# APPROVISIONNEMENT AXÉ SUR LE PROFIT DES CHAINES DE SERVICE RÉSEAU

Par

Walid RACHEG

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE LOGICIEL
M. Sc. A.

MONTRÉAL, LE 1er AOÛT 2017

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

**PRÉSENTATION DU JURY**

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Mohamed Faten Zhani, directeur de mémoire
Département de génie logiciel et TI à l'École de technologie supérieure

M. Abdelouahed Gherbi, président du jury
Département de génie logiciel et TI à l'École de technologie supérieure

Mme Halima Elbiaze, examinateur externe
Département d'informatique à l'Université du Québec à Montréal

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 27 JUILLET 2017

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

**REMERCIEMENTS**

Je tiens tout d'abord à exprimer ma gratitude et mes vifs remerciements à mon directeur de recherche, le professeur Mohamed Faten Zhani pour sa grande qualité d'encadrement, sa disponibilité, son dévouement, ses encouragements et ses conseils fort judicieux. Ses qualités humaines et scientifiques ont largement contribué au bon déroulement de ce mémoire. Je tiens à remercier les membres du jury d'avoir acceptés d'évaluer mon mémoire.

Je remercie tous les membres de ma famille qui m'ont apporté leur soutien et leur aide au cours de ces années d'études, et plus spécialement ma chère mère.

J'adresse aussi mes vifs remerciements à tous mes enseignants à l'école de technologie supérieure pour le savoir et les connaissances qu'ils m'ont inculqués tout au long de ces années d'études.

Je tiens également à remercier mes collègues au laboratoire LASI pour leur soutien technique et moral.

Je remercie aussi tous ceux qui ont participé de près ou de loin à l'élaboration de ce travail. Qu'ils trouvent dans ce travail l'expression de ma profonde gratitude.

# Approvisionnement Axé sur le Profit des Chaines de Service Réseau

Walid RACHEG

## RÉSUMÉ

La virtualisation des fonctions réseau (*Network Function Virtualization* - NFV) est un paradigme émergeant qui est en train de transformer la manière avec laquelle les services réseau sont approvisionnés et gérés. L'idée principale du NFV est de découpler les fonctions réseau des équipements réseau qui les exécutent. Ainsi, un service réseau peut être approvisionné à la demande comme étant une chaine de fonctions réseau virtuelles. Cela permettrait d'améliorer la flexibilité et l'évolutivité des services réseau et éventuellement de réduire les coûts de déploiement. Dans ce contexte, l'un des principaux défis des fournisseurs de nuage qui restent à résoudre est d'allouer efficacement les ressources pour les services réseau de manière à réduire les coûts opérationnels et qui maximise leurs profits. Dans ce travail, nous abordons ce défi en proposant un système d'approvisionnement de service réseaux conçu pour les infrastructures à grande échelle couvrant différents sites géographiquement distribués. Nous proposons trois algorithmes qui maximisent le profit du fournisseur de nuage en tenant compte de la consommation d'énergie de l'infrastructure et de la variabilité des prix de l'énergie dans les différentes régions. Nous montrons ensuite grâce à des simulations que ces algorithmes sont capables de trouver efficacement des allocations de ressources quasi-optimales avec un minimum de complexité de calcul et de maximiser le profit du fournisseur.

**Mots-clés :** infonuagique, virtualisation des fonctions réseau, efficacité énergétique, profit.

**PROFIT-DRIVEN RESOURCE PROVISIONING IN NFV-BASED ENVIRONMENTS**

Walid RACHEG

## ABSTRACT

Network Function Virtualization (NFV) is an emergent paradigm that is currently transforming the way network services are provisioned and managed. The main idea of NFV is to decouple network functions from the hardware running them. This allows to reduce deployment costs and to further improve the flexibility and the scalability of network services. Despite these benefits, a major challenge cloud providers are still facing is how to efficiently allocate resources for NFV-based services in a way that reduces operational costs and maximizes their profits. In this thesis, we address this particular challenge and propose an effective profit-driven service chain provisioning scheme designed for large-scale infrastructures spanning different geographically distributed sites. We hence propose three algorithms that maximize the provider's profit taking into consideration energy consumption of the infrastructure and the variability of energy prices in different locations. Through extensive simulations, we show that these algorithms are able to efficiently find near-optimal resource allocations and maximize the provider's profit with minimal computational complexity.

**Keywords:** cloud computing, network function virtualization, energy efficiency, cloud provider's profit.

## TABLE DES MATIÈRES

## LISTE DES TABLEAUX

Page

## LISTE DES FIGURES

Page

## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

| | |
|---|---|
| CPU | Central Processing Unit (Processeur) |
| DNS | Domain Name System |
| IaaS | Infrastructure as a Service |
| MV | Machine Virtuelle |
| NAT | Network Address Translation |
| NF | Network Function |
| NFV | Network Function Virtualisation |
| PaaS | Platform as a Service |
| POP | Point Of Presence |
| QoS | Quality of Service |
| SaaS | Software as a Service |
| SLA | Service Level Agreement |
| VNF | Virtual Network Function |

# INTRODUCTION

Avec l'apparition des services infonuagiques et la diversité des applications réseau, les réseaux informatiques sont appelés à transmettre une énorme quantité de données qui ne cesse de croitre et à offrir de nouveaux services et fonctionnalités réseau capable de s'adapter aux différentes exigences en matière de performance, sécurité et évolutivité. Malheureusement, les architectures réseau traditionnelles sont statiques, peu flexibles et centrées principalement sur les équipements hardware qui ne sont pas faciles à remplacer ou à faire évoluer. Ainsi, elles sont stagnantes et incapables d'offrir la flexibilité et l'agilité requises pour créer rapidement de nouveaux services réseau et de les adapter aux exigences des différentes applications.

La virtualisation des fonctions réseau est un nouveau concept, récemment apparu, qui permet de pallier aux limitations des réseaux traditionnels et de créer des infrastructures et des services réseau plus dynamiques et programmables. Ce nouveau concept exploite la technologie de virtualisation pour offrir les fonctions des équipements réseau (par ex., routeurs, équilibreur de charge, pare-feu) comme des composantes logicielles appelées des fonctions réseau virtuelles (*Virtual Network Function* – VNF) (Lee, Pack, Shin, & Paik, 2014). En d'autres termes, la virtualisation des fonctions réseau consiste à remplacer les équipements réseau dédiés (implémentés en hardware) par des serveurs standards puissant et moins coûteux où il sera possible de créer au besoin des fonctions réseau sans avoir à faire des changements au niveau des équipements. Ces fonctions peuvent être ensuite connectés et enchaînés pour construire ce qu'on appelle une chaîne de service réseau qui sera traversé par les paquets transmis dans le réseau.

La virtualisation des fonctions réseau apporte plusieurs avantages en termes de coûts, de flexibilité et d'évolutivité par rapport aux plates-formes de réseau existants où les fonctions réseau sont implémentées dans des équipements physiques dédiés qui sont généralement chers et très difficile à installer ou à remplacer. En effet, en termes de flexibilité et de performance, le VNF permet d'instancier, gérer et dimensionner les fonctions

réseau à la demande et en fonction du besoin et de la quantité du trafic à traiter. Il permet aussi de réduire les coûts opérationnels et la consommation d'énergie en hébergeant plusieurs fonctions virtuelles dans la même machine. Ainsi, il permet d'approvisionner dynamiquement des services réseau et de les redimensionner en fonction de la demande et des besoins actuels et d'optimiser davantage l'utilisation des ressources de l'infrastructure.

De plus, la virtualisation des fonctions réseau permet la location multiple où le propriétaire d'une infrastructure réseau (appelé ci-après fournisseur de VNFs ou de nuage) est capable de louer ses ressources à plusieurs fournisseurs de service. Ainsi, un fournisseur de service peut demander une chaine de service réseau auprès d'un fournisseur de VNFs. Le fournisseur de VNFs s'assurera d'approvisionner toutes les ressources nécessaires (cpu, mémoire et disque, et bande passante) pour chacune des fonctions virtuelles composants la chaine.

L'un des principaux défis auxquels sont confrontés les fournisseurs de VNFs consiste à maximiser leurs profits lors de l'allocation des ressources pour les chaînes de services demandées. Cela nécessite de trouver l'emplacement le plus approprié pour chaque fonction de réseau virtuel composant la chaîne de service tout en satisfaisant aux exigences de service en termes de chaînage (c.-à.-d., l'ordre des VNFs demandés), de bande passante, de délai de transmission et de quantité de ressources (CPU, mémoire et disque). Le problème devient encore plus difficile lorsque ces VNFs sont approvisionnés dans des infrastructures avec des capacités de ressources différentes et qui sont réparties dans divers endroits ayant des prix d'électricité différents.

## 0.1 Problématique

Le principal défi des fournisseurs de VNFs est de maximiser leurs profits lorsqu'ils allouent les ressources pour les chaînes de service demandées par les utilisateurs. Ce défi consiste à trouver l'emplacement le plus approprié pour chacune des fonctions réseau virtuelle (VNF) composants la chaîne de service tout en satisfaisant aux exigences en termes de délai.

Une chaîne de service est composée d'une séquence ordonnée de différents types de VNFs tels que des équilibreurs de charge, des pare-feu ou des systèmes de détection d'intrusion. La chaîne (également appelée requête de service) possède une source et une destination qui correspondent respectivement au POP (point de présence ; constituer d'un ensemble de serveurs connectés entre eux par des switches) qui déclenche le trafic et qui reçoit le trafic après avoir traversé la chaîne de service. La requête de service est également caractérisée par un délai de bout en bout qui désigne le temps total nécessaire pour que le trafic circule dans les VNFs. Il s'agit d'une exigence de performance spécifiée par l'utilisateur et doit être prise en compte lors de l'approvisionnement du service. De plus, chacun des VNFs composant le service a des exigences de CPU, de mémoire et de stockage qui doivent être satisfaites afin d'assurer un temps de traitement minimal du trafic pour chaque fonction. Enfin, les VNFs sont connectés par des liaisons virtuelles caractérisées par une exigence de bande passante fournie par l'utilisateur. Idéalement, la bande passante devrait être suffisante pour transporter la quantité de trafic qui doit traverser la chaîne de service.

Une fois que l'utilisateur définit la chaîne de service demandée et ses besoins en ressources. Le fournisseur de VNFs est responsable de placer les VNFs, de les relier et d'allouer les ressources nécessaires. Habituellement, un fournisseur de VNFs construit une infrastructure composée de plusieurs points de présence (*Point of Presence* – POP) qui sont répartis géographiquement sur plusieurs sites. Chaque POP contient un ensemble de serveurs physiques et au moins un routeur pour le connecter aux POPs voisins. Naturellement, le prix de l'électricité varie entre les POPs selon sa position géographique et les fournisseurs d'énergie locaux (Zhang, Zhu, Zhani, Boutaba, & Hellerstein, 2013). Les modèles de consommation d'énergie (c'est-à-dire la quantité d'énergie consommée en fonction des ressources utilisées) varient également d'un POP à un autre selon le type de serveurs utilisés.

Ce travail vise à trouver des solutions pour le problème d'allocation de ressources pour les chaines de service. En d'autres termes, il vise à trouver l'emplacement le plus approprié de VNFs de la chaîne de sorte que le profit du fournisseur de VNFs est maximisé.

Les solutions proposées dans ce mémoire permettent d'atteindre cet objectif en réduisant les coûts énergétiques de l'infrastructure tout en tenant compte des modèles de consommation de l'énergie des POPs et des différents prix de l'électricité dans leurs emplacements.

## 0.2 Objectif et méthodologie

L'objectif principal de ce mémoire est de proposer des solutions performantes permettant d'allouer efficacement les ressources des chaines de service afin de maximiser le profit des fournisseurs de VNFs tout en satisfaisant aux exigences de service en termes de quantité de ressources (cpu, mémoire, disque et bande passante) et de performance (notamment en termes de délai de bout en bout). Il faudrait aussi que ces solutions prennent en considération les prix de l'électricité dans les différents POPs et les modèles de consommation d'énergie des équipements utilisés.

Pour atteindre cet objectif, nous allons procéder comme suit :

- Nous allons commencer par décrire les concepts de base liés à l'infonuagique, et à la virtualisation des fonctions du réseau. Ensuite, nous allons établir l'état de l'art des solutions existantes pour le problème du placement et de chaînage de VNFs et analyser leurs limites.

- Nous allons ensuite modéliser mathématiquement le problème étudié et proposer des solutions qui permettront de trouver les emplacements optimaux pour les VNFs et les chaines de service.

- Finalement, nous allons comparer les solutions proposées et les tester à travers des simulations afin d'évaluer leurs performances en termes de taux d'acceptation des chaines de service, l'utilisation des ressources (mémoire, cpu, disque, et bande passante), délai de bout en bout, coût, profit du fournisseur et de temps d'exécution de l'algorithme.

## 0.3 Publications

Le travail présenté dans ce mémoire a fait l'objet d'une publication d'un article de conférence qui s'intitule : « *Profit-Driven Resource Provisioning in NFV-based Environments* » et qui a été publiée à IEEE *International Conference on Communications (ICC)* en mai 2017 à Paris en France.

## 0.4 Plan du mémoire

Le reste de ce rapport est organisé comme suit. Le premier chapitre fournit les notions de base reliées à l'infonuagique et la virtualisation des fonctions réseau, résume les travaux existants sur le problème du placement et du chaînage des VNFs et discute leurs limitations.

Dans le deuxième chapitre, nous fournissons la formulation mathématique du problème étudié et nous présentons les détails des solutions proposées.

Le troisième chapitre présente les environnements de simulation et les détails sur les différentes topologies et chaines de services simulées ainsi que les principaux résultats expérimentaux.

Le dernier chapitre présente quelques conclusions générales et identifie des perspectives pour des travaux futurs.

La Figure 0.1 montre une représentation graphique du contenu des différents chapitres de ce rapport.

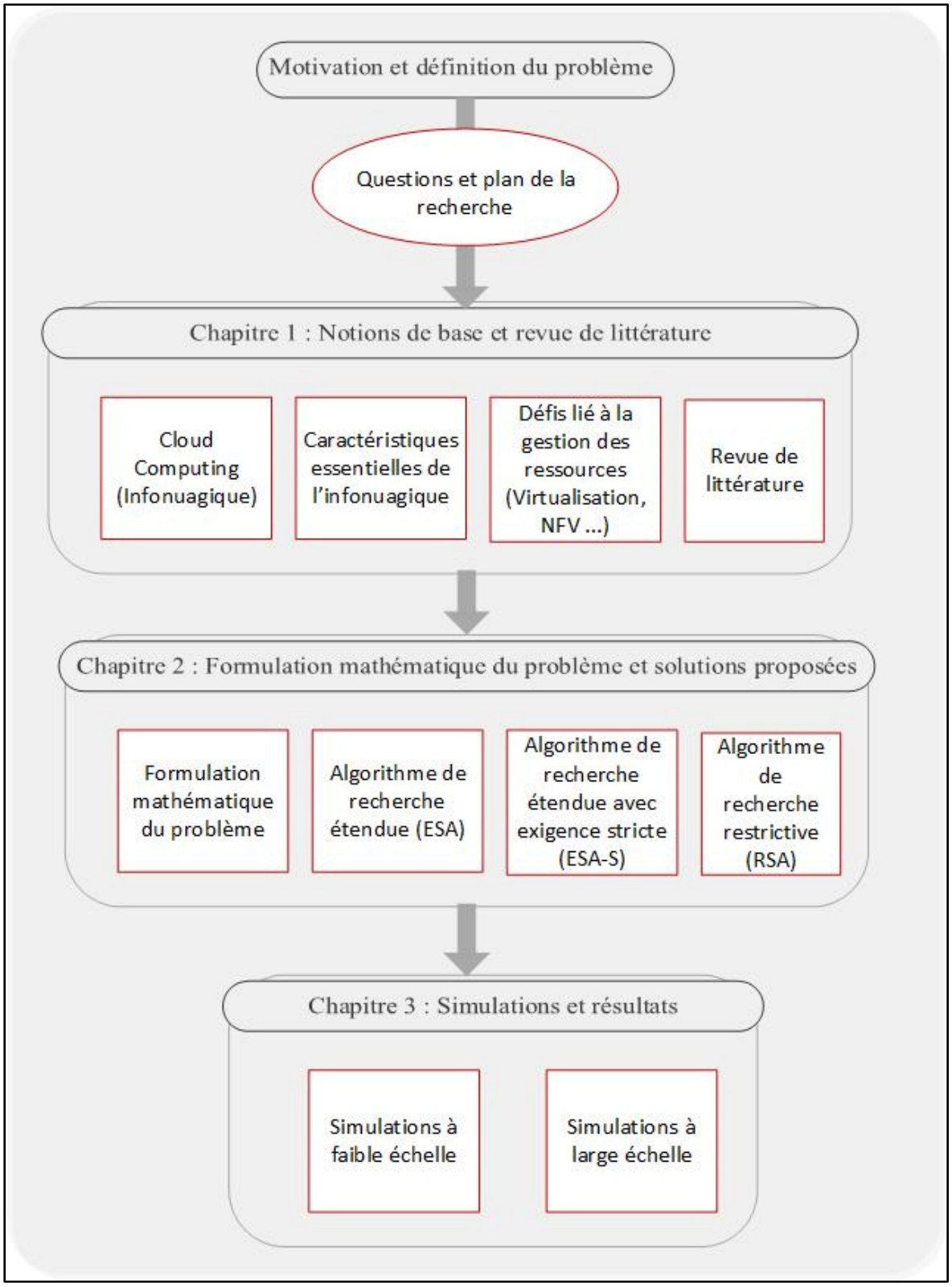


Figure 0.1 Diagramme des chapitres

CHAPITRE 1

# NOTIONS DE BASE ET REVUE DE LITTÉRATURE

## 1.1 Introduction

Ce chapitre présente une revue de littérature des différents aspects théoriques qui nous mènera à bien comprendre le problème de ce projet de recherche. En effet, ce chapitre s'appuie autour de deux grandes sections. Nous allons d'abord définir les notions de base reliées à notre domaine de recherche telles que l'infonuagique, la virtualisation, et le placement et chaînage des VNFs. Nous présentons par la suite les travaux de recherche récents sur le problème du placement et du chaînage des VNFs.

## 1.2 Infonuagique

D'après l'institut national des standard et des technologies (National Institute of Standards and Technology – NIST), l'infonuagique (*cloud computing*) est un modèle permettant d'accéder à un réseau partagé de ressources informatiques configurables (telles que les réseaux, les serveurs, le stockage, les applications et les services informatiques) qui peuvent être rapidement approvisionnées et libérées avec un minimum d'effort de gestion et d'interaction avec le fournisseur de service (Roadmap, NIST Cloud Computing Standards).

Dans ce qui suit, nous présentons les cinq caractéristiques essentielles de l'infonuagique, ses modèles de service et de déploiement tel que défini par le NIST.

### 1.2.1 Caractéristiques essentielles de l'infonuagique

Parmi les caractéristiques essentielles de l'infonuagique, on peut citer (Roadmap, NIST Cloud Computing Standards) :

- **Demande en libre-service :** un utilisateur du cloud peut demander unilatéralement et automatiquement des ressources informatiques (telles que le temps du serveur et le stockage) selon ses besoins et sans nécessairement avoir une interaction humaine

avec chaque fournisseur de service. Des exemples de ressources comprennent le stockage, le traitement, la mémoire et la bande passante du réseau.

- **Accès réseau étendu :** les ressources sont disponibles à travers le réseau et accessibles par l'intermédiaire de mécanismes standards (par ex., Internet) qui favorisent l'utilisation par des plates-formes client hétérogènes, minces ou épaisses (par exemple, téléphones mobiles, tablettes, ordinateurs portables et postes de travail).

- **Mise en commun des ressources :** les ressources informatiques du fournisseur sont mises en commun pour desservir plusieurs utilisateurs à l'aide d'un modèle multi-locataire. Les ressources physiques et virtuelles sont assignées dynamiquement et réaffectées aux utilisateurs selon leurs demandes. Ainsi, les utilisateurs n'ont généralement aucun contrôle ou connaissance sur l'emplacement exact des ressources fournies, mais ils peuvent être en mesure de spécifier cet emplacement à un niveau supérieur d'abstraction (par exemple, pays, état ou centre de données).

- **Élasticité rapide :** les ressources peuvent être approvisionnées et libérées en fonction de la demande. Cela peut être fait automatiquement dans certains cas. Pour l'utilisateur, les ressources disponibles pour approvisionnement semblent souvent être illimitées et peuvent être appropriées en n'importe quelle quantité et à tout moment.

- **Service mesuré :** les systèmes infonuagiques contrôlent et optimisent automatiquement l'utilisation des ressources en exploitant des compteurs définis à certains niveaux d'abstraction dépendamment du type de service (par exemple, stockage, traitement, bande passante, comptes d'utilisateurs actifs). L'utilisation des ressources peut être surveillée, contrôlée, vérifiée et déclarée, assurant la transparence tant pour le fournisseur que pour l'utilisateur du service utilisé.

### 1.2.2 Modèles de service du nuage

Les fournisseurs de nuage peuvent offrir leurs services selon trois modèles tel que défini par le NIST (Roadmap, NIST Cloud Computing Standards) . Ces modèles de service diffèrent en fonction du type de la ressource offerte à l'utilisateur du nuage et du niveau de contrôle qui lui est offert sur celle-ci. Le NIST identifie 3 modèle de services : le *Software as a Service* (SaaS), *Platform as a Service* (PaaS) et *Infrastructure as a service* (IaaS). La Figure 1.1 illustre les trois modèles de service ainsi que les différentes couches de ressources (logiciel, plate-forme, infrastructure et matériel) qui correspondent aux différents modèles de services.

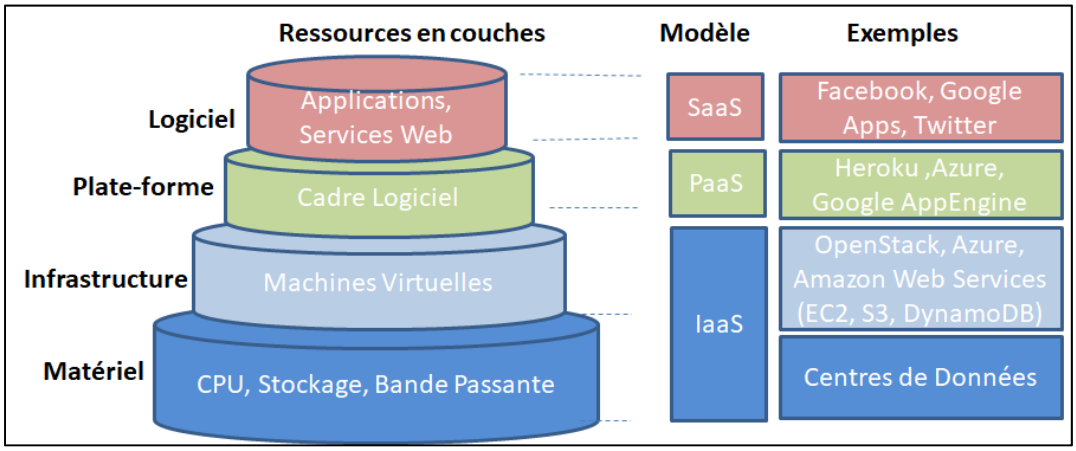


Figure 1.1 Modèles de services

Dans ce qui suit, nous donnons plus de détails sur chacun de ces trois modèles :

- Software as a Service (SaaS)

Les services d'application en nuage ou Software as a Service (SaaS) propose l'utilisation d'un logiciel à un client via l'internet. SaaS utilise le Web pour fournir des applications qui sont gérées par un fournisseur tiers dont l'interface est accessible du côté des clients (Kushida, Murray, & Zysman, 2011). La plupart des applications SaaS peuvent être exécutées directement à partir d'un navigateur Web sans avoir besoin de télécharger ou

installer l'application sur la machine du client. Dans certains cas, seulement un plugin peut être nécessaire pour faire fonctionner l'application.

Avec le SaaS, il est facile pour les entreprises de rationaliser leur maintenance et leur support, car tout peut être géré par les fournisseurs : applications, temps d'exécution, données, middleware, système d'exploitation, virtualisation, serveurs, stockage et mise en réseau.

Les offres SaaS les plus populaires sont le courrier électronique, les outils de collaboration web, la gestion des relations avec la clientèle et les applications liées à la santé. Facebook, les applications Google et Twitter sont des exemples d'application offertes aux utilisateurs du cloud comme service.

- Platform as a Service (PaaS)

Les services de plates-formes Cloud ou Platform as a Service (PaaS) met à la disposition des différents utilisateurs un environnement de développement et d'hébergement sur le nuage. PaaS rend le développement, le test et le déploiement des applications rapides, simples et rentables (Kushida, Murray, & Zysman, 2011). Dans ce cas, le fournisseur de nuage prend en charge la gestion du système d'exploitation, la virtualisation, les serveurs, le stockage, la mise en réseau et le logiciel PaaS lui-même. Cependant, les utilisateurs du nuage se chargeront du développement et de la gestion des applications.

Les applications utilisant PaaS héritent des caractéristiques de nuages telles que l'évolutivité, la haute disponibilité, la multi-location, l'activation de SaaS et plus encore. Les entreprises bénéficient de PaaS car elles réduisent la quantité de codage nécessaire, automatisent l'approvisionnement des applications et permettent de migrer les applications vers l'infonuage.

Heroku, Azure et Google AppEngine sont des exemples de plateformes cloud qui permettent aux utilisateurs et prennent en charge plusieurs langages de programmation.

- Infrastructure as a Service (IaaS)

Ce service est une offre standardisée et hautement automatisée, où les ressources de calcul, complétées par des capacités de stockage et de mise en réseau, appartiennent et sont hébergées par un fournisseur de services et offertes aux clients à la demande. Les clients peuvent s'auto-provisionner cette infrastructure, en utilisant une interface utilisateur graphique basée sur le Web qui sert comme une console de gestion des opérations informatiques pour l'environnement global.

Contrairement aux SaaS et PaaS, les utilisateurs d'IaaS sont responsables de la gestion des applications, des données, du temps d'exécution, des middlewares et des systèmes d'exploitation. Les fournisseurs gèrent toujours la virtualisation, les serveurs, les disques durs, le stockage et la mise en réseau. Beaucoup de fournisseurs d'IaaS offrent désormais des bases de données, des files d'attente de messagerie et d'autres services au-dessus de la couche de virtualisation. Généralement, ces fournisseurs ont recours à des plate-formes de gestion de ressources infonuagiques telles que OpenStack ou CloudStack qui leurs permettent de gérer leurs infrastructures et de l'offrir sous forme de ressources virtuelles.

Ce que les utilisateurs obtiennent avec IaaS est une infrastructure virtuelle (généralement des machines virtuelles) sur laquelle ils peuvent installer n'importe quelle plate-forme requise comprenant le système d'exploitation, les librairies et les environnements de développement, et les applications. Les utilisateurs sont responsables de la mise à jour de cette plateforme si de nouvelles versions deviennent disponibles (Kushida, Murray, & Zysman, 2011).

### 1.2.3 Modèles de déploiement du nuage

Les modèles de déploiement représentent les manières avec lesquelles le nuage informatique peut être implémentéé (Kushida, Murray, & Zysman, 2011). Les modèles de déploiement diffèrent en fonction du propriétaire du nuage, sa taille du nuage et les utilisateurs autorisés

à accéder ses ressources. Généralement, la littérature identifie quatre modèles de déploiement de nuage qui peuvent être résumés comme suit :

- **Nuage public**

Le nuage public est un type d'hébergement en nuage dans lequel les services en nuage sont fournis sur un réseau ouvert pour un usage public. Ce modèle est une véritable représentation de l'hébergement en nuage ; Dans ce cas, le fournisseur de services rend les services et l'infrastructure à divers clients. Les clients n'ont aucune distinction ni contrôle sur l'emplacement de l'infrastructure. Du point de vue technique, il peut y avoir une légère ou aucune différence entre le design structurel des nuages privés et publics, sauf dans le niveau de sécurité offert pour les différents services offerts aux abonnés publics en nuage par les fournisseurs d'hébergement en nuage.

En raison de la diminution des frais généraux et du coût opérationnel, ce modèle de nuage est économique. Le concessionnaire peut fournir le service gratuitement ou sous la forme de la politique de licence, comme la rémunération par utilisateur. Le coût est partagé par tous les utilisateurs, de sorte que le public cloud profite davantage aux clients en réalisant des économies d'échelle. Les installations de cloud public peuvent être utilisées gratuitement, un exemple d'un nuage public est Google.

- **Nuage privé**

L'infrastructure en nuage est fournie pour une utilisation exclusive par une seule organisation comprenant plusieurs consommateurs (par exemple, unités commerciales). Il peut appartenir, géré et exploité par l'organisation, un tiers, ou une combinaison d'entre eux, et il peut exister ou en dehors des locaux. Le cloud privé, car il permet uniquement aux utilisateurs autorisés, donne à l'organisation un contrôle plus important et plus direct de leurs données.

Dans un nuage privé, les ordinateurs physiques soient hébergés à l'interne ou à l'extérieur, ils fournissent les ressources d'un pool distinct aux services du nuage privé. Les entreprises qui ont des besoins dynamiques ou imprévus, les missions essentielles, les alarmes

de sécurité, les demandes de gestion et les conditions de disponibilité sont mieux adaptées pour adopter un cloud privé. Les obstacles en matière de sécurité peuvent être évités dans un nuage privé, mais en cas de catastrophe naturelle et de vol de données internes, le cloud privé risque d'être vulnérable (Kushida, Murray, & Zysman, 2011).

- **Nuage hybride**

Le nuage hybride est un type de nuage intégré qui combine deux ou plusieurs nuages de types différents, c'est-à-dire un nuage privé, public ou communautaire (Kushida, Murray, & Zysman, 2011). Dans un nuage hybride, les ressources peuvent être gérées et fournies en interne ou par des fournisseurs externes.

Le nuage hybride constitue une plates-forme dans laquelle les applications et les données peuvent être distribuées entre le cloud privé et le nuage public selon la demande et les besoins en termes de ressources, de confidentialité et de sécurité. Ainsi, les données et les applications non critiques peuvent être hébergées dans le cloud public appartenant à un fournisseur tiers, alors que celles qui sont critiques ou sensibles vont être logées en interne dans le nuage privé.

L'hébergement par nuage hybride offre plusieurs avantages tels que l'évolutivité, la flexibilité et la sécurité. Cependant, il présente quelques défis tels que l'incompatibilité d'interface de programme d'application et les éventuels problèmes de connectivité réseau.

- **Nuage communautaire**

Le nuage communautaire est un type de nuage dans lequel l'installation est partagée entre plusieurs organisations appartenant à une communauté particulière et qui partagent les mêmes intérêts, appréhensions et des objectifs informatiques (Kushida, Murray, & Zysman, 2011). Ainsi, les ressources du nuage sont partagées entre plusieurs organisations. Les membres de la communauté partagent généralement des soucis similaires en matière de de confidentialité, de performance et de sécurité.

Un nuage de communauté peut être géré en interne ou il peut être géré par un fournisseur tiers. Il peut aussi être hébergé à l'extérieur ou à l'interne des organisations impliquées. L'un des avantages principaux de ce type de nuage est que le coût est partagé par les organisations qui utilisent le nuage communautaire.

### 1.2.4 Modèle d'affaire

Dans un environnement infonuagique, le rôle du fournisseur de service internet traditionnel a été divisé en deux : un fournisseur d'infrastructure et un fournisseur de service (Bari, et al., 2013) :

- Fournisseur d'infrastructure (*InF Provider*) : le fournisseur d'infrastructure possède l'infrastructure physique (par ex., routeurs, liens optiques, serveurs et centres de données). A l'aide des technologies de virtualisation, il est capable de partitionner les ressources et de les distribuer aux fournisseurs de services sous forme de machines virtuelles et de routeurs virtuels connectés à travers des liens virtuels.

- Fournisseur de service (*Service Provider*) : le fournisseur de service loue les ressources virtuelles auprès du fournisseur d'infrastructure. Il est ainsi capable de configurer les machines virtuelles et les liens virtuels acquis et d'y installer les services et les applications qu'il veut offrir aux utilisateurs de l'internet. Les utilisateurs de l'internet utilisent ensuite ces services et applications.

D'une façon analogue, lorsque les fonctions réseau sont offertes en tant que service, on peut identifier trois acteurs :

- Fournisseur de VNFs (*VNFs Provider*) : le fournisseur de VNFs possède et gère l'infrastructure physique. Il loue les ressources de cette infrastructure sous forme de chaine de service aux fournisseurs de service.

- Fournisseur de service (*Service Provider*) : le fournisseur de service loue les ressources auprès du fournisseur de VNFs. Ces ressources sont principalement des chaines de services composées de fonctions virtuelles qui permettront de traiter le trafic des utilisateurs finaux. C'est le fournisseur de service qui définit les composants de la chaine et la quantité de ressources pour chaque VNFs en termes

de disque, cpu, mémoire et bande passante et envoie sa requête au fournisseur de VNFs.

- Utilisateur final (*End User*) : l'utilisateur final est l'utilisateur de l'internet qui demande l'accès à internet auprès d'un fournisseur de service. Ainsi, le trafic généré par l'utilisateur final sera transmis à travers les différents VNFs qui composent la chaine de service du fournisseur de service.

La Figure 1.2 présente les différentes parties prenantes du modèle d'affaire, ainsi les différentes relations entre eux.

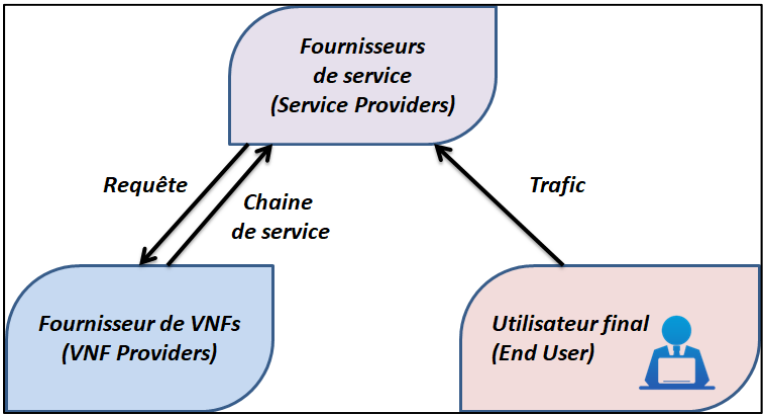


Figure 1.2 Modèle d'affaire

Dans le présent travail, on se positionne en tant que fournisseur de VNFs qui offre des ressources aux fournisseurs de service et qui cherche à allouer les ressources nécessaires pour approvisionner une chaine de service et donc résoudre le problème de placement et de chaînage des VNFs dans le but de maximiser ses profits.

## 1.3 Virtualisation

### 1.3.1 Définition

La virtualisation est une technique qui permet de créer des environnements simulés appelés « machine virtuelles » (MVs). Chacune agit comme une machine physique ayant son propre système d'exploitation et applications (Mijumbi, et al., Network Function Virtualization: State-of- the-Art and Research Challenges, 2016).

Cette technologie est utilisée principalement pour virtualiser des serveurs. Dans ce cas, une couche de logiciel appelée hyperviseur est installée dans le serveur pour imiter le matériel sous-jacent et pour permettre l'exécution de plusieurs machines virtuelles en parallèle. Cela comprend souvent la mémoire, la CPU et le trafic réseau. Le système d'exploitation invité, qui interagit habituellement avec un véritable matériel, le fait maintenant avec l'hyperviseur qui émule l'existence du matériel. Ainsi, généralement le système d'exploitation installé au niveau de la machine virtuelle est souvent inconscient de la virtualisation. Bien que la performance de ce système virtuel ne soit pas égale aux performances du système d'exploitation exécuté sur du matériel physique, la notion de virtualisation fonctionne car la plupart des systèmes d'exploitation invités et des applications n'ont pas besoin de l'utilisation complète des ressources du matériel sous-jacent. Cela permet un plus grand contrôle, souplesse et isolation en supprimant la dépendance sur une plate-forme matérielle donnée.

Cette technique peut s'appliquer aux serveurs traditionnels, aux serveurs de stockage et aux réseaux. Cela permet de réduire les couts informatiques tout en stimulant l'efficacité et la flexibilité des entreprises de toute taille. À travers la technique de virtualisation, il est possible de réduire d'une façon très efficace les dépenses informatiques en plaçant plusieurs machines virtuelles dans une même machine physique. Elle permet aussi d'offrir plus de flexibilité en permettant d'ajuster dynamiquement la capacité des machines virtuelles dépendamment de la demande et de les migrer en temps réel dans des serveurs physiques plus puissants.

Bien qu'initialement destiné à la virtualisation des serveurs, le concept de virtualisation s'est propagé aux applications, aux réseaux, aux fonctions réseau, aux données et aux ordinateurs de bureau (Mijumbi, et al., Network Function Virtualization: State-of- the-Art and Research Challenges, 2016). Dans ce qui suit, nous détaillons davantage le principe de la virtualisation des fonctions réseau.

### 1.3.2 Virtualisation des fonctions réseau

La virtualisation des fonctions réseau consiste à séparer les fonctions réseau (par ex., routage) des équipements qui les exécutent en utilisant la technologie de virtualisation. Les fonctions réseau tels que le routage, la traduction d'adresses réseau (NAT), le pare-feu, la détection d'intrusion, le service de noms de domaine (DNS) et la mise en cache peuvent être virtualisées et offertes par des logiciels hébergés dans des machines virtuelles au lieu d'avoir des équipements hardware dédiés.

Une fonction réseau virtuelle est une implémentation d'une fonction réseau (*Network Function* - NF) qui est déployée sur des ressources virtuelles telles qu'une machine virtuelle (Mijumbi, et al., Network Function Virtualization: State-of- the-Art and Research Challenges, 2016). La virtualisation des fonctions réseau implique la mise en œuvre de cette fonction dans un logiciel qui peut fonctionner sur une gamme de matériels de serveur standard de l'industrie et qui peut être facilement déplacé ou instancié dans plusieurs emplacements du réseau sans avoir besoin d'installer de nouveaux équipements ((NFV), Network Functions Virtualisation). Ces fonctions réseau sont ensuite connectées ensemble dans un ordre défini par le fournisseur de service pour créer une chaîne de service (Network Functions Virtualisation (NFV),Placement and chaining).

La virtualisation des fonctions réseau offre une nouvelle façon de concevoir, de déployer et de gérer les services réseau. Elle permettrait de transformer la façon dont les opérateurs de réseaux conçoivent leurs réseaux en exploitant la technologie de virtualisation pour consolider de nombreux types d'équipements de réseau sur des serveurs de haut volume, commutateurs et de stockage, qui pourrait être situé dans des centres de données, nœuds de réseau et dans les locaux des utilisateurs finaux.

La Figure 1.3 montre l'architecture de virtualisation des fonctions réseau qui se compose de trois couches (services, ressources virtuelles et ressources physiques), ainsi que les différentes utilisations de chaque couche, tout en assurant la gestion et l'orchestration du chacun d'eux.

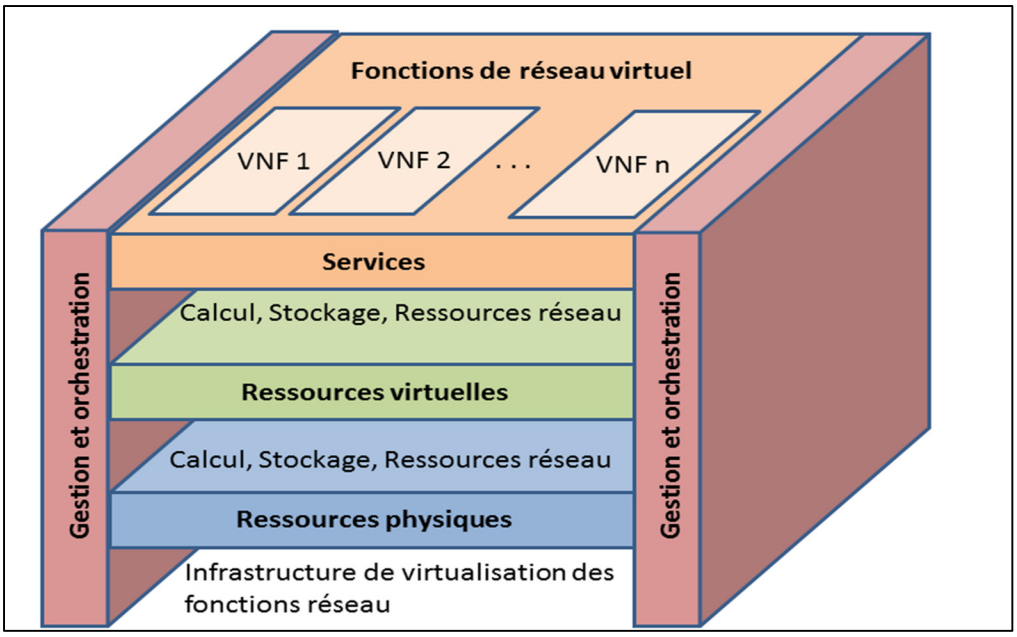


Figure 1.3 Architecture de virtualisation des fonctions réseau

Le concept de la virtualisation des fonctions réseau est né en octobre 2012, lorsqu'un certain nombre de leaders mondiaux ont rédigé conjointement un livre blanc (Guerzoni, 2012) appelant à des actions industrielles et de recherche. En novembre 2012, sept de ces opérateurs (AT & T, BT, Deutsche Telekom, Orange, Telecom Italia, Tele-fonica et Verizon) ont sélectionné l'Institut européen des normes de télécommunications (ETSI) ((ETSI), 2015) pour être le siège du Groupe de spécifications de l'industrie pour le NFV (ETSI ISG NFV). Maintenant, plus de quatre ans plus tard, une grande communauté d'experts travaille intensément pour élaborer les normes requises pour le NFV ainsi que pour partager leurs expériences de développement et de mise en œuvre précoce. L'adhésion à ETSI est passée à plus de 245 entreprises individuelles, dont 37 des principaux fournisseurs

de services au monde, ainsi que des représentants des fournisseurs de télécommunications et d'informatique ((ETSI), 2015).

### 1.3.3 Défis de la gestion des ressources virtuelles

Il y a un certain nombre de défis à mettre en œuvre concernant la virtualisation des fonctions réseau qui doivent être pris en charge (Chiosi, Clarke, Willis, & Reid, October 2012). Parmi ces défis, on peut citer :

- **Placement et chaînage des VNFs :** le placement et le chaînage consiste à trouver les meilleurs emplacements et hôtes pour les VNFs afin de diriger le trafic à travers ces fonctions réseau virtuelles afin de respecter les exigences en termes de performance et de délai. Ce problème peut être décomposé en deux sous-problèmes: le placement et le chaînage. Le placement consiste à sélectionner les nœuds qui hébergeront les VNFs requis tandis que le chaînage consiste à créer des chemins que le trafic doit suivre pour passer à travers les VNFs qui composent chaque chaine de service (Network Functions Virtualisation (NFV),Placement and chaining).

- **Garantie de performance :** étant donné que la virtualisation des fonctions du réseau est basée sur les serveurs standards, il est attendu qu'il y ait une diminution probable des performances. Dans ce contexte, un défi majeur consiste à minimiser la dégradation des performances en utilisant des hyperviseurs appropriés et des technologies logicielles qui permettent de minimiser les effets sur la latence, le débit et les temps de traitement.

- **Migration, coexistence de l'héritage et compatibilité avec les plates-formes existantes :** les fonctions réseau virtuelles doivent coexister avec les équipements réseau traditionnels qui sont déjà installés dans les réseaux des opérateurs internet. Ainsi, il est nécessaire de concevoir et mettre en place des protocoles qui permettent

aux VNFs d'être compatible avec les systèmes de gestion de réseau et les systèmes d'orchestration de service existants.

- **Automatisation de la gestion et orchestration de service :** l'automatisation de l'approvisionnement, la configuration et la gestion des services réseau devrait réduire les coûts et accélérer leurs performances. Cela nécessite une architecture cohérente de gestion et d'orchestration pour l'allocation automatique des ressources et le dimensionnement dynamique des fonctions virtuelles qui composent les chaines de services réseau en fonction de la demande.

- **Stabilité du réseau :** la gestion d'un grand nombre d'appliances virtuelles (une appliance virtuelle est une image de machine virtuelle préconfigurée et prête à fonctionner sur un hyperviseur) peut créer une certaine instabilité de service vu que les fonctions virtuelles sont reconfigurées et déplacées afin de les adapter aux différentes variations de trafic en fonction du temps. Cela peut engendrer des interruptions des services offerts par les appliances et peut créer un trafic supplémentaire dans le réseau.

- **Portabilité et interopérabilité :** ce défi consiste à pouvoir charger et d'exécuter des appliances virtuelles (une image de machine virtuelle préconfigurée, prêt à fonctionner sur un hyperviseur) dans des environnements de centres de données différents gérés par différents opérateurs.

- **Simplicité :** Assurer que les plates-formes de réseau virtualisées seront plus simples à utiliser que celles qui existent aujourd'hui.

- **Sécurité et résilience :** Les opérateurs de réseau doivent assurer que la sécurité, la résilience et la disponibilité de leurs réseaux ne sont pas altérées lorsque des fonctions de réseau virtualisées sont utilisées.

- **L'intégration :** L'intégration transparente de plusieurs appliances virtuelles sur des serveurs et des hyperviseurs différents est un défi majeur pour la virtualisation des fonctions réseau. Les opérateurs de réseau doivent être capable de faire fonctionner des serveurs de différents fournisseurs, des hyperviseurs de différents fournisseurs et des appliances virtuelles de différents fournisseurs sans encourir de coûts d'intégration importants ou impacter la performance des services offerts par les appliances.

- **Minimiser les coûts :** ce défi concerne la réduction des coûts des équipements et la réduction de la consommation de l'énergie électrique de l'infrastructure hébergeant les fonctions virtuelles et le coût de cette énergie. Afin d'atteindre cet objectif, des techniques de gestion et de consolidation de machines virtuelles doivent être mises en place en tenant compte de la consommation d'énergie et des prix de l'électricité.

Dans ce travail, nous nous intéressons à l'allocation de ressources pour les chaines de service et en particulier, au placement et de chaînage des VNFs dans le but de minimiser les différents coûts associés à cette allocation dans le but de maximiser les profits des fournisseurs de VNFs.

## 1.4 Revue de littérature

### 1.4.1 Résumé des travaux existants

Dans cette section, nous examinons les travaux de recherche existants qui ont abordé le problème du placement et du chaînage des VNFs.

Clayman et al. (Clayman, Mainiy, Galis, Manzaliniz, & Mazzocca, 2014) abordent le problème du placement de routeurs virtuels et proposent une architecture multicouche (couche d'application, couche d'orchestration, couche d'abstraction et une couche infrastructure) pour orchestrer et gérer les ressources d'infrastructure. Ils évaluent également

des algorithmes de placement de base tels que le *Least Used Host* (l'hôte le moins utilisé), qui tente d'équilibrer la charge à travers le réseau en plaçant des VNFs dans le nœud physique hébergeant le moins de VNFs, et le *Least Busy Host* (l'hôte le moins occupé), qui place les VNFs dans les hôtes ayant le moins de trafic de communication. Les résultats présentés démontrent que les différents algorithmes intégrés dans chacun des moteurs de placement ont des comportements différents et donnent des stratégies de placement très différentes pour les routeurs virtuels. Cependant, ces solutions sont basiques car elles ne tiennent pas compte des exigences de délai de bout en bout et des coûts d'énergie.

Bari et al. (Bari, Chowdhury, Ahmed, & Boutaba, 2015) résolvent le problème de l'allocation et du chaînage des VNFs. Les auteurs fournissent une heuristique (NFO-DP) pour déterminer le nombre de VNF requis et le meilleur placement pour eux. L'objectif principal de ce travail est de minimiser le coût opérationnel (peut être décrit comme les frais associés à l'exploitation d'une activité, d'un appareil, d'un composant d'une pièce d'équipement ou d'une installation) et l'utilisation du réseau, sans violer le contrat de service « SLA ». Le contrat de service est un document qui définit la qualité de service et la qualité de prestation prescrite entre un client et un fournisseur de service. Autrement dit, il s'agit d'une clause basée sur un contrat définissant les objectifs précis attendus et le niveau de service que souhaite obtenir un client de la part du prestataire). Les résultats montrent que l'utilisation de VNFs réduit la consommation d'énergie par rapport aux boîtiers intermédiaires matériels avec un temps beaucoup plus rapide que la solution optimale (environ 65 à 3500 fois plus rapide). NFO-DP favorise les nœuds de mappage de chaque demande sur le même nœud physique. Cela signifie que les bandes passantes sur les liens sont négligées car il n'y a pas de contraintes de lien lorsque les VNFs sont mappés sur le même nœud physique.

Mijumbi et al. (Mijumbi, et autres, Design and evaluation of algorithms for mapping and scheduling of virtual network functions, April 2015) abordent et formulent le problème du mappage en ligne et l'ordonnancement des VNFs. Les auteurs proposent trois algorithmes gloutons et une solution heuristique basée sur la recherche Tabu visant à mapper

et ordonnancer les VNFs. Ils les comparent ensuite en termes de taux d'acceptation, de temps total de traitement des services et de revenus. Toutefois, les délais de bout en bout et les coûts énergétiques n'ont pas été pris en compte.

Huang et al. (Huang, Li, & Wen, Oct 2015) abordent les problèmes de contention intra-chaîne et inter-chaîne. Le problème de contention intra-chaîne se rapporte à la surutilisation des mêmes liaisons par une chaîne de service, alors que le problème de contention inter-chaîne se produit lorsque certains liens sont surchargés par plusieurs chaînes de services. Pour résoudre ces problèmes de contention, les auteurs proposent un schéma d'orchestration en chaîne adapté au réseau (NACHOS) qui utilise la programmation linéaire en nombre entier et la programmation dynamique pour trouver les meilleurs itinéraires pour les chaînes de services qui maximisent la bande passante disponible dans le réseau. Cependant, cette solution ne garantit pas les exigences de délai de bout en bout et ne tient pas compte des coûts d'énergie et des revenus du fournisseur de VNFs.

Moens et De Turck (H. Moens and F. De Turck, Nov 2014) abordent le problème du placement de VNF dans l'infrastructure hybride NFV. L'environnement NFV hybride se compose de matériels basés sur des boîtiers intermédiaires et de VNFs. L'allocation de VNFs intervient lorsque les boîtiers intermédiaires matériels sont entièrement utilisés. Les auteurs proposent un modèle de placement VNF (VNF-P) qui vise à minimiser le nombre de nœuds utilisés pour héberger des VNFs et à réduire le délai de bout en bout. Cependant, ce modèle ne prend pas en compte les différents prix de l'électricité aux POPs et ne minimise pas les coûts de l'énergie ni maximise les revenus des fournisseurs de VNFs.

Mechtri et al. (M. Mechtri and C. Ghribi and D. Zeghlache, September 2016) traitent le problème de placement et chaînage des VNFs dans les environnements et les réseaux cloud en trouvant les meilleurs emplacements et hôtes pour les VNFs afin de diriger le trafic à travers ces fonctions (VNFs), tout en respectant les exigences des utilisateurs

et en maximisant les revenus des fournisseurs de VNFs. L'objectif principal des auteurs est de réduire les dépenses en CAPEX (correspond au total des dépenses d'investissement (corporel et incorporel) consacrées à l'achat d'équipement professionnel) et OPEX (les charges d'exploitation de l'entreprise) d'une part, et gagner en agilité de service d'autre part. Les simulations démontrent que cette approche offre une meilleure performance par rapport aux travaux précédents en termes de revenue et de taux d'acceptation. Cependant, la différence des coûts de l'électricité qui peut changer d'une zone géographique à une autre n'est pas prise en considération vu que le problème est résolu dans une seule localisation donc un seul prix d'électricité.

Luizelli et al. (Luizelli, Bays, Burio, Barcellos, & Gasparyl, 2015) abordent le problème du placement et de chaînage du VNF. Ils estiment d'abord le nombre minimal d'instances VNFs nécessaires pour satisfaire la demande attendue, puis ils les placent de telle sorte que les délais de bout en bout soient satisfaits. Ils proposent un programme linéaire en nombre entier (ILP) et un algorithme heuristique pour faire face à des infrastructures à grande échelle. Mehraghdam et al. (Mehraghdam, Keller, & Karl, 2014) formalisent le problème de chaînage VNFs en utilisant « un langage hors-contexte » (H. Lieberman and T. Selker, 2000) qui permet de spécifier les types de VNFs et leur ordre. Ils ont également mis en avant un programme à contrainte quadratique mixte (MIQCP) pour trouver le meilleur emplacement pour les VNFs qui maximise la bande passante non utilisée et minimise le nombre de nœuds utilisés et la latence de la chaîne VNFs. Cependant, les deux solutions (Luizelli, Bays, Burio, Barcellos, & Gasparyl, 2015), (Mehraghdam, Keller, & Karl, 2014) ne tiennent pas compte des coûts de l'énergie et des différents prix de l'énergie aux emplacements des POPs.

Sahhaf et al. (Sahhaf, et al., 2015) résolvent le problème du placement et de chaînage des services basés sur le NFV en tenant compte des exigences de la demande en termes de bande passante et de délai de bout en bout. Dans leur solution, ils considèrent la décomposition des Fonctions de Réseau (NF) en des NF plus raffinées (par exemple, un pare-feu est décomposé en firewall basé sur iptable et / ou un pare-feu à flux).

Ils soutiennent qu'une telle décomposition permettra de réutiliser ces composants raffinés pour construire de nouvelles fonctions de réseau. Leur algorithme de placement et de chaînage réduit la consommation de ressources et augmente le taux d'acceptation des demandes. Cependant, il ne tient pas compte des coûts de l'énergie et ne maximise les bénéfices des fournisseurs de VNFs.

Rankothge et al. (Rankothge, Ma, Le, Russo, & Lobo, May 2015) abordent le problème d'allocation et de placement des VNFs. Les auteurs proposent une approche basée sur «la programmation génétique» (G Gerules and C Janikow, 2016) pour résoudre le problème d'allocation et de gestion des VNFs. Ils ont construit une plate-forme expérimentale NFV (pour mieux comprendre et étudier les questions de recherche liées au NFV). Ils ont exécuté un ensemble d'expériences tout en respectant les exigences SLA et QoS. Les résultats montrent l'approche GA proposée peut calculer des configurations à trois ordres de grandeur plus rapidement que les solutions traditionnelles. Toutefois, allouer et placer plusieurs chaines de service de VNFs simultanément n'est pas pris en considération. Dans ce travail, les demandes du placement et de chainage des VNFs doivent être placées une demande à la fois.

### 1.4.2 Comparaison des solutions existantes

Tableau 1.1 Solutions existantes comparées aux solutions proposées

| Approche | Objectifs | | | |
|---|---|---|---|---|
| | **Minimiser le délai de bout en bout** | **Minimiser les coûts énergétiques** | **Minimiser le nombre de nœuds / Bande passante** | **maximiser les profits** |
| **Clayman et al.** | ✘ | ✘ | ✔ | ✘ |
| **Luizelli et al. , Mehraghdam et al.** | ✔ | ✘ | ✔ | ✘ |
| **Sahhaf et al.** | ✔ | ✘ | ✘ | ✘ |
| **Huang et al.** | ✘ | ✘ | ✔ | ✘ |
| **Mijumbi et al.** | ✘ | ✘ | ✘ | ✔ |
| **Moens et De Turck** | ✔ | ✘ | ✔ | ✘ |
| **Bari et al.** | ✔ | ✔ | ✘ | ✔ |
| **Mechtri et al.** | ✔ | ✔ | ✘ | ✔ |
| **Rankothge et al.** | ✔ | ✘ | ✔ | ✘ |
| ***Nos solutions*** | ✔ | ✔ | ✘ | ✔ |

Le

Tableau 1.1 fournit une comparaison entre les solutions existantes qui traitent du problème du placement et du chaînage des chaines de service et les solutions proposées dans le présent travail. Le tableau compare les différents objectifs considérés tels que la minimisation

du délai de bout en bout de transmission des paquets, la minimisation des coûts énergétiques, de bande passante et la maximisation de revenu des fournisseurs de VNFs.

Contrairement aux travaux existants, les solutions que nous proposons dans ce mémoire visent à atteindre simultanément les objectifs suivants : réduire les coûts énergétiques et maximiser les profits du fournisseur de VNFs tout en tenant compte des différents prix de l'électricité dans les emplacements couverts par l'infrastructure du fournisseur de VNFs et de l'exigence de délai de bout en bout.

## 1.5 Conclusion

Ce chapitre présente les concepts de base liés à l'infonuagique et la virtualisation des fonctions réseau. Il présente aussi une étude approfondie des solutions existante dans la littérature qui ont abordé le problème de placement et de chaînage des VNFs. Ces travaux sont classés en fonction de leurs objectifs : la satisfaction des performances requises en termes de délai de bout en bout, minimisation des coûts énergétiques, et la maximisation des revenus des fournisseurs de VNFs. Contrairement à ces travaux, les solutions proposées dans ce mémoire visent à atteindre tous ces objectifs et prennent en considération les prix de l'électricité dans les emplacements des différents nœuds composant l'infrastructure.

Dans le chapitre suivant, nous présentons la formulation mathématique du problème de placement et de chaînage des VNFs qui tient compte des objectifs de performance visés et des différentes contraintes d'emplacement et de coût. Ensuite, nous présentons des algorithmes heuristiques permettant d'atteindre les mêmes objectifs avec une complexité de calcul minimale.

**CHAPITRE 2**

# FORMULATION MATHÉMATIQUE ET SOLUTIONS PROPOSÉES

## 2.1 Introduction

Ce chapitre est composé de deux sections. Dans la première section, nous formulons le problème de placement et de chaînage des VNFs visant à maximiser le profit du fournisseur de VNFs en tant qu'un programme linéaire en nombres entiers (*Integer Linear Program* – ILP). Dans la deuxième section, nous proposons un algorithme pour l'allocation de ressource pour les chaine de service qui permet d'explorer efficacement tous les chemins de trafic possibles et les emplacements de VNFs capables de satisfaire les exigences de la chaîne de service et de maximiser le profit du fournisseur de VNFs. Nous proposons ensuite deux autres variantes de cet algorithme qui permettront de réduire davantage la complexité et le temps d'exécution sans affecter la qualité de la solution trouvée.

## 2.2 Formulation mathématique du problème

Comme l'infrastructure physique se compose de plusieurs points de présence (*Point of Presence* – POPs) situés dans différentes régions, elle est modélisée par un graphe $\overline{G} = (\overline{N}, \overline{L})$ où $\overline{N}$ désigne l'ensemble des POPs et $\overline{L}$ désigne l'ensemble des liens physiques qui les relient.

Chaque POP $\bar{n} \in \overline{N}$ a une capacité de ressources notée $c_{\bar{n}}^{ir}$ où $r \in R$ est le type de ressource. $R = \{1,2,3\}$ est l'ensemble des types de ressources représentant respectivement la CPU, la mémoire et le stockage.

En outre, chaque lien $\bar{l}$ a une capacité de bande passante $b_{\bar{l}}$ et un délai de propagation $t_{\bar{l}}$. De même, nous supposons que $G^i = (N^i, L^i)$ représente une chaîne de service demandée $i \in I$ où $N^i$ est l'ensemble de ses VNFs composantes et $L^i$ est l'ensemble des liens virtuels qui les relient.

On suppose que chaque VNF $n \in N^i$ a une exigence $c_{\bar{n}}^{ir}$ pour la ressource $r \in R$, et que chaque lien virtuel reliant une paire de VNFs a une exigence de bande passante de $b_{\bar{l}}$. La chaîne de services i a une exigence de délai de bout en bout définie par $\pi_d^i$. La variable de décision $x_{n\bar{n}}^i \in \{0,1\}$ indique si le VNF n appartenant à la chaîne de services i est placé dans le POP $\bar{n} \in \bar{N}$.

### 2.2.1 Contraintes

Pour trouver un placement réalisable, plusieurs contraintes doivent être satisfaites. Par exemple, pour s'assurer que les VNFs placés dans un POP ne dépassent pas sa capacité en termes de ressources $c_{\bar{n}}^{ir}$ (c'est-à-dire CPU, mémoire, stockage), la contrainte suivante doit être satisfaite :

$$\sum_{i \in I} \sum_{n \in N^i} x_{n\bar{n}}^i \, C_n^{ir} \leq C_{\bar{n}}^r \quad \forall \bar{n} \in \bar{N}, r \in R \tag{2.1}$$

Nous devons également veiller à ce que chaque VNF soit intégré dans un seul et unique POP :

$$\sum_{\bar{n}} x_{n\bar{n}}^i \leq 1 \; \forall i \forall n \tag{2.2}$$

La contrainte suivante garantit que la bande passante requise pour intégrer tous les liens virtuels dans un lien physique ne dépasse pas la capacité de la bande passante disponible :

$$\sum_{i \in I} \sum_{l \in L^i} f_{l\bar{l}}^i \leq b_{\bar{l}} \; \forall \bar{l} \in \bar{L} \tag{2.3}$$

### 2.2.2 Consommation d'énergie

Désignons $E_{\bar{n}}$ la consommation d'énergie de POP $\bar{n}$ qui contient un ensemble de serveurs $\bar{M}_{\bar{n}}$. Comme dans (Zhang, Zhani, Boutaba, & Hellerstein, 2013), nous utilisons

un modèle linéaire pour calculer la consommation d'énergie d'un seul serveur. Pour simplifier, nous supposons que tous les serveurs $\bar{m} \in \bar{M}_{\bar{n}}$ du même POP $\bar{n}$ sont homogènes en termes de modèle de consommation et de capacité CPU. En d'autres termes, un serveur $\bar{m} \in \bar{M}_{\bar{n}}$ a la même consommation d'énergie quand il est inactif (noté par $E_{\bar{n}}^{idle}$) et sa consommation d'énergie croît linéairement avec une pente $\gamma_{\bar{n}}$.

Ainsi, l'énergie consommée par le serveur $\bar{m} \in \bar{M}_{\bar{n}}$ est fournie par l'équation suivante (Zhang, Zhani, Boutaba, & Hellerstein, 2013) :

$$\boldsymbol{E_{\bar{m}} = E_{\bar{n}}^{idle} + \gamma_{\bar{n}} U_{\bar{m}}} \tag{2.4}$$

où $U_{\bar{m}}$ est l'utilisation de la CPU du serveur $\bar{m} \in \bar{M}_{\bar{n}}$ . La consommation d'énergie pour chaque POP peut alors être calculée comme la somme de l'énergie consommée par ses serveurs. Elle peut s'écrire comme suit:

$$\boldsymbol{E_{\bar{n}} = \sum_{\bar{m}=1}^{|\bar{M}_{\bar{n}}|} \left(E_{\bar{n}}^{idle} + \gamma_{\bar{n}} U_{\bar{m}}\right)} \tag{2.5}$$

Étant donné que tous les serveurs du même POP ont le même modèle de consommation d'énergie (c'est-à-dire la même consommation d'énergie lorsque le serveur est inactif $E_{\bar{n}}^{idle}$ et la même pente $\gamma_{\bar{n}}$), la consommation totale d'énergie du POP $\bar{n}$ peut s'écrire comme suit :

$$\boldsymbol{E_{\bar{n}} = |\bar{M}_{\bar{n}}| E_{\bar{n}}^{idle} + \gamma_{\bar{n}} \sum_{\bar{m}=1}^{|M_{\bar{n}}|} U_{\bar{m}}} \tag{2.6}$$

En outre, la somme de l'utilisation du CPU de tous les serveurs du POP $\bar{n}$ (c'est-à-dire $\sum_{\bar{m}=1}^{|M_{\bar{n}}|} U_{\bar{m}}$ ) peut être écrite comme la somme de toutes les ressources CPU consommées par les VNFs intégrés dans le POP $\bar{n}$ (c'est-à-dire $\sum_{i \in I} \sum_{n \in N^i} C_n^{i1}\, x_{n\bar{n}}^i$) divisée par le total de la ressource CPU disponible dans le POP $\bar{n}$ (soit $C_n^1$).

La consommation totale d'énergie dans un POP peut donc s'écrire :

$$\mathbf{E_{\bar{n}} = E_{\bar{n}}^{idle} + {\gamma_{\bar{n}}}/{c_{\bar{n}}^{1}} \left( \sum_{i \in I} \sum_{n \in N^i} c_n^{i1} \, x_{n\bar{n}}^{i} \right)} \quad (2.7)$$

### 2.2.3 Pénalité

Dans notre modèle, nous supposons qu'il est possible d'intégrer une chaîne de service même si l'exigence de délai de bout en bout n'est pas satisfaite. Toutefois, si cette exigence est violée, une pénalité devrait être payée par le fournisseur de VNFs. Notons par $\pi^i$ la latence de la requête $i$ après l'allocation et par $\pi_d^i$ la latence requise par le fournisseur de service pour la même requête. La pénalité $\rho_i$ associée à la requête i peut être calculée comme suit :

$$\mathbf{\rho_i = (\pi^i - \pi_d^i)^+ \beta} \quad (2.8)$$

où β est la pénalité payée par le fournisseur de VNFs pour chaque milliseconde dépassant la latence requise. La fonction $(x)^+$ est égale à 0 si x est négatif. Ainsi, si la latence de la chaine de service après l'allocation est moins élevée que la latence désirée, il n'y a aucune pénalité ; sinon, la pénalité est proportionnelle à la différence entre les deux latences : plus la violation est élevée plus la pénalité est élevée.

### 2.2.4 Revenus du fournisseur de VNFs

Les revenus du fournisseur de VNFs dépendent principalement des ressources demandées par les VNFs en termes de cpu, mémoire, et disque ainsi que la bande passante des liens virtuelles composant la chaîne de service.

Les revenus obtenus par le fournisseur de VNFs pour une chaîne de service $i$ peuvent être estimés comme suit :

$$\boldsymbol{\sigma_i} = \left(\sum_{\mathbf{n}\in\mathbf{N^i}} \sum_{\mathbf{r}\in\mathbf{R}} \mathbf{c_n^{ir}}\ \boldsymbol{\delta_r}\right) + \sum_{\mathbf{l}\in\mathbf{L}} \mathbf{b_l^i}\ \boldsymbol{\delta_b} \tag{2.9}$$

où $\delta_r$ et $\delta_b$ correspondent respectivement au prix de vente de la ressource $r \in R$ (cpu, mémoire ou disque) et au prix de vente de la bande passante.

### 2.2.5 Fonction objectif

La maximisation du profit du fournisseur de VNFs peut être atteinte en maximisant les revenus et en minimisant les coûts de l'énergie et la pénalité. La fonction d'objectif J peut donc être écrite comme suit :

$$\mathbf{J} = \left(\sum_{\mathbf{i}\in\mathbf{I}} \boldsymbol{\sigma_i}\, \mathbf{z_i} - \sum_{\bar{\mathbf{n}}\in\bar{\mathbf{N}}} \boldsymbol{\lambda_{\bar{n}}}\ \mathbf{E_{\bar{n}}} - \sum_{\mathbf{i}\in\mathbf{I}} \boldsymbol{\rho_i}\right) \tag{2.10}$$

où $z_i \in \{0,1\}$ est une variable booléenne qui indique si la chaine de service i est acceptée et $\lambda_{\bar{n}}$ représente le prix de l'énergie au POP $\bar{n}$. La fonction objectif est soumise aux contraintes précédemment mentionnées.

Ce problème est *NP-difficile* car il généralise le problème de « bin packing » (Cauwer, Mehta, & O'Sullivan, Novembre 2016). Par conséquent, il n'est pas possible de trouver une solution optimale dans un temps polynomial en raison de la taille de l'infrastructure et du grand nombre potentiel de demandes de chaîne de service dans les environnements de production. Nous proposons donc dans ce qui suit des algorithmes heuristiques capables de résoudre ce problème pour des scénarios de grande envergure et qui permettent d'explorer efficacement l'espace des solutions pour trouver une solution quasi-optimale.

## 2.3 Solutions Proposées

### 2.3.1 Algorithme de recherche étendue (ESA)

Dans cet algorithme, appelé algorithme de recherche étendue (ESA), nous essayons d'explorer intelligemment toutes les solutions possibles d'intégration de la requête de VNFs dans l'infrastructure et de sélectionner celui qui maximise le profit du fournisseur de VNFs ; ce le conduira à la solution optimale. Comme décrit dans l'algorithme 1, lorsqu'une requête i est reçue, l'algorithme trouve des chemins potentiels reliant la source et la destination de $i$. Pour ce premier algorithme, un chemin potentiel est défini comme un chemin capable de satisfaire la demande de bande passante de la requête (et non pas nécessairement les exigences en termes de latence).

Dans la suite, nous fournissons des détails sur la façon d'explorer les solutions possibles d'intégration de la requête de VNFs dans l'infrastructure pour une chaîne de services $i$ dans un chemin potentiel $P$.

Pour chaque VNF $\in N^i$, nous créons la liste $POPs\,(P,n)$ contenant les POPs appartenant à P et qui sont capables de satisfaire les besoins en ressources de VNF n. Chaque combinaison $(\mathrm{n}_1, \mathrm{n}_2, , \mathrm{n}_{|\mathrm{N}^i|} \in \mathrm{POPs}(\mathrm{P},1) \times \ldots \times \mathrm{POPs}\,(\mathrm{P}, |\mathrm{N}|^{\mathrm{i}})$ représente une solution possible d'intégration de la chaine de service dans l'infrastructure qui satisfait aux conditions suivantes :

1) un VNF $n_{i+1}$ devrait être incorporé soit dans le même POP que $n_i$ ou dans l'un de ses POPs suivants dans le chemin. Ainsi, nous avons :

$$\mathbf{OrderPOP}(\mathbf{n_{i+1}}) \geq \mathbf{OrderPOP}(\mathbf{n_i}) \tag{2.11}$$

où $\mathrm{OrderPOP}(\mathrm{n_i})$ est l'ordre du POP qui héberge le VNF $n_i$ dans le chemin P.

2) Si $POP(n_{i+1}) = POP(n_i)$, les ressources disponibles dans le POP qui héberge $n_i$ (noté par $POP(n_i)$) après l'incorporation de $n_i$ devraient être suffisantes pour héberger le VNF $n_{i+1}$.

**Algorithm 1** Extensive Search Algorithm (ESA)

1: $MaxPEmbed = \oslash$
2: At the receipt of request $i$.
3: **for** each potential path $P$ from source($i$) to destination($i$) **do**
4: **for** $n \in N^i$ **do**
5: $POPs(P,n)$: set of POPs in $P$ having enough resources to host $n$
6: **if** $POPs(P,n) = \oslash$ **then**
7: Reject request $i$
8: **end if**
9: **end for**
10: **for** each feasible embedding $embed$ **do**
11: **if** $Profit(embed) > Profit(MaxPEmbed)$ **then**
12: $MaxPEmbed = embed$
13: **end if**
14: **end for**
15: **end for**
16: **if** $MaxProfitEmbedding \neq \oslash$ **then**
17: Embed($MaxProfitEmbedding$)
18: **end if**

Figure 2.1 Extensive Search Algorithm (ESA)

Pour chaque solution possible pour l'allocation de ressource pour la chaine de service dans l'infrastructure, nous calculons le profit en utilisant la fonction objectif définie dans l'équation 10 (ligne 11). Après avoir analysé toutes les allocations possibles associées à chacun des chemins potentiels, nous sélectionnons celui qui maximise la fonction objectif.

En termes de complexité, le nombre d'opérations pour explorer une solution possible d'intégration de la requête de VNFs dans l'infrastructure d'une requête i à l'aide de l'algorithme ESA est $O(|Paths|.|POPs|.|N^i|))$ où $|Paths|$ est le nombre de chemins disponibles entre la source et la destination de la requête. $|POPs|$ est le nombre de POPs par chemin, et $|N^i|$est le nombre de VNF dans la chaîne de service.

### 2.3.2 Algorithme de recherche étendue avec exigence stricte (ESA-S)

Nous proposons également une autre variante de l'algorithme ESA appelée algorithme de recherche étendue avec exigence de latence stricte (*Extended Search Algorithm with Strict Guarantees* – ESA-S) où les chaines de service ne peuvent être intégrées que dans un chemin qui satisfait non seulement l'exigence en termes de ressources (cpu, mémoire et disque), de bande passante mais aussi de délai de bout en bout souhaité. Dans ce cas, la pénalité sera toujours égale à zéro puisque tous les chemins considérés doivent satisfaire la latence désirée.

Cette solution peut être utilisée lorsque le fournisseur de VNFs ne tolère aucune violation du contrat de service. Cela est particulièrement utile lorsque les fournisseurs de services ont nécessairement besoin de satisfaire les exigences en bande passante et de délai.

Bien que cette solution ait des avantages, elle pourrait réduire l'espace de recherche et augmenter le nombre de chaînes de service rejetées puisque seuls les chemins qui satisfont les exigences en termes de délai seront considérés.

### 2.3.3 Algorithme de recherche restrictive (RSA)

L'une des limites de l'algorithme ESA est sa complexité vu que tous les chemins possibles sont explorés afin de trouver la solution optimale qui permet de maximiser la fonction objectif. Pour réduire davantage la complexité, on pourrait réduire le nombre de chemins explorés et par suite diminuer le temps d'exécution de l'algorithme ESA. Pour se faire, nous proposons un troisième algorithme, appelé algorithme de recherche

restrictive (*Restrictive Search Algorithm* – RSA), qui restreint l'espace de recherche et explore uniquement les chemins avec une utilisation inférieure à un seuil prédéfini $U_{th}$. Cela est raisonnable car les chemins hautement utilisés (avec une utilisation de bande passante élevée et une utilisation de ressources élevée dans les POPs) sont moins susceptibles d'être en mesure de répondre aux demandes des nouvelles chaines de service. L'utilisation d'une ressource $r \in R$ pour un chemin particulier *P* peut être estimée en utilisant l'équation suivante :

$$\mathbf{U_r(P)} = \frac{\mathbf{1}}{\mathbf{|P|}} \sum_{\bar{\mathbf{n}} \in \mathbf{P}} \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{n} \in \mathbf{N^i}} \mathbf{x^i_{n\bar{n}}} \frac{\mathbf{c^{ir}_n}}{\mathbf{c^r_{\bar{n}}}} \tag{2.12}$$

Où |P| est le nombre de POPs dans le chemin considéré et $\frac{\mathbf{c^{ir}_n}}{\mathbf{c^r_{\bar{n}}}}$ est la portion de la ressource **r** consommé par le VNF $n$ dans le POP $\bar{n}$.

L'utilisation de la bande passante sur un chemin peut également être calculée en utilisant l'équation suivante :

$$\mathbf{U_b(P)} = \frac{\mathbf{1}}{\mathbf{|P| - 1}} \sum_{\bar{\mathbf{l}} \in \mathbf{P}} \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{l} \in \mathbf{L^i}} \frac{\mathbf{f^i_{l\bar{l}}}}{\mathbf{b_{\bar{l}}}} \tag{2.13}$$

où |P|-1 est le nombre de liens dans le chemin considéré et $\frac{\mathbf{f^i_{l\bar{l}}}}{\mathbf{b_{\bar{l}}}}$ est la portion de la bande passante consommé par le lien virtuel $l$ dans le lien physique $\bar{l}$.

### 2.3.4 Discussion

Tableau 2.1 Comparaison entre les algorithmes proposés

| Algorithme | Critères pour sélectionner les chemins potentiels | | |
|---|---|---|---|
| | **Besoin de bande passante** | **Exigence de latence** | **Seuil d'utilisation** |
| **ESA** | Oui | Avec pénalité | Non |
| **ESA-S** | Oui | Stricte | Non |
| **RSA** | Oui | Avec pénalité | Oui |

Le Tableau 2.1 résume les principales différences entre les trois solutions proposées :

- ESA : l'algorithme de recherche étendue qui permet d'explorer tout l'espace des solutions possibles afin d'identifier la solution qui maximise l'utilisation de l'infrastructure, le taux d'acceptation des chaines de service, et le profit du fournisseur de VNFs. Cet algorithme considère une pénalité de violation proportionnelle à la violation de l'exigence en latence.

- ESA-S : l'algorithme de recherche étendue avec exigence stricte qui permet d'explorer tout l'espace des solutions possibles. Cependant, il n'accepte que les allocations qui permettent de satisfaire les exigences en termes de délai.

- RSA : l'algorithme de recherche restrictive qui utilise l'utilisation des chemins comme un critère permettant de réduire l'espace de recherche de l'algorithme et ainsi réduire le temps d'exécution de l'algorithme d'ESA. Le RSA considère aussi une pénalité lorsque l'exigence en termes de délai n'est pas respectée.

## 2.4 Conclusion

Dans ce chapitre, nous avons présenté la formulation mathématique du problème sous forme d'un programme linéaire en nombre entier dont l'objectif est de maximiser les revenus des fournisseurs de VNFs tout en respectant les différentes contraintes

de capacité et de performance. Nous avons ensuite proposé trois algorithmes à savoir ESA, ESA-S et RSA pour résoudre le problème de placement et de chaînage avec une complexité réduite.

Nous étudions dans le chapitre suivant la performance de chacun des algorithmes proposés à travers plusieurs simulations à faible et à large échelle.

CHAPITRE 3

# SIMULATIONS ET RÉSULTATS

## 3.1 Introduction

Dans ce chapitre, nous allons évaluer les performances des algorithmes proposés ESA, ESA-S et RSA. Nous allons ainsi considérer plusieurs différents critères de performance tels que le taux d'acceptation, le taux d'utilisation moyenne des POPs et des liens, la latence moyenne, le coût moyen par requête et finalement le profit total du fournisseur de VNFs. Pour cela, nous allons simuler deux scenarios avec des topologies différentes afin d'évaluer la performance des algorithmes proposés à différentes échelles. Le premier scenario à faible échelle considère la topologie « NSFNet » qui a un faible nombre de POPs tandis que le deuxième scenario à large échelle considère la topologie « Giant » contient un nombre important de POPs.

## 3.2 Simulations à faible échelle

Dans cette section, nous allons commencer par décrire les paramètres des simulations à faible échelle que nous avons effectuées ainsi que les principaux résultats obtenus.

Pour évaluer les solutions proposées, nous avons développé un simulateur avec le langage JAVA qui permet de simuler l'ensemble de l'environnement qui comprend l'infrastructure physique, les POPs, les arrivées des chaînes de service et les trois algorithmes d'allocation des ressources proposés.

### 3.2.1 Environnement de simulation

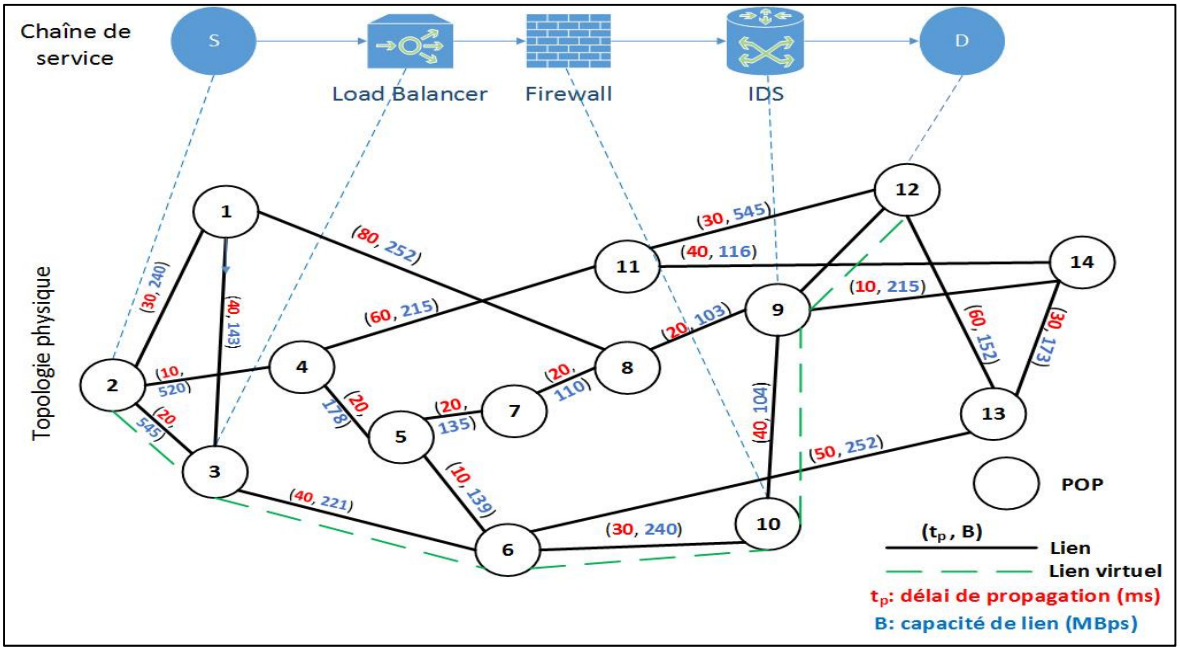


Figure 3.1 Mappage d'une chaîne de service dans les POPs de la topologie NSFNet

Dans cette première expérience, nous avons simulé la topologie NSFNet qui est composée de 14 POPs seulement qui sont connectés à travers 21 liens physiques comme le montre la Figure 3.1. Chaque lien est caractérisé par un délai de propagation (mesuré en ms) et une capacité (mesuré en Gbps) tel que décrit dans la figure. Nous avons choisi la capacité des liens comme la capacité des liens reliant les centres de données Amazon EC2 afin d'imiter un cas réaliste (Feng, Li, & Li, 2012).

Dans nos simulations, nous supposons aussi que chaque POP contient seulement deux serveurs du même modèle (c'est-à-dire, ayant les mêmes capacités en termes de cpu, mémoire et disque et les mêmes modèles de consommation d'énergie). Le Tableau 3.1 présente les caractéristiques de chaque POP de la topologie NSFNet à savoir le nombre de serveurs, leurs modèles, leurs capacités (normalisées entre 0 et 1) et les prix de l'électricité aux emplacements des POPs.

Tableau 3.1 Caractéristiques des POPs de la topologie NSFNet

| ID POP | Nombre de serveurs | Modèle de serveur | Capacité normalisée du serveur | | | Prix de l'électricité ($/heure) |
|---|---|---|---|---|---|---|
| | | | CPU | Mémoire | Stockage | |
| 1, 5, 9 | 2 | Dell Power-Edge R210 | 0.08 | 0.0625 | 0.9 | $2.16 \times 10^{-3}$ |
| 2, 6, 10, 13 | 2 | Dell Power-Edge R515 | 0.25 | 0.5 | 0.9 | $2.88 \times 10^{-3}$ |
| 3, 7, 11 | 2 | HP DL385 G7 | 0.5 | 0.25 | 1 | $3.6 \times 10^{-3}$ |
| 4, 8, 12, 14 | 2 | HP DL585 G7 | 1 | 1 | 1 | $4.32 \times 10^{-3}$ |

Pour simuler l'arrivée des chaines de service, nous avons généré les requêtes de chaines de service suivant un processus de Poisson avec un taux moyen de 0,2 requêtes par seconde. Chaque requête génère aléatoirement entre 1 et 5 VNFs qui peuvent être de même ou de différents types. Nous avons considéré 5 différents types de VNFs. Le prix d'un VNF va de 0,0065 à 0,104 dollars par heure. Ces prix sont similaires aux machines virtuelles proposées par Amazon EC2 (Amazon Web Services, Inc, 2017).

D'autre part, chaque VNF est placée dans une machine virtuelle qui possède un ensemble de ressources (CPU, mémoire et stockage) dont la quantité est normalisée et définie aléatoirement entre 0,01 et 0,04. Les exigences des chaines de service en termes de latence et de bande passante sont également générées de manière aléatoire dans les intervalles [100ms, 200ms] et [0,5MB, 10MB], respectivement. La durée de vie de la chaine de service est choisie aléatoirement entre 10 et 60 minutes.

Enfin, la pénalité payée par le fournisseur de VNFs pour chaque milliseconde excédant la latence requise est calculée comme 10% du revenu généré par la requête.

Toutes les simulations ont été effectuées sur un serveur avec un processeur Intel Core i7 3,33 GHz CPU et 32 Go de RAM sous Ubuntu 14.04.LTS 64-bit OS.

### 3.2.2 Résultats

- **ESA vs. ESA-S :**

Dans ce premier ensemble d'expériences, nous comparons l'algorithme de recherche étendue (ESA) et l'algorithme de recherche étendue avec exigence de latence stricte (ESA-S) dans le cas où la topologie NSFNet est utilisée.

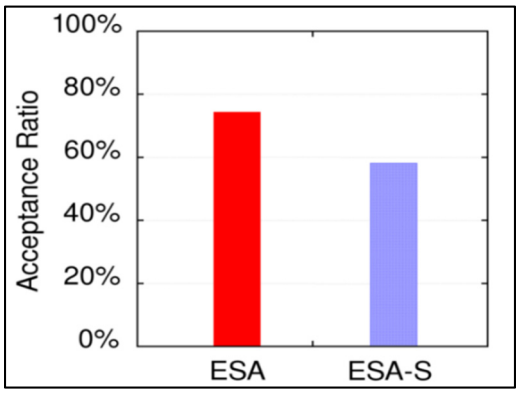


Figure 3.2 Taux d'acceptation

Nous commençons par comparer le taux d'acceptation des différents algorithmes proposés. Le taux d'acceptation est défini comme le nombre de chaines de service qui ont pu être acceptées et dont les ressources ont pu être allouées divisé par le nombre total de chaines de service qui ont été reçus.

La Figure 3.2 compare les algorithmes ESA et ESA-S en termes de taux d'acceptation. Il ressort clairement de la figure que l'algorithme ESA accepte 20% de demandes de chaînes

de service de plus que l'ESA-S. Cela est dû au fait que l'ESA peut accepter des demandes même si l'exigence de délai de bout en bout n'est pas satisfaite.

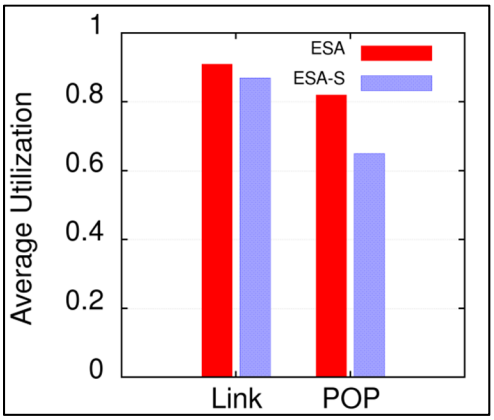


Figure 3.3 Utilisation moyenne

Ce taux d'acceptation élevé se traduit par une utilisation plus élevée des liens de l'infrastructure et des POPs, comme on peut le voir à la Figure 3.3. On peut voir qu'avec l'algorithme ESA, l'utilisation atteint 90% pour les liens et 80% pour les POPs contre 85% et 62% en utilisant l'ESA-S.

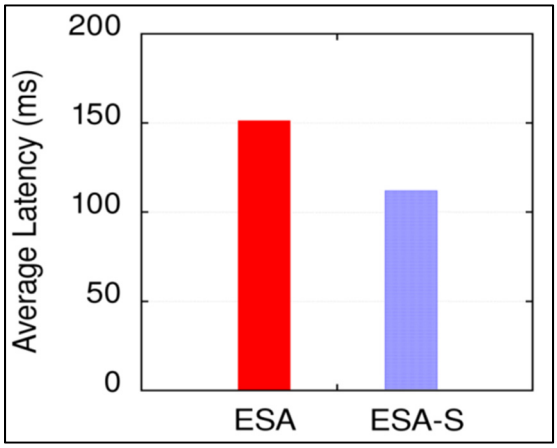


Figure 3.4 Latence moyenne par requête

La Figure 3.4 compare la latence moyenne par chaine de service après l'allocation en utilisant les algorithmes ESA et ESA-S. Il est clair que l'ESA augmente la latence de la chaîne de service par rapport à l'ESA-S, qui accepte uniquement les demandes qui vont satisfaire aux exigences de latence. Pour expliquer ce résultat, rappelons-nous que l'ESA augmente le profit généré par l'intégration de la requête de VNFs pour une chaîne de services dans le chemin encourir moins de coûts d'énergie et une pénalité minimale. En analysant tous les chemins potentiels, l'algorithme ESA trouve de longs chemins où les coûts d'énergie encourus par l'intégration de la requête de VNFs dans l'infrastructure pour une chaîne de services sont extrêmement faibles par rapport aux coûts lorsque des chemins plus courts sont utilisés. Même lorsque la pénalité de violation est considérée, les coûts d''intégration des VNFs des chaines de service dans l'infrastructure (la somme des coûts énergétiques et de la pénalité) sont considérés comme faible et donc le bénéfice (la différence entre les revenues et les coûts) pourrait encore être plus élevé. Ainsi, dans nos expériences, nous avons constaté que l'algorithme ESA privilégie les chemins longs. Cela explique la latence élevée qu'on peut observer dans la Figure 3.4.

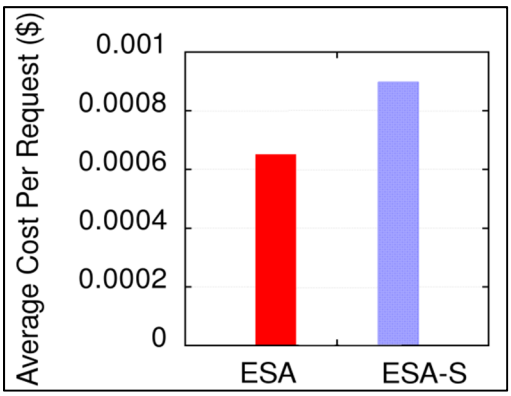


Figure 3.5 Coût moyen par requête
(ESA vs ESA-S)

Cette observation est corroborée par les résultats présentés dans la Figure 3.5 qui montre que le coût moyen par chaine de service lors de l'utilisation de l'algorithme ESA est inférieur de 31% par rapport à celui de l'ESA-S. Il est à noter que le coût moyen par chaine de service inclut les coûts de l'énergie lorsque les ressources ont été allouées pour la chaine de service

ainsi que le coût de pénalité payée par le fournisseur de VNFs dans le cas où l'exigence de latence est violée.

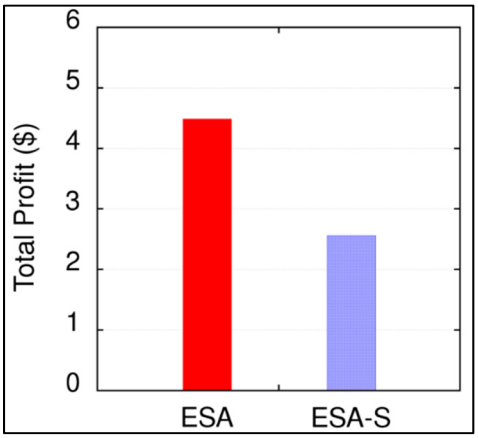


Figure 3.6 Profit total
(ESA vs ESA-S)

Enfin, la Figure 3.6 représente le profit total obtenu par le fournisseur de VNFs pour toutes les requêtes incorporées pendant toute la simulation. Il est clair que l'ESA améliore le profit du fournisseur de VNFs jusqu'à 44% par rapport à l'ESA-S. Ces résultats montrent l'efficacité de l'algorithme ESA d'avoir un revenu beaucoup plus élevé si un modèle de pénalité est utilisé en cas de violation des exigences de latence.

- **ESA vs. RSA :**

Dans ce qui suit, nous comparons l'algorithme de recherche étendue (ESA) avec l'algorithme de recherche restrictive (RSA). L'algorithme RSA est proposé pour réduire la complexité de calcul ESA en réduisant l'espace de recherche. Pour atteindre cet objectif, RSA limite la recherche aux chemins avec une utilisation de ressources inférieure à $U_{th}$. Intuitivement, lorsque l'utilisation du chemin moyen est élevée, il est moins probable qu'il soit capable d'héberger la requête. Dans nos expériences, nous fixons $U_{th}$= 80% parce que, selon la Figure 3.3, le lien moyen ou l'utilisation POP atteint au maximum 80%

lorsque l'algorithme de recherche étendue est utilisé. Cela suggère, qu'au-delà de ce seuil, il n'est pas facile de trouver assez de ressources pour les requêtes.

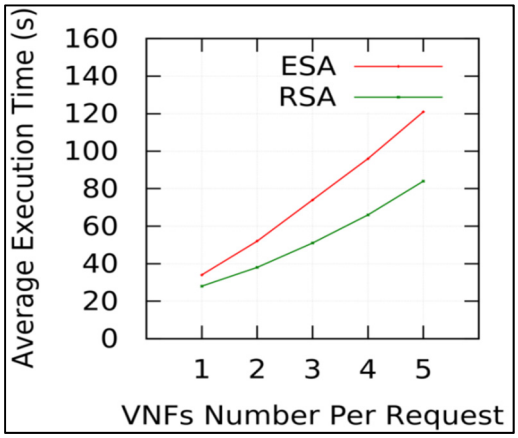


Figure 3.7 Temps d'exécution moyen par requête

La Figure 3.7 représente le temps d'exécution moyen pour ESA et RSA par rapport au nombre de VNFs par demande. Le temps d'exécution moyen est le temps nécessaire à l'algorithme pour explorer des solutions réalisables pour intégrer une seule requête dans l'infrastructure physique. La figure montre que RSA réduit le temps d'exécution par rapport à l'ESA, et en particulier lorsque le nombre de VNFs augmente. Par exemple, pour une requête de 5 VNFs, RSA améliore jusqu'à 40% le temps d'exécution par rapport à l'ESA.

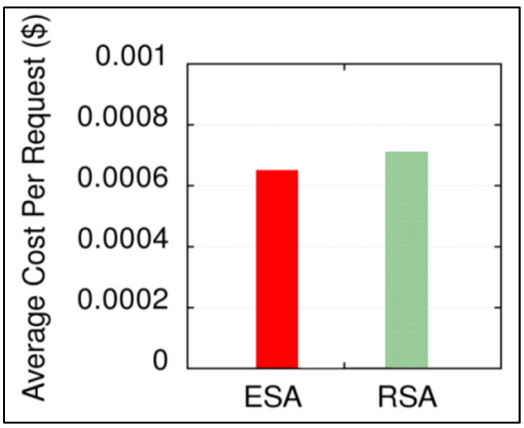


Figure 3.8 Coût moyen par demande (ESA vs RSA)

Ensuite, nous analysons les coûts d'intégration et de profit, la Figure 3.8 compare le coût moyen par demande pour l'ESA et la RSA et montre que RSA entraîne une légère augmentation de 10% du coût d'intégration par rapport à l'ESA.

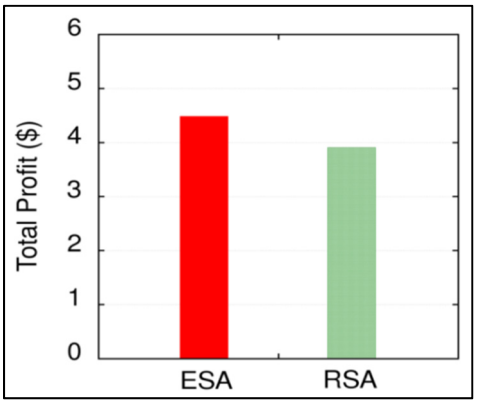


Figure 3.9 Profit total
(ESA vs RSA)

En termes de profit total gagné par le fournisseur de VNFs, RSA obtient des résultats similaires à ceux de l'ESA avec seulement 14% de moins de bénéfices que l'ESA (Figure 3.9).

En conclusion, l'algorithme de recherche restrictive peut améliorer le temps d'exécution jusqu'à 40% par rapport à l'ESA au détriment de 10 à15% de profit en moins. Cela rend cet algorithme plus approprié à la gestion des infrastructures à grande échelle ou lorsque le temps d'approvisionnement nécessaire pour allouer les ressources pour la chaîne de service est cruciale et doit être très faible.

## 3.3 Simulations à large échelle

Dans cette section, nous allons présenter les paramètres utilisés pour les simulations à large échelle ainsi les principaux résultats obtenus.

### 3.3.1 Environnement de simulation

Dans les simulations à large échelle, nous avons utilisé la topologie « Giant » décrite dans la Figure 3.10. Cette infrastructure est composée de 42 POPs connectés par 71 liens physiques. La capacité des liens physiques correspond à la capacité des liens reliant les centres de données Amazon EC2 (Amazon Web Services, Inc, 2017). Nous supposons que chaque POP contient 2 serveurs du même modèle. Le Tableau 3.2 présente les caractéristiques de chacun des POPs en termes de nombre de serveurs, de modèle de serveur, de capacité et de prix de l'électricité à l'emplacement du POP.

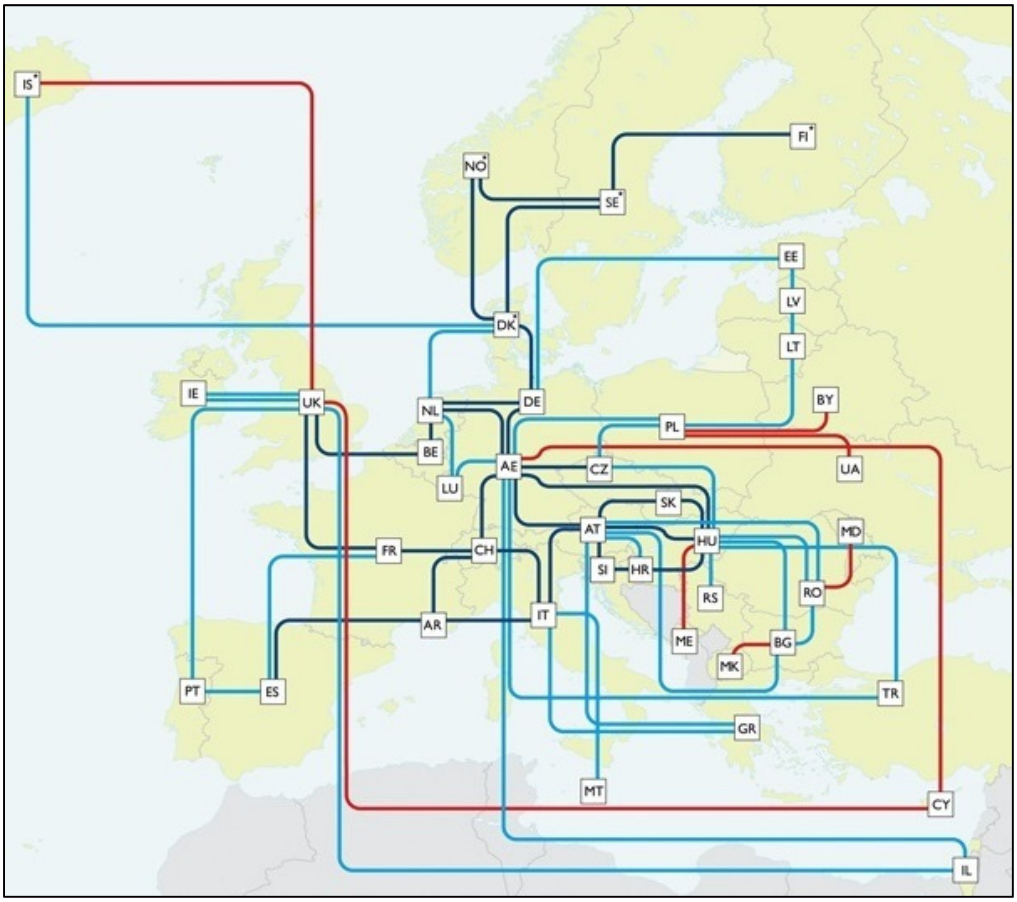


Figure 3.10 Topologie du réseau «GÉANT» (GÉANT topology , 2017)

Tableau 3.2 Caractéristiques des POPs de la topologie « Giant »

| ID du POP | Nombre de serveurs | Modèle de serveur | Capacité normalisée du serveur | | | Prix de l'électricité ($/heure) |
|---|---|---|---|---|---|---|
| | | | CPU | Mémoire | Stockage | |
| IS, DK, LT, RO, TR, GR, CY, CH, SI, MK, ES | 2 | Dell Power-Edge R210 | 0.08 | 0.0625 | 0.9 | $2.16 \times 10^{-3}$ |
| SE, HU, FR, DE, BE, MT, IL, SK, CZ | 2 | Dell Power-Edge R515 | 0.25 | 0.5 | 0.9 | $2.88 \times 10^{-3}$ |
| NO, EE, BY, UA, AT, HR, IE, PT, PL, MD | 2 | HP DL385 G7 | 0.5 | 0.25 | 1 | $3.6 \times 10^{-3}$ |
| BG, RS, ME, UK, NL, LU, FI, LV, IT. AR, AE | 2 | HP DL585 G7 | 1 | 1 | 1 | $4.32 \times 10^{-3}$ |

Les requêtes de chaines de service sont générées selon les mêmes règles utilisées dans les simulations à faible échelle. La seule différence est que le nombre aléatoire de VNFs générés est entre 1 et 30 pour chaque requête (au lieu d'être entre 1 et 5).

Toutes les simulations ont été effectuées sur un serveur avec un processeur Intel Core i7 3,33 GHz CPU et 32 Go de RAM sous Ubuntu 14.04.LTS 64-bit OS.

### 3.3.2 Résultats

- **ESA vs. ESA-S :**

Dans notre premier ensemble d'expériences, nous comparons l'algorithme de recherche étendue (ESA) avec l'algorithme de recherche étendue avec exigence de latence stricte (ESA-S).

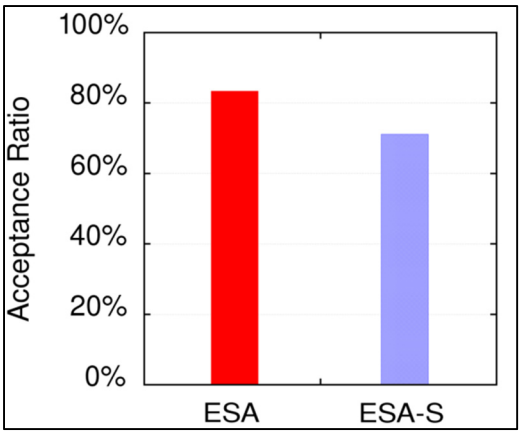


Figure 3.11 Ratio d'acceptation

Cette figure compare ESA et ESA-S en termes de taux d'acceptation. ESA accepte 17% de demandes de chaînes de services de plus que l'ESA-S. Cela est dû au fait que l'ESA peut accepter des chaines de service même si l'exigence de délai de bout en bout n'est pas satisfaite.

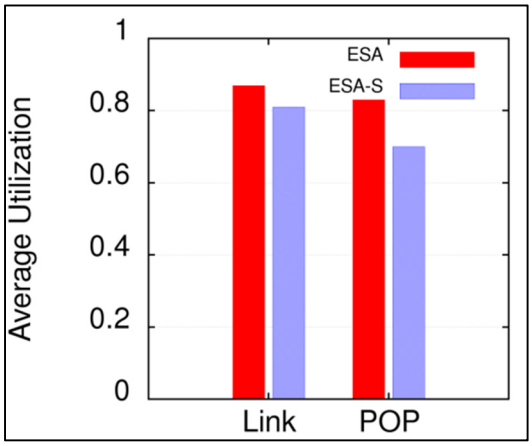


Figure 3.12 Utilisation moyenne

Cette figure montre que, pour l'algorithme ESA, l'utilisation des liens est de 87% et que l'utilisation des POPs est de 83% des POPs contre 81% et 70% pour l'algorithme ESA-S.

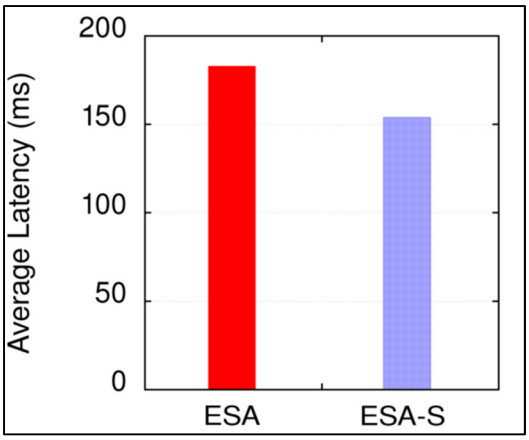


Figure 3.13 Latence moyenne
par chaine de service

Cette figue compare la latence moyenne par chaine de service pour l'ESA et l'ESA-S. Il est évident que l'ESA augmente la latence de la chaîne de service par rapport à l'ESA-S. Cela est attendu vu que l'ESA-S accepte uniquement les chaines de service qui satisfont aux exigences de latence. En effet, nous avons constaté que l'ESA privilège les longs chemins; cela explique la latence élevée indiquée dans la Figure 3.13 (environ 20% de plus par rapport ESA-S).

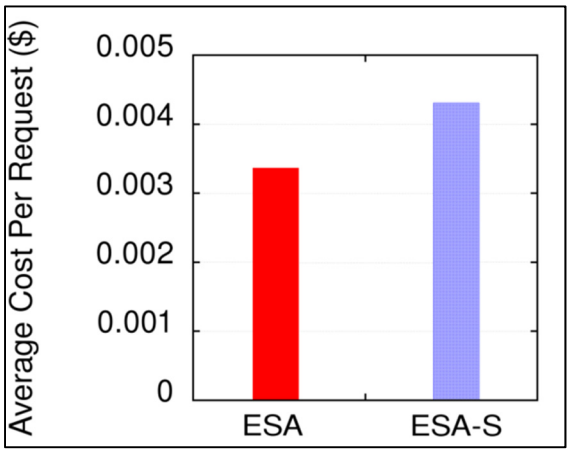


Figure 3.14 Coût moyen par chaine
de service (ESA vs ESA-S)

Cette figure montre que le coût moyen par chaine de service en utilisant l'ESA est 28% inférieur à celui utilisant l'ESA-S. Le coût moyen par requête inclut les coûts énergétiques

lorsque la demande est intégrée et le coût de pénalité en cas de violation de l'exigence de latence.

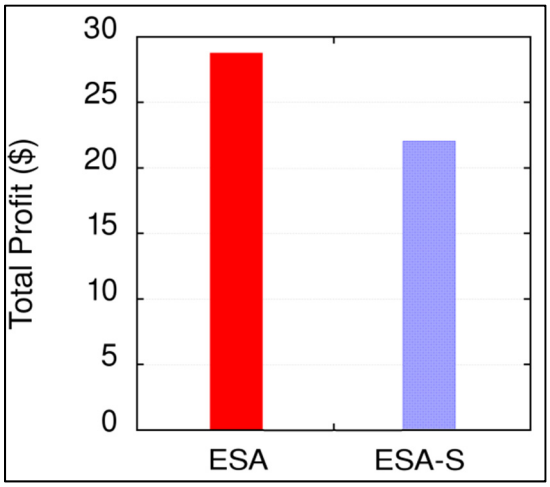


Figure 3.15 Profit total (ESA vs ESA-S)

ESA améliore le profit du fournisseur de VNFs jusqu'à 30% par rapport à l'ESA-S. Ces résultats montrent l'efficacité de l'algorithme ESA pour le fournisseur de VNFs d'avoir un revenu beaucoup plus élevé si un modèle de pénalité est utilisé en cas de violation des exigences de latence.

- **ESA vs. RSA :**

Dans ce qui suit, nous comparons l'algorithme de recherche étendue (ESA) avec l'algorithme de recherche restrictive (RSA) pour les simulations à large échelle.

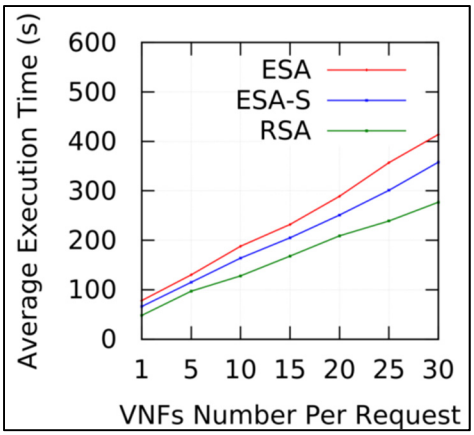


Figure 3.16 Temps d'exécution moyen par requête

Cette figure représente le temps d'exécution moyen pour les algorithmes ESA, ESA-S et RSA par rapport au nombre de VNFs par chaine de service. La figure montre que l'algorithme RSA réduit le temps d'exécution par rapport à ESA et ESA-S, notamment lorsque le nombre de VNFs augmente. Par exemple, pour une chaine de service composée de 30 VNFs, l'algorithme RSA améliore jusqu'à 45% le temps d'exécution par rapport à ESA et de 30% par rapport à ESA-S.

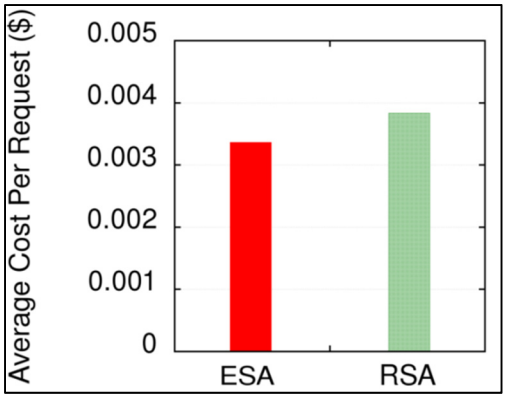


Figure 3.17 Coût moyen par chaine de service (ESA vs RSA)

Cette figure compare le coût moyen par chaine de service pour les algorithmes ESA et RSA. Elle montre que l'algorithme RSA entraîne une légère augmentation de 14% du coût

des ressources allouées par chaine de service par rapport à l'ESA. Cela est dû au fait que le RSA sélectionne seulement les chemins qui satisfont l'exigence en termes de délai.

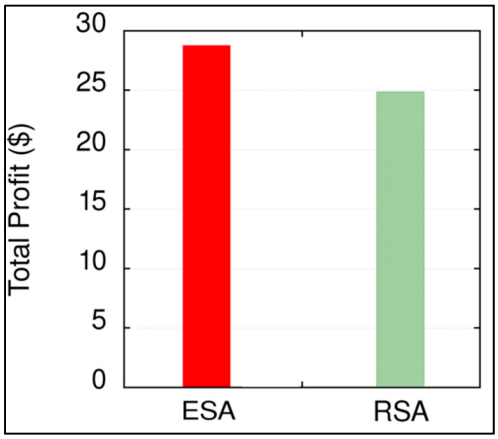


Figure 3.18 Profit total
(ESA vs RSA)

En termes de bénéfice total gagné par le fournisseur de VNFs, RSA obtient des résultats proches à ceux de l'ESA avec seulement 16% moins de profit que l'ESA.

## 3.4 Discussion

En se basant sur les résultats présentés ci-dessus, nous pouvons résumer les avantages et les inconvénients des différents algorithmes proposés comme suit :

- ESA vs ESA-S : en utilisant l'algorithme de recherche étendue qui considère la pénalité de violation (10% du prix de la demande lorsque l'exigence de latence est violée) plutôt que l'algorithme de recherche étendue avec exigence stricte permet de maximiser l'utilisation de l'infrastructure, le taux d'acceptation, et le bénéfice du fournisseur de VNFs.

- ESA vs RSA : l'algorithme de recherche restrictive qui exploite l'utilisation des chemins comme critère permet de réduire le temps d'exécution de l'algorithme d'une façon remarquable (RSA améliore jusqu'à 45% en temps d'exécution par rapport

à ESA) avec juste une faible diminution de profit par rapport à ESA (qui peut aller jusqu'a16% de moins de profit).

- Nos solutions vs. solutions existantes : contrairement aux travaux antérieurs, les solutions proposées dans le présent travail réduisent les coûts énergétiques, maximisent le bénéfice total du fournisseur de VNFs tout en tenant compte des différents prix de l'électricité dans les emplacements des POPs et de l'exigence de délai de bout en bout.

## 3.5 Conclusion

En se basant sur les résultats des simulations, nous déduisons que l'algorithme ESA comparé à ESA-S permet de d'augmenter davantage l'utilisation de l'infrastructure, le taux d'acceptation, et par conséquent le bénéfice du fournisseur de VNFs même si ce dernier paye des pénalités aux fournisseurs de service.

D'autre part, si on considère que le temps d'exécution est un critère important, notamment dans les cas où le temps nécessaire pour l'allocation de ressources est crucial, les simulations prouvent que l'algorithme RSA, bien qu'il génère moins de profit, permet de réduire le temps d'exécution d'une façon remarquable par rapport à l'algorithme ESA.

# CONCLUSION GÉNÉRALE

Dans ce mémoire, nous avons abordé l'un des principaux défis auxquels sont confrontés les fournisseurs de services cloud. Ce défi consiste à allouer efficacement des ressources des chaines de service réseau composées de VNFs de manière à réduire les coûts opérationnels et à maximiser leurs profits.

Nous avons d'abord formulé le problème de placement et de chaînage de VNFs comme un problème linéaire en nombres entiers et ensuite nous avons proposés trois nouveaux algorithmes (ESA, ESA-S et RSA) qui prennent en considération les différents prix d'énergie dans les emplacements des POPs et les modèles de consommation d'énergie des POPs.

Nous avons ensuite effectué plusieurs simulations qui démontrent que ces algorithmes sont capables de trouver des solutions quasi-optimales pour l'allocation de ressources des chaines de services dans l'infrastructure qui permettent de 1) maximiser le profit du fournisseur, 2) satisfaire les exigences de chaîne de service en termes de bande passante, de latence et de CPU, de mémoire et de disque et 3) minimiser le temps de calcul et de l'exécution de l'algorithme.

L'importance de notre travail se manifeste dans l'évolutivité des algorithmes proposés. Ainsi, quel que soit la taille de la topologie utilisée, les fournisseurs de VNFs pourront garantir un profit assez élevé tout en gardant une qualité de service acceptable en termes de performance.

Nous espérons que ce travail sera adopté par les fournisseurs de VNFs pour améliorer la qualité de leur service et augmenter leurs profits. Comme perspectives futures, nous comptons intégrer les solutions proposées dans les plateformes de gestion de nuage existantes comme OpenStack afin d'évaluer davantage leurs performances dans des environnements en production.

Nous comptons aussi étendre les solutions proposées afin de considérer l'optimisation en temps réel des emplacements des VNFs en fonction de la demande qui peut varier en fonction du temps. Dans ce cas, d'autres paramètres doivent être pris en compte tels que le cout de la migration des fonctions virtuelles et leur impact sur la continuité du service.

## LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES


(ETSI). (2015). *European telecommunications standards institute, Industry Specifica-tion Groups (ISG)—NFV*. Sophia-Antipolis Cedex, France.

(NFV), Network Functions Virtualisation. (s.d.). *Terminology for Main Concepts in NFV*. ETSI GS NFV 003.

Amazon Web Services, Inc. (2017). *amazon*. (Amazon Web Services, Inc) Consulté le 2017, sur https://aws.amazon.com/fr/ec2/pricing/

Bari, M. F., Boutaba, R., Esteves, R., Granville, L. Z., Podlesny, M., Rabbani, M. G., . Zhani, M. F. (2013). Data Center Network Virtualization: A Survey. *IEEE Communications Surveys and Tutorials, 15*(2), 909-928.

Bari, M. F., Chowdhury, S. R., Ahmed, R., & Boutaba, R. (2015). On orchestrating virtual network functions. *IEEE Conference on Network and Service Management (CNSM)*, (pp. 50-56). Barcelona.

Cauwer, M. D., Mehta, D., & O'Sullivan, B. (Novembre 2016). The Temporal Bin Packing Problem: An Application to Workload Management in Data Centres. *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, (pp. 157-164). San Francisco, CA, USA - United States USA - United States.

Chiosi, M., Clarke, D., Willis, P., & Reid, A. (October 2012). Network Functions Virtualisation. *SDN and OpenFlow World Congress*, (pp. 1-16). Darmstadt-Germany.

Clayman, S., Mainiy, E., Galis, A., Manzaliniz, A., & Mazzocca, N. (2014). The dynamic placement of virtual network functions. *IEEE Network Operations and Management Symposium (NOMS)*, (pp. 1-9). Krakow, Poland.

Feng, Y., Li, B., & Li, B. (2012). Jetway: Minimizing costs on inter-datacenter video traffic. *ACM international conference on Multimedia*, (pp. 259-268). Nara, Japan.

G Gerules and C Janikow. (2016). A survey of modularity in genetic programming. *IEEE Congress on Evolutionary Computation (CEC)*, (pp. 5034-5043). Vancouver, BC.

GÉANT topology . (2017, 01). *GÉANT topology map*. (GÉANT) Consulté le 2017, sur https://www.geant.org/Networks/Pan-European_network/Pages/GEANT_topology_map.aspx

Guerzoni, R. (2012). Network functions virtualisation: An introduction, bene-fits, enablers, challenges and call for action. Introductory white paper. *SDN & OpenFlow World Congress*, (pp. 1-16). Darmstadt-Germany.

H. Lieberman and T. Selker. (2000). Out of context: Computer systems that adapt to, and learn from, context. *IBM Systems Journal*, 617-632.

H. Moens and F. De Turck. (Nov 2014). VNF-P: A model for efficient placement of virtualized network functions. *International Conference on Network and Service Management (CNSM)*, 418–423.

Huang, P.-H., Li, K.-W., & Wen, C. H.-P. (Oct 2015). NACHOS: Network-aware chains orchestration selection for NFV in SDN datacenter. *IEEE International Conference on Cloud Networking (CloudNet)*, 205 - 208.

Kushida, K. E., Murray, J., & Zysman, J. (2011). Diffusing the Cloud: Cloud Computing and Implications for Public Policy. *SpringerLink*, 209--237.

Lee, S.-I., Pack, S., Shin, M.-K., & Paik, E. (2014). Resource management for dynamic service chain adaptation. *Internet Research Task Force (IRTF)*, 1-18.

Luizelli, M. C., Bays, L. R., Burio, L. S., Barcellos, M. P., & Gasparyl, L. P. (2015). Piecing together the NFV provisioning puzzle: efficient placement and chaining of virtual network functions. *IEEE International Symposium on Integrated Network Management (IM)*, 98-106.

M. Mechtri and C. Ghribi and D. Zeghlache. (September 2016). A Scalable Algorithm for the Placement of Service Function Chains. *IEEE Transactions on Network and Service Management*, 1 - 7.

Mehraghdam, S., Keller, M., & Karl, H. (2014). Specifying and placing chains of virtual network functions. *IEEE International Conference on Cloud Networking (CloudNet)*, (pp. 7 - 13). Luxembourg, Luxembourg.

Mijumbi, R., Serrat, J., Gorricho, J.-L., Bouten, N., Turck, F. D., & Boutaba, R. (2016). Network Function Virtualization: State-of- the-Art and Research Challenges. *IEEE Communications Surveys Tutorials*, 236 - 262. Consulté le 2017, sur http://www.vmware.com/fr/solutions/virtualization.html

Mijumbi, R., Serrat, J., Gorricho, J.-L., Bouten, N., Turck, F. D., & Davy, S. (April 2015). Design and evaluation of algorithms for mapping and scheduling of virtual network functions. *IEEE Conference on Network Softwarization (NetSoft)*, 1–9.

*Network Functions Virtualisation (NFV),Placement and chaining.* (s.d.). ETSI GS NFV 005.

Rankothge, W., Ma, J., Le, F., Russo, A., & Lobo, J. (May 2015). Towards making network function virtualization a cloud computing service. *IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 89–97.

Roadmap, NIST Cloud Computing Standards. (s.d.). *NIST Cloud Computing Standards Roadmap Dorking Group.* NIST Cloud Computing Program, Information Technology Laboratory.

Sahhaf, S., Tavernier, W., Rost, M., Schmid, S., Collea, D., Pickavet, M., & Demeestera, P. (2015). Network service chaining with efficient network function mapping based on service decompositions. *IEEE Conference on Network Softwarization (NetSoft)*, (pp. 1-5). London.

Zhang, Q., Zhani, M. F., Boutaba, R., & Hellerstein, J. L. (2013). Harmony: Dynamic heterogeneity-aware resource provisioning in the cloud. *IEEE International Conference on Distributed Computing Systems (ICDCS)*, (pp. 510 - 519). Philadelphia, PA, USA.

Zhang, Q., Zhu, Q., Zhani, M. F., Boutaba, R., & Hellerstein, J. M. (2013). Dynamic service placement in geographically distributed clouds. *IEEE Journal on Selected Areas in Communications (JSAC), 31*(10), 14 - 28.