



UNIVERSITE AMAR TELIDJI LAGHOUCAT  
FACULTE DES SCIENCES  
DEPARTEMENT D'INFORMATIQUE



Mémoire :

Présenté en vue de l'obtention du diplôme

**Master en Informatique**

Option : Réseaux, Systèmes et Applications Réparties

Présenté et soutenu publiquement par :

**Taha ElAmine HADJADJ**

Titre :

**VeGAS : Système d'aide à la décision basé  
sur l'agrégation des connaissances dans les  
réseaux inter-véhiculaires**

JURY

Pr. Youcef Ouinten	Président	Professeur, université de Laghouat
Dr. Younes Guellouma	Examineur	MCB, université de Laghouat
Dr. Mustapha Bouakkaz	Encadrant	MCB, université de Laghouat

2016/2017

# Dédicace

*A* la plus belle perle du monde... ma tendre mère

*A* celui qui a toujours garni mes chemins avec force et lumière... mon très cher  
père

*A* la mémoire de mes grands parents

Puisse dieu les accueillir dans son infinie miséricorde

*A* mes sœurs et mon frère

Je leur souhaite tout le succès... tout le bonheur

*A* toute ma famille pour l'amour et le respect qu'ils m'éprouvent

*A* tous ceux qui me sont très chères au cœur... pour tout le soutien et le bonheur  
apporté

*A* tous mes amis et mes camarades

Pour une sincérité si merveilleuse... jamais oubliable, en leur souhaitant tout  
succès... tout bonheur

*A* toute personne qui m'a aidé à franchir un horizon dans ma vie ...

Aimablement

Je dédie ce modeste travail...

*Taha*

# Remerciements

Je voudrais exprimer mes sentiments les plus sincères et mes considérations envers les personnes sans lesquelles ce travail n'aurait pas pu voir le jour. Leur aide, accompagnement et soutien m'ont été indispensables afin de pouvoir réaliser mon mémoire.

Je tiens à exprimer mes sincères remerciements à mon encadrant Mr. Mustapha Bouakkaz. Pour son aide sans limites et ses précieux conseils sans lesquels ce travail n'aurait pas vu le jour. Il m'a ainsi offert l'opportunité de faire mes premiers pas de scientifique sur un travail de recherche passionnant et prometteur : le monde de systèmes distribués et fouille de données.

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

Un grand merci à ma famille pour son soutien tout au long de ma scolarité, jusqu'à ce mémoire. Mes parents, Fatiha et Noureddine, pour les sacrifices qu'ils ont consentis dans mon éducation et sans lesquels je n'aurais pas pu réaliser ce mémoire.

Je présente mes sincères remerciements à tous les enseignants et responsables du département Math et Informatique de l'université de Laghouat pour leur formation de licence et master dont je suis très satisfait, et pour leur travail avec passion et patience.

A tous ceux qui m'ont aidé, que ce soit par une grande collaboration ou par une tendre parole.

---

## Résumé

Les travaux réalisés dans ce mémoire 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 ou avec une infrastructure. Avec de tels mécanismes de communication, un véhicule peut recevoir des informations à partir d'un serveur central via l'infrastructure, 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.). 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 l'infrastructure, un résumé (ou agrégat) qui fournit des données qui seront ensuite échangées avec les véhicules pour apporter des informations aux conducteurs qui vont les aider à prendre leurs décisions, y compris lorsqu'aucun véhicule communicant ne se trouve pas à proximité. Pour ce faire, nous proposons une structure spatio-temporelle et une approche d'agrégation permettant à un serveur de résumer l'ensemble des événements reçus et observés par les véhicules. Nous avons validé l'ensemble de nos propositions en utilisant un simulateur que nous avons développé en Java et 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 d'échanges entre véhicules.

**Mots clés :** Réseaux inter-véhicule - Agrégation des données – Construction de résumé - Protocole d'échange - Extraction de connaissances.

---

# Table des matières

Introduction générale.....	1
Chapitre 1 Concepts de base et état de l'art.....	5
1.1 Que ce qu'un système décisionnel ? .....	5
1.1.1 Système d'aide à la décision.....	5
1.1.2 Système d'information d'aide à la décision .....	6
1.1.3 Des systèmes d'aide à la décision aux entrepôts de données .....	8
1.1.4 Entrepôt et magasins de données.....	9
1.2 Que ce que l'agrégation des données ? .....	10
1.2.1 Solutions d'agrégation existantes .....	13
1.3 Comment un system décisionnel peut aider dans la conduite ? .....	18
1.3.1 Systèmes autonomes.....	18
1.3.2 Systèmes coopératifs .....	19
1.3.3 Systèmes interactifs .....	20
1.4 Synthèse .....	22
Chapitre 2 Contribution : L'agrégation des données inter-véhiculaires.....	24
2.1 Introduction .....	24
2.2 Architecture globale .....	26
2.3 Notre contribution VeGAS.....	27
2.4 Modèle spatio-temporel.....	29
2.4.1 Niveau Physique .....	29
2.4.2 Niveau logique.....	31
2.5 Construction des résumés des événements.....	33
2.5.1 L'Agrégation par graphe .....	34

2.5.2	Exemple .....	36
2.6	Structure d'agrégation .....	38
2.7	Conclusion.....	40
Chapitre 3 Implémentation et validation expérimentale.....		43
3.1	Introduction .....	43
3.2	Choix du simulateur .....	43
3.2.1	Les simulateurs existants .....	44
3.2.2	Simulateur « VeGASim ».....	45
3.2.3	Adaptation du simulateur .....	47
3.3	Paramètres de simulation .....	48
3.4	Critères et stratégies utilisés dans les simulations.....	49
3.5	Evaluation qualitative du résumé spatio-temporel .....	51
3.6	Évaluation du processus d'échange.....	56
3.7	Conclusion.....	58
Chapitre 4 Conclusion et perspectives.....		59
Bibliographies .....		62

# Table des Figures

Figure 1. Représentation systémique d'une organisation [8].....	6
Figure 2. Le SIAD dans le SI [8] .....	7
Figure 3. Le système d'Aide à la Décision [8] .....	8
Figure 4. Entrepôt et magasins de donnée.....	10
Figure 5. Classification des stratégies d'agrégation [1] .....	12
Figure 6. Niveau physique (portion de Laghouat) .....	30
Figure 7. Niveau logique.....	32
Figure 8. Graphe exemple .....	36
Figure 9. Construction des Cycles.....	38
Figure 10. Structure d'agrégation (serveur) .....	41
Figure 11. Structure d'agrégation (véhicule).....	42
Figure 12. Comparaison des différentes stratégies sans échange.....	53
Figure 13. Impact du processus d'échange .....	54

## Liste des tableaux

Tableau 1. Matrice des Fréquences (FM).....	36
Tableau 2. Matrice d'Adjacence (AdjM).....	37
Tableau 3. Matrice D'affinité (AffM).....	37
Tableau 4. Extraction des cycles à partir de la Matrice d'Affinité (AffM) ....	37
Tableau 5. Caractéristiques des 8 parkings considérés durant les simulations .....	48
Tableau 6. Nombre de messages échangé pour les différentes stratégies .....	57

# Glossaire

ADAS. *Advanced Driver-Assistance Systems*

AFIL. *Alerte de Franchissement Involontaire de Ligne*

API. *Application Programming Interface*

BI. *Business Intelligence*

DSRC. *Dedicated Short Range Communication*

ED. *Entrepôt de Données*

ETL. *Extraction Transformation Loading*

FCC. *Commission Fédérale des Communications*

GPS. *Global Positioning System*

I2V. *Infrastructure to Vehicle*

IEEE. *Institute of Electrical and Electronics Engineers*

ITS. *Intelligent Transportation System*

LAVIA. *Limiteur s'Adaptant à la Vitesse Autorisée*

LBAG. *Location Based Aggregation*

MD. *Magasin de Données*

RLSMP. *Region-based Location Service Management Protocol*

S.I.D.. *Systèmes d'Information Décisionnels*

SAD. *Système d'Aide à la Décision*

SI. *Système d'Information*

SIAD. *Système d'Information d'Aide à la Décision*

SO. *Système Opérant*

SP. *Système de Pilotage*

*TAG. Textual Aggregation by Graphs*

*V2I. Vehicle to Infrastructure*

*V2V. Vehicle to Vehicle*

*VANETs. Vehicular Ad-hoc NETWORKs*

*VeGAS. Vehicular Graphical Aggregation System*

*VeGASim. Vehicular Graphical Aggregation Simulator*

*VESPA. Vehicular Event Sharing with a mobile P2P Architecture*

## Introduction générale

De nos jours, la voiture est le moyen de transport le plus répandu. Aujourd'hui cette voiture est de plus en plus munie d'équipements électroniques et informatiques 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 comprendre 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 - 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 [1].

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 [1].

En Algérie, le nombre de véhicules a augmenté de 987193 en 2009 à 1397554 de véhicules en 2014 soit 41,5 % [2]. Pourtant, la population n'a augmenté que de 10,7 % au cours de la même période passant de 35,3 à 39,1 millions d'habitants [3]. Le nombre de véhicules a donc augmenté 4 fois plus vite que la population. Cette augmentation de la production en masse des véhicules a une influence sur le bon

déroulement du trafic et mène à des problèmes critiques tels que l'augmentation de la congestion et de la perturbation de circulation. Les problèmes de circulation et le besoin d'offrir plus de sécurité aux conducteurs poussent les chercheurs à intégrer de nouvelles technologies intelligentes dans le domaine du transport. Une nouvelle génération de ITS a été lancée pour répondre à l'exigence d'un développement rapide des transports. Cette nouvelle génération vise à améliorer la sécurité routière, réduire les émissions à effets de serres des véhicules et économiser de l'énergie.

Plusieurs grands projets de recherche ont été abordés tels que le comportement des conducteurs, la gestion de la sécurité du trafic, les véhicules intelligents sur la chaussée, les technologies de collecte de l'information, les systèmes de simulation d'accidents, les alertes de sécurité et d'urgence, les systèmes d'assistance au conducteur et d'aide à la prise de décision, etc. [4].

Les ITS sont l'une des réponses techniques au défi des transports du XXI<sup>e</sup> siècle. Ils ont été développés depuis plusieurs décennies dans le monde. Ils devraient permettre à une population de plus en plus nombreuse de se déplacer de plus en plus souvent, dans des meilleures conditions. Les ITS sont définis comme le point culminant de plusieurs services aux utilisateurs qui sont activés par les dernières améliorations en matière d'informatique, des technologies de détection et de communication véhiculaire afin de résoudre les problèmes de transports complexes.

Dans l'optique d'améliorer l'assistance au conducteur à travers une prise de décision adaptée, on aspire la mise sur pied d'une solution permettant d'accroître les connaissances qu'a le conducteur de son environnement et l'aider dans sa prise de décision face à prendre le chemin rapide vers leur destination et un espace de stationnement.

Le présent travail se concentre sur l'agrégation et l'extraction des connaissances dans les réseaux inter-véhiculaires en utilisant une approche d'agrégation adaptée de l'algorithme de TAG (Textual Aggregation by Graphs) [5]. Les principaux points sur lesquels notre travail s'articule sont :

1. **La proposition d'une structure d'agrégation spatio-temporelle** permettant au serveur de résumer l'ensemble des événements observés sur la route. Cette structure doit être capable de stocker des informations sur des événements observés, provenant des 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 conducteur (choix de zones spatio-temporelles à résumer) et aussi échangeable ;

2. **L'adaptation d'une approche d'agrégation** pour qu'elle sera utilisée par un serveur pour construire ses résumés. Cette approche doit nous générer des agrégats échangeables avec les véhicules et qui résument un type d'évènement donné dans un espace-temps donné.

3. **La validation des solutions proposées** tant du point de vue faisabilité que performance. Les contributions présentées dans ce travail ont été implémentées au sein du simulateur « VeGASim » que nous avons développé en conséquence.

Hormis l'introduction et la conclusion, ce mémoire est organisé en trois chapitres répartis de la manière suivante :

**Le chapitre 1, Concepts de base et 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 systèmes décisionnels et les systèmes d'aide à la décision aux entrepôt et magasins de données. Ensuite nous exposons les travaux liés à l'agrégation de données 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 2, Contribution : L'agrégation des données inter-véhiculaires**, constitue le cœur de notre travail. Dans ce chapitre nous commençons par décrire notamment les éléments sur lesquels nous nous appuyons pour définir nos contributions. Nous y identifions également nos besoins en termes d'agrégation et l'architecture globale de notre système. Après, Nous y détaillons nos approches. Nous présentons d'abord notre modèle spatio-temporel à deux niveaux ainsi que

notre structure d'agrégation. Ensuite, nous décrivons en détails notre approche d'agrégation et notre Système VeGAS (Vehicular Graphical Aggregation System).

**Le chapitre 3, simulation et validation expérimentales**, dans ce chapitre nous présentons les différents résultats expérimentaux obtenues issues de l'implémentation de nos contributions.

**Le chapitre 4, conclusions et perspectives**, Enfin nous terminons ce mémoire par des conclusions et tracer des perspectives à ce travail.

## Chapitre 1 Concepts de base et état de l'art

### 1.1 Que ce qu'un système décisionnel ?

"Le système d'information est l'ensemble des méthodes et moyens de recueil de contrôle et de distribution des informations nécessaires à l'exercice de l'activité en tout point de vue d'organisation. Il a pour fonction de produire et de mémoriser les informations, de l'activité du système opérant (système opérationnel), puis de les mettre à disposition du système de décision (système de pilotage)" [6].

Les différences qui existent entre le système de pilotage et le système opérationnel, du point de vue fonctionnel ou des tâches à effectuer, conduit à l'apparition des "systèmes d'information décisionnels" (S.I.D.) [7].

#### 1.1.1 Système d'aide à la décision

La modélisation systémique de toute organisation se décompose en trois sous-systèmes : Système Opérant (SO), Système d'Information (SI) et Système de Pilotage (SP). Le SO représente l'activité productrice de l'organisation étudiée. Cette activité consiste à transformer les flux primaires (matières, finance, personnel...) pour répondre aux besoins des clients. Le SP regroupe l'ensemble du personnel d'encadrement qui effectue les tâches de régulation, de pilotage et d'adaptation de l'organisation à son environnement. Le SI permet de collecter, mémoriser, traiter et restituer les différentes données de l'organisation afin de permettre au SP d'effectuer ses fonctions tout en assurant son couplage avec le SO. L'activité du SO produit des informations stockées dans le SI ; après traitement, la transmission de ces informations vers le SP permet à ce dernier de connaître l'activité du SO (flèches "informations" dans la Figure 1). Les décisions du SP seront répercutées vers le SI puis vers le SO pour permettre au SP d'en maîtriser le fonctionnement (flèches "décisions" dans la Figure 1) [8].

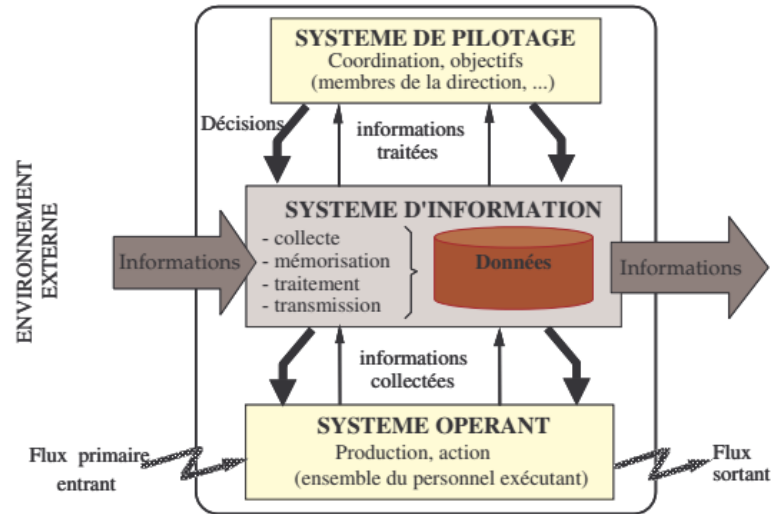


Figure 1. Représentation systémique d'une organisation [8]

Pour répondre aux besoins des décideurs, il est nécessaire de synthétiser, réorganiser et historier les données de production du SI afin d'en déterminer une sous-partie relative à l'aide à la décision. La suite de ce mémoire se centre sur cet aspect. Notamment, dans les sections suivantes, nous définissons les concepts de système d'information d'aide à la décision, de système d'aide à la décision, des entrepôts et des magasins de données.

### 1.1.2 Système d'information d'aide à la décision

Par analogie à la définition précédente d'un SI, nous proposons la définition du Système d'Information d'Aide à la Décision (SIAD) suivante [8]:

**Définition 1** : Un SIAD est la partie d'un système d'information permettant d'accompagner les décideurs dans le processus de prise de décision (Figure 2). Les fonctions d'un SIAD permettent de :

- Collecter, intégrer, synthétiser et transformer les données opérationnelles d'un SI, - mémoriser de manière adaptée les données décisionnelles,
- Traiter ces données (alimentation, rafraîchissement, pré-calculs.),
- Restituer de manière appropriée ces données afin de faciliter la prise de décision.

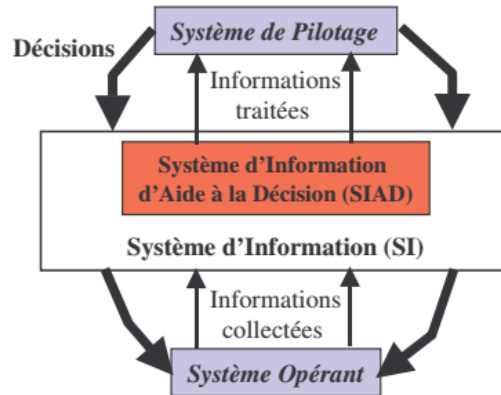


Figure 2. Le SIAD dans le SI [8]

De nos jours, l'ensemble des outils informatiques permettant de supporter un SIAD est qualifié de Business Intelligence (BI) ou de Système d'Aide à la Décision (SAD). Un SAD vise à exploiter les données opérationnelles d'une organisation afin de faciliter la prise de décision pour un pilotage éclairé. Afin d'être plus explicite, nous proposons pour les SAD la définition suivante :

**Définition 2 :** Un Système d'Aide à la Décision (SAD) regroupe l'ensemble des outils informatiques (matériels et logiciels) permettant :

- D'extraire, de transformer et de charger les données opérationnelles,
- De constituer un ou des espaces de stockage de données décisionnelles,
- De manipuler ces données au travers d'outils d'analyse ou d'interrogation destinés au pilotage des organisations.

La plupart des travaux déclinent ces applications informatiques en trois catégories [8]:

- Extraction, transformation et chargement (ou ETL acronyme de "Extraction Transformation Loading") des données opérationnelles (hétérogènes et disparates) pour alimenter et rafraîchir le système d'aide à la décision,
- Stockage et traitement des données décisionnelles,
- Restitution des données sous une forme adaptée aux utilisateurs (interrogations ou analyses décisionnelles).

Nous pouvons schématiser ces différents outils dans la Figure 3 suivante :

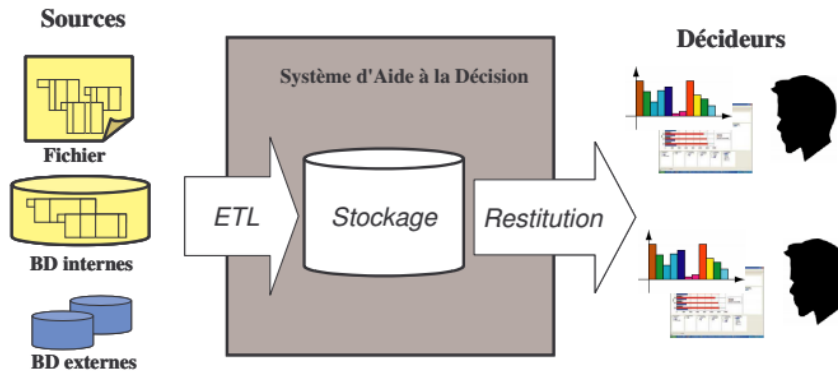


Figure 3. Le système d'Aide à la Décision [8]

### 1.1.3 Des systèmes d'aide à la décision aux entrepôts de données

De nos jours, les entrepôts de données constituent une solution adéquate pour construire un système d'aide à décision. Un entrepôt de données (ED) est défini comme étant "une collection de données intégrées, orientées sujet, non volatiles, historisées, résumées et disponibles pour l'interrogation et l'analyse" [9]. Cette définition met l'accent sur les caractéristiques suivantes [8]:

- **Intégrées** : les données alimentant l'entrepôt proviennent de sources multiples et hétérogènes. Les données des systèmes de production doivent être converties, reformatées et nettoyées de façon à avoir une vision globale dans l'entrepôt.

- **Orientées sujet** : contrairement aux systèmes de production structurant les données par processus fonctionnel, les données d'un ED s'organisent par thèmes d'analyse. L'intérêt de cette organisation est de disposer de l'ensemble des informations utiles sur un thème, le plus souvent transversales aux structures fonctionnelles et organisationnelles d'une entreprise. Cette orientation sujet permet de mettre en avant les indicateurs de performance pour chaque thème d'analyse.

- **Non volatiles** : après intégration, transformation et synthèse des données opérationnelles dans un ED, les seules actions que peuvent effectuer

des décideurs sont des interrogations et des analyses décisionnelles (pas de mise à jour).

- **Historisées** : l'alimentation et le rafraîchissement d'un ED consiste en l'intégration des données opérationnelles à différents points d'extraction. Cette intégration de données à des dates différentes permet de conserver "l'historisation" des données qui est vitale pour toute prise de décision.

- **Résumées** : les informations issues des sources doivent être transformées mais surtout agrégées pour faciliter le processus de prises de décision.

- **Disponible pour l'interrogation et l'analyse** : afin d'améliorer les performances d'une organisation, les décideurs doivent pouvoir consulter et analyser les données contenues dans un ED au travers d'outils interactifs.

#### 1.1.4 Entrepôt et magasins de données

L'ED doit permettre d'extraire, de transformer et de stocker un grand volume de données opérationnelles et, en même temps, de répondre à des requêtes utilisateurs concernant un thème d'analyse spécifique. En fait, cette définition regroupe deux problématiques que nous avons identifiées comme suit dès le début de nos travaux [8]:

- La gestion efficace des données "historisées", "centralisées" (intégration des sources),

- La définition d'un sous-ensemble de données autour d'un thème particulier afin de répondre aux besoins spécifiques de décideurs.

Aussi, l'architecture des systèmes d'aide à la décision que nous proposons est basée sur une dichotomie d'espaces de stockage : l'entrepôt et les magasins de données [7].

**Définition 3** : Un Entrepôt de Données (ED) est l'espace de stockage centralisé d'un extrait des sources pertinent pour les décideurs. Son organisation doit faciliter la

gestion des données et la conservation des évolutions nécessaires pour les prises de décision [10].

**Définition 4 :** Un Magasin de Données (MD) est un extrait de l'ED adapté à un thème d'analyse particulier et organisé selon un modèle adapté aux outils d'analyse et d'interrogation décisionnelle [10].

Dans la Figure 4 suivante, nous schématisons l'architecture des SAD telle que nous l'avons définie précédemment.

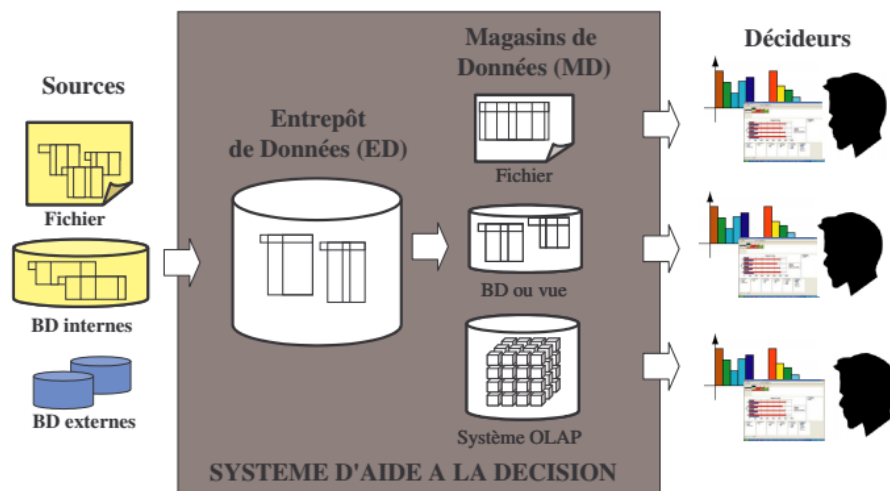


Figure 4. Entrepôt et magasins de donnée

Cette dichotomie des espaces de stockage a servi de guide pour nos travaux de développement.

## 1.2 Que ce que l'agrégation des données ?

Pour effectuer la communication dans les VANET, la Commission fédérale des communications (FCC) a consacré 75 MHz de la gamme de fréquence dans la bande de 5,850 GHz à 5,925 GHz à utiliser pour la communication V2V et V2I. La gamme de 5,9 GHz a été appelé Communication à courte portée dédiée (Dedicated Short Range Communication, DSRC) [11] et utilise IEEE 802.11p [12]. Les applications VANET doivent partager la bande passante allouée, ce qui en fait une ressource rare qui devrait être gérée très attentivement. Une diffusion de données insuffisante gaspille une grande quantité de bande passante qui, si elle est enregistrée, permettrait

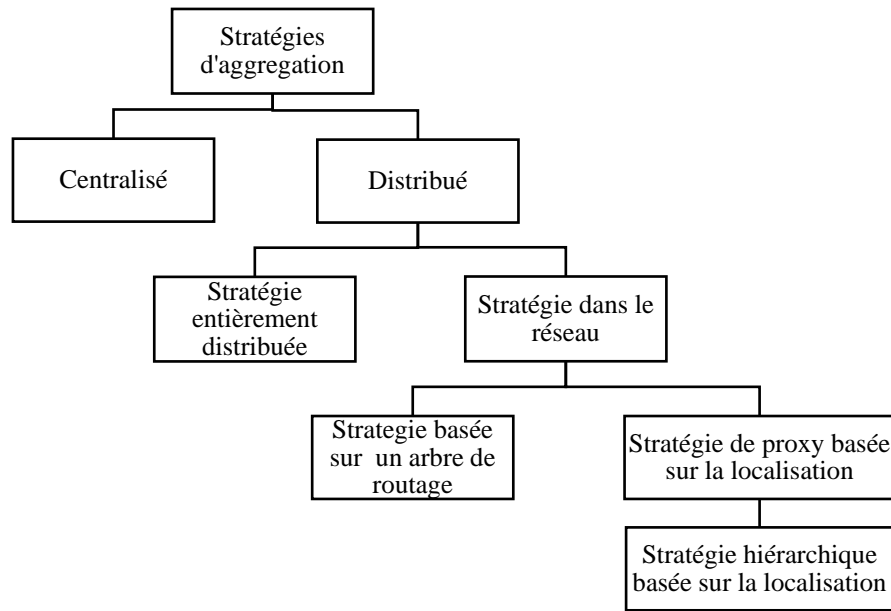
à plus d'applications de véhicules de coexister en plus de permettre la diffusion des données véhiculaires.

De nombreuses applications VANET exigent que chaque véhicule partage ses données (par exemple, la vitesse et l'emplacement) avec ses voisins en diffusant un message contenant ces données. L'envoi de ces messages à des distances plus éloignées va perdre la bande passante et peut causer un problème de tempête de diffusion en fonction de la densité de trafic [13]. Donc, pour partager efficacement les données avec les véhicules à de plus grandes distances, de nombreuses techniques d'agrégation de données ont été proposées.

L'agrégation des données a été proposée dans les VANET pour résoudre le problème d'utilisation de la bande passante. L'idée de base est de récolter des informations sur de nombreux véhicules dans un seul cadre. Dans la littérature les techniques d'agrégation sont classées en deux catégories : syntaxique ou sémantique [14].

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 termes de coût d'envoi [1].

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 [15]. Cette classification est montrée dans la Figure 5 [1].



**Figure 5. Classification des stratégies d'agrégation [1]**

Le classement présenté dans la Figure 5 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 : quels 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 [16] 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 [1].

### 1.2.1 Solutions d'agrégation existantes

L'agrégation de données a été particulièrement étudiée dans le contexte des réseaux de capteurs. Plusieurs techniques d'agrégation développées dans le cadre des données spatio-temporelles existent également dans la littérature [1] [17] [18] [19] [20] 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. Toutefois, ces mécanismes de regroupement ne peuvent pas être appliqués directement pour l'agrégation des données inter-véhiculaire 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, une hiérarchie d'arborescence ou un cluster. Dans des réseaux très dynamiques comme l'agrégation des données inter-véhiculaire, la construction de telles hiérarchies n'est pas triviale, leur maintenance encore moins ;
- Les architectures des réseaux de capteurs : 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, ne peuvent pas être déployés pour de nombreuses applications d'agrégation des données inter-véhiculaire 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 agrégations des données inter-véhiculaire [1].

Une proposition d'Eichler et al. [17] 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'utilisation 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 [18], 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 [14] 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.

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

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. 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 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. 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 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 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, une structure pour agrégation d'événements. Elle combine le aRB tree (aggregate RB-) avec le sketch Flajolet Martin. 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.

Un autre travail d'agrégation et extraction des connaissances dans les réseaux inter-véhiculaires inscrit dans le projet VESPA (Vehicular Event Sharing with a mobile P2P Architecture) [19] proposé par Dorsaf Zekri [1] dont ils présentent une structure d'agrégation d'événements spatio-temporels observés qui soit utilisable dans un contexte d'échanges inter-véhiculaires. Cette structure est construite sur un modèle spatio-temporel à deux niveaux. Pour ce faire, ils ont tout d'abord défini un modèle spatio-temporel qui 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, ils ont décrit une structure d'agrégation basée sur le modèle spatio-temporel à deux niveaux. Cette structure est propre à chaque véhicule ce qui leur permet de travailler dans un environnement complètement décentralisé. Après ils décrivent un 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, le processus s'appuie sur le fait que les résumés sont compacts (dû à la structure de données utilisée). Pour aller plus loin, ils ont 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. Et finalement ils ont proposé une 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 [20].

Nous avons présenté dans cette section 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 des données inter-véhiculaires. Nous avons mentionné que la première catégorie d'agrégation liée aux réseaux de capteurs ne peut 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. Seul le processus d'agrégation est une réduction basée sur la sémantique des données alors que la compression utilise des critères syntaxiques. Dans ce mémoire, 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 de construire un résumé d'événements produits sur la route qui serait exploité pour estimer si un événement est susceptible de survenir même sans observation directe. Un travail préalable 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.

### 1.3 Comment un system décisionnel peut aider dans la conduite ?

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 :

#### 1.3.1 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 concerne l'aide au contrôle de la vitesse comme :

**Speed Alert** : qui est un système embarqué d'information sur les limites de vitesse. Il aide le conducteur à rester dans les limites de vitesse en utilisant 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.

**LAVIA** : Limiteur s'Adaptant à la Vitesse Autorisée mise en ligne : mardi 2 janvier 2007. Le LAVIA (Limiteur s'adaptant à la vitesse autorisée) est un 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 SCANIA et VOLVO 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.

### 1.3.2 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 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** : 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 inter-distances 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** : 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*, *Apila* ou *PlaceLib* 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.

### 1.3.3 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.

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** : 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** : 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** : est une solution d'informations géo localisées 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).

**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 optimisé basé sur le concept de probabilité de rencontre, utilisé pour estimer la pertinence d'événements pour les véhicules [19].

## 1.4 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. Dans cette section, nous avons défini les différents concepts servant de support à nos travaux sur l'aide à la décision. A partir de la représentation systémique d'une organisation, nous avons identifié le concept de SIAD qui est la partie d'un SI permettant d'accompagner un ou plusieurs décideurs dans le processus de prise de décision. Un Système d'Aide à la Décision (SAD), partie informatisée d'un SIAD, regroupe l'ensemble des outils informatiques capables d'extraire les données opérationnelles afin de les transformer en informations pertinentes pour les décideurs. D'un point de vue architectural, nous avons identifié deux espaces de stockage des données dans un SAD : l'entrepôt (espace de stockage centralisé) et les magasins de données (espace de stockage extrait d'un ED et centré sur un thème d'analyse particulier). Après nous avons présenté les travaux existants pour l'assistance au conducteur que nous avons classés 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 certain nombre de tâches qui pourraient atténuer sa vigilance. Aucun système n'exploite l'environnement du conducteur pour l'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 e l'abonnement trimestriel) en plus des inconvénients de la centralisation des informations qui sont liés eux 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 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.

## Chapitre 2 Contribution : L'agrégation des données inter-véhiculaires

### 2.1 Introduction

Pourquoi agréger ? Même si les systèmes de communication inter-véhiculaires de type VESPA (voir 1.3.2) [19] peuvent fournir une assistance efficace au conducteur dans de nombreux cas, ils souffrent d'un inconvénient majeur. En effet, dans les VANets, une fois que les données reçues 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 l'exemple « recherche de la ressource : place de parking ». 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 disséminé par un ou plusieurs véhicules libérant leurs places en plus il n'y a aucune place libre visible pour 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 connecter à un serveur pour répondre à sa demande. Dans notre exemple « recherche d'une place de parking » il peut accéder au serveur « INDIGO Park »<sup>1</sup>. 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 INDIGO n'assure pas la couverture de toutes les places de parking de toutes les villes.

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

---

<sup>1</sup> <http://www.parkindigo.fr/>

peuvent être fournies par le système « COYOTE »<sup>2</sup> 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 INDIGO) 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. 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. Z.Dorsaf [1] propose 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. Et pour construire collaborativement ces résumés des données les véhicules ont besoin d'échanger leurs résumés et leurs connaissances. Cette approche est plus coûteuse qu'un simple message en termes de bande passante et nombre de messages dans les communications inter-véhiculaire.

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

Nous proposons donc que les données collectées par un véhicule ne servent pas uniquement pour produire un message d'alerte à destination du conducteur mais ces données seront périodiquement envoyées au serveur central, il est en effet possible d'utiliser a posteriori ces données collectées pour produire, au niveau du serveur, des connaissances sur l'environnement exploitables par les conducteurs des véhicules connectés ou par les responsables pour les aider dans la prise de décision. A titre d'exemple, les événements concernant les places de stationnement disponibles reçues

---

<sup>2</sup> <http://www.moncoyote.com/fr/comment-ca-marche>

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 notre travail à fournir une solution complète en termes d'événements qui peuvent se 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.

## 2.2 Architecture globale

Dans ce qui suit, nous décrivons l'architecture de notre contribution. Notre système d'aide à la décision est composé de trois parties principales :

1. Construction de résumé reçu par les véhicules, 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 aperçus, 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 l'agrégation des résumés au niveau du serveur central 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 du génération de l'événement agrégée ;

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 pour son application en bord ;

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 échangé entre serveur central et 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 ;

## 2.3 Notre contribution VeGAS

Partant des solutions d'agrégation décrites, nous avons noté qu'une première catégorie des travaux tels que [17], [18] considère l'agrégation dans les échanges inter-véhiculaires 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 [21], [22], [23]. 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 [14] 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 [24] l'utilisation des arbres de nœuds qui sont couteux en mémoire et difficiles à maintenir du fait de la forte mobilité des véhicules. Une autre solution [1] propose une structure d'agrégation qui supporte plusieurs types d'événements qui se produisent sur la route en s'appuyant sur un modèle spatiotemporel défini pour agréger les événements. Ce modèle permet à chaque conducteur de choisir ses zones d'intérêts. Une fois le modèle spatiotemporel défini, on utilise une méthode pour agréger le maximum d'événements détectés soit par observation directe, soit par échange avec d'autres véhicules par l'utilisation de la solution de Flajolet et Martin [25] 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.

Dans nos travaux, nous proposons une structure d'agrégation centrale qui supporte plusieurs types d'événements se produisant sur la route et qui répond aux besoins du processus d'agrégation exprimés dans l'Architecture globale que nous avons appelé VeGAS (Vehicular Graphical Aggregation System). Le principe général de la contribution est de s'appuyer sur un modèle spatiotemporel que nous définissons pour agréger les données sous forme de graphes. Ce modèle permet à chaque conducteur de choisir ses zones d'intérêts et le type d'évènements que son application de bord a besoin. Une fois le modèle spatio-temporel défini, nous avons choisi une méthode pour agréger le maximum d'événements détectés et reçu par les véhicules. Pour ce faire, nous utilisons une solution adaptée de l'algorithme de TAG [5] qui se base sur des cycles construits par le calcul des affinités entre les évènements de même type et leurs emplacement géographique sur la carte. Ces cycles permettent de connaître l'occurrence la plus probable d'un évènement dans l'espace-temps souhaité.

Le reste de ce chapitre est organisé comme suit. Nous allons d'abord décrire dans la section 2.4 le modèle spatio-temporel à deux niveaux utilisés pour construire le résumé d'événements. Ensuite, en s'appuyant sur notre modèle ainsi défini, nous

modélisons notre structure d'agrégation dans la section 2.6. Finalement, nous concluons dans la section 2.7.

## 2.4 Modèle spatio-temporel

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 2.2, 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 obtenir des mis à jour entre eux et le serveur central.

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 agrégées et va être utilisé comme référentiel commun entre les véhicules lors des échanges ;

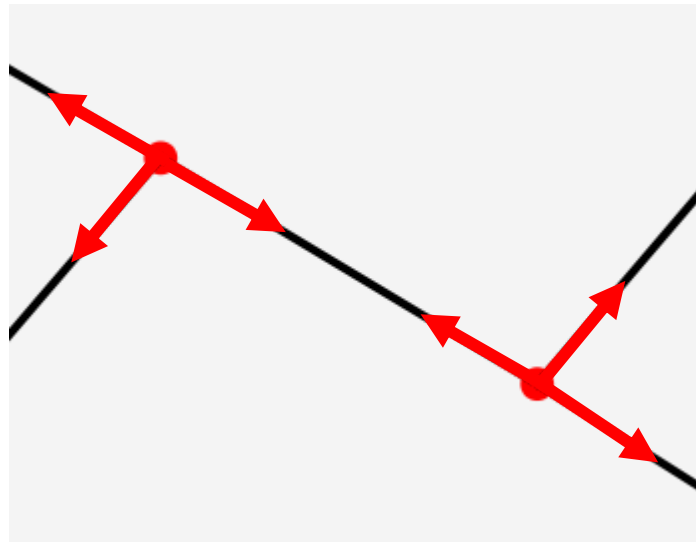
2. « le niveau logique » choisi par chaque conducteur en se basant sur le niveau physique.

### 2.4.1 Niveau Physique

Le niveau le plus bas, appelé niveau physique, constitue un référentiel partagé entre tous les véhicules. Il est représenté en nœuds et arcs physiques qui représentent les intersections et les directions, l'ensemble forme une partition complète de l'espace et uniforme pour tous les véhicules comme l'illustre la figure 6 sur une partie de la ville de Laghouat.



a) Niveau physique, un nœud.



b) Niveau physique, un arc.

**Figure 6. Niveau physique (portion de Laghouat)**

Chaque nœud représente une intersection, et est identifié par les coordonnées  $(X, Y)$ . Chaque arc est représenté par deux nœud et le sens giratoire. 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.

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 weekend dans un quartier d'affaires par exemple).

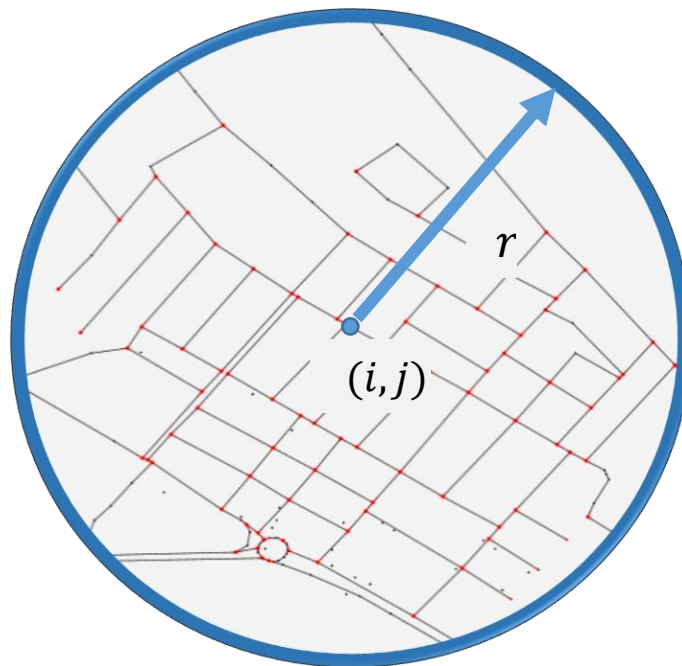
Le couple {arc, segment temporel} est la plus petite unité pouvant comptabiliser des occurrences d'événements. Cette structure permettrait d'avoir une résolution spatiale meilleure dans les zones urbaines (et donc une meilleure précision) car il y'a plus d'intersection, par contre nous aurons moins de couples dans les routes non urbaines ce qui nous donne un avantage de réduction de la taille des informations.

#### 2.4.2 Niveau logique

En se basant sur le niveau physique, chaque véhicule peut construire sa propre zone logique représentant les zones spatiales. Ce niveau logique est constitué d'un ensemble de cercles eux-mêmes constitués d'un ensemble de nœuds du niveau physique (les rectangles pointillés sur la Figure 7(a)).



a) Découpage spatial pour le niveau logique



b) Découpage spatial pour le niveau logique

**Figure 7. Niveau logique**

Une zone d'intérêt est définie comme un cercle construit sur un nombre entier de nœuds physiques (Figure 7(b)). Elle est définie par un couple  $((i, j), r)$  : coordonnées du nœud central  $(i, j)$  et le rayon  $r$ .

Nous supposons que nous avons  $g$  granularités temporelles basées sur le 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 gème granularité de 22h à 24h). Un conducteur peut ne pas s'intéresser à tout l'espace et le nombre de zones physiques réellement observées au niveau logique est donc très petit relativement au niveau physique.

## 2.5 Construction des résumés des événements

Nous nous appuyons sur les deux niveaux spatio-temporels du modèle que nous avons proposé dans la section précédente. Le résumé doit avoir l'apparition la plus probable 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. L'analyse de ces données nous estime la plus forte probabilité. 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 agréger et analyser les données récoltées, nous choisissons parmi les approches d'agrégation étudiés dans l'article [26] une approche adaptée de TAG (Textual Aggregation by Graph) que nous introduisons dans la section 2.5.1. Ensuite, nous détaillons notre structure d'agrégation des événements dans la section 2.6.

### 2.5.1 L'Agrégation par graphe

L'approche de TAG [26] est une méthode d'agrégation de mots clés de plusieurs documents basés sur la théorie des graphes. Cette fonction a comme but l'extraction à partir d'un ensemble de mots clés de l'ensemble des mots clés les plus représentatifs en utilisant un graphe. Cette fonction prend comme entrée l'ensemble de tous les termes extraits  $T$  et donne comme une sortie un ensemble ordonné,  $KW$ , qui contient les mots clés les plus représentatifs. Le processus d'agrégation se fait en 3 étapes :

1. Extraction des mots clés et leurs fréquences,
2. Construction de la matrice d'affinité et le graphe d'affinité, et
3. Construction des cycles et la sélection des mots clés agrégés.

Dans notre contexte nous adaptons ce processus pour traiter et agréger les données véhiculaires telles qu'au lieu d'extraire la fréquence d'utilisation des mot clés dans un document, notre fonction aura comme données d'entrées l'ensemble  $T$  des fréquences d'occurrence d'un évènement dans un arc organisé sous forme d'une matrice de fréquences  $FM(i, j) = FE_{ij}$  tel que  $FE$  : la fréquence d'occurrence de l'évènement  $E_i$  dans l'arc  $A_j$ . Cette matrice est reconstruite chaque fois que le nombre des évènements reçus atteint un seuil défini par l'administrateur du système. Ce seuil varie par rapport aux capacités hardware de la machine.

Deuxième étape consiste à construire la matrice d'affinités ( $AffM$ ) entre les arcs à partir de la table des fréquences  $FM$  telle quel  $AffM[a_1, a_2] = Aff_{a_1, a_2}$  l'affinité entre l'arc  $a_1$  et l'arc  $a_2$  est obtenu par la fonction de calcul (1), (2). Nous utilisons aussi la matrice d'adjacence ( $AdjM$ ) pour obtenir  $Adj_{a_1, a_2}$  l'adjacence entre l'arc  $a_1$  et l'arc  $a_2$  où

$$AdjM[a_1, a_2] = \begin{cases} 1 & \text{si } a_1 \text{ et } a_2 \text{ sont adjacents} \\ -1 & \text{sinon} \end{cases}$$

La fonction de calcul de l'affinité entre l'arc  $a_1$  et  $a_2$  est comme suit :

$$Aff_{a_1, a_2} = (f(k, a_1) + f(k, a_2)) \times AdjM[a_1, a_2] \quad (1)$$

Où  $k$  : nombre d'évènements enregistrés au serveur, et

$$f(k, j) = \begin{cases} FE_{1,j} & \text{si } k = 1 \\ FE_{k,j} \times \mu + f(k-1, j) \times (\mu - 1) & \end{cases} \quad (2)$$

Tel quel  $\mu \leq 0.5$

Le nombre d'évènements enregistré au serveur est limité par un seuil défini par l'administrateur du système. Ce seuil varie par rapport aux capacités hardware de la machine. L'affinité entre deux arcs  $a_1$  et  $a_2$  adjacent consiste à la somme des moyennes pondérées des fréquences d'occurrence de l'évènement. Le graphe d'affinité est formellement défini par un graphe pondéré  $G = (N, L, f)$ . Tel que  $N = \{A_1, A_2, \dots, A_n\}$  est l'ensemble des arêtes défini par les couples  $(a_i, a_j)$  et  $f$  est la fonction du poids défini par

$$f(a_i, a_j) = AffM[i, j].$$

En troisième étape, notre approche consiste d'agréger les affinités en cycles. Puis pour connaître la plus grande probabilité d'occurrence d'un évènement  $E$  à partir d'un arc  $a$ , il suffit de trouver le cycle avec le plus grand poids que  $a$  appartient. Le processus de construction de cycles en utilisant le graphe d'affinité comme suit :

1. Initialiser  $p = 1$ .
2. Commencer avec un cycle vide  $C_p = \emptyset$ , et initialiser  $m = 1$ .
3. Sélectionner le premier arc  $a_i$  aléatoirement de  $N$ . Mettre  $c_m = a_i$ ,  
 $C_p = C_p \cup \{c_m\}$ ,  $m++$ .
4. Si  $m < 2$  alors trouver  $a_i \neq a_j / f(a_i, a_j) = \text{Max } f(a_i, a_k)$  pour tout  $a_i \neq a_k$  et  $a_k \geq 0$  si non trouver  $a_i \neq a_j$  et  $a_j \neq c_{m-1} / f(a_i, a_j) = \text{Max } f(a_i, a_k)$  pour tout  $a_i \neq a_k$  et  $a_k \neq c_{m-1}$  et  $a_k \geq 0$ .
5.  $c_m = a_j$ ,  $C_p = C_p \cup \{c_m\}$ ,  $m++$ .
6. Si  $c_m \notin C_p$  alors aller à 4, sinon  $N = N - C_p$ ,  $p++$  aller à 2.

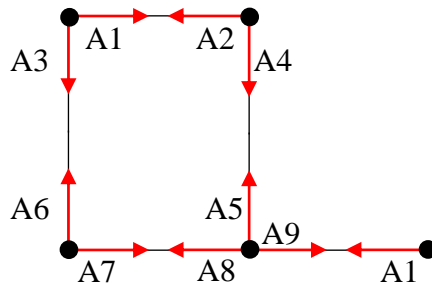
Après la construction des cycles on calcule l'affinité moyenne de chaque cycle et nous les sauvegardons dans le serveur. Chaque cycle sauvegardé est relié par l'espace-temps qui le concerne. Autrement dit, à la fin de la construction des cycles nous obtiendrons dans chaque segment de temps l'ensemble C des cycles construits à partir des évènements où ils ont eu occurrence.

### 2.5.2 Exemple

Nous prenons comme exemple l'observations des évènements E1 ... E13 de type espace parking libre entre cinq intersections, dans les arcs A1... A10 figurés dans le graphe (Figure 8) qui est, ensuite traduit en matrice d'adjacence (Tableau 2). Nous obtenons la matrice de fréquences Tableau 1 :

**Tableau 1. Matrice des Fréquences (FM)**

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
E1	10	9	22	15	9	20	15	9	28	39
E2	15	22	26	0	9	16	11	0	25	0
E3	5	15	0	15	22	0	15	0	0	0
E4	0	16	0	0	15	10	0	0	0	0
E5	16	12	2	13	16	12	0	12	2	0
E6	21	0	19	21	17	9	0	0	10	0
E7	13	0	14	0	0	15	1	0	17	0
E8	17	0	8	0	0	8	0	18	20	0
E9	22	14	0	0	14	21	0	17	0	0
E10	0	7	0	0	7	0	15	18	20	0
E11	5	18	10	5	15	15	15	18	20	0
E12	20	4	7	17	4	7	0	5	3	105
E13	1	10	11	1	10	17	0	16	10	0



**Figure 8. Graphe exemple**

Tableau 2. Matrice d'Adjacence (AdjM)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	-1	1	-1	-1	-1	1	-1	-1	-1	-1
A2	1	-1	-1	-1	1	-1	-1	-1	-1	-1
A3	-1	1	-1	-1	-1	1	-1	-1	-1	-1
A4	1	-1	-1	-1	1	-1	-1	-1	-1	-1
A5	-1	-1	-1	1	-1	-1	1	-1	-1	1
A6	-1	-1	1	-1	-1	-1	-1	1	-1	-1
A7	-1	-1	1	-1	-1	-1	-1	1	-1	-1
A8	-1	-1	-1	1	-1	-1	1	-1	-1	1
A9	-1	-1	-1	1	-1	-1	1	-1	-1	1
A10	-1	-1	-1	-1	-1	-1	-1	-1	1	-1

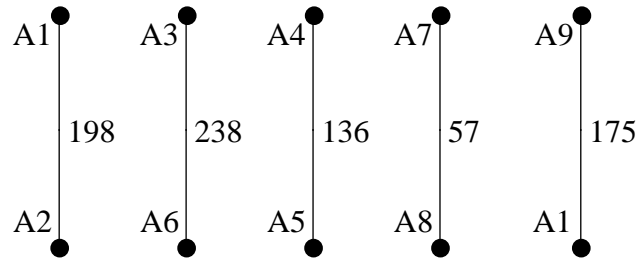
Comme expliqué dans la deuxième étape de la section 2.5.1 nous construisons la matrice d'affinité Tableau 3.

Tableau 3. Matrice D'affinité (AffM)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	-145	198	-259	-180	-274	200	-119	-125	-283	-175
A2	198	-127	-135	-141	200	-280	-120	-168	-197	-157
A3	-259	135	-119	-143	-155	238	-114	-112	-242	-173
A4	180	-141	-143	-87	136	-152	-80	-111	-145	-197
A5	-274	-200	-155	136	-129	-283	117	-171	-199	157
A6	-200	-280	238	-152	-283	-150	-108	-170	-283	-171
A7	-119	-120	114	-80	-117	-108	-67	57	-132	-54
A8	-125	-168	-112	111	-171	-170	57	-95	-161	158
A9	-283	-197	-242	145	-199	-283	132	-161	-143	175
A10	-175	-157	-173	-197	-157	-171	-540	-158	175	-144

Tableau 4. Extraction des cycles à partir de la Matrice d'Affinité (AffM)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	-145	198	-259	-180	-274	200	-119	-125	-283	-175
A2	198	-127	-135	-141	200	-280	-120	-168	-197	-157
A3	-259	135	-119	-143	-155	238	-114	-112	-242	-173
A4	180	-141	-143	-87	136	-152	-80	-111	-145	-197
A5	-274	-200	-155	136	-129	-283	117	-171	-199	157
A6	-200	-280	238	-152	-283	-150	-108	-170	-283	-171
A7	-119	-120	114	-80	-117	-108	-67	57	-132	-54
A8	-125	-168	-112	111	-171	-170	57	-95	-161	158
A9	-283	-197	-242	145	-199	-283	132	-161	-143	175
A10	-175	-157	-173	-197	-157	-171	-540	-158	175	-144



**Figure 9. Construction des Cycles**

La troisième étape est la construction des cycles (Tableau 4 ,Figure 9) comme expliqué dans l'algorithme dans la section 2.5.1. Dans cet exemple, nous n'avons pas pris un grand nombre d'arcs, ce qui a donné comme résultat des petits cycles de deux nœuds. Dans ce scénario, Si un véhicule situé dans l'arc A3 cherche un emplacement de parking libre, il est plus probable de trouver un, s'il tourne vers l'arc A6 dans la prochaine intersection.

## 2.6 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) :

- *tye* : il s'agit du type de l'événement observé (accident, libération place de parking, etc.) ;
- *loe* : 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 ;
- *idfe* : identifiant unique de l'occurrence d'événement observé. Nous supposons qu'une occurrence d'événement produit toujours le même identifiant sur le véhicule  $V$ .

Nous proposons de décrire les résumés par deux structures de données pour le serveur d'agrégation décrite dans la figure 10 et pour les véhicules, décrite dans la figure 11. La première est constituée de :

- Un ensemble de types d'évènements définis chacun par un identifiant unique ( $tye$ );
- $p$  Pointeurs sur un tableau de  $g$  segment de temps représentant les granularités temporelles. Dans notre étude nous avons pris des segments temporels de 2 heures, donc nous obtiendrons  $g = 84$  granularités temporelles durant la semaine.
- Chaque segment de temps contient deux pointeurs :
  - o Un pointeur sur la liste des cycles agrégeant les affinités des arcs, de type  $tye$ , observés dans la cellule pour une granularité temporelle  $g$ . Chaque affinité contient deux arcs  $a_1$   $a_2$  et sa valeur.
  - o Un deuxième pointeur sur la liste des évènements de type  $tye$ , observés dans la cellule pour une granularité temporelle  $g$ . Chaque évènement contient sa valeur et l'arc où il a eu occurrence.

La deuxième structure qui concerne les véhicules est constituée de :

- Un ensemble de zones d'intérêts définies chacune par un identifiant unique ( $idfz$ ). Chaque zone représente une zone logique  $((i, j), r)$ . Le tableau des zones d'intérêt est trié selon les valeurs croissantes de  $i$  ;
- $p$  Pointeurs sur un tableau de  $g$  segments de temps représentant les granularités temporelles.
- Chaque segment de temps contient un pointeur sur la liste des arcs avec la valeur du plus grand cycle où cet arc appartient.

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 zone associée et le segment de temps 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).

## 2.7 Conclusion

Ce chapitre décrit l'objectif de notre mémoire : proposer une solution pour construire le résumé d'événements. Nous avons présenté une structure d'agrégation de données spatio-temporels observés qui soit utilisable dans un contexte d'échanges de données inter-véhiculaires et serveur central. 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.

Le deuxième objectif de ce travail est d'implémenter l'approche et simuler le comportement des véhicules en utilisant l'expérience et les données agrégées reçu par le serveur. Nous présentons dans le chapitre suivant l'implémentation et la validation par simulation.

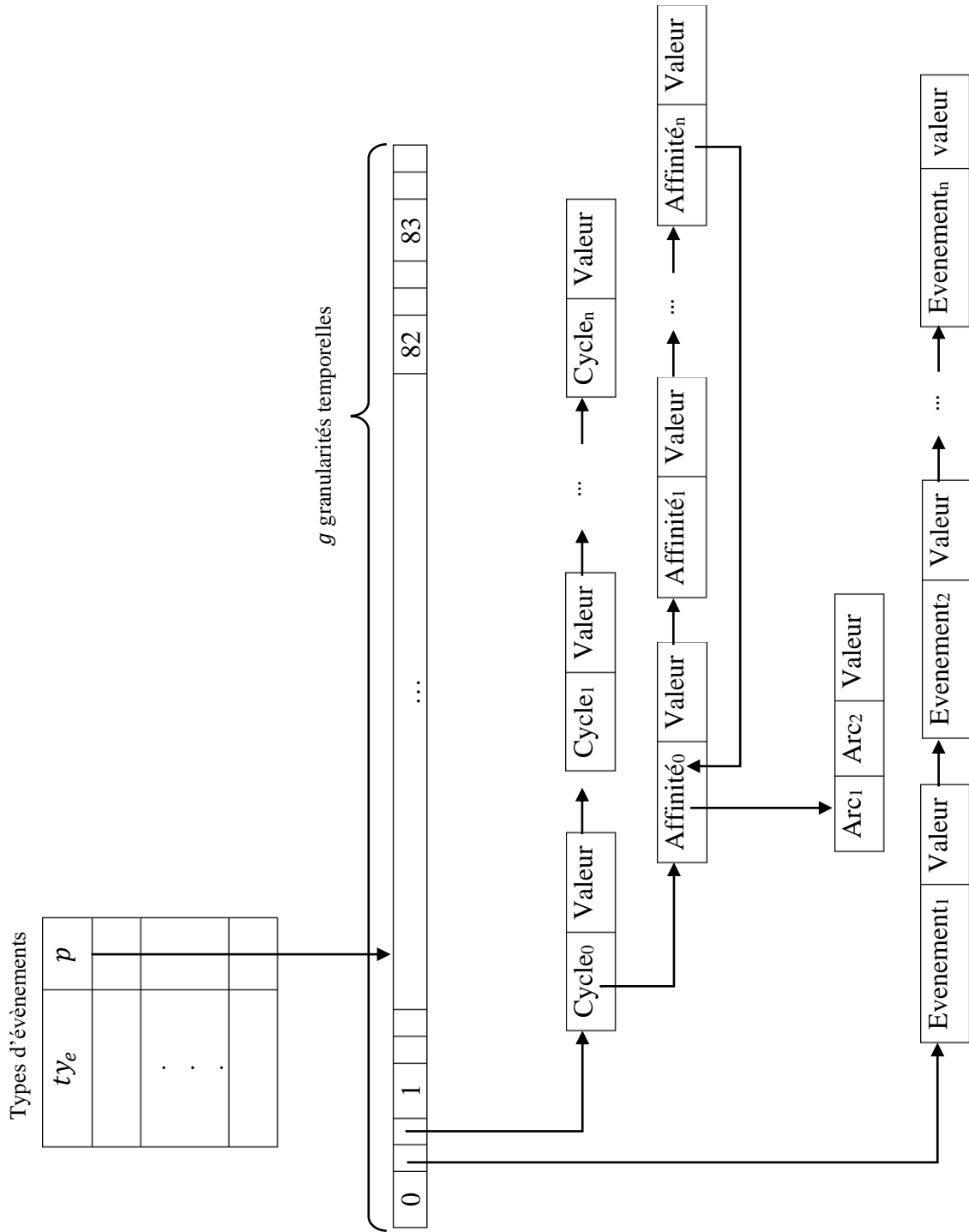


Figure 10. Structure d'agrégation (serveur)

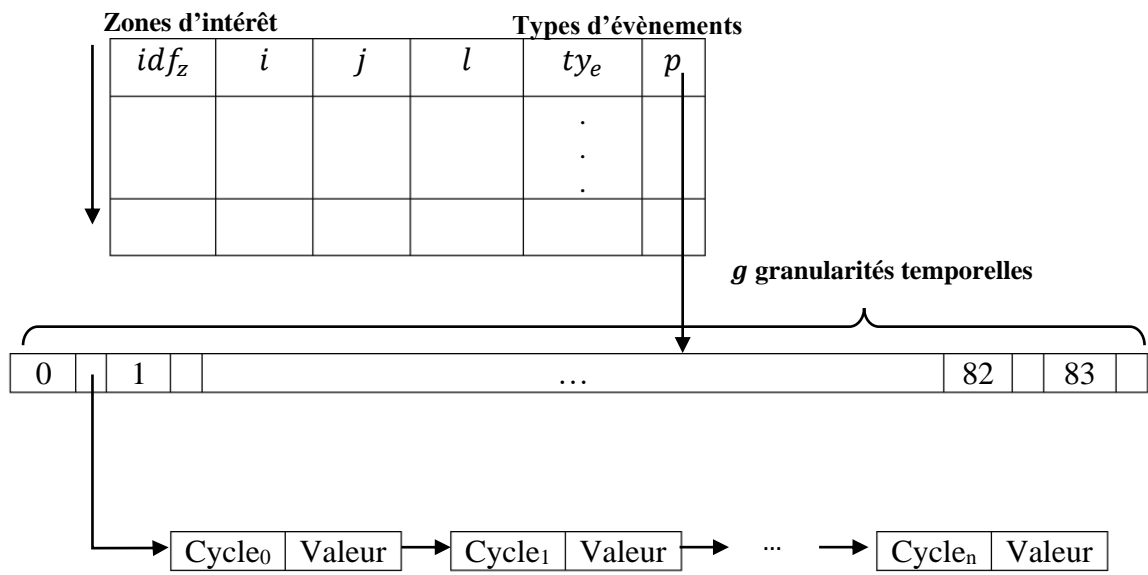


Figure 11. Structure d'agrégation (véhicule)

## Chapitre 3 Implémentation et validation expérimentale

### 3.1 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 de données, nous avons choisi dans ce chapitre de nous focaliser sur l'embouteillage.

Dans un premier temps, nous choisissons un simulateur et nous procédons à une étape de déploiement de nos modules implémentés afin d'adapter le simulateur à nos besoins en termes 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 3.2, nous exposons une étude des simulateurs existants qui nous a permis de choisir le simulateur le plus approprié décrit dans la section 3.2.1. Ensuite nous allons montrer comment nous avons adapté ce simulateur à nos besoins dans la section 3.2.3. Pour préparer nos simulations nous présentons dans la section 3.3 les paramètres évalués et dans la section 3.4 les stratégies utilisées. Enfin, les deux sections 3.5 et 3.6 montrent les différents résultats obtenus.

### 3.2 Choix du simulateur

Dans cette section, nous présentons les simulateurs existant, puis nous exposons le simulateur que nous avons développé pour évaluer notre structure d'agrégation proposée dans le Chapitre 2.

### 3.2.1 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 côté, 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<sup>3</sup> ou TRANSIMS<sup>4</sup> [27] et d'autres sont libres comme VanetMobiSim [28] ou GrooveNet<sup>5</sup> [29];
2. D'un autre côté, nous devons considérer les simulateurs généraux de réseaux. Les plus connus sont NS-2<sup>6</sup>, 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 [30], il est écrit en Parsec, langage dédié basé sur le C ; JiST-SWANS<sup>7</sup> [31]. Quelques travaux proposent d'intégrer les deux, comme TraNS<sup>8</sup> [32].

Le simulateur que nous devons choisir doit répondre à certains critères propres à nos besoins. En effet, le simulateur doit nous permettre de représenter l'environnement graphique qui permet de bien observer nos applications, 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

---

<sup>3</sup> <http://mctrans.ce.ufl.edu/featured/tsis/version5/corsim.htm> [37]

<sup>4</sup> <https://www.fhwa.dot.gov/planning/tmip/resources/transims/>

<sup>5</sup> <http://mlab.seas.upenn.edu/projectsites/groovenet/index.html>

<sup>6</sup> <http://www.isi.edu/nsnam/ns/> [36]

<sup>7</sup> <http://jist.ece.cornell.edu/>

<sup>8</sup> <http://lca.epfl.ch/projects/trans/>

l'événement et à quel moment. De plus, le comportement des conducteurs doit pouvoir être influencé par les données 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 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 développé un simulateur propre à nos contributions que nous avons nommé « VeGASim » (Vehicular Graphical Aggregation Simulator) pour qu'il corresponde aux critères propres à nos besoins. Dans ce qui suit nous allons décrire le simulateur « VeGASim » et l'extension réalisée.

### 3.2.2 Simulateur « VeGASim »

Ce simulateur que nous avons développé en Java est utilisé pour simuler des déplacements de véhicules dans un réseau routier et pour évaluer différents protocoles de routage dans différentes conditions de circulation afin d'étudier leurs impacts sur le trafic. Pour simuler le réseau et les protocoles de routage « VeGASim » utilise et communique avec le simulateur de réseau NS3<sup>9</sup> [38].

---

<sup>9</sup> <https://www.nsnam.org/>

Le simulateur VeGASim 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 OpenStreetMap<sup>10</sup> [33];
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 créées. Chaque route possède sa propre limitation de vitesse.

Pour les parkings de supermarché, le simulateur « VeGASim » 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.

---

<sup>10</sup> <http://www.openstreetmap.org>

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).

### 3.2.3 Adaptation du simulateur

Pour évaluer notre structure d'agrégation et notre processus d'échange, nous nous sommes concentrés sur deux 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, et pour trouver le chemin qui évite la congestion et l'embouteillage. Pour se faire, nous avons étendu le simulateur « VeGASim » 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 au niveau du serveur. Nous avons utilisé la fonction d'agrégation présentée dans la section 2.5.1 pour construire les résumés/agrégats des événements et des semaines et construire notre structure d'agrégation présentée dans la section 2.6;
2. Dans l'échange, nous nous sommes basés sur le mécanisme de Multi-Client - Serveur qui permet à plusieurs véhicules de communiquer avec le serveur. Nous avons intégré un module pour la définition des priorités de chaque véhicule que nous allons prendre en considération lors de l'échange des résumés ;
3. Dans l'exploitation, nous avons ajouté deux modules pour exploiter les résumés et pour envoyer au conducteur l'aide pour la prise de décision. Dans le premier module nous avons implémenté une fonction qui ajoute un poids calculé à partir des informations reçues du serveur de type **embouteillage** sur les arcs pour influencer la fonction de calcul du plus court chemin. Au deuxième module nous avons implémenté une fonction qui nous renvoie l'emplacement du **parking** dont la probabilité de trouver un emplacement libre est la plus grande en utilisant les informations reçues du serveur d'agrégation.

### 3.3 Paramètres de simulation

Dans nos simulations, nous avons considéré un environnement correspondant au centre de Laghouat, une ville située au centre de L'Algérie. Cette zone a été représentée par une surface de 2,5 km<sup>2</sup>. 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 5.

**Tableau 5. Caractéristiques des 8 parkings considérés durant les simulations**

<b>Paramètres</b>	<b>Initialisation</b>
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%

Les véhicules dans les simulations se déplacent à 30kmh. 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 avec le serveur 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 le serveur de la libération de cette place.

### 3.4 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] :

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 le 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 mise 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. « SummaryAggregation + Infrastructure » cette stratégie prend en compte l'information fournie par le serveur central d'agrégation pour les conducteurs. Cette information est sous la forme de la structure d'agrégation spatio-temporelle. Cette structure d'agrégation intégrée est remplie par les événements observés par le véhicule, récoltés et agrégés par le serveur depuis

le début de la simulation dans une zone logique souhaitée. Avec cette stratégie, le conducteur est guidé vers la zone la plus proche ayant la plus haute fréquence de places de stationnement libérées. Cette stratégie est représentée par notre system VeGAS.

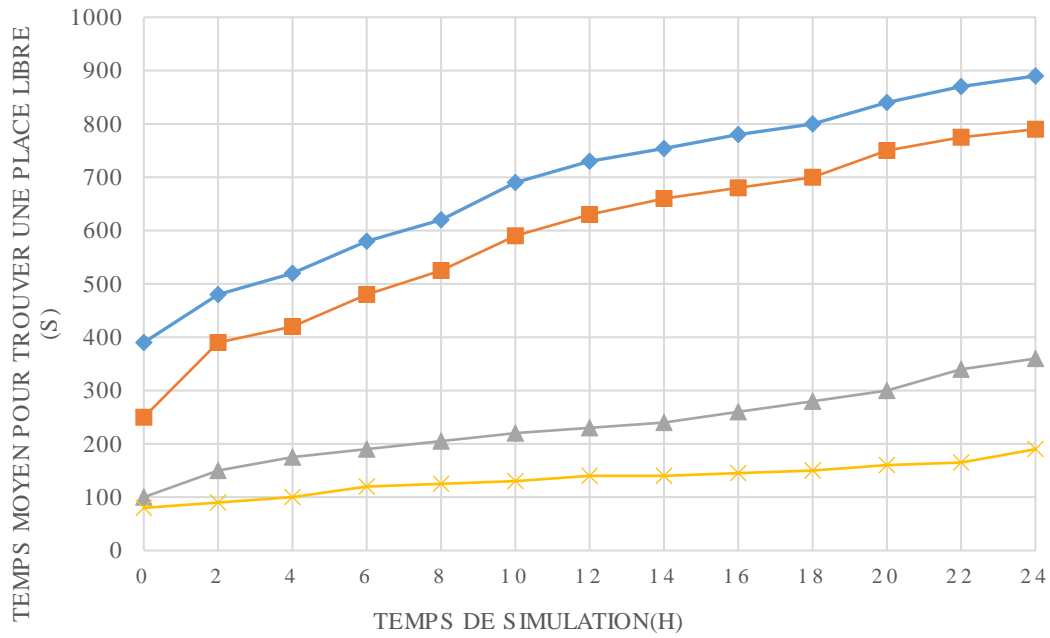
7. « View + Dissemination + SummaryAggregation » (resp. View + Dissemination + Infrastructure) : cette stratégie combine « View », « Dissemination » et « SummaryAggregation » (resp. « Infrastructure »). Ainsi, lorsqu'un conducteur cherche une place de stationnement libre, s'il n'observe aucune place disponible sur son chemin et ne reçoit pas de message d'un autre véhicule libérant sa place, la structure d'agrégation spatiotemporelle 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 », « View + Dissemination + SummaryAggregation » et « SummaryAggregation + Infrastructure »), 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. 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. Les stratégies 1, 2, 3, 4, 5 et 7 ont été implémenté et simulé dans [1] où Dorsaf et al. ont utilisé les mêmes paramètres de simulation que les notre, donc nous comparons directement les résultats de l'implémentation de notre stratégie 6 avec leurs.

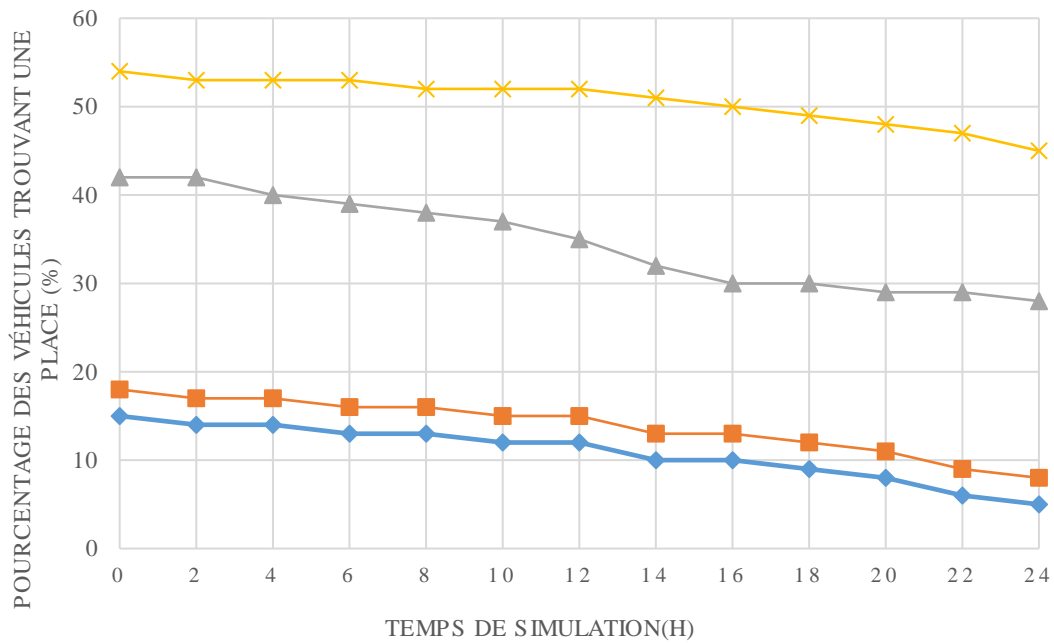
### 3.5 Evaluation qualitative du résumé spatio-temporel

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'évaluations des stratégies sans échange et avec échange. La Figure 12 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 ».



(a) Temps moyen



(b) Pourcentage de véhicules

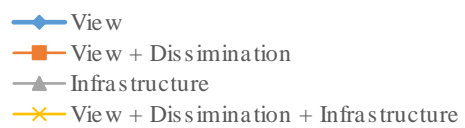
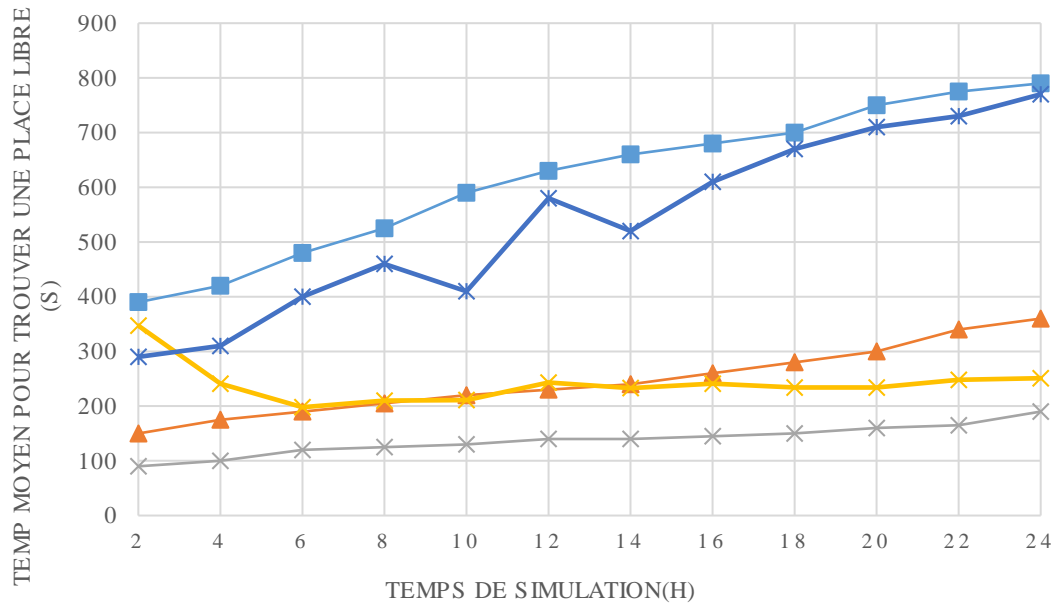
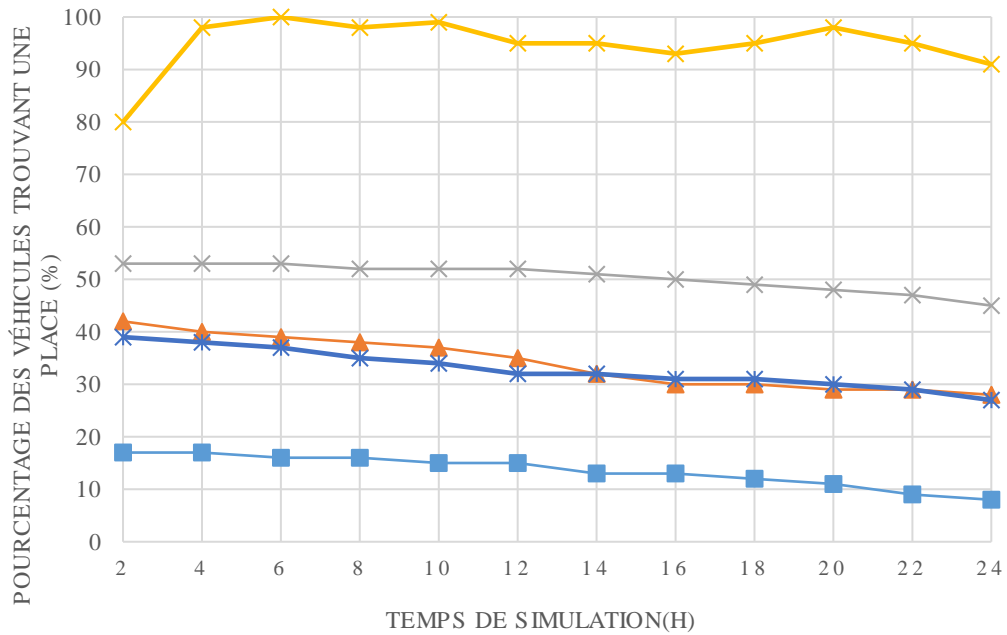


Figure 12. Comparaison des différentes stratégies sans échange



(a) Temps moyen



(b) Pourcentage de véhicules

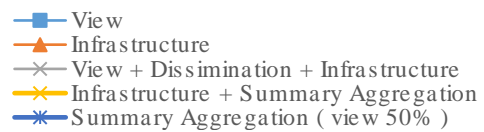


Figure 13. Impact du processus d'échange

Dans la Figure 12, 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 55). 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 12 montrent que la stratégie « 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 12(a)) que le pourcentage de véhicules trouvant une place de parking (Figure 12(b)). En outre, « View + Dissemination + Infrastructure » est la stratégie ayant les meilleurs résultats ce qui montre que les stratégies élémentaires correspondantes sont complémentaires.

Dans la figure 13, nous introduisons le processus d'agrégation partielle au niveau de chaque véhicule (par exemple, la stratégie 4) et au niveau du serveur (stratégie 6) et le comparons avec les stratégies étudiées dans la Figure 12.

Si nous comparons les résultats du « SummaryAggregation + Infrastructure » avec ceux obtenus pour la stratégie de « Infrastructure », nous constatons que les temps moyens sont plutôt proches, quand il n’y a pas d’embouteillage, et s’améliore après. Cela montre donc que notre approche apporte un gain, même avec une qualité moindre des résumés.

La Figure 13. Impact du processus d’échange montre que la réalisation d’échanges d’agrégats entre les véhicules et l’infrastructure améliore considérablement les résultats obtenus avec la stratégie « SummaryAggregation + infrastructure » surtout pour le pourcentage des véhicules trouvant une ressource. Les résultats sont alors très proches de ceux obtenus avec la stratégie « Infrastructure » au temps moyen pour trouver un emplacement libre. Les simulations avec échanges montrent que de bons résultats peuvent être obtenus avec les stratégies basées sur «SummaryAggregation + Infrastructure» basé sur notre système VeGAS, les résultats des simulations montrés dans la Figure 13(b) sont presque égaux à ceux obtenus avec « infrastructure » et s’améliore après que le nombre de véhicules augmente.

### 3.6 Évaluation du processus d’échange

Nous comparons ici le nombre des échanges des agrégat entre les véhicules avec les échanges véhicule - serveur. Comme le nombre des échanges de messages as un grand impact sur la surcharge de la bande passante, le débit de communication et surtout la consommation de l’énergie, l’analyse de cet aspect est importante. Pour cela nous avons réalisé d’autres séries de simulations. Nous reprenons le même environnement correspondant à la ville de Laghouat considéré dans les simulations précédentes avec les mêmes paramètres définis dans le Tableau 55. Comme dans les simulations précédentes une phase d’initialisation de la structure est réalisée (remplissage de la structure d’événements pendant 24h). Les stratégies utilisées d’aide au conducteur dans la comparaison sont VeGAS « SummaryAggregation + infrastructure » et « SummaryAggregation » avec portée de vision à 50% qui est étudié et simulé dans [1]. 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).

**Tableau 6. Nombre de messages échangé pour les différentes stratégies**

<b>Strategie</b>	<b>Nb nœds</b>	<b>Heure simulation</b>	<b>Nb échanges total</b>
VeGAS	100	1h	231
	200	2h	540
SummaryAggregation avec portée d'échange de 100 mètres [1]	100	1h	2014
	200	2h	7259
SummaryAggregation avec portée d'échange de 50 mètres [1]	100	1h	984
	200	2h	4275

Le Tableau 6 montre les échanges entre 100 (resp. 200) véhicules pendant une heure puis au bout de deux heures de simulations pour une distance maximale d'échange de 100 mètres, 50 mètres pour la stratégie du « SummaryAggregation » et pour notre stratégie VeGAS. Il faut tenir compte du fait que plus le nombre des véhicules est petit, et plus la stratégie VeGAS performe mieux que « SummaryAggregation » comme nous l'avons montré dans les résultats de la simulation précédente section 3.5. L'analyse nous montre que le nombre de messages échangé dans la stratégie VeGAS est beaucoup moins que le nombre de messages échangé des autres stratégies, car notre système n'a pas besoin de faire des échanges avec les autres véhicules pour construire une connaissance collaborative. Par contre la deuxième stratégie que [20] ont implémentée, pour construire une base de connaissance collaborative, un échange des agrégats est nécessaire chaque fois qu'un véhicule rencontre un autre. Alors que dans la stratégie VeGAS les véhicules ne font des échanges avec l'infrastructure que pour signaler un nouvel événement (la

libération d'une place de stationnement dans notre exemple) ou pour faire une mise à jour des connaissances quand la durée de vie des agrégats spatio-temporels du véhicule est achevée.

### 3.7 Conclusion

Ce chapitre a été consacré à la mise en œuvre et la validation de notre approche VeGAS (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 « VeGASim. » 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 le serveur et 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 nombre des messages nous ont montrées que nous arrivons facilement à avoir qu'on obtient avec notre système les meilleurs résultats avec très peu nombre de messages échangés par rapport à d'autres solutions d'agrégation aux réseaux véhiculaires.

---

## Chapitre 4 Conclusion et perspectives

L'objectif principal de ce travail 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 données spatio-temporels 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 notre contribution, une structure d'agrégation de données spatio-temporels observés qui seront utilisables dans un contexte d'échanges inter-véhiculaire qu'on a appelé VeGAS. Elle permet d'approximer la fréquence des types d'événements observés. Cette structure est construite sur un modèle spatiotemporel 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.

Pour la validation et la mise en œuvre de l'ensemble des solutions proposées, nous avons développé le simulateur « VeGASim. » 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 en développant deux applications, la première pour trouver un stationnement libre aux parkings et la deuxième pour éviter l'embouteillage. 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 les véhicules et le serveur améliore nettement la qualité des résumés et la quantité des messages échangés.

Finalement, le travail effectué dans ce mémoire 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.

## Perspectives

Dans nos futurs travaux, comme notre système est réalisable avec la disponibilité de la technologie internet mobile à tous les gens, nous volons le réaliser dans de vrais scénarios. Tout d'abord :

- Déployer un serveur VeGAS sur un support physique puissant qui peut traiter plusieurs requêtes en parallèle.
- Créer des applications mobiles téléchargeables par le grand public qui utilise les données du serveur pour plusieurs types d'évènements qui ont rapport avec les véhicules et le trafic routier.
- Développer et mettre en disposition des API et des bibliothèques en plusieurs langages de programmation pour les développeurs d'applications.
- Rendre notre VeGAS accessible en ligne pour le mettre en disponibilité aux contributeurs, les étudiants et les chercheurs pour l'améliorer

D'un autre côté nous voulons adapter notre approche d'agrégation dans d'autres domaines :

- Médical, en analysant les données d'un patient pour faire une prédiction d'une maladie, ce qui va permettre de la traiter avant son occurrence. Ou d'analyser les données de santé des gens d'une région pour prédire une contamination.

- Explorer l'usage des IoT, en analysant les données des périphériques connectées pour prédire l'occurrence d'un évènement.
- Les réseaux sociaux, en analysant l'activité d'un ou plusieurs individus pour prédire leur relation avec plusieurs groupe malicieux.

---

## Bibliographies

- [1] Z. Dorsaf, «Agrégation et extraction des connaissances dans les réseaux inter-véhicules." PhD dissertation, Institut National des Télécommunications, SudParis, 2013.» Paris, France, 2013.
- [2] M. K. BERRAH, «Parc National Automobile au 31/12/2014,» Juillet 2014.
- [3] Office Nationale des Statistique, «Démographie Algérienne 2014,» 2014.
- [4] N. Belghazi, «Système adaptatif multicritère d'avertissements véhiculaires basé sur le comportement des conducteurs,» Juin 2015.
- [5] M. Bouakkaz, S. Loudcher et Y. Ouinten, «A New Tool for Textual Aggregation In Information Retrieval,» chez *Proceedings of the 18th International Conference on Enterprise Information Systems*, Rome, Italy, 2016.
- [6] J.-L. L. Moigne, *Théorie du système général - Théorie de la modélisation*, France , 1977.
- [7] A. FILALI et S. KEDJNANE, «Conception et réalisation d'un Data Warehouse pour la mise en place d'un système décisionnel,» 2010.
- [8] F. Ravat, «Modèles et outils pour la conception et la manipulation de syst'emes d'aide à la décision,» Toulouse, France, 2009.
- [9] Inmon et W. H, *Building the Data Warehouse*, John Wiley & Sons, 1996.
- [10] F. RAVAT et O. TESTE, «Modélisation et manipulation de données historisées et archivées dans un entrepôt orienté objet,» chez

---

*Proceedings of the 11th International Conference on Database and Expert Systems Applications*, 2000.

- [11] US Department of Transportation, *Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems*, 2003.
- [12] «Status of Project IEEE 802.11p.» IEEE, Novembre 2006. [En ligne]. Available: [http://grouper.ieee.org/groups/802/11/Reports/tgp\\_update.htm](http://grouper.ieee.org/groups/802/11/Reports/tgp_update.htm).
- [13] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen et J.-P. Sheu, «The broadcast storm problem in a mobile ad hoc network,» chez *ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Seattle, WA, USA, Aout 1999.
- [14] F. Picconi, N. R. M. Gruteser et L. Iftode, «Probabilistic validation of aggregated data in vehicular ad-hoc networks,» chez *the 3rd international workshop on Vehicular ad hoc networks, ACM VANET*, New York, NY, USA, 2006.
- [15] C. Chen, «Location-based data aggregation in mobile ad hoc networks,» Stuttgart, Holzgartenstr, Germany, 2003.
- [16] C. Lochert, S. Bjorn et M. Mauve, «Probabilistic aggregation for data dissemination in VANETs,» chez *the fourth ACM international workshop on Vehicular ad hoc networks, VANET'07*, New York, NY, USA, 2007.
- [17] S. Eichler, C. Merkle et M. Strassberger, «Data Aggregation System for Distributing Inter-Vehicle Warning Messages,» chez *Proceedings. 2006 31st IEEE Conference on Local Computer Networks*, 2006.

- 
- [18] H. Saleet et B. Otman, «Location-based message aggregation,» chez *IEEE Global Communications Conference Workshops*, November 2007.
- [19] T. Delot, S. Ilarri, N. Cenerario et T. Hien, «Event Sharing in Vehicular Networks Using Geographic Vectors and Maps,» *Mobile Information Systems*, vol. 7, n° 11, p. 21–44, 2011.
- [20] Z. Dorsaf et B. Defude, «Évaluation de requêtes sur des résumés spatio-temporels pour l'assistance aux conducteurs de véhicules. (GEDSIP) Gestion des Données dans les Systèmes d'Information Pervasifs,» chez *INFORSID'10*, Marseille, France, 2010.
- [21] T. Nadeem, S. Dashtinezhad, C. Liao et Liviu, «Trafficview : Traffic data dissemination using car-to-car,» *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 8, n° 13, pp. 6-19, 2004.
- [22] A. Cheng et K. Rajan, «A digital map/GPS based routing and addressing scheme for wireless ad hoc networks,» *Intelligent Vehicle Symposium*, vol. 4, n° 13, pp. 17-20, 2003.
- [23] C. Lochert, B. Scheuermann, M. Caliskan et M. Mauve, «The Feasibility of Information Dissemination in Vehicular Ad-Hoc Networks,» chez *Forth Annual Conference on Wireless On demand Network Systems and Services (WONS'07)*, January 2007.
- [24] T. Yufei, K. George, C. Jeffrey, L. Feifei et P. Dimitris, «patio-Temporal Aggregation Using Sketches,» chez *In Proceedings of the 20th International Conference on Data Engineering, ICDE'04*, Washington, DC, USA, 2004.
- [25] P. Flajolet et M. Nigel, «Probabilistic Counting,» *J. Comput. Syst. Sci.*, vol. 31, n° 12, pp. 182-209, 1985.

- 
- [26] M. Bouakkaz, r. S. Loudche et Y. Ouinten, «Automatic textual aggregation approach of scientific articles in OLAP context,» chez *10th International Conference on Innovations in Information Technology (INNOVATIONS)*, 2014.
- [27] S. Eubank, «What makes a simulation useful? [TRANSIMS],» chez *IEEE International Conference on Systems, Man, and Cybernetics. IEEE SMC '99 Conference Proceedings.*, Tokyo, Japan, Japan, 1999.
- [28] M. Fiore, J. Harri, F. Filali et C. Bonnet, «Vehicular Mobility Simulation for VANETs,» chez *Proceedings of the 40th Annual Simulation Symposium, ANSS '07*, Washington, DC, USA, 2007.
- [29] R. Mangharam, DanielWeller, R. Rajkumar, P. Mudalige et F. Bai, «GrooveNet : A Hybrid Simulator for Vehicle-to-Vehicle Networks,» chez *Mobile and Ubiquitous Systems, Annual International Conference*, 2006.
- [30] X. Zeng, R. Bagrodia et M. Gerla, «GloMoSim: a library for parallel simulation of large-scale wireless networks,» chez *Twelfth Workshop on Parallel and Distributed Simulation, 1998. PADS 98. Proceedings.*, Banff, Alberta, Canada, Canada, 1998.
- [31] R. Barr, P. Z. J. Haas et D. R. v. Renesse, «JiST: An efficient approach to simulation using virtual machines,» *Software Practice & Experience*, vol. 35, n° 16, pp. 539-576, May 2005.
- [32] M. Piorowski, M. Raya, A. L. Lugo, P. Papadimitratos, M. Grossglauser et J.-P. Hubaux, «TraNS: Realistic Joint Traffic and Network Simulator for VANETs,» chez *ACM SIGMOBILE Mobile Computing and Communications Review*, 2008.

- 
- [33] OpenStreetMap Foundation, [En ligne]. Available: [/www.openstreetmap.org](http://www.openstreetmap.org). [Accès le 26 4 2017].
- [34] C. BUET, M. DAVID, B. DEAL et T. MASI, «L'évolution technologique et le progrès de la voiture pour l'environnement.,» [En ligne]. Available: <http://tpe-automobile.e-monsite.com/pages/introduction.html>.
- [35] «GrooveNet,» mLab at the University of Pennsylvania, [En ligne]. Available: <http://mlab.seas.upenn.edu/projectsites/groovenet/index.html>. [Accès le 24 April 2017].
- [36] «ns-2 - The Network Simulator,» [En ligne]. Available: <http://www.isi.edu/nsnam/ns/>. [Accès le 24 April 2017].
- [37] «CORSIM,» McTrans Center, University of Florida, [En ligne]. Available: <http://mctrans.ce.ufl.edu/featured/tsis/version5/corsim.htm>. [Accès le 24 April 2017].