Avis de Soutenance

Madame Manel GHERARI

Spécialité : Informatique et Applications

Soutiendra publiquement ses travaux de thèse intitulés
« Solution architecturale pour l'automatisation du plan de gestion des réseaux INC (In-Network Computing) »

Travaux dirigés par Monsieur Yacine GHAMRI DOUDANE et Madame halima Elbiaze
Cotutelle avec l'université "Université du Québec à Montréal" (Canada)

Soutenance prévue le **jeudi 19 novembre *2026***

Lieu : Pavillon Président-Kennedy 201,
avenue du Président-Kennedy,
local PK-2150 Montréal (Québec) H2X 3Y7
Salle : PK-5115

Composition du jury proposé

M. Yacine GHAMRI DOUDANE	Professeur des universités	Université de La Rochelle	Directeur de thèse
Mme Halima ELBIAZE	Professeure des universités	Université du Québec à Montréal	Co-directrice de thèse
Mme Melike EROL-KANTARCI	Professeure des universités	Université d'Ottawa	Examinatrice
Mme Diala NABOULSI	Professeure des universités	École de technologie supérieure Montréal	Rapporteure
M. Stefano SECCI	Professeur des universités	Conservatoire national des arts et métiers — Cnam, Paris	Rapporteur
M. Florent AVELLANEDA	Professeur	Université du Québec à Montréal	Examinateur
M. ElMAHDI DRIOUCH	Université du Québec à Montréal	Invité	

Résumé :

Les applications immersives, telles que la communication holographique et le métavers, se caractérisent par la nature volumineuse, dynamique et hétérogène des données qu'elles manipulent. Elles reposent sur deux catégories d'actifs numériques aux caractéristiques contrastées. Les actifs liés au service, comme les environnements 3D, sont volumineux et stables. À l'inverse, les actifs liés à l'utilisateur, tels que les avatars, évoluent rapidement selon les mouvements, les gestes et l'activité physique de l'utilisateur. Cette dualité, renforcée par un pipeline intégrant l'encodage, le décodage, le rendu et la mise à jour de l'état utilisateur, impose des exigences conjointes en communication, en calcul et en mise en cache (3C). Afin de maintenir la sensation d'immersion, le délai entre l'action de l'utilisateur et la restitution du contenu doit rester de l'ordre de 20 ms. Cette contrainte laisse une marge limitée pour le transfert des données, l'exécution des fonctions du pipeline et la transmission du contenu rendu. Les infrastructures cloud-edge actuelles peinent toutefois à satisfaire ces exigences : le cloud centralisé peut introduire des délais de propagation importants, tandis que l'edge dispose de capacités de calcul et de stockage plus limitées. Dans ce contexte, l'In-Network Computing (INC) constitue une extension prometteuse du continuum cloud-edge, en permettant à certains équipements réseau programmables de contribuer au traitement et à la mise en cache des données. La première contribution, présentée dans l'article « Enhancing Holographic Video Communication with In-Network Computing: An Efficient 3C Resource Allocation Framework », propose une allocation conjointe des ressources 3C dans le continuum cloud-edge-INC. Le pipeline immersif est modélisé sous la forme d'une chaîne de fonctions de service (SFC) intégrant conjointement le placement des fonctions et des actifs numériques. Le problème est formulé comme un programme non linéaire en nombres entiers mixtes et résolu à l'aide d'une approche hybride combinant l'optimisation par essaim particulaire et une stratégie gloutonne. Les résultats démontrent la valeur ajoutée de l'INC en améliorant le taux de rafraîchissement et en maintenant une latence inférieure à 22 ms. La deuxième contribution, présentée dans l'article « Optimizing Immersive Services with Parallel In-Network Rendering and Deep RL », approfondit l'orchestration des actifs liés au service au moyen d'une ferme de rendu distribuée fondée sur l'INC. Lorsqu'un actif atteint un niveau suffisant de demande et de popularité, son rendu est exécuté en parallèle. Les résultats partiels sont ensuite agrégés, puis le contenu rendu est répliqué et placé en pré-cache. Les requêtes ultérieures peuvent ainsi accéder directement aux répliques sans réexécuter l'ensemble du pipeline de traitement. Ce problème d'optimisation multi-objectif est résolu par l'apprentissage par renforcement profond multi-agent. Les résultats montrent une latence moyenne inférieure à 17 ms et un taux d'acceptation d'environ 90 %. Ils montrent que la parallélisation accélère le traitement initial, tandis que la réplication et le pré-cache réduisent la latence des accès ultérieurs. La troisième contribution, présentée dans l'article « SenseGate: Human Activity Sensing-Driven SFC Migration Stability for Immersive Networks », étend le cadre aux actifs propres à l'utilisateur, dont l'évolution dépend du contexte d'activité humaine. Leur orchestration est modélisée par une SFC pouvant être migrée dans le continuum cloud-edge-INC. SenseGate combine un apprentissage par renforcement multi-agent coopératif, qui guide le choix des destinations selon la latence, l'état des ressources et le contexte utilisateur, avec un mécanisme de stabilisation qui filtre les décisions afin de limiter les migrations impulsives et les oscillations. Les résultats montrent une réduction du nombre total de migrations avec une latence moyenne de 15 ms.