



**DOCTORAT EN CO-ACCREDITATION
TELECOM SUDPARIS ET L'UNIVERSITE EVRY VAL D'ESSONNE**

Spécialité : Informatique

Ecole doctorale : Sciences et Ingénierie

Présentée Par : Dorsaf ZEKRI

**Pour obtenir le grade de
DOCTEUR DE TELECOM SUDPARIS**

Agrégation et extraction des connaissances dans les réseaux inter-véhicules

Soutenue le 17 janvier 2013 devant le jury composé de :

Rapporteurs :

Mme Karine Zeitouni Professeur à l'Université de Versailles Saint-Quentin
M. Jérôme Gensel Professeur à l'Université Pierre Mendès France

Examineurs :

Mme Anne Doucet Professeur au laboratoire d'informatique de Paris 6
M. Talel Abdessalem Professeur à Télécom ParisTech

Directeur de thèse :

M. Bruno Defude Professeur à Télécom SudParis

Co-Directeur de thèse :

M. Thierry Delot Maître de Conférences (HDR) à l'université de Valenciennes

Résumé : Les travaux réalisés dans cette thèse traitent de la gestion des données dans les réseaux inter-véhiculaires (VANETs). Ces derniers sont constitués d'un ensemble d'objets mobiles qui communiquent entre eux à l'aide de réseaux sans fil de type IEEE 802.11, Bluetooth, ou Ultra Wide Band (UWB). Avec de tels mécanismes de communication, un véhicule peut recevoir des informations de ses voisins proches ou d'autres plus distants, grâce aux techniques de multi-sauts qui exploitent dans ce cas des objets intermédiaires comme relais. De nombreuses informations peuvent être échangées dans le contexte des « VANETs », notamment pour alerter les conducteurs lorsqu'un événement survient (accident, freinage d'urgence, véhicule quittant une place de stationnement et souhaitant en informer les autres, etc.). Au fur et à mesure de leurs déplacements, les véhicules sont ensuite « contaminés » par les informations transmises par d'autres. Dans ce travail, nous voulons exploiter les données de manière sensiblement différente par rapport aux travaux existants. Ces derniers visent en effet à utiliser les données échangées pour produire des alertes aux conducteurs. Une fois ces données utilisées, elles deviennent obsolètes et sont détruites. Dans ce travail, nous cherchons à générer dynamiquement à partir des données collectées par les véhicules au cours de leur trajet, un résumé (ou agrégat) qui fournit des informations aux conducteurs, y compris lorsqu'aucun véhicule communicant ne se trouve pas à proximité. Pour ce faire, nous proposons tout d'abord une structure d'agrégation spatio-temporelle permettant à un véhicule de résumer l'ensemble des événements observés. Ensuite, nous définissons un protocole d'échange des résumés entre véhicules sans l'intermédiaire d'une infrastructure, permettant à un véhicule d'améliorer sa base de connaissances locale par échange avec ses voisins. Enfin, nous définissons nos stratégies d'exploitation de résumé afin d'aider le conducteur dans la prise de décision. Nous avons validé l'ensemble de nos propositions en utilisant le simulateur « VESPA » en l'étendant pour prendre en compte la notion de résumés. Les résultats de simulation montrent que notre approche permet effectivement d'aider les conducteurs à prendre de bonnes décisions, sans avoir besoin de recourir à une infrastructure centralisatrice.

Mots clés : Réseaux inter-véhicule - Agrégation d'événements - Construction de résumé - Protocole d'échange - Extraction de connaissances.

Abstract : The works in this thesis focus on data management in inter-vehicular networks (VANETs). These networks consist of a set of moving objects that communicate with wireless networks IEEE 802.11, Bluetooth, or Ultra Wide Band (UWB). With such communication mechanisms, a vehicle may receive information from its close neighbors or other more remote, thanks to multi-jump techniques that operate in this case intermediate objects as relays. A lot of information can be exchanged in the context of « VANETs », especially to alert drivers when an event occurs (accident, emergency braking, vehicle leaving a parking place and want to inform others, etc.). In their move vehicles are then « contaminated » by the information provided by others. In this work, we use the data substantially different from the existing work. These are, in fact, use the data exchanged to produce alerts drivers. Once these data are used, they become obsolete and are destroyed. In this work, we seek to generate dynamically from data collected by vehicles in their path, a summary (or aggregate) which provides information to drivers, including when no communicating vehicle is nearby.

To do this, we first propose a spatio-temporal aggregation structure enabling a vehicle to summarize all the observed events. Next, we define a protocol for exchanging summaries between vehicles without the mediation of an infrastructure, allowing a vehicle to improve its local knowledge base by exchange with its neighbors. Finally, we define our operating strategies of the summary to assist the driver in making decision. We validated all of our proposals using the « VESPA » simulator by extending it to take into account the concept of summaries. Simulation results show that our approach can effectively help drivers make good decisions without the need to use a centralized infrastructure.

Keywords : Inter-vehicle network - Event Aggregation - Building Summaries - Exchange Protocol - Extracting knowledge.

Remerciements

Je voudrais exprimer mes sentiments les plus spontanés envers les personnes sans lesquelles ce travail de thèse n'aurait pas pu voir le jour. Leur aide, accompagnement et soutien m'ont été indispensables afin de pouvoir aboutir aux contributions de ma thèse.

Je voudrais tout d'abord exprimer ma reconnaissance envers tous les membres du jury pour la grande attention qu'ils ont bien voulu porter à mon travail.

Je suis très reconnaissante à *Bruno Defude*, mon directeur de thèse, qui m'a encadré et dirigé dans mes recherches tout au long de ces années. Je lui dis ma gratitude pour l'aide compétente qu'il m'a apportée, pour ses encouragements et pour la confiance qu'il m'a toujours témoignée.

Mes plus chaleureux remerciements vont également à *Thierry Delot*, qui a su m'aider pour mener à bien le présent travail. Sa disponibilité et ses conseils m'ont permis de mieux cerner les problèmes et ainsi de mieux y répondre.

Je remercie *Amel Mammar* pour ses efforts lors de la relecture de ce mémoire et pour le temps qu'elle m'a accordé et les remarques constructives qu'elle m'a données.

Je remercie les membres de l'équipe SIMBAD et du département informatique de Télécom SudParis : Brigitte Houassine qui nous simplifie énormément la vie au sein du département. Samir Tata pour ces conseils. Je remercie aussi toutes les personnes que j'ai connu au cours de cette thèse et qui l'ont rendu plus agréable : Alda, Amel, Nedja, Gabriel, Jérôme, Leon, Marie, Mohamed, Mouna, Mourad, Amin, Chan, Rami, Rim, Yassin, Zahra, Zied, Amina, Olfa, Ines, Salma, Amira.

Un grand merci à ma famille pour son soutien tout au long de cette thèse. Mes parents, Najet et Noureddine, pour les sacrifices qu'ils ont faits en faveur de mon éducation et sans qui je n'aurais pas pu réaliser cette thèse. Je tiens également à remercier mon frère Mehdi.

Table des matières

1	Introduction Générale	1
1.1	Contexte	1
1.2	Problématiques	3
1.3	Contributions	5
1.4	Organisation du mémoire	6
2	État de l'art	9
2.1	Introduction	9
2.2	Protocoles de communication pour VANETs	11
2.2.1	Dissémination d'informations dans les VANETs	12
2.2.2	Synthèse	13
2.3	Agrégation dans les VANETs	14
2.3.1	Type d'agrégation dans les VANETs	14
2.3.2	Solutions d'agrégation existantes	16
2.3.3	Synthèse	19
2.4	Assistance au conducteur dans les VANETs	19
2.4.1	Systèmes autonomes	20
2.4.2	Systèmes coopératifs	21
2.4.3	Systèmes interactifs	22
2.4.4	Synthèse	24
2.5	Conclusion	25
3	Approche générale	27
3.1	Introduction	27

3.2	VESPA	28
3.2.1	Principes de VESPA	28
3.2.2	Classification et représentation des événements	31
3.3	Pourquoi agréger ?	34
3.4	Architecture globale	36
3.5	Conclusion	37
4	Construction du résumé	41
4.1	Introduction	41
4.2	Modèle spatio-temporel à deux niveaux	42
4.2.1	Niveau physique	44
4.2.2	Niveau logique	45
4.3	Construction des résumés des événements	47
4.3.1	Les Sketches	48
4.3.2	Structure d'agrégation	52
4.3.3	Evaluation théorique de la taille de la structure d'agrégation	55
4.4	Conclusion	57
5	Processus d'échange de résumés	59
5.1	Introduction	59
5.2	Propriétés du processus d'échange	60
5.3	Principe du processus d'échange	62
5.3.1	Notion de priorité	64
5.3.2	Matching des priorités	69
5.3.3	Fusion des résumés	71
5.4	Trace des derniers véhicules et échanges	72

5.5	Dynamique du processus d'échange	74
5.6	Conclusion	77
6	Exploitation des résumés	79
6.1	Introduction	79
6.2	Architecture d'exploitation du résumé	81
6.3	Niveau d'accès aux résumés	82
6.3.1	Cas de recherche d'une ressource	83
6.3.2	Cas de production d'alerte	85
6.4	Niveau d'assistance aux conducteurs	86
6.4.1	Recherche de ressources	86
6.4.2	Production d'alertes	90
6.5	Conclusion	91
7	Simulation et validation expérimentales	93
7.1	Introduction	93
7.2	Choix du simulateur	94
7.2.1	Les simulateurs existants	94
7.2.2	Simulateur « VESPA »	96
7.2.3	Adaptation du simulateur	97
7.3	Paramètres de simulations	98
7.4	Critères et stratégies utilisés dans les simulations	100
7.5	Évaluation qualitative du résumé spatio-temporel	102
7.6	Évaluation du processus d'échange	106
7.7	Dynamique des échanges	108
7.8	Conclusion	114

8 Conclusions et perspectives	117
8.1 Conclusions	117
8.2 Perspectives	119
Bibliographie	123

Table des figures

1.1	Exemple du réseau VANET [DKHIL 2009]	3
2.1	Classification des stratégies d'agrégation	15
3.1	l'architecture du <i>VESPA</i>	29
3.2	L'aide à la décision en utilisant <i>VESPA</i>	31
3.3	L'architecture globale	39
4.1	Echange des connaissances entre les deux véhicules	43
4.2	Découpage spatial au niveau physique (portion de Paris)	44
4.3	Niveau logique	46
4.4	Insertion d'un événement dans le sketch d'événement	50
4.5	Estimation de nombre d'événements dans le sketch S_1	51
4.6	Structure d'agrégation	54
5.1	Principe d'échange	64
5.2	Environnement d'échange	66
5.3	Matching des priorités de V_1 avec les résumés de V_2	69
5.4	fusion des sketches dans V_1	71
5.5	Extrait du fichier XML des traces des derniers véhicules et échanges	73
5.6	Nombre de véhicules rencontrés pour la variation de $\alpha\beta$	75
5.7	Nombre de véhicules rencontrés pour la variation de γ	76
6.1	Modèle d'assistance au conducteur	81
6.2	Cellules à distance 1	87

6.3	Moyenne des cellules voisines	88
6.4	Cellules à distance 1 et plus	89
6.5	Interface d’alerte des zones dangereuses	90
7.1	Interface graphique du simulateur	99
7.2	Mesures de qualité pour les différentes stratégies	104
7.3	Impact de la portée d’observation	105
7.4	Impact du processus d’échange	107
7.5	Évolution du graphe d’échange distance 100 m pendant la simulation	110
7.6	Évolution du graphe d’échange distance 50 m pendant la simulation	111
7.7	Résultats avec diminution du rayon d’échange	113

Liste des tableaux

3.1	Représentation des événements	33
3.2	Exemple de message généré pour annoncer la libération d'une place de parking	33
4.1	Résultat de l'évaluation de la taille de la structure	56
5.1	Représentation temporelle d'échange de résumés	72
7.1	Caractéristiques des 8 parkings considérés durant les simulations	99
7.2	Analyse des graphes 7.5(a) et 7.5(b)	112
7.3	Analyse des graphes 7.6(a) et 7.6(b)	112

Introduction Générale

Sommaire

1.1	Contexte	1
1.2	Problématiques	3
1.3	Contributions	5
1.4	Organisation du mémoire	6

1.1 Contexte

La voiture est de nos jours le moyen de locomotion le plus répandu. Aujourd'hui cette voiture est de plus en plus munie d'équipements électroniques et informatique qui ont conduit à l'apparition du terme « *véhicule intelligent* ». Ainsi, armé d'un ensemble de radars, de caméras, de scanners et de capteurs, un véhicule intelligent se trouve capable de reconnaître les objets, de détecter les marquages au sol, de "lire" les panneaux de signalisation, d'éviter les cyclistes et les dangers, d'interpréter les comportements des piétons ou encore de s'adapter aux conditions météo et à la surface du sol. Un ensemble de véhicules intelligents construit ce que nous appelons « *les systèmes de transport intelligents* » (*ITS - Intelligent Transportation System*) qui visent à améliorer la sécurité et l'efficacité dans les transports routiers ou encore à réduire les émissions de gaz polluants.

Dans l'environnement routier, les systèmes de gestion de trafic sont basés sur des infrastructures centralisées où des caméras et des capteurs implantés sur la route qui collectent des informations sur la densité et l'état du trafic et

transmettent ces données à une unité centrale pour les traiter et prendre les décisions adéquates. De tels systèmes exhibent un coût de déploiement assez important et se caractérisent par un temps de réaction long pour le traitement et le transfert des informations, dans un contexte où le délai de transmission de l'information revêt une importance majeure. De plus, les équipements mis en place sur les routes nécessitent une maintenance périodique et chère. Par conséquent, pour un déploiement à grande échelle, un important investissement dans l'infrastructure de communication et de capteurs est nécessaire. Cependant, avec le développement rapide des technologies de communication sans fil, des systèmes de localisation et de collecte d'information via différents capteurs, une nouvelle architecture décentralisée (ou semi-centralisée) basée sur des communications véhicule à véhicule (V2V - Vehicle-to-Vehicle) suscite ces dernières années un réel intérêt auprès des constructeurs automobiles, de la communauté R&D et des opérateurs Télécoms. Ce type d'architecture s'appuie sur un système distribué et autonome, formé par les véhicules eux-mêmes, sans l'appui nécessaire d'une infrastructure préexistante pour la diffusion des données et des messages. Nous parlons dans ce cas d'un réseau Ad-Hoc de véhicules (VANET - Vehicular Ad-Hoc NETwork) qui vise à permettre les communications au sein d'un groupe de véhicules à portée les uns des autres (Inter-Vehicle Communication - IVC) et avec les équipements fixes de la route à portée pour fournir des informations opportunes aux conducteurs et aux autorités intéressées. Un exemple de réseau VANET urbain est illustré dans la figure 1.1.

De nouveaux services s'appuient sur cette capacité d'échanger et de propager de l'information et ensemencent un nouvel écosystème mêlant les transports, les opérateurs de télécommunication et les services. Ces avancées sont le plus souvent au service de la sécurité routière et du confort. Le « véhicule intelligent » comprend des dispositifs permettant d'assister le conducteur dans sa conduite avec l'objectif de lui offrir d'avantage de confort et de sécurité. Le conducteur sera donc plus assisté, gèrera mieux ses temps de trajets et consommera moins de carburant.

Nos travaux se déroulent plus précisément dans le cadre du projet *VESPA*¹

1. Une description détaillée est présentée dans <http://www.univ-valenciennes.fr/ROI/SID/tdeLOT/vespa/>

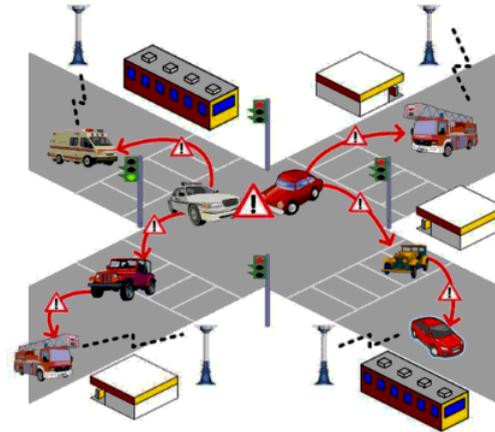


FIGURE 1.1 – Exemple du réseau VANET [DKHIL 2009]

(Vehicular Event Sharing with a mobile P2P Architecture) qui est un système d'assistance aux conducteurs qui exploite les réseaux ad-hoc inter-véhicules pour échanger des évènements entre véhicules [Delot 2010].

1.2 Problématiques

Il existe aujourd'hui une quantité très importante d'informations utiles pour alerter les conducteurs lorsqu'un événement survient (les accidents, les embouteillages, les travaux, les freinages d'urgence, les places de stationnement disponibles, la présence de radar de police, les véhicules d'intervention d'urgence). Ces événements peuvent être observés directement et en temps réel par le véhicule. Il est par ailleurs intéressant que toutes ces données soient partagées entre les véhicules dans le cadre d'une « conduite collaborative » basée sur un échange de renseignements entre véhicules munis d'instruments (capteurs par exemple) leur permettant de percevoir ce qui les entoure et de collaborer en groupes formés dynamiquement. Le conducteur peut ainsi obtenir une meilleure connaissance des caractéristiques du réseau routier sur lequel il évolue.

Des contributions intéressantes concernant l'échange d'informations entre les véhicules ont été proposées récemment dans le projet « VESPA » [Cenerario 2007,

Cenerario 2008] et d'autres projets de recherche relatifs à l'environnement routier [Lochert 2007, Lopez 2005, Morsink 2003, Nadeem 2006].

Cependant ces événements ont une durée de vie en général assez courte. Ces travaux ne considèrent en effet les données échangées entre les véhicules que comme des éléments à transmettre. Une fois ces données utilisées, elles sont considérées obsolètes et détruites.

Selon nous, il existe au contraire un intérêt à conserver ces données, en les agrégeant pour limiter l'espace nécessaire, pour pouvoir les utiliser à posteriori et produire, au niveau du véhicule, des connaissances exploitables sur l'environnement. A titre d'exemple, les événements concernant les places de stationnement disponibles reçues par un véhicule peuvent être exploitées, lorsqu'il n'y a aucune place disponible diffusée par les autres véhicules, pour déterminer l'endroit où la probabilité de trouver une place libre est la plus importante (en fonction du jour et de l'heure par exemple). Dans un autre contexte, grâce à la corrélation des différents messages reçus sur les accidents et les freinages d'urgence, les zones dangereuses peuvent être dynamiquement détectées et indiquées au conducteur, qu'elles soient continuellement dangereuses ou temporairement du fait des conditions climatiques par exemple. Dans cette situation un automobiliste peut être prévenu de la présence d'une zone dangereuse bien avant d'arriver à cette zone ce qui lui permet de prendre les précautions nécessaires (ex. diminuer la vitesse, allumer les feux, etc.) et d'éviter un possible accident.

La construction du résumé dans le contexte des réseaux sans fil n'est pas une idée récente. Outre les réseaux véhiculaires, ce problème a été largement étudié dans les réseaux de capteurs. Plusieurs travaux [Madden 2002, Yi 2008, Rajagopalan 2006] sont allés dans le sens de fournir un résumé des données qui est essentiellement vu comme une méthode de compression d'informations pour optimiser le stockage et réduire les besoins en bande passante. En effet, les véhicules peuvent être équipés de terminaux embarqués dont la capacité de stockage et les ressources CPU peuvent être limitées. Ces contraintes tendent à disparaître avec les avancées technologiques telles que les cartes de stockages externes « Secure Digital » par exemple tout en restant limité dans l'espace de stockage à quelques gigaoctet (*Go*).

L'originalité de notre approche réside clairement dans le fait de considérer le résumé comme une base de connaissances permettant la prise de décision même en absence d'informations précises. Comme pour les événements observés directement, les agrégats peuvent également être échangés pour élargir les connaissances de chaque véhicule. Plusieurs défis doivent être cependant surmontés pour permettre l'implantation de la communication dans les systèmes de transport intelligents. Les véhicules, sur la route se déplacent potentiellement vite. Ainsi, entre deux instants proches, la topologie du réseau peut être très différente. Des véhicules proches à un instant donné peuvent ne plus l'être l'instant d'après. Cela signifie donc que des véhicules qui peuvent communiquer à un instant donné ne le pourront peut-être plus quelques millisecondes plus tard. Il faut donc choisir la meilleure façon d'utiliser les communications pour récolter le plus d'informations utiles. La grande difficulté vient du nombre pharamineux de facteurs et de possibilités à choisir et ajuster pour optimiser la quantité d'informations reçues et sa fiabilité.

1.3 Contributions

L'objectif de nos recherches est que les données reçues par un véhicule ne servent pas uniquement pour produire un message d'alerte à destination du conducteur. Une fois stockées sur un véhicule, il est en effet possible d'utiliser à posteriori ces données collectées pour produire, au niveau du véhicule, des connaissances sur l'environnement exploitables par le conducteur. Ceci se décline en trois contributions majeures qui seront évaluées dans une quatrième partie :

1. **la proposition d'une structure d'agrégation spatio-temporelle** permettant à un véhicule de résumer l'ensemble des événements observés sur la route. Cette structure doit être capable de stocker des informations sur des événements observés directement par le véhicule, provenant de l'infrastructure (ex. capteur sur les routes) ou de l'échange avec d'autres véhicules. Elle doit être prête à gérer les différents types d'événements produits sur la route donc incrémentale en construction tout en restant compacte. Notre structure d'agrégation doit être adaptable au conduc-

- teur (choix de zones spatio-temporelles à résumer) et aussi échangeable avec d'autres véhicules (ce qui suppose qu'elle soit capable de détecter les doublons) ;
2. **la proposition d'une solution d'échange de résumés que nous appelons « processus d'échange »** qui cherche à échanger les agrégats entre les véhicules pour améliorer les connaissances locales à un coût raisonnable. Les propriétés du processus d'échange que nous proposons diffèrent de celles des protocoles existants dédiés à l'environnement routier dont le but est de spécialiser le routage des messages en fonction de critères de localisation géographique ;
 3. **la définition d'une approche pour exploiter les résumés** et aider le conducteur à prendre une décision sur la route. Une approche déclarative d'exploitation des résumés basée sur l'utilisation d'un langage de requêtes doit permettre à la fois le filtrage simple sur un type d'événement (rechercher la probabilité d'occurrence d'un événement dans une zone spatio-temporelle donnée), mais aussi d'aller plus loin en recherchant notamment des corrélations entre événements. Les types de requêtes à supporter sont des requêtes spatio-temporelles classiques pour trouver une ressource donnée par exemple une place de parking à un instant donné dans une zone donnée. Les requêtes spatio-temporelles continues permettent quant à elles d'avertir le conducteur dès qu'une probabilité forte de danger peut survenir ;
 4. **la validation des solutions proposées** tant du point de vue faisabilité que performance. Les contributions présentées dans cette thèse ont été implémentées au sein du simulateur « *VESPA* » qui a été modifié en conséquence.

1.4 Organisation du mémoire

Hormis l'introduction et la conclusion, ce mémoire comporte quatre parties présentées en six chapitres répartis de la manière suivante :

Le chapitre 2 : état de l'art, présente un état de l'art des solutions existantes dans la littérature. Il comporte trois parties. Tout d'abord, nous

présentons les protocoles de dissémination dans les VANETs. Ensuite nous exposons les travaux liés à l'agrégation d'événements spatio-temporels. Une troisième partie consiste à décrire les travaux de recherche menés pour l'aide à la décision du conducteur. Enfin nous clôturons ce chapitre par une synthèse des principaux travaux.

Le chapitre 3 : Approche générale, présente les principes du système *VESPA* dans lequel s'intègre notre travail. Ce chapitre décrit notamment les éléments sur lesquels nous nous appuyons pour définir nos contributions. Nous y identifions également nos besoins en terme d'agrégation et l'architecture globale de notre système.

Les chapitres 4 : construction du résumé, 5 : processus d'échange de résumés et 6 : exploitation des résumés, constituent le cœur de notre travail. Nous y détaillons respectivement nos approches. Nous présentons d'abord notre modèle spatio-temporel à deux niveaux ainsi que notre structure d'agrégation dans le chapitre 4. Ensuite dans le chapitre 5, nous décrivons notre deuxième contribution qui consiste en un processus d'échange des agrégats/résumés entre les véhicules. Enfin notre troisième contribution est l'objet du chapitre 6. Nous y explicitons nos stratégies d'exploitation de résumés pour aider le conducteur à prendre une décision sur la route.

Le chapitre 7 : simulation et validation expérimentales, présente, quant à lui, les différentes conclusions issues de l'implémentation des trois contributions et les résultats expérimentaux menés pour évaluer notre structure d'agrégation et le comportement du protocole d'échange.

Le chapitre 8 : conclusions et perspectives, présente une synthèse des aspects abordés dans cette thèse, dresse un bilan des principaux résultats de nos travaux et conclut sur les perspectives des travaux présentés.

État de l'art

Sommaire

2.1	Introduction	9
2.2	Protocoles de communication pour VANETs	11
2.2.1	Dissémination d'informations dans les VANETs	12
2.2.2	Synthèse	13
2.3	Agrégation dans les VANETs	14
2.3.1	Type d'agrégation dans les VANETs	14
2.3.2	Solutions d'agrégation existantes	16
2.3.3	Synthèse	19
2.4	Assistance au conducteur dans les VANETs	19
2.4.1	Systèmes autonomes	20
2.4.2	Systèmes coopératifs	21
2.4.3	Systèmes interactifs	22
2.4.4	Synthèse	24
2.5	Conclusion	25

2.1 Introduction

Les systèmes de transport intelligent (ITS - Intelligent transportation system) [Richard 2000, Figueiredo 2001, Yan 2007] ont reçu une attention toute particulière ces dernières années. Dans ce contexte, le développement des technologies de l'information et de la communication (e.g., dispositifs mobiles, réseaux sans fil) a ouvert de nouvelles perspectives avec l'essor des réseaux

véhiculaires. Les communications inter-véhicules (IVC) ou les réseaux ad hoc inter-véhicules (VANETs) permettent en effet d'échanger des données entre véhicules afin de fournir aux automobilistes des informations relatives à (1) la sécurité routière, (2) l'aide à la conduite, ou (3) le confort du conducteur et des passagers.

Dans le contexte de la communication inter-véhicules, un certain nombre de projets d'envergure avec des objectifs variés ont été menés à bien depuis le début des années 2000. Ainsi, le projet Fleetnet¹ - Internet on the Road [Festag 2004], est un projet allemand initié par un consortium d'industriels associé à trois universités. L'objectif de FleetNet était de développer une plateforme de communication pour les réseaux de véhicules et de standardiser les solutions proposées afin d'améliorer la sécurité et d'assurer un meilleur confort au conducteur et à ses passagers. L'architecture FleetNet repose sur un mécanisme de routage basé sur un système de localisation et de navigation, et considère les communications véhicule-à-infrastructure afin de fournir un accès à l'Internet.

Le consortium de communication Car2Car² a été initié par six constructeurs automobile européens et ouvert à d'autres partenaires comme des organismes de recherches. Le consortium Car2Car s'est fixé comme objectif d'améliorer la sécurité routière, de gérer efficacement le trafic à travers l'utilisation des communications inter-véhicules. Les principales missions du consortium de communication Car2Car sont les suivantes : (1) la création d'un standard européen ouvert pour les communications véhicule-à-véhicule basé sur des composants LAN sans fil, (2) le développement de prototypes et de démonstrateurs des systèmes V2V pour les applications de sécurité routière, (3) l'attribution d'une bande de fréquence exclusive libre pour les applications de Car2Car en Europe, et (4) le développement de stratégies de déploiement et des modèles économiques pour la pénétration du marché.

Le projet européen IST CarTalk2000³ (coordonné par le constructeur Daimler Chrysler entre 2001 et 2004) [Morsink 2003] avait pour but de dé-

1. FleetNet project- Internet on the road, <http://www.et2.tu-harburg.de/fleetnet>

2. Car2Car Communication Consortium, www.car-to-car.org

3. Safe and comfortable driving based upon inter-vehicle communication, <http://www.cartalk2000.net>

velopper des systèmes coopératifs d'aide à la conduite et de mettre en œuvre un réseau pur ad hoc sans fil auto-organisé. Outre les aspects technologiques, le projet a également étudié des facteurs liés aux stratégies d'introduction dans le marché comprenant des analyses de coûts et des aspects légaux.

Le projet NOW (Network-on-Wheels)⁴ (2004-2008) est un projet allemand du ministère fédéral de l'éducation et de la recherche, impliquant des constructeurs automobile, des acteurs de télécommunication et des académiques. NOW soutient et coopère fortement avec le consortium Car2Car. Les protocoles de communication développés pour le projet sont aussi bien dédiés aux applications de sécurité qu'aux applications de divertissement et fournissent ainsi une plate-forme de communication ouverte pour un large éventail d'applications. L'un des principaux objectifs du projet NOW est la mise en œuvre de protocoles de communications et d'algorithmes de sécurité des données dans les réseaux de véhicules. En se basant sur la technologie sans fil IEEE 802.11 et un routage basé sur la position dans un contexte de communication V2V ou véhicule à infrastructure, le but du projet NOW était d'implémenter un système de référence et de contribuer à la standardisation d'une telle solution en Europe en collaboration avec le consortium Car2Car.

Dans la suite de ce chapitre, nous nous intéressons tout d'abord aux protocoles de dissémination dans les VANETs. Nous verrons ensuite dans la section 2.3 les travaux qui ont été menés sur l'agrégation des données et nous identifions ces limites. Nous continuons dans la section 2.4 avec une description des différentes solutions actuelles liées à l'assistance aux conducteurs. Nous concluons dans la section 2.5, et nous présentons la direction de nos recherches.

2.2 Protocoles de communication pour VANETs

Dans les VANETs, la dissémination consiste à acheminer, au travers de communications multi-sauts, une information d'une source (e.g., véhicule ou

4. NOW (Network-On-Wheels), www.network-on-wheels.de

hotspot) vers une ou plusieurs destinations (véhicule(s)), en limitant le délai d'acheminement et en maximisant les garanties de bon acheminement.

2.2.1 Dissémination d'informations dans les VANETs

De nombreux protocoles ont été proposés ces dernières années pour acheminer des informations à un ensemble de destinataires via un réseau véhiculaire.

Ainsi, IVG (Inter-Vehicle Geocast) [Bachir 2004] est une méthode de diffusion qui permet de surmonter les problèmes de dynamique du réseau grâce à des relais dynamiques introduits pour rediffuser périodiquement les messages d'alerte.

Dans [Wu 2004], les auteurs introduisent MDDV (Mobility-Centric Data Dissemination Algorithm for Vehicular Networks), un algorithme de diffusion considérant que les véhicules ne disposent pas des positions des véhicules voisins. Dans cette approche, le réseau routier est modélisé comme un graphe orienté où les nœuds représentent les intersections, et les liens les segments routiers. Un poids est associé à chaque lien pour refléter la distance et la densité de trafic correspondante. MDDV utilise une trajectoire de relayage spécifiée comme le chemin ayant la plus petite somme des poids d'une source vers la "région destination" dans le graphe orienté.

Dans [Lochert 2007], Lochert et al. se concentrent sur les environnements urbains. Ils font notamment une distinction intéressante entre le transport des informations via locomotion (avec des stratégies dénommées « store-and-forward » [Adler 2006] ou « carry-and-forward » [Zhao 2006] dans lesquelles les véhicules transportent une information dans une zone géographique où elle peut être diffusée) et via dissémination (diffusions successives). Cette distinction permet notamment de supporter des densités de véhicules différentes lors de la dissémination.

Tonguz et al. [Wisitpongphan 2007] fournissent un cadre global (DV-CAST) pour gérer la diffusion dans les VANETs en considérant trois densités de trafic possible (dense, régulier, et rare). Dans un trafic dense, ils suggèrent d'utiliser une technique nommée timer-based [Wisitpongphan 2007], tandis que dans le trafic rare, ils vous suggèrent d'utiliser multi-diffusion role-based [Briesemeister 2000].

Ils ne suggèrent pas une technique spécifique pour qu'elle soit utilisée en cas de trafic à densité régulière car ils définissent la densité régulière comme un mélange de certains véhicules de détection de trafic dense et d'autres de détection du trafic rare. Un travail dans [Rahim 2009] consiste à proposer un mécanisme d'ordonnancement des messages à diffuser. Il explique le fait qu'une information soit pertinente n'est pas suffisant dans le cas où plusieurs messages doivent être diffusés. Cet ordonnancement est modélisé par des files d'attente qui permettent de donner la priorité aux événements les plus importantes.

Dans [Xu 2004b], les auteurs proposent une approche inspirée de l'épidémiologie pour la dissémination d'informations sur des ressources (e.g., places de stationnement) entre véhicules. Un véhicule en possession d'une information peut ainsi en contaminer d'autres rencontrés sur son trajet. Pour contenir les informations disséminées dans une zone spatio-temporelle déterminée dans laquelle elle est pertinente, les auteurs utilisent une fonction de pertinence évaluée par chaque véhicule avant de relayer l'information.

Dans [Delot 2011], les auteurs introduisent une probabilité de rencontre qui permet notamment d'adapter la dissémination des événements dans le réseau en fonction du type d'informations considérées (e.g., place de stationnement, accident, freinage d'urgence, etc.) [Cenerario 2011]. Cette probabilité est exploitée à la fois pour déterminer si une information reçue par un véhicule doit être communiquée au conducteur ou relayée vers d'autres véhicules. Ce travail s'intègre dans le projet VESPA⁵.

2.2.2 Synthèse

La dissémination d'informations dans un réseau inter-vehicules avec les différents protocoles mentionnés dans la section 2.2.1 repose sur une approche coopérative où chaque noeud du réseau accepte de relayer les informations afin par exemple d'améliorer la sécurité sur les routes ou de réduire les temps de parcours. Ces protocoles doivent veiller à limiter le nombre de véhicules relayant l'information dans le réseau afin notamment d'éviter l'inondation de celui-ci [Lochert 2007]. La plupart des protocoles utilisent pour ce faire des

5. [urlhttp://www.univ-valenciennes.fr/ROI/SID/tdelot/vespa/](http://www.univ-valenciennes.fr/ROI/SID/tdelot/vespa/)

mécanismes de contention [Füßler 2003]. D'autres approches utilisent des techniques de "machine learning" afin d'adapter le processus de dissémination en fonction de la disponibilité d'une information dans le réseau [Szczurek 2010].

Les protocoles présentés ici ne considèrent que l'échange de très faibles volumes de données. Les événements sont en effet représentés par des messages de très petites tailles contenant quelques attributs (e.g., localisation et type de l'événement traité). L'échange entre véhicules de volumes de données plus important pose évidemment de gros problèmes dans des environnements aussi dynamiques que les réseaux véhiculaires. Dans la section suivante, nous discutons des mécanismes d'agrégation généralement utilisés dans ce cas pour réduire la quantité d'informations à échanger.

2.3 Agrégation dans les VANETs

2.3.1 Type d'agrégation dans les VANETs

Dans la littérature les techniques d'agrégation sont classées en deux catégories : syntaxique ou sémantique. [Fabio 2006] L'agrégation syntaxique consiste à compresser ou encoder les données provenant de plusieurs véhicules dans le but de rendre les données dans un même format. Dans ce cas la structure d'agrégation comporte la localisation, la vitesse et l'identifiant du véhicule. Ce type d'agrégation réduit essentiellement la surcharge de messages entraînant un gain en terme de coût d'envoi.

L'agrégation sémantique, est applicable uniquement sur des types d'informations comme la localisation et la vitesse. Avec ce type d'agrégation, au lieu de rapporter la position exacte de cinq véhicules, seul le fait que cinq véhicules existent sont signalés. Le compromis permet de générer un message à diffuser beaucoup plus petit au prix toutefois d'une dégradation de la précision des données.

Une classification des stratégies d'agrégation est également introduite en fonction du lieu d'exécution et de la manière dont elles sont appliquées [Chen 2003]. Cette classification est montrée dans la figure 2.1.

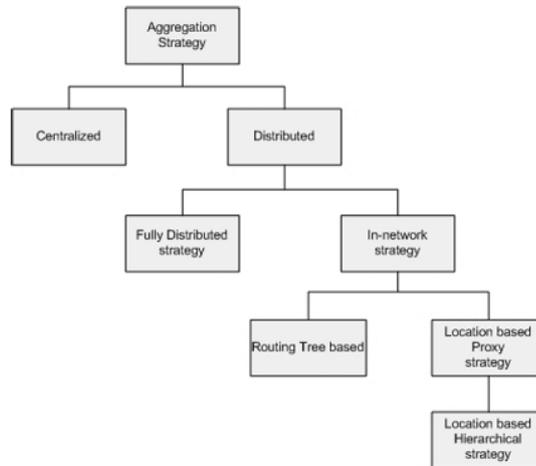


FIGURE 2.1 – Classification des stratégies d’agrégation

Le classement présenté dans la figure 2.1 distingue plusieurs niveaux. Le premier niveau est « Qui effectue l’agrégation de données : l’expéditeur de requêtes ou d’autres nœuds ». Si l’expéditeur de requête est celui qui effectue l’agrégation, la stratégie est dite centralisée, sinon la stratégie est dite distribuée. Les stratégies d’agrégation distribuées sont classées selon un deuxième critère de classification : « quelles nœuds effectuent l’agrégation ? (i.e., tous les nœuds interrogés ou certains autres nœuds) ». Si tous les intervenants interrogés agrègent les données, la stratégie d’agrégation est entièrement distribuée. Dans le cas où certains nœuds seulement sont concernés par l’agrégation alors il s’agit d’une agrégation dite en réseau, ce qui signifie que les données brutes sont agrégées quand elles sont envoyées à l’expéditeur de la requête.

Un autre type d’agrégation est étudié dans [Christian 2007] correspondant à des situations où l’utilisateur a besoin d’une connaissance fine de son voisinage et où cette connaissance peut devenir plus grossière au fur et à mesure que nous nous éloignons de ce voisinage. Ce type d’agrégation est appelé agrégation hiérarchique. Avec ce type d’agrégation nous partons de N individus nous agrégeons les deux les plus proches, et nous considérons le couple agrégé comme un nouvel individu. Nous sommes alors ramenés au problème précédent avec maintenant $(N - 1)$ individus. Nous agrégeons à nouveau les deux plus proches et nous itérons le processus jusqu’à n’avoir plus qu’un seul

individu, qui sera le sommet de l'« arbre hiérarchique ».

2.3.2 Solutions d'agrégation existantes

L'agrégation de données a été particulièrement étudiée dans le contexte des réseaux de capteurs [Madden 2002, Yi 2008, Rajagopalan 2006]. Plusieurs techniques d'agrégation développées dans le cadre des données spatio-temporelles existent également dans la littérature [Lopez 2005]. Dans [Stefan 2009] l'objectif principal de l'agrégation est de minimiser la consommation d'énergie en combinant des données provenant de nœuds de capteurs différents afin de réduire la redondance et minimiser ainsi le coût de la communication [Bhaskar 2002]. Toutefois, ces mécanismes de regroupement ne peuvent pas être appliqués directement pour VANETs pour plusieurs raisons :

- la plupart des mécanismes utilisent les nœuds de capteurs fixes afin de réaliser l'agrégation hiérarchique. Par exemple, les données sont agrégées sur une chaîne de nœuds [Lindsey 2002], une hiérarchie d'arborescence [Stephanie 2001, Haowen 2006] ou un cluster [Rabiner 2000]. Dans des réseaux très dynamiques comme les VANETs, la construction de telles hiérarchies n'est pas triviale, leur maintenance encore moins ;
- les architectures des réseaux de capteurs : de nombreuses applications considèrent dans ce contexte une station de base qui interroge les capteurs. Après une telle requête, les nœuds de capteurs propagent les informations collectées vers la station de base. Les protocoles qui utilisent ce mécanisme [Intanagonwiwat 2000], ne peuvent pas être déployés pour de nombreuses applications VANET puisqu'il n'y a pas de station de base dédiée. Tous les véhicules dans une région spécifique sont censés être des récepteurs.

Hormis les travaux sur l'agrégation dans les réseaux de capteurs, plusieurs contributions ont spécifiquement considéré le problème de l'agrégation dans les VANETs.

Une proposition d'Eichler et al. [Eichler 2006] consiste à considérer des véhicules qui agrègent des données relatives à des notifications s'ils reçoivent plusieurs messages relatifs au même événement. Ils proposent également l'uti-

lisation de messages d'invalidation lorsqu'un véhicule ne détecte pas un danger dans une zone définie comme étant dangereuse d'après les informations agrégées.

L'agrégation de messages peut se faire en fonction de zones géographiques comme dans [Saleet 2007], où le protocole RLSMP (Region-based Location Service Management Protocol) a été proposé. Le but est de réduire les mises à jour des positions et le nombre de messages générés pour la gestion des localisations des véhicules. Saleet et al. précisent toutefois que même si l'agrégation améliore la montée en charge, elle peut également conduire à :

- plus de collisions de paquets et donc à plus de retransmissions (essentiellement parce que les paquets échangés sont de plus grande taille) ;
- des délais plus importants, du fait des traitements réalisés sur les données avant qu'elles puissent être effectivement envoyées.

Si l'agrégation peut réduire les problèmes de consommation de bande passante, elle rend aussi les problèmes de sécurité plus compliqués à gérer. La solution proposée dans [Fabio 2006] traite de ces aspects sécurité et plus précisément de la détection des attaques consistant à disséminer des messages contenant des données agrégées erronées. La solution proposée consiste en un service infalsifiable déployé sur les véhicules.

Dans [Kulik 2002], l'agrégation des données est définie par une technique utilisée pour surmonter deux problèmes : *l'implosion* (i.e., les données détectées par un noeud sont dupliquées dans le réseau en raison des stratégies de routage utilisées) et *les recouvrements* (i.e., deux noeuds différents diffusent les mêmes données).

Dans [Nadeem 2004], Nadeem et al. présentent le système TrafficView qui utilise l'agrégation sémantique. Les auteurs présentent deux techniques pour l'agrégation : *ratio-based* et *cost-based*.

- l'algorithme « ratio-based » considère une division de la route devant le véhicule en un ensemble de segments. A chacun de ces segments est ensuite associé un ratio d'agrégation ;
- l'algorithme « cost-based » vise quant à lui à minimiser le coût d'agrégation des enregistrements (basé sur l'erreur introduite durant la fusion,

le nombre de véhicules affectés par l'agrégation, etc.).

Une autre solution pour l'agrégation est présentée dans [Fabio 2006]. Cette solution agrège seulement des informations concernant les voitures, telles que les vitesses ou les localisations. Elle ne peut pas être utilisée pour des événements tels que des accidents. De même, la réagrégation n'est pas considérée.

Un autre protocole d'agrégation a été proposé dans [Cheng 2003] appelé LBAG (Location Based Aggregation) dans lequel l'agrégation des données repose sur une hiérarchie de localisations statiques au lieu de considérer une structure d'arbre de nœuds qui serait particulièrement difficile à maintenir du fait de la forte mobilité des véhicules. Dans le but d'éviter de compter plusieurs fois une même occurrence d'événement observé par deux véhicules différents, Lochert et al. [Lochert 2007] proposent une étude de l'agrégation hiérarchique basée sur l'utilisation des sketches. La motivation de cette approche est qu'un véhicule a besoin d'informations détaillées concernant son voisinage. Les informations relatives à des zones plus lointaines peuvent par contre être moins détaillées. Un algorithme basé sur l'utilisation de Flajolet-Martin sketches [Flajolet 1985] est proposé pour stocker des informations approximatives. Par exemple, il est possible de fusionner deux agrégats (même s'il y a un certain recouvrement entre les deux) tout en évitant l'apparition de doublons. La hiérarchie d'agrégation est prédéfinie dans la carte des données groupant des zones en fonction de leurs relations naturelles (e.g. par district ou routes). Dans le contexte des applications spatio-temporelles, les sketches sont également utilisés dans [Yufei 2004] comme un moyen d'éviter le problème du comptage pour les requêtes définies avec des fonctions d'agrégat « count » ou « sum ». Tao et al. proposent, dans [Yufei 2004], une structure pour agrégation d'événements. Elle combine le aRB tree (aggregate RB-) [Papadias 2002] avec le sketch Flajolet Martin [Flajolet 1985]. Ce travail propose une méthode pour l'indexation spatio-temporelle basée sur un R-tree pour la partie spatiale et un B-tree pour le temps partiel. La valeur mémorisée dans une cellule de l'arbre est un sketch pas un simple entier.

2.3.3 Synthèse

Nous avons présenté dans cette section 2.3 des travaux sur l'agrégation en soulignant leurs avantages et leurs limites. Nous pouvons diviser les solutions d'agrégation existantes, en deux catégories : agrégation pour le réseau de capteurs et agrégation pour les réseaux ad-hoc véhiculaires.

Nous avons mentionné que la première catégorie d'agrégation liée aux réseaux de capteurs ne peuvent pas être appliqués directement dans le contexte des VANETs vu leur forte dynamique et le besoin d'une station de base.

La deuxième catégorie des travaux porte sur l'agrégation dans les VANETs et considère l'agrégation comme un moyen d'optimiser le stockage et la compression des données afin de réduire les besoins en bande passante. L'agrégation et la compression sont distinguées dans [Nadeem 2004]. Seul le processus d'agrégation est une réduction basée sur la sémantique des données alors que la compression utilise des critères syntaxiques.

Dans cette thèse, nous nous intéressons aux résumés comme une base de connaissance d'une information approchée permettant la prise de décision même en l'absence d'informations précises. En d'autres termes le but n'est pas de compresser les données pour optimiser la quantité d'informations dans la bande passante mais plutôt construire un résumé d'événements produits sur la route qui sera exploité pour estimer si un événement est susceptible de survenir même sans observation directe.

Un travail préalable [Defude 2008] nous a permis de spécifier une première structure d'agrégation, basée sur un simple comptage des événements dans une cellule spatio-temporelle. Dans nos travaux nous voulons aller plus loin en proposant une structure d'agrégation plus efficace.

2.4 Assistance au conducteur dans les VANETs

La densité croissante du trafic routier, les normes de sécurité plus poussées, les lois environnementales plus strictes et les changements démographiques amèneront l'industrie automobile à relever des défis de plus en plus décisifs.

Les systèmes d'assistance au conducteur d'aujourd'hui contribuent considérablement à relever ces défis puisqu'ils assistent les conducteurs dans des situations dangereuses, déclenchent des procédures autonomes afin d'éviter des accidents ou de réduire la consommation de carburant.

Sous le vocable outils adaptés pour l'assistance, les solutions existantes dans l'assistance au conducteur se répartissent en trois catégories :

2.4.1 Systèmes autonomes

Ces systèmes sont les plus connus pour l'assistance au conducteur. Dans cette catégorie, le véhicule dispose simplement de capteurs (e.g., caméras, radars, etc.) embarqués capables d'apporter une connaissance sur l'environnement immédiat du véhicule. Il n'interagit pas avec les autres véhicules directement. Cependant, le véhicule dispose de système d'assistance à la conduite (ADAS) lui permettant par exemple de mesurer les distances par rapport aux véhicules qui le précèdent, de repérer les obstacles, de freiner automatiquement dans les situations d'urgence (freinage automatique à une seconde du choc) de détecter les écarts de trajectoire par détection du marquage au sol. Pour ces fonctions, il s'appuie sur des capteurs d'images et de télémétrie laser. Il peut également embarquer un système de localisation lié à une cartographie numérique.

Plusieurs consortiums se sont formés entre différents groupes de chercheurs, d'acteurs de l'industrie et des gouvernements pour faire avancer la recherche sur les systèmes d'assistances autonomes.

Une première catégorie de systèmes concernent *l'aide au contrôle de la vitesse* comme :

« **Speed Alert** »⁶ qui est un systèmes embarqués d'information sur les limites de vitesse. Il aide le conducteur à rester dans les limites de vitesse en utilisent une caméra qui peut lire les informations de limite de vitesse et de rappeler aux conducteurs de la limite de vitesse via un affichage sur le tableau de bord.

6. <http://www.tc.gc.ca/eng/roadsafety/safevehicles-1190.htm>

« **LAVIA** »⁷ : Limiteur s'Adaptant à la Vitesse Autorisée mise en ligne : mardi 2 janvier 2007. Le LAVIA (Limiter s'adaptant à la vitesse autorisée) est système qui permet au conducteur d'adapter sa vitesse aux changements de limitation de vitesse qui rythment sa route. C'est un dispositif d'aide à la conduite qui peut fonctionner selon plusieurs modes.

Des régulateurs de vitesse standards proposés par SCANIA⁸ et VOLVO⁹ qui essaient de maintenir une certaine vitesse sur une pente. Chaque conducteur peut désormais optimiser sa vitesse, y compris sur des routes inconnues et en pleine obscurité.

Une deuxième catégorie de systèmes concernent *l'avertissement des dangers* comme :

« **AFIL** » L'alerte de franchissement involontaire de ligne est un dispositif qui avertit le conducteur d'une automobile quand il franchit involontairement (sans actionner ses clignotants) une ligne continue ou discontinue. L'avertissement se manifeste par une vibration dans le siège du conducteur du côté où le franchissement a lieu. Le principe de ce système AFIL réside dans l'utilisation de diodes infrarouges et de capteurs sensibles à ces rayonnements, placés sous le bouclier avant du véhicule. Les variations de réflexion des rayonnements infrarouges sur les bandes blanches engendrent le déclenchement du système via le calculateur du véhicule. Ce système est proposé par Citroën puis Peugeot et adopté par Toyota BMW et Opel.

2.4.2 Systèmes coopératifs

Avec le déploiement de l'infrastructure et l'apparition des techniques de communication dans les réseaux véhiculaires, une deuxième catégorie de solution d'assistance au conducteur apparaît. Le conducteur dispose d'une meilleure connaissance de l'environnement, grâce à une communication avec l'infrastructure (V2I) qui permet, par exemple, de savoir les conditions de circulation en

7. <http://www.lavia.fr/>

8. <http://www.scania.fr/buses-coaches/safety-technology/safety-support-systems/index.aspx>

9. <http://www.volvotrucks.com/trucks/switzerland-market/fr-ch/trucks/volvo-fh-series/key-features/Pages/driver-support-systems.aspx>

aval, le niveau d'adhérence et plus globalement d'avoir des caractéristiques de la voie courante. Il dispose également d'une perception plus fine des véhicules environnants sur la section et sur les voies adjacentes.

« **ARCOS** »¹⁰ : Ce projet vise à réduire le nombre d'accidents sur la route. Il s'appuie sur l'approche triangulaire "infrastructure-véhicule-conducteur" afin d'organiser la perception de manière globale. Quatre fonctions identifiées comme contribuant fortement à l'amélioration de la sécurité routière ont été développées durant ce projet : alerter les véhicules en amont d'incident, gérer les interdistances entre les véhicules arrêtés ou lents, prévenir les collisions sur obstacles fixes et prévenir les sorties de voies. Différents modes de partage de la conduite, plus ou moins intrusifs, entre systèmes d'assistance et conducteurs ont été explorés.

« **COOPERS** »¹¹ : les véhicules circulant sur autoroutes sont connectés à l'infrastructure via une communication sans fil continue ; Ils échangent des données et des informations pertinentes relatives à la section d'autoroute traversée afin d'améliorer la sécurité et permettre une gestion coopérative du trafic.

Plus récemment, des applications pour smartphones visant à assister les conducteurs sont également apparues. Ces applications comme *Waze*¹², *Apila*¹³ ou *PlaceLib*¹⁴ exploitent les réseaux de téléphonie mobile et les récepteurs GPS embarqués dans les smartphones pour faciliter l'échange entre membres d'une communauté d'informations relatives aux conditions de trafic ou à la localisation de places de stationnement disponibles.

2.4.3 Systèmes interactifs

Tout véhicule échange des informations avec les autres véhicules environnants (V2V), tout comme avec l'infrastructure (V2I et I2V) et les opérateurs. Des véhicules traceurs, qui circulent en amont, font remonter des données.

10. Action de Recherche pour une Conduite Sécurisée : <http://www.arcos2004.com/>

11. Co-operative systems for intelligent road safety : <http://www.coopers-ip.eu/>

12. <http://www.waze.com>

13. <http://www.apila.fr>

14. <http://www.placelib.com>

L'analyse de ces informations permet de diagnostiquer des risques d'accidents liés soit à l'infrastructure, soit aux comportements des conducteurs. Dans ce niveau, le conducteur bénéficie d'un accès temps réel aux caractéristiques de la route et des conditions de circulation. Sa perception de l'environnement est optimisée par une fusion des différentes sources d'information (capteurs autonomes, systèmes coopératifs, échanges d'informations). Le véhicule peut, pour sa part, détecter en temps réel les défaillances pour la trajectoire et le comportement dynamique.

« **CVIS** »¹⁵ : Le projet a pour objectif de fournir un système ouvert pour le développement d'applications et de services coopératifs, en rendant transparent l'accès à une multitude de systèmes de communication à travers l'architecture CALM. Le principal objectif de CVIS est d'améliorer la qualité et la sécurité du réseau routier par la mise en place de systèmes et d'applications coopératives.

« **SAFESPOT** »¹⁶ [Roberto 2006] : En combinant les informations provenant des capteurs embarqués dans les véhicules et déployés sur l'infrastructure, le projet SAFESPOT permet une amélioration du délai de détection des risques d'incidents potentiels. La transmission d'alertes et de recommandations aux véhicules grâce aux communications véhicules-véhicules et véhicules-infrastructure, permet d'améliorer dans le temps et l'espace la vision qu'a le conducteur de son environnement.

COYOTE »¹⁷ est une solution d'informations géolocalisées permettant de communiquer en temps réel et en fonction du trajet de l'automobiliste, les variations de limitation de vitesse, la présence de zones dangereuses et les perturbations routières. « *COYOTE* » communique en permanence avec un serveur central qui transmet les informations pertinentes à l'utilisateur. Il permet aussi à chaque automobiliste de signaler en temps réel, à l'ensemble des autres membres de la communauté Coyote, la présence d'aléas de la route (dangers permanents, risques temporaires et perturbations routières).

15. Co-operative vehicle-infrastructure systems : <http://www.cvisproject.org/>

16. Cooperative vehicles and road infrastructure for road safety : <http://www.safespot-eu.org/>

17. Pour plus d'informations, vous pouvez visiter <http://www.moncoyote.com/fr/comment-ca-marche>

2.4.4 Synthèse

Les systèmes d'aide à la conduite sont un secteur en pleine mutation. La plupart des systèmes actuels présents en série dans les véhicules ont pour objectif de pallier aux défaillances du conducteur. Les systèmes d'aide à la conduite agissent sur la sécurité soit en aidant le conducteur à éviter un accident ou une situation à risque, soit en cherchant à en minimiser les conséquences.

Nous avons présenté les travaux existants pour l'assistance au conducteur que nous avons classé en trois catégories en se basant sur les outils adaptés pour l'assistance. Nous avons aussi souligné pour chaque solution son rôle dans l'assistance au conducteur. Nous pouvons constater que toutes les solutions existantes visent à aider le conducteur pour éviter l'apparition d'une situation dangereuse risquant d'aboutir à l'accident et libérer le conducteur d'un certains nombres de tâches qui pourraient atténuer sa vigilance. Aucun système exploite l'environnement du conducteur pour lui aider à prendre une décision liée par exemple à la recherche d'une place libre pour se garer. Dans ce contexte nous pouvons tirer profit du serveur « *VINCI Park* »¹⁸ la solution qui regroupe l'ensemble des activités dans le secteur du stationnement et des services qui lui sont liés pour l'opérateur VINCI. Cette solution va coûter chère à l'utilisateur (environ 322.00 € l'abonnement trimestriel) en plus des inconvénients de la centralisation des informations qui sont liés aussi aux solutions pour les systèmes coopératifs et interactifs. En effet, de tels systèmes exhibent un coût de déploiement assez important et se caractérisent par un temps de réaction long pour le traitement et le transfert des informations, dans un contexte où le délai de transmission de l'information est vital et revêt d'une importance majeure dans ce type de systèmes. De plus, les équipements mis en place sur les routes nécessitent une maintenance périodique et chère. Par conséquent, pour un large déploiement d'un tel système un important investissement dans l'infrastructure de communication et de capteurs est nécessaire.

Ce bilan des travaux d'assistance au conducteur nous permet de spécifier

18. Pour plus d'informations, vous pouvez visiter http://www.vincipark.com/vincipark.nsf/fr/les_services/pages/vincipark.com.htm

notre direction de recherche qui consiste à proposer une solution d'assistance interagissant avec le conducteur (basée sur de la communication V2V) visant non seulement l'alerte du conducteur pour éviter un accident ou une situation à risque mais aussi l'aide dans la prise de décision sur la route.

2.5 Conclusion

Dans ce chapitre nous avons présenté les travaux liés à l'une des principales composantes des systèmes ITS qui est la communication inter-véhicules IVC ou ce que nous appelons réseaux ad hoc mobiles inter-véhiculaires (VANETs). Des applications de ces systèmes ainsi qu'un ensemble de projets et travaux de recherche se rapportant à ce domaine ont été présentés. Un premier bilan nous a permis de conclure que les données reçues sont utilisées en temps réel pour produire un message d'alerte à destination du conducteur, une fois utilisées elles sont considérées obsolètes puis détruites d'où l'idée de construire une base de connaissance pour chaque véhicule.

Dans ce chapitre nous avons également présenté dans une deuxième section des travaux sur l'agrégation dans les réseaux de capteurs puis les réseaux VANET. Un deuxième bilan nous a permis de conclure que les solutions existantes visent la compression des événements, pour réduire la taille en bande passante, plutôt que la réutilisation en cas d'absence d'événements disséminés.

La troisième partie du chapitre consiste à décrire les solutions d'assistance au conducteur que nous avons classées en trois catégories selon l'outil adapté pour l'assistance. La majorité des solutions proposées respresentent un système d'alerte au conducteur et reposent sur une infrastructure très coûteuse d'où la nécessité de proposer un système interprétant des situations routières encore plus complexes et en mesure d'intervenir plus librement dans les actions du conducteur.

Dans les différents projets mentionnés dans ce chapitre, les données échangées entre les véhicules ne sont considérées que comme un élément à transmettre. Une fois ces données reçues par un véhicule, leur pertinence est évaluée, généralement sur des critères spatiaux et temporels, et, le cas échéant,

le conducteur est informé de la présence d'un événement. Une fois les données utilisées, elles sont considérées obsolètes et détruites. Pour palier à cette limite, nos travaux décrits dans la suite de ce manuscrit consistent à proposer une solution que nous pouvons l'utiliser avec la dissémination pour que les données reçues par un véhicule ne servent pas uniquement pour produire un message d'alerte à destination du conducteur. Une fois stockées sur un véhicule, il est en effet possible d'utiliser á posteriori ces données collectées pour produire, au niveau du véhicule, des connaissances sur l'environnement exploitables par le conducteur.

Approche générale

Sommaire

3.1	Introduction	27
3.2	VESPA	28
3.2.1	Principes de VESPA	28
3.2.2	Classification et représentation des événements	31
3.3	Pourquoi agréger ?	34
3.4	Architecture globale	36
3.5	Conclusion	37

3.1 Introduction

Les systèmes de transport intelligents [Festag 2004, Christian 2007] sont de plus en plus étudiés dans le milieu de la recherche et de l'industrie (automobile, télécommunications, etc.). En effet, le développement d'applications de transport intelligent a introduit les technologies de l'information et de la communication dans les systèmes de transport (IVC : inter-vehicule communication). L'objectif principal de l'IVC est d'améliorer la sécurité routière par le développement de nouvelles aides pour les conducteurs en échangeant des informations entre les véhicules à proximité. Par exemple, la communication inter-véhicules pourrait être utilisée pour informer les conducteurs qu'un accident s'est produit ou qu'un obstacle est apparu sur la route à quelques centaines de mètres plus loin.

Notre travail s'inscrit dans le projet *VESPA*¹ (Vehicular Event Sharing with a mobile P2P Architecture) [Delot 2011], un système conçu pour le partage d'informations entre véhicules en réseaux ad-hoc.

Même si nos travaux peuvent s'inscrire en dehors du contexte *VESPA*, ils supposent néanmoins la production et la dissémination d'événements entre les véhicules. Cette dissémination peut ne pas utiliser d'infrastructure comme dans *VESPA* ou bien utiliser partiellement une infrastructure soit pour relayer des informations entre les véhicules, soit pour fournir une partie des informations. Dans la suite, nous décrivons *VESPA* comme archétype d'un système basé sur la dissémination des événements et comme substrat de nos contributions (nous reprenons la même représentation des événements par exemple). Il n'en reste pas moins que tout système semblable fonctionnellement à *VESPA* pourrait servir de base à notre approche.

Le reste de ce chapitre est organisé comme suit. Dans la section 3.2 nous présentons les principes du système *VESPA* puis nous focalisons sur la représentation des événements utilisée dans *VESPA* et que nous reprenons dans nos travaux. Dans la section 3.3, nous identifions nos besoins en terme d'agrégation. Avant de conclure, nous présentons notre architecture globale dans la section 3.4.

3.2 *VESPA*

3.2.1 Principes de *VESPA*

VESPA : Vehicular Event Sharing with a mobile P2P Architecture, est un système permettant l'échange d'informations liées à des événements observés entre des véhicules par le biais d'une architecture pair-à-pair. L'originalité de *VESPA* est de permettre l'échange de tout type d'événement dans le réseau et de minimiser le nombre de messages échangés lors de la dissémination d'un événement. Par conséquent, *VESPA* propose un protocole de diffusion opti-

1. Une description détaillée est présentée dans <http://www.univ-valenciennes.fr/ROI/SID/tdelot/vespa/>

misé basé sur le concept de probabilité de rencontre, utilisé pour estimer la pertinence d'événements pour les véhicules [Delot 2008].

L'architecture de *VESPA* est présentée dans la figure 3.1 où nous pouvons voir les éléments principaux dans *VESPA*.

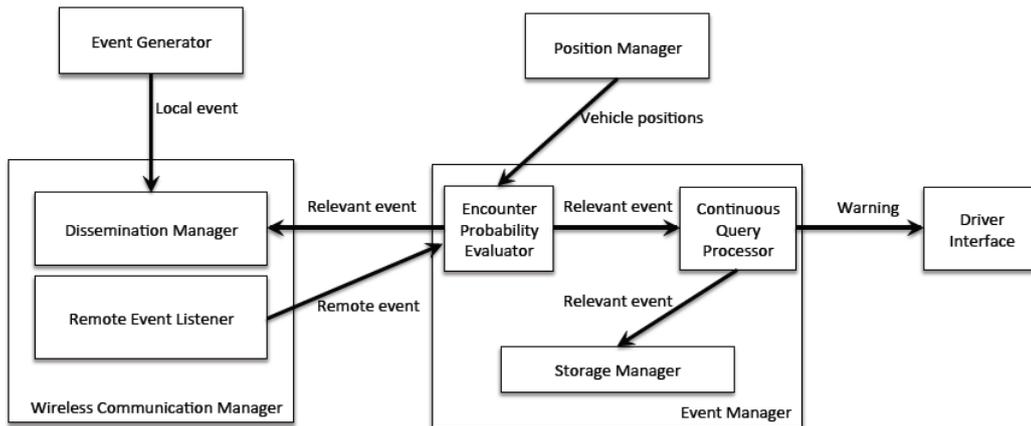
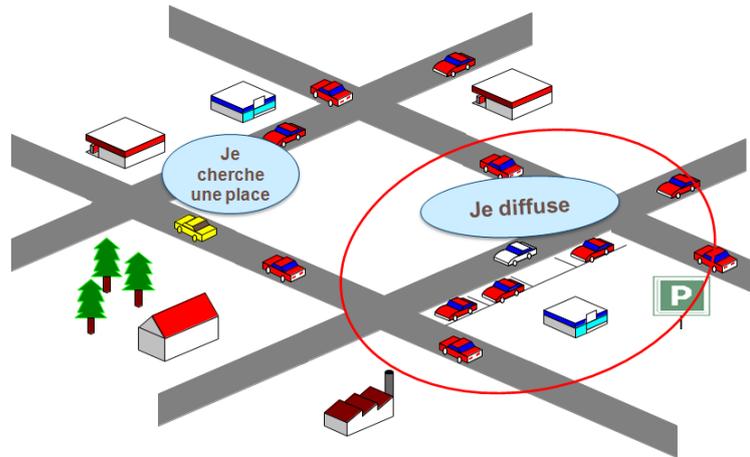


FIGURE 3.1 – l'architecture du *VESPA*

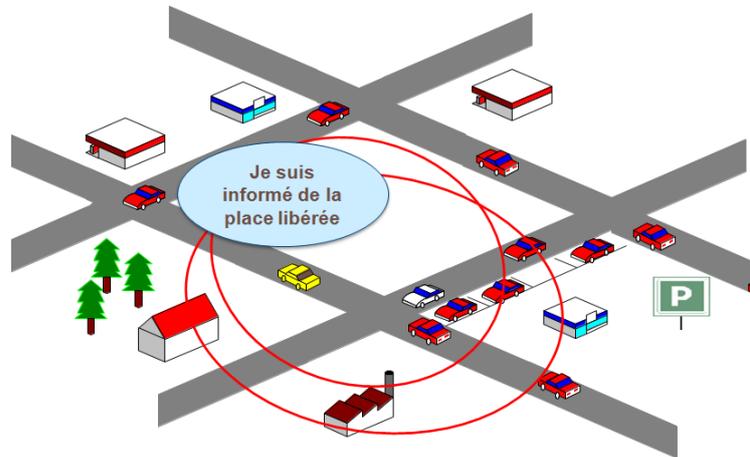
- **Wireless Communication Manager** s'occupe de la réception et de la transmission des événements. Ce module est composé du *Dissemination Manager* qui s'occupe de diffuser les événements et du *Remote Event Listener* qui écoute les événements transmis par les véhicules voisins. C'est donc ici que se fait la gestion de la dissémination. Un message est transmis aux véhicules circulant dans le voisinage de la place de parking pendant une période de temps limitée [Xu 2004a, Cenerario 2008]. La diffusion de messages entre les véhicules ne repose pas sur un serveur central mais s'appuie plutôt sur des échanges directs entre les véhicules suffisamment proches ;
- **Event Manager** traite les événements reçus par le véhicule. Il est composé du *Continuous Query Processor* qui exécute les requêtes continues actives répondant aux intérêts du conducteur. Une requête continue est une requête qui n'est pas évaluée une seule fois lors de son émission mais qui est évaluée de manière continue en mettant à jour les résultats en fonction des nouvelles données. Un exemple dans le domaine des véhicules est une requête de type « donnez moi tous les événements de type ac-

- cident observés dans un rayon de un kilomètre autour du véhicule ». A noter que cette requête, en plus d'être continue, est dépendante de la position du véhicule. L'*Event Manager* utilise un *Encounter Probability Evaluator*, et le *Storage Manager*, dont la tâche est de décider si les événements doivent être stockés ou effacés ;
- ***Driver Interface*** est une interface graphique permettant d'interagir avec le conducteur. Il s'agit donc de présenter les événements au conducteur, de lui offrir la possibilité d'activer des requêtes continues (par exemple lorsqu'il se met en recherche de place de stationnement) grâce à un écran tactile et une interface ergonomique ;
 - ***Position Manager*** est le lien entre le système et le module de positionnement, pour nous, le récepteur GPS. Le *Position Manager* sert donc à récupérer la position du véhicule ;
 - le ***Event Generator*** génère les événements détectés par le véhicule. De nombreux événements peuvent être générés en utilisant les différents capteurs présent dans le véhicule (par exemple, en couplant le déclenchement de l'airbag avec la création d'un événement représentant un accident). D'autres événements peuvent être générés grâce à des capteurs sur la route. Cela permet d'une part d'éviter un grand nombre d'interactions avec le conducteur et d'autre part d'éviter que les conducteurs ne créent des faux événements pour leur propre bénéfice. Par exemple, un conducteur peut créer une fausse libération de place de stationnement pour tromper les autres conducteurs en recherche et ainsi avoir une place de stationnement plus facilement.

Dans ce qui suit, nous allons donner l'exemple d'aide à la décision d'un véhicule cherchant une place de parking en utilisant le système *VESPA*. Dans *VESPA* et comme le montre la figure 3.2 un véhicule entre dans une zone en recherche de place de parking libre. A cet instant d'autres véhicules dans la même zone libèrent leurs places de parking et diffusent un message pour notifier les véhicules à proximité qu'il y a une place disponible. Ce message se propage à l'aide du mécanisme de dissémination de *VESPA* jusqu'à atteindre les véhicules en recherche d'une place. Une fois que l'information sur l'événement « libération de la place disponible » est utilisée, elle est détruite.



(a) Recherche d'une place de parking



(b) S'informer d'une place de parking par dissémination

FIGURE 3.2 – L'aide à la décision en utilisant *VESPA*

3.2.2 Classification et représentation des événements

Dans le domaine de la communication V2V (Vehicle To Vehicle), nous considérons deux sous-ensembles de types d'événements, à savoir « les événements immobiles » (Accidents, embouteillages, situations de freinage d'urgence, informations sur les prix des stations services, obstacles sur la chaussée, place de parking) et « les événements mobiles » (véhicules d'intervention d'urgence tels que les ambulances ou encore le comportement dangereux de certains conducteurs). Ces événements très diversifiés peuvent être classés selon d'autres critères (comme dans [Delot 2008] selon la mobilité et la dépendance

de la direction, etc.). Dans nos travaux nous allons adopter la classification donnée dans [Delot 2010]. Nous pouvons classer les informations utilisées dans un système d'aide aux conducteurs en deux catégories :

- *les alertes* : il s'agit des informations sur tous les dangers présents sur la route et que l'utilisateur risque de rencontrer (accident, zone dangereuse, embouteillage etc.). Ces données doivent être diffusées aux conducteurs en « *mode push* » et sont prioritaires pour assurer la sécurité du conducteur. Les réponses attendues pour aider le conducteur dans ces types d'événements, doivent être précises en terme de localisation ;
- *les ressources* : il s'agit des données associées à des objets recherchés par l'utilisateur tels que les places de stationnement. Ces données sont demandées par l'utilisateur donc récupérées en « *mode pull* ». Les réponses attendues peuvent être moins précises en terme de localisation.

Nous décrivons, dans la suite, comment ces différents événements sont représentés lors de leur génération afin qu'ils puissent être exploités par les véhicules :

- Tout d'abord, chaque événement doit être caractérisé par une « *Clef* » unique définie par la concaténation d'un identifiant unique de véhicule (par exemple, son adresse MAC) avec un identifiant local d'événement unique. Nous supposons qu'une instance d'un événement produit toujours le même identifiant sur le véhicule V_i et sur tous les autres véhicules. La génération d'un identifiant unique pour les événements "observés" par plusieurs véhicules (par exemple, les différents véhicules coincés dans un embouteillage) est un problème très difficile. Des idées intéressantes pour le résoudre ont été proposées dans le domaine de la fusion de l'information [Faouzi 2011, Kruijff 2006] ;
- Pour déterminer si un événement doit être présenté de manière urgente au conducteur, chaque événement possède une « *Priorité* ». Par exemple, si un véhicule reçoit un message indiquant un accident à quelques centaines de mètres plus loin, il est essentiel d'avertir le conducteur pour qu'il puisse agir de manière adéquate. La priorité associée à cet événement devrait donc être fixée au maximum. A l'inverse, les événements ayant une priorité faible ne devraient pas nécessairement être fournis au

- conducteur ;
- lors de sa génération, chaque événement possède aussi une date et une position en 3 dimensions. Ces informations constituent un vecteur à 4 dimensions représenté par l'attribut « *Position courante* » ;
 - Pour terminer, une « *Description* » permet de décrire plus précisément l'événement présenté (e.g. un accident, un freinage d'urgence, une place de stationnement disponible, etc.). Cet attribut permet de définir concrètement l'information qui doit être transmise au conducteur au moment où il doit être prévenu.

Les attributs utilisés pour décrire les événements sont résumés dans le tableau 3.1. Les positions données par l'attribut de type PositionAndTime correspondent à des relevés GPS incluant la position en 3 dimensions ainsi que l'information sur le temps GPS du relevé. L'utilisation du temps GPS nous évite les problèmes liés à la synchronisation des horloges de chaque véhicule.

Nom de l'attribut	Abréviation	Type
Clef unique	Key	String
Importance	I	Int
Position courante	CurrentPos	PositionAndTime
Description	Desc	EventDescription

TABLE 3.1 – Représentation des événements

Le tableau 3.2 présente un exemple de message échangé entre les véhicules pour annoncer un espace de stationnement disponible. Ce message contient un identifiant unique, une priorité, la position et une description de la ressource considérée.

Identifiant	Priorité	Position& Date	Description
2011300910251750191591N305111EAD	<i>low</i>	50°19'15.91 N 3°30'51.11 E 10h25m17s 2011-30-09	Libération d'une place de parking

TABLE 3.2 – Exemple de message généré pour annoncer la libération d'une place de parking

Habituellement, les événements diffusés dans le réseau des véhicules ont une durée de vie assez courte, allant de quelques secondes à quelques heures

selon le type d'événement. Dans *VESPA*, une fois les données reçues par un véhicule, leur pertinence est évaluée, généralement sur des critères spatiaux et temporels. Le cas échéant, le conducteur est informé de la présence d'un événement. Aucune autre exploitation des données n'a été investiguée. Une fois les données utilisées, elles sont considérées obsolètes et détruites.

3.3 Pourquoi agréger ?

Même si les systèmes de communication inter-véhicules de type *VESPA* peuvent fournir une assistance efficace au conducteur dans de nombreux cas, ils souffrent d'un inconvénient majeur. En effet, une fois que les événements sont utilisés, ils sont considérés comme obsolètes et détruits. Cependant, lorsque aucune information n'a été récemment produite par le véhicule ou les véhicules à proximité, ces systèmes ne peuvent pas aider le conducteur à prendre une décision sur la route.

Dans ce qui suit nous prenons le même exemple « recherche de la ressource : place de parking » que nous avons décrit dans la section précédente 3.2. Un véhicule entre dans une zone et il est en recherche d'une place de parking. A cet instant il n'y a aucun message disseminé par un ou plusieurs véhicules libérant leurs places en plus il n'y a aucune place libre visible par le conducteur. Dans ce cas en se basant sur le protocole de dissémination à la *VESPA* et la vision directe, le conducteur peut ne pas trouver une place pour se garer et il doit continuer à chercher (le cas échéant longtemps).

Face à cette situation un conducteur peut avoir recours à l'infrastructure et se connecte à un serveur pour répondre à sa demande. Dans notre exemple « recherche d'une place de parking » il peut accéder au serveur « *VINCI Park* »². En plus des problèmes dus à la centralisation de l'information (difficulté d'accès aux serveurs, temps de réponse, couverture spatiale limitée, etc.) l'opérateur VINCI n'assure pas la couverture de toutes les places de parking de toutes les villes.

2. Pour plus d'informations, vous pouvez visiter http://www.vincipark.com/vincipark.nsf/fr/les_services/pages/vincipark.com.htm

Dans un environnement routier, une information très pertinente que l'utilisateur veut recevoir durant son trajet c'est « l'alerte » dans le cas où il y a un accident, embouteillage, freinage d'urgence ou les zones dangereuses. Ces informations peuvent être fournies par le système « *COYOTE* »³ cependant ce système est limité aux alertes et ne prend pas en compte des événements de type ressource. Il complète les approches de type infrastructure (type VINCI) en ajoutant à l'envoi de requêtes d'un véhicule vers le serveur central, l'acquisition d'informations provenant des véhicules et à destination du serveur central. Dans les deux cas cependant, il n'y a pas de communication inter-véhicules et donc l'accès à l'information dépend de la connection avec le serveur central. Ces approches favorisent la qualité de l'information obtenue grâce à la vision globale, mais ne garantissent pas la couverture spatio-temporelle observée et surtout ne permettent pas de fonctionner en mode déconnecté.

Pour pallier les limitations dues aux systèmes basés sur la dissémination et aux systèmes construits sur un serveur central, nous proposons d'étendre les systèmes basés sur la dissémination en ajoutant un mécanisme de construction collaboratif de résumés des événements.

Nous proposons donc que les données reçues par un véhicule ne servent pas uniquement pour produire un message d'alerte à destination du conducteur. Une fois stockées sur un véhicule, il est en effet possible d'utiliser a posteriori ces données collectées pour produire, au niveau du véhicule, des connaissances sur l'environnement exploitables par le conducteur. A titre d'exemple, les événements concernant les places de stationnement disponibles reçues par un véhicule peuvent être exploités, lorsqu'il n'y a aucune place disponible diffusée par les autres véhicules, pour déterminer l'endroit où la probabilité de trouver une place libre est la plus importante (en fonction du jour et de l'heure par exemple). Dans un autre contexte, grâce à la corrélation des différents messages reçus sur les accidents et les freinages d'urgence, les zones dangereuses peuvent être dynamiquement détectées et indiquées au conducteur, qu'elles soient d'ailleurs continuellement dangereuses ou seulement temporairement du fait des conditions climatiques par exemple. En outre nous cherchons dans nos travaux à fournir une solution complète en terme d'événements qui peuvent se

3. Pour plus d'informations, vous pouvez visiter <http://www.moncoyote.com/fr/comment-ca-marche>

produire sur la route : les ressources comme les places de parking et les alertes telles que les embouteillages, freinages d'urgence, accidents, zones dangereuses, etc.

3.4 Architecture globale

Dans ce qui suit, nous décrivons l'architecture de notre système que nous avons modélisé dans la figure 3.3. Notre système d'aide à la décision est composé de trois parties principales :

1. *Construction de résumé* où chaque véhicule peut obtenir de l'information de multiples sources : capteurs embarqués sur le véhicule, événements obtenus par dissémination à partir des véhicules voisins, informations obtenues par une infrastructure. L'objectif ici est de construire des résumés de tous les événements perçus via les différentes sources, de manière à pouvoir estimer des probabilités d'apparition d'une occurrence d'événement dans une zone spatio-temporelle donnée. Comme chaque véhicule ne peut percevoir qu'une fraction limitée des événements produits, une approche de construction collaborative basée sur un échange des résumés entre les véhicules est proposée ;
2. *Niveau d'accès au résumé* où le résumé peut être vu comme une base de données spatio-temporelle stockant des agrégations d'événements relativement à un découpage spatio-temporel. Ces agrégats peuvent être interrogés par un langage de requêtes ;
3. *Niveau d'assistance au conducteur* où les réponses aux requêtes du niveau précédent, vont ensuite être interprétées pour fournir des aides à la décision aux conducteurs.

Le processus d'agrégation doit avoir les propriétés suivantes :

1. Résumer les événements en favorisant les dimensions fondamentales que sont la localisation et le temps. En effet, les attributs utilisés pour décrire les événements doivent contenir la position et le temps de la génération de l'événement agrégé ;
2. il doit être incrémental (croissant en nombre d'événements stockés) avec

un volume de données stockées faible (de l'ordre de quelques centaines de Mo allant jusqu'à quelques Go) ;

3. chaque conducteur doit pouvoir choisir à quels types d'événements il s'intéresse et avec quelles échelles spatiale et temporelle il veut agréger ;
4. les données agrégées doivent pouvoir être partagées entre les véhicules pour enrichir leurs connaissances respectives. En outre des messages sur les événements échangés entre véhicules, les résumés construits peuvent être aussi disseminés entre les véhicules pour des résumés plus riches et une base de connaissance locale plus complète ;
5. il doit pouvoir s'adapter à tous les types d'événements ;
6. le processus d'agrégation doit être peu coûteux (coût stockage, coût d'accès aux données pour l'exploitation).

3.5 Conclusion

Nous avons présenté dans ce chapitre le système *VESPA* comme un exemple typique de système basé sur la dissémination des événements sur lequel nous pouvons construire nos contributions. Nous avons ensuite montré- via des exemples de recherche de ressources ou de production d'alertes- que les systèmes actuels, soit basés sur la dissémination (*VESPA*), soit basés sur une infrastructure centralisée (*VINCI*, *COYOTTE*) ne peuvent fournir une assistance efficace au conducteur dans de nombreux cas.

La première catégorie de systèmes ne peut fournir d'aide en l'absence d'événements « actifs », alors que la seconde catégorie ne peut fonctionner en mode déconnecté et n'assure pas la couverture tant en type d'événements, qu'en espace spatio-temporel.

L'objectif de nos recherches est d'aller plus loin en utilisant à posteriori les données collectées pour estimer la probabilité d'occurrence d'un type d'événement donné en l'absence d'observation. Le résumé dans notre travail n'est plus « un moyen de compression » comme dans les travaux décrits dans le chapitre 2. Il a comme but « d'agréger les événements passés de manière à disposer d'une base de connaissance permettant d'estimer si un événement est susceptible de survenir même sans observation ».

Nous avons présenté notre architecture globale du système d'aide à la décision basé sur les résumés qui comporte trois modules : 1- construction du résumé qui fera l'objectif des chapitres 4 et 5, 2- interrogation du résumé et 3- interprétation des résultats qui seront détaillés dans le chapitre 6.

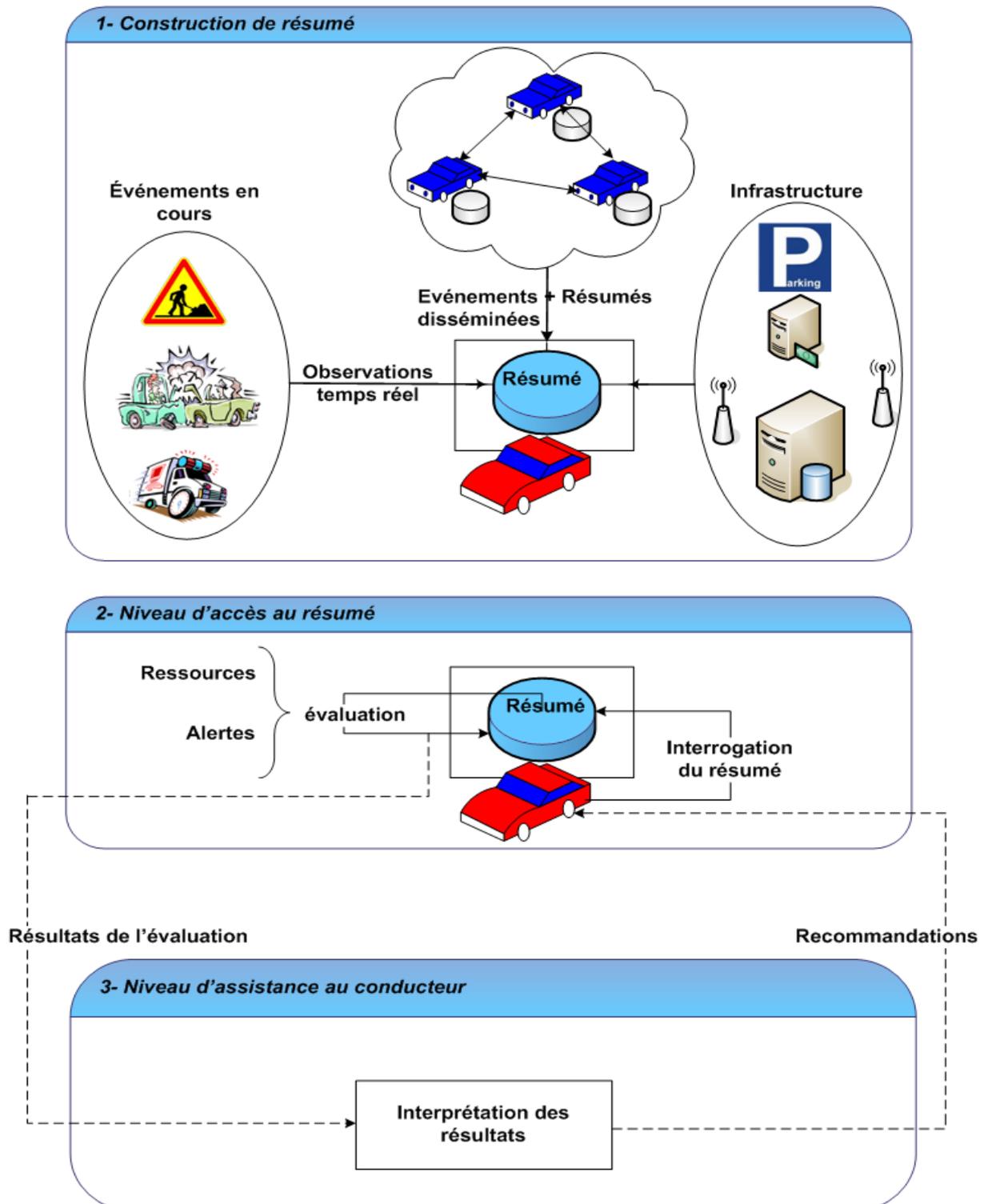


FIGURE 3.3 – L'architecture globale

Construction du résumé

Sommaire

4.1	Introduction	41
4.2	Modèle spatio-temporel à deux niveaux	42
4.2.1	Niveau physique	44
4.2.2	Niveau logique	45
4.3	Construction des résumés des événements	47
4.3.1	Les Sketches	48
4.3.2	Structure d'agrégation	52
4.3.3	Evaluation théorique de la taille de la structure d'agrégation	55
4.4	Conclusion	57

4.1 Introduction

Partant des solutions d'agrégation décrites dans le chapitre 2, nous avons noté qu'une première catégorie des travaux tels que [Eichler 2006, Saleet 2007] considère l'agrégation dans les échanges inter-véhicules comme un moyen de compression des données afin de réduire les besoins en bande passante. L'agrégation est vue par une deuxième catégorie de solutions comme un moyen pour construire un résumé de données citons [Nadeem 2004, Cheng 2003, Lochert 2007]. Toutefois, la forte mobilité et la concentration importante des véhicules dans certaines zones géographiques rendent très difficile l'utilisation de ces solutions qui présentent encore soit des lacunes à combler comme

dans [Fabio 2006] où la solution d'agrégation ne supporte pas tous les types d'événements, soit des inconvénients à éviter dans la construction de résumé par exemple dans [Yufei 2004] l'utilisation des arbres de nœuds qui sont coûteux en mémoire et difficiles à maintenir du fait de la forte mobilité des véhicules.

Dans nos travaux, nous proposons une structure d'agrégation qui supporte plusieurs types d'événements se produisant sur la route et qui répond aux besoins du processus d'agrégation exprimés dans le chapitre précédent 3. Le principe général de la contribution est de s'appuyer sur un modèle spatio-temporel que nous avons défini pour agréger les événements. Ce modèle permet à chaque conducteur de choisir ses zones d'intérêts. Une fois le modèle spatio-temporel défini, nous avons choisi une méthode pour agréger le maximum d'événements détectés soit par observation directe, soit par échange avec d'autres véhicules. Cette solution d'agrégation doit pouvoir détecter les doublons. Autrement dit, elle devrait ne pas comptabiliser plusieurs fois une même occurrence d'événements observés par deux véhicules différents. Pour ce faire, nous avons utilisé la solution de Flajolet et Martin [Flajolet 1985] qui se base sur des sketches (une suite de bit à 0 et 1) construits par l'application d'une fonction de hachage sur l'identifiant de l'événement observé.

Le reste de ce chapitre est organisé comme suit. Nous allons d'abord décrire dans la section 4.2 le modèle spatio-temporel à deux niveaux utilisé pour construire le résumé d'événements. Ensuite, en s'appuyant sur notre modèle ainsi défini, nous modélisons notre structure d'agrégation dans la section 4.3. Finalement, nous concluons dans la section 4.4.

4.2 Modèle spatio-temporel à deux niveaux

Chaque véhicule doit conserver, en plus des événements actifs reçus (c'est à dire ceux dont la durée de vie n'est pas encore terminée), un résumé des événements passés. Pour répondre aux besoins du processus d'agrégation exprimés dans la section 3.4, nous proposons un modèle spatio-temporel à deux niveaux. Il faut à la fois permettre à chaque conducteur de choisir ses propres zones d'intérêt et d'échanger un résumé entre deux véhicules sans perte d'in-

formations. La figure 4.1 donne un exemple de perte d'information suite à un échange entre deux véhicules.

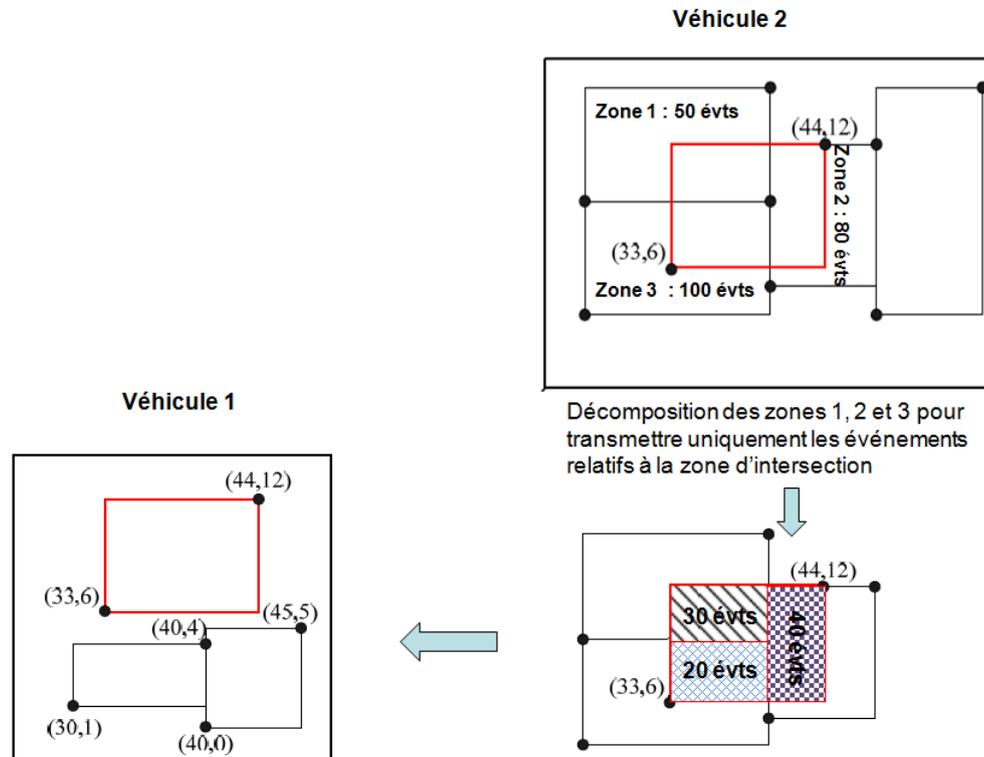


FIGURE 4.1 – Echange des connaissances entre les deux véhicules

Le véhicule 1 veut échanger avec le véhicule 2. Chaque véhicule a choisi son propre découpage de l'espace spatio-temporel pour agréger les événements. Pour être capable d'agréger les deux résumés à la suite d'un échange, il va falloir calculer l'intersection spatio-temporelle de leurs zones d'intérêt. Comme chaque véhicule dispose de son propre découpage, l'intersection spatio-temporelle va amener à découper les zones d'intérêt en sous-zones communes. Le problème dans ce cas là, est que nous ne savons pas précisément comment répartir la valeur agrégée du nombre d'événements observés dans les sous-zones. Dans notre exemple, zone 1 est découpée en deux sous-zones mais nous ne pouvons précisément répartir les 50 événements observés entre les deux sous-zones (nous pouvons juste approximer en supposant une distribution uniforme par exemple).

Pour prendre en compte ces deux besoins, nous proposons un modèle

spatio-temporel à deux niveaux :

1. « *le niveau physique* » qui représente un référentiel commun pour tous les véhicules et qui va servir à stocker les valeurs agrégés et va être utilisé comme référentiel commun entre les véhicules lors des échanges ;
2. « *le niveau logique* » choisi par chaque conducteur en se basant sur le niveau physique.

4.2.1 Niveau physique

Le niveau le plus bas, appelé niveau physique, constitue un référentiel partagé entre tous les véhicules. Il est découpé en carrés de taille fixe qui forment une partition complète de l'espace et uniforme pour tous les véhicules comme l'illustre la figure 4.2 sur une partie de la ville de Paris. Chaque carré représente une cellule physique.

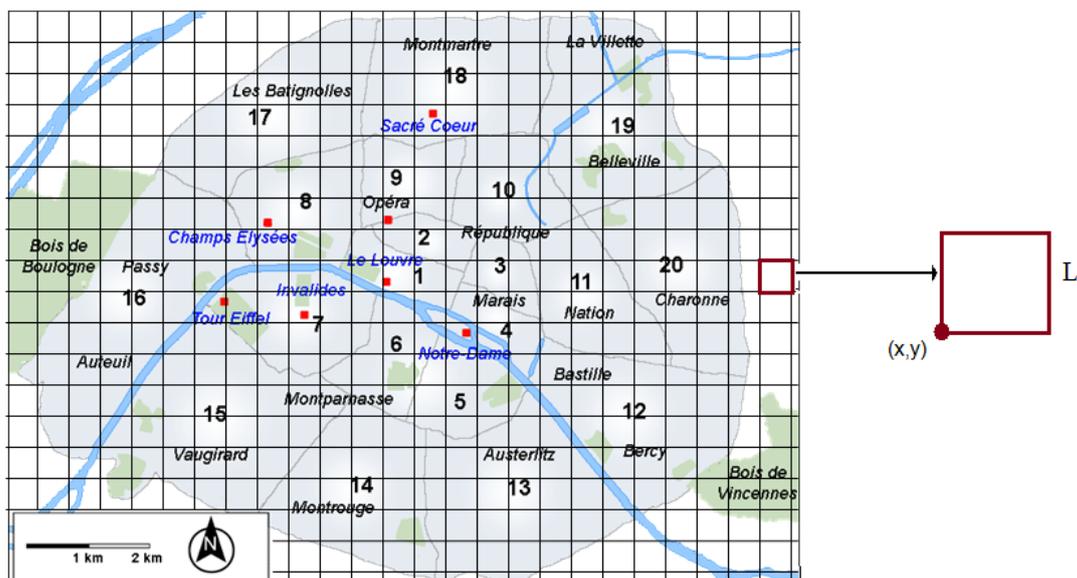


FIGURE 4.2 – Découpage spatial au niveau physique (portion de Paris)

Chaque cellule physique est identifiée par les coordonnées (X, Y) du point bas gauche et la longueur L du côté (figure 4.2). Pour représenter les coordonnées des cellules physiques dans notre modèle spatial à deux niveaux,

nous utilisons un système de coordonnées cartésiennes. Ces coordonnées cartésiennes sont obtenues par projection des coordonnées obtenues par le GPS décrit dans la section 3.2.2 dans le chapitre 3.

Nous procédons de même pour la dimension temporelle : le temps est découpé en segments qui forment une partition complète. Nous utilisons le découpage de l'année en 52 semaines et chaque semaine en 7 jours. Nous divisons chaque jour en tranche de 2 heures, soient au total 84 segments temporels par semaine. Ce découpage temporel permet de mettre en évidence des phénomènes « saisonniers » (plus de places libres de parking le week end dans un quartier d'affaires par exemple).

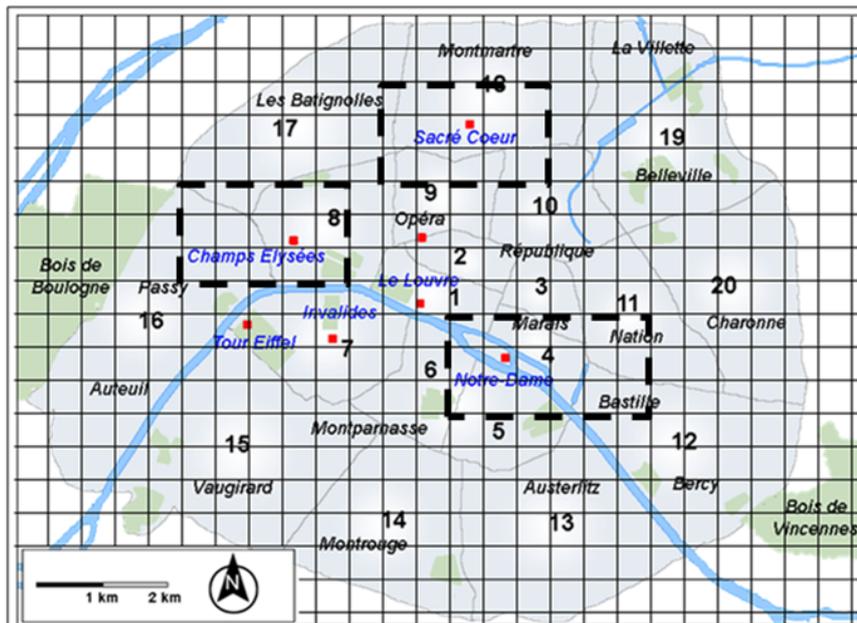
Le couple {cellule physique, segment temporel} est la plus petite unité pouvant comptabiliser des occurrences d'événements. Cet espace physique est très grand : si nous supposons que la taille d'une cellule spatiale est de $1km^2$ et que nous avons 10 segments temporels, la couverture de la France serait de l'ordre de 6 millions de couples. Ce nombre pourrait être réduit si nous structurons l'espace en zones de tailles non fixes, ce qui permettrait d'avoir une résolution spatiale meilleure dans les zones urbaines (et donc une meilleure précision). Dans la suite nous considérons la cellule physique de côté $L = 200m$. Avec une telle borne un véhicule peut avoir une vision globale d'une cellule physique. Aussi, plus nous diminuons la taille de la cellule physique plus nous augmentons la précision du résumé associé.

4.2.2 Niveau logique

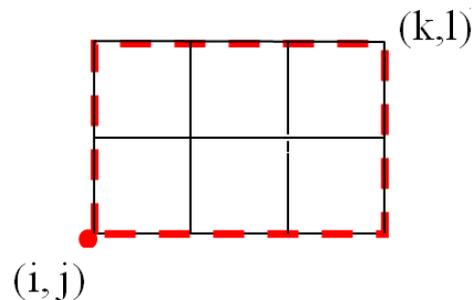
En se basant sur le niveau physique, chaque véhicule peut construire son propre découpage logique représentant les zones spatiales. Ce niveau logique est constitué d'un ensemble de rectangles eux-mêmes constitués d'un ensemble de cellules du niveau physique (les rectangles pointillés sur la figure 4.3(a)).

Une zone d'intérêt est définie comme un rectangle construit sur un nombre entier de cellules physiques (Figure 4.3(b)). Elle est définie par un couple de cellules physiques : bas gauche de coordonnées (i, j) et haut droit de coordonnées (k, l) .

Nous supposons que nous avons g granularités temporelles basées sur le



(a) Découpage spatial pour le niveau logique



(b) Zone d'intérêt (zone logique)

FIGURE 4.3 – Niveau logique

découpage temporel au niveau physique (ordonnées de lundi à dimanche et chaque jour commence par la période de 0h à 2h qui représente la première granularité jusqu'à la $g^{\text{ème}}$ granularité de 22h à 24h). Un conducteur peut ne pas s'intéresser à tout l'espace et le nombre de cellules physiques réellement observées au niveau logique est donc très petit relativement au niveau physique. Par exemple, si un conducteur s'intéresse à une centaine de zones spatiales couvrant en moyenne 20km^2 , le nombre de couples sera de l'ordre

de 2000.

Nous posons les notations suivantes utilisées dans la représentation de la structure d'agrégation à la section 4.3.2. L'espace est découpé en carrés de côté C_{cp} de taille L . L'environnement comporte N cellules physiques sur l'axe des X et P sur l'axe des Y . Le point origine des deux axes a pour coordonnées $(X_{origine}, Y_{origine})$.

4.3 Construction des résumés des événements

Le deuxième critère du processus d'agrégation, cité dans la section 3.2.2 du chapitre 3, exige une représentation compacte de la structure d'agrégation pour limiter l'espace mémoire occupé. Il implique aussi que le mécanisme d'agrégation détecte les doublons. Par conséquent, il doit reconnaître un même événement observé par deux véhicules différents afin de ne pas le considérer comme deux événements différents à regrouper.

Pour atteindre ces objectifs, nous nous appuyons sur les deux niveaux spatio-temporels du modèle que nous avons proposé dans la section 4.2. Le résumé doit permettre d'estimer la fréquence d'apparition d'un type événement donné dans une zone spatio-temporelle donnée. Pour cela, nous proposons d'estimer d'une part le nombre d'occurrences d'un type d'événement et d'autre part la durée d'observation. Le ratio entre ces deux valeurs nous estime la fréquence. Cela revient donc à maintenir deux informations par cellule physique. Pour estimer la durée d'observation, nous avons choisi de travailler au niveau des semaines ce qui offre un compromis raisonnable entre la précision et la place occupée. Pour cela, nous supposons que tous les véhicules sont synchronisés sur la même semaine de départ (numérotée 1) et ensuite maintiennent un simple compteur du numéro de semaine. Pour détecter les doublons, nous choisissons parmi les sketches étudiés dans la section 4.3.1 les sketches de Flajolet-Martin que nous introduisons dans la section 4.3.1.3. Ensuite, nous détaillons notre structure d'agrégation des événements dans la section 4.3.2. La taille de notre structure d'agrégation sera évaluée dans la section 4.3.3.

4.3.1 Les Sketches

Ce sont Alon Matias et Szegedy [Noga 1996] qui ont introduit la notion de sketch aléatoire pour la première fois. Ce sketch aléatoire est un résumé de flux de données très compact utilisé afin d'estimer la réponse à certaines requêtes à travers l'ensemble des données du flux. Cette technique vise à projeter aléatoirement chaque élément dans un espace particulier en utilisant des fonctions de hachage et de ne garder que les composantes les plus pertinentes, minimisant ainsi de l'espace tout en préservant l'essentiel de l'information. Plusieurs types de sketches peuvent être appliqués pour agréger des événements. Nous citons *le filtre de Bloom*, *le Count sketch* et *le Flajolet Martin sketch*.

4.3.1.1 Count sketch

Ce sketch, proposé par [Moses 2002], permet de répondre aux requêtes de type 'top-k', c'est à dire trouver la liste des k éléments les plus fréquents dans le flux de données respectant la contrainte $n_i > (1 - \varepsilon)n_k$ où n_i est la fréquence d'un élément donné d'indice i dans la liste des éléments classés par fréquence et ε la précision souhaitée. Le sketch est représenté sous la forme d'une matrice de comptage de taille $t \times b$. Nous disposons de t fonctions de hachage s_i de M dans $\{-1,1\}$ et de t fonctions de hachage h_i de M dans $[1,b]$. A l'arrivée d'un nouvel élément e , nous mettons à jour la matrice par

$$\forall i = 1, \dots, t, h_i(e) + = s_i(e) \quad (4.1)$$

Nous estimons la fréquence d'un élément donné e en évaluant la médiane de $h_i(e) \times s_i(e)$. Enfin, pour obtenir les 'top-k' éléments, nous trions incrémentalement les fréquences des éléments et nous gardons dans une liste celles qui respectent la contrainte. Ainsi, chaque nouvel élément est inséré dans le sketch, s'il se trouve par ailleurs dans la liste, sa fréquence dans la liste est mise à jour et la liste est triée. S'il n'est pas dans la liste, sa fréquence est estimée à l'aide du sketch. Si cette dernière est plus grande que celle de l'élément de plus petite fréquence présent dans la liste, elle est ajoutée.

4.3.1.2 Filtre de Bloom

Un filtre de Bloom [H.Bloom 1970] est un tableau de bits qui permet de tester d'une manière rapide l'appartenance d'un élément à un ensemble d'éléments. Le filtre de Bloom comporte deux composants : (1) un ensemble de k fonctions de hachage et (2) un vecteur de bits d'une taille L donnée. Toutes les fonctions de hachage sont configurées de telle sorte que leurs intervalles correspondent à la taille du vecteur. Par exemple, si le vecteur de bits est de taille 200, alors toutes les fonctions de hachage doivent retourner une adresse entre 1 et 200. Les fonctions de hachage garantissent que les adresses générées sont réparties de façon équiprobable sur toutes les valeurs possibles. Pour introduire un terme dans un filtre de Bloom, nous calculons les valeurs des fonctions de hachage et nous mettons à 1 les bits du vecteur correspondant. Pour tester l'appartenance d'un élément x à un ensemble M introduit dans le filtre de Bloom, nous lui appliquons les fonctions de hachage. Si au moins un des bits est à 0 alors x n'appartient pas à M . Par contre, si tous les bits sont à 1 alors x appartient probablement à M avec un taux de faux positif moyen donné par la relation suivante :

$$f = (1 - e^{kM/L})^k \quad (4.2)$$

où L est la taille du vecteur du filtre de Bloom, M le nombre de termes indexés et k le nombre de fonctions de hachage. Pour minimiser ce taux, nous choisissons un nombre de fonctions de hachage k respectant la relation suivante :

$$k = \ln 2(L/M) \quad (4.3)$$

Un certain nombre d'améliorations ont été développées autour des Filtres de Bloom afin de palier aux limites de ces résumés. *Le Counting Bloom Filter* permet de gérer la suppression d'un élément du sketch. Cela le rend applicable aux fenêtres glissantes où des éléments ne faisant pas partie de la fenêtre sont considérés comme expirés et ne doivent plus faire partie du sketch. Les Scalable Bloom Filters [Almeida 2007] permettent de construire des filtres de Bloom

dont la taille peut augmenter de façon dynamique en garantissant un taux d'erreur maximal fixe.

4.3.1.3 Flajolet-Martin sketch

Le sketch de Flajolet et Martin (FM sketch) [Flajolet 1985] permet d'estimer le nombre d'éléments distincts dans un flux de données. Il a été conçu en 1985 pour résumer des données massives. Il fournit une représentation compacte pour estimer le nombre d'occurrences d'événements distincts. Ce sketch a le mérite d'être incrémental en terme de construction et rapide en manipulation, ce qui le rend utilisable pour la gestion des flux de données mais ses performances en qualité d'estimation se dégradent avec le nombre d'éléments différents insérés. Toutefois, les bornes permettent de fixer les paramètres en fonction du nombre d'occurrences attendues.

L'ajout d'un élément dans ce sketch se fait comme suit : soient un tableau binaire représentant le sketch S_1 de taille L initialisé à 0 et une fonction de hachage $h(X)$ qui répartit uniformément les éléments d'un ensemble M dans l'ensemble des mots binaires de taille L . Soit $\rho f \rho(h(X))$ la position de 1 de plus faible poids dans la représentation binaire de X suite à l'application de la fonction de hachage. Le bit d'indice $\rho f \rho(h(X))$ est alors mis à 1 dans le sketch s'il était encore à 0. La figure 4.4 illustre l'insertion d'un événement X_1 dans le sketch. X_1 peut être la libération d'une place de stationnement ou un accident par exemple.

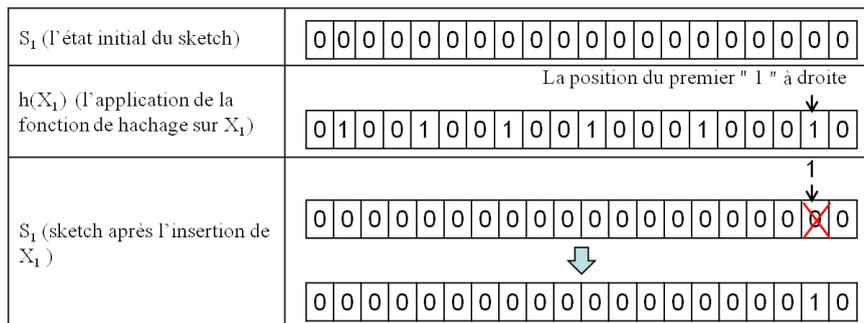


FIGURE 4.4 – Insertion d'un événement dans le sketch d'événement

Une fois le sketch construit, le nombre de valeurs distinctes dans M est estimé par p la position du 0 de plus faible poids dans le tableau binaire. Le nombre d'objets distincts est estimé par $n = 1,29 \times 2^k$ (avec k la position du premier bit dans le sketch qui est resté égal à 0). Nous pouvons prouver que $E(p) = \log_2(\phi n)$ avec $\phi = 0,77351$ et $\phi(r) = 1,12$. Pour plus de détails sur l'obtention de ces bornes, le lecteur peut se référer à [Flajolet 1985].

Nous partons du sketch S_1 , après insertion de n éléments, données par la figure 4.5. Nous appliquons la fonction d'estimation $n = 1,29 \times 2^k$ avec la position du premier bit dans le sketch qui est resté égal à 0 (dans l'exemple $k=8$) nous estimons un nombre d'éléments presque égal à 330.

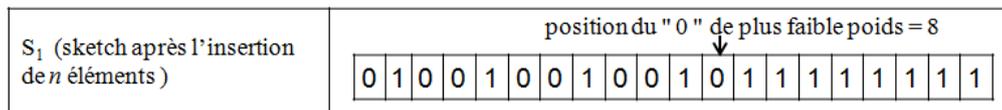


FIGURE 4.5 – Estimation de nombre d'événements dans le sketch S_1

Un tel sketch est intéressant car il détecte un double exemplaire du même événement par construction. En effet, deux instances d'un même événement ont la même image calculée par la fonction de hachage.

4.3.1.4 Synthèse

Parmi les techniques déjà citées, nous avons choisi la technique de Flajolet et Martin pour agréger les identifiants des événements. En effet, les sketches de Flajolet et Martin permettent de détecter les duplicats. Par conséquent, lorsque deux véhicules observent le même événement la technique du sketch de Flajolet et Martin l'insère une seule fois contrairement à *count sketch* qui cherche essentiellement à faciliter le comptage des éléments stockés et incrémente la fréquence d'un élément existant dans le sketch. Un filtre de Bloom cherche à vérifier si un élément appartient ou non à un ensemble d'événements et est donc capable de détecter les duplicats mais avec un taux d'erreur important.

4.3.2 Structure d'agrégation

Nous supposons que chaque véhicule V peut observer directement un ensemble d'événements E . Un événement « e » de E est caractérisé par les informations suivantes (ce sont les seules prises en compte pour le résumé, mais d'autres informations peuvent être utiles pour la gestion d'alertes ou la dissémination des messages) :

- ty_e : il s'agit du type d'événement observé (accident, libération place de parking, etc.) ;
- lo_e : il s'agit de la localisation géographique de l'événement ainsi que son estampille temporelle (le jour et l'heure de production de l'événement). Ces informations sont celles classiquement données par un système de positionnement de type GPS ;
- idf_e : identifiant unique de l'occurrence d'événement observé. Celui-ci est la base de la détection de dupliquat. Nous supposons qu'une occurrence d'événement produit toujours le même identifiant sur le véhicule V mais aussi sur les autres.

Nous proposons de décrire les résumés par la structure de données décrite dans la Figure 4.6. Celle-ci est constituée de :

- un ensemble de zones d'intérêts définies chacune par un identifiant unique (idf_z). Chaque zone est délimitée par deux cellules physiques (i, j, k et l). Le tableau des zones d'intérêt est trié selon les valeurs croissantes de i ;
- p pointeurs sur une liste chaînée qui a N éléments avec N égal au nombre de cellules physiques dans la zone d'intérêt. Chaque élément de la liste chaînée est caractérisé par :
 1. Une cellule physique (X_1, Y_1) ;
 2. Une liste de 7 pointeurs sur les jours de la semaine. Chaque élément de la liste des jours est caractérisé par :
 - un premier pointeur sur un sketch de type Flajolet-Martin agrégeant le nombre de semaines d'observation pour les événements

agrégés. Le sketch est construit en appliquant une fonction de hachage sur le numéro de la semaine. Si sketch détecte les doublons de semaine puisque la fonction de hachage utilisée est le numéro de semaine qui est unique. La fonction de hachage choisie sera présentée dans le chapitre 7 ;

- un deuxième pointeur sur g sketches de type Flajolet-Martin agrégeant les événements, de type ty_e donné, observés dans la cellule pour une granularité temporelle donnée. Le sketch est construit en appliquant une fonction de hachage sur l'identifiant idf_e de chaque événement observé.

Les opérations suivantes sont définies sur cette structure de données et se rapportent à un véhicule :

- Ajouter une zone d'intérêt : revient à créer une nouvelle entrée dans la table des zones ;
- Supprimer une zone d'intérêt : revient à supprimer l'entrée correspondante dans la table des zones ainsi que les listes chaînées associées ;
- Ajouter un événement observé : il faut retrouver la (ou les) zone(s) d'intérêt correspondant à la localisation de l'événement. S'il y en a (au moins) une, nous recherchons la cellule physique associée et le couple de sketches correspondant est mis à jour.

La structure d'agrégation proposée est définie pour optimiser l'accès aux résumés à partir de la dimension spatiale (zones d'intérêt). Nous considérons en effet que tout accès à un résumé se fait en utilisant tout d'abord le critère spatial. La recherche d'une zone est logarithmique (recherche dans une liste triée). L'accès relativement au temps est ensuite séquentiel (mais la taille est réduite).

Pour augmenter la précision de l'estimation, nous pouvons appliquer m fonctions de hachage pour produire m sketches distincts. Afin de minimiser le coût, nous pouvons choisir aléatoirement une fonction de hachage à appliquer sur chaque identifiant d'événement. Dans ce cas, le nombre d'événements sera évalué par la somme retournée par chaque sketch. L'erreur-type est en $O(m^{-1/2})$, donc avec $m = 4$, nous obtenons une bonne précision.

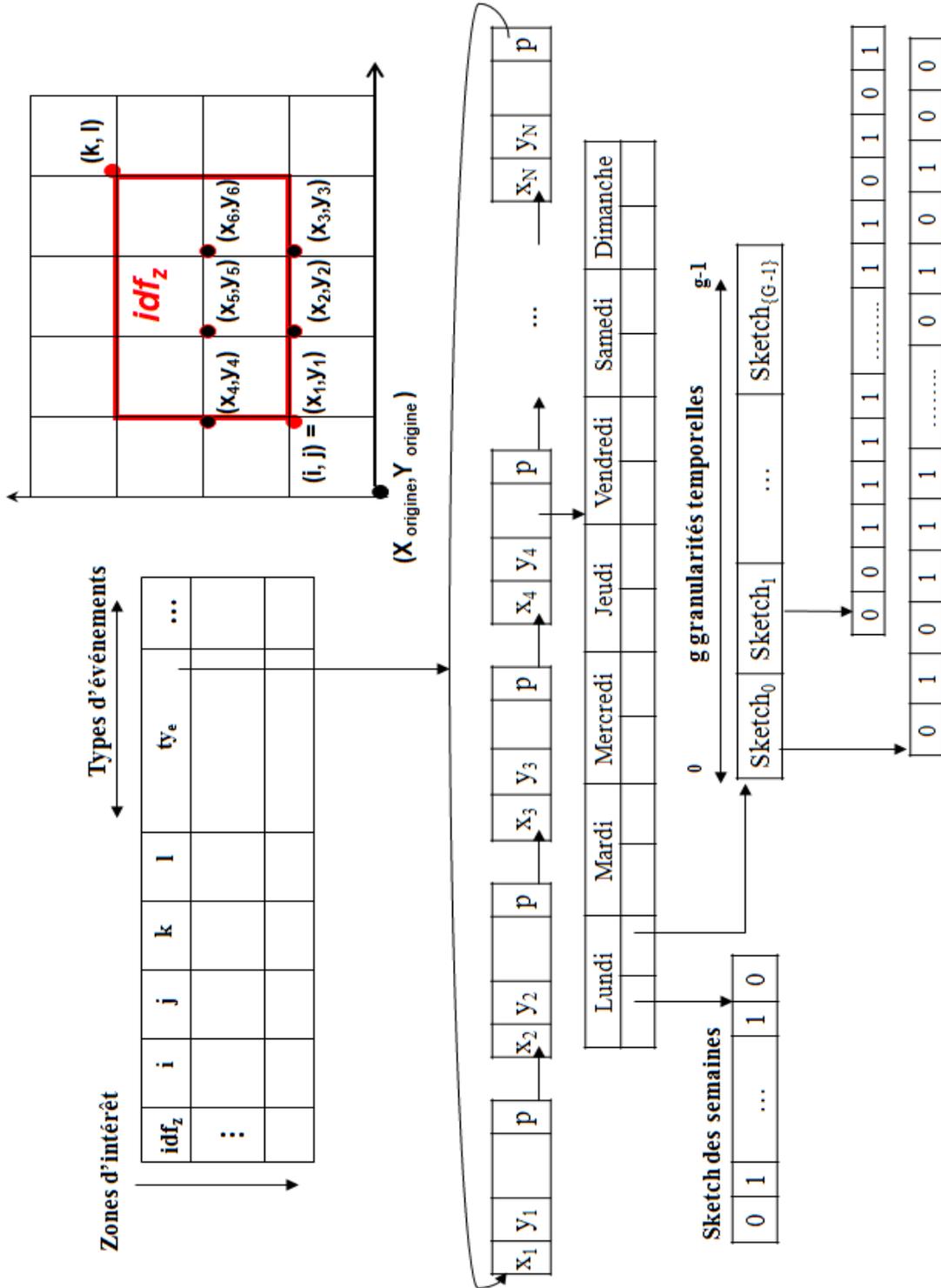


FIGURE 4.6 – Structure d'agrégation

4.3.3 Evaluation théorique de la taille de la structure d'agrégation

Dans ce qui suit, comme une première évaluation de notre structure, nous allons évaluer théoriquement sa taille pour deux types d'environnement :

1. Une ville (Paris) qui s'étend sur $\simeq 11km$ du nord au sud constituée dans notre référentiel spatial de $\simeq 3025$ cellules physiques dont chaque cellule est de côté $200m$;
2. Un pays (la France) qui s'étend sur $1000km$ du nord au sud et d'est en ouest, donc constitué dans notre référentiel spatial de $25.000.000$ cellules physiques de côté $200m$ chacune.

Dans notre structure d'agrégation, le sketch d'événements est construit en appliquant une fonction de hachage sur l' idf_e : l'identifiant de chaque événement observé. En ce qui concerne le sketch des semaines, il est construit en appliquant une fonction de hachage sur le numéro de la semaine. La taille du sketch dépend du nombre maximal d'éléments qu'il doit contenir.

Dans notre cas, nous considérons que nous avons à stocker jusqu'à 50.000 occurrences d'événements pour chaque type d'événement dans une cellule. Par conséquent, 16 bits sont nécessaires pour chaque sketch d'événement. Pour les sketches des semaines, si nous considérons 256 semaines, 8 bits sont nécessaires pour le sketch de semaines associé à un type d'événement donné. En résumé, pour chaque cellule physique et à chaque type d'événement, nous associons 4 sketches de taille $k_1 = 16$ bits pour les événements et 4 sketches de taille $k_2 = 8$ bits pour les semaines.

Dans l'espace, nous supposons qu'un véhicule observe P zones d'intérêts composées en moyenne de M cellules physiques avec E types d'événements agrégés sur toutes les granularités temporelles. Toutes ces observations sont faites sur toute la semaine (7 jours). La taille de la structure d'agrégation est

donnée par l'équation 5.2 :

$$\begin{aligned}
 \text{TailleStructure} = P \times [& ((id + i + j + k + l) \text{ octets} + E \text{ bits} + E \text{ pointeurs}) \\
 & + E \times M((i + j) \text{ octets} + 1 \text{ pointeur} + 7 \times (1 \text{ octet} + 1 \text{ Sketches_Semaines} \\
 & \quad + m \times g \text{ Sketches_Evenements}))]
 \end{aligned}
 \tag{4.4}$$

En fixant les valeurs suivantes pour les constantes :

- $id + i + j + k + l = 5$;
- $m = 4$ fonctions de hachage ;
- $i + j = 2$;
- $E = 4$ événements ;
- $\text{Taille_pointeur} = 4$ octets ;
- $\text{Sketches_Semaines} = 8$ bits ;
- $\text{Sketches_Evenements} = 16$ bits ;
- $g = 12$ granules temporelles.

Dans le premier cas, c'est à dire la ville de Paris, nous avons :

- $P = 64$ zones logiques ;
- $M = 48$ cellules physiques/zone logique.

Notre structure d'agrégation occupe donc 8,504704 M Octets.

Dans le deuxième cas, la France, nous avons :

- $P = 2000$ zones logiques ;
- $M = 48$ cellules physiques/zone logique.

Notre structure d'agrégation occupe donc 265,772 M Octets.

Nous pouvons récapituler les résultats dans le tableau 4.1 suivant :

Type d'environnement	Ville (e.g Paris)	Pays (e.g France)
Nb zones logique	64	2000
Nb cellules physiques/- zone logique	48	48
Nb événements	2	2
Nb granules temporelles	12	12
Taille de la structure	8,504704 M Octets	265,772 M Octets

TABLE 4.1 – Résultat de l'évaluation de la taille de la structure

L'évaluation de la taille de notre structure nous a permis de valider la propriété -2- dans le processus d'agregation cité dans la section 3.2.2 du chapitre 3. Avec un nombre élevé de zones d'intérêt couvrant un espace géographique important tel que la France, notre structure ne dépasse pas les 300 MOctets. Avec la présence et le développement des cartes de stockage externes « Secure Digital » ou « Compact Flash » par exemple qui ont une mémoire pouvant atteindre jusqu'à 32Go (pour « Secure Digital »), la taille de notre structure ne présente donc pas vraiment une limite au niveau du stockage.

4.4 Conclusion

Ce chapitre décrit le premier objectif de notre thèse : proposer une solution pour construire le résumé d'événements. Nous avons présenté une structure d'agrégation d'événements spatio-temporels observés qui soit utilisable dans un contexte d'échanges inter-véhicules. Cette structure est construite sur un modèle spatio-temporel à deux niveaux. Pour ce faire, nous avons tout d'abord défini notre modèle spatio-temporel qui nous permet d'avoir le même référentiel spatio-temporel physique pour tous les véhicules, ce qui permet notamment d'échanger facilement et sans perte d'informations des résumés. Dans ce modèle, chaque conducteur peut choisir ses propres zones d'intérêt.

Par la suite, nous avons décrit notre structure d'agrégation basée sur le modèle spatio-temporel à deux niveaux. Cette structure est propre à chaque véhicule ce qui nous permet de travailler dans un environnement complètement décentralisé.

Enfin, nous avons évalué la taille de notre structure qui est réaliste si nous maîtrisons le nombre de dimensions temporelles. La complexité des opérations d'accès à la structure est également raisonnable (logarithmique ou linéaire). Les travaux présentés dans ce chapitre ont été validés par deux publications [Defude 2009, Zekri 2010b].

Le deuxième objectif de ce travail est d'échanger la structure pour enrichir le résumé local de chaque véhicule. Nous proposons dans le chapitre suivant un processus d'échange qui répond à ce besoin.

Processus d'échange de résumés

Sommaire

5.1	Introduction	59
5.2	Propriétés du processus d'échange	60
5.3	Principe du processus d'échange	62
5.3.1	Notion de priorité	64
5.3.2	Matching des priorités	69
5.3.3	Fusion des résumés	71
5.4	Trace des derniers véhicules et échanges	72
5.5	Dynamique du processus d'échange	74
5.6	Conclusion	77

5.1 Introduction

La structure d'agrégation construite dans le chapitre 4 par chaque véhicule est restreinte car elle ne tient compte que des événements observés directement ou indirectement (par dissémination). De manière à avoir une information plus large, les véhicules ont tout intérêt à échanger leurs résumés entre eux. Dans ce chapitre, nous présentons notre solution d'échange de résumés que nous appelons « processus d'échange ». L'objectif du processus est de permettre aux véhicules d'échanger leurs agrégats.

Ce processus intègre les contraintes propres aux réseaux inter-véhiculaires comme mentionné dans le chapitre 1 section 1.2. En effet, les véhicules se

déplacent potentiellement vite ce qui explique que la période où deux véhicules sont suffisamment proches pour qu'ils puissent s'échanger des données ne dépasse pas quelques millisecondes.

Dans ce chapitre nous définissons les propriétés de notre processus d'échange dans la section 5.2 ensuite nous introduisons son principe dans la section 5.3. La section 5.3.1 détaille comment vont être ordonnés les résumés à échanger entre deux véhicules pour essayer d'échanger d'abord les résumés les plus intéressants. Pour expliquer les étapes du processus dans la section 5.3, nous utilisons, tout au long du chapitre un exemple d'échange entre deux véhicules. Enfin, la section 5.6 conclut ce chapitre.

5.2 Propriétés du processus d'échange

Une fois le résumé construit, chaque véhicule a une vision partielle de ses zones d'intérêts. En se basant sur cette vue partielle le système d'aide à la décision au conducteur peut souffrir du manque de connaissances sur des zones spatiales et temporelles ce qui va entraîner l'incapacité à fournir une aide au conducteur pertinente.

Nous pouvons prendre l'exemple d'un véhicule qui entre dans une zone pour la première fois. Il l'ajoute alors à ses zones d'intérêt et donc la partie de la structure d'agrégation associée à cette zone est vide. Ce véhicule cherche à se garer. A cet instant il n'y a aucun véhicule qui produit un événement de type libération de place et le conducteur ne voit pas de places libres devant lui. Une recherche dans les résumés ne lui sera d'aucune aide puisque celui-ci sera vide. Le même problème se pose pour tous les types d'événements, y compris les zones dangereuses.

Ces exemples montrent que lorsqu'un système d'aide à la décision se base uniquement sur les agrégats résultant de l'observation directe ou des événements disséminés entre les véhicules, il se trouve, dans certain cas, incapable d'aider le conducteur à prendre une décision. Comme solution à ce problème, nous proposons d'étendre notre mécanisme de construction de résumés par un processus d'échange entre véhicules. Pour cela, chaque véhicule décide de pu-

blier tout ou partie de ses résumés vers d'autres véhicules et réciproquement peut être intéressé à souscrire à tout ou partie des résumés des autres.

Plusieurs contributions ont déjà été proposées dans différents projets, concernant les protocoles d'échange d'informations en particulier pour les réseaux « *VANET* ». Ces travaux se concentrent sur « comment les informations vont transiter entre les différentes entités ». Dans [Lin 2010], Lin et al offrent un aperçu des différents protocoles de communication dans les VANETs. Ils fournissent une taxonomie des réseaux ad hoc véhiculaires où il y a principalement 3 familles de protocoles :

1. les protocoles de routage Unicast [Bernsen 2009] : Il s'agit des protocoles dont l'objectif est de transmettre une donnée d'une source connue à une destination connue ;
2. les protocoles multicast et Geocast [Füßler 2002, Maihöfer 2004] qui sont utilisés dans VESPA. Il s'agit de protocoles dont l'objectif est de diffuser une information à tous les véhicules présents dans une certaine zone géographique spécifique ;
3. les protocoles de Broadcast [Tonguz 2010] : dont l'objectif est principalement de diffuser des messages d'urgences à tous les véhicules.

Le but de ces protocoles de communication est uniquement de spécialiser le routage des messages en fonction de critères de localisation géographique. Les propriétés recherchées dans notre processus d'échange sont d'une part l'efficacité au niveau réseau (minimisation en bande passante) et la précision pour que la fusion de deux résumés soit la plus fidèle possible. Pour assurer la première propriété, nous nous appuyons déjà sur le fait que les résumés sont compacts. Pour aller plus loin, nous avons proposé un mécanisme de souscription qui permet à chaque véhicule d'annoncer les types d'événements et les zones spatio-temporelles qui l'intéressent. Couplé à un système de priorités, cela permet de n'échanger que l'information utile et de la plus prioritaire vers la moins prioritaire. Par conséquent si l'échange est interrompu, les données les plus prioritaires auront été échangées en premier.

5.3 Principe du processus d'échange

Chaque véhicule peut décider de publier la totalité ou une partie de ses résumés vers d'autres véhicules et réciproquement, peut être intéressé à souscrire à tout ou partie des résumés des autres. Vu que le but ici est d'enrichir les résumés de tous les véhicules, les publications et souscriptions sont publiques (nous publions/souscrivons vers tous les véhicules).

- « La publication » consiste à définir les résumés qui devraient être publiés ;
- « La souscription » consiste à définir des filtres indiquant les types d'événements qui nous intéressent en y ajoutant le cas échéant des critères spatiaux et temporels. Par exemple, nous sommes intéressés par le type « accidents » dans la zone « Paris » sur « le mois dernier ». L'échange d'informations entre véhicules peut ensuite se faire soit par le biais de relais (par exemple des serveurs localisés le long des routes), soit directement entre véhicules. Nous nous intéressons dans cette thèse à ce deuxième cas, pour lequel il n'est pas sûr que la durée d'échange soit suffisante pour permettre l'échange de tous les résumés pertinents. Pour cela, nous proposons d'utiliser un mécanisme de priorités (dit aussi de préférences et expliqué dans la section 5.3.1) afin de construire un ordre reposant sur les préférences utilisateur et d'utiliser cet ordre pour prioriser les résumés à échanger. Il faut noter que ce mécanisme de priorités doit être suffisamment rapide pour qu'il ne réduise pas trop le temps disponible pour l'échange des résumés.

Lorsque le véhicule V_i rencontre le véhicule V_j et veut récupérer des résumés de celui-ci, il commence par lui envoyer les informations concernant ses priorités. V_j calcule ensuite l'intersection entre ses résumés et les priorités envoyées et si elle est non vide, il envoie les données dans l'ordre de priorités demandé. En fonction de la durée de la connexion, la totalité ou une partie des échanges sera réalisée. L'opération de base à assurer ici est l'intersection entre deux zones d'intérêt (une par véhicule). Cette intersection renvoie soit l'ensemble vide si les deux zones sont distinctes, soit un ensemble de cellules physiques. Le coût de ce calcul d'intersection est logarithmique en nombre de zones (pour déterminer les p zones s'intersectant) et linéaire en nombre

de cellules physiques (en supposant qu'un véhicule observe P zones d'intérêts composées en moyenne de M cellules physiques) :

$$O(\log P + p \times M). \quad (5.1)$$

Un autre écueil à éviter dans l'échange concerne les dupliques : il ne faut pas comptabiliser plusieurs fois un même événement appartenant au résumé de V_i et V_j . Ce problème est réglé par les approches de comptage basées sur le hachage des clés d'événements (expliqué dans le chapitre 4 section 4.3). En effet, si deux véhicules V_i et V_j observent la même occurrence d'événement idf_e , nous appliquons alors la même fonction de hachage h sur les deux véhicules et nous obtenons $h_{(V_i)}(idf_e) = h_{(V_j)}(idf_e)$. Lors de l'échange des sketches, le « OU inclusif » ne conservera qu'une seule occurrence.

Le processus d'échange est modélisé dans la figure 5.1. Dans la figure 5.1(a) nous considérons 6 véhicules suffisamment proches pour pouvoir communiquer, sachant que les échanges vont se dérouler successivement sur les temps ($t_0 < t_1 < t_2 < \dots < t_6$).

Nous fixons r la portée de communication et d la distance entre le véhicule émetteur et le véhicule receveur. La distance d est une valeur comprise entre 0 et r puisque, théoriquement, la distance entre les deux véhicules ne peut pas être supérieure à la portée de communication.

Dans la figure 5.1(a), une arrête orientée entre les véhicules V_1 et V_2 décrit le fait que le résumé du véhicule V_1 a été mis à jour par des sketches échangés avec le véhicule V_2 . Le processus d'échange entre les véhicules V_1 et V_2 , où V_1 est l'expéditeur et V_2 le récepteur, se déroule en 2 phases décrites par la figure 5.1(b), et données par les algorithmes 1 et 2 :

1. *Étape 1* : dès que les deux véhicules V_1 et V_2 se trouvent à une distance inférieure à d et que le temps de rencontre donné par $\Delta t = d/v$ est supérieur à quelques milli secondes, V_1 demande l'échange en envoyant ses priorités à V_2 (appel de la fonction *demandeEchange* de l'algorithme 1). V_2 stocke les priorités de V_1 puis les compare avec ses résumés et

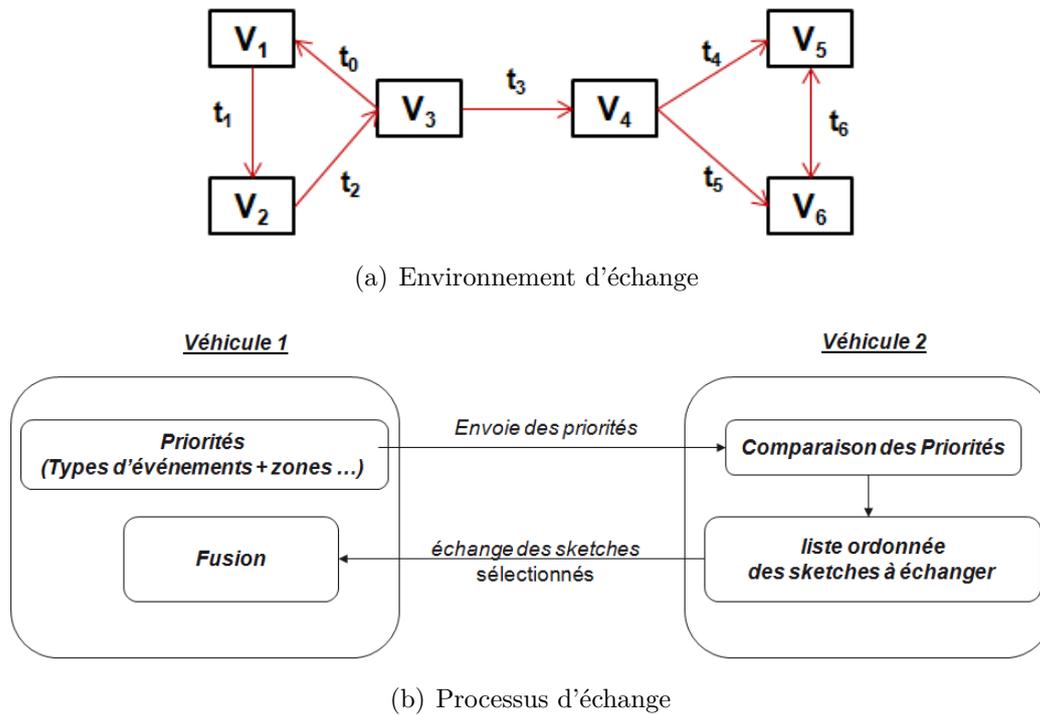


FIGURE 5.1 – Principe d'échange

retourne une liste ordonnée de sketchs à échanger ;

2. *Étape 2* : cette étape consiste à fusionner des couples de sketchs ($\text{sketch}(V_1)$, $\text{sketch}(V_2)$) en utilisant le « OU inclusif » (la fonction $\text{fusion_sk}()$). Si le temps de connexion $\Delta t = d/v$ est suffisamment important, tous les sketchs sélectionnés sont échangés mais si la connexion est interrompue avant la fin de l'échange complet des sketchs sélectionnés, seuls les sketchs les plus prioritaires se trouvent alors échangés.

5.3.1 Notion de priorité

Les priorités ou préférences désignent un ensemble de règles définissant un ordre entre deux éléments. Nous proposons d'utiliser comme éléments les différents types d'événements, les différentes granularités temporelles pour traiter la dimension temporelle et les différentes zones d'intérêt pour traiter la dimension spatiale. La notion de priorités permet au conducteur de 1- définir les événements qu'il veut récupérer dans l'échange 2- et dans quel ordre il

Algorithme 1 *Echange*(V_i, V_j)

Entrée : V_i, V_j /* le véhicule i demande à véhicule j les résumés correspondants à ses priorités*/

1: **debut**

/* Précondition : $Distance(V_1.pos, V_2.pos) \leq MAX_dis \ \&\& \ T_courant \leq (T_initial + \Delta t) \ \&\& \ V_2.idf \not\subseteq idf_echange$ */

2: $sketchs_recus \leftarrow V_j \rightarrow demandeEchange(V_i.priorités)$ /* demande d'échange avec V_j avec envoi des préférences de V_i , appel bloquant en attente de la terminaison de l'appel ou d'une perte de la connexion*/

3: **si** $sketchs_recus \neq NULL$ **alors**

4: $fusion_sk(Resume, sketchs_recus)$ /* Fusion des sketches dans V_i */

5: $ajoutCache(V_j)$

/* V_j est ajouté dans le cache de V_i */

6: **fin**

veux les récupérer.

Les préférences d'un véhicule sont décrites par un ensemble de préférences élémentaires :

$PREF_{VEHICULE} ::= PREF_ELEM^*$

Si cet ensemble est vide, cela signifie que le véhicule n'est intéressé par rien.

Une préférence élémentaire est définie de la manière suivante :

$PREF_ELEM ::= ITEM | ITEM > ITEM$

$ITEM ::= TYPE_EVT | ZONE_INTERET | GRANULE_TEMPORELLE$

Une préférence élémentaire est définie par un *ITEM* ou un ordre sur deux *ITEMs*. Les items cités sont ceux qui intéressent le véhicule.

A titre d'exemple, nous allons définir dans ce qui suit les préférences d'un véhicule V_i . L'expression suivante décrit les types d'événements pour lesquels V_i veut recevoir de l'information (ici accident en premier puis espace de stationnement, les autres ne l'intéressent pas) :

(Exp 1) $Accident > espace \ de \ stationnement$

Si nous supposons 10 granularités temporelles (g_1 représente la période la plus récente et g_{10} la plus ancienne), l'expression suivante permet de spécifier que V_i s'intéresse systématiquement aux résumés les plus récents (les granularités g_6 à g_{10} ne feront pas l'objet d'échanges) :

(Exp 2) $g_1 > g_2 > g_3 > g_4 > g_5$

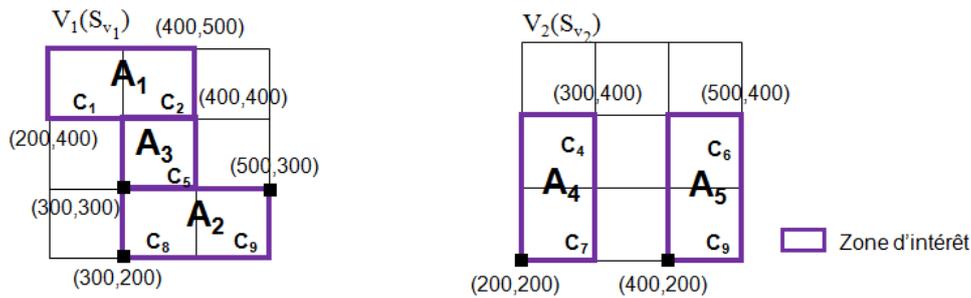
De même, si nous supposons 10 zones d'intérêt pour V_i , l'expression suivante permet de définir un ordre sur celles-ci :

$$(Exp\ 3) \quad ZI_1 > ZI_3; ZI_2 > ZI_4; ZI_4 > ZI_6; ZI_6 > ZI_8$$

Dans cet exemple, nous avons un ordre partiel avec ZI_1 et ZI_2 qui sont les zones prioritaires, puis ZI_3 et ZI_4 , puis ZI_6 et enfin ZI_8 . Les zones non citées ne sont pas concernées par l'échange.

Globalement, les expressions *Exp1*, *Exp2* et *Exp3* définissent la priorité à suivre lorsque le véhicule V_i échange des données avec un autre véhicule.

Les étapes du processus d'échange sont détaillées dans le reste du chapitre en les appliquant à un exemple d'échange entre deux véhicules V_1 et V_2 à l'instant t_1 . Une représentation des structures d'agrégation des zones d'intérêt de V_1 et V_2 est présentée dans la figure 5.2.



(a) Modèle spatial

Structure des zones d'intérêt de V_1							Structure des zones d'intérêt de V_2									
↓	A_1	200	400	400	500	Place de stationnement	Accident	↓	A_4	200	200	300	400	Place de stationnement	Accident	Embouteillage
	A_2	300	200	500	300	Place de stationnement	Accident		A_5	400	200	500	400	Place de stationnement	Accident	Embouteillage
	A_3	300	300	400	400	Place de stationnement	Accident									

(b) Structures d'agrégation associées

FIGURE 5.2 – Environnement d'échange

- A l'instant t_0 , V_1 ayant un résumé R_1 , s'intéresse à deux types d'événements (*Accidents et espaces de stationnement disponibles*) dans trois zones logiques A_1 , A_2 et A_3 . Chaque zone est composée de cellules physiques $A_1(C_1, C_2)$, $A_2(C_8, C_9)$ et $A_3(C_5)$;

- A l'instant t_0 , V_2 possédant un résumé R_2 est intéressé par trois types d'événements (*accidents, espaces de stationnement et les embouteillages*) dans deux zones logiques A_4 et A_5 . Chaque zone est composée de deux cellules physiques A_4 (C_4, C_7) et A_5 (C_6, C_9) comme le montre la figure 5.2(a) pour le modèle spatial et la figure 5.2(b) pour les résumés associés.

Ces véhicules ont aussi des priorités sur les types d'événements, les zones d'intérêt et le temps (granularité temporelle) qu'ils veulent suivre. Ces priorités pour V_1 et V_2 sont exprimées comme suit :

$$Priorités\ de\ V_1 \left\{ \begin{array}{l} Accident > Espace\ de\ stationnement \\ g_1 > g_2 > \dots > g_{12} \\ A_2 > A_1; A_3 > A_1 \end{array} \right.$$

$$Priorités\ de\ V_2 \left\{ \begin{array}{l} Embouteillage \\ g_1 > g_2 > \dots > g_{12} \\ A_3 > A_4 \end{array} \right.$$

La définition de ces priorités ne forme pas forcément un ordre total mais un ordre partiel. Par exemple, pour V_1 , A_2 et A_3 ne sont pas comparables. Il faut donc transformer l'ordre partiel induit par les priorités en un ordre total en utilisant des ordres implicites. Plus précisément, nous considérons que les priorités sont structurées selon trois dimensions : les événements, les zones d'intérêt et les granularités temporelles. Nous définissons un ordre total sur ces trois dimensions : événements $>$ zones d'intérêt $>$ granularités temporelles.

Ensuite, au niveau de chaque dimension il faut également avoir un ordre total, construit de manière compatible avec l'ordre partiel donné par les priorités. Pour la dimension événements, nous proposons de donner la priorité la plus forte aux événements correspondants aux alertes (accidents, freinage d'urgence, etc.) et la plus faible aux autres événements de type ressource (libération d'une place de stationnement, ...). A l'intérieur de chacune de ces deux catégories, nous utilisons l'ordre lexicographique (accident $<$ freinage par exemple).

Pour la dimension zones d'intérêt, nous proposons d'utiliser la proximité relativement à la position courante de l'échange : les zones vont être classées selon leur proximité (en nombre de cellules physiques) relativement à la cellule courante au moment de l'échange. En cas d'égalité, les zones sont classées selon leur identifiant.

Pour la dimension granularités temporelles, nous proposons d'utiliser le même principe que pour les zones d'intérêt, c'est à dire d'ordonner les granularités relativement au temps courant.

Le principe de transformation de l'ordre partiel en ordre total est le suivant :

- pour chaque dimension, les éléments non comparables selon les priorités sont ordonnés selon les règles définies ci dessus ;
- l'ordre total est construit à partir de l'ordre total sur les dimensions et des ordres totaux définis sur chaque dimension.

Dans notre exemple, les ordres partiels définis par les priorités sont :

- dimension événement : Accident $>$ Espace de stationnement ;
- dimension zones d'intérêt : $A_2 > A_1; A_3 > A_1$;
- dimension granularités temporelles : $g_1 > g_2 > \dots > g_{12}$.

Ici seule la dimension zones d'intérêt n'est pas totalement ordonnée (A_2 et A_3 ne sont pas comparables). Si l'on suppose que la position courante de V_1 à l'instant de l'échange est la cellule physique C_8 , alors $A_3 > A_2$.

Finalement, l'ordre total est le suivant :

(Accident, A_3 , g_1), ..., (Accident, A_3 , g_{12}), (Accident, A_2 , g_1), ..., (Accident, A_2 , g_{12}), (Accident, A_1 , g_1), ..., (Accident, A_1 , g_{12}), (Espace de stationnement, A_3 , g_1), ..., (Espace de stationnement, A_3 , g_{12}), (Espace de stationnement, A_2 , g_1), ..., (Espace de stationnement, A_2 , g_{12}), (Espace de stationnement, A_1 , g_1), ..., (Espace de stationnement, A_1 , g_{12})

Cet ordre total doit être calculé lors de chaque échange puisqu'il dépend (partiellement) de la position courante et de l'instant courant du véhicule qui veut échanger. Il va être envoyé lors de l'échange avec un autre véhicule sous forme de la séquence des trois ordres totaux. Les identifiants de zones d'intérêt seront juste remplacés par les coordonnées au niveau physique de la zone.

5.3.2 Matching des priorités

Après réception des priorités de V_1 , V_2 cherche une intersection entre son résumé et les priorités reçues. V_2 sélectionne les sketches de son résumé local correspondants aux types d'événements requis par V_1 (par exemple, un accident et une place de stationnement disponible) pour les zones spatiales et les granularités temporelles données par les priorités de V_2 .

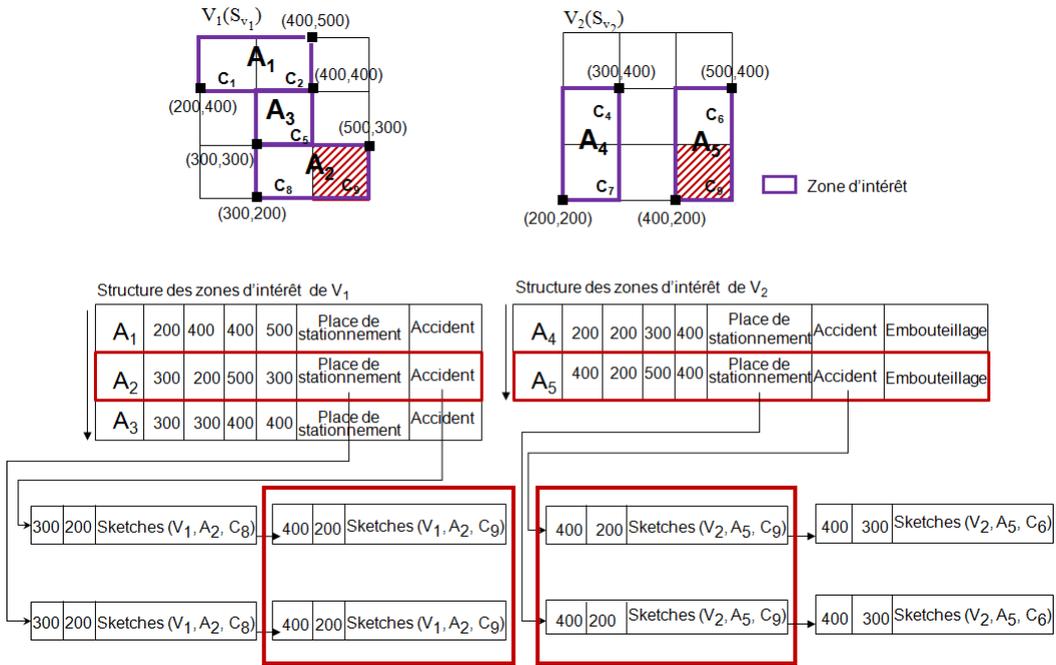


FIGURE 5.3 – Matching des priorités de V_1 avec les résumés de V_2

Dans notre exemple précédent, il y a une intersection entre les zones d'intérêt de V_1 (A_1, A_2 et A_3) et celles de V_2 (A_4 et A_5). Comme le montre la figure 5.3, V_2 identifie une seule cellule physique en commun avec V_1 donnée par $A_2 \cap A_4 = C_9$. Puis, V_2 identifie les sketches de l'échange (qui sont associés soit à un accident ou à une place de stationnement disponible) pour toutes les granularités temporelles correspondantes à la cellule C_9 (figure 5.2(b)). Il n'y a pas d'autres d'intersections entre A_1, A_2 et A_3 sur V_1 et A_4, A_5 sur V_2 . La sélection des sketches d'intersection dans V_2 est assurée par l'algorithme 2. Si V_2 voulait échanger avec V_1 , ses priorités sont comparées avec les résumés de V_1 , mais l'intersection est vide, car ils ne sont pas intéressés par les mêmes

types d'événements.

L'algorithme 2 permet de sélectionner les sketches qui seront envoyés du véhicule V_2 au véhicule V_1 .

Cet algorithme prend en entrée les priorités de V_1 représentées par $V_i.priorités$ et il retourne $liste_sketchs$ la liste des cellules physiques avec les sketches associés.

L'algorithme 2 énumère les priorités données en entrée et pour chacune d'entre elle calcule son intersection avec les résumés de V_j . Ce calcul d'intersection recherche dans les résumés de V_j ceux dont la zone d'intérêt a une intersection non vide avec celle de la priorité courante. Si oui, il renvoie les sketches associés à l'événement de la priorité courante (si cet événement existe dans ce résumé). Tous les sketches sélectionnés sont agrégés dans la liste renvoyée à V_i (ce retour peut se faire au fur et à mesure du remplissage de la liste sans attendre la terminaison de l'algorithme).

Algorithme 2 *demandeEchange($V_i.priorités$)*

Entrée : $V_i.priorites$ /* liste des priorités provenant de V_i */
 /* le véhicule j traite la demande d'échange du véhicule j en renvoyant les résumés correspondants dans l'ordre des priorités demandées */

Sortie : $liste_sketchs$ /* liste ordonnée des cellules physiques avec les sketches associés */

- 1: **debut**
- 2: $liste_sketchs \leftarrow NULL$
- 3: **pour** $j = 1 \rightarrow V_i.priorites.length()$ **faire**
- 4: $sketchs \leftarrow intersection(V_i.priorités[j], V_j.resumes)$
 /* calcul de l'intersection du jème item de priorités de V_i dans les résumés de V_j .
 Retourne pour chaque cellule physique dans l'intersection, les sketches associés */
- 5: **si** $sketchs \neq NULL$ **alors**
- 6: $liste_sketchs \leftarrow liste_sketchs + sketchs$
- 7: **retour** $liste_sketchs$
- 8: **fin**

La complexité de l'algorithme 2 est de l'ordre de $\mathcal{O}(P \times \log(N))$ avec N la taille du résumé de V_j et P le nombre de priorités de V_i .

5.3.3 Fusion des résumés

Dans la deuxième étape, V_2 envoie les sketches sélectionnés dans l'ordre correspondant aux préférences à V_1 (d'abord les sketches associés aux accidents, puis ceux sur les places de stationnement disponibles). Puis, une opération de fusion avec un « OU inclusif » est effectuée localement sur V_1 . Le résultat de cette opération est présenté dans la figure 5.4.

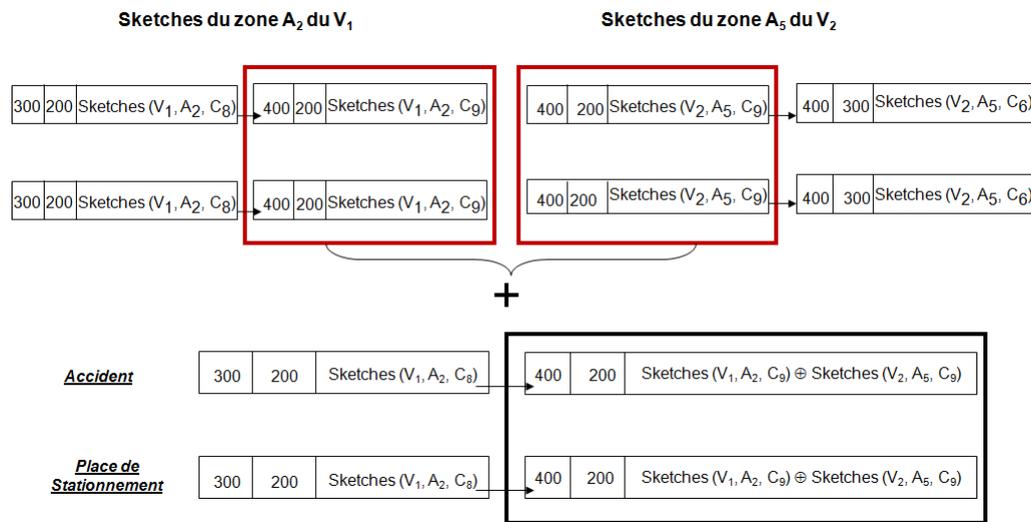


FIGURE 5.4 – fusion des sketches dans V_1

Ainsi, à $t_1 + \Delta t$, le résumé de la cellule physique commune dans V_1 change du sketch(V_1, A_2, C_9) à la fusion de deux sketches (V_1, A_2, C_9) + sketches (V_2, A_5, C_9).

Algorithme 3 *Fusion_sk(Résumé, sketches_recus)*

Entrée : *Résumé, sketches_recus* /* résumé de V_i et sketches envoyés par V_j */

Sortie : *Résumé* /* résumé de V_i modifié par la prise en compte des sketches reçus de V_j */

1: **debut**

2: **pour** $i = 1 \rightarrow sketches_recus.length()$ **faire**

3: *oulogique*(*Résumé, sketches_recus*[i]) /* on met à jour le résumé correspondant au sketch courant en faisant le ou logique des sketches */

4: **fin**

L'algorithme 3 prend en entrée la liste des sketches *sketchs_recus* retournée par l'algorithme 2. Pour chaque combinaison (événement, cellule physique, granularité temporelle) de *sketchs_recus* trouvée dans *Résumé*, il va y avoir une mise à jour des sketches événements et semaine de l'élément de *Résumé* par le calcul du ou logique avec l'élément courant de *sketchs_recus*. La sortie de l'algorithme 3 est le résumé de V_i mis à jour.

En se basant sur l'exemple de la figure 5.4, nous pouvons généraliser et donner une représentation temporelle de l'échange des résumés de l'environnement de la figure 5.1(a) dans le tableau 5.1.

$t_i \setminus V_j$	V_1	V_2	V_3	V_4	V_5	V_6
t_0	R_1	R_2	R_3+R_1	R_4	R_5	R_6
t_1	R_1+R_2	R_2	R_3+R_1	R_4	R_5	R_6
t_2	R_1+R_2	$R_2+R_3+R_1$	R_3+R_1	R_4	R_5	R_6
t_3	R_1+R_2	$R_2+R_3+R_1$	$R_3+R_1+R_4$	R_4	R_5	R_6
t_4	R_1+R_2	$R_2+R_3+R_1$	$R_3+R_1+R_4$	R_4+R_5	R_5	R_6
t_5	R_1+R_2	$R_2+R_3+R_1$	$R_3+R_1+R_4$	R_4+R_5	R_5+R_6	R_6
t_5	R_1+R_2	$R_2+R_3+R_1$	$R_3+R_1+R_4$	R_4+R_5	R_5+R_6	R_6+R_5

TABLE 5.1 – Représentation temporelle d'échange de résumés

Une cellule (i, j) du tableau donne la valeur du résumé d'un véhicule V_j à un instant t_i . Cela illustre le fait que l'échange permet d'améliorer les résumés initiaux des véhicules : par exemple, V_4 passe d'un résumé initial à t_0 égal à R_4 à un résumé résultat de la fusion de $R_4 + R_5 + R_6$ à t_6 . Notons également que le processus d'échange peut être bidirectionnel. Ceci est illustré par les deux arêtes entre V_5 et V_6 à t_6 dans la figure 5.1(a).

5.4 Trace des derniers véhicules et échanges

Dans le processus d'échange, il faut éviter d'échanger continuellement avec les mêmes véhicules (notamment quand plusieurs véhicules cherchent une place disponible dans une même zone). Pour ce faire, nous stockons au niveau de chaque véhicule dans une liste de taille fixe les identifiants des N derniers véhicules avec lesquels un échange a eu lieu, ainsi que l'instant de l'échange

le plus récent. La liste est appelée *idf_echange* (ligne 7 algorithme 1). Si un véhicule V lors d'un échange à un instant t se trouve déjà dans cette liste et que le temps courant est légèrement supérieur ou égal à $t + \Delta t$, donc proche de l'instant du dernier échange, alors V ne sera pas sélectionné pour l'échange. Nous avons choisis de modéliser, dans un fichier *XML* (figure 5.5), les traces

```
1: <?xml version="1.0" encoding="UTF-8"?>
2: <Echange>
3:   <Liste>
4:     <idf>101</idf>
5:
6:     <Vehicule>
7:       <idf>12</idf>
8:       <Date>22/04/2012</Date>
9:       <Temps>09:03:12:07</Temps>
10:    </Vehicule>
11:
12:    <Vehicule>
13:      <idf>222</idf>
14:      <Date>22/04/2012</Date>
15:      <Temps>23:22:54:12</Temps>
16:    </Vehicule>
17:
18:    <Vehicule>
19:      <idf>154</idf>
20:      <Date>22/04/2012</Date>
21:      <Temps>19:14:32:50</Temps>
22:    </Vehicule>
23:
24:    <Vehicule>
25:      <idf>876</idf>
26:      <Date>22/04/2012</Date>
27:      <Temps>11:11:33:10</Temps>
28:    </Vehicule>
29:    .
30:    .
31:    .
32:    .
33:  </Liste>
34: </Echange>
```

FIGURE 5.5 – Extrait du fichier XML des traces des derniers véhicules et échanges

des derniers échanges avec 1- la liste d'identifiants des véhicules ainsi que 2- les dates et les instants du début d'échange associés à chaque véhicule :

5.5 Dynamique du processus d'échange

Dans cette section, nous évaluons la capacité de notre processus d'échange pour diffuser les résumés auprès des véhicules intéressés. Cette évaluation est théorique mais nous donnerons dans le chapitre d'expérimentations une évaluation expérimentale de la dynamique des échanges à partir des résultats des simulations. Nous définissons la formule permettant de donner le nombre de véhicules atteints après i périodes d'échanges.

$$U_i = \sum_{i=1}^n U_{i-1} \times N\alpha\beta - \gamma \times \left(\sum_{i=1}^n U_{i-1} \times N\alpha\beta \right) \quad (5.2)$$

avec :

- N le nombre de véhicules dans l'environnement routier ;
- α le pourcentage de véhicules rencontrés ;
- β le pourcentage de véhicules intéressés par le résumé ;
- γ le pourcentage des échanges redondants (γ s'incrémente après chaque période) ;
- U_{i-1} est le nombre de véhicules recevant le résumé dans la période $(i-1)$;
- $U_0 = U_1 = 0$.

Dans ce modèle, nous formalisons la diffusion des résumés comme un processus itératif. A chaque période, les véhicules ayant un résumé local R l'échange avec un nombre constant de véhicules définis par $\alpha\beta N$. Pour faire face aux redondances (un véhicule peut obtenir un même résumé R de plusieurs véhicules), nous introduisons γ comme un facteur de redondance qui diminue U_i . En utilisant cette formule, nous mettons en évidence l'impact de deux facteurs :

1. $\alpha\beta$, le pourcentage des véhicules recevant le résumé ;
2. γ , pourcentage des échanges redondants.

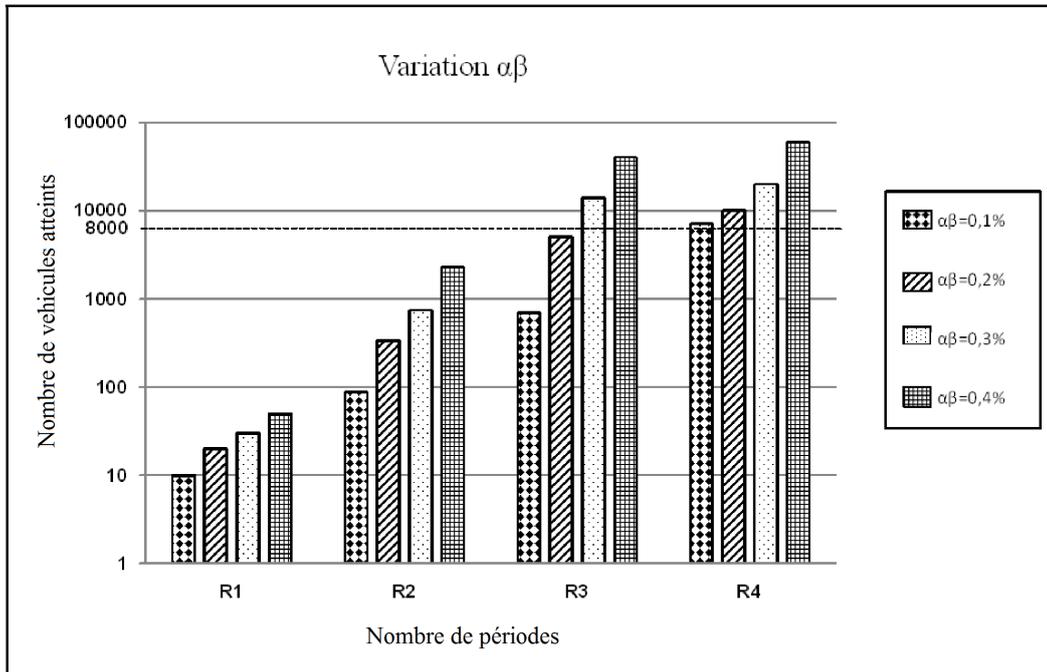


FIGURE 5.6 – Nombre de véhicules rencontrés pour la variation de $\alpha\beta$

Le résultat de cette variation est représenté dans un axe logarithmique dans les figures 5.6 et 5.7.

Dans ce qui suit, nous considérons un cas standard où l'objectif de dissémination d'un résumé est de 80% (soit 8000 véhicules pour N égal à 10000), autrement dit que 8000 véhicules parmi 10000 sont intéressés par le résumé. Nous supposons de plus que chaque période correspond à une semaine. La figure 5.6 montre l'impact du paramètre $\alpha\beta$ avec le même facteur de redondance γ (initialement fixée à 20% et augmenté de 10% dans chaque période). Notons que le nombre de véhicules atteints augmente de façon exponentielle (linéaire dans l'axe logarithmique). En outre, une valeur élevée de $\alpha\beta$ nécessite un nombre de périodes minimal pour atteindre les véhicules intéressés par le résumé. Comme le montre la figure 5.6, pour $N = 10000$, tous les véhicules intéressés peuvent être atteints en 4 périodes (y compris dans le cas d'une faible valeur de $\alpha\beta = 0,1\%$).

L'impact de γ est illustré à la figure 5.7. Nous avons assigné à $\alpha\beta$ une valeur constante et avons étudié la variation du paramètre γ avec deux valeurs

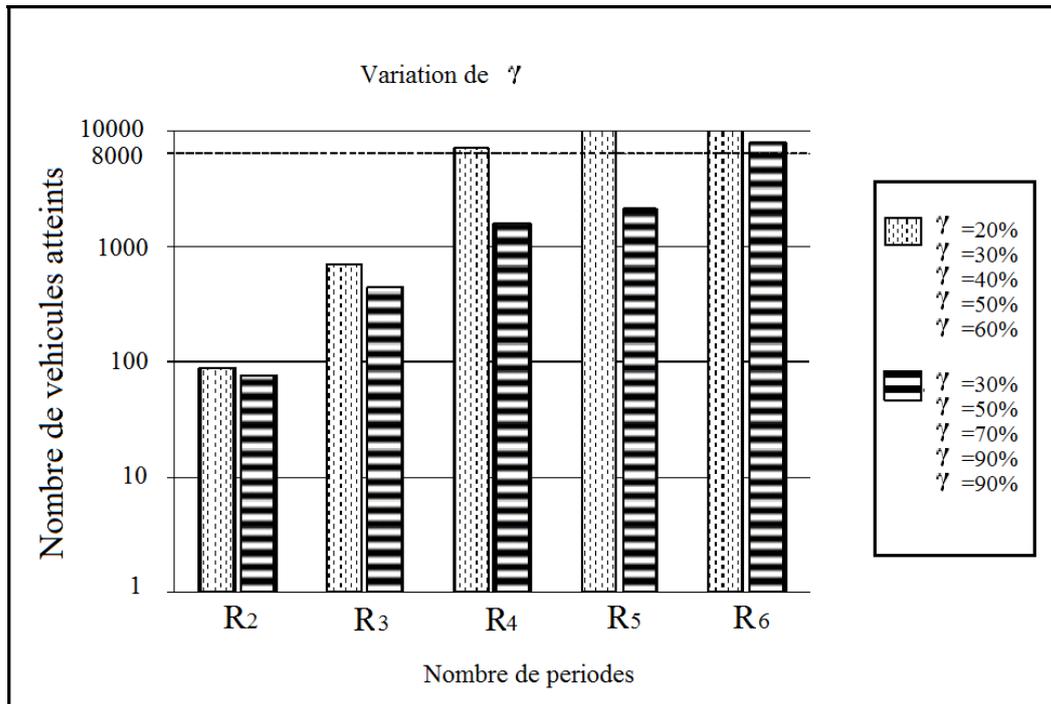


FIGURE 5.7 – Nombre de véhicules rencontrés pour la variation de γ

différentes.

1. Dans le premier cas, nous choisissons une valeur faible pour γ qui est initialement fixée à 20%. Cette valeur augmente de 10% après chaque période ;
2. Dans un deuxième cas, un pourcentage plus élevé est utilisé pour γ (initialement fixée à 30% et augmenté de 20% après chaque période).

Les deux scénarios correspondent au cas où de nombreux véhicules sont à la recherche d'une place de stationnement disponible. La probabilité qu'un véhicule rencontre à plusieurs reprises le même véhicule est alors élevée ce qui peut conduire à plusieurs redondances. Dans la figure 5.7, nous constatons que dans de telles configurations et pour un nombre de véhicules N égal à 10000, le nombre de périodes nécessaires pour atteindre tous les véhicules intéressés reste limité quand une valeur élevée de γ est utilisée. Donc pour communiquer le résumé aux 8000 véhicules intéressés il faut 4 semaines dans le premier scénario avec $\gamma=60\%$ et 6 semaines dans le deuxième scénario avec

$\gamma=90\%$).

En conclusion, notre processus d'échange est efficace car il assure la diffusion du résumé vers tous les véhicules concernés dans un nombre limité de périodes. Par ailleurs, la variation du paramètre γ a un impact moins important que la variation des paramètres $\alpha\beta$ sur le nombre de périodes nécessaires pour atteindre les véhicules intéressés par le résumé.

5.6 Conclusion

Dans ce chapitre nous avons décrit notre processus d'échange des résumés ou agrégats. Les propriétés recherchées sont d'une part l'efficacité au niveau réseau, d'autre part, la précision afin que la fusion de deux résumés soit la plus fidèle possible.

Pour assurer la première propriété, nous nous appuyons sur le fait que les résumés sont compacts (du à la structure de données utilisée). Pour aller plus loin, nous avons proposé un mécanisme de souscription qui permet à chaque véhicule d'annoncer les types d'événements et les zones spatio-temporelles qui l'intéressent. Couplé à un système de préférences, cela permet de n'échanger que l'information utile et la plus prioritaire. Ainsi, si l'échange est interrompu, les données les plus prioritaires auront été échangées en premier.

Dans la deuxième partie de ce chapitre, nous avons détaillé les deux étapes de notre processus d'échange et avons défini les traitements à assurer pour chacun des deux véhicules à chaque échange de résumé.

Une troisième partie du chapitre consiste à étudier le dynamique de notre processus d'échange ce qui nous a permis de conclure que notre processus est efficace car il assure la diffusion du résumé à tous les véhicules concernés dans un nombre limité de périodes.

Les travaux présentés dans ce chapitre ont été validés par la publications [Zekri 2012]. Dans le chapitre suivant, nous nous intéressons aux méthodes d'exploitation de résumés pour l'aide à la décision.

Exploitation des résumés

Sommaire

6.1	Introduction	79
6.2	Architecture d'exploitation du résumé	81
6.3	Niveau d'accès aux résumés	82
6.3.1	Cas de recherche d'une ressource	83
6.3.2	Cas de production d'alerte	85
6.4	Niveau d'assistance aux conducteurs	86
6.4.1	Recherche de ressources	86
6.4.2	Production d'alertes	90
6.5	Conclusion	91

6.1 Introduction

Dans les deux chapitres précédents, nous avons assuré les deux premiers objectifs de notre thèse, à savoir proposer « une structure pour construire les résumés » et « un processus pour les échanger entre les véhicules ». Dans ce chapitre, notre objectif est de répondre au troisième et dernier objectif de notre thèse en proposant « une approche pour exploiter les résumés et aider le conducteur à prendre une décision sur la route ».

Nous avons choisi une approche déclarative d'exploitation des résumés basée sur l'utilisation d'un langage de requêtes. La problématique de ce chapitre est la définition d'un langage de requêtes sur les résumés et son modèle d'évaluation associé. Les types de requêtes à supporter sont des requêtes spatio-

temporelles classiques (trouver une ressource, par exemple une place de parking à un instant et dans une zone donnée) et les requêtes spatio-temporelles continues (avertir le conducteur dès qu'une probabilité forte d'un événement peut survenir, entrée dans une zone à risque par exemple). La principale difficulté vient du fait que les résumés ne donnent qu'une information partielle qui est difficile à quantifier et qu'il faut à tout prix éviter de donner trop de fausses alertes au conducteur, tout en lui donnant néanmoins une aide.

Le problème principal de notre approche est celui de l'interprétation des résumés. Le résumé que nous construisons évalue la fréquence d'occurrences d'événements relativement à une zone spatio-temporelle. Il faut ensuite être capable de passer de cette valeur à une probabilité (ou tout autre modèle de l'incertain) d'occurrence d'un événement. Un autre problème est celui de la détermination de la (ou les) zone(s) spatio-temporelle(s) à traiter, sachant qu'un véhicule ne va pas forcément changer de direction à tout moment. Par exemple, lorsque nous recherchons une place de stationnement libre (ou plus exactement lorsque nous demandons au système de nous guider vers une zone ayant une probabilité élevée d'avoir une place libre), selon que nous raisonnons à une cellule de distance ou à plusieurs, nous n'allons pas forcément sélectionner la même direction.

Par ailleurs, dans le reste de ce chapitre, nous supposons qu'il existe toujours une route qui relie deux cellules physiques voisines et donc un véhicule peut toujours aller d'une cellule à une autre voisine. C'est bien évidemment une hypothèse très réductrice et il conviendrait de prendre en compte des informations cartographiques pour améliorer notre approche.

Le reste de ce chapitre est organisé comme suit. Dans la section 6.2, nous présentons l'architecture d'exploitation du résumé. L'exploitation est constituée de deux niveaux : 1- « le niveau d'accès au résumé » expliqué dans la section 6.3 où nous allons interroger le résumé avec des requêtes qui varient selon le type d'informations recherché. Les résultats du niveau d'accès au résumé sont alors envoyés au niveau 2- « le niveau d'assistance au conducteur » où les résultats seront interprétés pour envoyer les recommandations au conducteur. Nous concluons ce chapitre dans la section 6.5.

6.2 Architecture d'exploitation du résumé

Nous exposons dans cette section comment exploiter les résumés pour aider le conducteur dans la prise de décision. Ce travail se concentre dans la deuxième et la troisième partie de l'architecture globale de notre système présentée dans la section 3.4 du chapitre 3 et que nous reprenons dans la figure 6.1.

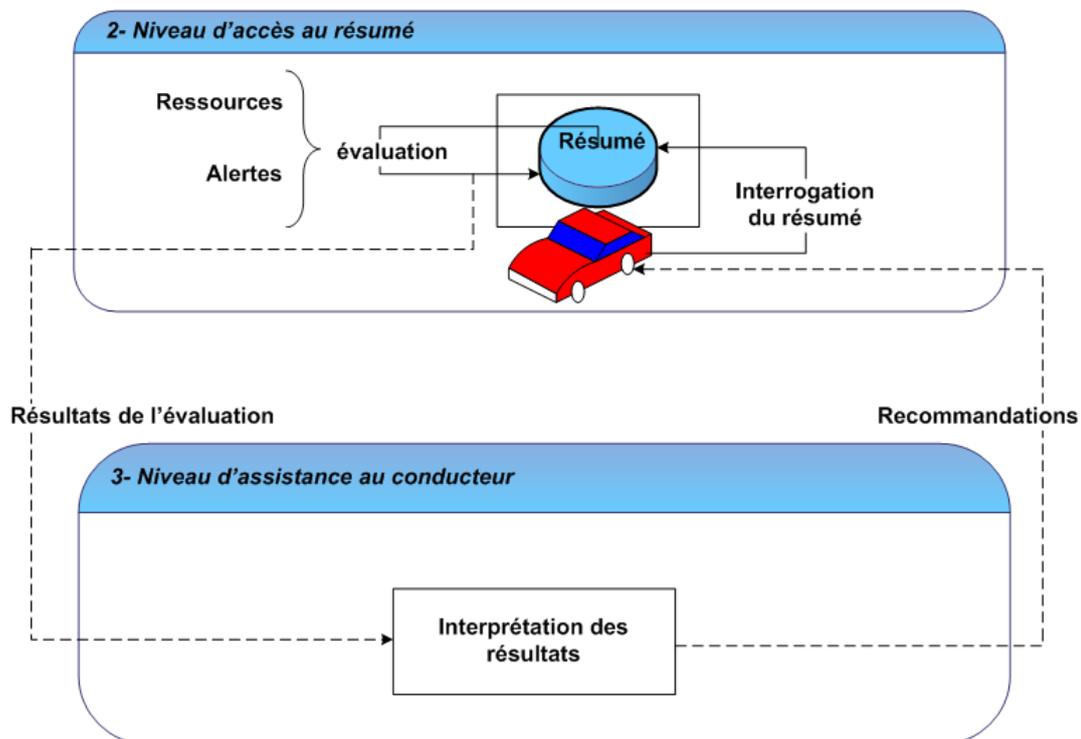


FIGURE 6.1 – Modèle d'assistance au conducteur

1. Au plus bas niveau, le résumé de chaque véhicule est vu comme une simple base de données relationnelle (constituée d'une relation unique) et que nous allons pouvoir interroger avec des requêtes spatio-temporelles. Ce niveau est détaillé dans la section 6.3 ;
2. Nous avons besoin au-dessus de cette couche d'accès aux données des résumés pour mieux interpréter les résultats en fonction des besoins applicatifs et c'est le rôle du « niveau d'assistance aux conducteurs » expliqué dans la section 6.4.

Nous décrivons dans les sections suivantes successivement ces deux niveaux.

6.3 Niveau d'accès aux résumés

Un résumé peut être vu comme une base de données relationnelle constituée d'une seule relation :

RÉSUMÉ (cellule_physique, temporalité, type_événement,
nombre_occurrences, nombre_semaines)

Cette base de données peut être interrogée par un langage de requêtes de type SQL étendu au spatio-temporel et avec deux modes d'évaluation (classique et continue).

Nous rappelons que nous avons classé les informations utilisées dans un système d'aide aux conducteurs en deux catégories (voir la section 3.2.2 dans le chapitre 3) :

1. les ressources : il s'agit des données associées à des objets recherchés par l'utilisateur tels que les places de stationnement. Ces données sont demandées selon ses besoins par le conducteur donc accédées en « mode pull ». Pour répondre aux besoins du conducteur nous allons interroger le résumé en utilisant des requêtes spatio-temporelles en mode d'évaluation classique ;
2. les alertes : il s'agit des informations sur tous les dangers présents sur la route et que le conducteur risque de rencontrer (accident, zone dangereuse, embouteillage etc.). Ce type de données est demandé d'une façon continue par le conducteur donc il les demande en « mode push ». Pour se faire nous allons utiliser les requêtes spatio-temporelles en mode d'évaluation continue.

6.3.1 Cas de recherche d'une ressource

Dans cette section nous allons nous intéresser à l'interrogation du résumé dans le cas d'une recherche de ressource. Nous prenons l'exemple de la recherche de places de stationnement disponibles.

A titre d'exemple, voici une première requête Q_1 qui recherche les cellules voisines de la position courante du véhicule et où la probabilité de trouver une place de parking sera soit élevée soit très faible. Cette requête n'est pas de type continu et sera évaluée à la demande du conducteur :

Requête Q_1 _____

```
SELECT cellule_physique, distance, nombre_occurrences/nombre_semaines,
'POSITIF'
```

```
FROM RESUME
```

```
WHERE cellule_physique IN VOISINAGE(current_position(), distancemax)
and CORRESPONDS (temporalité, current_date()) and type_événement =
'libération parking' and nombre_occurrences/nombre_semaines >= seuilFre-
quenceMax and nombre_semaines >= seuilSemaineMin
```

```
UNION
```

```
SELECT cellule_physique, distance, nombre_occurrences/nombre_semaines,
'NEGATIF'
```

```
FROM RESUME
```

```
WHERE cellule_physique IN VOISINAGE (current_position(), distancemax)
and CORRESPONDS (temporalité, current_date()) and type_événement =
'libération parking' and nombre_occurrences/nombre_semaines <= seuilFre-
quenceMin and nombre_semaines >= seuilSemaineMin
```

Q_1 est une union de deux sous-requêtes qui renvoient respectivement les

cellules avec une fréquence d'occurrences d'événements (ratio entre nombre d'occurrences et nombre de semaines d'observation) élevée (respectivement faible). Ces cellules sont filtrées (fonction VOISINAGE) par leurs positions relatives vis-à-vis de la position courante du véhicule et de la définition d'un carré englobant dont le côté est fixé par un nombre de cellules (distance maximale). Nous supposons que cette fonction renvoie pour chaque cellule filtrée la distance (en nombre de cellules) relativement à la position courante. Nous ajoutons également un filtre sur la temporalité (fonction CORRESPONDS) et sur le nombre de semaines d'observations (nous ne calculons que les fréquences obtenues avec un nombre suffisant de semaines d'observation). Des seuils doivent être définis pour déterminer la probabilité élevée ou faible (seuil-FrequenceMax et seuilFrequenceMin) ainsi que le nombre suffisant de semaines (seuilSemaineMin).

Dans ce type de requêtes, nous ne nous intéressons pas seulement aux zones traversées mais aussi aux régions « proches » de la zone d'intérêt. Il est également intéressant d'avoir des réponses positives (zones où la probabilité de trouver la ressource recherchée est élevée) mais aussi négatives (zones où la probabilité de trouver la ressource recherchée est très faible).

Les réponses à la requête Q_1 sur le résumé d'un véhicule qui cherche une place pour se garer est représentée par un ensemble de tuples triés donnant pour chaque cellule physique l'orientation par rapport à la position courante du véhicule, sa distance au véhicule et sa fréquence d'occurrences de l'événement.

1. (Haut Droite, 30, 2);
2. (Bas Droite, 20, 2);
3. (Haut Gauche, 12.5, 2);
4. (Haut Gauche, 10, 1);
5. (Bas Gauche, 7, 2);
6. (Bas Gauche, 5, 1);
7. (Haut Droite, 2, 1);
8. (Bas Droite, 0, 1).

Ces résultats vont être ensuite interprétés par « le niveau d'assistance aux conducteurs ».

6.3.2 Cas de production d'alerte

Dans cette section nous allons présenter le type de requête nécessaire pour alerter les conducteurs avant d'entrer dans une zone dangereuse ou les avertir de la présence d'un accident ou un embouteillage sur la route.

Dans ce qui suit nous traitons le cas d'une requête Q_2 permettant d'être informé des dangers présents sur la route et que le conducteur risque de rencontrer. Q_2 doit être évaluée de manière continue avec un pas correspondant à la zone d'anticipation :

```
Requête  $Q_2$ _____
SELECT cellule_physique, distance, type_événement,
nombre_occurrences/nombre_semaines
FROM RESUME
WHERE cellule_physique IN TRAVERSEE(current_position(), distancemax)
and CORRESPONDS (temporalité, current_date()) and type_événement IN
('accident', 'freinage') and nombre_occurrences/nombre_semaines >= seuil-
FrequenceMax and nombre_semaines >= seuilSemaineMin
CONTINUOUS
```

Ici les cellules sont filtrées par leur temporalité (fonction CORRESPONDS) et au niveau spatial par la fonction TRAVERSEE qui sélectionne les cellules qui vont être traversées par le véhicule à une distance maximale donnée. Cette fonction suppose que nous connaissons la direction de véhicule. Cette requête doit être réévaluée à la fois pour tenir compte du déplacement du véhicule et

aussi des modifications éventuelles du résumé.

Pour ce type de requêtes, les réponses attendues doivent être précises en terme de localisation (nous ne nous intéressons qu'aux zones que nous allons effectivement traverser et pas aux zones proches). Nous pouvons considérer que toutes les requêtes d'alertes sont définies une fois pour toutes au moment de l'installation du système d'assistance au conducteur.

6.4 Niveau d'assistance aux conducteurs

Nous montrons ici comment utiliser les informations brutes renvoyées par « le niveau d'accès aux données ». Ces résultats vont être interprétés puis le système utilise cette interprétation pour la recommandation au conducteur.

6.4.1 Recherche de ressources

Le premier niveau nous renvoie un ensemble de résultats positifs et négatifs.

Une première exploitation directe de ces résultats serait de ne considérer que les cellules les plus proches (à distance de 1) en les triant par ordre décroissant de fréquence d'occurrences. Ces résultats sont représentés dans l'exemple de la figure 6.2.

Dans notre exemple, la zone haut gauche (HG) a la fréquence la plus élevée (10) et est donc recommandée au conducteur.

La liste triée permet à l'utilisateur de savoir quelles sont les zones où la probabilité de trouver une place disponible à cette instant est presque égale à 0. Dans notre exemple la zone bas droit est à éviter puisque sa fréquence selon le résumé est égale à 0.

Cependant, comme nous ne disposons pas d'informations en temps réel mais d'estimations basées sur de simples résumés du passé, il est intéressant de raisonner non pas uniquement sur les cellules les plus proches, mais aussi sur des régions plus larges qu'une cellule. Par exemple, sur la figure 6.3, nous

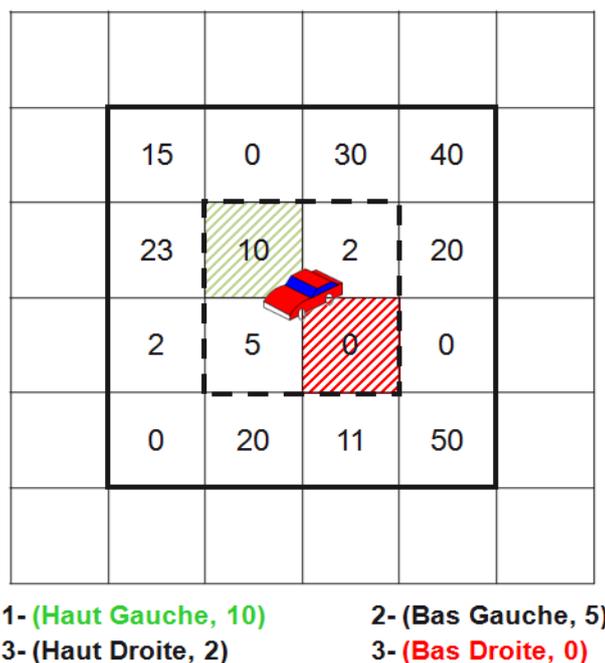
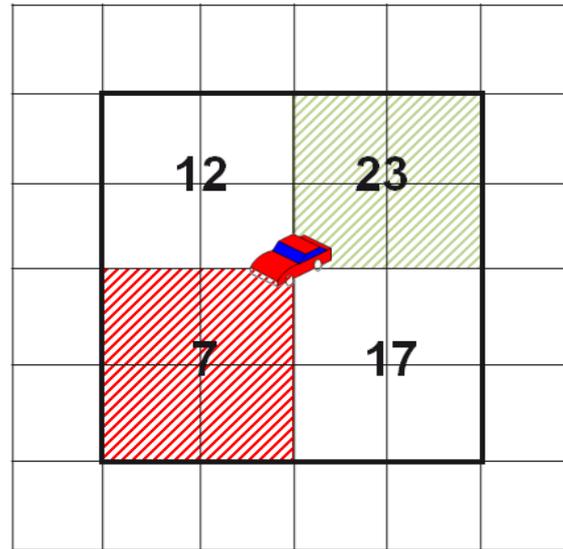


FIGURE 6.2 – Cellules à distance 1

raisonnons sur les carrés de côté 2 autour du véhicule. Cela permet aussi de donner plus de chances au conducteur de trouver une ressource disponible sans changer de direction (ce que ne garantit pas une recherche basée sur une distance de une cellule).

Dans cette deuxième solution nous associons à chaque carré la moyenne des valeurs des cellules constituantes. Ces moyennes sont renvoyées, comme montré sur le premier résultat, dans une liste triée dans l'ordre décroissant montrée dans la figure 6.4. En suivant le même principe de recommandation que celui où nous considérons les cellules à distance 1, la région à recommander est celle en haut à droite (valeur moyenne la plus élevée 23), et celle à éviter en bas à gauche (valeur moyenne la plus basse 7).

Nous pouvons également appliquer un autre mode de calcul qui prend en compte la distance que nous avons à faire pour arriver sur les cellules recommandées. Sur notre exemple cela donne l'ensemble de triplets suivants triés en ordre croissant des places de stationnement libres dans chaque cellule



1- (Haut Droite, 23, 1) 3- (Haut Gauche, 12, 1)
 2- (Bas Droite, 17, 1) 4- (Bas Gauche, 7, 1)

FIGURE 6.3 – Moyenne des cellules voisines

physique :

1. (Haut Droite, 30, 2);
2. (Bas Droite, 20, 2);
3. (Haut Gauche, 12.5, 2);
4. (Haut Gauche, 10, 1);
5. (Bas Gauche, 7, 2);
6. (Bas Gauche, 5, 1);
7. (Haut Droite, 2, 1);
8. (Bas Droite, 0, 1).

Par exemple, le quatrième triplet indique que dans la direction haut gauche (premier élément du triplet), à une distance de une cellule (troisième élément), nous avons observé 10 occurrences (deuxième élément). Le premier triplet indique dans son premier élément la direction « haut droite », à une distance de deux cellules (troisième élément), nous avons observé une moyenne de 30

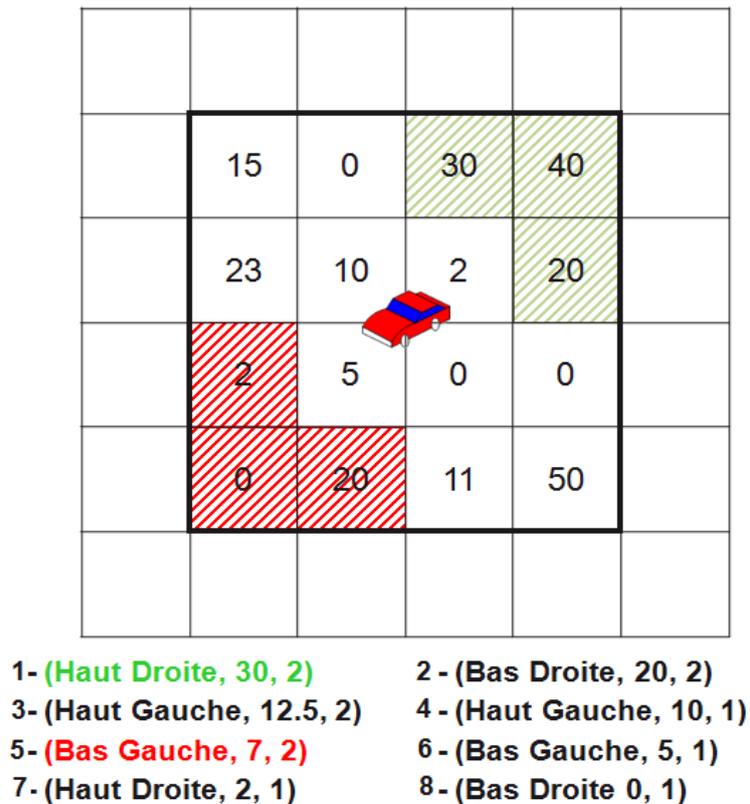


FIGURE 6.4 – Cellules à distance 1 et plus

occurrences (30+40+20 divisé par 3) représenté par le deuxième élément du triplet.

Nous avons généralisé cette approche à la distance souhaitée. Sur cette base, nous pouvons alors construire divers ordres, comme par exemple un ordre basé sur une combinaison des distances donnant plus d'importance aux cellules les plus proches. La région sélectionnée est alors donnée par l'équation 6.1.

$$\text{Région sélectionnée} = \text{Max} (2 \times (\text{valeur à distance 1}) + \text{valeur à distance 2}) \quad (6.1)$$

Dans notre exemple nous avons :

$$1. \text{ HG} = 2 \times 10 + 12.5 = 32.5$$

2. $HD = 2*2 + 30 = 34$
3. $BG = 2*5 + 7 = 17$
4. $BD = 2*0 + 20 = 20$

Ce qui nous permet de recommander positivement HD (plus éventuellement HG) et de ne recommander négativement aucune région.

6.4.2 Production d'alertes

Les résultats fournis par les requêtes continues du type Q_2 vont permettre de produire directement des alertes aux conducteurs.

Nous pouvons prendre l'exemple des « zones dangereuses » qui représente l'événement prioritaire que le conducteur veut recevoir. L'idée est de pouvoir faire une interface de type « feu tricolore » (rouge : zone dangereuse, orange : zone potentiellement dangereuse, vert : zone a priori sans danger connu) montrée dans la figure 6.5.

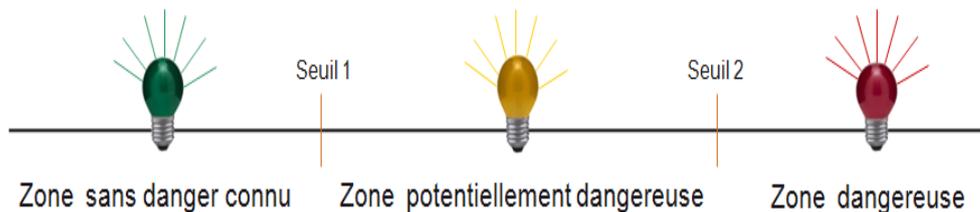


FIGURE 6.5 – Interface d'alerte des zones dangereuses

Il faut cependant décider des seuils délimitant la frontière entre les trois couleurs. Plutôt que de fixer des seuils absolus, il paraît plus intéressant de les définir de manière relative en cherchant à discriminer les valeurs élevées et basses.

Nous proposons ici de discriminer les 5% de cellules ayant les valeurs les plus élevées (respectivement basses). Cette information peut être dynamique en fonction des évolutions des valeurs dans le résumé ou maintenue statiquement.

6.5 Conclusion

Dans ce chapitre, nous avons assuré le troisième objectif de notre thèse en proposant notre approche à deux niveaux pour l'interrogation des résumés et l'interprétation des résultats.

Le premier niveau appelé « niveau d'accès au résumé » considère le résumé comme une base de données que nous allons interroger avec des requêtes spatio-temporelles évaluées classiquement dans le cas de recherche des ressources ou évaluées en mode continu dans le cas d'alertes.

Les résultats obtenus dans « le niveau d'accès au résumé » sont transmis au « niveau d'interprétation des résultats » pour proposer au conducteur une aide à la décision. Nous avons décrit plusieurs méthodes d'assistance dans le cas de la recherche des ressources qui se distinguent essentiellement par la taille de leur fenêtre de recherche (une cellule ou une zone composée de plusieurs cellules). Pour les alertes, l'assistance se résume à l'établissement de seuils.

Cette approche d'exploitation de résumés a été publiée dans [Zekri 2010a].

Simulation et validation expérimentales

Sommaire

7.1	Introduction	93
7.2	Choix du simulateur	94
7.2.1	Les simulateurs existants	94
7.2.2	Simulateur « VESPA »	96
7.2.3	Adaptation du simulateur	97
7.3	Paramètres de simulations	98
7.4	Critères et stratégies utilisés dans les simulations	100
7.5	Évaluation qualitative du résumé spatio-temporel	102
7.6	Évaluation du processus d'échange	106
7.7	Dynamique des échanges	108
7.8	Conclusion	114

7.1 Introduction

Ce chapitre présente les implémentations et les expérimentations réalisées pour valider les solutions présentées dans ce manuscrit. Notre objectif est de valider de manière expérimentale les différentes contributions. Dans les expérimentations, même si notre proposition d'agrégation peut contenir plusieurs types d'événements, nous avons choisi dans ce chapitre de nous focaliser sur les places de stationnements

Dans un premier temps, nous choisissons un simulateur et nous procédons à une étape de déploiement de nos modules implementés afin d'adapter le simulateur à nos besoins en terme de construction de résumés, d'échange et d'exploitation. Ensuite nous présentons les résultats obtenus grâce à nos implémentations afin de valider nos solutions.

Le reste de ce chapitre est organisé comme suit. Dans la section 7.2, nous exposons une étude des simulateurs existants qui nous a permis de choisir le simulateur le plus approprié décrit dans la section 7.2.2. Ensuite nous allons montrer comment nous avons adapté ce simulateur à nos besoins dans la section 7.2.3. Pour préparer nos simulations nous présentons dans la section 7.3 les paramètres évalués et dans la section 7.4 les stratégies utilisées. Enfin, les deux sections 7.5 et 7.6 montrent les différents résultats obtenus.

7.2 Choix du simulateur

Dans cette section, nous présentons le simulateur que nous avons choisi pour évaluer notre structure d'agrégation proposée dans le chapitre 4 et le processus d'échange dans le chapitre 5.

7.2.1 Les simulateurs existants

Les performances des applications destinées aux réseaux inter-véhiculaires sont souvent évaluées grâce à des simulateurs. Deux composants importants sont à prendre en compte pour cela :

1. D'un coté, nous devons considérer des modèles de mobilité réalistes. Dans la réalité, les véhicules ne se déplacent pas aléatoirement. Chaque automobiliste suit une route bien précise en fonction de l'endroit où il veut se rendre. De plus, le réseau routier est un environnement fortement contraint où les chemins, les vitesses, les ordres de passages sont réglementés. Pour cela, il existe un certain nombre de simulateurs de

trafic. Certains sont commerciaux comme *CORSIM*¹ ou *TRANSIMS*² et d'autres sont libres comme *VanetMobiSim* [Fiore 2007] ou *GrooveNet*³ [Mangharam 2006];

2. D'un autre côté, nous devons considérer les simulateurs généraux de réseaux. Les plus connus sont *NS-2*⁴ (qui ne permet pas de visualiser le résultat des expérimentations. Il permet uniquement de stocker une trace de la simulation, de sorte qu'elle puisse être exploitée par un autre logiciel), *GloMoSim*⁵ (il est écrit en Parsec, langage dédié basé sur le C), ou *JiST-SWANS*⁶. Quelques travaux proposent d'intégrer les deux, comme *TraNS*⁷.

Le simulateur choisi doit répondre à certains critères propres à nos besoins. En effet, le simulateur doit nous permettre de représenter l'environnement, de définir différentes routes pour les véhicules, de simuler des véhicules sans destination précise (pour simuler la recherche de place de stationnement), d'intégrer différentes vitesses et différentes conditions de circulation. Lorsqu'un événement survient sur la route, il est nécessaire de pouvoir identifier les véhicules qui ont rencontré l'événement et à quel moment. De plus, le comportement des conducteurs doit pouvoir être influencé par les événements reçus. Le partage d'information implique une exploitation de ces informations dans le simulateur. Le comportement des automobilistes doit être fortement lié aux informations transmises par le système simulé. Par exemple, si un conducteur reçoit un événement sur un embouteillage, il doit essayer de modifier sa route. De même, lorsqu'un conducteur cherche une place de stationnement, il doit circuler aléatoirement dans la zone de recherche et se diriger en fonction des événements concernant des libérations de place de stationnement lorsqu'il en reçoit. La simulation du réseau et la simulation du trafic ne peuvent donc pas être indépendantes. Elles s'influencent mutuellement puisque d'une part, les informations reçues doivent modifier les itinéraires des véhicules et d'autre

1. <http://mctrans.ce.ufl.edu/featured/tsis/version5/corsim.htm>

2. <http://transims.tsasa.lanl.gov/>

3. <http://www.seas.upenn.edu/~rahulm/Research/GrooveNet/>

4. <http://www.isi.edu/nsnam/ns/>

5. <http://pcl.cs.ucla.edu/projects/glomosim/>

6. <http://jist.ece.cornell.edu/>

7. <http://trans.epfl.ch/>

part, la modification des itinéraires provoquent des variations dans la topologie du réseau.

Une difficulté supplémentaire liée à l'utilisation d'un simulateur de la liste citée ci-dessus est la récupération des résultats spécifiques.

Vu les contraintes énumérées et nos besoins pour nos évaluations, nous avons besoin d'un simulateur qui assure le couple « réseau et trafic ». Nous avons choisi le simulateur « *VESPA* » que nous avons adapté pour qu'il corresponde aux critères propres à nos besoins. Dans ce qui suit nous allons décrire le simulateur « *VESPA* » et l'extension réalisée.

7.2.2 Simulateur « *VESPA* »

Ce simulateur a été développé au laboratoire LAMIH, équipe Systèmes Mobiles et Embarqués (SyME), de l'université de Valenciennes. Ce simulateur est développé en Java. Il est utilisé pour simuler des déplacements de véhicules sur un réseau routier et pour évaluer différents protocoles de routage dans différentes conditions de circulation afin d'étudier leurs impacts sur le trafic.

Le simulateur *VESPA* [Cenerario 2010] nous a permis de simuler des contextes réalistes. Pour construire notre réseau routier, nous avons deux possibilités :

1. Extraction de données cartographiques d'une base de données Télé Atlas ;
2. Création manuelle des routes.

De cette manière, nous pouvons utiliser des cartes réelles pour nos simulations sur une autoroute ou dans un centre ville ou bien nous pouvons créer nos propres configurations pour des cas particuliers (par exemple les parkings d'hypermarchés qui ne sont pas cartographiés).

Cela nous a permis d'obtenir un modèle de mobilité proche de la réalité et d'éviter un déplacement rectiligne des véhicules. Un ensemble de routes avec différents points d'intersection peuvent ainsi être créés. Chaque route possède sa propre limitation de vitesse.

Pour les parkings de supermarché, le « *simulateur VESPA* » possède un

générateur de configuration de routes qui nous permet d'avoir un environnement proche de la réalité. Chaque véhicule suit donc ces routes en partant d'un point de départ pour rejoindre un point d'arrivée.

Pour déterminer le chemin du véhicule, le plus court chemin est utilisé. Pour éviter d'avoir des trains de véhicules trop régulier (ce qui ne serait pas réaliste), nous pouvons associer à chaque véhicule une nervosité. Cette nervosité représente la manière de conduire du conducteur. Ainsi, pendant que certains véhicules roulent moins vite que la limite autorisée, d'autres peuvent faire des excès de vitesse. Cela permet d'avoir un comportement plus proche de la réalité. De plus, pour les variations de vitesse, les véhicules accélèrent et décélèrent de manière souple. En utilisant une fonction d'accélération et de freinage, nous parvenons à imiter le comportement des véhicules sur les routes.

Pour l'émission et la réception des informations, nous considérons que lorsqu'un message est envoyé, tous les véhicules suffisamment proches reçoivent l'information (cela dépend de la portée de communication choisie).

7.2.3 Adaptation du simulateur

Pour évaluer notre structure d'agrégation et notre processus d'échange, nous nous sommes concentrés sur un seul type d'événements. Nous avons choisi d'évaluer la valeur ajoutée de notre processus d'agrégation sur les véhicules à la recherche d'une place de parking disponible. Pour se faire, nous avons étendu le simulateur « *VESPA* » avec des modules permettant de construire, échanger et exploiter des résumés/agrégats d'événements :

1. le module de construction de résumé est implémenté en Java et représente 1382 lignes de code. Nous avons utilisé la fonction de hachage SHA-2 pour construire les sketches des événements et des semaines et construire notre structure d'agrégation présentée dans le chapitre 4. Pour modéliser les cellules physiques, nous avons défini le point d'origine de simulation et la valeur du côté du carré représentant la cellule physique ;
2. Dans l'échange, nous nous sommes basés sur le mécanisme de dissémination dans « *VESPA* » qui permet à un véhicule de communiquer

- avec les véhicules à proximité (dans un cercle dont le rayon est égal au rayon de communication défini selon le réseau sans fil utilisé). Nous avons intégré un module (574 lignes de code) pour la définition des priorités de chaque véhicule que nous allons prendre en considération lors de l'échange des résumés. Enfin nous avons modifié les entrées des modules assurant l'échange de façon à prendre le résumé et les priorités définies dans le chapitre 5 comme paramètres d'entrée ;
3. Dans l'exploitation, nous avons ajouté un module (355 lignes de code) pour exploiter les résumés et pour envoyer au conducteur l'aide pour la prise de décision. Dans ce module nous n'avons expérimenté que la première stratégie d'exploitation des résultats dans le cas de recherche de ressource expliquée dans la section 6.4.1 du chapitre 6 (recommandation de la cellule la plus proche ayant la plus forte fréquence d'occurrences d'événements).

La figure 7.1 montre l'interface graphique du simulateur où nous considérons deux parkings, chacun étant dans une cellule physique. Les voitures sont colorées en fonction de leurs états. Les véhicules en recherche sont initialement jaunes et s'ils trouvent une place pour se garer, ils deviennent rouges. Les véhicules qui dépassent leur temps de recherche sans trouver une place libre deviennent gris. Cela nous permet d'évaluer, visuellement, la qualité de notre système.

7.3 Paramètres de simulations

Dans nos simulations, nous avons considéré un environnement correspondant au centre de Valenciennes, une ville dans le nord de la France. Cette zone a été représentée par 64 cellules physiques, où chaque cellule physique est un carré de côté de 200 mètres. Pour évaluer notre processus d'agrégation, nous avons placé 8 parkings dans la ville. Chaque parking est situé dans une cellule physique. En outre, chaque parking présente une capacité prédéfinie et un taux de remplissage indiqué dans le tableau 7.1.

Les véhicules dans les simulations se déplacent à 30kmh et dès qu'ils entrent

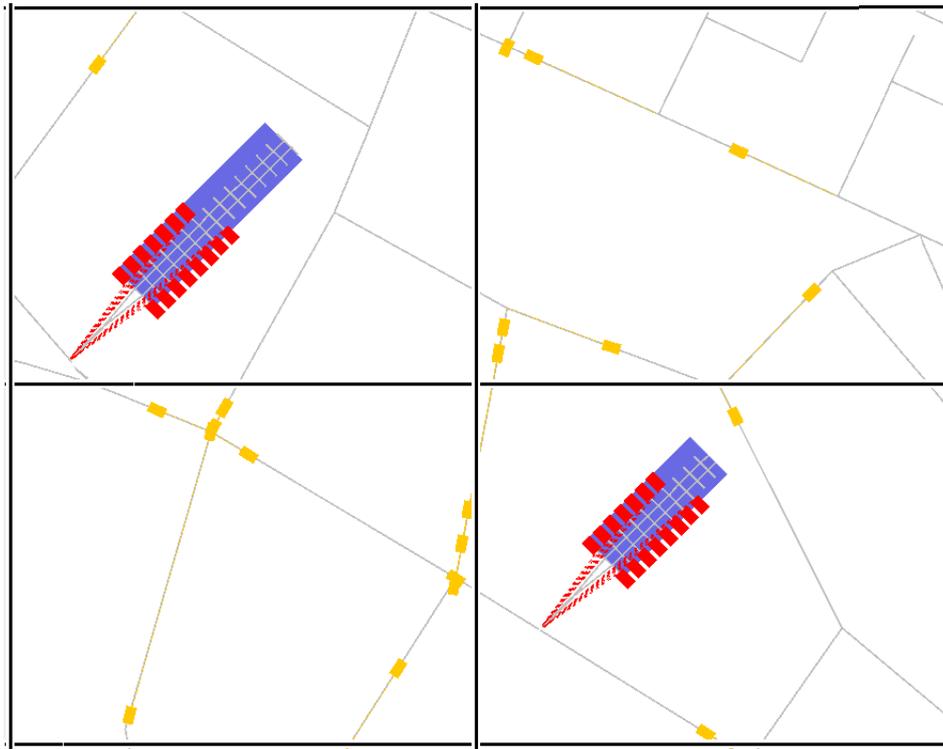


FIGURE 7.1 – Interface graphique du simulateur

Paramètres	Initialisation
Nombre de cellules physiques	64
Nombre de parkings	8
Nombre de places/parking	P1=50, P2=20, P3=40, P4=50, P5=50, P6=20, P7=50, P8=40
Nombre de véhicules/h	100
Remplissage initial /parking	P1=70%, P2=85%, P3=80%, P4=66%, P5=76%, P6=65%, P7=64%, P8=75%

TABLE 7.1 – Caractéristiques des 8 parkings considérés durant les simulations

dans un parking leur vitesse diminue à 10 km/h. Chaque heure, N véhicules avec $N = 100$ entrent au centre ville et commencent à chercher une place de parking pendant 1000s. S'ils ne trouvent pas une place de stationnement libre pendant la période de recherche, ils s'arrêtent de chercher mais continuent toutefois à échanger les agrégats et les messages dissimulés avec les autres véhicules jusqu'à ce qu'ils sortent de la simulation (10% des véhicules qui entrent chaque heure quittent la simulation après une heure de simulation).

Une fois qu'un véhicule a trouvé une place de parking, il reste stationné pendant une période (déterminée aléatoirement) allant de 1 heure à 4 heures. Ensuite, le véhicule quitte la place de stationnement et informe les véhicules au voisinage de la libération de cette place.

7.4 Critères et stratégies utilisés dans les simulations

Dans cette section, nous présentons les résultats obtenus pour différentes stratégies. Nous avons étudié l'évolution de deux critères importants lors de la recherche d'une place de stationnement disponible :

- le temps nécessaire à chaque véhicule pour trouver une place de stationnement libre ;
- le pourcentage de véhicules trouvant une place libre pendant la période de recherche déterminée.

Pour la recherche d'une place de stationnement disponible, nous avons considéré plusieurs stratégies élémentaires, et les avons ensuite combinées :

1. « *View* » : cette stratégie est basée sur l'observation directe d'un événement. Notre but ici est de modéliser le comportement classique d'un conducteur à la recherche d'une place de parking disponible et qui se gare quand il voit une place de stationnement vide (sans aucune assistance particulière) ;
2. « *Dissemination* » : cette stratégie ne considère que les messages diffusés par un véhicule qui quitte sa place de parking. Les véhicules reçoivent des messages disséminés par d'autres véhicules et ils seront ensuite guidés vers cette place libre. Cette stratégie n'utilise donc pas les résumés. Cette stratégie permet de guider les véhicules vers des places de stationnement libres. Elle peut toutefois induire de la compétition entre les véhicules ;
3. « *Infrastructure* » : cette stratégie prend en compte l'information fournie par une infrastructure (par exemple un serveur central stockant tous les événements survenant à proximité) pour les conducteurs. Cette information consiste en un ensemble de statistiques fiables sur la fréquence

de tous les événements et l'ensemble des zones concernées. Cette stratégie est mis en œuvre dans le simulateur par l'utilisation d'une structure d'agrégation de données spatio-temporelle complète, contenant par l'ensemble des événements observés sur toute la zone d'intérêt (autrement dit le rayon de visibilité : représentant la distance sur laquelle un véhicule peut voir les événements produits, égale au rayon de la zone d'intérêt). L'objectif de cette stratégie est d'évaluer une structure d'agrégation « parfaite » puisque contenant tous les événements survenus dans toute la zone ;

4. « *SummaryAggregation* » : cette stratégie prend en compte la structure d'agrégation des événements spatio-temporelle, mais uniquement au niveau de chaque véhicule. Cette structure d'agrégation intégrée est remplie par les événements observés par le véhicule depuis le début de la simulation dans un voisinage de portée maximale fixe (autrement dit un cercle autour du véhicule). Cette portée fixe est un paramètre de la simulation. Nous avons utilisé les valeurs de 50% ou 25%, désignant respectivement un rayon correspondant à la moitié (respectivement un quart du rayon de la zone d'intérêt du conducteur). Avec cette stratégie, le conducteur est guidé vers la zone la plus proche ayant la plus haute fréquence de places de stationnement libérées. La stratégie « *SummaryAggregation* » peut être utilisée avec ou sans échanges entre les véhicules. Pour évaluer l'efficacité de notre structure d'agrégation nous utilisons « *SummaryAggregation* » sans échange (un véhicule n'obtient aucune information des autres). Pour évaluer l'efficacité du processus d'échange d'agrégats nous utilisons la même stratégie « *SummaryAggregation* » avec échange où le véhicule échange tout ou une partie de sa structure d'agrégation avec d'autres véhicules situés dans un rayon de 100m de la position courante ;
5. « *View + Dissemination* » : cette stratégie combine l'observation directe et la stratégie de dissémination. Un véhicule utilise la dissémination en premier pour aller vers une place de stationnement libre (sous réserve qu'aucun autre véhicule ne l'atteigne avant). Le conducteur peut choisir une place libre rencontrée sur le chemin le conduisant vers la place proposée la suite de la dissémination ;

6. « *View + Dissemination + SummaryAggregation* » (resp. *View + Dissemination + Infrastructure*) : cette stratégie combine « *View* », « *Dissemination* » et « *SummaryAggregation* » (resp. « *Infrastructure* »). Ainsi, lorsqu'un conducteur cherche une place de stationnement libre, s'il n'observe aucune place disponible sur son chemin et ne reçoit pas de message d'un autre véhicule libérant sa place, la structure d'agrégation spatio-temporelle de données (ou l'infrastructure) est utilisée pour recommander la meilleure zone où rechercher une place libre.

Pour les stratégies utilisant la structure d'agrégation (« *SummaryAggregation* » et « *View + Dissemination + SummaryAggregation* »), une phase d'initialisation de la structure d'agrégation précède les simulations. Il s'agit de faire une simulation de 24 heures pour remplir la structure en respectant la portée d'observation dans les zones d'intérêt (une vue partielle de 25% ou de 50% de la zone ou une vue globale de 100% de la zone). Une vue globale agrège tous les événements produits dans la simulation, alors qu'une vue partielle de $X\%$ n'agrège que les événements observés dans un rayon de $X\%$ de la portée globale autour du véhicule.

Une vue partielle de $X\%$ de la zone d'intérêt correspond à l'observation que le conducteur s'intéresse dans sa collecte d'information à une partie de la zone d'intérêt avec un rayon de vision égal à $X\%$ du rayon de la zone considérée. Ensuite, après la phase d'initialisation, cette structure est mise à jour lors des simulations réalisées pour chacune des stratégies utilisant des paramètres différents. Les résultats présentés dans la suite ont été obtenus en calculant la moyenne de 10 simulations pour chaque stratégie.

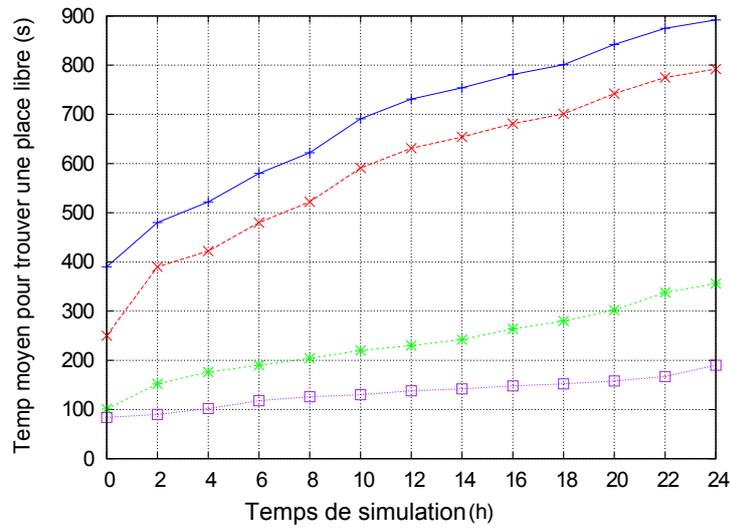
7.5 Évaluation qualitative du résumé spatio-temporel

Notre premier objectif est de mettre en évidence l'intérêt des agrégats spatio-temporels pour les véhicules à la recherche d'une place de stationnement disponible. A ce niveau de simulation nous allons comparer les résultats relatifs aux deux paramètres : le temps nécessaire pour trouver une place de

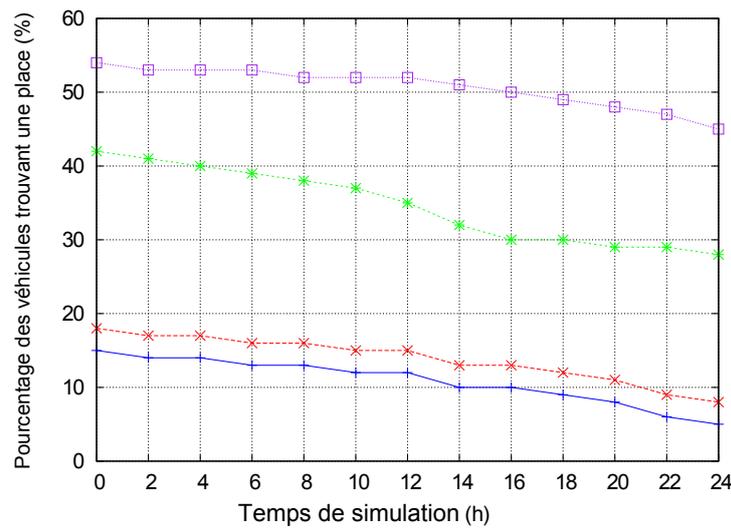
stationnement libre et le pourcentage de véhicules trouvant une place libre, obtenus dans l'évaluation des stratégies sans échange et avec échange. La figure 7.2 montre les résultats relatifs à la durée moyenne nécessaire à un véhicule pour trouver une place de parking et le pourcentage de véhicules qui ont trouvé une place libre, produits par les stratégies : « *View* », « *Dissemination* » et « *Infrastructure* ».

Dans la figure 7.2, nous pouvons observer que « *Infrastructure* » donne la borne supérieure de ce que nous pouvons obtenir avec notre approche (car correspond à un résumé « parfait » sur chaque véhicule) et que « *View* » représente la borne inférieure (aucune information n'est apportée au conducteur). Nous observons tout d'abord que quelle que soit la stratégie utilisée, le temps moyen pour trouver une place de stationnement disponible augmente au cours du temps. De la même façon, le pourcentage de véhicules trouvant une place libre pour se garer diminue avec le temps. La raison de ces deux observations, c'est que le nombre de véhicules entrant dans la simulation à la recherche d'une place de stationnement disponible est supérieur au nombre de places libres (selon les paramètres initiaux définis pour les simulations et présentés dans le tableau 7.1). En effet, le nombre de places de stationnement libérées sur les 8 parkings est d'au moins 20% inférieur au nombre de véhicules à la recherche d'une place de parking disponible. Par conséquent, il y a un problème de famine qui devient de plus en plus important au cours du temps de simulation et le pourcentage de véhicules trouvant une place disponible ne peut pas atteindre 100%. La raison pour laquelle nous avons choisi de présenter les résultats d'un tel environnement encombré est que les systèmes d'assistance devraient être particulièrement efficaces dans de telles configurations où il est très difficile pour les conducteurs de trouver une place pour se garer.

Les résultats présentés dans la figure 7.2 montrent que la stratégie d'« *Infrastructure* » donne des résultats significativement meilleurs que les stratégies « *View* » ou « *View + Dissemination* » montrant l'intérêt de la structure d'agrégation des événements. Ces observations sont valables aussi bien en considérant le temps moyen pour trouver une place de parking (figure 7.2(a)) que le pourcentage de véhicules trouvant une place de parking (figure 7.2(b)). En outre, « *View + Dissemination + Infrastructure* » est la stratégie ayant les meilleurs résultats ce qui montre que les stratégies élémentaires corres-



(a) Temps moyen



(b) Pourcentage de véhicules

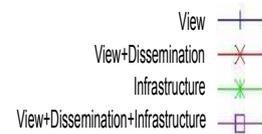
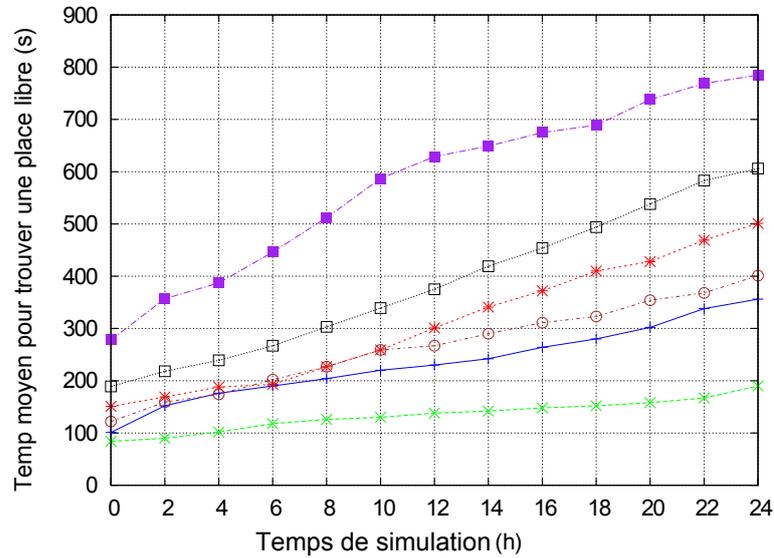


FIGURE 7.2 – Mesures de qualité pour les différentes stratégies

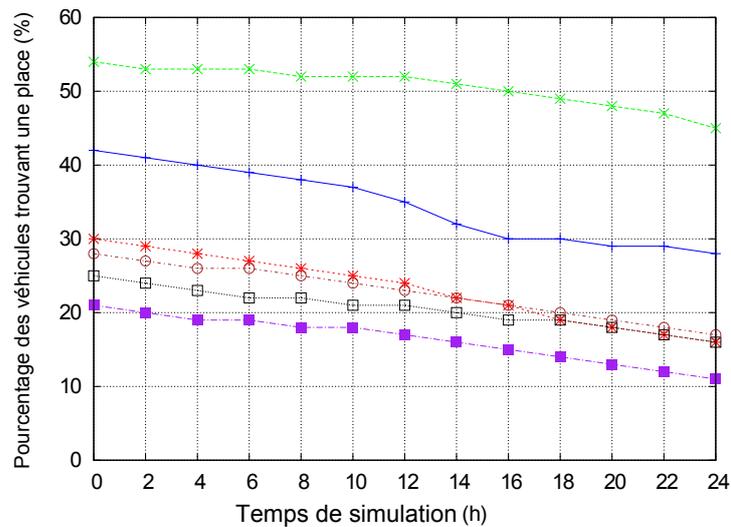
pondantes sont complémentaires.

Dans la figure 7.3, nous introduisons le processus d'agrégation partielle au

niveau de chaque véhicule (par exemple, la stratégie « *SummaryAggregation* ») et le comparons avec les stratégies étudiées dans la figure 7.2.



(a) Temps moyen



(b) Pourcentage de véhicules

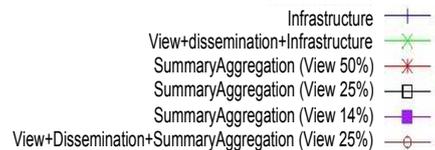


FIGURE 7.3 – Impact de la portée d'observation

Ici, nous n'avons pas tenu compte de l'échange de résumés entre les vé-

hicules, les avantages du processus d'échange sur la qualité de la structure d'agrégation seront évalués ultérieurement dans la section 7.6. Nous avons analysé l'impact du paramètre « portée de l'observation ». Nous pouvons noter que la stratégie d'« *Infrastructure* » peut être considérée comme la « *SummaryAggregation* » avec une portée de vision égale à 100% du rayon de la zone d'intérêt (c'est à dire, la structure d'agrégation contient tous les événements produits dans la zone d'intérêt).

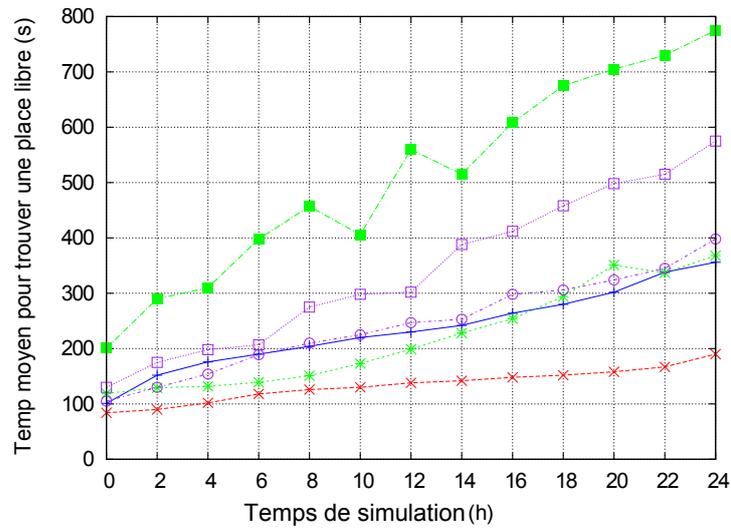
Si nous comparons les résultats du « *SummaryAggregation* » avec une portée de 50% avec ceux obtenus pour la stratégie de l'« *Infrastructure* », nous constatons que même si la « qualité » de la structure des agrégats est divisée par 2, le temps moyen et le pourcentage des véhicules trouvant une ressource ne varient pas dans la même proportion qui est plutôt proche de 1,5. Cela montre donc que notre approche apporte un gain, même avec une qualité moindre des résumés.

Dans cette section nous avons évalué notre structure d'agrégation. Une évaluation du processus d'échange proposé dans le chapitre 5 est présentée dans la section suivante.

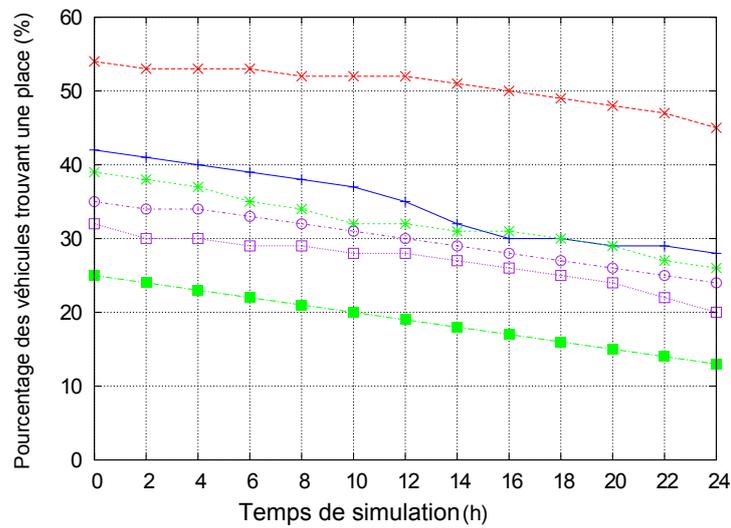
7.6 Évaluation du processus d'échange

Dans cette section, nous évaluons l'impact de l'échange des résumés sur la qualité des agrégats produits (et donc l'efficacité des prédictions faites avec ces agrégats). Nous avons fixé le paramètre de la portée de vision à 50% pour la stratégie « *SummaryAggregation* » et nous comparons ici les résultats avec et sans échanges d'agrégats. L'échange se fait entre deux véhicules se rencontrant dans une portée de 100 mètres et nous supposons qu'il n'y a pas d'interruption (donc pas de perte d'informations) lors du transfert des résumés.

La figure 7.4 montre que la réalisation d'échanges d'agrégats entre les véhicules améliore considérablement les résultats obtenus avec la stratégie « *SummaryAggregation* ». Les résultats sont alors très proches de ceux obtenus avec la stratégie « *Infrastructure* ». Les simulations avec échanges montrent que de bons résultats peuvent être obtenus avec les stratégies basées sur « *Sum-*



(a) Temps moyen



(b) Pourcentage de véhicules

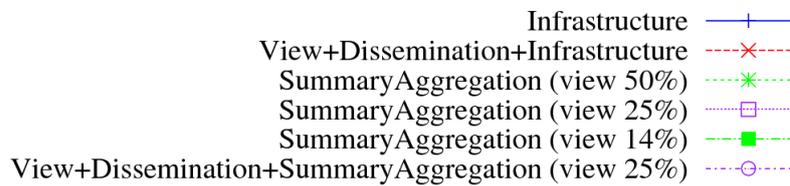


FIGURE 7.4 – Impact du processus d'échange

maryAggregation » même avec une faible portée de vision. En effet avec une

portée de vision égale à 14% en utilisant la stratégie « *SummaryAggregation* », les résultats des simulations montrés dans la figure 7.4(b) sont presque égaux à ceux obtenus avec une portée de vision égale à 25% avec la même stratégie mais sans échange (comme montré par la figure 7.3(b)).

Cette approche coopérative se compare ainsi avec une approche centralisée, comme celle définie par la stratégie « *Infrastructure* ».

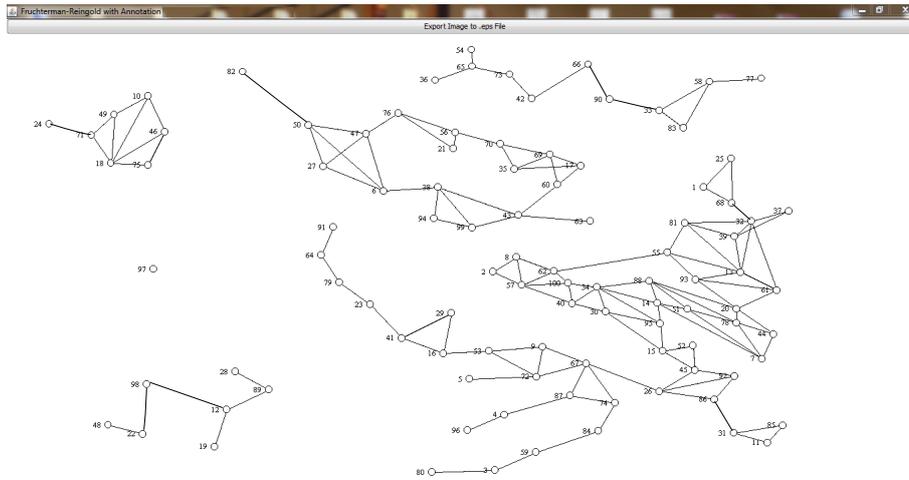
7.7 Dynamique des échanges

Nous analysons ici la dynamique des échanges entre les véhicules. Nous voulons répondre à trois questions principales (i) est ce que le processus d'échange permet à deux véhicules éloignés d'échanger entre eux (ii) comment le processus d'échange évolue t'il dans le temps (iii) quel est l'impact du nombre d'échanges sur la qualité de l'aide au conducteur ? Pour cela nous avons réalisé d'autres séries de simulations. Nous reprenons le même environnement correspondant au centre de Valenciennes considéré dans les simulations précédentes avec les mêmes paramètres définis dans le tableau 7.1. Comme dans les simulations précédentes une phase d'initialisation de la structure est réalisée (remplissage de la structure d'événements pendant 24h). La stratégie utilisée d'aide au conducteur est « *SummaryAggregation* » avec portée de vision à 50%. Nous avons essentiellement fait varier deux paramètres, à savoir la durée de la simulation pour comparer l'évolution des échanges dans le temps (nous avons utilisé deux valeurs, une heure et deux heures) et la portée de l'échange ou distance maximale entre deux véhicules pour qu'ils puissent échanger (nous avons utilisé également deux valeurs, 100 mètres et 50 mètres). La variation sur la valeur de la portée d'échange permet d'évaluer soit la qualité du lien de communication, soit la taille globale de l'espace considéré (lorsque la valeur est de 50 mètres, nous pouvons considérer que cela double la taille de l'espace relativement à une portée de 100 mètres).

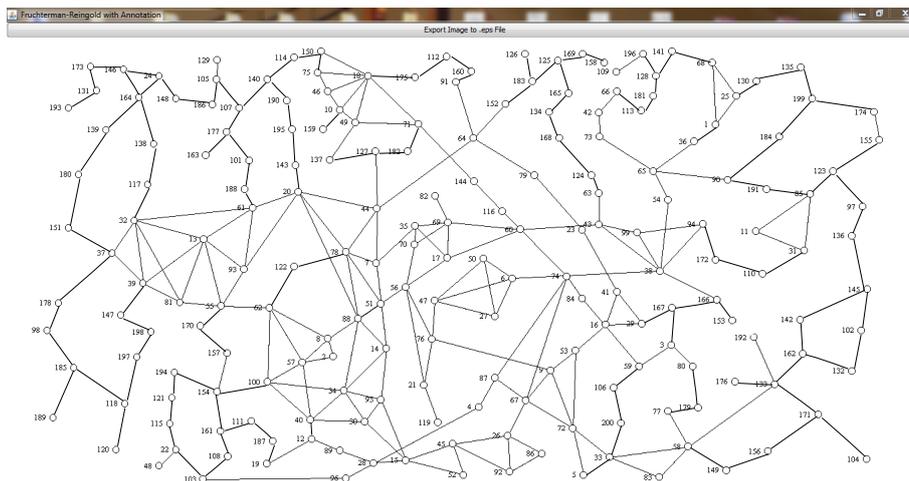
Les figures 7.5(a) et 7.5(b) montrent les échanges entre 100 (resp. 200) véhicules dans 64 cellules physiques respectivement pendant une heure puis au bout de deux heures de simulations pour une distance maximale d'échange de 100 mètres. Les figures 7.6(a) et 7.6(b) montrent la même chose pour une

distance maximale d'échanges de 50 mètres. Nous avons choisi de représenter ces échanges sous forme d'un graphe non orienté dont les noeuds représentent les véhicules et les arêtes les échanges directs. Pour des raisons de clarté, nous avons choisi de ne représenter qu'un seul échange entre deux véhicules, autrement dit si un véhicule a échangé plusieurs fois son résumé avec le même véhicule, ces échanges sont représentés par une seule arête dans le graphe. Ces graphes sont visualisés et analysés avec l'outil « Network Workbench Tool (NWT) » qui prend en entrée un fichier "log" contenant les noeuds avec leurs "Id" et les relations d'échanges et permet de visualiser et analyser le graphe résultant.

La figure 7.6 visualise le graphe obtenu avec une portée de 50 mètres alors que la figure 7.5 visualise le graphe obtenu avec une portée de 100 mètres. L'analyse nous montre que dans les deux cas le nombre de classes de connexité diminue fortement avec le temps (à 100 mètres nous passons de 5 classes à une seule alors qu'à 50 mètres nous passons de 6 à 2). Sans surprise, au fur et à mesure du temps, les véhicules, en se déplaçant, rencontrent de plus en plus de nouveaux véhicules. Au bout d'un moment, nous tendons vers une seule classe de connexité ce qui veut alors dire qu'il y a un échange direct ou indirect entre n'importe quel couple de véhicules. Si nous analysons plus finement les graphes, le graphe à 100 mètres passe de 2000 arêtes en une heure en plus de 7200 en deux heures, ce qui veut dire qu'en doublant le temps nous avons un facteur d'accroissement des arêtes de plus de 3,5. Pour l'arité des noeuds, la dispersion entre les noeuds est assez forte au bout d'une heure (arité maximale à 39 et minimale à 0 pour une moyenne à 22), alors que nous observons un meilleur équilibre entre les noeuds au bout de deux heures (arité maximale à 52 et minimale à 31 pour une moyenne à 41). Le même phénomène est observé pour les graphes à 50 mètres avec un nombre d'arêtes s'accroissant d'un facteur 4,7 à deux heures et une dispersion entre les noeuds se lissant au fil du temps (même si la dispersion initiale est moins forte à 50 mètres qu'à 100 mètres). A 100 mètres, la longueur moyenne du chemin entre deux noeuds de la même classe de connexité passe de 8 (1 heure) à 11 (2 heures). La proximité entre les noeuds est donc bonne même avec une seule classe de connexité. Nous observons à peu près la même chose à 50 mètres (longueur moyenne passe de 5 à 9 avec deux classes de connexité à



(a) Graphe d'échange distance 100 m après 1h de simulation

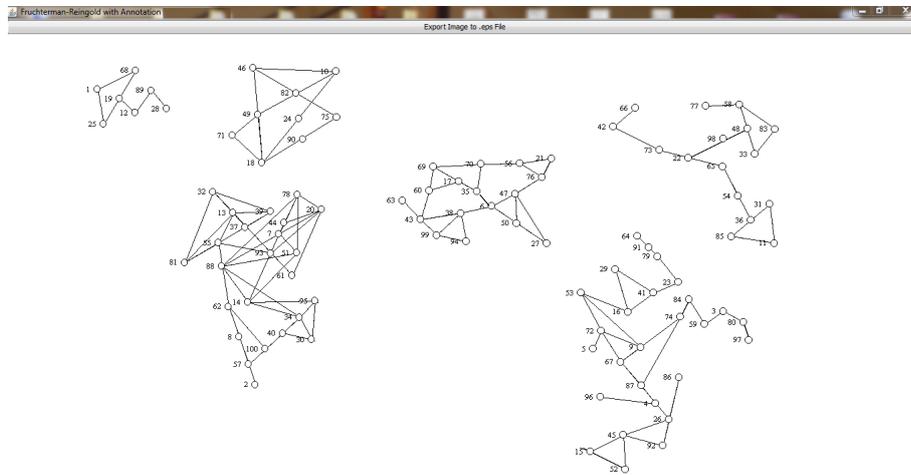


(b) Graphe d'échange distance 100 m après 2h (fin de simulation)

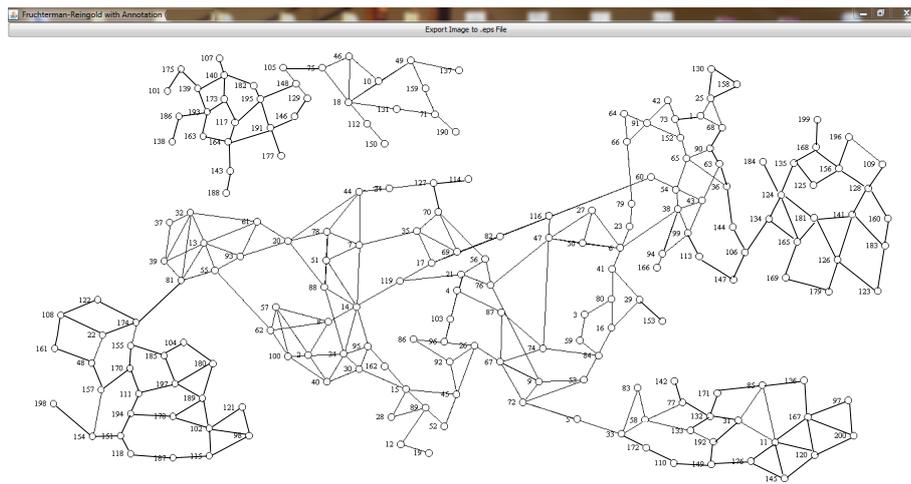
FIGURE 7.5 – Évolution du graphe d'échange distance 100 m pendant la simulation

2 heures). Sans surprise, si nous diminuons la distance d'échange maximale entre deux véhicules le nombre d'échanges diminue dans la même proportion (à une heure, 2014 échanges pour une distance de 100 mètres contre 984 pour une distance de 50 mètres). Par contre, la baisse du nombre d'échanges est moindre à deux heures (nous passons de 7259 à 4275 soit un ratio de 1,7 et non de 2). L'impact de la distance maximale d'échange devrait donc baisser au fur du temps (même si ce n'est pas forcément ce que nous voyons sur la

figure 7.7). L'analyse des deux graphes à 100 mètres de la figure 7.5 est donnée



(a) Graphe d'échange distance 50 m après 1h de simulation



(b) Graphe d'échange distance 50 m après 2h (fin de simulation)

FIGURE 7.6 – Évolution du graphe d'échange distance 50 m pendant la simulation

par le tableau 7.2 alors que pour les graphes à 50 mètres de la figure 7.6 c'est donné par le tableau 7.3.

Il est intéressant aussi de voir si les échanges se font seulement entre des véhicules initialement proches ou bien si deux véhicules initialement éloignés peuvent échanger directement en fonction de leurs déplacements. Nous ana-

	Nb nœuds	heure simulation	Nb échanges total	Arité min	Arité max	Arité moyenne	longueur chemin
Graphe 7.5(a)	100	1h	2014	0	39	22	8
Graphe 7.5(b)	200	2h	7259	31	52	41	11

TABLE 7.2 – Analyse des graphes 7.5(a) et 7.5(b)

lysons ceci par la notion d'échange inter-cellules. Un échange inter-cellule est un échange direct entre deux véhicules qui se trouvaient dans deux cellules différentes au moment de leur entrée dans la simulation.

- Nombre d'échanges « *inter-cellules* » obtenus après une heure pour une distance maximale à 100 mètres égal à 259 (175 pour une distance maximale de 50 mètres) $\simeq 1/7$ nombre total d'échanges (1/6 à 50 mètres) ;
- Nombre d'échanges « *inter-cellules* » obtenus après deux heures pour une distance maximale à 100 mètres égal à 1609 (837 pour une distance maximale de 50 mètres) $\simeq 1/4$ nombre total d'échanges (1/5 à 50 mètres).

Nous voyons donc que la part relative de ces échanges inter-cellules augmente en fonction du temps de simulation et qu'en doublant la durée de simulation leur quantité est multipliée par un facteur 7 à 100 mètres (5 à 50 mètres). Une analyse plus fine serait nécessaire pour comprendre si cet accroissement dépend fortement du modèle de mobilité des véhicules ou non.

	Nb nœuds	heure simulation	Nb échanges total	Arité min	Arité max	Arité moyenne	longueur chemin
Graphe 7.6(a)	100	1h	984	1	24	10	5
Graphe 7.6(b)	200	2h	4275	20	43	27	9

TABLE 7.3 – Analyse des graphes 7.6(a) et 7.6(b)

La figure 7.7 présente une comparaison des résultats relatifs aux critères présentés dans la section 7.4 (temps de recherche et pourcentage de véhicules trouvant une place) et nous ajoutons un critère "le pourcentage des informations utiles" qui représente le pourcentage des places proposées par le système et réellement prises par le conducteur. Ces résultats sont fournis en utilisant la stratégie « *SummaryAggregation* » avec portée de vision à 50% et en prenant un rayon d'échange $R = 100m$ puis $R = 50m$. Les résultats présentés sont la

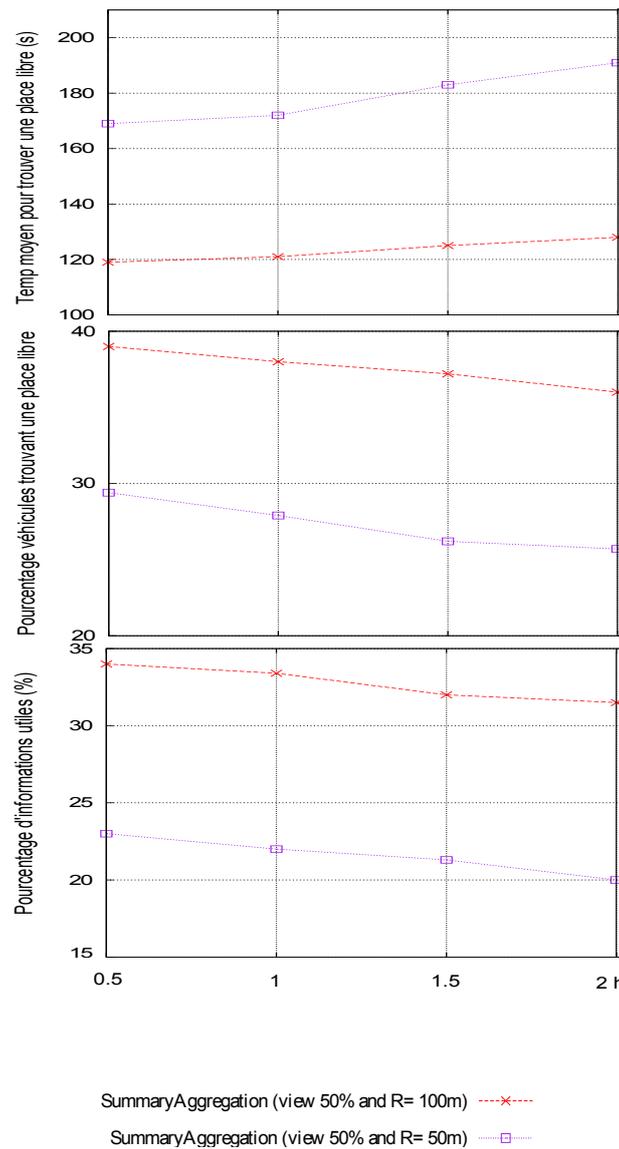


FIGURE 7.7 – Résultats avec diminution du rayon d'échange

moyenne de 10 simulations.

La figure 7.7 montre que la diminution du rayon d'échange a une influence significative et très rapide sur la qualité de l'assistance au conducteur. En effet, au bout d'une demi-heure seulement de simulation l'écart est déjà significatif. Nous pouvons cependant noter que la diminution du rayon (ou distance maximale d'échange) de moitié induit une baisse de moins de la moitié des performances du système (le temps moyen pour trouver une place libre augmente

d'environ 40%, le pourcentage de véhicules trouvant une place libre baisse de 25% environ alors que le pourcentage d'informations utiles baisse de 33% environ). L'évolution des paramètres est ensuite la même entre une demi-heure et deux heures de simulation pour un rayon de 100 mètres comparativement à 50 mètres, même si l'écart augmente légèrement et semble indiquer une dégradation pour un rayon de 50 mètres. Cela montre en tous cas l'importance des échanges dans la qualité des résumés et donc dans la qualité de l'assistance aux conducteurs. Il serait donc intéressant d'étudier d'autres méthodes pour l'échange des résumés que la seule rencontre opportuniste entre deux véhicules (utiliser des véhicules relais ou bien l'infrastructure).

7.8 Conclusion

Ce chapitre a été consacré à la mise en œuvre et la validation de notre approche (qualité des agrégats et apport des échanges). Pour réaliser nos expérimentations, nous avons commencé par choisir un simulateur que nous avons adapté pour simuler nos scénarios. Une étude des simulateurs existants a montré que le simulateur « VESPA » est le plus adapté à notre contexte. Nous l'avons donc étendu par un module d'agrégation qui permet de construire la structure d'agrégation dans chaque véhicule et de l'exploiter ensuite.

Dans une deuxième partie nous avons défini l'environnement de simulation, les critères et les stratégies simulées. Une phase préalable d'initialisation de la structure doit être faite avant chaque simulation.

Dans une troisième partie du chapitre les expérimentations pour valider la structure d'agrégation et le processus d'échange ont été présentées. Elles montrent d'une part que les résumés permettent d'aider les conducteurs à prendre de bonnes décisions sous réserve d'atteindre une bonne qualité des résumés via le processus d'échange. Les études menées sur le processus d'échange et sa dynamique nous ont montrées que nous arrivons facilement à avoir un graphe d'échange fortement connexe (ou presque). Il n'en reste pas moins que le processus d'échange est un élément très important de la qualité des résumés et que pour être sûr de son efficacité il faudrait sûrement le sécuriser en utilisant des véhicules relais (pas directement intéressés par le résumé mais

permettant la continuité d'échange) et/ou de l'infrastructure.

Conclusions et perspectives

8.1 Conclusions

L'objectif principal de ce travail de thèse était de fournir une solution pour l'aide à la décision au conducteur en se basant sur des résumés d'événements dans le cas d'absence d'informations sur les événements de la route en temps réel. Pour assurer cet objectif, nous avons proposé : une structure d'agrégation des événements spatio-temporel dans un environnement routier, un processus d'échange des résumés construits et des stratégies d'exploitation tirant profit de ces résumés. Ces contributions ont été validées par simulation.

Une étude des travaux existants nous a permis de conclure que l'agrégation est simplement vue comme une méthode de compression d'informations dans les travaux existants. Notre approche s'appuie sur une vision de l'agrégation comme un résumé, donnant une information approchée permettant la prise de décision même en absence d'informations précises. Nous avons donc présenté, dans une première contribution, une structure d'agrégation d'événements spatio-temporels observés qui sera utilisable dans un contexte d'échanges inter-véhicules. Elle permet d'approximer la fréquence des types d'événements observés. Cette structure est construite sur un modèle spatio-temporel à deux niveaux permettant d'avoir le même référentiel spatio-temporel physique pour tous les véhicules, ce qui permet notamment d'échanger facilement et sans perte d'informations des résumés.

La structure d'agrégation construite par chaque véhicule est restreinte car elle ne tient compte que des événements observés directement ou indirectement (par dissémination). Pour avoir une information plus large, les véhicules ont tout intérêt à échanger leurs résumés entre eux. Dans une deuxième partie

de notre travail nous avons présenté notre « processus d'échange » dont les propriétés recherchées sont l'efficacité au niveau réseau et la précision afin que la fusion de deux résumés soit la plus fidèle possible. Pour tenir compte de la faible qualité des réseaux VANET, nous avons proposé un mécanisme de souscription pour limiter les volumes échangés qui permet à chaque véhicule d'annoncer les types d'événements et les zones spatio-temporelles qui l'intéressent. Un système de préférences complète la souscription permettant aux données les plus prioritaires d'être échangées en premier en cas de rupture de la connexion durant l'échange.

Les résumés construits et échangés doivent être exploités pour aider le conducteur à prendre une décision sur la route. Notre travail dans cette troisième contribution consiste à proposer une approche déclarative d'exploitation des résumés basée sur l'utilisation d'un langage de requêtes. Cette approche est à deux niveaux : (1) « niveau d'accès au résumé » où le résumé est considéré comme une base de données que nous allons interroger avec des requêtes spatio-temporelles, (2) « niveau d'interprétation des résultats » pour proposer au conducteur une aide à la décision. Nous distinguons ici deux modes d'assistance. Le premier est réalisé à la demande du conducteur lorsqu'il recherche une ressource (place de stationnement libre par exemple). Ici le niveau d'assistance va s'exprimer par une requête spatio-temporelle classique sur les résumés. Le second mode correspond à la gestion des alertes et s'exécute en continu sans intervention du conducteur. Le niveau d'assistance s'exprime alors par une requête spatio-temporelle continue sur les résumés.

Pour la validation et la mise en œuvre de l'ensemble des solutions proposées, nous avons choisi le simulateur « VESPA » et nous l'avons donc étendu par un module d'agrégation qui permet de construire la structure d'agrégation dans chaque véhicule et de l'exploiter ensuite. Les expérimentations menées avec ce simulateur montrent bien la faisabilité de nos approches. L'utilisation des résumés augmente effectivement le nombre de véhicules recherchant une place libre et diminue le temps de recherche. Ce gain est obtenu même avec des résumés n'observant qu'une faible partie des événements. Enfin, l'échange des résumés entre véhicules améliore nettement la qualité des résumés.

Finalement, le travail effectué dans cette thèse s'est avéré très enrichissant

car il se situe à la jonction de différents domaines de recherche. En effet, nous avons pu aborder différents problèmes liés aux domaines systèmes distribués et fouille de données spatio-temporelles.

8.2 Perspectives

Dans cette section, nous présentons quelques directions de travaux futurs. Nous présentons les perspectives liées aux trois principales contributions de notre thèse : construction du résumé, l'échange des agrégats et l'exploitation des résumés pour l'aide à la décision du conducteur.

1. Perspectives liées à la construction du résumé

La structure d'agrégation que nous avons proposée tire profit des événements provenant de l'observation directe ou indirecte via dissémination d'événements ainsi que de l'échange des résumés avec d'autres véhicules. Il serait intéressant également d'utiliser les informations en provenance d'une infrastructure. Cette infrastructure pourrait être composée par des terminaux disséminés le long des routes et fournissant à la fois de l'information précise aux véhicules et servant de relais pour échanger des informations entre véhicules. Cela permettrait d'éviter que de l'information ne soit pas transmise par manque de véhicules rencontrés intéressés. Cela introduit également une autre thématique liée à la confiance envers les sources d'informations. En effet, une information en provenance d'un fournisseur de parkings est fiable, alors que nos résumés sont de simples approximations. Il serait donc pertinent de prendre en compte cette confiance dans notre système.

Enfin, une dernière perspective consiste à envisager une solution de rafraîchissement de la structure d'agrégation proposée. En effet, les conditions du monde réel peuvent changer : par exemple, une zone dangereuse à cause des conditions climatiques ou des travaux peut ne pas l'être pour toujours, une zone péri-urbaine peut être transformée par de grands programmes immobiliers. Notre structure d'agrégation devrait pouvoir être mise à jour pour tenir compte de changements importants dans certaines zones et ne pas faire de mauvaises recommandations.

2. Perspectives liées à l'échange des agrégats

Dans le chapitre 5 le processus proposé permet l'échange des agrégats. Cet échange est purement opportuniste puisqu'il concerne deux véhicules qui se rencontrent et partagent (au moins) un intérêt commun. Il serait intéressant de caractériser de manière plus fine l'intérêt que peut avoir un véhicule à échanger avec un autre, de manière à pouvoir choisir le meilleur lorsqu'il y a le choix. De plus, ce principe d'échange basé sur un intérêt commun, ne garantit pas qu'une information ne puisse pas disparaître. En effet, si un véhicule ne rencontre pas d'autres véhicules partageant le même intérêt (tout simplement parce qu'il n'y en a pas dans la zone où il se déplace), alors l'information peut être perdue (en tous cas pas échangée avec les véhicules intéressés). Pour éviter cela, il serait intéressant d'utiliser des véhicules relais dont le rôle sera de transporter l'information vers d'autres (même si eux mêmes ne sont pas intéressés).

3. Perspectives liées à l'exploitation des résumés

Pour l'exploitation des résumés, nous avons choisi une approche déclarative basée sur l'utilisation d'un langage de requêtes. Pour les deux cas de requêtes définis pour les ressources et les alertes, des seuils doivent être définis pour déterminer une probabilité élevée ou faible (seuilMax et seuilMin) des valeurs envoyées au niveau d'assistance au conducteur. Ces seuils peuvent être fixes, mais il paraît plus intéressant de les définir de manière relative en cherchant à discriminer les valeurs élevées et basses.

Au niveau de la recommandation nous nous sommes basés que sur les valeurs renvoyées du résumé sans considérer la direction du déplacement du conducteur. Ainsi les propositions dans l'aide à la décision peuvent ne pas répondre parfaitement aux besoins du conducteur qui va être obligé de changer de direction et de s'éloigner de son but d'arrivée. Nous envisageons d'introduire au niveau « interprétation d'autres caractéristiques telles que la direction du déplacement. Autre contrainte que nous devons prendre en compte dans l'exploitation est la cartographie. En effet dans notre proposition nous avons supposé qu'il y a toujours une route permettant le passage d'une cellule physique à une autre voisine ce qui n'est pas toujours vrai bien évidemment. Le

temps de trajet est également un élément qui peut être pris en compte si nous utilisons la cartographie (il est plus rapide de prendre une rue à quatre voies sans feux tricolores, qu'une succession de rues étroites avec des feux et/ou des stops).

Si nous veut allons plus loin dans l'aide à la décision, nous pouvons essayer d'améliorer la qualité de la prédiction. Plusieurs pistes sont envisageables pour cela. Tout d'abord, nous pouvons essayer d'améliorer la précision de nos agrégats en stockant non pas seulement de simples fréquences mais en stockant plutôt des distributions. Les problèmes à résoudre ici portent sur la non sensibilité des constructions de distributions aux dupliquats et sur la compacité des distributions. Une autre piste consiste à tenir compte de plusieurs sources d'informations (agrégats et informations en provenance de l'infrastructure par exemple) et de corrélérer ces informations, éventuellement en leur donnant des niveaux de confiance différents.

Bibliographie

- [Adler 2006] C. Adler. Information dissemination in vehicular ad hoc networks. Master's thesis, Univ. of Munich, 2006. (Cité en page 12.)
- [Almeida 2007] Paulo Sérgio Almeida, Carlos Baquero, Nuno Preguiça and David Hutchison. *Scalable Bloom Filters*. Inf. Process. Lett., vol. 101, no. 6, pages 255–261, Mars 2007. (Cité en page 49.)
- [Bachir 2004] Abdelmalik Bachir. *Multicast dans les réseaux Ad-Hoc mobiles - (GIV) Geodiffusion Inter-Véhicules*. PhD thesis, Université d'Avignon, LIA, 2004. (Cité en page 12.)
- [Bernsen 2009] James Bernsen and D. Manivannan. *Unicast routing protocols for vehicular ad hoc networks : A critical comparison and classification*. Pervasive and Mobile Computing, pages 1–18, 2009. (Cité en page 61.)
- [Bhaskar 2002] Krishnamachari Bhaskar, Estrin Deborah and Wicker Stephen B. *The Impact of Data Aggregation in Wireless Sensor Networks*. In Proceedings of the 22nd International Conference on Distributed Computing Systems, ICDCSW'02, pages 575–578, Washington, DC, USA, 2002. IEEE Computer Society. (Cité en page 16.)
- [Briesemeister 2000] Linda Briesemeister and Gunter Hommel. *Role-based multicast in highly mobile but sparsely connected ad hoc networks*. page 4550. In Proceedings of ACM MobiHoc Boston, MA, August 2000. (Cité en page 12.)
- [Cenerario 2007] Nicolas Cenerario and Thierry Delot. *Evaluation continue de requêtes dans les réseaux de communication inter-véhicules*. Actes de l'Atelier Gestion de Données dans les Systèmes d'Information Pervasifs (GEDSIP) organise dans le cadre du XXVème congrès INFORSID, Perros-Guirec (France), mai 2007. (Cité en page 4.)
- [Cenerario 2008] Nicolas Cenerario, Thierry Delot and Sergio Ilarri. *Dissemination of Information in Inter-Vehicle Ad Hoc Networks*. In Intelligent Vehicles Symposium (IV'08), pages 763–768. IEEE Computer Society, June 2008. (Cité en pages 4 et 29.)

- [Cenerario 2010] Nicolas Cenerario. *Partage d'information dans les réseaux de communication inter-véhicules*. PhD thesis, Université de Valenciennes UVHC, 2010. (Cité en page 96.)
- [Cenerario 2011] Nicolas Cenerario, Thierry Delot and Sergio Ilarri. *A Content-Based Dissemination Protocol for VANETs : Exploiting the Encounter Probability*. IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 3, pages 771–782, September 2011. Special Issue on Exploiting Wireless Communication Technologies in Vehicular Transportation Networks. (Cité en page 13.)
- [Chen 2003] Chi Chen. Location-based data aggregation in mobile ad hoc networks. Master's thesis, Universitat Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, 2003. (Cité en page 14.)
- [Cheng 2003] Albert Cheng and Kanhirodan Rajan. *A digital map/GPS based routing and addressing scheme for wireless ad hoc networks*. In Intelligent Vehicle Symposium (IV'03), pages 17–20. IEEE, 2003. (Cité en pages 18 et 41.)
- [Christian 2007] Lochert Christian, Scheuermann Bjorn and Mauve Martin. *Probabilistic aggregation for data dissemination in VANETs*. In Proceedings of the fourth ACM international workshop on Vehicular ad hoc networks, VANET'07, pages 1–8, New York, NY, USA, 2007. ACM. (Cité en pages 15 et 27.)
- [Defude 2008] Bruno Defude, Thierry Delot, Sergio Ilarri, Jose-Luis Zechinelli and Nicolas Cenerario. *Data Aggregation in VANETs : the VESPA Approach*. In First International Workshop on Computational Transportation Science (IWCTS'08), in conjunction with the 5th Annual International Conference on Mobile and Ubiquitous Systems : Networks and Services (MOBIQUITOUS'08), Dublin (Ireland), page 6. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ISBN 978-963-9799-27-1, July 2008. (Cité en page 19.)
- [Defude 2009] Bruno Defude, Thierry Delot and Dorsaf Zekri. *Résumé de flux d'événements dans les réseaux inter-véhiculaires*. Proceedings of the 5th French-Speaking Conference on Mobility and Ubiquity Computing, pages 53–60, 2009. (Cité en page 57.)

- [Delot 2008] Thierry Delot, Nicolas Cenerario and Sergio Ilarri. *Estimating the Relevance of Information in Inter-Vehicle Ad Hoc Networks*. In Proceedings of the Ninth International Conference on Mobile Data Management Workshops, MDMW '08, pages 151–158, Washington, DC, USA, 2008. IEEE Computer Society. (Cité en pages 29 et 31.)
- [Delot 2010] Thierry Delot, Nicolas Cenerario and Sergio Ilarri. *Vehicular Event Sharing with a mobile Peer-to-peer Architecture*. Transportation Research - Part C (Emerging Technologies),18(4), pages 584–598, August 2010. (Cité en pages 3 et 32.)
- [Delot 2011] Thierry Delot, Sergio Ilarri, Nicolas Cenerario and Thomas Hien. *Event Sharing in Vehicular Networks Using Geographic Vectors and Maps*. Mobile Information Systems, vol. 7, no. 1, pages 21–44, 2011. (Cité en pages 13 et 28.)
- [DKHIL 2009] Hassen DKHIL. Greedy perimeter stateless routing sur omnet++. Master's thesis, Ecole nationale supérieur d'informatique, Tunisie, 2009. (Cité en pages vii et 3.)
- [Eichler 2006] Stephan Eichler, Christian Merkle and Markus Strassberger. *Data Aggregation System for Distributing Inter-Vehicle Warning Messages*. pages 543–544, November 2006. (Cité en pages 16 et 41.)
- [Fabio 2006] Picconi Fabio, Ravi Nishkam, Gruteser Marco and Iftode Liviu. *Probabilistic validation of aggregated data in vehicular ad-hoc networks*. In Proceedings of the 3rd international workshop on Vehicular ad hoc networks, VANET'06, pages 76–85, New York, NY, USA, 2006. ACM. (Cité en pages 14, 17, 18 et 42.)
- [Faouzi 2011] Nour-Eddin El Faouzi, Henry Leung and Ajeesh Kurian. *Data fusion in intelligent transportation systems : Progress and challenges - A survey*. Information Fusion, vol. 12, no. 1, pages 4–10, 2011. (Cité en page 32.)
- [Festag 2004] Andreas Festag, Holger Füßler, Hannes Hartenstein, Amardeo Sarma and Ralf Schmitz. *FLEETNET : Bringing car-to-car communication into the real world*. Computer, vol. 4, no. L15, page 16, 2004. (Cité en pages 10 et 27.)

- [Figueiredo 2001] Lino Figueiredo, Isabel Jesus, Tenreiro Machado, Jose Rui Ferreira, and Martins Carvalho. *Towards the development of intelligent transportation systems*. In *Intelligent Transportation Systems*, pages 1206–1211. IEEE Computer Society, 2001. (Cité en page 9.)
- [Fiore 2007] Marco Fiore, Jerome Harri, Fethi Filali and Christian Bonnet. *Vehicular Mobility Simulation for VANETs*. In *Proceedings of the 40th Annual Simulation Symposium, ANSS '07*, pages 301–309, Washington, DC, USA, 2007. IEEE Computer Society. (Cité en page 95.)
- [Flajolet 1985] Philippe Flajolet and Nigel Martin. *Probabilistic Counting Algorithms for Data Base Applications*. *J. Comput. Syst. Sci.*, vol. 31, no. 2, pages 182–209, 1985. (Cité en pages 18, 42, 50 et 51.)
- [Füßler 2002] Holger Füßler, Hannes Hartenstein, Martin Mauve, Michael Käsemann and Dieter Vollmer. *Location-Based Routing for Vehicular Ad-Hoc Networks*. *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (ACM MobiCom'02)*, September 2002. (Cité en page 61.)
- [Füßler 2003] Holger Füßler, Jörg Widmer, Michael Käsemann, Martin Mauve and Hannes Hartenstein. *Contention-based forwarding for mobile ad hoc networks*. *Ad Hoc Networks*, vol. 1, no. 4, pages 351 – 369, 2003. (Cité en page 14.)
- [Haowen 2006] Chan Haowen, Perrig Adrian and Song Dawn. *Secure hierarchical in-network aggregation in sensor networks*. In *Proceedings of the 13th ACM conference on Computer and communications security, CCS'06*, pages 278–287, New York, NY, USA, 2006. (Cité en page 16.)
- [H.Bloom 1970] Burton H.Bloom. *Space/time trade-offs in hash coding with allowable errors*. *Commun. ACM*, vol. 13, no. 7, pages 422–426, Juillet 1970. (Cité en page 49.)
- [Intanagonwiwat 2000] Chalermek Intanagonwiwat, Ramesh Govindan and Deborah Estrin. *Directed Diffusion : A scalable and robust communication paradigm for sensor networks*. In *Proceedings of the 6th annual international conference on Mobile computing and networking, MobiCom '00*, pages 56–67, New York, NY, USA, 2000. ACM. (Cité en page 16.)

- [Kruijff 2006] Geert Kruijff, John D. Kelleher and Nick Hawes. *Information fusion for visual reference resolution in dynamic situated dialogue*. Perception and Interactive Technologies : International Tutorial and Research Workshop, PIT 2006, volume 4021 of Lecture Notes in Computer Science, 2006. (Cité en page 32.)
- [Kulik 2002] Joanna Kulik, Wendi Heinzelman and Hari Balakrishnan. *Negotiation-based protocols for disseminating information in wireless sensor networks*. Wireless Networks, vol. 8, no. 2/3, pages 169–185, Mars 2002. (Cité en page 17.)
- [Lin 2010] Yun-Wei Lin, Yuh-Shyan Chen and Sing-Ling Lee. *Routing Protocols in Vehicular Ad Hoc Networks : A Survey and Future Perspectives*. J. Inf. Sci. Eng., vol. 26, no. 3, pages 913–932, 2010. (Cité en page 61.)
- [Lindsey 2002] Stephanie Lindsey and Cauligi Raghavendra. *PEGASIS : Power-efficient gathering in sensor information systems*. Systems Research (IEEE), vol. 3, pages 3–1125, 2002. (Cité en page 16.)
- [Lochert 2007] Christian Lochert, Björn Scheuermann, Murat Caliskan and Martin Mauve. *The Feasibility of Information Dissemination in Vehicular Ad-Hoc Networks*. In Forth Annual Conference on Wireless On demand Network Systems and Services (WONS'07), pages 92–99. IEEE Computer Society, January 2007. (Cité en pages 4, 12, 13, 18 et 41.)
- [Lopez 2005] Ines Fernando Vega Lopez, Richard Snodgrass and Bongki Moon. *Spatiotemporal aggregate computation : A survey*. IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pages 271–286, February 2005. (Cité en pages 4 et 16.)
- [Madden 2002] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein and Wei Hong. *TAG : a Tiny AGgregation service for ad-hoc sensor networks*. SIGOPS Oper. Syst. Rev., vol. 36, no. SI, pages 131–146, Décembre 2002. (Cité en pages 4 et 16.)
- [Maihöfer 2004] Christian Maihöfer. *A survey of geocast routing protocols*. IEEE Communications Surveys and Tutorials, vol. 6, no. 1-4, pages 32–42, 2004. (Cité en page 61.)

- [Mangharam 2006] Rahul Mangharam, Daniel Weller, Raj Rajkumar, Priyanka Mudalige and Fan Bai. *GrooveNet : A Hybrid Simulator for Vehicle-to-Vehicle Networks*. Mobile and Ubiquitous Systems, Annual International Conference on IEEE Computer Society, vol. 0, pages 1–8, 2006. (Cité en page 95.)
- [Morsink 2003] Paul Morsink, Redouane Hallouzi, Leonard Cseh, Christoph Schafers, Martin Nelisse and David de Bruin. *Cartalk 2000 : Development of a cooperative adas based on vehicle to vehicle communication*. In Proc. of the 10th World Congress and Exhibition in intelligent Transport Systems and Services, Saint Malo, France, 2003. (Cité en pages 4 et 10.)
- [Moses 2002] Charikar Moses, Chen Kevin and Farach-Colton Martin. *Finding Frequent Items in Data Streams*. In Proceedings of the 29th International Colloquium on Automata, Languages and Programming, ICALP '02, pages 693–703, London, UK, UK, 2002. Springer-Verlag. (Cité en page 48.)
- [Nadeem 2004] Tamer Nadeem, Sasan Dashtinezhad, Chunyuan Liao and Liviu Iftode. *Trafficview : Traffic data dissemination using car-to-car communication*. ACM SIGMOBILE Mobile Computing and Communications Review, vol. 8(3), pages 6–19, 2004. (Cité en pages 17, 19 et 41.)
- [Nadeem 2006] Tamer Nadeem, Pravin Shankar and Liviu Iftode. *A Comparative Study of Data Dissemination Models for VANETs*. In Int. Conf. on Mobile and Ubiquitous Systems (MOBIQUITOUS'06), pages 1–10. IEEE Computer Society, July 2006. (Cité en page 4.)
- [Noga 1996] Alon Noga, Matias Yossi and Szegedy Mario. *The space complexity of approximating the frequency moments*. In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, STOC '96, pages 20–29, New York, NY, USA, 1996. (Cité en page 48.)
- [Papadias 2002] Dimitris Papadias, Yufei Tao, Panos Kalnis and Jun Zhang. *Indexing Spatio-Temporal Data Warehouses*. In Proceedings of the 18th International Conference on Data Engineering, ICDE '02, pages 166–, Washington, DC, USA, 2002. IEEE Computer Society. (Cité en page 18.)

- [Rabiner 2000] Heinzelman Wendi Rabiner, Chandrakasan Anantha and Balakrishnan Hari. *Energy-Efficient Communication Protocol for Wireless Microsensor Networks*. In Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 8, HICSS'00, page 8020, Washington, DC, USA, 2000. IEEE Computer Society. (Cité en page 16.)
- [Rahim 2009] Aneel Rahim, Mehboob Yasin, Imran Ahmad, Zeeshan Shafi Khan and Muhammad Sher. *Relevance based approach with virtual queue for vehicular ad hoc networks*. In Computer Control and Communication 2nd International Conference(IC4'09), page 14, 2009. (Cité en page 13.)
- [Rajagopalan 2006] Ramesh Rajagopalan and Pramod Varshney. *Data aggregation techniques in sensor networks : A survey*. Comm. Surveys et Tutorials, IEEE, vol. 8, pages 48–63, 2006. (Cité en pages 4 et 16.)
- [Richard 2000] Weiland Richard and Purser Lara. *Intelligent Transportation Systems*. A5009 committee on Intelligent Transportation Systems, TRB, 2000. (Cité en page 9.)
- [Roberto 2006] Brignolo Roberto. *Co-operative road safety - the SAFESPOT integrated project*. APSN - APROSYS Conference. Advanced Passive Safety Network 6th Annual Conference, May 2006. (Cité en page 23.)
- [Saleet 2007] Hanan Saleet and Otman Basir. *Location-based message aggregation in vehicular ad hoc networks*. pages 1–7. IEEE Global Communications Conference Workshops, November 2007. (Cité en pages 17 et 41.)
- [Stefan 2009] Dietzel Stefan, Schoch Elmar, Bako Boto and Kargl Frank. *A Structure-free Aggregation Framework for Vehicular Ad Hoc Networks*. In 6th International Workshop on Intelligent Transportation (WIT 2009), Hamburg, Germany, March 2009. (Cité en page 16.)
- [Stephanie 2001] Lindsey Stephanie, Raghavendra Cauligi and Sivalingam Krishna. *Data gathering in sensor networks using the energy delay metric*, volume 00. IEEE Comput. Soc, 2001. (Cité en page 16.)

- [Szcurek 2010] Piotr Szcurek, Bo Xu, Jie Lin and Ouri Wolfson. *Spatio-temporal Information Ranking in VANET Applications*. IJNGC, vol. 1, no. 1, 2010. (Cité en page 14.)
- [Tonguz 2010] Ozan K. Tonguz, Nawaporn Wisitpongphan and Fan Bai. *DV-CAST : a distributed vehicular broadcast protocol for vehicular ad hoc networks*. Wireless Commun., vol. 17, no. 2, pages 47–56, Avril 2010. (Cité en page 61.)
- [Wisitpongphan 2007] NawaPorn Wisitpongphan, Ozan Tonguz, Jayendra Parikh, Priyantha Mudalige, Fan Bai and Varsha Sadekar. *Broadcast storm mitigation techniques in vehicular ad hoc networks*. Wireless Commun., vol. 14, no. 6, pages 84–94, Décembre 2007. (Cité en page 12.)
- [Wu 2004] Hao Wu, Richard Fujimoto, Randall Guensler and Michael Hunter. *MDDV : a mobility-centric data dissemination algorithm for vehicular networks*. In Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks, VANET '04, pages 47–56, New York, NY, USA, 2004. ACM. (Cité en page 12.)
- [Xu 2004a] Bo Xu, Aris Ouksel and Ouri Wolfson. *Opportunistic Resource Exchange in Inter-Vehicle Ad-Hoc Networks*. In Fifth International Conference on Mobile Data Management (MDM'04), IEEE Computer Society, Berkeley, California, pages 4–12, January 2004. (Cité en page 29.)
- [Xu 2004b] Bo Xu, Aris M. Ouksel and Ouri Wolfson. *Opportunistic Resource Exchange in Inter-Vehicle Ad-Hoc Networks*. In Mobile Data Management, pages 4–12, 2004. (Cité en page 13.)
- [Yan 2007] Lei Yan, Maria Kihl, Yuefeng Liu and Lina He. *The application of Inter-Vehicle Communication system to ITS*. Geographical Information Science Research Conference GISRUK, 2007. (Cité en page 9.)
- [Yi 2008] Yang Yi, Wang Xinran, Zhu Sencun and Cao Guohong. *SDAP : A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks*. ACM Trans. Inf. Syst. Secur., vol. 11, no. 4, pages 18 :1–18 :43, Juillet 2008. (Cité en pages 4 et 16.)

- [Yufei 2004] Tao Yufei, Kollios George, Considine Jeffrey, Li Feifei and Papadias Dimitris. *Spatio-Temporal Aggregation Using Sketches*. In Proceedings of the 20th International Conference on Data Engineering, ICDE'04, page 214, Washington, DC, USA, 2004. IEEE Computer Society. (Cité en pages 18 et 42.)
- [Zekri 2010a] Dorsaf Zekri and Bruno Defude. *Évaluation de requêtes sur des résumés spatio-temporels pour l'assistance aux conducteurs de véhicules*. Gestion des Données dans les Systèmes d'Information Pervasifs (GEDSIP) in INFORSID'10, Marseille, France, 2010. (Cité en page 91.)
- [Zekri 2010b] Dorsaf Zekri, Bruno Defude and Thierry Delot. *Summarizing sensors data in vehicular ad hoc networks*. RAIRO - Operations Research, vol. 44, no. 4, pages 345–364, October 2010. (Cité en page 57.)
- [Zekri 2012] Dorsaf Zekri, Bruno Defude and Thierry Delot. *A cooperative scheme to aggregate spatio-temporal events in VANETs*. IDEAS '12 : 16th International Database Engineering & Applications Symposium, Prague, Czech Republic, pages 100–109, august 2012. (Cité en page 77.)
- [Zhao 2006] Jing Zhao and Guohong Cao. *VADD : Vehicle-assisted data delivery in vehicular ad hoc networks*. In Int. Conf. on Computer Communications (INFOCOM'06), 2006. (Cité en page 12.)