

INFORSID 2018 Forum Jeunes Chercheuses Jeunes Chercheurs

Inforsid2018 - Forum of junior researchers

Junior Dongo¹, Soufiane Faieq², Franck Jeveme Panta³, and Thomas Polacsek⁴

¹Université Paris-Est - Laboratoire d'Algorithmique, Complexité et Logique, 61 Avenue du Général de Gaulle 94010 Créteil
junior.dongo@u-pec.fr

²Université Grenoble Alpes, CNRS, Grenoble INP*, LIG CS 40700 - 38058 Grenoble cedex 9 F-38000, Grenoble
soufiane.faieq@univ-grenoble-alpes.fr

LRIT, Research Unit Associated to the CNRST (URAC 29), Faculty of Sciences

Mohammed V University in Rabat, Morocco

³Université de Toulouse III, Paul Sabatier Institut de Recherche en Informatique de Toulouse (IRIT), CNRS, 118 Route de
Narbonne 31062 Toulouse Cedex 9

Franck.Panta@irit.fr

⁴ONERA, 2, avenue Edouard Belin 31000 Toulouse

thomas.polacsek@onera.fr

RÉSUMÉ. Reprenant trois des contributions présentées lors du Forum Jeunes Chercheuses Jeunes Chercheurs d'Inforsid 2018, le présent article s'interroge les apports possibles des architectures réseaux centrée sur la donnée, cherche à établir une composition de services sensible au contexte et vise à proposer une modélisation des mécanismes de filtrage de grandes collections de vidéo dans le cadre des systèmes de vidéosurveillance.

ABSTRACT. This article presents three of the contributions presented at the Inforsid 2018 Young Researchers Forum. This article examines the possible contributions of data-oriented network architectures, it seeks to establish a context-sensitive composition of services and it aims to propose a modeling of the filtering mechanisms of large video collections in CCTV systems.

MOTS-CLÉS. Big data, Réseau centré données, Protocole par nommage de données, Composition de Services, Réutilisation de Services, Vidéosurveillance, forensic, métadonnées.

KEYWORDS. Big data, Content Centric Networking, Named Data Networking, Service Composition, Service Reuse, CCTV system, forensic, metadata.

1. Introduction

Le Forum Jeunes Chercheuses Jeunes Chercheurs (JCJC) d'Inforsid 2018 a eu lieu à Nantes, lors du congrès annuel de l'association Inforsid. Événement bisannuel, le JCJC permet à des doctorantes de première ou deuxième année de venir présenter leurs travaux. Pour la communauté des chercheuses centrées sur l'informatique des organisations et les systèmes d'information, les JCJC sont l'occasion de prendre le pouls de la recherche francophone, de voir les thèmes de recherche qui animent la communauté. Les JCJC sont un instantané qui donne à voir un panorama des questionnements scientifiques en cours et la neuvième édition du Forum n'a pas dérogé à la règle. Ainsi, ce sont six doctorantes et doctorants qui sont venus présenter leurs travaux en session plénière puis devant leur poster.

La première impression qui se dégage des travaux présentés est l'importance des thématiques relatives aux données massives. En effet, si les données massives représentent l'un des grands défis informatiques de la décennie, comme le souligne Stéphane Mallat, titulaire d'une Chaire « Sciences des données » créée cette année 2018 au Collège de France, les questions portant sur la collecte, la gestion, le traitement et l'emploi de ces données sont au cœur des préoccupations de la communauté Inforsid. D'ailleurs trois présentations des JCJC s'inscrivent dans cette thématique. Partant de données brutes et non structurées, la thèse de Rabah Tighilt Ferhat vise à extraire et à restructurer ces données dans le but d'effectuer des

traitements décisionnels. Concernant le traitement, et plus précisément le traitement efficient, la thèse de Nabil El Malki se focalise sur l'optimisation des accès complexes et répétitifs appliqués à des données massives. Toujours sur la problématique des traitements massifs, mais avec une approche centré infrastructure, la thèse de Junior Dongo interroge les apports possibles des architectures réseaux centrée sur la donnée. Se focalisant non plus sur le traitement brut, mais sur la donnée et les informations afférentes la thèse de Franck Jeveme Panta vise à proposer une modélisation des mécanismes de filtrage de grandes collections de vidéo dans le cadre des systèmes de vidéosurveillance. Toujours sur le thème de la gestion de l'information, la thèse de Tiphaine Van de Weghe cherche à apporter des solutions aux chercheuses et chercheurs en Sciences Humaines et Sociales pour collecter, organiser et utiliser leurs données. Pour finir, parmi l'ensemble des données produites, ce trouve aussi les profils utilisateurs que possèdent les différents services que nous utilisons et ces justement à partir de ces profils que Soufiane Faieq, dans le cadre de sa thèse, cherche à établir une composition de services sensible au contexte.

Comme nous le voyons au travers de ces six sujets de thèses présenté ici, la communauté s'est très largement emparée de la question de la massification des données, de leur traitement et, en parallèle, continue son travail sur la gestion de l'information. Cependant, en marge de ces travaux essentiels, nous pouvons espérer que, dans un avenir proche, les questions relatives à l'impact de ces collectes massives de données et aux capacités sans cesse accru de traitements seront elles aussi embrassées par la communauté. Le présent article reprend trois des contributions présentées lors des JCJC.

Proposée pour être une des futures architectures d'Internet, Named-Data Networking (NDN) apporte un changement de paradigme dans le mode de communication réseau en passant d'un adressage du matériel hôte à un adressage centré sur la donnée par l'utilisation de noms. L'adoption de cette architecture implique un changement dans la manière de penser et de concevoir les nouvelles applications. En tant que système applicatif distribué, le Big Data est fortement concerné par ce changement. Une réflexion est donc indispensable afin d'anticiper le changement en termes d'infrastructure réseau et de permettre une transition transparente pour les applications Big Data vers NDN. La section 2 de cet article présente notre proposition en terme d'approche architecturale permettant d'adapter les squelettes de programmations classiques Big Data au paradigme NDN.

Les environnements intelligents de demain sont en particulier caractérisés par une vision centrée sur l'utilisateur et le grand nombre de services disponibles. Ceci rend les approches classiques de l'Informatique Orientée Services inappropriées pour la création des services dans ces environnements. La section 3 de cet article porte sur l'intégration du contexte de l'utilisateur ainsi que sur les techniques de l'apprentissage automatique pour la composition de services. Dans un premier temps, un cadre de travail centré sur la notion du contexte est proposé pour le développement des services dans les environnements intelligents. Ce cadre de travail permet d'intégrer les technologies récentes dans la création des services par composition. Dans un deuxième temps, un système basé sur les techniques de recommandation pour la composition de services permet d'améliorer la qualité des services recommandés en analysant les traces d'exécution des applications à base de services. L'approche est illustrée avec le cas d'étude du projet SMARTROAD relevant du domaine du transport intelligent.

L'évolution des systèmes de vidéosurveillances ainsi que le déploiement massif des caméras de surveillance dans les métropoles sont à l'origine de l'énorme croissance des données vidéo. En raison de cette croissance rapide, la gestion et l'analyse des contenus vidéo deviennent complexe et très coûteux en terme temps. Alors, il devient très important d'extraire les informations ou métadonnées utiles, qui

offrent de nombreuses possibilités aux opérateurs de vidéosurveillance. Dans la section 4 de cet article, nous proposons une approche qui consiste en la modélisation de données générées par les systèmes de vidéosurveillance et la mise en œuvre des algorithmes de requêtage basés métadonnées pour le filtrage de grands volumes de données vidéo. Les résultats du filtrage permettront de faciliter l’analyse des contenus vidéo en réduisant le nombre de de vidéo à analyser et implicitement le temps de traitement.

2. Repenser le Big Data pour les Réseaux du Futur

2.1. Contexte

Dans la communauté Big Data, l’utilisation du squelette de programmation MapReduce a été considérée comme l’une des principales approches permettant de répondre à une demande permanente et croissante des ressources informatiques imposées par des données massives. Son importance s’explique par l’évolutivité du paradigme MapReduce qui permet une exécution massivement parallèle et distribuée sur un grand nombre de nœuds de calcul. L’émergence de nouvelles architectures réseaux telles que Information-Centric Networking (ICN) offre des perspectives nouvelles pour déléguer le support du Big Data depuis la couche applicative vers la couche réseau. L’une des architectures réseaux ICN dominantes est Named Data Networking (NDN) (L. Zhang *et al.*, 2010) qui est une architecture centrée sur la donnée. L’objectif ultime de cette architecture est de proposer une transition des communications centrées sur les machines hôtes vers un modèle de communications centrées sur les données.

L’initiative NDN a été financée principalement par la Fondation Nationale Américaine pour la science (NSF) dans le cadre du projet Future Internet Architecture (FIA). NDN dispose d’une spécification solide du protocole qui est en phase d’être standardisée en plus de plusieurs implémentations et d’un déploiement sous forme d’un réseau universitaire multinational.

Plusieurs défis techniques et fonctionnels sont liés au design d’une architecture Big Data sur les réseaux NDN. Dans notre effort, nous nous focalisons sur trois principaux défis qui sont la fiabilité du système, la conformité aux interfaces de programmations existantes, et les performances durant l’exécution. Dans ce cadre, la contribution au domaine du Big Data par cet article est de proposer une nouvelle approche pour :

- une architecture Big Data totalement distribuée et tolérante aux pannes ;
- une implémentation respectant les principales interfaces et squelette de programmations existants ;
- une utilisation optimisée des mécanismes NDN pour la mise en cache et l’agrégation des données pour obtenir une efficacité en termes de communication entre les nœuds.

2.2. État de l’art

Le traitement Big Data peut être considéré en deux grandes phases : le stockage des données, celui-ci s’effectue généralement sur un Système de Fichiers Distribués (DFS) et le lancement de calculs distribués sur les données. Différentes approches de calcul distribué ont été proposées dans le cadre des traitements Big Data, le plus utilisé étant le MapReduce (Zhao, Pjesivac-Grbovic, 2009). Dans cette approche, les données sont généralement stockées sur le système de fichiers distribués Hadoop (HDFS) (Shvachko *et al.*, 2010) et les calculs sur les données effectués en utilisant le paradigme MapReduce.

L'architecture NDN présente deux principaux types de composants : des Consumers et Producers. Un Producer a pour rôle de produire la donnée. Il expose le préfixe des noms de données qu'il est capable de servir. Un Consumer envoie un intérêt pour une donnée en la nommant. Chaque élément de données est identifié par un nom unique. Les deux types de paquet sur NDN sont : Interest et Data, comme le montre la structure de la Figure 1. Aucune information concernant la source ou la destination d'une donnée n'est contenue dans un intérêt ou une data. NDN permet la mise en cache de la donnée sur le réseau. Chaque paquet étant signé lors de sa création, l'intégrité des données peut être facilement vérifiée par n'importe quel nœud.

Interest Packet	Data Packet
Name	
Selectors (order preference, publisher filter, exclude filters...)	MetaInfo (content type, freshness period, ...)
Nonce	Content
Guiders (Scope, Interest lifetime)	Signature (signature type, key locator, signature bits, ...)

Figure 1. Types de paquets NDN

Dans un effort d'adapter une implémentation Big Data classique au réseau NDN, une approche Big Data centrée sur la donnée a été proposée par les travaux réalisés à l'Université de L'Arizona (Gibbens *et al.*, 2017). Leur approche consiste en un portage de Hadoop sur NDN. En effet NDN étant basé sur la donnée et non sur les adresses, une profonde modification de Hadoop a été nécessaire afin de permettre une exécution sur NDN. Cependant l'une des principales limitations de Hadoop dans sa version 2 concerne la structure client serveur du système de fichiers distribué. Un seul NameNode gère les index des blocs de données, lesquels sont réparties sur les DataNodes. Le secondary NameNode joue quant à lui un rôle clé pour l'importation de nouvelles données dans le système HDFS (Hadoop Distributed File System). Ainsi, toute défaillance du NameNode a un impact direct sur les calcul Map/Reduce en cours ou à venir.

Une approche de système de fichiers distribués a été proposée pour les réseaux NDN en vue de faciliter les traitements Big Data (Chen *et al.*, 2015). Cette approche n'utilise pas de réPLICATION de données, qui est primordiale dans le calcul Big Data, et nécessite une profonde refonte de l'architecture NDN.

2.3. Problématique

Notre effort a pour objectif d'adresser l'intégration et l'amélioration des notions du Big Date, incluant les systèmes de fichiers distribués et les squelettes de programmations, pour les technologies réseau du futur. Pour ce faire, nous avons identifié deux problématiques cruciales qui sont le gaspillage de la bande passante et le point central de défaillance. En effet, l'utilisation du mode actuel de communication d'hôte à hôte, laisse apparaître des problèmes en terme de duplication dans le transfert des données identiques échangées. Par exemple dans le cadre d'une réPLICATION de données, si nous avons trois nœuds qui doivent réPLiquer une donnée depuis un nœud source, cela produira trois paquets de la source vers les trois

destinations. Il n'y a donc pas d'optimisation de paquet, vu qu'en l'espèce, il s'agit de la même donnée contenue dans les trois paquets.

De plus, la majorité des systèmes de fichiers actuels utilise un composant central pour la gestion des données. Cela est le cas par exemple de Hadoop, où le NameNode stocke les métadonnées ainsi que l'arborescence de tous les fichiers du système de fichiers. Ainsi, en cas de perte du NameNode, il y a un risque de perte de toutes les données. Il représente donc un point unique de défaillance.

L'objectif est donc d'apporter une solution pleinement distribuée qui évite l'utilisation d'un composant central et qui permet une agrégation des paquets réseaux afin de réduire la consommation réseau.

2.4. Actions réalisées

Notre première action a consisté en la définition d'un système de fichiers distribués basé sur NDN. Ce DFS a les caractéristiques suivantes :

- Pleinement distribué : ne contient pas un unique point de défaillance ;
- Résilient : mécanisme de recouvrement intégré pour tous les composants ;
- Sécurisé : chaque donnée est signée et encryptée ;
- Adaptable : supporte les données de petite et de grande taille, accepte des données de formats différents.

Notre système est composé de 3 composants :

- Client : utilisé par l'utilisateur pour initier une sauvegarde de donnée (demande de réPLICATION). La donnée est découpée en segments et rendue disponible sur le réseau. Un intérêt pour une demande de réPLICATION est émis à destination des nœuds de stockage.
- Storage : il s'agit du composant responsable du stockage des réPLICAS de données. Basé sur la capacité de stockage du nœud, il répond aux demandes de réPLICATION émises par le client. Lors du traitement d'une demande de réPLICATION, si un autre réPLICA est nécessaire, le storage prend en charge la tâche d'effectuer la demande de réPLICATION. Il est donc responsable dans ce cas de vérifier la disponibilité du nœud ayant pris en charge sa demande de réPLICATION en utilisant le composant de heartbeat.
- Heartbeat : Un mécanisme utilisé pour vérifier la disponibilité des réPLICAS sur le réseau. L'algorithme proposé dans notre approche, est un algorithme dans lequel un nœud vérifie un autre nœud. Il y a donc le maintien d'un pointeur circulaire sur les nœuds qui répliquent la donnée. Cela aide à avoir cet aspect complètement distribué. En effet, les réPLICAS n'ont pas à rapporter périodiquement leur présence à un composant central. Chaque nœud est responsable de la vérification d'un autre nœud. Cela se fait en envoyant périodiquement un heartbeat Interest. Le nœud vérifié, si disponible répondra avec un Data packet ; sinon un timeout sera déclenché et le nœud storage de contrôle devra trouver un autre nœud (Storage) pour répliquer les données de celui défaillant.

Pour la conception de notre proposition de DFS, nous avons choisi d'exprimer une première vue de notre algorithme avec un diagramme de collaboration UML. Ceux-ci sont utilisés lorsque nous voulons comprendre le flux de messages et l'organisation structurelle. Il montre explicitement les demandes et aussi les noms exposés. La Figure 2 affiche une topologie avec 4 nœuds : le Client pilote le chargement et

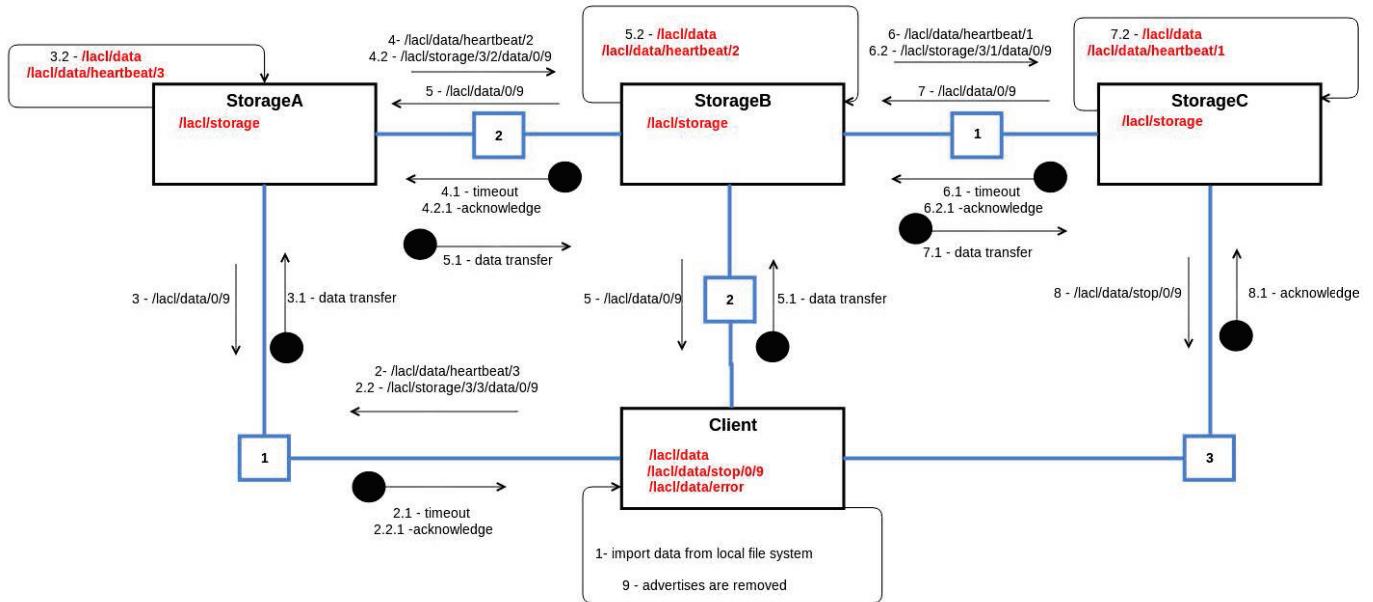


Figure 2. RéPLICATION de données

la réPLICATION des données sur le réseau NDN. Les liens entre les nœuds jouent le rôle de la configuration réseau. Le carré sur la ligne affiche le poids de ce lien dans une stratégie de routage des données. Les requêtes sont représentées par des flèches avec un numéro de séquence. Les réponses sont affichées sous la forme d'une autre flèche dans le sens opposé, avec une puce du côté source. Un numéro hiérarchique montre la relation entre la requête et la réponse.

Des outils en ligne de commande ont été développés afin de permettre le chargement des données depuis le système de fichier local vers le système de fichier distribué. Ces outils sont comparables à ceux utilisés par Hadoop pour le chargement de fichier. Après l'importation des données à partir de son système de fichiers local, le nœud Client envoie un intérêt sur le réseau avec le nom /lacl/data/heartbeat/3. Cet intérêt permet au nœud Client de vérifier si un nœud a déjà exposé ce nom, cela signifierait que la 3^{ème} réPLICATION des données a déjà été effectuée. Après un timeout (flèche 2.1), il sait qu'un tel nœud n'existe pas. Par conséquent, le nœud Client émet l'intérêt suivant /lacl/storage/3/3/data/0/9 où :

- /lacl est le nom de domaine,
- /storage est la demande précise. Il exprime un intérêt de capacité de stockage pour le chargement des données,
- /3/3 est le facteur de réPLICATION, suivi du rang dans le processus de réPLICATION. Ce rang est essentiel pour la stratégie du heartbeat,
- /data/0/9 est le nom des données suivi par l'index du premier et du dernier segment.

La stratégie de routage des messages suit le chemin de coût minimum. Lorsqu'un intérêt est transmis, la demande n'est pas diffusée sur le réseau (broadcast), mais un chemin est sélectionné en fonction du coût le plus bas. Ce choix se fait dynamiquement en s'appuyant sur le protocole de routage NLSR (Named-data Link State Routing protocol) (Hoque *et al.*, 2013).

Au début du scénario, 3 nœuds exposent /lacl/storage. Avec le calcul du chemin de coût le plus faible, StorageA reçoit l'intérêt (flèche 2.2) et un accusé de réception est reçu par le nœud Client (flèche 2.2.1). En conséquence, le nœud StorageA, envoie un intérêt pour les données à enregistrer (/lacl/data/0) pour le premier segment et (/lacl/data/9) pour le dernier segment.

Seul le nœud Client a exposé le bon nom, il retourne les données et une 1^{ère} réPLICATION est effectuée. Ensuite, le nœud StorageA expose un nouveau nom de données en rapport aux données reçues (flèche 3.2) et un autre nom pour l'algorithme de heartbeat (/lacl/data/heartbeat/3). Ce nom sera utilisé pour vérifier que le nœud est toujours actif. S'il est arrêté, un autre nœud de stockage enverra un intérêt de stockage pour trouver un nœud capable de le remplacer. Parallèlement, le nœud StorageA envoie un intérêt /lacl/data/heartbeat/2 sur l'existence d'une seconde réPLICATION des données. Lorsqu'il reçoit un timeout (flèche 4.1), il crée un autre intérêt de stockage avec un facteur de réPLICATION décrémenté qui a la valeur 2 (flèche 4.2). Comme précédemment, un nœud avec la bonne capacité répond (flèche 4.2.1). Le processus de message suit le même modèle que précédemment et les données sont répliquées une 2^{ème} fois sur le nœud StorageB (flèche 5.1).

Étant donné que le facteur de réPLICATION est supérieur à 0, le nœud StorageB envoie un intérêt pour vérifier l'existence de la première réPLICATION (flèche 6). Après un timeout, une dernière requête de réPLICATION est envoyée par StorageB sur le réseau (flèche 6.2). Lorsque le nœud StorageC reçoit les données (flèche 7-1), il expose également de nouveaux noms pour l'accès aux données et le heartbeat et diminue le facteur de réPLICATION. Sa valeur devient 0 et a pour effet l'arrêt du processus de réPLICATION. Le nœud StorageC envoie un intérêt afin de mettre fin au processus (flèche 8) avec le nom /lacl/data/stop/0/9. Seulement le nœud Client peut filtrer cette requête, il supprime tous les noms exposés et met fin à son activité. La surveillance de la vivacité continue et chaque nœud de stockage envoie périodiquement un intérêt pour vérifier une existence de données de réPLICATION. Par exemple, le nœud StorageA envoie l'intérêt /lacl/data/heartbeat/2 (flèche 4) pour contrôler que le nœud StorageB est toujours actif.

Le heartbeat ne s'applique pas au Client, car il représente une tâche de chargement de fichier lancée par l'administrateur du système.

Une implémentation de l'approche en utilisant le simulateur réseau ndnSim (Afanashev *et al.*, 2012) basé sur ns-3 (Lacage, Henderson, 2006) a été réalisée. Lors de cette simulation, nous avons utilisé plusieurs paramètres (nombre de nœud, nombre d'utilisateurs, nombre de répliques, taille des données, ...) en entrée afin de déterminer leurs impacts sur le système. Nous avons obtenu les résultats suivants :

- une augmentation du nombre d'utilisateurs permet de réduire le temps nécessaire pour récupérer une donnée. Cela s'explique par l'utilisation de cache réseau au niveau de NDN.
- à partir d'un réplica disponible sur le réseau, le système est capable de se reconstruire à un état stable si des nœuds de stockage sont disponibles.

Afin de valider notre approche, nous avons considéré l'Internet des Objets (IoT) qui est l'un des domaines prometteurs pour le Big Data vue la quantité de données générée par les larges déploiements de capteurs (Bhatt *et al.*, 2017). Afin de répondre aux contraintes de latences relatives à l'IoT, nous avons opté pour un déploiement dans des machines sur le edge (Satyanarayanan, 2017). Ce contexte nous a permis d'avoir une étude de cas réaliste que nous avons utilisé pour faire une étude comparative entre une approche basée sur IP et celle basée sur NDN (Dongo, Atik *et al.*, 2018).

Cette étude de cas nous a permis de valider notre approche, car nous avons pu montrer que notre approche adresse efficacement les deux problématiques discutés dans la section précédente. En effet, nous avons montré que l'utilisation de notre approche permet une optimisation jusqu'à 80% du trafic réseau dans le cas de collecte distribuée depuis le même capteur connecté. De plus, grâce à l'aspect totalement distribuée de notre approche, nous avons pu montrer qu'un déploiement sur des réseaux isolés est

possible et que le système tolère très bien les défaillances causées par les interruptions de communication entre les composants du système IoT.

En plus des études de cas, nous avons effectué des études statistiques basées sur les paquets échangés sur le réseau. Les résultats ont permis de montrer que l'impact du facteur de réPLICATION sur la demande des données depuis la source est linéaire. Autrement dit, grâce à l'optimisation introduite par l'agrégation des paquets, le nombre de répliques des mêmes données a un impact mineur sur notre système à l'opposé des systèmes existants sur IP sur lesquels le nombre de réPLICATION a un impact majeur. Cela montre l'efficience du système en terme de communication et de coût de déploiement.

2.5. *Actions futures*

L'étude d'une approche de calcul distribué de type MapReduce basée sur notre approche de DFS sur NDN a fait l'objet d'une implémentation sur ndnSim et est en cours de validation. L'objectif est de distribuer le calcul, mais aussi d'éviter d'exécuter un même calcul plusieurs fois si celui-ci a déjà été exécuté par un autre nœud (mise à disposition des résultats des calculs intermédiaires). Nous ferons une analyse des résultats issus de l'implémentation cette approche dans le simulateur ndnSim.

Une implémentation et expérimentation de l'approche sur des machines physiques sont prévues afin de confirmer les résultats obtenus, mais aussi une comparaison de ceux-ci avec ceux de Hadoop.

Vue que les performances sont un facteur crucial dans le choix de solution Big Data, nous analysons le comportement du forwarder NDN afin de proposer des améliorations en termes de transmissions de paquets. Nous expérimontons un potentiel déploiement en tant de forwarder élastique. Nous prévoyons d'intégrer notre forwarder dans l'architecture Big Data NDN afin de surpasser les limitations en performance causées par les implémentations existantes (Dongo, Mahmoudi, Mourlin, 2018).

3. Vers un framework de composition de services sensible au contexte pour les environnements intelligents

Résumé Mots clé en français.

3.1. *Contexte*

La ville intelligente est apparue pour faire face aux défis relevés par l'urbanisation et l'augmentation de la population mondiale. Alors que plusieurs définitions du concept existent et dépendent généralement du point de vue de l'auteur, il est clair qu'une ville intelligente doit offrir des services "intelligents" à ces habitants. En effet, c'est l'utilisation efficiente des services intelligents fournis par les parties prenantes dans les différents secteurs qui est supposée contribuer à l'amélioration de la qualité de vie des citoyens et à l'optimisation des ressources utilisées. Le secteur du transport, à titre d'exemple, est un secteur vital dans n'importe quelle ville. Un dysfonctionnement du réseau de transport a des retombées à la fois sur l'économie, l'environnement et la mobilité. Le projet Toubkal SMARTROAD¹, cadre dans lequel ce

1. Partenariat Hubert Curien (PHC) "TOUBKAL", N° Toubkal/17/45

travail de recherche est effectué, vise à contribuer aux méthodologies et aux outils de la sécurité et la fluidification du trafic routier dans le cadre d'une ville intelligente.

Dans ce contexte, l'Informatique Orientée Services (Service Oriented Computing) a fait ses preuves comme un paradigme permettant l'amélioration de l'agilité des organisations, ainsi que la facilitation de la coopération inter-organisationnelle avec l'utilisation des principes de conception comme la réutilisabilité, la virtualisation et la composition de services. Les organisations (privées et publiques) cherchent de plus en plus de moyens pour capitaliser sur la richesse des données disponibles grâce aux avancées de la technologie pour améliorer leurs services offerts. En même temps, les attentes des utilisateurs par rapport à la pertinence et la qualité des services offerts deviennent de plus en plus élevées. Cela est particulièrement vrai pour les tâches complexes et qui dépendent du contexte de l'utilisateur, telles que la planification de voyages impliquant de nombreuses activités (par exemple, trouver et réserver des hébergements, transports, événements intéressants, etc.) où l'utilisateur doit passer au crible une grande quantité d'informations (offres) et prendre beaucoup de décisions pour réaliser son objectif.

Nos travaux de recherche visent à explorer l'utilisation de la notion de contexte issue de l'informatique ubiquitaire dans la composition de services. Ces services doivent être capables de satisfaire les besoins des utilisateurs, particulièrement dans le contexte des environnements intelligents qui sont riches en information de contexte et où plusieurs services sont offerts. Nous visons donc à exploiter les techniques d'apprentissage automatique, particulièrement ceux liés à la recommandation, ainsi que les informations contextuelles pour recommander des services pertinents aux utilisateurs.

3.2. *État de l'art*

L'intégration des besoins de l'utilisateur et de son contexte dans le processus de création de services est rarement étudié dans la littérature (Sheng *et al.*, 2014). En effet, c'est le fournisseur qui est au centre de l'Architecture Orientée Services puisque les utilisateurs ne peuvent qu'utiliser les services répertoriés par le fournisseur.

Les travaux proposés dans la littérature pour supporter les utilisateurs et les intégrer dans la création de services opèrent généralement aux différentes phases de la composition de services. Parmi ces travaux, certaines approches se basent sur les ontologies comme celle de (Xiao *et al.*, 2011), où les auteurs présentent une approche pour masquer la complexité du modèle de la composition des services et supporter les utilisateurs dans l'exécution de ce processus. Dans cette approche, les ontologies sont utilisées pour encapsuler les connaissances du domaine ainsi que pour déduire des flux de contrôle afin d'améliorer la génération des processus de composition. Une autre famille d'approches se base sur les techniques de l'apprentissage automatique pour déduire des modèles qui ont pour but d'améliorer un ou plusieurs aspects de la composition(Y. Zhang *et al.*, 2017). Alors que certains travaux cherchent à modéliser et intégrer le comportement de l'utilisateur pour prédire ses actions futures et donc être capable de lui offrir des services pertinents (He *et al.*, 2015), d'autres cherchent à offrir les différents services les plus pertinents pour l'utilisateur grâce à l'utilisation des systèmes de recommandation(Zheng *et al.*, 2011).

L'application des systèmes de recommandation pour optimiser la sélection des services a été suffisamment étudiée et donne généralement de bons résultats. Dans ce sens, la majorité des travaux repose sur des techniques se basant sur le filtrage collaboratif (Zheng *et al.*, 2011 ; Jiang *et al.*, 2011). Le principe est de recommander un service s à un utilisateur u , si un autre utilisateur u' qui est similaire à l'utilisateur

u a déjà utilisé le même service *s*. Le cœur de ces approches réside généralement dans les fonctions de similarité utilisées pour calculer le degré de similarité entre les utilisateurs où encore entre les services. Pour les utilisateurs, ces fonctions peuvent utiliser des paramètres comme les profils des utilisateurs, les activités qu'ils réalisent, etc. Pour les services, ces fonctions sont définies dépendamment de la méthode avec laquelle les services sont décrits (texte, fichier de description) ou en ajoutant encore des paramètres liés à qualité de ces services (temps de réponse, disponibilité).

3.3. Problématique

La création de services dans l'Architecture Orientée Services évoque naturellement une relation fournisseur-consommateur (utilisateur). Cependant, l'analyse de la littérature montre que l'on se focalise généralement beaucoup plus sur la vision du fournisseur que sur la vision de l'utilisateur. Les utilisateurs se retrouvent donc avec des services qui ne correspondent pas à leurs besoins. Ce problème est encore plus prononcé dans les environnements ubiquitaires d'aujourd'hui où les équipements numériques se sont immiscés partout dans l'environnement et fournissent des services cruciaux et personnels aux utilisateurs, à l'exemple des environnements intelligents.

Pour avoir des services plus pertinents et utilisables, le rôle de l'utilisateur, ses besoins et ses préférences doivent être aussi définis et intégrés dans le processus de composition de services. À cet égard, l'importance de l'information contextuelle a été reconnue par les chercheurs et les praticiens de nombreuses disciplines. L'incorporation du contexte dans les systèmes de recommandation permet d'améliorer la qualité des recommandations (Adomavicius, Tuzhilin, 2011). Cependant, les applications de ces systèmes nommés "Systèmes de Recommandation sensibles au contexte" restent restreintes sur des items de nature simple comme des films, de la musique ou des lieux à visiter et leur application sur les services reste peu explorée (Cao *et al.*, 2013). La tâche devient encore plus difficile quand on considère des services de nature composite.

Dans ce contexte, nous visons à répondre aux questions suivantes :

- Comment peut-on intégrer l'utilisateur, à travers son contexte, dans le processus de composition de services dans les environnements intelligents ?
- Quel serait l'impact de l'intégration des informations de contexte dans les algorithmes d'apprentissage automatique pour la recommandation de services sur la qualité des services recommandés ?

3.4. Actions réalisées

Nos recherches jusqu'alors se sont focalisées sur la réponse à la première question et à l'apport du contexte dans le processus de création de services dans les environnements ubiquitaires ou intelligents supportés par l'informatique en nuage (Cloud Computing), tout en examinant le rôle des utilisateurs dans ce processus.

Dans un premier temps, nous nous sommes concentrés sur l'importance du contexte dans le processus de composition de services. Nous avons proposé dans (Faieq *et al.*, 2017b), une architecture pour la composition sensible au contexte de services. Nous défendons l'importance du contexte pour (i) l'utilisateur, dans la minimisation des interactions avec les services ou le système de composition de services (ii) le système de composition pour améliorer les performances de ses systèmes ainsi que (iii) la satisfaction de

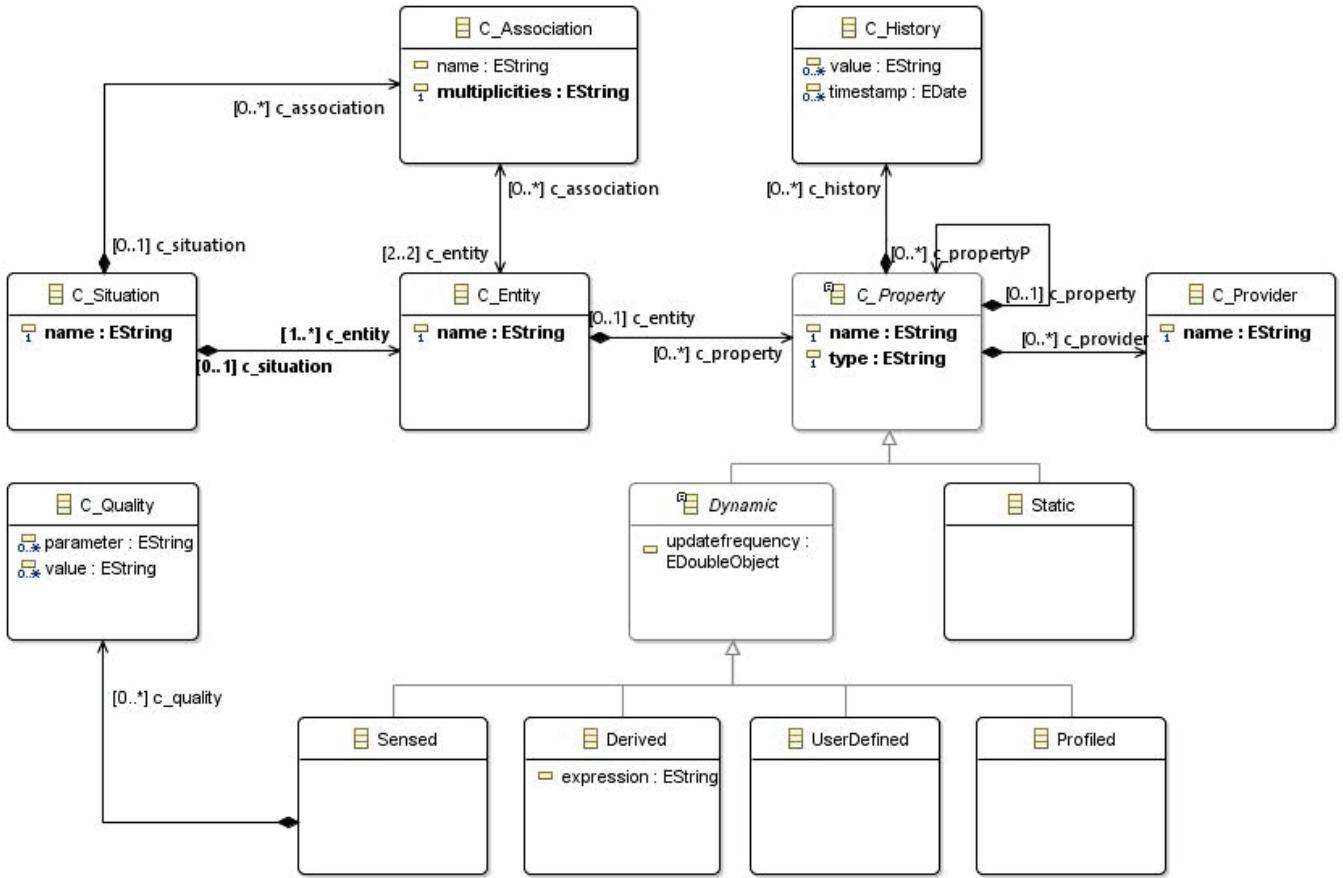


Figure 3. Méta-modèle du contexte conforme aux spécifications du MOF

l'utilisateur du service composite résultant. Cette architecture est appuyée par un méta-modèle (cf. Figure 3) du contexte pour représenter la situation des entités participant dans la composition (utilisateurs et services).

Dans un second temps, nous avons étendu cette architecture en étudiant les synergies existantes entre les différentes technologies que nous avons jugées principales dans le développement des services dans les environnements intelligents à savoir, Big data, Internet des objets, Informatique en nuage et sensibilité au contexte. Ces environnements présentent un nouveau paradigme où les services sont fournis non seulement par des fournisseurs de services mais aussi par des équipements et appareils contenant des capteurs et des actionneurs (comme les smartphones, smartWatch, smartTV, etc.). Nous avons présenté dans (Faieq *et al.*, 2017a) un framework centré sur la notion du contexte pour le développement des services dans les environnements intelligents, dans lequel nous expliquons comment ces technologies peuvent collaborer pour créer des services qui sont puissants, pervasifs et pertinents au contexte de l'utilisateur.

En se basant sur les constats tirés des travaux précédents et pour répondre à la deuxième question de la problématique, nous proposons une approche pour la mise en place d'un système de recommandation sensible au contexte pour les services composites. L'objectif est d'améliorer la qualité des services recommandés dans les différentes situations en analysant les traces d'exécution des applications à base de services. L'approche générale est présentée dans la Figure 4.

La première phase consiste à analyser les traces d'exécution des applications à base de services. Cette analyse nous permet de construire une matrice d'évaluation utilisateur/service où les lignes représentent

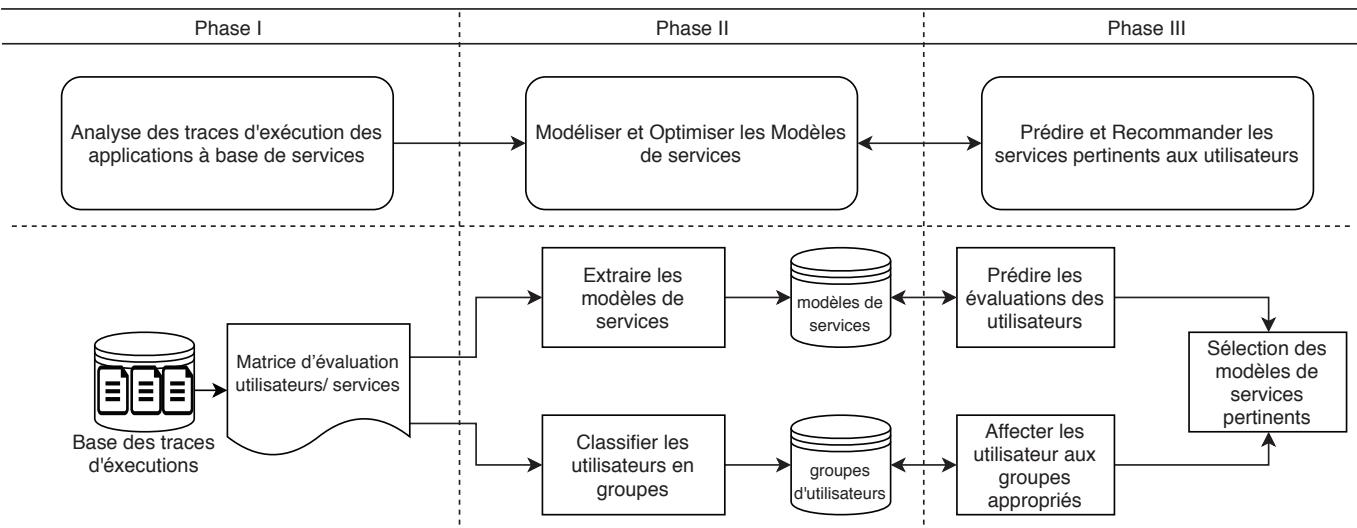


Figure 4. Approche générale pour une recommandation de services sensible au contexte

les utilisateurs, les colonnes représentent les services et les cellules sont les valeurs de qualité de service pour chaque service invoqué par l'utilisateur. Nous notons qu'un même utilisateur peut avoir plusieurs profils dépendamment de son contexte au moment de l'invocation. Dans la deuxième phase, nous utilisons la matrice d'évaluation pour (i) dégager les meilleurs modèles de services pour chaque contexte possible pour les utilisateurs du système (ii) extraire les groupes des utilisateurs similaires. La troisième phase finit par la sélection du n meilleurs modèles de services pour un nouvel utilisateur, qui sont les plus pertinents par rapport à son contexte. Elle repose sur l'utilisation des modèles et des groupes extraits pour premièrement l'affecter à un groupe d'utilisateurs approprié et puis de prédire la qualité des modèles de services extraits s'ils sont invoqués par cet utilisateur.

Nous avons proposé un cas d'étude relevant du cadre du projet SMARTROAD. Dans ce cas d'étude, un certain nombre de services qui sont fournis par les différentes entités impliquées dans un écosystème de transport routier intelligent (e.g. les conducteurs, les véhicules, les infrastructures et l'environnement) et pourront servir à améliorer la sécurité et le trafic routier. Un ensemble non exhaustif de ces services est présenté dans la Tableau 1.

Acronyme	Nom du service	Entité
AEVW	Approaching Emergency Vehicle Warning	Infrastructure
DHMS	Driver Health Monitoring Service	Driver or Vehicle
ISA	Intelligent Speed Adaptation	Vehicle
IVS	InVehicle Signage	Vehicle
ERP	Electronic Road Panel Service	Infrastructure
WWS	Weather Warning Service	Environment
RSS	Route Selection Service	Vehicle
RWW	Road Works Warning	Infrastructure
ERS	Emergency Rescue Service	Infrastructure
NAAS	Nearby Area Alarming Service	Infrastructure

Tableau 1. Exemples de services fournis

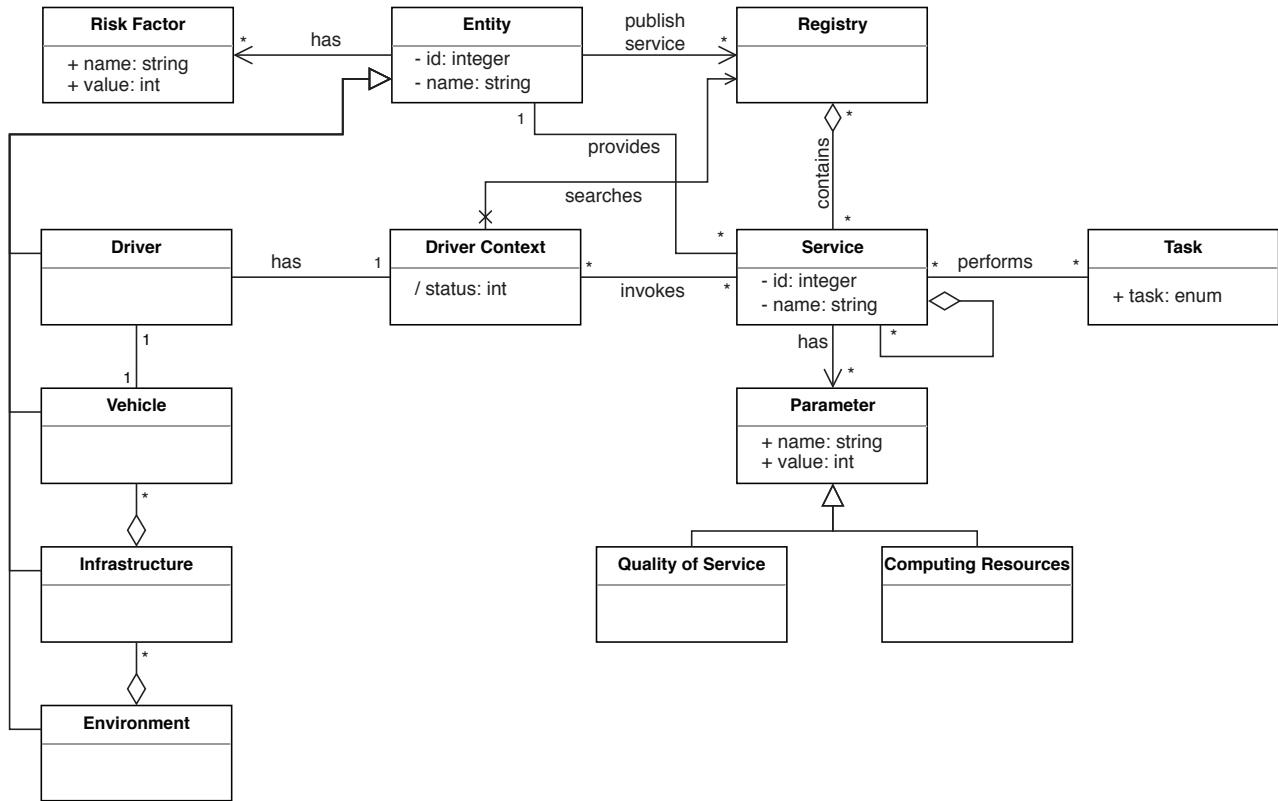


Figure 5. modèle du système de surveillance du trafic routier

Nous nous sommes particulièrement intéressés à la sécurité routière en définissant les situations qui peuvent être considérées comme anormales et les facteurs de risque qui peuvent contribuer à l'occurrence de ces situations anormales. Pour simplifier, nous considérons que chaque facteur de risque peut causer une situation dangereuse. Ces facteurs de risque sont liés aux différentes entités qui participent dans l'écosystème du transport routier (i.e. véhicule, conducteur, infrastructure et environnement) et peuvent généralement être surveillés à travers des capteurs. La vitesse, la drogue et la météo son des exemples de ces facteurs de risque. Nous avons ensuite lié chaque situation à une orchestration de services (composition) pour informer les entités qui peuvent être influencées par la dite situation. Le but est de sensibiliser ces entités aux dangers et de permettre aux autorités d'exécuter les démarches nécessaires pour revenir à un état normal de circulation. Le contexte est l'agrégation des statuts (valeurs) de chaque facteur de risque dans un instant t . L'invocation des services appropriée se fait une fois une situation anormale est détectée et les informations concernant cette invocation est stocké. La Figure 5 présente le modèle du prototype de notre cas d'étude.

Un exemple qui illustre le déroulement de l'exécution du système de recommandation proposé est présenté dans la Figure 6. Nous considérons dans cet exemple deux situations contextuelles. La première est liée à l'environnement et est causée par un mauvais temps (BadWeather). La deuxième est liée au conducteur et elle est causée par son état de santé, par exemple si le conducteur est sous influence de drogues ou d'alcool (RiskyDriverHealth). Dans la première phase, le système de recommandation commence par l'analyse des traces d'exécution du prototype SMARTROAD. Cette analyse résulte à la construction de la matrice d'évaluation utilisateur/service contenant la qualité de chaque service invoqué par l'utilisateur et son contexte. Cette matrice est utilisée dans la deuxième phase pour dégager (i) les différents groupes de conducteurs (drivers) à partir de leurs contextes (e.g. des utilisateurs qui sont fatigués et qui conduisent dans de bonnes conditions météorologiques) (ii) les modèles de services (pour simplifier, on parlera de

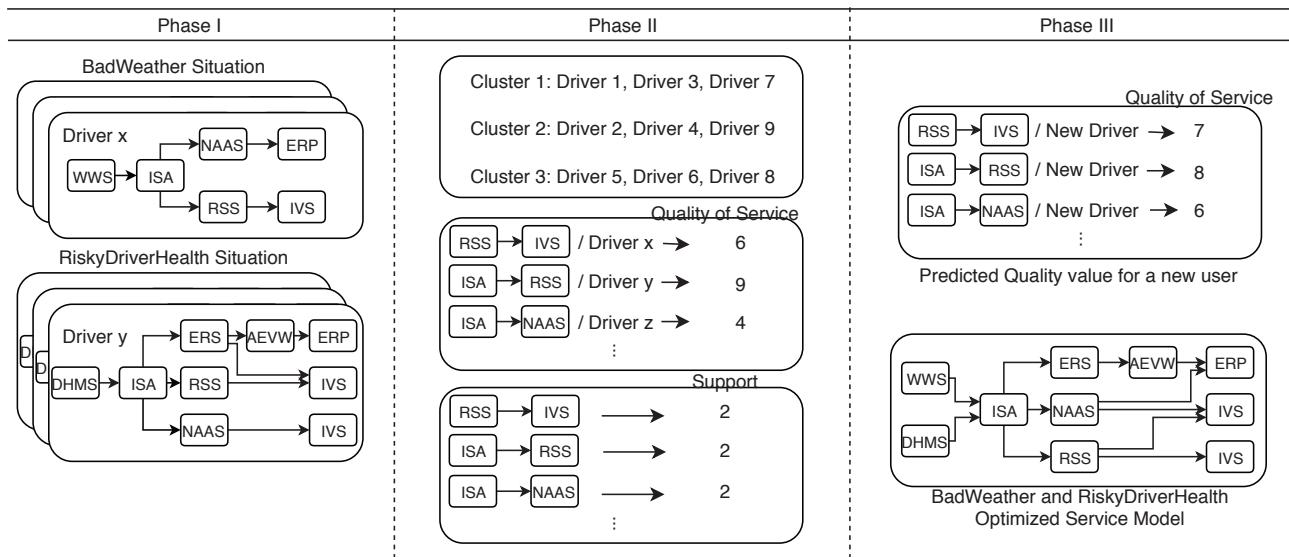


Figure 6. Exemple d'exécution du processus de recommandation

tâches) les plus utilisés (i.e. les tâches qui sont souvent effectuées seules ou ensemble) (iii) la qualité des instances de services invoqués dans chaque situation détectée. La troisième phase est déclenchée lorsqu'un nouvel utilisateur s'introduit au système. Dans cette phase, une optimisation des modèles de services (orchestrations) est générée, s'il le faut, pour éviter l'invocation multiple des mêmes services. Ceci est fait en utilisant les modèles extraits dans la deuxième phase. En même temps, des prédictions sur la qualité des services répertoriés sont effectuées pour le nouvel utilisateur. Les résultats des deux opérations précédentes sont agrégées pour permettre de lier les meilleurs services en termes de qualité au modèle de service optimisé pour être recommandé à l'utilisateur.

3.5. Actions futures

Dans les environnements intelligents, les services sont souvent exécutés dans des environnements d'exécutions avec des ressources de traitement variantes, ce qui affecte leurs paramètres de qualité de service. Ceci pourrait être aggravé par l'utilisation d'un système de recommandation puisque le système pourrait recommander le même service à plusieurs utilisateurs, générant plus de requêtes et donc plus de charges de traitement. Ce pourquoi, il est important de prendre en compte ces paramètres dans la phase d'apprentissage pour améliorer la qualité des prédictions, en étudiant la corrélation entre le contexte de l'utilisateur, les ressources de traitement et la qualité de service; afin d'étudier l'impact du système de recommandation sur la performance des services, en utilisant les simulateurs d'infrastructure cloud (e.g. cloudsim).

4. Modélisation des (méta)données hétérogènes et filtrage des contenus de vidéosurveillance : application au Forensic

4.1. Introduction

Au cours des dernières années, le taux de criminalité a augmenté très rapidement dans le monde. Le nombre de caméras de vidéosurveillance censées assurer la sécurité des citoyens dans les grandes villes croît régulièrement et de manière significative. Ces caméras génèrent une masse importante de don-

nées vidéos (big data) à analyser afin d’élucider les différents crimes (terrorisme, homicide, enlèvement, ...) survenus. Le processus actuellement utilisé par les enquêteurs (en France), consiste à appliquer des traitements vidéo (détection de visages, détection de véhicules, détection et lecture des plaques d’immatriculation...) sur l’ensemble des séquences vidéos récupérées en amont et préalablement indexées (conversion au bon format, géolocalisation des caméras, gestion des horodatages...). Les traitements sont appliqués sur l’ensemble des vidéos ou uniquement sur un sous-ensemble géographique ou temporel. Après ces traitements et une extraction des données, les enquêteurs effectuent des recherches (véhicules, plaques, personnes, visages...), visionnent les séquences d’intérêt et lancent de nouveaux traitements. Les affaires récentes (terrorisme, enlèvement, homicide) ont nécessité l’analyse de plusieurs dizaines de milliers d’heures de vidéo. Le gain de temps fourni par les outils d’analyse actuels reste insuffisant dans un contexte opérationnel. Beaucoup des vidéos analysées pour une enquête s’avèrent inexploitables (par ex., problème de luminosité) pour l’enquête. Écarter ces séquences inéligibles permettra d’optimiser le temps d’exploitation. Cette recherche a pour objectif de proposer une (méta)modélisation et les mécanismes afférents de filtrage de grandes collections de vidéo liées à la recherche de preuves à posteriori. Cette contribution basée sur les (méta)données passe par la révision de la norme ISO 22311/IEC 79 qui décrit les exigences fonctionnelles à saisir par les systèmes de vidéosurveillance.

4.2. *État de l’art*

L’analyse vidéo automatique est actuellement confrontée à trois défis majeurs : (i) problème de gestion du grand nombre de caméras largement distribuées, (ii) difficulté de stockage et d’organisation de l’énorme quantité de données générées par ces caméras, (iii) analyse, annotation et recherche demeurent très coûteuses en temps.

De nombreux travaux ont été menés dans le but de proposer des solutions aux problèmes d’analyse vidéo ; notre objectif n’est pas d’en faire un exposé exhaustif ni d’apporter une contribution à l’analyse de contenu proprement dite : le verrou adressé est celui qui consiste à enrichir la modélisation des (méta)données, actuellement pas ou peu exploitées dans ce domaine.

Une méthode efficace et efficiente pour la construction et l’extraction des métadonnées descriptives des vidéos Web est présentée dans (Algur *et al.*, 2014) qui propose un modèle de métadonnées descriptives pouvant faciliter le traitement des contenus vidéos. Ce modèle est orienté "classification des vidéos" et nécessite d’autres spécifications pour être utilisé dans notre contexte. Un framework utilisant des informations géographiques pour la récupération des vidéos sur le web est présenté dans (Han *et al.*, 2016). Les auteurs décrivent et utilisent des données géographiques liées à la vidéo, telles que la localisation vidéo, le champ de vue de la caméra et les trajectoires pour effectuer des recherches dans les vidéos. Cependant, la qualité d’images/vidéos n’est pas prise en compte.

Les récents travaux de notre équipe se sont focalisés sur la modélisation des métadonnées pour la gestion distribuée des documents multimédias et l’interopérabilité des systèmes de vidéo surveillance. (Manzat, 2013) propose un format de métadonnées associées à tout type de contenu multimédia et permettant à plusieurs systèmes d’échanger des données et de les utiliser sans effectuer de manipulations spéciales. Ce format de métadonnées a été validé dans le projet LINDO². Une étude des métadonnées utiles pour la recherche dans les collections de vidéos est présentée dans (Codreanu, 2015). Cette étude

2. <http://lindo-itea.eu/>

a été appliquée dans le cadre du projet ANR METHODEO³. L'approche utilisée dans (Codreanu *et al.*, 2015) vise à aider les opérateurs humains de vidéosurveillance dans la recherche de séquences vidéo d'intérêt en leur proposant un ensemble de caméras susceptibles d'avoir filmé une scène recherchée. Une extension de cette approche est décrite dans (Panta, Sèdes, 2016).

Globalement, les solutions proposées dans la littérature utilisent des approches d'interrogation de contenus basées sur des métadonnées (techniques, liées aux caméras, spatio-temporelles) mais ne proposent pas d'interrogation basée sur le contenu via les descripteurs de qualité images/vidéos ou d'autres caractéristiques liées aux contenus.

4.3. Problématique

Un rapide état de l'art montre donc la nécessité d'exploiter les informations concises des contenus vidéo à partir de tous les éléments disponibles, *a priori* ou *a posteriori* en fonction des (méta)données potentiellement élicitables. Ces informations constituent des métadonnées issues d'algorithmes d'analyse de contenu (caractéristiques de bas niveau, présence et/ou nombre de personnes dans la scène, ...) et peuvent être modélisées selon différents niveaux de sémantique et de granularité. La collaboration des différents niveaux de métadonnées est un verrou essentiel dans notre approche. Cette approche consiste à effectuer un filtrage de contenus vidéo avec le meilleur compromis qualité/précision/temps de réponse, en combinant les métadonnées techniques, les métadonnées décrivant le mouvement et le champ de vue de la caméra, les métadonnées issues d'algorithmes d'analyse de contenu, et les métadonnées de contexte (réseaux sociaux, open data, géolocalisation indoor/outdoor, jour/nuit). Une telle approche s'apparente à un problème de modélisation multicouches et est caractérisée par l'hétérogénéité des données dont elle est issue.

Nous souhaitons lever les verrous relatifs à (i) la proposition de "descripteurs" de qualité, (ii) l'interopérabilité des métadonnées associées aux contenus vidéo, (iii) la recherche dans de grands volumes de données vidéo sans recourir aux contenus intégraux, ces problématiques visant à mettre en oeuvre le filtrage de contenus pertinents pour les enquêteurs.

4.4. Actions réalisées

Notre travail s'applique à la recherche des preuves *a posteriori* (forensic) dans de grandes collections de vidéos et se déroule dans le cadre du projet ANR FILTER2⁴. L'objectif est d'aider les enquêteurs dans leur tâche en leur proposant des outils permettant d'identifier rapidement et automatiquement, parmi la masse des fichiers de vidéoprotection, les données les plus exploitables et pertinentes afin d'en optimiser le traitement. Afin d'atteindre cet objectif et lever les verrous scientifiques mentionnés dans la problématique, nous proposons une approche qui consiste en l'intégration sémantique des données/métadonnées hétérogènes liées à la vidéosurveillance. L'intégration sémantique de données se défini ici comme étant le processus d'utilisation d'une représentation conceptuelle de données et de leurs relations pour résoudre le problème d'hétérogénéité.

L'approche proposée peut se résumer en quatre étapes comme présenté à la figure 7.

3. <http://www.agence-nationale-recherche.fr/Projet-ANR-10-SECU-0006>

4. FILtrage négaTif des contEnus de vidéopRotectiOn

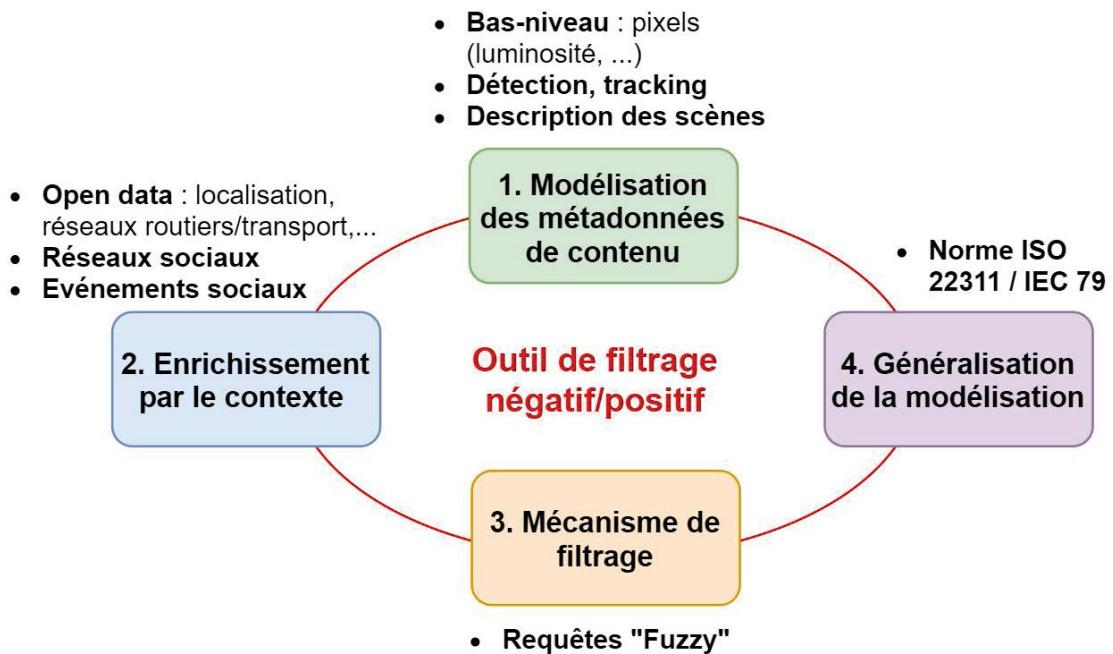


Figure 7. Vue d'ensemble de notre approche.

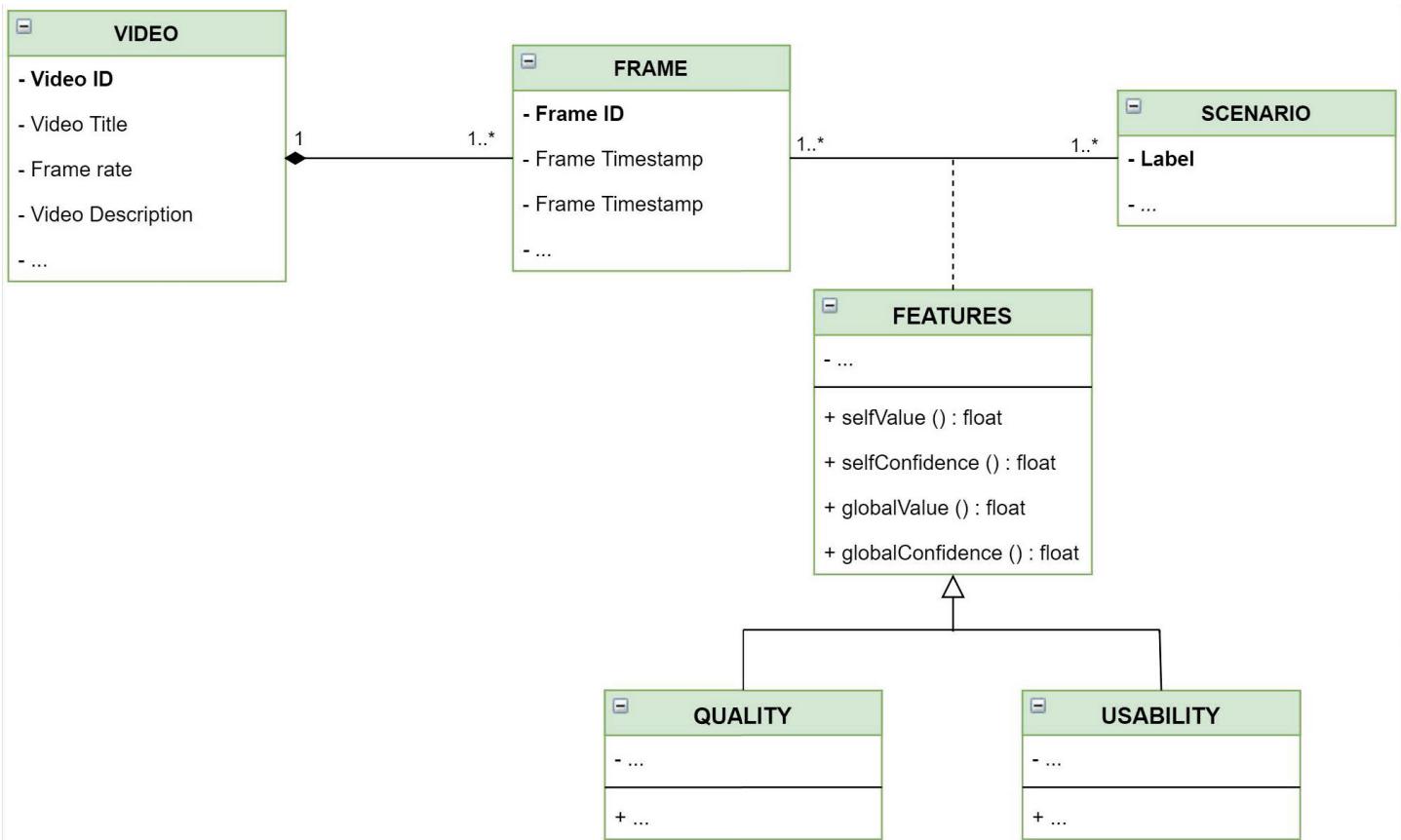


Figure 8. Modèle de données prenant en compte les critères de qualité et d'utilisabilité.

Étape 1 : Modélisation des métadonnées de contenu. Ici, on s'intéresse aux métadonnées issues des algorithmes d'analyse des contenus vidéo tels que la détection et le tracking des objets dans les vidéos, la description des scènes, et les caractéristiques de bas niveau liées aux pixels (intensité lumineuse, contraste, ...). L'objectif est de définir des critères de qualité vidéo et d'utilisabilité qui seront utilisés

dans le processus de filtrage, en prenant en compte les différents traitements (scénarios) et les modes d'analyse (urgent et investigation) auxquels sont confrontés les enquêteurs. Trois traitements sont retenus dans le cadre de FILTER2 : la détection de visages, la détection et la lecture automatique des plaques, et la détection des véhicules. Nous proposons une version initiale du modèle de données (Figure 8) dans lequel, les opérations (*selfValue()*, *selfConfidence()*, *globalValue()*, *globalConfidence()*) de la classe "FEATURES" permettent de calculer les valeurs et niveaux de confiance des caractéristiques de qualité et d'utilisabilité/d'utilité de chaque "FRAME" en fonction de chaque "SCENARIO".

Ce modèle de données sera utilisé dans le mécanisme de requêtage.

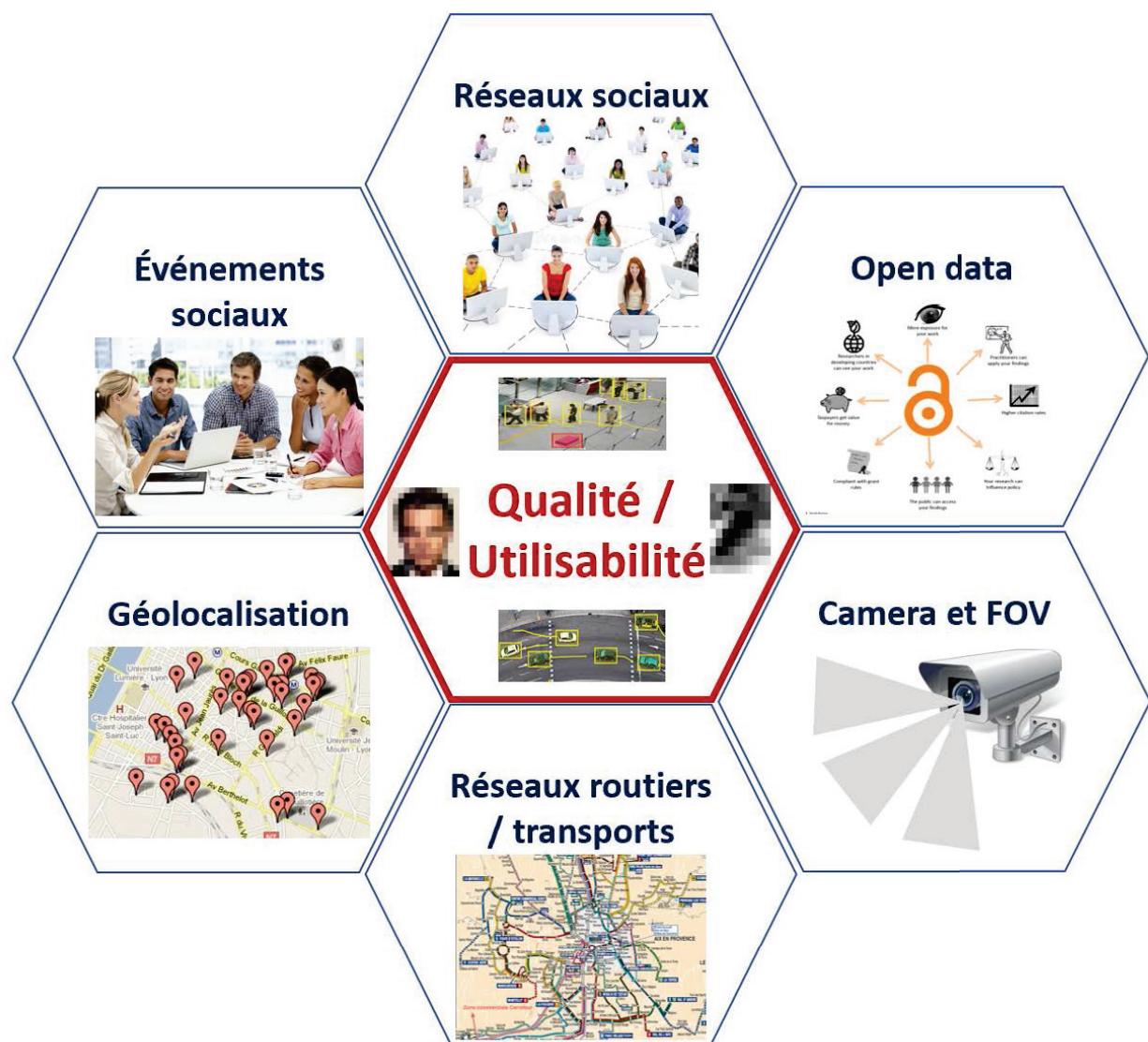


Figure 9. Enrichissement par le contexte.

Étape 2 : Enrichissement par le contexte. L'enrichissement vise à améliorer le filtrage basé sur les métriques de qualité et d'utilisabilité vidéo en intégrant des métadonnées de contexte telles que les réseaux sociaux (liens entre utilisateurs, groupe d'utilisateurs, ...), les évènements sociaux, l'open data, les réseaux routiers, les réseaux de transports en commun, la localisation indoor/outdoor, la mobilité, et les métadonnées décrivant le mouvement et les champs de vue des caméras(Figure 9). L'intégration de toutes ces métadonnées constitue une modélisation multicouche qui permettra d'effectuer un filtrage à différents niveaux.

Étape 3 : Mise en oeuvre d'un mécanisme de requêtage pour le filtrage. Cette étape consiste à filtrer des grands volumes de données vidéo à partir des métadonnées modélisées aux étapes précédentes. Le résultat du filtrage négatif/positif pour une vidéo donnée est obtenu en comparant les caractéristiques de chaque frame de la vidéo aux différents seuils de compatibilité de chaque mode d'analyse.

Algorithme 1 : Filtering algorithm based on quality and usability metadata for urgent mode

Input : Video V , Scenario Sc , compatibility threshold P
Output : List of compatibles video segments for urgent mode analysis

```
1 foreach framei of video V do
2   if getFeature(Sc, framei) ≥ P then
3     add(temporalSegment, framei);
4   else if temporalSegment is not empty then
5     add(listOfCompatibleSegment, temporalSegment);
6     empty temporalSegment;
7   else
8     empty temporalSegment;
9   end if
10 end foreach
11 if temporalSegment is not empty then
12   add(listOfCompatibleSegment, temporalSegment);
13   empty temporalSegment;
14 end if
15 return listOfCompatibleSegment;
```

Il est important de définir des plages de valeurs optimales pour les seuils de compatibilité. Nous proposons actuellement des algorithmes de filtrage basés sur des métadonnées de qualité et d'utilisabilité avec des seuils de compatibilité paramétrables. Le résultat pour le mode urgent est une liste de segments vidéo compatibles (Algorithme 1). Pour le mode investigation(Algorithme 2), le résultat est constitué d'une liste de segments vidéo compatibles et moyennement compatibles : l'enquêteur décidera s'il est nécessaire de traiter les résultats moyennement compatibles.

Étape 4 : Généralisation de la modélisation. L'objectif ici est d'enrichir et d'étendre la norme ISO 22311/IEC 79 afin de tendre vers la fourniture d'une boîte à outils standards pour le filtrage « amont » et l'interrogation intelligente et adaptable des corpus de vidéoprotection.

4.5. Conclusions et perspectives

Les travaux réalisés concernent la modélisation et le filtrage basé sur les métadonnées de qualité et d'utilisabilité. L'enrichissement par le contexte est la prochaine étape de notre démarche. Nous envisageons de proposer une modélisation et l'intégration des métadonnées de contexte afin d'effectuer un filtrage à différents niveaux. Les algorithmes de requêtage seront définis et mis à jour progressivement, ainsi que l'extension de la norme ISO 22311/IEC79.

Algorithme 2 : Filtering algorithm based on quality and usability metadata for investigation mode

Input : Video V , Scenario Sc , compatibility thresholds P_1, P_2

Output : List of compatible and average compatible video segments for urgent mode analysis

```
1 foreach framei of video V do
2   if getFeature(Sc, framei) ≥ P1 then
3     add(compatibleSegment, framei);
4   else if getFeature(Sc, framei) ≥ P2 then
5     add(avCompatibleSegment, framei);
6   else
7     if compatibleSegment is not empty then
8       add(listOfCompatibleSegment, compatibleSegment);
9       empty compatibleSegment;
10    end if
11  else
12    empty compatibleSegment;
13  end if
14  if avCompatibleSegment is not empty then
15    add(listOfAvCompatibleSegment, avCompatibleSegment);
16    empty avCompatibleSegment;
17  end if
18  else
19    empty avCompatibleSegment;
20  end if
21 end if
22 end foreach
23 if compatibleSegment is not empty then
24   add(listOfCompatibleSegment, compatibleSegment);
25   empty compatibleSegment;
26 end if
27 if avCompatibleSegment is not empty then
28   add(listOfAvCompatibleSegment, avCompatibleSegment);
29   empty avCompatibleSegment;
30 end if
31 return listOfCompatibleSegment;
32 return listOfAvCompatibleSegment;
```

Bibliographie

- Adomavicius G., Tuzhilin A. (2011). Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), *Recommender systems handbook*, p. 217–253. Boston, MA, Springer US.
- Afanasyev A., Moiseenko I., Zhang L. et al. (2012). ndnsim : Ndn simulator for ns-3. *University of California, Los Angeles, Tech. Rep.*, vol. 4.
- Algur S. P., Bhat P., Jain S. (2014). Metadata construction model for web videos : A domain specific approach. *International Journal of Engineering and Computer Science*, vol. 3, n° 12.

- Bhatt C., Dey N., Ashour A. S. (2017). Internet of things and big data technologies for next generation healthcare.
- Cao B., Liu J., Tang M., Zheng Z., Wang G. (2013). Mashup service recommendation based on user interest and social network. In *Proceedings of the 2013 ieee 20th international conference on web services*, p. 99–106. Washington, DC, USA, IEEE Computer Society.
- Chen S., Cao J., Zhu L. (2015, Sept). Ndss : A named data storage system. In *2015 international conference on cloud and autonomic computing*, p. 196-199.
- Codreanu D. (2015). *Modélisation des métadonnées spatio-temporelles associées aux contenus vidéos et interrogation de ces métadonnées à partir des trajectoires hybrides : application dans le contexte de la vidéosurveillance*. Thèse de doctorat, Université de Toulouse, Université Toulouse III-Paul Sabatier.
- Codreanu D., Peninou A., Sedes F. (2015). Video spatio-temporal filtering based on cameras and target objects trajectories—videosurveillance forensic framework. In *Availability, reliability and security, 2015 10th international conference on*, p. 611–617.
- Dongo J., Atik Y., Mahmoudi C., Mourlin F. (2018, June). Distributed file system for ndn : an iot application. In *2018 international conference on selected topics in mobile and wireless networking (mownet)*, p. 138-141.
- Dongo J., Mahmoudi C., Mourlin F. (2018). Ndn log analysis using big data techniques : Nfd performance assessment. In *2018 ieee fourth international conference on big data computing service and applications (bigdataservice)*.
- Faieq S., Saidi R., Elghazi H., Rahmani M. D. (2017a). C2iot : A framework for cloud-based context-aware internet of things services for smart cities. *Procedia Computer Science*, vol. 110, p. 151 - 158.
- Faieq S., Saidi R., Elghazi H., Rahmani M. D. (2017b). A conceptual architecture for a cloud-based context-aware service composition. In *Advances in ubiquitous networking 2*, p. 235–246. Singapore, Springer Singapore.
- Gibbens M., Gniady C., Ye L., Zhang B. (2017, juin). Hadoop on named data networking : Experience and results. *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 1, n° 1, p. 2 :1–2 :21.
- Han Z., Cui C., Kong Y., Qin F., Fu P. (2016). Video data model and retrieval service framework using geographic information. *Transactions in GIS*, vol. 20, n° 5, p. 701–717.
- He W., Ren G., Cui L., Li H. (2015). User behavioral context-aware service recommendation for personalized mashups in pervasive environments. In *Web technologies and applications*, p. 683–694. Cham, Springer International Publishing.
- Hoque A., Amin S. O., Alyyan A., Zhang B., Zhang L., Wang L. (2013). Nlsr : named-data link state routing protocol. In *Proceedings of the 3rd acm sigcomm workshop on information-centric networking*, p. 15–20.
- Jiang Y., Liu J., Tang M., Liu X. F. (2011). An effective web service recommendation method based on personalized collaborative filtering. In *Proceedings of the 2011 ieee international conference on web services*, p. 211–218. Washington, DC, USA, IEEE Computer Society.
- Lacage M., Henderson T. R. (2006). Yet another network simulator. In *Proceeding from the 2006 workshop on ns-2 : the ip network simulator*, p. 12.
- Manzat A.-M. (2013). *Contribution à la modélisation des métadonnées associées aux documents multimédias et à leur enrichissement par l'usage*. Thèse de doctorat, Université de Toulouse, Université Toulouse III-Paul Sabatier.
- Panta F. J., Sèdes F. (2016). Mobile objects in indoor environment : Trajectories reconstruction. In *Proceedings of the 14th international conference on advances in mobile computing and multi media*, p. 332–336.
- Satyanarayanan M. (2017). The emergence of edge computing. *Computer*, vol. 50, n° 1, p. 30–39.
- Sheng Q. Z., Qiao X., Vasilakos A. V., Szabo C., Bourne S., Xu X. (2014). Web services composition : A decade's overview. *Information Sciences*, vol. 280, p. 218 - 238.
- Shvachko K., Kuang H., Radia S., Chansler R. (2010). The hadoop distributed file system. In *Mass storage systems and technologies (msst), 2010 ieee 26th symposium on*, p. 1–10.
- Xiao H., Zou Y., Tang R., Ng J., Nigul L. (2011, 24 Mar). Ontology-driven service composition for end-users. *Service Oriented Computing and Applications*, vol. 5, n° 3, p. 159.
- Zhang L., Estrin D., Burke J., Jacobson V., Thornton J. D., Smetters D. K. et al. (2010). Named data networking (ndn) project. *Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC*.

- Zhang Y., Zhang M., Zheng X., Perry D. E. (2017, June). Service2vec : A vector representation for web services. In *2017 ieee international conference on web services (icws)*, p. 890-893. IEEE Computer Society.
- Zhao J., Pjesivac-Grbovic J. (2009). *Mapreduce : The programming model and practice*. (Tutorial)
- Zheng Z., Ma H., Lyu M. R., King I. (2011, April). Qos-aware web service recommendation by collaborative filtering. *IEEE Transactions on Services Computing*, vol. 4, n° 2, p. 140-152.