for Smart Grids: Survey of Architectures, Spectrum Sensing Mechanisms and Networking Protocols. *IEEE Communications Surveys & Tutorials*, 18(1), 860–898.

Kiani, A., Ansari, N. (2018). Profit maximization for geographically dispersed green data centers. *IEEE Transactions on Smart Grid*, 9(2), 703–711.

Kliazovich, D., Bouvry, P., Audzevich, Y., Khan, S.U. (2010). Greencloud: A packet-level simulator of energy-aware cloud computing data centers. Dans *IEEE Global Telecommunications Conference GLOBECOM 2010*, 1–5.

Liu, L., Sun, H., Li, C., Hu, Y., Li, T., Zheng, N. (2018). Exploring customizable heterogeneous power distribution and management for datacenter. *IEEE Transactions on Parallel and Distributed Systems*, 29(12), 2798–2813.

Markovic, D.S., Zivkovic, D., Branovic, I., Popovic, R., Cvetkovic, D. (2013). Smart power grid and cloud computing. *Renewable and Sustainable Energy Reviews*, 24, 566–577.

Marrouchi, S., Ben Saber, S. (2014). A comparative study of fuzzy logic, genetic algorithm, and gradient-genetic algorithm optimization methods for solving the unit commitment problem. *Mathematical Problems in Engineering*, 2014, 1–14.

Mohsenian-Rad, A., Wong, V.W., Jatskevich, J., Schober, R., Leon-Garcia, A. (2010). Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans. Smart Grid*, 1(3), 320–331.

Rao, L., Liu, X., Xie, L., Liu, W. (2010). Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricity-market environment. *Proc. INFOCOM*, 1–9.

Samadi, P., Mohsenian-Rad, H., Schober, R., Wong, V.W. (2012). Advanced demand side management for the future smart grid using mechanism design. *IEEE Trans. Smart Grid*, 3(3), 1170–1180.

Sharma, N.K., Reddy, G.R.M. (2019). Multi-objective energy efficient virtual machines allocation at the cloud data center. *IEEE Transactions on Services Computing*, 12(1), 158–171.

Somani, A., Joshi, Y.K. (2009). Data Center Cooling Optimization: Ambient Intelligence Based Load Management. Dans *Proceedings of the ASME Summer Heat Transfer Conference 2009*.

Terzija, V., Valverde, G., Cai, D., Regulski, P., Madani, V., Fitch, J., Skok, S., Begovic, M.M., Phadke, A. (2011). Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks. *Proceedings of the IEEE*, 99(1), 80–93.

Vardakas, J.S., Zorba, N., Verikoukis, C.V. (2015). A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms. *IEEE Communications Surveys & Tutorials*, 17(1), 152–178.

Wang, H., Ye, Z. (2016). Renewable energy-aware demand response for distributed data centers in smart grid. Dans *IEEE Green Energy and Systems Conference (IGSEC)*. IEEE, Long Beach, 1–8.

Wang, D., Ren, C., Sivasubramaniam, A. (2013). Virtualizing power distribution in datacenters. *SIGARCH Computer Architecture News*, 41(3), 595–606.

Wang, Y., Lin, X., Pedram, M. (2014a). Coordination of the smart grid and distributed data centers: A nested game-based optimization framework. Dans *ISGT 2014*. IEEE, Washington, 1–5.

Wang, Y., Saad, W., Han, Z., Poor, H.V., Baar, T. (2014b). A game-theoretic approach to energy trading in the smart grid. *IEEE Transactions on Smart Grid*, 5(3), 1439–1450.

Wang, H., Huang, J., Lin, X., Mohsenian-Rad, H. (2016). Proactive demand response for data centers: A win-win solution. *IEEE Transactions on Smart Grid*, 7(3), 1584–1596.

Yan, Y., Qian, Y., Sharif, H., Tipper, D. (2013). A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements and Challenges. *IEEE Communications Surveys & Tutorials*, 15(1), 5–20.

Yang, J., Zhang, S., Wu, X., Ran, Y., Xi, H. (2017). Online Learning-Based Server Provisioning for Electricity Cost Reduction in Data Center. *IEEE Transactions on Control Systems Technology*, 25(3), 1044–1051.

Yildizoglu, M. (2011). *Introduction à la théorie des jeux.* Dunod, Paris.

Yu, L., Jiang, T., Cao, Y., Qi, Q. (2014). Carbon-aware energy cost minimization for distributed internet data centers in smart microgrids. *IEEE Internet of Things Journal*, 1(3), 255–264.

Yu, L., Jiang, T., Cao, Y. (2015). Energy cost minimization for distributed internet data centers in smart microgrids considering power outages. *IEEE Transactions on Parallel and Distributed Systems*, 26(1), 120–130.

Yu, L., Jiang, T., Zou, Y. (2016). Real-time energy management for cloud data centers in smart microgrids. *IEEE Access*, 4, 941–950.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.

8

Vers de nouvelles architectures intelligentes pour l'Internet des véhicules

Léo MENDIBOURE[1], Mohamed Aymen CHALOUF[2] et Francine KRIEF[3]

[1] *LaBRI, Bordeaux, France*

[2] *IRISA, Rennes, France*

[3] *ENSEIRB-MATMECA, Bordeaux, France*

8.1. Introduction

L'architecture de référence des systèmes de transport intelligents coopératifs pour *Cooperative-Intelligent Transport Systems* (C-ITS) (ETSI 2010) a été définie par le travail conjoint de différents organismes de standardisation : IEEE, ISO, ETSI, etc. Cette architecture, basée sur trois plans principaux (gestion, contrôle, sécurité), doit permettre le déploiement d'un système de communication véhiculaire à large échelle. Toutefois, comme cela a été démontré par les auteurs de (Kaiwartya *et al.* 2016), elle présente, en l'état actuel, différentes limitations : capacité à traiter un volume important de données, interopérabilité entre différents réseaux, garantie de la qualité de service (QoS), garantie de la sécurité et respect de la vie privée, etc. C'est pourquoi les réseaux véhiculaires ont évolué vers un nouveau paradigme : l'Internet des véhicules (IdV).

Se servant des principes de l'Internet des objets (IdO), l'IdV doit permettre de répondre aux limitations des réseaux véhiculaires *ad hoc*, encore appelés « VANET » (*Vehicular Ad hoc NETwork*). Pour ce faire, l'IdV s'appuie sur le développement de nouveaux types de communication : véhicule à piéton (VaP),

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

véhicule à objet (VaO), etc. Ceci doit rendre possible l'amélioration des services de sécurité routière, le développement de services de gestion du trafic global ainsi que la conception de nouveaux services de divertissement (multimédia, publicité, etc.).

L'architecture de communication devant permettre l'avènement de l'IdV n'a jusqu'ici pas été normalisée. Cette architecture doit remplir de nombreux critères : sécurité et protection de la vie privée, gestion simplifiée, diffusion optimisée de l'information, mise à l'échelle, etc. (Kaiwartya *et al.* 2016). Différentes architectures, basées sur l'architecture C-ITS de référence, ont été proposées dans la littérature afin de permettre ces différentes améliorations. Ces architectures se basent sur l'intégration de différentes technologies : *Network Function Virtualization* (NFV), *Software Defined Network* (SDN), *Blockchain*, *Edge Computing* (EC), etc.

L'ensemble de ces architectures présente un point commun : l'idée d'intégrer des techniques d'intelligence artificielle (IA) au sein du plan de contrôle. En effet, quelle que soit l'approche choisie (NFV, SDN, EC, etc.), le partage de connaissance et la prise de décisions intelligentes sont présentés comme des points essentiels. L'utilisation de l'IA, en complément d'autres technologies, pourrait apporter de nombreuses améliorations au plan de contrôle : routage dynamique, prédiction du mouvement des véhicules et contrôle de congestion.

Toutefois, l'ensemble des architectures décrites dans la littérature se concentrent uniquement sur une amélioration du plan de contrôle. Or, des prises de décisions intelligentes pourraient également renforcer le plan de sécurité et la protection de la vie privée (détection d'intrusion, routage sécurisé, etc.) ainsi que le plan de gestion de l'architecture (équilibrage de charge, allocation de ressources virtuelles, etc.). Aussi, comme nous l'avons suggéré dans (Mendiboure *et al.* 2019a), l'ajout d'un quatrième plan, le plan de connaissance, au service de l'ensemble des autres plans (gestion, contrôle, sécurité) pourrait présenter de nombreux avantages.

Dans ce chapitre, nous nous proposons de décrire et de comparer les travaux visant à améliorer le fonctionnement des réseaux véhiculaires grâce à l'intégration de l'IA au sein du plan de contrôle. Par la suite, nous définirons une architecture permettant de compléter ces travaux et d'offrir des services intelligents, non seulement au plan de contrôle mais également au plan de gestion et de sécurité. Pour chacun de ces plans, au travers de travaux existants, nous présenterons les avantages de l'utilisation des techniques d'IA et listerons certains défis qui restent à relever.

Ainsi, nous présentons tout d'abord dans la section 8.2 le contexte : le principe de l'IdV ainsi que ses applications. Par la suite, nous décrivons dans la section 8.3 les différents travaux intégrant des techniques d'IA au sein de l'architecture de communication véhiculaire. Dans cette même section, nous justifions l'intérêt de la

définition d'une nouvelle architecture pour l'IdV. Dans la continuité, la section 8.4 détaille l'architecture proposée, ses principaux composants, ainsi que ses avantages. Finalement, la section 8.5 présente les enjeux de l'architecture proposée et la section 8.6 représente la conclusion et les perspectives concernant l'intégration de techniques d'IA au sein de l'architecture de communication véhiculaire.

8.2. Internet des véhicules (IdV)

Cette section vise à présenter les caractéristiques de l'IdV ainsi que les applications rendues possibles par cette nouvelle architecture véhiculaire.

8.2.1. *Positionnement*

Comme indiqué dans (Kaiwartya *et al.* 2016), le développement de l'IdV a trois motivations principales :

– répondre aux limitations des réseaux véhiculaires *ad hoc*, à savoir :

- le manque d'interopérabilité : l'architecture véhiculaire actuelle ne permet pas l'interconnexion de réseaux hétérogènes (Wi-Fi, cellulaire, Li-Fi, etc.), empêchant ainsi le déploiement de services de transport intelligents globaux et fiables ;

- l'absence de connectivité internet : l'architecture véhiculaire actuelle ne garantit pas une connectivité internet aux véhicules. Par conséquent, le développement de services commerciaux (publicité, divertissement, etc.) ou de services de gestion du trafic efficace est impossible ;

- les prises de décisions intelligentes limitées : en raison de la connectivité internet limitée et de capacités de calcul/stockage réduites au niveau des véhicules, le traitement de volumes de données importants est impossible ;

- le manque d'échanges : les réseaux véhiculaires *ad hoc* étant conçus pour permettre une interconnexion des véhicules, l'intégration de nouveaux types d'objets connectés (caméras, téléphones, etc.) n'est pas prévue alors que celle-ci pourrait apporter de nombreux bénéfices (détection de piétons et cyclistes, création de carte coopérative, etc.) ;

– répondre à l'augmentation du nombre d'accidents sur les routes : les systèmes de transport intelligents visent premièrement à assurer et à améliorer la sécurité routière. Avec un nombre d'accidents et de morts sur les routes toujours croissant (Kaiwartya *et al.* 2016), le développement de services de sécurité routière est

essentiel et doit reposer sur une architecture permettant le développement de services globaux ;

– offrir de nouvelles opportunités : l'intégration de nouveaux objets au sein des réseaux véhiculaires, l'interopérabilité entre différents réseaux ainsi qu'une connectivité internet assurée pourraient permettre de générer d'importants volumes de données, d'offrir de nouveaux services et par conséquent d'attirer de nouveaux investisseurs : fournisseurs de services, constructeurs automobiles, etc.

8.2.2. *Caractéristiques*

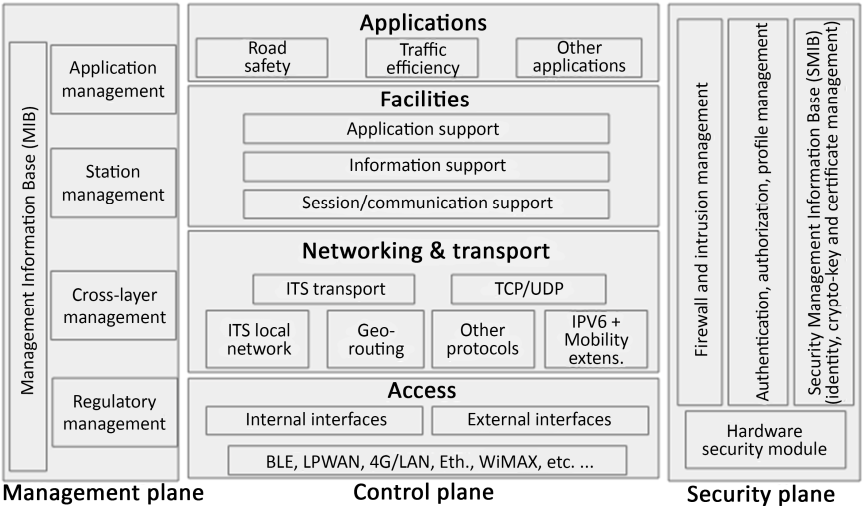


Figure 8.1. *Présentation de l'architecture C-ITS*

Le développement de l'IdV et l'amélioration des performances des réseaux de communication véhiculaires s'appuient sur deux grandes idées :

– l'intégration de nouveaux types d'objets au sein des réseaux véhiculaires au travers de nouveaux types de communication, on parle ainsi de communication V2X pour *Vehicle to Everything* :

- véhicule à piéton (VaP) ;
- véhicule à objet (VaO) ;
- véhicule à réseau électrique, encore appelé *Vehicle To Grid* (V2G) ;
- véhicule à *Cloud* (VaC) ;

– l'intégration de nouvelles technologies au sein de l'architecture C-ITS (figure 8.2) devant chacune permettre de s'affranchir de certaines des limitations des réseaux véhiculaires *ad hoc* :

- virtualisation de fonctions réseau, encore appelée *Network Function Virtualisation* (NFV) : virtualisation des fonctions et services réseau devant permettre la baisse des coûts (matériel standard), un déploiement plus rapide ainsi qu'un passage à l'échelle simplifié (flexibilité) ;

- réseau logiciel, encore appelé *Software Defined Networking* (SDN) : technique de découplage du plan de contrôle et du plan de données devant offrir au réseau une programmabilité et une vision d'ensemble, permettant ainsi une grande adaptation au contexte (dynamicité) ;

- *Edge Computing* : déploiement de capacités de calcul et de stockage en bordure du réseau, extension des capacités cloud, devant permettre de limiter les temps de latence (meilleure expérience utilisateur) et de décharger le cœur du réseau (passage à l'échelle) ;

- *blockchain* : technologie distribuée permettant le stockage et l'échange d'informations de façon transparente, sécurisée et infalsifiable, pouvant permettre le développement de solutions de sécurité efficaces, robustes et supportant le passage à l'échelle ;

- intelligence artificielle : techniques d'IA devant permettre d'améliorer l'architecture existante, mais également les technologies nouvellement intégrées (EC, NFV, SDN), au travers d'aides à la prise de décision et d'automatisation de processus.

8.2.3. *Principales applications*

En répondant aux limites des réseaux véhiculaires *ad hoc*, en intégrant de nouveaux types d'objets et en assurant une connectivité internet, l'IdV doit pouvoir permettre de nombreuses nouvelles applications que l'on peut classer en différentes catégories (Mendiboure *et al.* 2019b) :

– gestion des transports : gestion du trafic en temps réel, réservation de places de parking, points de recharge de véhicules électriques, etc. ;

– sécurité routière : gestion d'intersections, prévention de collisions coopérative, détection d'obstacles, etc. ;

– gestion du véhicule : assistant personnel, système de maintenance à distance, etc. ;

– assistance au conducteur : aide au freinage d'urgence pour *Advanced Driver Assistance Systems* (ADAS), aide au parking, création de cartes coopératives, etc. ;

– santé connectée : détection de fatigue, confort de l'utilisateur, assistance médicale, etc. ;

– nouveaux modes de transport : *platooning*, création de convois de véhicules connectés ou *car sharing*, partage de véhicule, nouveaux modes de transport visant à réduire le nombre de véhicules sur les routes ainsi que la consommation d'essence, et par conséquent la pollution.

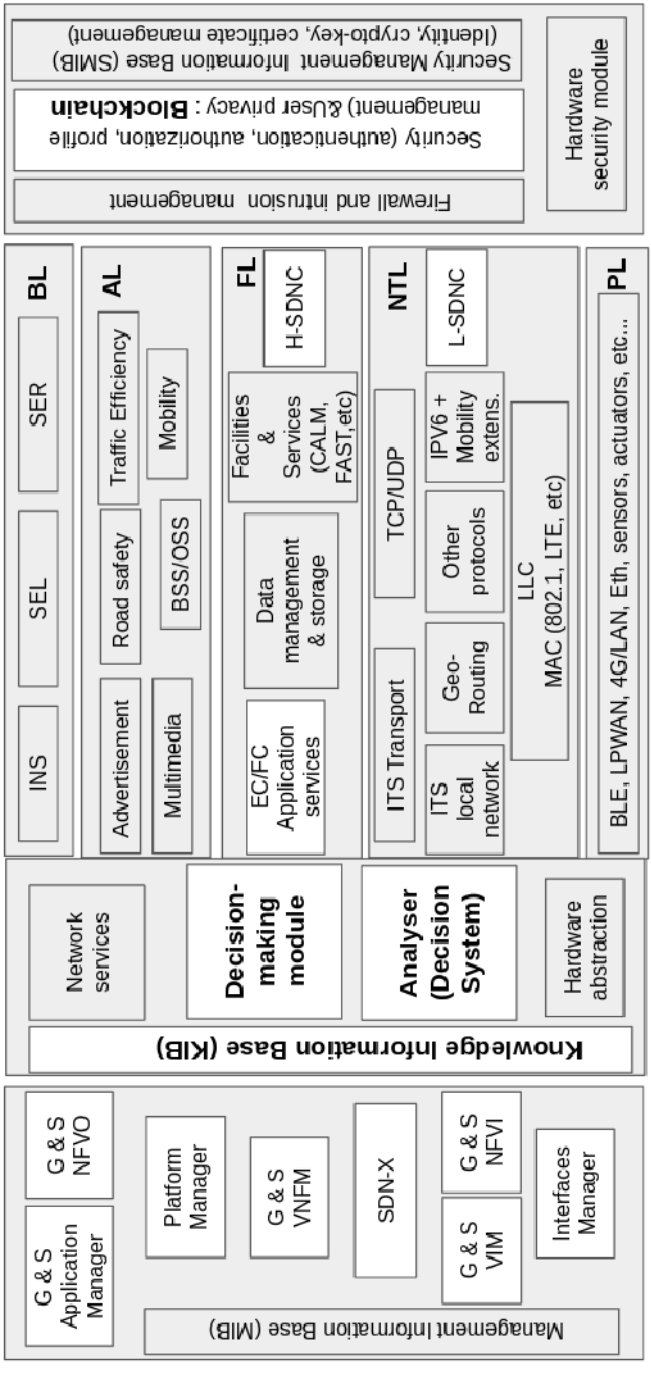
8.3. Les architectures IdV proposées dans la littérature

Pour surmonter les limites des réseaux véhiculaires *ad hoc* et offrir de nouveaux services, de nombreuses évolutions de l'architecture C-ITS ont été proposées dans la littérature. Cette section vise à présenter ces différentes architectures en mettant en avant un point essentiel : l'intégration de techniques d'IA au sein de ces systèmes. Ainsi, les avantages et limites de ces architectures sont évalués en fonction de leur niveau d'automatisation.

8.3.1. *Intégration de techniques d'IA au sein d'une couche du plan de contrôle*

La première approche proposée dans différents papiers consiste à intégrer au sein de l'architecture véhiculaire une nouvelle couche, une couche d'IA, devant permettre la récupération, le stockage, l'analyse et le traitement des données provenant des couches inférieures (Kaiwartya *et al.* 2016 ; Aliyu *et al*. 2017 ; Borcoci *et al.* 2017).

Les auteurs de (Aliyu *et al*. 2017, figure 8.2(2)) définissent une nouvelle architecture composée de quatre couches et spécifiquement destinée au développement de services de *Cloud Computing* dans l'environnement véhiculaire. La couche d'IA doit permettre le traitement des données provenant des couches inférieures (perception, coordination) et rendre ainsi possible des applications nécessitant le traitement de volumes massifs de données, notamment pour la gestion du trafic en temps réel. Cette couche vise également à améliorer le fonctionnement des services cloud en augmentant la capacité de calcul des véhicules utilisant des services cloud (VuC pour *Vehicle using Cloud*) et en optimisant l'utilisation des ressources des véhicules intégrés au cloud (VCC pour *Vehicular Cloud Computing*).



MOP: Management and Operation Plane; KP: Knowledge Plane; CDP: Control and Data Plane; SPP: Security and Privacy Plane
BL: Business Layer; AL: Application Layer; FL: Facilities Layer; NTL: Network and Transport Layer; PL: Perception Layer

Figure 8.2. *Présentation des architectures IdV décrites dans la littérature*

Une architecture similaire, devant servir de base à l'IdV, est proposée par les auteurs de (Kaiwartya *et al.* 2016, figure 8.2(6)). Cette couche, utilisant différentes techniques d'IA (systèmes experts, VCC, analyse de données massives), a les mêmes objectifs que celle présentée par les auteurs de (Aliyu *et al.* 2017) : traitement de volumes de données massives et amélioration du fonctionnement des applications. Néanmoins, selon les auteurs, cette couche pourrait également permettre la commercialisation des réseaux véhiculaires. En effet, la définition et le développement de modèles économiques (couche *Business*) pour les fournisseurs de services pourraient s'appuyer sur l'analyse statistique des données d'utilisation des applications fournies par la couche d'IA.

Cette définition d'une architecture à cinq couches est reprise par les auteurs de (Borcoci *et al.* 2017, figure 8.2(3)). Toutefois, elle est étendue dans deux directions : intégration de capacités d'EC et intégration de SDN au sein de l'architecture. Les capacités de calcul en bordure du réseau (EC) pourraient permettre le déploiement de services intelligents au plus proche de l'utilisateur, améliorant ainsi la qualité d'expérience de l'utilisateur et limitant la remontée d'informations. De plus, l'utilisation de SDN pourrait permettre d'améliorer la gestion de ces ressources : positionnement d'applications et réservation de ressources (bande passante, capacité de stockage et de calcul, etc.).

8.3.2. *Intégration de techniques d'IA au sein de plusieurs couches du plan de contrôle*

Toutefois, l'intérêt de l'intelligence artificielle ne se limite pas à la gestion des ressources cloud et au traitement d'informations destinées aux applications.

Ainsi, les auteurs de (Contreras-Castillo *et al.* 2017, figure 8.2(4)) définissent une architecture à sept couches destinée à l'IdV. L'intelligence est ici intégrée à différents niveaux : au niveau de la couche de prétraitement, au niveau de la couche de communication et au niveau de la couche de contrôle. Au niveau de la couche de prétraitement, le traitement des données doit permettre de limiter la quantité de données transmises et d'en améliorer la pertinence, diminuant ainsi la charge du réseau. Au niveau de la couche de communication, l'IA doit permettre un choix optimal de la technologie de transmission en fonction de la QoS nécessaire. Enfin, au niveau de la couche de contrôle, l'IA doit permettre une gestion des fournisseurs de services et de l'échange de données, comme cela était déjà proposé dans les travaux précédents (section 8.3.1).

Cette idée d'application de techniques d'IA à différents niveaux se retrouve également dans (Chen *et al.* 2018, figure 8.2(5)). Les auteurs introduisent l'idée

d'Internet des véhicules intelligent (CIoV pour *Cognitive Internet of Vehicles*). Selon eux, l'absence de prises de décisions intelligentes limiterait l'impact de l'IdV : gestion sous-optimale des ressources, manque de fiabilité du processus de décision, faible potentiel marketing, etc. Ainsi, l'IA s'applique ici non seulement à une gestion du réseau (gestion du trafic réseau, gestion de la sécurité, allocation de ressources), mais également à une protection des usagers de la route : autoréparation du véhicule, surveillance de la santé du conducteur, contrôle de la sécurité du conducteur, analyse des émotions, etc. L'application de l'IA est par conséquent étendue à différents niveaux (communication, cognition, contrôle), mais également à différents domaines (humain et réseau).

8.3.3. *Définition d'un plan de connaissance associé au plan de contrôle*

Pour permettre des prises de décisions inter couches, les auteurs de (Jiacheng *et al*. 2016, figure 8.2(1)) proposent l'ajout d'un plan de connaissance. Dans cette architecture basée sur SDN, le plan de connaissance doit rendre possible le traitement des informations provenant des couches les plus basses (couche de données) en fonction des besoins des autres couches (application, contrôle). Ainsi, ce plan doit aider à surmonter les principaux défis liés à l'intégration de SDN : gestion de la mobilité et hétérogénéité des ressources au travers de différentes fonctions (contrôle de la transmission des paquets, virtualisation du réseau, gestion du *handover*, etc.).

8.3.4. *Comparaison des architectures et positionnement*

Les travaux présentés dans cette section visent à améliorer l'architecture de communication véhiculaire grâce à l'intégration de techniques d'IA.

Architecture	Apport			Multicouche	Plan de connaissance
	Gestion	Contrôle	Sécurité		
(Aliyu *et al*. 2017)	Cloud	Oui	Non	Non	Non
(Kaiwartya *et al.* 2016)	Cloud	Oui	Non	Non	Non
(Borcoci *et al.* 2017)	Cloud	Oui	Non	Non	Non
(Contreras-Castillo *et al.* 2017)	Non	Oui	Non	Oui	Non
(Chen *et al*. 2018)	Oui	Oui	Non	Oui	Non
(Jiacheng *et al*. 2016)	Non	Oui	Non	Oui	Oui
Proposition	Oui	Oui	Oui	Oui	Oui

Tableau 8.1. *Comparaison des différentes architectures proposées*

L'ensemble des travaux présentés visent avant tout à permettre une amélioration des performances du plan de contrôle (tableau 8.1). Certains de ces travaux se concentrent sur l'évolution d'une seule couche, la couche applicative, au travers de l'ajout d'une couche d'IA de prétraitement et d'analyse des données au service des applications (Kaiwartya *et al.* 2016 ; Aliyu *et al*. 2017 ; Borcoci *et al.* 2017). D'autres proposent des approches multicouches destinées également à un meilleur transport des données (interopérabilité, routage, optimisation de bande passante) (Contreras-Castillo *et al.* 2017 ; Chen *et al.* 2018). Un seul de ces papiers (Jiacheng *et al*. 2016) propose l'ajout d'un plan de connaissance qui pourrait permettre une prise de décisions inter couches et un traitement centralisé et optimisé des informations.

Certains travaux proposent également d'utiliser l'IA au service du plan de gestion (gestion des ressources cloud, gestion des ressources en bordure du réseau, positionnement des applications) (Kaiwartya *et al.* 2016 ; Aliyu *et al*. 2017 ; Borcoci *et al.* 2017 ; Chen *et al.* 2018). Toutefois, les auteurs de (Jiacheng *et al*. 2016) n'abordent pas cette question et se concentrent sur les interactions entre le plan de contrôle SDN et le plan de connaissance. Aussi, l'idée de prises de décisions inter couches et d'optimisation à plusieurs niveaux n'a été jusqu'ici proposée qu'au service du plan de contrôle. De plus, l'idée d'appliquer l'IA au service du plan de sécurité et de protection de la vie privée (intrusions, routage, etc.) n'a été abordée dans aucun des travaux cités précédemment, malgré la pertinence de cette approche.

Pour répondre à ces limites, nous nous proposons dans la suite de ce document de définir une nouvelle architecture pour l'IdV et d'en étudier les principaux avantages, ainsi que les principaux verrous en termes de contrôle, gestion et sécurité et vie privée.

8.4. Notre proposition d'architecture IdV intelligente

Dans cette section, l'architecture IdV proposée ainsi que les différents plans la composant (contrôle, gestion, sécurité et vie privée, et connaissance) sont présentés. De plus, les différentes applications de l'IA au service du contrôle, de la gestion et de la sécurité et de la vie privée présentes dans la littérature sont analysées.

8.4.1. *Présentation*

L'architecture IdV intelligente que nous proposons est présentée figure 8.3. Elle comprend les quatre plans suivants :

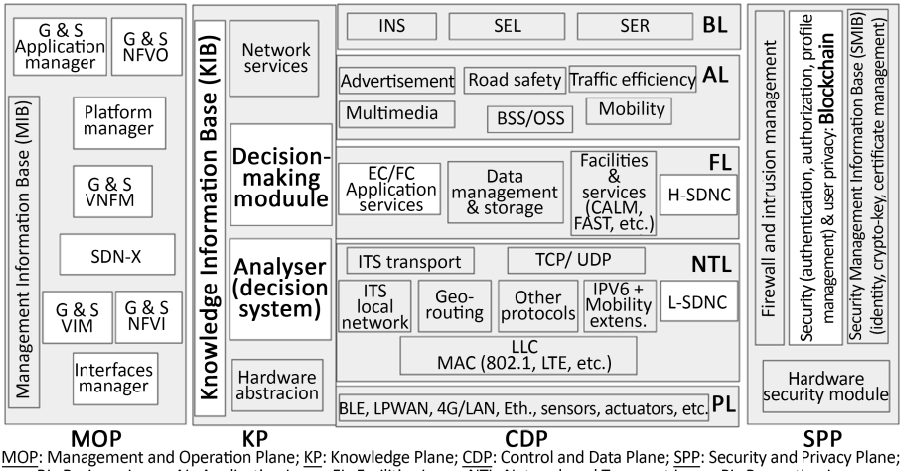
– *Management and Orchestration Plane* (MOP) : ce plan est chargé de la gestion globale de l'architecture et des ressources ;

– *Security and Privacy Plane* (SPP) : ce plan offre des services de sécurité et de protection de la vie privée à l'ensemble des autres plans : autorisations, contrôle d'accès, gestion de profils, etc. ;

– *Knowledge Plane* (KP) : ce plan offre des services de traitement et d'analyse de l'information et de prises de décisions à l'ensemble des autres plans ;

– *Control and Data Plane* (CDP) : ce plan est chargé de la récupération, du transport et de la distribution optimale de l'information.

Cette architecture est la première à proposer un plan de connaissance pour l'IdV. Les premiers travaux sur le plan de connaissance datent de 2003 et avaient pour objectif de permettre aux réseaux de prendre des décisions, en particulier de gestion, sans intervention humaine (Clark *et al*. 2003). Les tâches de gestion étaient effectuées par le réseau lui-même, chaque entité autonome étant capable de s'auto-configurer, s'auto-optimiser, s'auto-protéger et s'auto-réparer (Krief 2010). Les difficultés de mise en œuvre ont limité les réalisations du plan de connaissance à des preuves de concept, mais les avancées actuelles, principalement en matière d'intelligence artificielle, permettent d'envisager aujourd'hui son déploiement.



MOP: Management and Operation Plane; KP: Knowledge Plane; CDP: Control and Data Plane; SPP: Security and Privacy Plane; BL: Business Layer; AL: Application Layer; FL: Facilities Layer; NTL: Network and Transport Layer; PL: Perception Layer

Figure 8.3. *Notre proposition*

8.4.2. *Un plan de connaissance au service du transport de données*

La mobilité des véhicules a un impact important sur le transport des données. En effet, la variation du nombre de véhicules en fonction du temps dans un espace

donné implique une optimisation constante du transport des données pour garantir des performances élevées : routage, congestion, etc. L'application de techniques d'IA au plan de contrôle de l'architecture a été proposée pour différentes applications, notamment la prédiction du mouvement des flux de véhicules, le contrôle de congestion et le routage dynamique.

La prédiction du mouvement des flux de véhicules est un premier point important. En effet, celle-ci doit permettre d'améliorer le fonctionnement de nombreuses applications ITS : gestion de la congestion du trafic routier, réduction des émissions de gaz à effet de serre, réduction de la consommation de carburant ou amélioration de services basés sur la localisation. Les auteurs de (Huang *et al*. 2014) sont les premiers à avoir proposé une approche basée sur l'apprentissage profond pour la prédiction du mouvement des flux de véhicules. Cette solution, basée sur une pile de machines de Boltzmann restreintes, intégrant l'idée d'apprentissage multitâche, vise à améliorer les performances des méthodes basées sur des réseaux statiques préalablement utilisées et à démontrer l'intérêt d''une approche dynamique. Néanmoins, le système proposé pourrait être amélioré. En effet, la prédiction du trafic n'est ici pas couplée à l'information temporelle. De la même manière, les auteurs de (Lv *et al*. 2015) ont proposé une autre approche basée sur l'apprentissage profond devant permettre une prédiction fine du mouvement des véhicules. Grâce à l'apprentissage non supervisé d'un modèle d'auto-encodeurs, cette approche doit ainsi permettre de réaliser des corrélations spatiales et temporelles non linéaires en analysant le flux de données. La comparaison de cette méthode avec d'autres approches existantes (rétropropagation, marche aléatoire, machines à vecteurs de support) démontre son efficacité en termes de prédiction du trafic routier. Enfin, dans (Ide *et al*. 2015), les auteurs se penchent à la fois sur la question de la prédiction du mouvement des flux de véhicules et sur celle de la connectivité des communications LTE. Ainsi, grâce aux données physiques remontées des véhicules et à l'utilisation de réseaux de dépendances de Poisson, ils construisent un modèle leur permettant de prédire le mouvement des véhicules. Grâce à des informations de mesure de connectivité cellulaire, le système proposé permet également, en utilisant des arbres de régression de Poisson, de prédire de façon distribuée la connectivité cellulaire pour chaque véhicule. Grâce à cela, il doit être possible d'améliorer le processus de *handover* vertical et d'optimiser les communications machine-à-machine.

Un second point important est le contrôle de congestion. En effet, pour limiter la perte de paquets, le délai de transmission de bout en bout, et pour augmenter la fiabilité des réseaux véhiculaires, limiter la congestion des données est essentiel. Les stratégies de contrôle de congestion peuvent être basées sur différentes approches visant à optimiser différents paramètres : gestion de la puissance, gestion du débit,

prioritisation et ordonnancement, etc. L'idée d'un contrôle de congestion intelligent, pour améliorer ces solutions, a été proposée dans de nombreux papiers. Par exemple, les auteurs de (Taherkhani et Pierre 2016) discutent une solution basée sur un partitionnement en k-moyennes (*k-means*) devant permettre de gérer la congestion au niveau des intersections en utilisant les équipements de bord de route (RSU pour *RoadSide Units*) pour contrôler le niveau de congestion des canaux radio. Ces RSU récupèrent les données transmises par l'ensemble des véhicules au niveau de l'intersection et les regroupent en fonction de différents paramètres : taille et type des messages, validité, position du receveur, distance entre le véhicule et l'infrastructure de bord de route, etc. Par la suite, en fonction des ressources disponibles, des paramètres de communication sont affectés à chacun de ces groupes : délai de transmission, puissance de transmission, taille de la fenêtre de congestion, etc. Ceci permet par conséquent de limiter la congestion. Néanmoins, la solution proposée présente différentes limitations, notamment la capacité de calcul et la latence liées au traitement de ces informations. Pour assurer le bon fonctionnement des services de sécurité routière coopératifs, les véhicules diffusent des messages courts aux véhicules environnants (*beacon*) contenant différentes informations : position, vitesse, etc. Dans une zone contenant un nombre important de véhicules, la transmission de ces messages réguliers pourrait entraîner des problèmes de congestion. C'est pourquoi les auteurs de (Toutouh et Alba 2016, 2018) ont proposé des méthodes de contrôle de congestion basées sur une intelligence distribuée (*swarm intelligence*) visant à optimiser au mieux la fréquence de ces messages. Ces méthodes doivent permettre de maintenir le taux d'utilisation des canaux de communication radio à un niveau permettant d'assurer le bon fonctionnement du réseau, en évitant la congestion, tout en maximisant le nombre de messages échangés et, par conséquent, en améliorant la fiabilité des informations reçues par chacun des véhicules. Toutefois, comme cela est souligné par les auteurs, les performances de ces deux algorithmes doivent encore être démontrées dans un environnement réel.

Enfin, pour établir une communication entre deux véhicules, un calcul du chemin prenant en compte les caractéristiques des différents liens (durée de vie, bande passante, délai, etc.) est nécessaire. Pour réaliser un calcul optimal, de nombreuses approches basées sur des techniques d'IA ont été proposées. Par exemple, les auteurs de (Zhang *et al*. 2018a) proposent une solution basée sur des algorithmes génétiques devant permettre un routage optimal visant à garantir une QoS élevée. Ces algorithmes cherchent à trouver rapidement une solution approchée à un problème d'optimisation. En utilisant les informations liées au trafic des véhicules (vitesse, direction, etc.), les auteurs décrivent un mécanisme visant à calculer dynamiquement le chemin de communication optimal entre deux véhicules. La solution proposée doit permettre de gérer la dynamicité de la connectivité (mouvement des véhicules, cassure de liens) et d'optimiser le délai de transmission des paquets. Néanmoins, le processus proposé est

complexe et l'optimisation du temps de recherche du chemin optimal devrait être considérée. Dans (Lai *et al.* 2015), une approche basée sur le *Machine Learning* est proposée. Pour la diffusion de données, trois problèmes principaux sont considérés par les auteurs : prédiction du mouvement des véhicules, évaluation de la capacité de transmission des équipements et évaluation de la direction de transmission. L'étude de ces différentes problématiques est assurée au niveau des RSU, qui partagent entre eux les informations concernant la position des véhicules et la capacité de leurs liens de communication. Un calcul intelligent du chemin optimal est ensuite réalisé par le système. Les auteurs souhaitent ainsi démontrer la pertinence de l'usage de techniques d'IA au service du routage, assurant un taux de perte de paquets et une latence plus faibles que celles des solutions existantes. Enfin, dans (Yao *et al.* 2018) est décrit un mécanisme devant permettre de déterminer les positions futures des véhicules et de trouver le chemin optimal pour atteindre ces positions à venir. Pour ce faire, les auteurs proposent un système basé sur un modèle de Markov caché et sur l'historique des mouvements passés des véhicules. En effet, en analysant les lieux où ces véhicules se sont précédemment rendus et en faisant une corrélation entre le temps et l'espace, il est possible de déterminer où les véhicules sont en train de se rendre et, par conséquent, quels chemins ils vont utiliser. Grâce à cette approche, il semble possible d'améliorer les délais de transmission des paquets et d'en réduire les pertes. Toutefois, pour vérifier les performances de la solution proposée par les auteurs, une implémentation dans un environnement réel est nécessaire.

8.4.3. *Un plan de connaissance au service de la gestion de l'architecture IdV*

L'architecture de communication véhiculaire doit permettre une gestion optimale des ressources à disposition : calcul, stockage et communication. Cette gestion doit intervenir à différents niveaux, afin de permettre un équilibrage de charge entre différents réseaux hétérogènes, entre différents opérateurs ainsi qu'entre différents serveurs de calcul. Ceci doit permettre d'assurer des performances élevées pour l'ensemble des utilisateurs, offrant la possibilité de minimiser le délai et la perte de paquets et de maximiser la bande passante. Tout comme pour la transmission de données, pour une optimisation de l'utilisation des ressources, l'approche IA a été proposée dans de nombreux papiers. Les travaux existants peuvent être classifiés en trois grandes catégories correspondant à différents apports : équilibrage de charge entre différents réseaux hétérogènes, équilibrage de charge pour les communications intervéhiculaires et gestion des ressources de calcul et de stockage (*Edge Computing*).

En raison du déplacement des véhicules, des zones et, par conséquent, des stations de base peuvent avoir à gérer un nombre important de communications,

pouvant entraîner une baisse de la QoS. C'est pourquoi des solutions intelligentes ont été développées. Ces solutions permettent un équilibrage de charge et un contrôle vertical dans les réseaux véhiculaires. Par exemple, les auteurs de (Li *et al.* 2017) proposent une solution visant à permettre une association efficace de l'utilisateur à des stations de base hétérogènes, grâce à l'apprentissage par renforcement. En effet, pour pouvoir garantir une bonne QoS, une répartition optimisée des véhicules entre les différentes stations de base est nécessaire. Ainsi, l'association est dans un premier temps réalisée en fonction des informations actuelles des stations de base et de leurs capacités. Par la suite, les associations sont optimisées en fonction des informations collectées par ces stations de base (heures de pic de trafic, mouvement des véhicules, etc.), permettant de déterminer le nombre de véhicules devant être associés dans une zone donnée et, par conséquent, comment les associer. Afin de démontrer les bénéfices de cette approche, les auteurs l'ont comparée à des méthodes couramment utilisées : rapport signal sur bruit maximal (*max-SINR*) et méthode d'optimisation de grands systèmes (*Distributed Dual Decomposition Optimization*). Dans le cas de réseaux hétérogènes, l'équilibrage de charge peut également être utilisé pour garantir une QoS maximale à tout moment, comme cela est montré dans (Xu *et al.* 2014). Les auteurs de ce papier proposent également d'utiliser une méthode d'apprentissage par renforcement (*Fuzzy Q-learning*) devant offrir une adaptabilité importante à l'évolution des conditions de trafic. En effet, celle-ci doit permettre de déterminer le réseau optimal pour la transmission de données en fonction de quatre paramètres : rapport signal sur bruit, nombre de véhicules connectés au réseau visé, vitesse du véhicule et quantité de données à transmettre. Les auteurs démontrent la pertinence de ce service pour les applications multimédias. Grâce à la technique d'équilibrage de charge proposée, le débit est amélioré et une faible latence est assurée. Enfin, et dans un souci de limiter le déploiement de stations de bases et d'utiliser au mieux les ressources disponibles, les auteurs de (Zheng *et al*. 2016) ont défini un système devant permettre d'optimiser le délai de transmission des paquets dans un environnement virtuel (SDN). Ce système, reposant sur des chaînes de Markov et des équations de Bellman, est composé de deux parties principales. La première est une gestion macroscopique de l'allocation des ressources virtuelles (MaVRA) basée sur l'analyse de paramètres globaux et étendus dans le temps, notamment la densité du trafic. La seconde partie, quant à elle, consiste en une gestion microscopique de l'allocation des ressources virtuelles (MiVRA) en fonction de paramètres évoluant rapidement, notamment l'état des files d'attente des équipements réseau et l'état des canaux de communication. Au travers d'une simulation, les auteurs de ce papier démontrent les gains en termes de délai de l'approche proposée.

Certaines applications, notamment les applications liées à la sécurité routière, ont des contraintes de latence fortes et se basent principalement sur des communications

inter véhiculaires. Toutefois, pour maintenir une QoS élevée, il est important d'utiliser au mieux l'ensemble des ressources disponibles. C'est pourquoi les auteurs de (Ashraf *et al*. 2016) ont introduit un système permettant un équilibrage de charge entre les différents chemins VaV (véhicule à véhicule). Pour ce faire, les auteurs définissent un mécanisme introduisant la formation de groupes de véhicules en fonction de leurs déplacements (position, direction, vitesse). La formation de ces groupes ont permis l'optimisation des échanges intergroupes. Cette optimisation a également lieu à l'intérieur de ces groupes grâce à la formation de paires de véhicules en fonction de la stabilité et de la QoS des liens de communication. Ainsi, ce système permet d'utiliser au mieux les liens VaV disponibles. De la même manière, les auteurs de (Ye *et al.* 2019) définissent un mécanisme décentralisé permettant à un véhicule d'utiliser au mieux les ressources en communication à disposition. Ce système, basé sur un apprentissage par renforcement et sur un apprentissage profond, respecte les contraintes des communications VaV, tout en minimisant les interférences avec les communications VaI (véhicule à infrastructure).

Enfin, pour décharger le cœur du réseau et offrir une latence plus faible ainsi qu'un débit plus élevé, l'idée de déployer des serveurs de stockage et de traitement de l'information en bordure du réseau (*Edge Computing*) revient dans de nombreux travaux. Toutefois, que ces ressources soient disponibles au niveau de l'unité de bord de route (RSU, stations de base) ou à l'intérieur des véhicules, de nombreuses problématiques liées à l'utilisation de ces ressources se posent. C'est pourquoi les auteurs de (Sun *et al.* 2018) s'intéressent à la question du partage de ressources de calcul entre véhicules. Ainsi, si un véhicule dispose de ressources qu'il n'utilise pas à un moment donné, ces ressources pourront être utilisées par des véhicules environnants. Toutefois, le problème de la QoS de ce type de service se pose, et c'est pourquoi les auteurs proposent un système devant permettre de minimiser le temps de latence lors de l'externalisation du traitement des données. Cette approche repose sur un algorithme de réplication des tâches basé sur la théorie des jeux (*Combinatorial Multi-Armed Bandit Theory*). Ce système doit permettre à chaque véhicule souhaitant externaliser des tâches de calcul de savoir combien de temps prendrait chacun des véhicules environnant pour réaliser cette tâche. Grâce à cela, il est possible pour ces véhicules de sélectionner un véhicule offrant une QoS suffisante pour l'application visée. Les auteurs se penchent également sur la question de la réplication des tâches, lorsque la densité de véhicules est importante, et sur l'évaluation des performances de ce système. Ces tâches peuvent également être traitées au niveau des RSU. Toutefois, pour éviter la surcharge d'un équipement, la collaboration entre ces équipements est primordiale. C'est pourquoi, dans (Li *et al.* 2019), est introduit un système basé sur l'apprentissage par renforcement, afin de permettre un équilibrage de charge entre différents serveurs de bord de route. Ce

système se base sur une analyse du trafic routier et une prédiction du trafic routier à venir pour estimer le niveau de charge de chacun des serveurs des équipements de bord de route et, par conséquent, pour optimiser la répartition des tâches (traitement de données, stockage, etc.), notamment pour les applications de carte HD (haute définition) devant améliorer les conditions du trafic routier.

8.4.4. *Un plan de connaissance au service de la sécurisation de l'architecture IdV*

Pour garantir la sécurité des usagers de la route et contrer les actions d'entités malveillantes, la sécurisation des communications véhiculaires est également une question essentielle. En effet, l'envoi de messages erronés, la suppression de messages ou encore la perturbation des communications pourraient fortement détériorer les performances du système IdV, voire même le rendre dangereux pour la sécurité routière. Grâce à l'intégration de techniques d'IA, le niveau de sécurité de l'architecture IdV pourrait être amélioré, notamment au travers de la détection d'intrusions, de la prévention des attaques de routage et de la protection contre les attaques de déni de service (DoS pour *Denial of Service*).

La détection d'intrusion doit permettre de mettre en évidence des comportements anormaux, et ce à différents niveaux : véhicule, équipements de bord de route, etc. Les auteurs de (Kang et Kang 2016) proposent un système devant permettre de détecter les intrusions à l'intérieur du véhicule, grâce à des techniques d'apprentissage profond. Plus particulièrement, avec ce système, il doit être possible de détecter des comportements anormaux au sein du bus de données CAN (*Controller Area Network*) du véhicule. Ce bus CAN est un système utilisé pour établir une communication entre les différents composants du véhicule. Sa sécurité est par conséquent essentielle au fonctionnement du véhicule autonome. La méthode de classification non supervisée proposée et la définition de vecteurs de caractéristiques basés sur l'analyse des paquets transitant dans le véhicule doivent permettre, d'après l'évaluation effectuée par les auteurs, une détection efficace des attaques en temps réel. Les auteurs de (Zeng *et al.* 2018) visent, quant à eux, à proposer un système devant permettre grâce au *Machine Learning* de détecter des intrusions localement et globalement. La détection globale s'effectue au niveau des RSU, qui disposent d'une capacité de calcul importante et d'une vue globale du réseau grâce à leurs échanges. La détection locale s'effectue, quant à elle, au niveau des véhicules. Les véhicules étant organisés en *cluster*, ce système vise principalement à identifier les comportements anormaux de deux types de véhicules : les leaders (CH pour *Cluster Head*) et les véhicules échangeant avec d'autres *clusters* (MPR pour *MultiPoint Relay*). Au niveau des RSU, les données concernant le

contexte et le comportement des CH sont remontées et analysées en utilisant un réseau neuronal, afin de déterminer si le comportement du leader est normal ou malicieux. Localement, les véhicules sont en charge de contrôler le comportement des véhicules communiquant avec d'autres *clusters*, les MPR. Leurs capacités de calcul étant limitées, ces véhicules se contentent de comparer le nombre de paquets reçus et transmis par ces véhicules MPR. Ainsi, la combinaison de ces deux systèmes (locaux et globaux) doit permettre la sécurisation des communications véhiculaires et la suppression des véhicules malveillants. Pour finir, dans (Alheeti *et al.* 2015) sont comparées les performances de différentes techniques d'IA pour la détection d'intrusions et de comportements anormaux de véhicules (*Grey Hole*) dans les réseaux véhiculaires. La première approche est basée sur un système de machine à vecteur de support (SVM pour *Support Vector Machine*) et la seconde sur un système de réseaux de neurones à propagation avant (FFNN pour *FeedForward Neural Network*). Cette comparaison démontre la grande efficacité des réseaux de neurones pour la détection d'anomalies et la limitation du nombre de faux négatifs. Toutefois, les performances de cette approche en termes de délais s'avèrent bien moindres que celles de l'approche SVM.

La détection et la prévention d'attaques de déni de service (DoS) est un autre point essentiel. En effet, la perturbation des services de communication véhiculaire pouvant empêcher les échanges entre véhicules ou entre véhicules et infrastructure pourraient mettre en danger la sécurité des usagers de la route. C'est pourquoi les auteurs de (Nyabuga *et al.* 2016) proposent de détecter ces attaques grâce à une optimisation par essaims particulaires. En effet, pour la détection d'attaques de DoS sur les communications VaI et VaV, l'approche *swarm* pourrait offrir des probabilités de détection supérieures à des approches utilisant des algorithmes génétiques, comme cela est démontré par les auteurs. De leur côté, les auteurs d'(Alheeti *et al.* 2016) proposent la première approche basée sur des réseaux de Petri (*Fuzzy Petri Nets*) pour la détection d'intrusions et d'attaques de DoS dans un environnement véhiculaire. L'approche proposée doit permettre, grâce à l'extraction de différentes caractéristiques (nombre de paquets transmis, nombre de paquets jetés), de déterminer les attaques externes (au niveau d'autres véhicules) et internes (à l'intérieur du véhicule) avec une probabilité élevée et un taux de fausses alarmes limité. Néanmoins, des améliorations de la solution proposée (extension à d'autres attaques, évaluation, utilisation des ressources) sont encore possibles. Enfin, dans (Lyamin *et al.* 2018), les auteurs s'intéressent à la détection d'attaque par brouillage radio au sein de convois de véhicules (*platooning*). En effet, la formation et le bon fonctionnement de ces convois reposent essentiellement sur l'échange d'informations entre véhicules et une perturbation de ces communications, simple à mettre en place, pourrait représenter un réel danger. Ainsi, les auteurs comparent premièrement les méthodes de référence existantes, basées sur des modèles ou des données. Dans un second temps, ils proposent une nouvelle

méthode hybride basée sur l'analyse statistique du trafic réseau et sur des méthodes d'exploration de données. Cette nouvelle approche intelligente doit permettre d'assurer des performances acceptables, même dans le cas d'une gigue variable durant l'émission des messages d'informations. Dans ce même papier est proposée une discussion intéressante autour de l'application de l'apprentissage profond aux réseaux véhiculaires et des limites actuelles de ces approches, notamment la quantité des données nécessaires à l'apprentissage.

Enfin, lors du calcul de chemins de communication, afin d'éviter la perturbation du réseau, il semble important de prendre en compte les nœuds pouvant être malveillants ou infectés, afin de les éliminer du processus de routage. Ainsi, les auteurs de (Krundyshev *et al.* 2018) s'intéressent aux nœuds malicieux transmettant de fausses informations de routage et indiquant qu'ils offrent le chemin de communication le plus court jusqu'au nœud destinataire, sans que ce soit le cas. Les attaques de routage dans les réseaux véhiculaires sont nombreuses : DoS, trou noir, *wormhole*, *Sybil*, etc. La solution proposée doit permettre d'éviter ces différentes attaques en analysant le comportement des nœuds. Pour cela, elle est basée sur un modèle de confiance et sur un algorithme d'intelligence distribuée (algorithme des gouttes d'eau intelligentes). Grâce à une définition d'un indice de confiance pour chacun des véhicules, basé sur l'avis d'autres véhicules, et grâce à une transmission rapide de cette information, la solution proposée permet de maintenir un délai de transmission faible, ainsi qu'un débit et un pourcentage de paquets reçus élevés, pour différents types d'attaques (*wormhole* et trou noir). De la même manière, les auteurs de (Zhang *et al.* 2018b) cherchent à définir une approche permettant d'offrir des taux de délivrance de paquets élevés et des débits faibles dans des situations d'attaques, sans entraîner un surcoût de communication important. Cette méthode est basée sur l'algorithme de routage adaptatif *AntHocNet*, auquel a été appliqué un système de détection avec logique floue pour exclure un processus de décision et de routage, les nœuds pouvant s'avérer malveillants. Ainsi, afin de déterminer quels sont les nœuds malveillants, chacun des véhicules analyse le comportement des véhicules environnants en fonction de différents paramètres, notamment le taux de paquets transmis (reçus/envoyés) sur une certaine période de temps. Grâce à cela, cette méthode permet d'améliorer les performances de l'algorithme *AntHocNet* dans le cas d'attaques de type trou noir. Néanmoins, la possibilité de faire passer à l'échelle cette solution et l'analyse de ses performances pour d'autres types d'attaques restent encore à étudier.

8.5. Enjeux

L'architecture proposée ici, basée sur la définition d'un plan de connaissance au service de l'ensemble des autres plans (contrôle, gestion, sécurité et vie privée) doit

apporter de nombreux bénéfices : détection d'intrusions, équilibrage de charge, contrôle de congestion, prédiction du mouvement des véhicules, etc. Néanmoins, pour que l'implémentation de l'architecture proposée soit envisageable, de nombreux verrous restent à lever, notamment la sécurité et la protection de la vie privée, la complexité des méthodes de calcul, l'apprentissage distribué et les mouvements de flux de véhicules.

8.5.1. *Sécurité et vie privée*

Comme cela est indiqué dans (Liang *et al*. 2019), bien que des approches intelligentes puissent apporter de nombreux bénéfices au plan de sécurité (détection d'intrusion, routage sécurisé, etc.), l'intégration de techniques d'IA au sein de l'architecture véhiculaire pourrait soulever d'importantes questions de sécurité et de confiance. En effet, l'entraînement de ces algorithmes étant basé sur l'analyse de données existantes ainsi que sur l'intégration de données erronées ou manipulées par une entité malveillante, pourrait modifier le comportement attendu du système intelligent. Ainsi, dans un environnement aussi sensible que l'environnement véhiculaire, l'étude de la fiabilité et de la robustesse de ces techniques d'IA semble essentielle. De plus, le développement de ces approches nécessite le partage des données des utilisateurs et un traitement externalisé en bordure de réseau ou dans le Cloud. Le développement de solutions assurant la protection de la vie privée des utilisateurs (anonymat, traitement au plus près, etc.) est également une question primordiale.

8.5.2. *Apprentissage distribué*

Les techniques d'apprentissage distribué (*swarm*) fonctionnent grâce à la collecte d'informations par chacun des nœuds du réseau et par l'échange d'informations entre ces nœuds. En effet, pour que chacun des nœuds puisse disposer d'une vision globale, ou partiellement globale, la coopération est un point primordial. Or, dans l'environnement véhiculaire, la transmission d'informations au service de cette intelligence pourrait nécessiter des ressources de communication importantes et non nécessairement disponibles. En effet, les approches d'apprentissage distribué destinées aux communications véhiculaires telles que (Toutouh et Alba 2016, 2018) ou (Nyabuga *et al.* 2016) ne prennent pas en compte les limitations liées à l'environnement véhiculaire. Aussi, pour que ces méthodes soient applicables à l'environnement véhiculaire, l'idée de systèmes de coopération prenant en compte les contraintes physiques des équipements et les contraintes de l'environnement (délais, canaux de communication, bande passante, puissance de transmission, etc.) devra être abordée.

8.5.3. *Complexité des méthodes de calcul*

La question de la complexité de calcul est également une problématique importante. En effet, les solutions basées sur l'apprentissage profond développées hors de l'environnement véhiculaire nécessitent des capacités de calcul et de stockage importantes pour l'apprentissage et l'entraînement des algorithmes. Ceci doit permettre d'offrir des résultats aux performances toujours plus élevées. Toutefois, les véhicules ont des capacités de calcul et de stockage limitées et ces solutions ne pourront pas y être déployées. C'est pourquoi les auteurs de (Zeng *et al.* 2018) proposent une solution de détection d'intrusions adaptée aux capacités des véhicules. La prise en compte de ces capacités et le développement de solutions moins consommatrices de puissance de calcul sont essentiels. En effet, les communications véhiculaires nécessitant une latence faible, de nombreuses prises de décisions et de nombreux calculs devront être réalisés au niveau des véhicules, sans aide extérieure. Ainsi, pour s'adapter à ces ressources, des méthodes de calcul avec une complexité réduite, mais permettant de garantir des niveaux de performance élevés, devront être conçues.

8.5.4. *Mouvement des flux de véhicules*

Bien que différentes solutions pour la prédiction du mouvement des flux de véhicules soient introduites dans (Huang *et al*. 2014 ; Ide *et al*. 2015 ; Lv *et al*. 2015), aucune d'entre elles n'a pour l'instant été testée dans un environnement réel. De plus, de nombreuses améliorations pourraient être considérées, afin d'en augmenter les performances (Huang *et al*. 2014). Or, qu'il s'agisse de sécuriser, contrôler ou gérer les communications, la prédiction du mouvement des flux de véhicules constitue la base de l'ensemble de ces travaux. Aussi, le développement d'une solution fiable, normalisée et offrant des performances optimales, notamment en terme de latence, est un point primordial nécessaire à l'ensemble de l'architecture véhiculaire. De plus, l'analyse du comportement des conducteurs et des utilisateurs des véhicules pourrait permettre de développer des solutions d'une précision encore plus importante. Il s'agit, par conséquent, d'un domaine ouvert pour lequel l'évaluation de propositions dans un environnement réel est primordiale.

8.6. Conclusion

L'architecture de référence des systèmes de transport intelligents coopératifs (C-ITS) présente, en l'état actuel, différentes limitations : capacité à traiter un volume important de données, interopérabilité, garanties de qualité de service (QoS), etc. C'est pourquoi les réseaux véhiculaires ont évolué vers un nouveau paradigme :

l'Internet des véhicules (IdV). L'IdV devrait permettre de répondre aux limites des réseaux véhiculaires et de développer de nouveaux services : assistance au conducteur, sécurité routière, santé, etc.

L'évolution de l'architecture véhiculaire et l'avènement de l'IdV se basent sur l'intégration de nouvelles technologies au sein de l'architecture véhiculaire. L'intelligence artificielle fait partie de ces technologies. En effet, celle-ci pourrait offrir de nombreux avantages, notamment au travers de la prise de décisions intelligentes et de l'automatisation de nombreux services. C'est pourquoi de nombreux travaux ont déjà proposé la définition d'architectures de communication véhiculaires intégrant des techniques d'IA. Néanmoins, ces architectures se concentrent essentiellement sur une évolution du plan de contrôle de l'architecture de communication véhiculaire et non sur une évolution des autres plans, à savoir les plans de gestion et de sécurité pour lesquels l'intégration de techniques d'IA pourrait également présenter de nombreux bénéfices.

C'est pourquoi, dans ce chapitre, nous proposons d'ajouter un quatrième plan, le plan de connaissance, au service de l'ensemble des plans. Ce plan pourrait également permettre une prise de décision multi niveaux, un traitement plus efficace des informations et une amélioration de nombreux services. Dans la section 8.4 de ce chapitre, ce plan de connaissance est présenté, ainsi que ses bénéfices pour chacun des plans : gestion (équilibrage de charge, gestion des ressources distribuées, etc.), contrôle (mouvement, contrôle de congestion, routage dynamique) et sécurité (détection d'intrusions, déni de service, routage sécurisé). Néanmoins, pour que l'IA s'intègre dans l'architecture véhiculaire, différents points devront être considérés, notamment la sécurité et la protection de la vie privée, la complexité des méthodes de calcul, l'apprentissage distribué et la prédiction du mouvement des véhicules.

8.7. Bibliographie

Alheeti, K.M.A., Gruebler, A., McDonald-Maier, K.D. (2015). On the detection of grey hole and rushing attacks in self-driving vehicular networks. Dans *7th Computer science and Electronic Engineering Conference (CEEC)*, 231–236.

Alheeti, K.M.A., Gruebler, A., McDonald-Maier, K.D., Fernando, A. (2016). Prediction of DoS attacks in external communication for self-driving vehicles using a fuzzy petri net model. Dans *IEEE International Conference on Consumer Electronics (ICCE)*, 502–503.

Aliyu, A., Abdullah, A.H., Kaiwartya, O., Cao, Y., Usman, M.J., Kumar, S., Lobiyal, D.K., Raw, R.S. (2017). Cloud computing in VANETs: architecture, taxonomy, and challenges. *IETE Technical Review*, 35(5), 523–547.

Ashraf, M.I., Bennis, M., Perfecto, C., Saad, W. (2016). Dynamic proximity-aware resource allocation in vehicle-to-vehicle (V2V) communications. *IEEE Globecom Workshops*, 1–6.

Borcoci, E., Obreja, S., Vochin, M. (2017). Internet of Vehicles Functional Architectures-Comparative Critical Study. Dans *The Ninth International Conference on Advances in Future Internet (AFIN)*, 10–14.

Chen, M., Tian, Y., Fortino, G., Zhang, J., Humar, I. (2018). Cognitive internet of vehicles. *Computer Communications*, 120, 58–70.

Clark, D.D., Partridge, C, Ramming, C., Wroclawski, J.T. (2003). A Knowledge plane for Internet. Dans *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 3–10.

Contreras-Castillo, J., Zeadally, S., Guerrero Ibáñez, J.A. (2017). A seven-layered model architecture for Internet of Vehicles. *Journal of Information and Telecommunication*, 1(1), 4–22.

ETSI (2010). Intelligent Transport Systems (ITS), Communications Architecture. Document, European Telecommunications Standard Institute, Sophia Antipolis.

Huang, W., Song, G., Hong, H., Xie, K. (2014). Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems*, 15(5), 2191–2201.

Ide, C., Hadiji, F., Habel, L., Molina, A., Zaksek, T., Schreckenberg, M., Kersting, K., Wietfeld, C. (2015). LTE connectivity and vehicular traffic prediction based on machine learning approaches. Dans *IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 1–5.

Jiacheng, C.H.E.N., Haibo, Z.H.O.U., Ning, Z., Peng, Y., Lin, G., Xuemin, S. (2016). Software defined Internet of vehicles: architecture, challenges and solutions. *Journal Of Communications And Information Networks*, 1(1), 14–26.

Kaiwartya, O., Abdullah, A.H., Cao, Y., Altameem, A., Prasad, M., Lin, C.T., Liu, X. (2016). Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects. *IEEE Access*, 4, 5356–5373.

Kang, M.J., Kang, J.W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. *PLOS ONE*, 11(6).

Krief, F. (dir.) (2010). *Communicating embedded Networks: Network Applications*. ISTE Ltd, Londres et Wiley, New York.

Krundyshev, V., Kalinin, M., Zegzhda, P. (2018). Artificial swarm algorithm for VANET protection against routing attacks. *IEEE Industrial Cyber-Physical Systems (ICPS)*, 795–800.

Lai, W.K., Lin, M.T., Yang, Y.H. (2015). A machine learning system for routing decision-making in urban vehicular ad hoc networks. *International Journal of Distributed Sensor Networks*, 11(3), 374391.

Li, Z., Wang, C., Jiang, C.J. (2017). User association for load balancing in vehicular networks: An online reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 18(8), 2217–2228.

Li, J., Luo, G., Cheng, N., Yuan, Q., Wu, Z., Gao, S., Liu, Z. (2019). An end-to-end load balancer based on deep learning for vehicular network traffic control. *IEEE Internet of Things Journal*, 6(1), 953–966.

Liang, L., Ye, H., Li, G.Y. (2019). Toward Intelligent Vehicular Networks: A Machine Learning Framework. *IEEE Internet of Things Journal*, 6(1), 124–135.

Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y. (2015). Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865–873.

Lyamin, N., Kleyko, D., Delooz, Q., Vinel, A. (2018). AI-Based Malicious Network Traffic Detection in VANETs. *IEEE Network*, 32(6), 15–21.

Macedo, D.F., Dos Santos, A.L., Nogueira, J.M.S., Pujolle, G. (2008). A knowledge plane for autonomic context-aware wireless mobile ad hoc networks. Dans *IFIP/IEEE International Conference on Management of Multimedia Networks and Services (MMNS)*, 1–13.

Mendiboure, L., Chalouf, M.A., Krief, F. (2019a). Towards a 5G vehicular architecture. Dans *Communication Technologies for Vehicles – Proceedings of the 14th International Workshop*. Nets4Workshops, Colmar.

Mendiboure, L., Chalouf, M.A., Krief, F. (2019b). Edge Computing Based Application in Vehicular Environments: Comparative Study and Main Issues. *Journal of Computer Science and Technology*, 34(4), 869–886.

Movahedi, Z., Ayari, M., Langar, R., Pujolle, G. (2012). A Survey of Autonomic Network Architectures and Evaluation Criteria. *IEEE Communications Surveys & Tutorials*, 14(2).

Nyabuga, S.M., Cheruiyot, W., Kimwele, M. (2016). Using particle swarm optimization (PSO) algorithm to protect vehicular ad hoc networks (VANETS) from denial of service (DoS) attack. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 5(3).

Sun, Y., Song, J., Zhou, S., Guo, X., Niu, Z. (2018). Task replication for vehicular edge computing: A combinatorial multi-armed bandit based approach. Dans *2018 IEEE Global Communications Conference (GLOBECOM)*, 1–7.

Taherkhani, N., Pierre, S. (2016). Centralized and localized data congestion control strategy for vehicular ad hoc networks using a machine learning clustering algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 17(11), 3275–3285.

Toutouh, J., Alba, E. (2016). Distributed fair rate congestion control for vehicular networks. Dans *Distributed Computing and Artificial Intelligence, 13th International Conference (DCAI)*, 433–442.

Toutouh, J., Alba, E. (2018). A swarm algorithm for collaborative traffic in vehicular networks. *Vehicular Communications*, 12, 127–137.

Xu, Y., Li, L., Soong, B.H., Li, C. (2014). Fuzzy Q-learning based vertical handoff control for vehicular heterogeneous wireless network. Dans *2014 IEEE International Conference on Communications (ICC)*, 5653–5658.

Yao, L., Wang, J., Wang, X., Chen, A., Wang, Y. (2018). V2X routing in a VANET based on the hidden Markov model. *IEEE Transactions on Intelligent Transportation Systems*, 19(3), 889–899.

Ye, H., Liang, L., Li, G.Y., Kim, J., Lu, L., Wu, M. (2018). Machine learning for vehicular networks: Recent advances and application examples. *IEEE Vehicular Technology Magazine*, 13(2), 94–101.

Ye, H., Li, Y.G., Juang, B.H.F. (2019). Deep reinforcement learning for resource allocation in V2V communications. *IEEE Transactions on Vehicular Technology*.

Zeng, Y., Qiu, M., Ming, Z., Liu, M. (2018). Senior2Local: A Machine Learning Based Intrusion Detection Method for VANETs. Dans *International Conference on Smart Computing and Communication*, 417–426.

Zhang, G., Wu, M., Duan, W., Huang, X. (2018a). Genetic Algorithm Based QoS Perception Routing Protocol for VANETs. *Wireless Communications and Mobile Computing*.

Zhang, H., Bochem, A., Sun, X., Hogrefe, D. (2018b). A security aware fuzzy enhanced reliable ant colony optimization routing in vehicular ad hoc networks. Dans *2018 IEEE Intelligent Vehicles Symposium (IV)*, 1071–1078.

Zheng, Q., Zheng, K., Zhang, H., Leung, V.C. (2016). Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning. *IEEE Transactions on Vehicular Technology*, 65(10), 7857–7867.

PARTIE 5

Les communications de radio intelligente

9

Application de l'intelligence artificielle dans les réseaux de radio cognitive

Badr BENMAMMAR et Asma AMRAOUI
Université Abou Bekr Belkaid, Tlemcen, Algérie

9.1. Introduction

Actuellement et dans le domaine des réseaux sans-fil, les ressources en termes de disponibilité de fréquences et de bande passante sont de plus en plus rares. Par conséquent, de nouvelles solutions ont été nécessaires pour minimiser la consommation d'énergie et optimiser l'allocation des ressources radio.

Pour un accès flexible au spectre, Joseph Mitola III et Gerald Q. Maguire ont introduit la radio cognitive (RC) en se basant sur la radio logicielle en 1999 (Mitola et Maguire 1999). La radio logicielle est une radio qui a su réaliser sous forme logicielle les fonctions typiques de l'interface radio généralement réalisées sous forme matérielle, telles que la fréquence porteuse, la largeur de bande du signal et la modulation. En effet, les auteurs dans (Mitola et Maguire 1999) ont combiné leurs expériences de la radio logicielle ainsi que leurs passions pour l'apprentissage automatique et l'intelligence artificielle (IA) pour mettre en place la technologie de la radio cognitive. D'après Joseph Mitola III, une radio cognitive peut connaître, percevoir et apprendre de son environnement, puis agir pour simplifier la vie de l'utilisateur (Mitola 2000). En 2005, Simon Haykin avait passé en revue le concept de la radio cognitive et l'avait traité comme une communication sans fil dotée d'un pouvoir cérébral (Haykin 2005).

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

La RC est une technologie qui détecte l'environnement, analyse ses paramètres de transmission, puis prend des décisions en matière d'allocation et de gestion de ressource. Toutefois, les formulations d'optimisation pour l'allocation des ressources offrent des solutions optimales parfois au détriment du temps de calcul et de la complexité du traitement. Pour réduire cette complexité et obtenir une allocation de ressources en temps raisonnable, les réseaux radio cognitifs (RRC) doivent être dotés de capacités d'apprentissage et de raisonnement. Le moteur cognitif doit coordonner les actions de la RC en utilisant des techniques d'apprentissage automatique.

Par conséquent, une RC doit être intelligente et capable d'apprendre de son expérience en interagissant avec son environnement radiofréquence. Par conséquent, l'apprentissage est un élément indispensable de la RC, qui peut être fourni en utilisant l'intelligence artificielle et les techniques d'apprentissage automatique. En effet, l'application de l'intelligence artificielle, et en particulier l'apprentissage automatique dans les RRC, a récemment suscité un vif intérêt dans la littérature.

L'apprentissage vise à faire en sorte que les machines effectuent des tâches similaires à celles d'un expert. La machine intelligente percevra son environnement et prendra des mesures pour maximiser sa propre utilité. L'intelligence artificielle se focalise sur la déduction, le raisonnement, la résolution de problèmes, la représentation des connaissances et l'apprentissage (Woods 1986).

Le processus d'apprentissage dans les RRC est illustré dans la figure 9.1 et peut être présenté comme suit (Abbas *et al*. 2015) :

– détection des paramètres de radiofréquence tels que la qualité du canal ;

– observation de l'environnement et analyse de ses réactions ;

– apprentissage ;

– conservation des décisions et des observations pour la mise à jour du modèle ;

– décision sur les problèmes de gestion des ressources et l'ajustement des erreurs de transmission en conséquence (Bkassiny *et al*. 2013 ; Russell 2016).

Dans la littérature, il y a quelques travaux qui se sont intéressés à l'application de l'IA dans les RRC. Dans (Zhao et Morales-Tirado 2012), les auteurs ont introduit l'utilisation des techniques de l'intelligence artificielle et de l'apprentissage automatique dans la RC. Ils ont présenté aussi les applications possibles et les idées fondamentales autours de la RC. Dans (Bkassiny *et al*. 2013), les auteurs ont présenté un *survey* sur les différentes techniques de l'intelligence artificielle, telles que l'apprentissage par renforcement, la théorie des jeux, les réseaux de neurones, les machines à vecteurs de support et le modèle de Markov. Le *survey* discute les points forts et les points faibles de ces techniques, ainsi que les difficultés rencontrées dans

leurs applications dans le domaine de la RC. Dans (Gavrilovska *et al*. 2013), les auteurs ont étudié la théorie des jeux, l'apprentissage par renforcement et des approches de raisonnement telles que les réseaux bayésiens, la logique floue et le raisonnement à partir de cas dans les RRC. En 2015, le *survey* présenté dans (Abbas *et al*. 2015) a été consacré à la logique floue, aux algorithmes génétiques, aux réseaux de neurones, à la théorie des jeux, à l'apprentissage par renforcement, aux machines à vecteurs de support, au raisonnement à partir de cas, aux arbres de décision, aux réseaux bayésiens, aux modèles de Markov, aux systèmes multiagents et aux algorithmes de colonies d'abeilles artificielles.

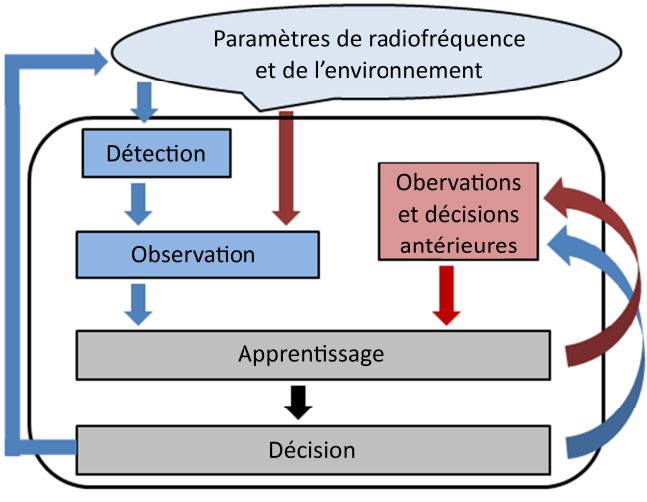


Figure 9.1. *Processus d'apprentissage dans les réseaux de radio cognitive (Abbas et al. 2015)*

Toutefois, nous avons remarqué que le papier (Abbas *et al*. 2015) n'a pas discuté l'application de l'optimisation par essaim de particules, une métaheuristique très utilisée dans la RC. D'autres métaheuristiques plus récentes et également utilisées dans la RC ont aussi été ignorées dans (Abbas *et al*. 2015), comme l'algorithme des lucioles, la recherche coucou et l'algorithme de recherche gravitationnelle.

Dans ce chapitre, nous nous intéressons aux techniques de l'intelligence artificielle qui ont été les plus utilisées dans les trois dernières années dans la RC (l'état de l'art des travaux ultérieurs à 2015 a été déjà réalisé dans (Abbas *et al*. 2015)).

Nous nous intéressons à des métaheuristiques qui n'étaient pas discutées dans les précédents *survey*, comme l'algorithme des lucioles, la recherche coucou, l'algorithme

de recherche gravitationnelle et l'optimisation par essaim de particules. Nous présentons également les travaux récents liés à l'application des autres techniques de l'intelligence artificielle dans la RC, à savoir : les algorithmes génétiques, les algorithmes de colonies d'abeilles, la logique floue, la théorie des jeux, les réseaux de neurones, les modèles de Markov, les machines à vecteurs de support, le raisonnement à partir de cas, les arbres de décision, les réseaux bayésiens, les systèmes multiagents et l'apprentissage par renforcement. En tout, dix-sept techniques d'IA seront étudiées dans ce chapitre.

Les points fondamentaux traités dans ce chapitre portent sur :

– la présentation d'une étude complète sur les techniques de l'IA, leurs définitions et leurs applications dans la RC ; à noter que quelques techniques n'ont jamais été discutées auparavant dans ce domaine ;

– la présentation des tâches principales de la RC et leurs défis correspondants ;

– la catégorisation des techniques présentées selon le type d'apprentissage (supervisé ou non supervisé) et leurs applications en fonction des tâches de la RC.

Le reste de ce chapitre est organisé comme suit : la section 9.2 présente le cycle de cognition, les principales tâches de la RC et leurs défis correspondants. La section 9.3 offre un état de l'art sur l'application des méthodes de l'intelligence artificielle dans la RC. La section 9.4 propose une catégorisation des techniques présentées selon le type d'apprentissage (supervisé ou non supervisé) et présente leurs applications en fonction des tâches de la RC. Finalement, la section 9.5 est consacrée à une conclusion.

9.2. La radio cognitive

9.2.1. *Cycle de cognition*

Un RRC suit le cycle cognitif pour optimiser ses performances (voir figure 9.1). Il commence par le *sensing* de l'environnement, puis continue par l'analyse des paramètres extérieurs, puis finit par prendre des décisions relatives à l'allocation et à la gestion dynamiques des ressources, afin d'améliorer l'efficacité spectrale (Biglieri *et al*. 2013).

Sensing de l'environnement : dans la RC, le réseau secondaire peut utiliser le spectre disponible, mais sans causer d'interférence au réseau primaire. Par conséquent, le réseau secondaire doit auparavant détecter les paramètres de son environnement, tels que la disponibilité des trous de spectre dans la fréquence.

Analyse des paramètres de l'environnement : les paramètres de l'environnement détectés seront utilisés comme entrées pour la gestion des ressources. Cette dernière peut inclure la minimisation de la consommation d'énergie, la minimisation des interférences, l'optimisation du débit, l'amélioration de la QoS et la maximisation de l'efficacité spectrale (Wyglinski *et al*. 2009).

Prise de décision : dans la RC, la prise de décision peut être basée sur des algorithmes d'optimisation. Cependant, et afin de réduire la complexité et d'obtenir une allocation de ressources en temps raisonnable, les RRC utilisent l'apprentissage automatique et l'intelligence artificielle (Qiu *et al*. 2012).

9.2.2. *Tâches de la radio cognitive et défis correspondants*

La RC se base généralement sur deux tâches principales :

– la tâche « cognitive », qui peut être obtenue en utilisant des techniques de détection de spectre. Les principaux défis auxquels se heurtent ces techniques sont la précision de la décision concernant la disponibilité du spectre, la durée de détection, la fréquence de détection, l'incertitude de la puissance du bruit ambiant, en particulier à faible SNR (*Signal-to-Noise Ratio*) dû à l'atténuation multichemins et au *shadowing*. Pour améliorer les performances de détection du spectre, des technologies coopératives de détection et de géolocalisation ont été proposées dans la littérature (Ghasemi et Sousa 2008 ; Wang et Liu 2011 ; Umar et Sheikh 2012) ;

– la tâche « reconfigurable », qui est utilisée pour ajuster de manière dynamique les paramètres de transmission afin d'améliorer les performances du réseau. Elle est basée sur la prise de décision, basée elle-même sur des algorithmes d'optimisation. Le principal défi de cette tâche concerne la complexité et la convergence de ces techniques dans un temps limité. Une chose qui peut être résolue en utilisant les techniques de l'intelligence artificielle et de l'apprentissage automatique, afin de construire des modèles d'apprentissage pour la prise de décision. Par conséquent, le choix d'une technique d'apprentissage pour effectuer une tâche spécifique de la RC est considéré en lui-même comme un défi.

9.3. Application de l'intelligence artificielle dans la radio cognitive

9.3.1. *Les métaheuristiques*

Dans la littérature, les métaheuristiques sont classées en deux sous-catégories : celles à solution unique et celles à base de population. Les métaheuristiques à solution unique sont des approches itératives qui débutent avec une solution initiale

unique et l'améliorent d'une itération à une autre en exploitant son voisinage. Les métaheuristiques à base de population explorent l'espace de recherche en utilisant un ensemble de solutions appelées « population ». Cette dernière catégorie est aussi classée en deux sous catégories : les algorithmes évolutionnaires et les algorithmes basés sur l'intelligence en essaim.

Les algorithmes évolutionnaires s'inspirent de la manière dont les espèces évoluent dans la nature, et plus précisément du principe de la sélection naturelle énoncée dans la « théorie de l'évolution », développée par Charles Darwin (Darwin 2009) dans son ouvrage intitulé *The Origin of Species : By Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. Les algorithmes basés sur l'intelligence en essaim ont été introduits pour la première fois par Beni et Wang (Gerardo et Wang 1993) dans leur article intitulé « Swarm Intelligence in Cellular Robotics Systems », qui décrit le comportement d'un groupe de robots coopérant pour accomplir une tâche ou résoudre un problème.

Dans ce qui suit, nous présentons l'application des six métaheuristiques les plus utilisées dans les RRC, à savoir : l'algorithme des lucioles, la recherche coucou, les algorithmes de colonies d'abeilles, les algorithmes génétiques, l'algorithme de recherche gravitationnelle et l'optimisation par essaim de particules.

9.3.1.1. *L'algorithme des lucioles*

L'algorithme des lucioles (FA pour *Firefly Algorithm*) est une approche d'optimisation basée sur l'intelligence en essaim qui a été proposé par Xin-She Yang en 2008 (Surafel et Medard T Ngnotchouye 2017). Son principe est inspiré du comportement lumineux des lucioles (*fireflies*). En effet, chaque solution potentielle est assimilée à une luciole, dont la luminosité est proportionnellement liée à sa qualité (qualité de la solution). Les travaux suivants se sont intéressés à l'application de l'algorithme des lucioles dans les RRC.

L'algorithme des lucioles a été adapté par les auteurs dans un RRC basé sur l'*Orthogonal Frequency-Division Multiplexing* (OFDM) (Saoucha et Benmammar 2017). L'optimisation multi-objectifs a été également utilisée afin d'optimiser la qualité de la communication des utilisateurs secondaires (SU). Les performances de l'algorithme des lucioles ont été validées à travers une comparaison avec l'algorithme d'optimisation par essaim de particules (PSO) et avec l'entropie croisée en termes de vitesse de convergence, de qualité de la solution et de stabilité.

L'algorithme des lucioles a été utilisé dans (Tounsi et Babes 2017) pour résoudre le problème du contrôle de la puissance et d'allocation de canaux dans les RRC. Une version

modifiée de l'algorithme des lucioles, utilisant le nouveau facteur d'attractivité, est proposée pour résoudre ce problème. Une analyse théorique est présentée dans ce papier pour prouver l'efficacité et l'existence de l'équilibre de Nash concernant la stratégie proposée. Les résultats présents dans ce papier montrent que la méthode proposée surpasse les approches existantes dans la littérature en termes de vitesse de convergence.

Pour estimer les amplitudes des sous-porteuses d'annulation, les auteurs dans (Elahi *et al*. 2017) ont proposé deux algorithmes de recherche : les algorithmes génétiques et l'algorithme des lucioles. Les résultats de la simulation montrent que les algorithmes proposés permettent une meilleure réduction des lobes secondaires par rapport aux techniques existantes dans la littérature.

Dans (Ghanem *et al*. 2016), le problème de l'attaque par émulation de l'utilisateur primaire (PUE pour *Primary User Emulation*) est résolu à l'aide d'un modèle de défense de localisation basé sur l'utilisation de l'algorithme des lucioles. Les utilisateurs de la RC coopèrent pour détecter et localiser l'attaquant en comparant son emplacement avec la position de l'utilisateur primaire (PU). Les résultats de la simulation sont comparés avec les méthodes existantes et montrent que l'algorithme des lucioles réduit l'erreur de localisation et nécessite moins d'utilisateurs secondaires à la coopération.

9.3.1.2. *La recherche coucou*

La recherche coucou (CS pour *Cucko Search*) est une approche d'optimisation basée sur l'intelligence en essaim, qui a été proposée par Yang et Deb en 2009 (Yang et Deb 2017). Elle a été développée en s'inspirant du comportement parasitaire de certaines espèces d'oiseaux. Dans le monde réel, les coucous pondent leurs œufs dans les nids des autres espèces d'oiseaux. Dans la plupart des cas, l'oiseau hôte croit que les œufs déposés sont les siens et, par conséquent, il en prend soin. Cependant, dans certains cas, les œufs déposés sont découverts et sont ainsi jetés par l'oiseau hôte, ou bien le nid est abandonné par ce dernier.

Les travaux suivants se sont intéressés à l'application de la recherche coucou dans les RRC.

Une nouvelle méthode pour l'estimation de l'état du canal dans un RRC basée sur l'OFDM a été proposée dans (Manjith 2016). La méthode est une hybridation entre l'optimisation par recherche bactérienne (BFO pour *Bacterial Foraging Optimization*) et un algorithme modifié de la recherche coucou.

Dans (Kaur *et al.* 2018a), les auteurs ont proposé un nouveau système de radio cognitive multi-utilisateur, ainsi que son optimisation à l'aide de l'algorithme de

recherche coucou. Les paramètres de transmission de plusieurs utilisateurs secondaires sont considérés sous la norme IEEE 802.22 WRAN. Les résultats d'optimisation ont été comparés à une autre technique d'optimisation efficace basée sur la biogéographie et au recuit simulé.

9.3.1.3. *Algorithme de colonie d'abeilles*

Différents groupes de chercheurs ont participé au développement des algorithmes d'abeilles de façon indépendante au cours de ces dix dernières années. Craig A. Tovey, à Georgia Tech, en collaboration avec Sunil Nakrani, à l'Université d'Oxford, ont proposé pour la première fois en 2004 l'algorithme Honey Bee (Nakrani et Tovey 2004). L'algorithme Virtual Beea été créé en 2005 par Xin-She Yang à l'université de Cambridge pour résoudre les problèmes d'optimisation numériques (Yang 2005). Haddad, Afshar et leurs collègues ont présenté en 2005 l'algorithme Honey-bee mating optimization (Haddad *et al*. 2006).

L'algorithme des colonies d'abeilles artificielles (ABC pour *Artificial Bee Colony*) a été développé en 2005 par Karaboga pour l'optimisation des fonctions numériques (Karaboga 2005).

Les travaux suivants se sont intéressés à l'application des algorithmes de colonie d'abeilles dans les RRC.

Un nouvel algorithme de *handover* spectral basé sur l'algorithme des colonies d'abeilles artificielles dans un RRC a été proposé dans (Bayrakdar et Calhan 2018). Dans l'algorithme proposé par les auteurs, la caractéristique de disponibilité du spectre est observée sur la base des missions des abeilles, afin de minimiser le délai du *handover* spectral et de maximiser la probabilité de trouver un canal inactif. L'avantage principal de cet algorithme est que le délai du *handover* spectral des SU est considérablement réduit pour un nombre différent d'utilisateurs, sans réduire la probabilité de trouver un canal disponible.

Un algorithme hybride entre les colonies d'abeilles artificielles et les algorithmes génétiques a été proposé dans (Elghamrawy 2018) pour optimiser l'utilisation du spectre en détectant les attaques d'émulation de l'utilisateur primaire et en augmentant la probabilité de détection. L'algorithme proposé intègre les opérateurs génétiques avec l'algorithme des colonies d'abeilles artificielles, pour atteindre l'équilibre entre l'exploitation et l'exploration, afin de trouver la solution optimale. Les résultats des simulations montrent les performances prometteuses de l'algorithme proposé pour optimiser la détection du spectre, par rapport aux algorithmes de détection récents.

L'objectif principal des auteurs dans (Zaheer *et al*. 2016) est de minimiser les puissances de transmission et ainsi de réduire les interférences dans un RRC en utilisant l'algorithme de colonie d'abeilles artificielles.

9.3.1.4. *Les algorithmes génétiques*

Le principe des algorithmes génétiques (GA pour *Genetic Algorithm*) a été introduit par John Holland de l'université du Michigan aux États-Unis, dans les années 1960 (Holland 1992), et mis en avant par l'ouvrage de référence de David E. Goldberg (Goldberg 1989). Dans un algorithme génétique, une population est constituée d'un ensemble d'individus, dont chacun est identifié par un ensemble de gènes appelés « chromosomes ». La reproduction est réalisée en recombinant les chromosomes de deux individus primaires, donnant ainsi naissance à des individus enfants ayant une empreinte génétique héritée des parents. Cependant, le code génétique des enfants peut contenir des gènes inexistants chez les parents, modélisant ainsi le phénomène génétique de la mutation. Ce dernier permet des changements dans la morphologie des espèces, toujours dans le sens d'une meilleure adaptation au milieu naturel.

Les travaux suivants se sont intéressés à l'application des algorithmes génétiques dans les RRC.

La formulation du problème, le développement et l'utilisation d'un algorithme génétique pour l'attribution de canaux dans un RRC a été présenté dans (Elhachmi et Guennoun 2016). Cette approche offre un moyen efficace d'accéder au spectre disponible pour les utilisateurs primaires et secondaires. Par rapport aux méthodes existantes, les résultats de la simulation démontrent que l'algorithme proposé produit des résultats satisfaisants en termes d'interférences et de débit.

Dans (Jiao et Joe 2016), les auteurs ont considéré un nouveau modèle de RRC dans lequel les réseaux d'utilisateurs primaires sont constitués d'utilisateurs primaires hétérogènes. Les auteurs considèrent le problème d'allocation de ressources économe en énergie pour les utilisateurs RC possédant une zone de couverture dans laquelle des utilisateurs primaires hétérogènes fonctionnent simultanément *via* une technologie d'accès multiradio. Les auteurs ont proposé un schéma de recherche basé sur les algorithmes génétiques croisés à deux niveaux pour obtenir une solution optimale en termes de puissance et de débit. Les résultats des simulations montrent que l'algorithme proposé par les auteurs est stable et que sa convergence est plus rapide.

9.3.1.5. *L'algorithme de recherche gravitationnelle*

En 2009, Rashedi *et al*. (2009) ont développé l'algorithme de recherche gravitationnelle (GSA pour *Gravitational Search Algorithm*) qui est une

métaheuristique d'optimisation inspirée de la nature. GSA est basé sur la loi de gravité de Newton, qui décrit la gravitation comme une attraction entre des corps ayant une masse. Les masses des objets (solutions) sont proportionnelles à leurs valeurs de fonctions objectifs (coûts). À chaque itération, les masses s'attirent entre elles, par les forces de gravitation. La masse la plus lourde a la force d'attraction la plus grande. Par conséquent, les masses les plus lourdes, qui sont probablement près de l'optimum global, attirent les autres masses selon leurs distances. Chaque objet est déterminé par quatre spécifications : position ; inertie ; masse gravitationnelle active et masse gravitationnelle passive. La position correspond à une solution du problème ; l'inertie et les masses gravitationnelles sont déterminées en utilisant la fonction objectif.

Les travaux suivants se sont intéressés à l'application de l'algorithme de recherche gravitationnelle dans les RRC.

Les auteurs dans (Guo *et al*. 2018) présentent une nouvelle méthode pour résoudre le problème du gaspillage de spectre dans les RRC. Cette méthode est basée sur le problème de la coloration des graphes et sur l'algorithme de recherche gravitationnelle. Les auteurs ont comparé les performances de leur algorithme avec l'optimisation par essaim de particules (PSO) et avec les algorithmes génétiques.

Dans (Kaur *et al*. 2018b), un algorithme hybride de PSO et de recherche gravitationnelle est présenté pour optimiser un RRC. Un nouvel environnement de RC est proposé, permettant à plusieurs SU d'accéder au spectre alors que leurs canaux subissent une atténuation de type Nakagami-m. Les facteurs de transmission appartenant à plusieurs SU et reposant sur la norme IEEE 802.22 WRAN sont optimisés pour atteindre plusieurs objectifs liés à la QoS attendue en utilisant PSO, GSA et l'algorithme hybride PSOGSA. Des fonctions objectifs modifiées et influencées par l'atténuation sont établies pour la tâche d'optimisation. Les résultats de l'optimisation indiquent une performance améliorée de l'algorithme hybride par rapport aux deux autres techniques de base.

9.3.1.6. *L'optimisation par essaim de particules*

L'optimisation par essaim de particules (PSO pour *Particle Swarm Optimization*) est une métaheuristique d'optimisation dans la famille des algorithmes évolutionnaires. Russel Eberhart (ingénieur en électricité) et James Kennedy (sociopsychologue) l'ont proposée en 1995 (Kennedy et Eberhart 1995). La PSO trouve sa source dans les observations faites lors des simulations informatiques de vols d'oiseaux groupés et de bancs de poissons (Craig 1987 ; Heppner et Grenander 1990). En effet, la PSO s'inspire fortement de l'observation des relations grégaires d'oiseaux migrateurs qui, pour parcourir de « longues distances » (migration, quête de nourriture, parades

aériennes, etc.), doivent optimiser leurs déplacements en termes d'énergie dépensée, de temps (etc.), comme par exemple la formation en forme de V présentée dans la figure 9.2.

Figure 9.2. *Volée d'Anser en formation en V (Bestaoui 2015)*

Les travaux suivants se sont intéressés à l'application de l'optimisation par essaim de particules dans les RRC.

Une technique basée sur la PSO et sur l'indicateur d'intensité du signal reçu pour la détection de la position de l'utilisateur primaire (PU pour *Primary User*) et de l'attaquant par émulation de l'utilisateur primaire a été proposée dans (Fihri *et al*. 2018). Les auteurs visent à augmenter la précision de la détection et de réduire le risque de fausses alarmes.

L'effet des valeurs propres de la matrice de covariance des échantillons reçus sur la méthode d'estimation du SNR a été analysé dans (Manesh *et al*. 2017). Les auteurs ont proposé l'utilisation de PSO dans la technique d'estimation du SNR basée sur des valeurs propres afin d'optimiser ces paramètres. Les résultats de la méthode proposée sont comparés à ceux de la méthode d'estimation du SNR originale et les résultats valident l'amélioration obtenue par la technique proposée par rapport à la technique d'origine.

L'information sans-fil simultanée et le transfert de puissance multi-utilisateurs pour les RRC en se basant sur la PSO et sur la relaxation semi-finie a été étudiée dans (Tuan et Koo 2017). Un émetteur secondaire doté d'un réseau d'antennes fournit des informations et de l'énergie à plusieurs récepteurs secondaires à une seule antenne. Les auteurs ont démontré à travers les simulations que leur algorithme présente une convergence rapide et de meilleures performances par rapport aux autres systèmes existants.

Dans (Zhai et Wang 2017), les auteurs ont utilisé la PSO pour résoudre le paradigme du *crowdsourcing* consistant à attribuer aux utilisateurs de mobile la tâche de détection du spectre. Les résultats de la simulation montrent que l'algorithme proposé atteint des performances supérieures à celles des autres algorithmes.

Dans (Tang et Xin 2016), les auteurs ont utilisé la PSO pour étudier le compromis entre l'utilité et la consommation d'énergie dans un RRC basé sur l'OFDM. Compte tenu de la faible convergence de la PSO d'origine autour des *optima* locaux, une version améliorée combinant la théorie du chaos est proposée dans cette étude, afin d'aider la PSO à rechercher des solutions autour des meilleurs résultats globaux. Les auteurs, à travers les simulations, ont démontré que leur algorithme nécessite un nombre inférieur d'itérations et peut atteindre une efficacité énergétique supérieure à celle des autres algorithmes.

Les auteurs dans (Tuan et Koo 2016) ont proposé une méthode hybride basée sur la PSO et la recherche de force brute (BFS pour *Brute-Force Serach*). Cette méthode est utilisée pour maximiser le débit de l'utilisateur secondaire (SU pour *Secondary User*) dans un RRC en *full-duplex*, lorsqu'il dispose de deux antennes distinctes et d'une capacité d'autosupprimer ses interférences. Les simulations montrent que, pour certaines valeurs de paramètres, le système en question fournit un débit beaucoup plus élevé que les systèmes proposés précédemment.

Les auteurs dans (Alhammadi *et al*. 2016) ont discuté les trois mécanismes du *handover* spectral (proactif, réactif et hybride) utilisés pour réduire le délai du *handover*. Le papier contient une mise en œuvre de l'algorithme PSO pour minimiser le temps de service total du *handover* spectral à la valeur optimale. Les résultats numériques montrent que la PSO réduit considérablement le temps de service total par rapport aux autres systèmes de *handover* spectral.

9.3.2. *La logique floue*

Lotfi Zadeh, en 1965 et en se basant sur sa théorie mathématique des ensembles flous, a créé la logique floue (FL pour *Fuzzy Logic*), représentant une extension de la logique booléenne (Zadeh 1965). La logique floue rend possible la prise en compte des imprécisions et des incertitudes, car elle confère une flexibilité très appréciable aux raisonnements qui l'utilisent.

Le travail suivant s'est intéressé à l'application de la logique floue dans les RRC.

Les auteurs dans (Banerjee *et al*. 2017) ont proposé une nouvelle méthode de prise de décision basée sur la logique floue pour la sélection de relais, contrairement

à de nombreux travaux existants dans lesquels le rapport signal sur interférence plus bruit (SINR) est considéré comme le seul paramètre de sélection de relais. Pour trouver le meilleur relais, les auteurs ont mené une vaste étude de simulation. Les résultats de la simulation révèlent l'impact de différents paramètres sur la sélection du meilleur relais.

9.3.3. *La théorie des jeux*

La première discussion connue sur la théorie des jeux (GT pour *Game Theory*) a eu lieu dans une lettre écrite par James Waldegrave en 1713. La théorie des jeux est utilisée comme une technique de prise de décision, où plusieurs joueurs doivent faire des choix et, par conséquent, affecter les intérêts des autres joueurs. Chaque joueur décide de ses actions en fonction de l'historique des actions sélectionnées par les autres joueurs lors des tours précédents de la partie.

Dans les RRC, les nœuds sont les acteurs du jeu et les actions sont les paramètres de l'environnement radio, tels que la puissance d'émission et la sélection du canal. Ces actions seront entreprises par les nœuds sur la base d'observations représentées par des paramètres d'environnement, tels que la disponibilité du canal, la qualité du canal et les interférences. Par conséquent, chaque nœud tirera des leçons de ses actions passées, observera les actions des autres nœuds et modifiera ses actions en conséquence (Bellhouse 2017).

Les travaux suivants se sont intéressés à l'application de la théorie des jeux dans les RRC.

Une approche basée sur la théorie des jeux utilisant le jeu de Stackelberg pour sécuriser un réseau de capteurs de radio cognitif contre l'attaque de falsification de données de détection de spectre a été proposée dans (Abdalzaher *et al*. 2017) ; cette attaque a pour but de corrompre les décisions de spectre communiquées par les nœuds de capteurs au centre de fusion en imposant une puissance d'interférence. Les simulations indiquent l'amélioration des performances du modèle de protection proposé par rapport à deux mécanismes de défense de base, à savoir les mécanismes de défense aléatoires et ceux à protection égale avec rapport signal sur bruit statique.

Les problèmes liés à la sécurité de la couche physique et à l'efficacité énergétique grâce au contrôle de puissance et à la coopération de relais, où les protocoles *decode-and-forward* et *amplify-and-forward* sont considérés, ont été étudiés dans (Fang *et al*. 2017). Les auteurs ont proposé un modèle de jeu de Stackelberg à un leader et à un suiveur en présence de multiples écouteurs, dans lequel une stratégie optimale d'allocation de puissance et de tarification peut être déterminée afin de maximiser les

utilités des joueurs. Les simulations réalisées par les auteurs démontrent que le modèle de jeu proposé améliore l'efficacité énergétique du réseau et offre de meilleures performances contre les attaques d'écoute, par rapport aux systèmes d'équilibre de Nash, de *rand* et de transmission directe.

Dans (Roy *et al*. 2017), les auteurs ont utilisé la théorie des jeux pour étudier le conflit et la coopération entre deux niveaux d'utilisateurs secondaires (temps réel et non temps réel). Un modèle de jeu de vente aux enchères est proposé afin d'analyser le processus de prise de décision et d'allouer efficacement un canal inactif à une paire de SU (temps réel et non temps réel) appartenant à un groupe d'utilisateurs.

Un modèle de jeu de Stackelberg à deux niveaux, dans lequel les PU et les SU agissent respectivement en tant que leaders et suiveurs pour améliorer l'efficacité énergétique des nœuds dans un RRC à sauts multiples a été proposé dans (Shu *et al*. 2016). Les simulations effectuées ont montré la pertinence des propositions des auteurs.

9.3.4. *Les réseaux de neurones*

Les réseaux de neurones (NN pour *Neural Networks*) ont été introduits par Warren McCulloch et Walter Pitts en 1943 et s'inspiraient du système nerveux central. Semblable au réseau de neurones biologiques, le réseau de neurones artificiels sera formé de nœuds, également appelés « neurones » ou « éléments de traitement », qui sont connectés ensemble pour former un réseau.

Le réseau de neurones artificiels reçoit des informations de tous les neurones voisins et fournit une sortie en fonction de son poids et de ses fonctions d'activation. Les poids adaptatifs peuvent représenter les forces de connexion entre les neurones. Pour accomplir le processus d'apprentissage, les poids doivent être ajustés jusqu'à ce que la sortie du réseau soit approximativement égale à la sortie souhaitée.

Les réseaux de neurones artificiels ont été utilisés pour permettre à la RC d'apprendre de l'environnement et de prendre des décisions afin d'améliorer la QoS du système de communication (Haykin 2008 ; Rojas 2013).

Les travaux suivants se sont intéressés à l'application des réseaux de neurones dans les RRC.

Les auteurs dans (Supraja et Pitchai 2019) ont présenté un système hybride constitué d'un algorithme génétique, de PSO et d'un réseau de neurones de type *back-propagation* en tant que nouvel algorithme d'apprentissage supervisé permettant de prédire les profils de spectre dans les RRC.

Les auteurs dans (Zhang *et al*. 2017) ont utilisé le réseau de neurones convolutifs dans un système automatique pour reconnaître les formes d'ondes radio cognitives. Le système proposé peut identifier jusqu'à huit types de signaux. Les résultats de la classification ont été démontrés par les auteurs à travers des simulations.

Les auteurs dans (Liu *et al*. 2019) ont étudié le compromis entre l'efficacité énergétique et l'efficacité spectrale pour les PU et les SU dans un RRC. Un réseau de neurones *feed-forward* est conçu et un algorithme analogue de type *back-propagation* est développé pour apprendre les paramètres optimaux de l'algorithme proposé par les auteurs. Les simulations sont fournies pour confirmer l'efficacité de l'algorithme proposé.

9.3.5. *Les modèles de Markov*

Le modèle de Markov (MM pour *Markov Model*) est utilisé pour modéliser des processus aléatoires passant d'un état à un autre dans le temps. Le processus aléatoire est sans mémoire ou les états futurs dépendent uniquement de l'état présent (Norris 1998 ; Ching et Ng 2006). Dans les modèles de Markov, les états sont visibles pour l'observateur ; cependant, dans le modèle de Markov caché (HMM pour *Hidden Markov Model*) certains états sont masqués ou ne sont pas explicitement visibles (Fraser 2008).

Les travaux suivants se sont intéressés à l'application des modèles de Markov dans les RRC.

Pour résoudre les problèmes du contrôle distribué de la puissance dans un réseau de capteurs cognitifs sans-fil, un mécanisme de contrôle de puissance basé sur le modèle de Markov caché est proposé dans (Zhu *et al*. 2017) en fonction de la différence et de l'indépendance des résultats de *sensing* du canal parmi les utilisateurs du réseau. Les simulations indiquent que le mécanisme de contrôle de puissance basé sur le modèle HMM non seulement améliore l'efficacité énergétique, mais également respecte mieux le SINR cible par rapport aux autres méthodes.

Une méthode adaptative de détection d'énergie à double seuil basée sur le modèle de Markov a été proposée dans (Liu *et al*. 2017). Lors de l'utilisation de cette méthode, le modèle de Markov modifié tient compte de la caractéristique variable dans le temps de l'occupation du canal pour résoudre l'état du canal « en confusion ». Les simulations montrent que la méthode proposée donne de meilleures performances en termes de détection par rapport à d'autres méthodes existantes.

Une méthode pour construire une carte d'environnement radio (REM pour *Radio Environment Map*) dans un environnement avec plusieurs PU a été proposée dans

(Ichikawa et Fujii 2017). Le REM fournit des informations statistiques sur l'activité du PU à chaque endroit. Il permet également au SU d'accéder à la bande sous licence de manière dynamique. Les simulations montrent que la méthode proposée présente de meilleures performances que la méthode de classification non supervisée existante.

9.3.6. *Les machines à vecteurs de support*

Les machines à vecteurs de support (SVM pour *Support Vector Machine*) sont une approche d'apprentissage automatique qui utilise un classifieur linéaire non probabiliste pour classer les données en deux catégories.

Les travaux suivants se sont intéressés à l'application des machines à vecteurs de support dans les RRC.

Quatre techniques d'apprentissage automatique supervisées, deux provenant des réseaux de neurones et deux provenant des machines à vecteurs de support, ont été utilisées pour étudier la prédiction de l'activité du PU dans (Agarwal *et al*. 2016). Les résultats mettent en évidence l'analyse des techniques d'apprentissage en fonction de diverses statistiques de trafic et suggèrent le meilleur modèle d'apprentissage permettant la prédiction avec précision de l'activité de l'utilisateur primaire.

Un vecteur de probabilité de faible dimension pour la détection coopérative du spectre basée sur des techniques d'apprentissage automatique dans un RRC a été proposé dans (Lu *et al*. 2016). L'algorithme de classification *K-means*, les machines à vecteurs de support et le vecteur de probabilité ont été étudiés par les auteurs. Considérant un RRC avec un PU et N SU, le vecteur de probabilité proposé pourrait réduire la dimension du vecteur d'énergie existant de N dimension à deux dimensions, entraînant une précision de détection égale sinon meilleure, une durée d'apprentissage plus courte et un temps de classification plus court.

9.3.7. *Le raisonnement à partir de cas*

Le raisonnement à partir de cas (CBR pour *Case-Based Reasoning*) est un paradigme de résolution de problèmes fondé sur la réutilisation d'expériences passées pour résoudre de nouveaux problèmes. CBR construit une base de données d'informations sur les situations passées, les problèmes, leurs solutions et leurs avantages. Les nouveaux problèmes sont ensuite résolus en trouvant le cas le plus similaire en mémoire et en déduisant la solution de la situation actuelle (Kolodner 2014).

Le travail suivant s'est intéressé à l'application du raisonnement à partir de cas dans les RRC.

Une méthode *Q-learning*, à partir de cas pour l'accès dynamique au spectre améliorant et stabilisant les performances des systèmes cellulaires cognitifs dotés de topologies dynamiques, a été proposée dans (Morozs *et al*. 2016). L'approche proposée est la combinaison d'un *Q-learning* distribué classique et d'une nouvelle implémentation de l'algorithme du raisonnement à partir de cas visant à faciliter un certain nombre de processus d'apprentissage exécutés en parallèle. Les simulations montrent que l'approche *Q-learning* proposée à partir de cas permet une amélioration constante de la qualité de service (QoS) dans des conditions de topologie de réseau dynamique et asymétrique et avec une charge de trafic.

9.3.8. *Les arbres de décisions*

Les arbres de décision (DT pour *Decision Tree*) sont l'une des structures de données majeures de l'apprentissage automatique. Leur fonctionnement repose sur des heuristiques qui, tout en satisfaisant l'intuition, donnent des résultats remarquables en pratique. La structure arborescente des arbres de décision les rend également lisibles par un être humain, contrairement à d'autres approches où le classifieur construit est une « boîte noire », comme dans le cas des réseaux de neurones.

Les travaux suivants se sont intéressés à l'application des arbres de décision dans les RRC.

Un nouvel algorithme qui combine la forêt aléatoire (plusieurs arbres de décision) pour réduire les interférences a été proposé dans (Wang et Yang 2016a). Une avancée qui a permis également d'améliorer considérablement le débit du réseau.

Un nouveau schéma de détection de spectre en se basant sur les arbres de décision pour la classification du protocole de la couche MAC a été proposé dans (Wang et Yang 2016b). Les simulations confirment que la nouvelle méthode proposée a pu améliorer considérablement le débit du réseau.

9.3.9. *Les réseaux bayésiens*

Les réseaux bayésiens (BN pour *Bayesian Networks*) sont des modèles probabilistes graphiques qui s'appuient sur l'interaction entre les différents nœuds pour obtenir un apprentissage pour et à partir de chaque nœud impliqué dans le

processus. Les réseaux bayésiens jouent un rôle dans le processus de prise de décision s'ils sont associés à d'autres outils afin de former des diagrammes d'influence (Bolstad 2004).

Les travaux suivants se sont intéressés à l'application des réseaux bayésiens dans les RRC.

Un modèle basé sur un réseau bayésien afin de traiter la modularité et l'incertitude a été proposé dans (Elderini *et al*. 2017). Le modèle bayésien permet d'ajouter qualitativement et quantitativement des paramètres qui ont un effet sur la probabilité de panne et le SINR dans un RRC.

Dans (Salahdine *et al*. 2017a), les réseaux bayésiens sont considérés parmi les techniques de gestion de l'incertitude dans les RRC. Un cas d'utilisation consiste à modéliser le RRC par un graphe dans lequel chaque nœud représente un SU et chaque arête représente le lien de communication entre les deux nœuds correspondants.

Une méthode hydride entre un modèle bayésien et une technique de trilatération, qui est utilisée pour obtenir une bonne approximation de la position de l'attaquant par émulation de l'utilisateur primaire (PUE), a été proposée dans (Fihri *et al*. 2017). La théorie de la décision bayésienne basée sur la fonction de perte et la probabilité conditionnelle permettant de déterminer l'existence de l'attaquant PUE dans la zone d'incertitude.

Une méthode efficace et rapide de *sensing*, dans laquelle la matrice de *sensing* de Toeplitz et le modèle bayésien sont combinés pour traiter les incertitudes et réduire le caractère aléatoire des mesures, a été proposée dans (Salahdine *et al*. 2017b). La méthode proposée a été implémentée et testée de manière approfondie, donnant ainsi des résultats satisfaisants.

Une approche bayésienne non paramétrique pour le *clustering* des sous-canaux dans les RRC basés sur l'OFDMA a été proposée dans (Ahmed *et al*. 2017). L'approche exploite les fonctionnalités de trafic de chaque sous-canal pour obtenir des statistiques sur sa période d'inactivité/occupé. Sur la base de l'énergie récoltée, le SU détermine le seuil de détection d'énergie de sorte qu'il puisse maximiser son taux d'utilisation de spectre tout en minimisant les interférences avec le PU.

9.3.10. *Les systèmes multiagents et l'apprentissage par renforcement*

Jacques Ferber a présenté les systèmes multiagents (MAS pour *Multiagent Systems*) en tant qu'entité intelligente consciente de son environnement, capable

d'agir avec habileté et de communiquer de façon autonome. Ils contiennent l'environnement, les objets, les agents et les différentes relations entre ces entités (Ferber et Weiss 1999). Avec les SMA, les utilisateurs peuvent interagir, négocier et coopérer pour assurer une communication plus efficace entre les entités du réseau.

Leur utilisation dans les RRC permettra aux utilisateurs de gérer leur propre spectre de manière dynamique et décentralisée. Les agents percevront leur environnement et réagiront en conséquence. L'association des SMA avec la RC permet une meilleure exploitation du spectre non utilisé et une gestion optimale des ressources radio, tout en réduisant les risques d'interférences.

De l'autre côté, l'apprentissage par renforcement (RL pour *Reinforcement Learning*) est un domaine de l'apprentissage automatique (ML pour *Machine Learning*) permettant de résoudre des problèmes de décisions séquentielles dans l'incertain. Il joue un rôle essentiel avec les SM, car il permet aux agents de découvrir la situation et de prendre des mesures à l'aide d'essais et d'erreurs, afin de maximiser la récompense cumulée. Dans l'apprentissage par renforcement, un agent doit prendre en compte les avantages immédiats et les conséquences de ses actions pour optimiser les performances du système à long terme (Wiering et Van Otterlo 2012).

Le *Deep Reinforcement Learning* (DRL) utilise les principes du *Deep Learning* et de l'apprentissage par renforcement pour créer des algorithmes efficaces pouvant être appliqués à des domaines tels que la robotique, les jeux vidéo, la finance et les soins de santé (François-Lavet *et al*. 2018). En mettant en œuvre une architecture de *Deep Learning* (réseaux de neurones profonds, etc.) avec un algorithme d'apprentissage par renforcement (*Q-learning*, etc.), il est possible de créer un modèle puissant de *Deep Reinforcement Learning* capable d'adapter des problèmes auparavant insolubles.

Les travaux suivants se sont intéressés à l'application des systèmes multiagents, de l'apprentissage par renforcement et du *Deep Reinforcement Learning* dans les RRC.

Le problème d'optimisation de l'accessibilité opportuniste des canaux dans les environnements de RC ferroviaire a été étudié dans (Yin *et al*. 2017). Le modèle proposé par les auteurs consiste en une inférence bayésienne permettant de calculer la probabilité de réussite de la transmission sur une station unique, ainsi qu'une collaboration en équipe visant à optimiser les performances du réseau au sein d'un groupe de stations de base.

Un mode de transmission efficace basé sur un algorithme *Q-learning* est proposé dans les RRC coopératifs dans (Rahman *et al*. 2016). L'état, l'action et la

récompense sont définis pour obtenir de bonnes performances en termes de délai et d'efficacité énergétique lors de la transmission de données, ainsi que pour l'interférence causée aux PU lors des transmissions réalisées par les SU. Les simulations montrent que le schéma proposé par les auteurs peut efficacement prendre en charge la détermination du mode de transmission et surpasse les schémas classiques existants dans la littérature.

L'allocation efficace de la puissance de transmission entre les SU, sans créer d'interférence pour les PU, est l'objectif des auteurs dans (Lall *et al*. 2016). Trois stratégies mixtes (équilibre corrélé) ont été utilisées pour contrôler la puissance de transmission pendant l'apprentissage. Les résultats expérimentaux ont indiqué que l'algorithme proposé surpasse de loin ses homologues classiques.

Le problème du routage de plusieurs flux générés par les SU vers une destination donnée, en tenant compte de la présence des PU, a été abordé dans (Pourpeighambar *et al*. 2017). Chaque SU souhaiterait minimiser égoïstement le délai de bout en bout de son flux, tout en respectant les exigences de QoS des PU. Pour une adaptation rapide de la décision de routage des SU aux changements d'environnement et à leur interaction non coopérative, les auteurs ont formulé le problème de routage en tant que processus d'apprentissage stochastique représenté par des jeux non coopératifs. Ensuite, ils ont proposé un système basé sur l'apprentissage par renforcement distribué pour résoudre le problème de routage et éviter ainsi les échanges d'informations entre les SU concurrents. Les simulations montrent que le schéma proposé converge de manière prouvable et démontrent son efficacité pour réduire le délai tout en respectant les exigences de QoS des PU.

Un algorithme basé sur l'apprentissage par renforcement pour gérer le problème d'attribution de puissance du canal de transmission et du canal de contrôle dans les RRC a été proposé dans (Lin *et al*. 2016). Les résultats de simulation montrent que ce nouvel algorithme apporte une amélioration significative en termes de compromis entre la fiabilité du canal de contrôle et l'efficacité du canal de transmission.

Un algorithme pour sélectionner un canal pour la transmission de données et pour prédire la durée pendant laquelle il restera inoccupé afin de minimiser le temps consacré à sa détection est proposé dans (Raj *et al*. 2018). Ledit algorithme apprend en deux étapes : une approche d'apprentissage par renforcement pour la sélectionner des canaux et une approche bayésienne pour déterminer la durée pendant laquelle la détection peut être ignorée.

Le problème du partage de spectre dans un RRC composé d'un PU et d'un SU a été étudié dans (Xingjian *et al*. 2018). Le PU et le SU travaillent de manière non coopérative. L'objectif des auteurs est de développer une méthode de contrôle de

puissance basée sur l'apprentissage pour le SU, afin de partager le spectre commun avec le PU. Les auteurs ont développé une méthode basée sur le *Deep Reinforcement Learning*, que le SU peut utiliser pour ajuster intelligemment sa puissance d'émission, de telle sorte qu'après deux cycles d'interaction avec le PU, les deux utilisateurs puissent transmettre leurs propres donnés avec la QoS demandées.

9.4. Catégorisation et utilisation des techniques dans la radio cognitive

La disponibilité des données affecte le choix de la technique d'apprentissage à utiliser. L'apprentissage supervisé est utilisé lorsque les données d'apprentissage sont étiquetées et que la RC dispose d'informations préalables sur l'environnement. Les algorithmes non supervisés, quant à eux, ne nécessitent pas de données d'apprentissage étiquetées. L'apprentissage non supervisé est utilisé lorsque certains composants RF ne sont pas connus du RRC, ce qui lui permet de fonctionner de manière autonome sans aucune connaissance préalable.

Les techniques d'apprentissage se différencient dans leurs points forts, leurs limites, leurs défis et leurs applications dans les RRC.

Les arbres de décision, les réseaux de neurones, les machines à vecteurs de support et le raisonnement à partir de cas sont considérés comme des techniques d'apprentissage supervisé. La théorie des jeux, l'apprentissage par renforcement, les modèles de Markov, les métaheuristiques, la logique floue et les réseaux bayésiens sont considérés comme des techniques d'apprentissage non supervisé.

À noter que toutes les techniques étudiées dans ce chapitre sont utilisées à la fois pour le *sensing* et pour la décision dans la RC, sauf pour les métaheuristiques et pour le raisonnement à partir de cas, qui sont utilisés exclusivement pour la décision. Les arbres de décision, la théorie des jeux et la logique floue sont à utiliser pour le *sensing* en tenant compte de la capacité de la technique de détection du spectre utilisée.

Les SMA peuvent être utilisés en combinaison avec toutes les techniques présentées dans ce chapitre.

9.5. Conclusion

Ce chapitre présente une étude complète sur les techniques de l'intelligence artificielle utilisées dans les réseaux de radio cognitive. Les définitions des

différentes techniques et leurs applications dans la RC ont été aussi discutés dans ce travail. En effet, un état de l'art complet sur l'application de l'algorithme des lucioles, la recherche coucou, l'algorithme de recherche gravitationnelle, l'optimisation par essaim de particules, les algorithmes génétiques, les algorithmes de colonies d'abeilles, la logique floue, la théorie des jeux, les réseaux de neurones, les modèles de Markov, les machines à vecteurs de support, le raisonnement à partir de cas, les arbres de décision, les réseaux bayésiens, les systèmes multiagents et l'apprentissage par renforcement dans les RRC ont été présentés dans ce chapitre. Ce dernier présente également les tâches principales de la RC et leurs défis correspondants, catégorise les techniques présentées selon le type d'apprentissage (supervisé ou non supervisé) et présente leurs applications en fonction des deux tâches de la RC (le *sensing* et la décision).

9.6. Bibliographie

Abbas, N., Nasser, Y., El Ahmad, K. (2015). Recent advances on artificial intelligence and learning techniques in cognitive radio networks. *EURASIP Journal on Wireless Communications and Networking*, 174.

Abdalzaher, M.S., Seddik, K., Muta, O. (2017). Using Stackelberg game to enhance cognitive radio sensor networks security. *Iet Communications*, 11(9), 1503–1511.

Agarwal, A. *et al*. (2016). Learning based primary user activity prediction in cognitive radio networks for efficient dynamic spectrum access. Dans *International Conference on Signal Processing and Communications*.

Ahmed, M.E. *et al*. (2017). Energy-arrival-aware detection threshold in wireless-powered cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 66(10), 9201–9213.

Alhammadi, A., Roslee, M., Yusoff Alias, M. (2016). Analysis of spectrum handoff schemes in cognitive radio network using particle swarm optimization. Dans *IEEE – 3rd International Symposium on Telecommunication Technologies*.

Banerjee, J.S., Chakraborty, A., Chattopadhyay, A. (2017). Fuzzy based relay selection for secondary transmission in cooperative cognitive radio networks. *Advances in Optical Science and Engineering*, 279–287.

Bayrakdar, M.E., Çalhan, A. (2018). Artificial bee colony–based spectrum handoff algorithm in wireless cognitive radio networks. *International Journal of Communication Systems*, 31(5).

Bellhouse, D. (2007). The problem of Waldegrave. *Electronic Journal for the History of Probability and Statistics*, 3(2), 1–12.

Bestaoui, A.A. (2015). Gestion de spectre dans un réseau de radio cognitive en utilisant l'algorithme d'optimisation par essaim de particules. PFE de master en informatique, Université de Tlemcen, Tlemcen.

Biglieri, E. *et al*. (2013). *Principles of cognitive radio*. Cambridge University Press, Cambridge.

Bkassiny, M., Li, Y., Jayaweera, S.K. (2013). A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys & Tutorials*, 15(3), 1136–1159.

Bolstad, W.M. (2004). *Introduction to Bayesian Statistics*. John Wiley & Sons, Hoboken.

Ching, W.-K., Ng, M.K. (2006). *Markov chains. Models, algorithms and applications*. Springer, Berlin.

Craig, W.R. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4), 25–34.

Darwin, C. (2009) *The Origin of Species: By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 6e édition. Cambridge University Press, Cambridge.

Elahi, A. *et al*. (2017). Interference reduction in Cognitive radio networks using Genetic and Firefly Algorithms. Dans *International Conference on Communication, Computing and Digital Systems*.

Elderini, T., Kaabouch, N., Reyes, H. (2017). Outage probability estimation technique based on a bayesian model for cognitive radio networks. Dans *IEEE 7th Annual Computing and Communication Workshop and Conference*.

Elghamrawy, S.M. (2018). Security in cognitive radio network: defense against primary user emulation attacks using genetic artificial bee colony (GABC) algorithm. *Future generation computer systems*.

Elhachmi, J., Guennoun, Z. (2016). Cognitive radio spectrum allocation using genetic algorithm. *EURASIP Journal on Wireless Communications and Networking*, 133.

Fang, H., Xu, L., Raymond Choo, K.-K. (2017). Stackelberg game based relay selection for physical layer security and energy efficiency enhancement in cognitive radio networks. *Applied Mathematics and Computation*, 296, 153–167.

Ferber, J., Weiss, G. (1999). *Multiagent systems: an introduction to distributed artificial intelligence*, volume 1. Addison-Wesley, Boston.

Fihri, W.F. *et al*. (2017). Bayesian decision model with trilateration for primary user emulation attack localization in cognitive radio networks. Dans *International Symposium on Networks, Computers and Communications*.

Fihri, W.F. *et al*. (2018). A particle swarm optimization based algorithm for primary user emulation attack detection. Dans *IEEE 8th Annual Computing and Communication Workshop and Conference*.

François-Lavet, V. *et al*. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, 11(3/4), 219–354.

Fraser, A.M. (2008). *Hidden Markov Models and Dynamical Systems*, volume 107. Siam, Philadelphia.

Gavrilovska, L., Atanasovski, V., Macaluso, I., Da Silva, L. (2013). Learning and reasoning in cognitive radio networks. *IEEE Communications Surveys Tutor*, 15(4), 1761–1777.

Gerardo, B., Wang, J. (1993). Swarm intelligence in cellular robotic systems. Dans *Robots and biological systems: towards a new bionics? Conference proceedings*, 703–712.

Ghanem, W.R., Shokair, M., Desouky, M.I. (2016). An improved primary user emulation attack detection in cognitive radio networks based on firefly optimization algorithm. Dans *33rd National Radio Science Conference*.

Ghasemi, A., Sousa, E.S. (2008). Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. *IEEE Communications magazine*, 46(4), 32–39.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing, Boston.

Guo, L., Chen, Z., Huang, L. (2018). A novel cognitive radio spectrum allocation scheme with chaotic gravitational search algorithm. *International Journal of Embedded Systems*, 10(2), 161–167.

Haddad, O.B., Afshar, A., Mariño, M.A. (2006). Honey-bees mating optimization (HBMO) algorithm: a new heuristic approach for water resources optimization. *Water Resources Management*, 20(5), 661–680.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications. *IEEE journal on selected areas in communications*, 23(2), 201–220.

Haykin, S. (2008). *Neural Networks and Learning Machines*. Pearson, Londres.

Heppner, F., Grenander, U. (1990). *A stochastic nonlinear odel for coordinated bird flocks*. AAAS Publication, Washington.

Holland, J.H. *et al*. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, Cambridge.

Ichikawa, K., Fujii, T. (2017). Radio environment map construction using hidden Markov model in multiple primary user environment. Dans *International conference on computing, networking and communications*.

Jiao, Y., Joe, I. (2016). Energy-efficient resource allocation for heterogeneous cognitive radio network based on two-tier crossover genetic algorithm. *Journal of Communications and Networks*, 18(1), 112–122.

Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization, volume 200. Rapport technique, Erciyes University.

Kaur, K., Rattan, M., Singh Patterh, M. (2018a). Cuckoo search based optimization of multiuser cognitive radio system under the effect of shadowing. *Wireless Personal Communications*, 99(3), 1217–1230.

Kaur, K., Rattan, M., Singh Patterh, M. (2018b). Cognitive Radio Design Optimization Over Fading Channels Using PSO, GSA and Hybrid PSOGSA. Dans *Second International Conference on Intelligent Computing and Control Systems*.

Kennedy, J., Eberhart, R.C. (1995). Particle Swarm Optimization. Dans *Proceedings of the IEEE International Conference on Neural Networks IV*, 1942–1948.

Kolodner, J. (2014). *Case-based reasoning*. Morgan Kaufmann, Burlington.

Lall, S. *et al*. (2016). Multiagent reinfocement learning for stochastic power management in cognitive radio network. Dans *International Conference on Microelectronics, Computing and Communications*.

Lin, Y. *et al*. (2016). A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks. *Sensors*, 16(10), 1675.

Liu, Y. *et al*. (2017). Adaptive double threshold energy detection based on Markov model for cognitive radio. *PLOS ONE*, 12(5).

Liu, M. *et al*. (2019). Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 68(1), 641–65.

Lu, Y. *et al*. (2016). Machine learning techniques with probability vector for cooperative spectrum sensing in cognitive radio networks. Dans *IEEE Wireless Communications and Networking Conference*.

Manesh, M.R. *et al*. (2017). An optimized SNR estimation technique using particle swarm optimization algorithm. Dans *IEEE 7th Annual Computing and Communication Work-shop and Conference*.

Manjith, R. (2016). A hybrid of BFO and MCS algorithms for channel estimation of cognitive radio system. *Arabian Journal for Science and Engineering*, 41(3), 841–852.

Mitola, J. (2000). Cognitive radio: an integrated agent architecture for software defined radio. Thèse de doctorat, Royal Institute of Technology, Stockholm.

Mitola, J., Maguire, G.Q. (1999). Cognitive radio: making software radios more personal. *IEEE personal communications*, 6(4), 13–18.

Morozs, N., Clarke, T., Grace, D. (2016). Cognitive spectrum management in dynamic cellular environments: A case-based Q-learning approach. *Engineering Applications of Artificial Intelligence*, 55, 239–249.

Nakrani, S., Tovey, C. (2004). On honey bees and dynamic server allocation in internet hosting centers. *Adaptive Behavior*, 12(3–4), 223–240.

Norris, J.R. (1998). *Markov chains*, numéro 2. Cambridge University Press, Cambridge.

Pourpeighambar, B., Dehghan, M., Sabaei, M. (2017). Non-cooperative reinforcement learning based routing in cognitive radio networks. *Computer communications*, 106, 11–23.

Qiu, R. *et al*. (2012) *Cognitive radio communication and networking: Principles and practice*. John Wiley & Sons, Hoboken.

Rahman, M.A., Lee, Y.-D., Koo, I. (2016). An efficient transmission mode selection based on reinforcement learning for cooperative cognitive radio networks. *Human-centric Computing and Information Sciences*, 6(1), 2.

Raj, V. *et al*. (2018). Spectrum access in cognitive radio using a two-stage reinforcement learning approach. *IEEE Journal of Selected Topics in Signal Processing*, 12(1), 20–34.

Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S. (2009). GSA: a gravitational search algorithm. *Information sciences*, 179(13), 2232–2248.

Rojas, R. (2013). *Neural networks: a systematic introduction*. Springer Science & Business Media, Berlin.

Roy, A. *et al*. (2017). Optimized secondary user selection for quality of service enhancement of Two-Tier multi-user Cognitive Radio Network: A game theoretic approach. *Computer Networks*, 123, 1–18.

Russell, S.J., Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson Malaysia, Kuala Lumpur.

Salahdine, F., Kaabouch, N., El Ghazi, H. (2017a). Techniques for dealing with uncertainty in cognitive radio networks. Dans *IEEE 7th Annual Computing and Communication Workshop and Conference*.

Salahdine, F., Kaabouch, N., El Ghazi, H. (2017b). A Bayesian recovery technique with Toeplitz matrix for compressive spectrum sensing in cognitive radio networks. *International Journal of Communication Systems*, 30(15).

Saoucha, N.A., Benmammar, B. (2017). Adapting radio resources in multicarrier cognitive radio using discrete firefly approach. *International Journal of Wireless and Mobile Computing*, 13(1), 39–44.

Shu, Z. *et al*. (2016). A game theoretic approach for energy-efficient communications in multi-hop cognitive radio networks. *Wireless communications and mobile computing*, 16(14), 2131–2143.

Supraja, P., Pitchai, R. (2019). Spectrum prediction in cognitive radio with hybrid optimized neural network. *Mobile Networks and Applications*, 24(2), 357–364.

Surafel, L.T., Ngnotchouye, J.M.T. (2017). Firefly algorithm for discrete optimization problems : A survey. *KSCE Journal of Civil Engineering*, 21(2), 535–545.

Tang, M., Xin, Y. (2016). Energy efficient power allocation in cognitive radio network using coevolution chaotic particle swarm optimization. *Computer Networks*, 100, 1–11.

Tounsi, A., Babes, M. (2017). An efficient joint spectrum and power allocation in cognitive radio networks using a modified firefly algorithm. *International Journal of Communication Networks and Distributed Systems*, 19(2), 214–236.

Tuan, P., Koo, I. (2016). Throughput maximisation by optimising detection thresholds in full-duplex cognitive radio networks. *IET Communications*, 10(11), 1355–1364.

Tuan, P., Koo, I. (2017). Robust weighted sum harvested energy maximization for SWIPT cognitive radio networks based on particle swarm optimization. *Sensors*, 17(10), 2275.

Umar, R., Sheikh, A.U.H. (2012). Cognitive radio oriented wireless networks: challenges and solutions. Dans *Proceedings of the International Conference on Multimedia Computing and Systems*.

Wang, B., Liu, K.J.R. (2011). Advances in cognitive radio networks: a survey. *IEEE J. Selected Topics Signal Process*, 5(1), 5–23.

Wang, D., Yang, Z. (2016a). A novel spectrum sensing scheme combined with machine learning. Dans *9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*.

Wang, D., Yang, Z. (2016b). An advanced scheme with decision tree for the improvement of spectrum sensing efficiency in dynamic network. Dans *9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics.*

Wiering, M., Van Otterlo, M. (2012). Reinforcement learning. *Adaptation, learning, and optimization*, 12, 3.

Woods, W.A. (1986). Important issues in knowledge representation. *Proceedings of the IEEE*, 74(1), 1322–1334.

Wyglinski, A.M., Nekovee, M., Hou, T. (dir.) (2009). *Cognitive radio communications and networks: principles and practice*. Academic Press, Cambridge.

Xingjian, L. *et al*. (2018). Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach. *IEEE Access*, 6, 25463–25473.

Yang, X.-S. (2005). Engineering optimizations via nature-inspired virtual bee algorithms. Dans *International Work-Conference on the Interplay Between Natural and Artificial Computation*.

Yang, X.-S., Deb, S. (2017). Cuckoo search : state-of-the-art and opportunities. Dans *IEEE 4th International Conference on Soft Computing & Machine Intelligence*, 55–59.

Yin, Z., Wang, Y., Wu, C. (2017). A Multiagent Collaborative Model for Bayesian Opportunistic Channel Accessibility in Railway Cognitive Radio. *International Journal of Performability Engineering*, 13(4).

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.

Zaheer, M. *et al*. (2016). Interference control in cognitive radio using joint beamforming and power optimization by applying artificial bee colony. Dans *19th International Multi-Topic Conference*.

Zhai, L., Wang. H. (2017). Crowdsensing task assignment based on particle swarm optimization in cognitive radio networks. *Wireless Communications and Mobile Computing*.

Zhang, M., Diao, M., Guo, L. (2017). Convolutional neural networks for automatic cognitive radio waveform recognition. *IEEE Access*, 5, 11074–11082.

Zhao, Y., Morales-Tirado, L. (2012). Cognitive radio technology: Principles and practice. Dans *International Conference on Computing, Networking and Communications*.

Zhu, J. *et al*. (2017). A game-theoretic power control mechanism based on hidden Markov model in cognitive wireless sensor network with imperfect information. *Neurocomputing*, 220, 76–83.

10

Apport de la radio intelligente pour répondre aux besoins de communication sur route des véhicules autonomes

Francine KRIEF[1], Hasnaâ ANISS[2], Marion BERBINEAU[2] et Killian LE PAGE[3]

[1] *ENSEIRB-MATMECA, Bordeaux, France*

[2] *IFSTTAR, Bordeaux, France*

[3] *ALTEN, Boulogne-Billancourt, France*

10.1. Introduction

Les besoins de communication sur route vont évoluer et connaître une forte croissance grâce à l'arrivée du véhicule connecté et autonome. La connectivité véhicule à véhicule ou *Vehicle-to-Vehicle* (V2V), véhicule à infrastructure ou *Vehicle-to-infrastructure* (V2I), véhicule à piétons ou *Vehicle-to-Pedestrian* (V2P), et plus largement véhicule à X, permet le déploiement d'une large variété d'applications visant en premier lieu à améliorer la sécurité routière mais également le confort des usagers lors de leurs déplacements. Dans le contexte des véhicules autonomes, de nouveaux services verront le jour comme la conduite en convoi ou *platooning*, qui permettra à un véhicule avec chauffeur de guider des véhicules autonomes rassemblés en convoi sur le réseau urbain. Le partage de véhicules et l'Internet des objets vont encore élargir le champ des services offerts. Ces applications, toujours plus nombreuses, auront des exigences très variées en termes de qualité de service et de sécurité des communications, auxquelles il conviendra de répondre. La fiabilité du lien

Gestion et contrôle intelligents des réseaux,
coordonné par Badr BENMAMMAR.

de communication nécessitera une auto-adaptation de la technologie d'accès radio, qui peut être assurée grâce à l'utilisation de la radio intelligente, technologie capable de détecter les bandes de fréquences libres et d'adapter ses paramètres de transmission en fonction des besoins et des contraintes de la communication. Ce concept a été introduit par Mitola en 1999 (Mitola et Maguire 1999).

La radio intelligente (RI) se définit par ses capacités de perception, d'adaptation et de cognition. Elle se caractérise par des capacités à l'interopérabilité spectrale, l'efficacité, l'optimisation des ressources radio et l'amélioration de la fiabilité des communications, qui présentent un intérêt considérable pour le véhicule connecté et autonome. Des premiers prototypes ont vu le jour avec succès pour répondre aux exigences des communications militaires, ainsi que dans le domaine de la sécurité publique. En plus d'être capable de percevoir et de s'adapter à son environnement radioélectrique, la radio intelligente a aussi des capacités d'apprentissage *via* les techniques de l'intelligence artificielle.

Dans la section 10.2 nous introduisons le véhicule autonome, ainsi que ses principaux composants. Par la suite, nous décrivons dans la section 10.3 des applications de communication véhiculaires et leurs contraintes en termes de qualité de service (QoS). La section 10.4 présente les différentes architectures de communication. La section 10.5 met en avant l'apport de la radio intelligente dans le domaine véhiculaire. Nous présentons également dans cette section un état de l'art des principaux travaux de recherche dans le domaine de la RI pour les réseaux véhiculaires. Nous spécifions dans la section 10.6 notre positionnement au sein du projet SERENA. Enfin, la section 10.7 conclut ce chapitre.

10.2. Le véhicule autonome

Le véhicule autonome (VA), véhicule dont la conduite est en partie ou entièrement automatisée, représente l'un des grands défis technologiques actuels. Son arrivée bouleversera la mobilité, la sécurité et le comportement des usagers de la route.

La recherche sur le véhicule autonome, ou plutôt automatisé et connecté, est très active et couvre de nombreux domaines, que ce soit la connectivité à travers une infrastructure numérique, les technologies de perception de l'environnement, la localisation sûre et précise, la cartographie haute définition, le juridique comprenant tous les aspects règlementaires et légaux relatifs à la responsabilité (les propriétaires du véhicule, les constructeurs, l'État, les collectivités) mais aussi la philosophie et l'éthique (quelle est la responsabilité morale du développeur de l'intelligence artificielle derrière le véhicule ?).

10.2.1. *Les niveaux d'automatisation*

La Society of Automotive Engineers (SAE) identifie 6 niveaux d'assistance à la conduite allant de 0 à 5. Nous pourrons parler réellement de véhicule autonome à partir du niveau 5. Avant cela, le véhicule n'est pas encore considéré comme autonome puisqu'il nécessite la présence du conducteur. Les niveaux d'automatisation sont décrits dans le tableau 10.1, issu de (Shladover *et al*. 2014).

10.2.2. *Les principaux composants*

Une voiture autonome désigne une voiture dont la conduite est en partie ou entièrement automatisée, grâce à différents capteurs permettant la perception de l'environnement et une connectivité à l'infrastructure routière, afin d'anticiper les événements de la scène routière. Les différents éléments permettant au VA de percevoir, de se localiser et de communiquer peuvent être résumés par la figure 10.1.

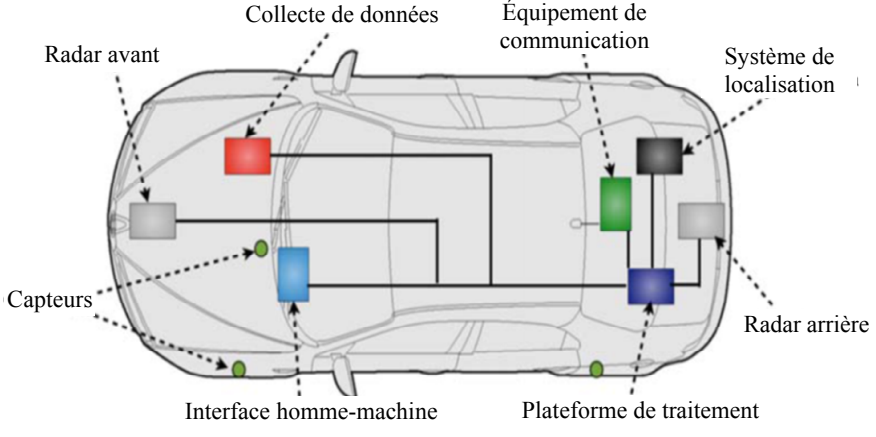


Figure 10.1. *Véhicule autonome (Hubaux 2005)*

Les systèmes de perception à l'avant et à l'arrière sont capables de percevoir l'environnement physique, afin de prévenir d'une éventuelle collision. Ces systèmes sont de type multi capteur et sont composés entre autres de radar anticollision, de lidar et de systèmes d'imagerie dans le domaine visible et parfois infrarouge. Certains systèmes considèrent également des capteurs audio. Le module de collecte et de traitement des données est responsable de la récupération des données. Il est directement relié au bus CAN (*Control Area Network*) du véhicule et permet d'accéder aux données, telles que la vitesse, l'accélération ou encore la température et l'humidité, mais aussi l'état du véhicule et de son environnement.

Niveau SAE	Nom	Description	Exécution de la conduite (volant, accélération, freinage)		Surveillance de l'environnement de conduite	Reprise de la conduite	Capacité du système
L'humain surveille l'environnement de conduite							
0	Aucune automatisation	Exécution à plein temps par le conducteur humain de tous les aspects de la tâche de conduite dynamique, même lorsqu'elle est renforcée par des systèmes d'alerte ou d'intervention	Conducteur humain	Conducteur humain	Conducteur humain	N/A	
1	Assistance à la conduite	Exécution (mode de conduite spécifique) par le système d'aide à la conduite, des manœuvres, soit sur la direction, soit sur l'accélération/décélération, à l'aide d'informations sur l'environnement de conduite. Le conducteur humain exécute tous les autres aspects de la tâche dynamique de la conduite	Conducteur humain et système	Conducteur humain	Conducteur humain	Certains modes de conduite	

Niveau SAE	Nom	Description	Exécution de la conduite (volant, accélération, freinage)	Surveillance de l'environnement de conduite	Reprise de la conduite	Capacité du système
2	Automatisation partielle	Exécution (mode de conduite spécifique) par un ou plusieurs systèmes d'aide à la conduite d'actions à la fois sur la direction et l'accélération/décélération à l'aide d'informations sur l'environnement de conduite. Le conducteur humain exécute tous les autres aspects de la tâche de conduite dynamique	Système	Conducteur humain	Conducteur humain	Certains modes de conduite
Le système de conduite automatisé surveille l'environnement de conduite						
3	Automatisation conditionnelle	Exécution (mode de conduite spécifique) de tous les aspects de la tâche de conduite dynamique par un système de conduite automatisé avec l'attente que le conducteur humain répondra de manière appropriée à une demande d'intervention	Système	Système	Conducteur humain	Certains modes de conduite

Niveau SAE	Nom	Description	Exécution de la conduite (volant, accélération, freinage)	Surveillance de l'environnement de conduite	Reprise de la conduite	Capacité du système
4	Forte automatisation	Exécution (mode de conduite spécifique) de tous les aspects de la tâche de conduite dynamique par un système de conduite automatisé, même si le conducteur humain ne répond pas de manière appropriée à une demande d'intervention	Système	Système	Système	Certains modes de conduite
5	Automatisation complète	Exécution à temps plein par un système de conduite automatisé de tous les aspects de la tâche de conduite dynamique dans toutes les conditions routières et environnementales pouvant être gérées par un conducteur humain	Système	Système	Système	Tous les modes de conduite

Tableau 10.1. *Les différents niveaux d'automatisation SAE (source : 2014 SAE International)*

L'équipement de communication s'appuie aujourd'hui sur le système ITS-G5 et les systèmes cellulaires existants. Il permet la communication entre les véhicules et l'infrastructure, mais aussi entre les véhicules. Il est notamment responsable de la transmission des alertes vers des centres de contrôle en cas de détection de situations potentiellement dangereuses.

L'interface homme-machine est l'interface utilisable par le conducteur pour interagir avec le véhicule.

Le système de localisation permet le géo positionnement du VA, qui sera d'autant plus précis qu'il est associé à une cartographie haute définition et à une fusion de données issues des multiples capteurs.

Pour plus de détails, le lecteur peut consulter le dossier thématique « Regards croisés sur le véhicule autonome » réalisé par l'IFSTTAR (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux)[1].

10.3. Le véhicule connecté

Les systèmes de transport intelligents coopératifs ou *Cooperative Intelligent Transport System* (C-ITS) basés sur les communications véhiculaires permettent le déploiement de nouvelles applications. Celles-ci peuvent être classées en plusieurs catégories ; citons ici la sécurité routière et le divertissement. Les besoins en qualité de service (QoS) et les performances varient en fonction du type d'application.

10.3.1. *Les applications de sécurité routière*

Les applications de sécurité routière sont les plus critiques. Il s'agit d'une des catégories avec les plus grandes exigences en termes de qualité de service. Les exemples d'applications peuvent aller (de façon non exhaustive) de la détection de collision à la gestion d'accident en passant par la détection de piétons sur la route (Dar *et al.* 2010 ; Cunha *et al*. 2016).

– Détection de collision : l'objectif est de détecter en amont les risques de collision avec un obstacle quelconque. Les points les plus critiques pour la communication sont la latence, qui doit être la plus faible possible afin de transmettre l'information au plus vite, et la précision du positionnement du véhicule (mobile ou statique) et de l'obstacle.

1. https://www.ifsttar.fr/ressources-en-ligne/espace-science-et-societe/mobilites/dossiers-thematiques/vehicule-autonome/.

– Gestion d'accident : l'objectif est de détecter les accidents en cours, ou prévisibles à très court terme, pour changer le comportement du conducteur, par exemple en lui indiquant un nouvel itinéraire contournant l'accident. Dans ce cas, une portée radio suffisamment grande est nécessaire pour permettre d'anticiper au mieux. La latence doit aussi être très faible pour prendre une décision le plus rapidement possible. Enfin, concernant le débit, celui-ci doit être suffisamment important, afin de transmettre la totalité des informations nécessaires, suite à l'accident rencontré.

– Détection de piétons traversant la route : l'objectif est de notifier au conducteur la présence d'un piéton arrivant sur la route de manière inattendue. Le conducteur (ou le VA) peut ainsi freiner ou dévier son véhicule de la trajectoire du piéton pour éviter tout risque d'accident avec ce dernier. La latence doit donc être suffisamment faible pour permettre de réagir en temps réel et éviter ainsi l'accident. La portée doit être au minimum de 200 mètres pour anticiper au mieux le passage du piéton.

Les travaux référencés (Dar *et al*. 2010 ; Amjad *et al*. 2018 ; MacHardy 2018 ; Mir et Filali 2018) proposent certaines caractéristiques de performances nécessaires à la réalisation de ces cas d'usage, résumées dans le tableau 10.2.

Application	Latence (ms)	Débit (Mo/s)	Distance (m)
Détection de collision	10 à 100	De 0,0625 à 87,5	Milieu urbain : 500 Autoroute : 2 000
Gestion d'accident	De l'ordre de la milliseconde	De 0,125 à 0,75	De moins de 500 à environ 1 000
Détection de piéton	20 à 100	De 0,0625 à 87,5	≥ 200

Tableau 10.2. *Contraintes exprimées pour des applications véhiculaires de sécurité routière*

10.3.2. *Les applications de divertissement*

Les applications de divertissement sont les moins critiques en termes de qualité de service. Effectivement, de par leur nature, ces applications ne nécessitent pas de priorités très élevées contrairement aux applications précédemment citées. Les

besoins seront à considérer au cas par cas. Prenons deux exemples différents issus de (Campolo *et al*. 2017) : le *streaming* vidéo et la navigation sur Internet.

– *Streaming* vidéo : cette application consiste à diffuser ou à lire des flux vidéo. La latence doit être la plus faible possible pour garantir une bonne qualité d'expérience utilisateur. Le débit doit être dimensionné en fonction de la qualité souhaitée. Pour une très bonne qualité vidéo, le débit devra être important, alors que ce dernier pourra être moindre pour une qualité vidéo plus faible. La portée radio doit être suffisamment grande pour permettre un accès au contenu pendant un long parcours et éviter toute interruption.

– Navigation Internet : il s'agit de permettre à l'utilisateur de consulter des pages internet. La latence est moins importante comparée aux autres applications, tout en étant suffisamment faible pour satisfaire la qualité d'expérience de l'utilisateur. Le débit nécessaire pour une navigation fluide peut dépendre du contenu visité, mais ce dernier n'a pas besoin d'être aussi important que pour les autres applications. Enfin, la portée radio doit aussi être importante pour permettre une navigation en continu.

Les travaux de (Campolo *et al*. 2017) et (Dar *et al*. 2010) estiment les performances minimales nécessaires, résumées dans le tableau 10.3.

Application	Latence (ms)	Débit (Mo/s)	Portée (m)
***Streaming* vidéo**	Plus faible possible	1,875 (pour de la vidéo en UHD)	> 1 000
Navigation Internet	100 ms	0,0625	> 1 000

Tableau 10.3. *Contraintes exprimées pour des applications véhiculaires de divertissement*

10.4. Les architectures de communication

Les réseaux véhiculaires C-ITS s'appuient sur un ensemble de standards à l'ISO (International Organization for Standardization) et à l'ETSI (European Telecommunication Standards Institute) définissant l'architecture de chaque composant (véhicule, unité de bord de route, centre de gestion routier, etc., voir figure 10.2) sur la base d'une pile protocolaire ITS (*Intelligent Transport System*) et d'une pile OSI (*Open Systems Interconnexion*).

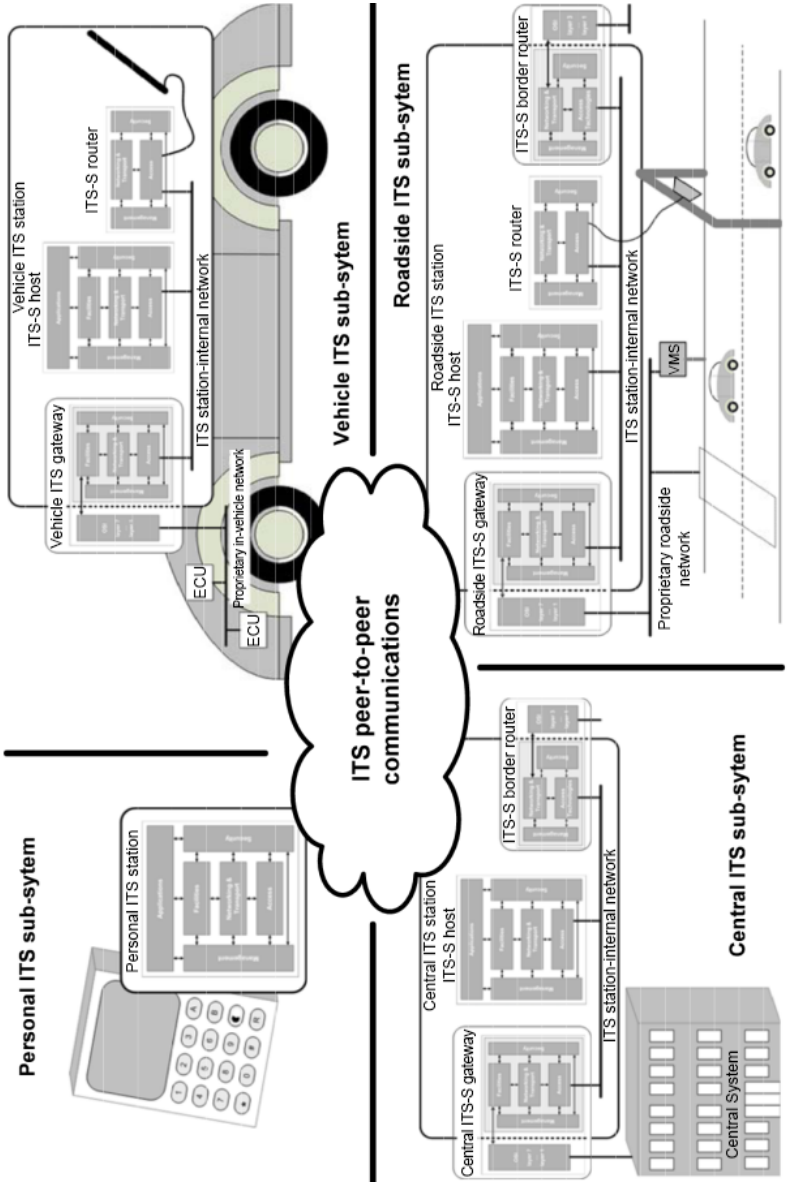


Figure 10.2. *Illustration des systèmes C-ITS (ETSI 2010)*

Chaque sous-système illustré figure 10.2 est défini à partir de la même architecture, composée (voir figure 10.3) :

– d'une couche « Accès », qui représente les couches OSI 1 et 2 ;

– d'une couche « Réseau et transport », qui représente les couches OSI 3 et 4 ;

– d'une couche « Installation », qui représente les couches OSI 5, 6 et 7 ;

– d'une couche « Application », qui gère la production du service C-ITS ;

– d'une couche transversale « Sécurité », pour la sécurisation du message ;

– d'une couche transversale pour le « Management », qui gère l'interaction entre les différentes couches.

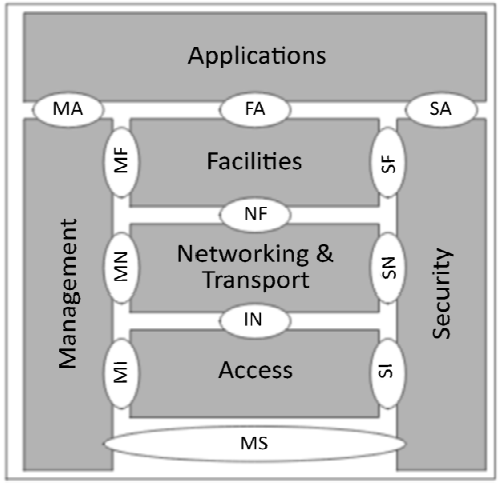


Figure 10.3. *Architecture de référence d'une station ITS (ETSI 2010)*

En termes d'accès, deux types de communication sont utilisés :

– en mode *ad hoc*, l'ITS-G5 ; dans ce cas, le transport se fait soit en IP soit, plus généralement, en utilisant le *geonetworking* (ETSI 2014) ;

– le cellulaire (3G/4G) ; le lien est alors exclusivement IP.

Ces deux modes d'accès sont parfois couplés, afin d'augmenter le niveau de couverture des services ; c'est alors le mode hybride.

10.4.1. *ITS-G5*

Le système ITS-G5 est régi par un ensemble de normes de l'ETSI basées, pour la partie couche physique/contrôle d'accès au support, sur la norme IEEE 802.11p (IEEE 2010).

L'ETSI a alloué trois bandes de fréquences dans la bande des 5 GHz pour les systèmes C-ITS. Chaque bande est divisée en canaux de 10 MHz. La première bande, dite ITS-G5A, de 30 MHz, est dédiée aux applications de sécurité routière. La deuxième bande de 20 MHz, dite ITS-G5B, est destinée aux autres applications. La dernière bande de 20 MHz (ITS-G5D) est réservée à un usage futur.

	Intervalle de fréquence	Usage	Réglementation	Standard harmonisé
ITS-G5A	5 875 à 5 905	Applications ITS pour la sécurité routière	Décision de la commission	EN 302571
ITS-G5B	5 855 à 5 875	Applications ITS non liées à la sécurité routière	Recommandation ECC	EN 302571
ITS-G5C	5 470 à 5 725	RLAN (BRAN, WLAN)	Décision ERC	EN 301893
ITS-G5D	5 905 à 5 925	Applications ITS futures	Décision ECC	EN 302571

Tableau 10.4. *Le système ITS-G5*

Type de canal	Fréquence centrale (MHz)	Numéro de canal IEEE 802.11	Espacement canal (MHz)	Débit par défaut (Mbit/s)	Tx puissance limite (dBm PIRE)	Tx densité de puissance limite (dBm/MHz)
G5-CCH	5 900	180	10	6	33	23
G5-SCH1	5 880	176	10	6	33	23
G5-SCH2	5 890	178	10	12	23	13

Type de canal	Fréquence centrale (MHz)	Numéro de canal IEEE 802.11	Espacement canal (MHz)	Débit par défaut (Mbit/s)	Tx puissance limite (dBm PIRE)	Tx densité de puissance limite (dBm/MHz)
G5-SCH3	5 870	174	10	6	23	13
G5-SCH4	5 860	172	10	6	0	– 10
G5-SCH5	5 850	182	10	6	0	– 10
G5-SCH6	5 910	184	10	6	0	– 10
G5-SCH7	Comme décrit dans la bande : 5 470 MHZ à 5 725 MHz	94 à 145	Plusieurs	Dépend de l'espacement des canaux	30 (DFS maître)	17
					23 (DFS esclave)	10

Tableau 10.5. *Table d'allocation des fréquences ITS-G5 en Europe (IEEE 2010)*

La prise en charge de débits numériques de 3 Mbps, 6 Mbps et 12 Mbps est obligatoire pour les stations ITS.

Contrôle d'accès au support (MAC)

Pour permettre l'utilisation de la norme 802.11 dans le cadre des communications véhiculaires, les fonctionnalités suivantes ont été modifiées :

– l'authentification de la sous-couche MAC et les procédures d'association sont désactivées ;

– l'économie d'énergie n'est pas autorisée ;

– la sécurité 802.11 n'est pas supportée.

10.4.2. *LTE-V2X*

En 2016, le groupe de standardisation 3GPP a publié des spécifications pour les communications V2X s'appuyant sur la technologie LTE (*Long Term Evolution*),

appelé « LTE V2X » (ou C-V2X) pour se différencier de la technologie V2X 802.11p. En plus de la communication directe (V2V, V2I), le C-V2X supporte également la communication longue distance sur un réseau cellulaire (V2N). Le standard LTE-V2X supporte deux interfaces radio :

– l'interface cellulaire Uu ;

– l'interface PC5 en charge des communications V2V. Dans le mode 3, le réseau cellulaire affecte et gère les ressources radio utilisées par les véhicules pour leurs communications. Par contre dans le mode 4, les véhicules sélectionnent de manière autonome les ressources radio.

Fin 2018, la 3GPP a publié la *release 15* : première spécification pour la 5G NR (*New Radio*). La 5G succédera au LTE avec une période transitoire ou le LTE bénéficiera d'améliorations issues de la 5G. L'une des différences radio avec le LTE est une bande de fréquence plus étendue (de 700 MHz à 70 GHz), segmentée en fonction des types d'applications ou des environnements (urbain dense à rural). De plus, l'information sera traitée au plus proche de l'utilisateur final, afin d'améliorer les performances du système. On annonce des latences de moins de 1 ms.

10.4.3. *Communication hybride*

Dans (Mannoni *et al*. 2019), les auteurs après simulation des différents modes de communication (LTE-C-V2X, ITS-G5) montrent que, malgré une portée plus importante et un débit souvent plus élevé pour le mode cellulaire, l'ITS-G5 reste plus performant pour les applications nécessitant une faible latence en tout temps, à savoir les applications de sécurité routière. Toutefois, afin d'augmenter le taux de pénétration des services C-ITS, il peut être pertinent d'utiliser les communications courte et longue portées simultanément pour les applications de divertissement ou d'information. Les communications courte portée restent dédiées aux services liés à la sécurité du transport. Le principe majeur de l'hybridation est de permettre de recevoir, quel que soit le réseau utilisé, la même information. Les messages C-ITS sont les mêmes, quel que soit le médium de communication choisi, et transitent à travers une unité centrale permettant de faire le lien géographique entre toutes les unités cellulaires. Ceci implique que les couches supérieures de la pile C-ITS soient identiques (couches Application, Installation et Sécurité). Les variations sont au niveau de la couche Réseau et transport, et Accès.

10.5. Apport de la radio intelligente dans les réseaux véhiculaires

La radio intelligente (RI), ou encore *Cognitive Radio* en anglais, est la combinaison d'une radio dite « logicielle » avec un module de prises de décision. Cette combinaison permet d'adapter dynamiquement le système radio à son environnement

électromagnétique. Comme déjà mentionné, le concept de radio intelligente a été introduit par J. Mitola en 1999 (Mitola et Maguire 1999). La RI peut être très utile dans le domaine véhiculaire, afin de rester toujours connecté malgré la mobilité et la variation du type de réseau de télécommunication disponible à un instant donné le long du trajet du véhicule. Nous allons tout d'abord décrire les spécificités de la radio intelligente, puis nous présenterons son apport dans les réseaux véhiculaires. Nous poursuivrons par un état de l'art des principaux travaux de recherche dans le domaine de la RI pour les réseaux véhiculaires.

10.5.1. *La radio intelligente*

L'apport principal de la RI est une meilleure gestion du spectre radioélectrique, en exploitant des bandes de fréquences laissées vacantes à un instant donné, grâce à un accès dynamique au spectre. Pour ce faire, la radio intelligente distingue généralement deux catégories d'utilisateur. Les utilisateurs dits *primaires* qui peuvent utiliser les bandes licenciées (bandes de fréquences des opérateurs mobiles, etc.) à tout moment, à l'aide d'un abonnement adéquat. La seconde catégorie d'utilisateurs est celle des utilisateurs *secondaires*. Dans cette catégorie, les utilisateurs ne peuvent utiliser les bandes licenciées qu'à condition qu'elles soient libres au moment de l'utilisation, qu'il n'y ait pas de risque de générer d'interférence et d'être capable de changer de bande dès qu'un utilisateur primaire se présente. Afin de pouvoir accéder au spectre dynamiquement, le système de RI tient compte dans ses prises de décision de la politique d'allocation du spectre définie par les autorités régulatrices. De manière générale, la radio RI peut s'adapter à son environnement grâce à des modules de perception de l'environnement électromagnétique, d'analyse du spectre, de reconnaissance des formes d'ondes, de prises de décision et grâce à la capacité de reconfiguration dynamique des paramètres du système radio (débit, fréquences, modulation, codage, etc.). Le fonctionnement de la radio intelligente suit un cycle connu sous la terminologie « boucle de cognition », illustré figure 10.4.

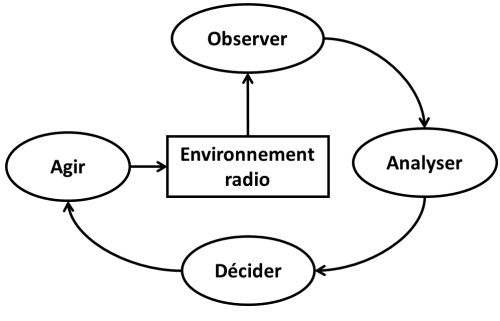


Figure 10.4. *Boucle de cognition*

La radio intelligente est un domaine de recherche très actif. En effet, la capacité d'un système radio à pouvoir analyser son environnement, à choisir les bandes dans lesquelles il peut émettre et à se reconfigurer sont des caractéristiques de plus en plus importantes pour la cohabitation de différents systèmes de télécommunication sans-fil, pour l'optimisation des services en fonction de la mobilité et de la charge d'un réseau, pour le routage de l'information au sein de différents réseaux, etc.

Le lecteur intéressé trouvera plus de détails sur les systèmes de radio intelligente dans (Arslan 2007 ; Doyle 2009 ; Palicot *et al*. 2010).

Dans le cadre du développement de la 5G, le groupe 5GPPP financé par la Commission européenne joue un rôle très important pour la pré standardisation depuis 2015. Dans ce cadre, l'accès dynamique à différentes bandes de fréquences et à différentes technologies d'accès radio sera possible (*extended Dynamic Spectrum Access* (eDSA), MAC *framework*). L'architecture de ce nouveau protocole d'accès multi technologies est en cours de développement et s'appuie sur le LTE-A. Le gestionnaire de ressources radio centralisé (cRRM), le gestionnaire de spectre (*Spectrum Manager*) et l'*Operation and Administration and Management* (OAM) jouent un rôle fondamental, qui pourrait s'apparenter au *Cognitive Manager* que nous introduirons plus tard (5G PPP Architecture Working Group 2017).

10.5.2. *Les CR-VANET*

CR-VANET pour *Cognitive Radio for Vehicular Ad-hoc Network* ou CRAVENET pour *Cognitive Radio Assisted Vehicular Network* est une évolution du réseau véhiculaire *ad hoc*.

Ce type de réseau utilise les capacités de la radio intelligente pour permettre une meilleure gestion du spectre radio, une garantie de connectivité, une amélioration de la bande passante disponible et de la qualité de service.

(Singh *et al*. 2014 ; Eze *et al*. 2017) proposent un aperçu des travaux de recherche sur les CR-VANET. C'est un domaine très actif. Dans (Singh *et al*. 2014), les auteurs proposent une taxonomie des principaux problèmes étudiés dans la littérature. Dans (Eze *et al*. 2017), des travaux dans le domaine du routage, de la couche MAC et de la sécurité complètent la taxonomie. Les travaux relatifs aux simulateurs et aux plateformes d'évaluation sont aussi présentés.

La figure 10.5 illustre les principaux domaines de recherche traités.

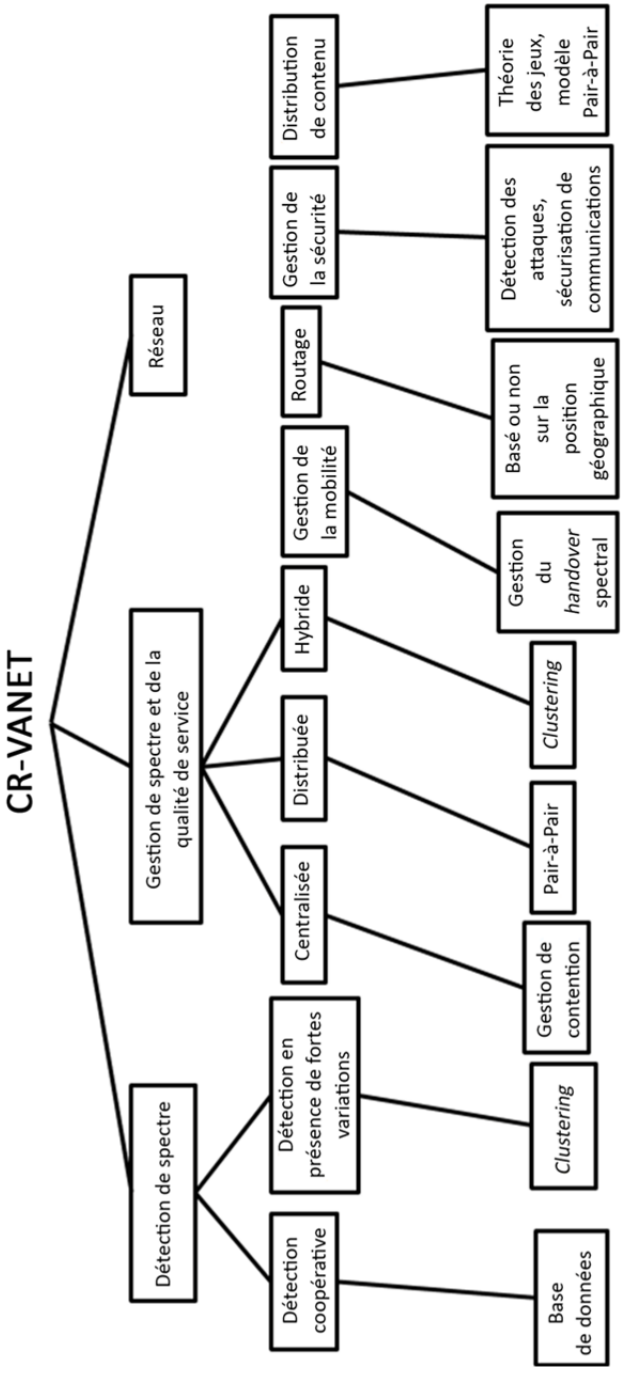


Figure 10.5. *Les principaux problèmes liés à la RI traités dans la littérature (Singh et al. 2014 ; Eze et al. 2017)*

Dans les sections suivantes, nous listons les principaux travaux menés dans les domaines indiqués sur la figure 10.5, à savoir les travaux sur la détection de l'occupation du spectre (*Spectrum Sensing*), l'identification du type de système présent dans la bande analysée, la gestion du spectre, la QoS et les travaux relatifs au réseau.

10.5.2.1. *Détection de l'occupation spectrale*

La problématique principale de la détection de spectre, encore appelée « sondage spectral », est de réaliser la meilleure détection possible de l'occupation du spectre par des utilisateurs primaires, en prenant en compte les caractéristiques véhiculaires. Ces travaux sont très nombreux. De façon générale, on peut distinguer des approches bande étroite ou large bande, des approches coopératives ou non et des approches aveugles ou non. Lorsqu'un utilisateur est détecté dans la bande sondée, il convient alors de détecter le type de modulation, afin de pouvoir identifier le réseau auquel appartient l'utilisateur primaire détecté.

Le domaine du sondage spectral et de la détection de la modulation est un domaine très actif ces dernières années. Un standard a même été proposé, le standard IEEE 802.22 (IEEE 2019). Un certain nombre de travaux se sont intéressés à la détection de l'environnement radio lorsque l'activité des canaux varie fortement au cours du temps, en particulier dans le domaine ferroviaire. Des méthodes aveugles de sondage spectral ont été développées par (Hassan *et al*. 2014), notamment en présence de bruit impulsif, et par (Bouallegue *et al*. 2018). Différentes méthodes de reconnaissance de la modulation ont été proposées dans (Hassan *et al*. 2010 ; Hassan *et al*. 2012 ; Kharbech *et al*. 2013, Kharbech 2018). Cette problématique est importante puisque l'environnement mobile composé par les véhicules fait que l'occupation du spectre peut varier très fréquemment, en particulier en fonction de l'environnement (autoroute, route rurale, etc.).

L'approche coopérative s'appuie sur l'utilisation d'une base de données ou d'un *clustering* comme proposé dans (Li *et al*. 2018). Le résultat de la détection effectuée par plusieurs véhicules est combiné et analysé afin d'attribuer au mieux les fréquences entre les véhicules. Cette approche est d'autant plus intéressante que les bandes de fréquences disponibles peuvent varier en fonction de la position géographique du véhicule. La fusion de données peut s'effectuer soit de manière distribuée entre véhicules ou bien de manière centralisée en passant par une infrastructure telle que les unités de bord de route.

10.5.2.2. *Gestion du spectre et de la QoS*

Dans cette catégorie, les travaux de recherche se concentrent sur les méthodes permettant de gérer au mieux le spectre en respectant une certaine QoS. Les méthodes proposées sont centralisées, distribuées ou bien encore hybrides.

La méthode centralisée consiste à transmettre les différentes informations liées au spectre et à la QoS à une infrastructure qui sera en charge de récolter les informations et de les analyser pour décider du comportement des véhicules.

La méthode distribuée consiste à s'appuyer sur ses pairs et sur soi-même pour gérer au mieux le spectre et la QoS. Un exemple traitant de la distribution en pair-à-pair, et donc de manière distribuée, est présentée dans (Bradai *et al*. 2014). Ici, un contenu vidéo est diffusé entre véhicules en choisissant le meilleur canal disponible à un instant t, en effectuant une analyse dynamique de la qualité des canaux et en comptant sur un réseau pair-à-pair pour la transmission des données. La sélection du canal se base tout d'abord sur le meilleur canal DSRC (*Dedicated Short Range Communications*) disponible. Autrement, la radio intelligente intervient pour choisir le meilleur canal parmi les canaux disponibles n'appartenant pas au DSRC. Le critère de sélection se fait sur la qualité du lien RSSI (*Received Signal Strength Indication*).

Concernant la méthode hybride, telle qu'utilisée dans (Niyato *et al*. 2011), il s'agit de s'appuyer sur une structure comme l'infrastructure routière ainsi que sur les véhicules formant un réseau opportuniste pour gérer au mieux le spectre et la QoS. Dans (Niyato *et al*. 2011) les auteurs cherchent à réduire la charge de l'infrastructure en comptant sur les chefs de *cluster* pour communiquer entre *clusters* et avec l'infrastructure, afin de prendre les meilleures décisions.

10.5.2.3. *Réseau*

Ici, nous retrouvons toutes les problématiques réseau classiques appliquées au véhiculaire, à savoir la gestion de la mobilité, le routage, la distribution de contenu ou bien encore la gestion de la sécurité.

Dans le cadre de la gestion de la mobilité, il s'agit de prendre en compte le *handover* pour garder une connexion active tout en évitant de parasiter les canaux. Ainsi, dans (Kumar *et al*. 2017), les auteurs proposent une solution pour la gestion du *handover* spectral.

Plusieurs méthodes sont proposées pour le routage des messages : basées sur la position géographique, sans connaissance de la position géographique, etc. Ainsi, dans (Usha et Ramakrishnan 2019), les auteurs améliorent l'algorithme MPR OLSR en prenant en compte les canaux inactifs, afin de les utiliser tout en choisissant les nœuds relais avec le plus grand nombre de nouveaux voisins.

Les travaux sur la distribution de contenu s'intéressent en particulier à l'utilisation de la théorie des jeux ou à l'approche pair-à-pair. Ainsi, dans (Tian *et al*.

2019), les auteurs utilisent une approche de jeu évolutif de sorte à attribuer les canaux licenciés aux utilisateurs secondaires sur une base de prix évoluant au cours du temps et d'équilibrage de charge en utilisant ce canal. Concernant l'utilisation du modèle pair-à-pair, nous pouvons citer les travaux de (Bradai *et al*. 2014), qui s'appuient sur la notion de voisinage pour transmettre des fragments de vidéo et permettre ainsi la diffusion de contenu vidéo au sein du réseau.

Enfin, la sécurité réseau et la protection de la vie privée de l'utilisateur constituent un sujet de recherche très important. En effet, les messages de sécurité routière sont essentiels pour le fonctionnement du réseau CR-VANET, tout autant que dans un réseau VANET (Mitra *et al*. 2016 ; Wei *et al*. 2016). Il faut donc s'assurer que ces messages (gestion des accidents, gestion des collisions, etc.) arrivent bien à destination. Dans (Mitra *et al*. 2016), il est question de la détection des attaques de types trous noirs dans le réseau et de leur suppression. Ces attaques consistent en la création d'un nœud virtuel fictif par lequel toutes les communications passeraient, mais ne ressortiraient pas, cela dans le but de perturber les communications entre véhicules. La solution proposée est la détection de ces attaques et, une fois le trou noir repéré, la mise en place d'un chemin alternatif contournant le trou noir et permettant le rétablissement des communications. Dans (Wei *et al*. 2016), les auteurs ont proposé l'utilisation d'un *Cloud* léger en association avec une infrastructure d'unités de bord de routes (UBR) pour permettre la sécurisation des communications dans un CR-VANET. La sécurisation des communications passe aussi par l'utilisation d'un nouveau service, nommé « Spectrum Sensing as a Service » (SaaS), permettant une détection de spectre coopératif en utilisant le Cloud mis en place.

Dans la suite, nous présentons un travail de recherche que nous allons mener dans le domaine de la RI pour les réseaux véhiculaires.

10.6. Projet SERENA : sélection auto-adaptative des technologies d'accès radio en utilisant la radio intelligente

Le projet SERENA est un projet de recherche collaboratif entre le LaBRI (Laboratoire bordelais de recherche en informatique) et l'IFSTTAR (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux). Il bénéficie également du soutien du *cluster* SysNum (Systèmes numériques) de l'Idex Bordeaux (Initiative d'excellence de l'université de Bordeaux). L'objectif général de ce projet, qui va débuter fin 2019, est l'amélioration des communications sur route afin de permettre le déploiement de nouveaux services, et en particulier ceux

nécessaires pour le véhicule autonome. Il s'agit, en particulier, de définir un mécanisme de sélection auto-adaptative des technologies d'accès radio, afin d'être toujours connecté au mieux.

Le mécanisme d'auto-adaptation proposé reposera sur la spécification de nouveaux algorithmes de prises de décision capables de prendre en compte des contraintes variées, telles que la QoS, la sécurité, la consommation d'énergie ou encore les préférences utilisateurs (comme le coût de la communication par exemple). La prédiction de l'évolution probable du niveau de service permettra également d'agir de manière proactive dans le but d'assurer la continuité de service en mobilité.

Le projet SERENA permettra donc d'optimiser et d'améliorer la qualité des échanges de données entre les véhicules et avec l'infrastructure. De plus, l'approche retenue permettra la prise en charge des évolutions des technologies d'accès radio, grâce à l'utilisation de la radio intelligente et de la virtualisation des fonctions réseau.

10.6.1. *Présentation et positionnement*

Le projet SERENA consiste à proposer une solution qui doit permettre au véhicule connecté de sélectionner, de manière autonome et en temps réel, des technologies d'accès radio qui répondent le mieux aux besoins de chaque type de flux de communication (V2I et/ou V2V) et pour des cas d'usage bien précis. Pour ce faire, la solution imaginée s'appuiera sur les concepts de la radio intelligente et sur l'introduction d'un *Cognitive Manager* approprié situé dans le terminal embarqué. Cette solution sera validée d'abord par simulations, en considérant plusieurs cas d'usage se rapportant au véhicule autonome, puis en vraie grandeur nature en considérant des cartes de radio logicielle.

Parmi les travaux de recherche sur la sélection dynamique du meilleur réseau d'accès pour les communications véhiculaires, les travaux de (Singh *et al*. 2014 ; Kumar *et al*. 2017) portent essentiellement sur les communications V2V en utilisant la radio intelligente. Dans le cadre du projet ANR PLATA (Plateforme télématique multistandard programmable pour l'automobile), l'utilisation de systèmes SDR (*Software Defined Radio*) embarqués pour les communications V2V et V2I a été considérée (Haziza *et al*. 2013).

Dans le ferroviaire, une solution sur la couche applicative à base d'intergiciel pour choisir le meilleur réseau sans-fil disponible pour les communications V2I,

en fonction de différents critères, a été proposée (Billion *et al*. 2008). Ce concept est aussi proposé dans (Amanna *et al*. 2010). Dans le projet CORRIDOR (*COgnitive Radio for RaIlway through Dynamic and Opportunistic spectrum Reuse*), des travaux ont été menés pour l'auto-adaptation de la technologie d'accès radio pour des applications de communication V2I pour les trains à grande vitesse (Berbineau *et al*. 2014).

Dans le domaine routier, il n'y a pas, à notre connaissance, aujourd'hui de travaux portant sur l'auto-adaptation de la technologie d'accès radio aux besoins des communications V2X (V2I et V2V) en utilisant la radio intelligence et la virtualisation des fonctions réseau.

10.6.2. *Architecture générale visée*

L'architecture générale du système proposé reposera sur l'utilisation de la radio intelligente et la virtualisation des fonctions réseau.

10.6.2.1. *Boucle de cognition et* Cognitive Manager

Comme mentionné précédemment, la radio intelligente est une technologie qui est bien adaptée à la problématique d'auto-adaptation de la technologie d'accès radio pour répondre aux besoins de communication sur route.

L'objectif est ici d'adapter le fonctionnement du module de communication afin qu'il puisse, d'une part, sélectionner la/les meilleure(s) technologie(s) d'accès en présence de contraintes et de cas d'usage et, d'autre part, adapter ses paramètres de communication en conséquence. Pour cela, une boucle de cognition ainsi qu'un *Cognitive Manager* approprié devront être spécifiés.

Spécification de la boucle de cognition

En utilisant les connaissances obtenues lors des étapes d'analyse de l'environnement électromagnétique, la RI prend des décisions de reconfiguration de manière dynamique, en fonction de certains objectifs prédéfinis, de façon à améliorer l'efficacité de l'utilisation du spectre sans intervention humaine (boucle de cognition).

Dans le cadre du projet SERENA, la prise de décision pourra aboutir, par exemple, à l'arrêt de certains flux moins prioritaires afin de protéger des flux plus critiques, ou encore à une modification du niveau de sécurité des communications. Il conviendra donc de spécifier la boucle de cognition pour qu'elle puisse répondre à la problématique du projet SERENA telle qu'illustrée figure 10.6.

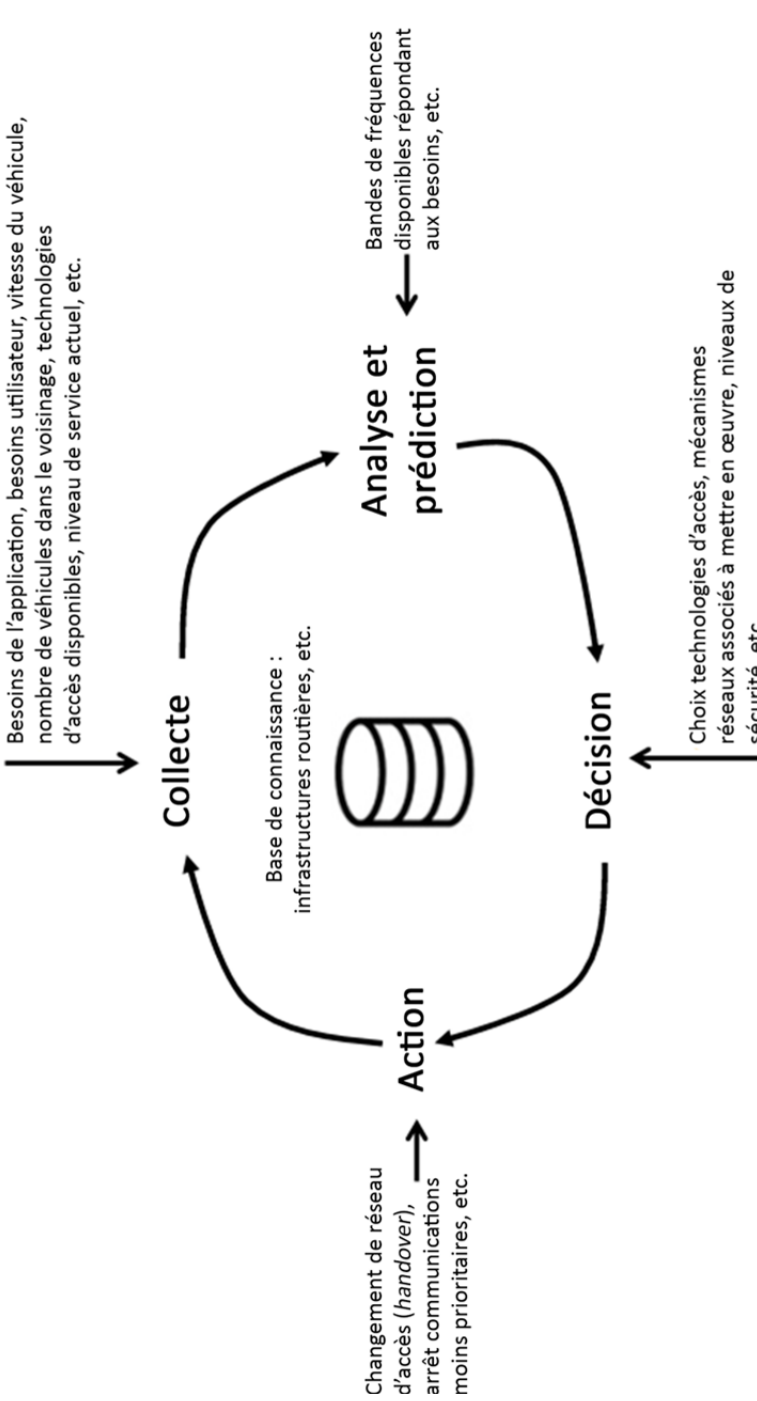


Figure 10.6. *Exemple de boucle de cognition du projet SERENA*

Cette boucle de cognition, comme déjà indiqué, comprend quatre grandes étapes :

– collecte (observation) : c'est la phase de collecte des données qui seront utiles à la prise de décision. Comme données utiles, nous pouvons citer : la vitesse, la position du véhicule, l'environnement radio (les réseaux cellulaires présents dans l'environnement). Les données statiques comme les contraintes de QoS associées à un type d'application sont préalablement stockées dans la base de connaissances ;

– analyse et prédiction : il s'agit d'effectuer une analyse des données récoltées en s'aidant de la base de connaissances. On pourra faire appel aux techniques de *Machine Learning* pour prédire, par exemple, le temps de disponibilité des différentes bandes de fréquence. Un premier filtrage pourra ainsi être réalisé en ne tenant compte que des bandes de fréquence susceptibles de répondre aux besoins de communication ;

– décision : cette étape permet de sélectionner la technologie radio et les protocoles de communication à utiliser, ou encore le niveau de sécurité à appliquer en s'appuyant sur les données issues de l'étape précédente. La décision devra tenir compte de la prévision à court terme de la disponibilité du lien entre deux véhicules ou de la disponibilité de l'infrastructure de télécommunications, afin d'assurer la continuité de service. La prise en compte du futur proche pourrait s'appuyer sur des informations sur le mouvement du véhicule, sur les mouvements des véhicules voisins ou encore sur la connaissance de l'infrastructure (informations contenues dans la base de connaissances) ;

– action : cette étape permet de réaliser un ensemble d'actions telles que reconfigurer la radio (changement de bande de fréquence, modification éventuelle des paramètres de transmission, etc.) conformément aux caractéristiques de la technologie d'accès retenue lors de l'étape précédente, arrêter les communications non prioritaires pour bénéficier de plus de bande passante, augmenter le niveau de sécurité, etc.

Spécification du *Cognitive Manager*

Le *Cognitive Manager* utilisera différents moteurs de l'intelligence (Ouattara 2014) pour mettre en œuvre la boucle de cognition, tels que les moteurs de mesure, d'analyse, de raisonnement, d'apprentissage, de prises de décision ou encore d'adaptation. Le contenu de chacun des différents moteurs nécessaires à la réalisation de la boucle de cognition du projet SERENA devra par conséquent être spécifié. Il conviendra également de spécifier le contenu de la base de connaissances. Celle-ci contiendra, d'une part, des informations statiques telles que le placement des

unités de bord de route ou encore les réseaux présents dans la zone et, d'autre part, des informations calculées ou prédites provenant des moteurs de l'intelligence.

10.6.2.2. *Virtualisation des fonctions réseau*

La virtualisation des fonctions réseau ou *Network Function Virtualization* (NFV) constitue aujourd'hui une évolution majeure des réseaux. Couplée à la technique SDN (*Software Defined Networking*), elle offre de nombreux avantages aux opérateurs réseaux, en particulier un déploiement rapide, une plus grande flexibilité et une meilleure adaptation au contexte. L'introduction de ces deux technologies est en train de transformer en profondeur les réseaux et permettra d'accélérer l'introduction de la voiture autonome (Mendiboure *et al*. 2019). En particulier, la 5G s'appuie sur SDN/NFV pour mettre en œuvre le *network slicing*, c'est-à-dire le découpage logique du réseau, afin d'être capable de prendre en charge des catégories de services diversifiés tels que les services ITS à latence faible et à haute disponibilité/fiabilité. De nombreux défis demeurent, tels que l'orchestration des *slices*, la gestion des fonctions réseau ou encore la sécurité des réseaux SDN (Foukas *et al*. 2017).

La durée de vie d'un véhicule est relativement longue (plusieurs années) au regard de l'évolution actuelle des technologies radio. Ainsi, plusieurs nouvelles technologies de communication peuvent voir le jour pendant ce laps de temps. L'approche de virtualisation des fonctions réseau constitue un atout important pour le projet SERENA en permettant de pérenniser la solution retenue vis-à-vis des évolutions technologiques, en particulier des nouveaux moyens d'accès radio. En effet, en virtualisant le module de communication du véhicule, il sera possible d'ajouter facilement de nouvelles fonctionnalités de communication par ajout de fonctions logicielles. Cette approche est d'autant plus facilitée que les communications radio s'appuient dans le cadre de ce projet sur la radio définie de manière logicielle (SDR).

10.6.3. *Principaux enjeux*

Le projet SERENA soulève plusieurs défis, tels que l'identification et la sélection en temps réel de la technologie d'accès dans un environnement fortement changeant (variation de la vitesse du véhicule, de la densité de connectivité, du nombre de réseaux d'accès disponibles, etc.), la capacité à supporter plusieurs technologies d'accès simultanées pour une même communication (pour sécuriser le lien radio dans le cas du véhicule autonome par exemple), la surveillance en temps réel du niveau de service offert et la prise de décision proactive afin d'assurer le maintien du niveau de service requis.

Le principal risque est d'obtenir des performances insuffisantes pour répondre à des besoins de forte mobilité. Cependant, la vitesse maximum autorisée de la voiture autonome est aujourd'hui limitée à quelques dizaines de kilomètres par heure.

Le second risque concerne les difficultés d'intégration actuelles de la radio intelligente avec les technologies d'accès qui seront retenues. Ce risque est cependant amené à disparaître au fur et à mesure du déploiement des solutions NFV (OPNGV)[2] et de la virtualisation des protocoles d'accès en particulier (Riggio *et al*. 2016).

10.7. Conclusion

Le véhicule autonome est en plein développement et de nouvelles applications apparaissent. Ces applications ont des besoins de communication spécifiques, que ce soit en termes de latence, de portée ou de bande passante. Pour que ces applications fonctionnent au mieux, il convient de sélectionner la technologie d'accès radio la mieux adaptée à leurs besoins et donc d'étudier les avantages et inconvénients des architectures sur lesquelles ces technologies d'accès reposent.

Le spectre radio étant mal exploité, l'utilisation de la radio intelligente permet d'apporter une solution aux besoins de continuité de communication dans l'environnement véhiculaire. Dans ce chapitre, nous avons illustré les principaux apports de la radio intelligente dans le domaine véhiculaire, tels que l'amélioration de la couverture réseau ou bien encore l'amélioration de la bande passante disponible.

Dans les CR-VANET, la radio intelligente peut apporter son lot d'enjeux. Par exemple, dans (Singh *et al*. 2014), les enjeux peuvent être de l'ordre de la fiabilité de la communication des messages d'urgences, de la gestion d'une topologie dynamique, de la rareté des bandes disponibles, de la gestion des nouvelles technologies de communication ainsi que de l'adaptation aux différents environnements (autoroute, ville, rural, etc.) et à la répartition du spectre.

Dans le projet SERENA, la radio intelligence associée à la virtualisation (NFV) et à la reconfiguration (SDN) des fonctions réseau permettra au module de communication du véhicule de sélectionner la meilleure technologie d'accès radio, d'adapter ses paramètres radio et d'être résilient aux évolutions technologiques.

2. OPNFV (*Open Platform for Network Function Virtualization*), disponible à l'adresse : https://www.opnfv.org/.

L'utilisation des outils de l'IA comme outils d'aide à la décision va de pair avec le développement du véhicule autonome et les algorithmes de *Machine Learning*, en particulier, vont continuer à connaître un engouement important dans ce domaine.

10.8. Bibliographie

5G PPP Architecture Working Group (2017). View of 5G Architecture (version 2.0) [En ligne]. Disponible à l'adresse : https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017_For-Public-Consultation.pdf.

Amanna, A., Gadniok, M., Price, M.J., Reed, J.H., Siriwongpairat, W.P., Himsoon, T.K. (2010). Railway Cognitive Radio. *IEEE Vehicular Technology Magazine*, 5(3), 82–89.

Amjad, Z., Sikora, A., Hilt, B., Lauffenburger, J.-P. (2018). Low Latency V2X Applications and Network Requirements Performance Evaluation. Dans *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Changsu.

Arslan, H. (2007). *Cognitive Radio, Software Defined radio, and Adaptative Wireless Systems*. Springer, Dordrecht.

Berbineau, M. *et al*. (2014). Cognitive Radio for High Speed Railway through Dynamic and Opportunistic Spectrum Reuse. Dans *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*. TRA, Paris, 1–10.

Billion, J., Van den Abeele, D., Gransart, C., Berbineau, M. (2008). ICOM: Toward Integrated Communications for Global Railway Systems. Dans *Word Congress on Railway Research*. WCRR, Séoul.

Bouallegue, K., Dayoub, I., Gharbi, M., Hassan, K. (2018). Blind Spectrum Sensing Using Extreme Eigenvalues for Cognitive Radio Networks. *IEEE Communications Letters*, 22(7), 1386–1389.

Bradai, A., Ahmed, T., Benslimane, A. (2014). ViCoV: Efficient video streaming for cognitive radio VANET. *Vehicular Communications*, 1(3), 105–122.

Campolo, C., Molinaro, A., Iera, A., Menichella, F. (2017). 5G Network Slicing For Vehicle-To-Everything Services. *IEEE Wireless Communications*, décembre.

Cunha, F.D. *et al*. (2016). Data Communication in VANETs: Survey, Applications and Challenges. *Ad Hoc Networks*, 44, 90–103.

Dar, K., Bakhouya, M., Gaber, J., Wack, M. (2010). Wireless Communication Technologies for ITS Applications. *IEEE Communications Magazine*, mai, 156–162.

Doyle, L. (2009). *Essentials of Cognitive Radio*. Cambridge University Press, Cambridge.

ETSI (2010). Intelligent Transport Systems (ITS) ; Communications Architecture. Document, ETSI, Sophia Antipolis.

ETSI (2014). Vehicular Communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 1: Media-Independent Functionality. Document, ETSI, Sophia Antipolis.

Eze, J., Zhang, S., Liu, E., Eze, E. (2017). Cognitive Radio Technology assisted Vehicular AdHoc Networks (VANETs): Current Status Challenges, and Research Trends. Dans *23rd International Conference on Automation and Computing*. ICAC, Huddersfield, 1–6.

Foukas, X., Elmokashfi, A., Patounas, G., Marina, M.K. (2017). Network Slicing in 5G: Survey and Challenges. *IEEE Communications Magazine*, février.

Hassan, K., Dayoub, I., Hamouda, W., Berbineau, M. (2010). Automatic modulation recognition using wavelet transform and neural networks in wireless system. *EURASIP Journal on Advances in Signal Processing*.

Hassan, K., Dayoub, I, Hamouda, W., Nzeza, C.N., Berbineau, M. (2012). Blind digital modulation identification for spatially-correlated MIMO systems. *IEEE Transactions on Wireless Communications*, 11, 683–693.

Hassan, K., Gautier, R., Dayoub, I., Berbineau, M., Radoi, E. (2014). Multiple-Antenna-Based Blind Spectrum Sensing in the Presence of Impulsive Noise. *IEEE Transactions on Vehicular Technology*, 63(5), 2248–2257.

Haziza, N., Kassab, M., Knopp, R., Harri, J., Kaltenberger, F. *et al*. (2013). Multi-technology vehicular cooperative system based on Software Defined Radio (SDR). Dans *Fifth workshop on Communication Technologies for Vehicles*. Nets4cars, Vilnius, 84–95.

Hubaux, J.-P. (2005). Vehicular Networks: How to Secure Them. Dans *MiNEMA Summer School*. MiNEMA, Klagenfurt.

IEEE (2010). 802.11p. Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments. Document, IEEE.

IEEE (2019). IEEE 802.22. Draft Standard for Information Technology – Local and Metropolitan Area Networks – Specific Requirements – Part 22: Cognitive Radio Wireless Regional Area Networks (WRAN) Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the Bands that Allow Spectrum Sharing where the Communications Devices may Opportunistically Operate in the Spectrum of the Primary Service. Document, IEEE.

Kharbech, S., Dayoub, I., Simon, E., Zwingelstein-Colin, M. (2013). Blind digital modulation detector for MIMO systems over high-speed railway channels. *Communication Technologies for Vehicles*, 232–241.

Kharbech, S., Dayoub, I., Zwingelstein-Colin, M., Simon, E.P. (2018). Blind Digital Modulation Identification for MIMO Systems in Railway Environments With High-Speed Channels and Impulsive Noise. *IEEE Transactions on Vehicular Technology*, 67(8), 7370–7379.

Kumar, K., Prakash, A., Tripathi, R. (2017). A Spectrum Handoff Scheme for Optimal Network Selection in NEMO Based Cognitive Radio Vehicular Networks. *Wireless Communications and Mobile Computing*, 2017.

Li, X., Song, T., Zhang, Y., Chen, G., Hu, J. (2018). A Hybrid Cooperative Spectrum Sensing Scheme Based on spatial-Temporal Correlation for CR-VANET. Dans *IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, Porto.

MacHardy, Z., Khan, A., Obana, K., Iwashina, S. (2018). V2X Access Technologies: Regulation, Research, and Remaining Challenges. *IEEE Communications Surveys & Tutorials*, 20(3), 1858–1877.

Mannoni, V., Berg, V., Sesia, S., Perraud, E. (2019). A Comparison of the V2X Communication Systems: ITS-G5 and C-V2X. Dans *IEEE Vehicular Technology Conference (VTC Spring)*. IEEE, Kuala-Lumpur.

Mendiboure, L., Chalouf, M.A., Krief, F. (2019). Edge computing based applications in vehicular environments: Comparative study and main issues. *Journal of Computer Science and Technology*, 34(4), 869–886.

Mir, Z.H., Filali, F. (2018). Applications, Requirements, and Design Guidelines for Multi-tiered Vehicular Network Architecture. Dans *10th Wireless Days Conference*. IEEE, Dubai.

Mitola, J., Maguire, G.Q. (1999). Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6(4), 13–18.

Mitra, S., Jana, B., Poray, J. (2016). A novel scheme to detect and remove black hole attack in cognitive radio vehicular ad-hoc networks (CR-VANETs). Dans *International Conference on Computer, Electrical & Communication Engineering*. ICCECE, Kolkata.

Niyato, D., Hossain, E., Wang, P. (2011). Optimal channel access management with QoS support for cognitive vehicular networks. *IEEE Transactions on Mobile Computing*, 10(4), 573–591.

Ouattara, D. (2014). Apport des réseaux intelligents aux usages et pratiques en e-santé : Une architecture flexible basée sur la technologie radio cognitive pour un suivi efficace et temps réel des patients. Thèse de doctorat, Université de Bordeaux, Bordeaux.

Palicot, J. *et al*. (2010). *De la radio logicielle à la radio intelligente*. Hermes Sciences-Lavoisier, Paris.

Riggio, R., Bradai, A., Harutyunyan, D., Rasheed, T., Ahmed, T. (2016). Scheduling Wireless Virtual Networks Functions. *IEEE Transactions on Network and Service Management*.

Shladover, S.E., Lappin, J., Denaro, R.P., Smith, B.W. (2014). Introduction: The Transportation Research Board's 2013 Workshop on Road Vehicle Automation. Dans *Road Vehicle Automation*, Meyer, G., Beiker, S. (dir.). Springer, Cham.

Singh, K.D., Rawat, P., Bonnin, J.-M. (2014). Cognitive radio for vehicular ad hoc networks (CR-VANETs): approaches and challenges. *EURASIP Journal on Wireless Communications and Networking*.

Tian, D., Zhou, J., Wang, Y., Sheng, Z., Duan, X., Leung, V.C.M. (2019). Channel Access Optimization with Adaptive Congestion Pricing for Cognitive Vehicular Networks: An Evolutionary Game Approach. *IEEE Transactions on Mobile Computing*, février.

Usha, M., Ramakrishnan, B. (2019). An enhanced MPR OLSR Protocol for Efficient Node Selection Process in Cognitive Radio Based VANET. *Wireless Personal Communications*, février.

Wei, Z., Yu, F.R., Tang, H., Liang, C., Yan, Q. (2016). Securing cognitive radio vehicular Ad hoc networks with trusted lightweight cloud computing. Dans *IEEE Conference on Communication and Network Security*. IEEE, Philadelphie.

Liste des auteurs

Asma AMRAOUI
Université Abou Bekr Belkaid
Tlemcen
Algérie

Hasnaâ ANISS
IFSTTAR
Bordeaux

Mohammed Anis BENBLIDIA
Université de technologie de Troyes

Fayssal BENDAOUD
ESI-SBA
Sidi Bel Abbès
Algérie

Badr BENMAMMAR
Université Abou Bekr Belkaid
Tlemcen
Algérie

Marion BERBINEAU
IFSTTAR
Bordeaux

Bouziane BRIK
CESI
Rouen

Mohamed Aymen CHALOUF
IRISA
Rennes

Moez ESSEGHIR
Université de technologie de Troyes

Omessaad HAMDI
IMT Atlantique
Rennes

Francine KRIEF
ENSEIRB-MATMECA
Bordeaux

Killian LE PAGE
ALTEN
Boulogne-Billancourt

Maïssa MBAYE
Université Gaston Berger
Saint-Louis
Sénégal

Léo MENDIBOURE
LaBRI
Bordeaux

Leila MERGHEM-BOULAHIA
Université de technologie de Troyes

Zeinab MOVAHEDI
Iran University of Science
and Technology
Téhéran
Iran

Abderrazaq SEMMOUD
Université Abou Bekr Belkaid
Tlemcen
Algérie

Ahmed Khalid Yassine SETTOUTI
Université Abou Bekr Belkaid
Tlemcen
Algérie

Index

E

F, G

H

I

K

L

M

N, O

P

Q, R

S

T

V

W, Z

La gestion et le contrôle des réseaux ne peuvent plus être envisagés sans l'introduction de l'intelligence artificielle dans l'ensemble de leurs étapes.

Cet ouvrage traite des thèmes d'actualité qui sont liés principalement à la sécurité intelligente des réseaux informatiques, au déploiement de services de sécurité dans les réseaux SDN (*Software-Defined Networking*), à l'optimisation des réseaux à l'aide des techniques de l'intelligence artificielle et aux méthodes d'optimisation multicritères pour la sélection des réseaux dans un environnement hétérogène.

Gestion et contrôle intelligents des réseaux s'intéresse également à la sélection des services *Cloud Computing*, au déchargement intelligent des calculs dans le contexte du *Mobile Cloud Computing*, à la gestion intelligente des ressources dans un système *Smart Grid-Cloud* pour une meilleure efficacité énergétique, à l'*Internet of Vehicles* (IoV) en se basant sur ses nouvelles architectures, à l'application de l'intelligence artificielle dans les réseaux de radio cognitive et à l'apport de la radio intelligente pour répondre aux besoins de communication sur route des véhicules autonomes.

Le coordonnateur

Badr Benmammar est professeur d'informatique à l'Université Abou Bekr Belkaid Tlemcen (Algérie) et chercheur au Laboratoire de télécommunications Tlemcen (LTT).

iSTE editions

9 781789 480085