Misbehaviour detection in vehicular networks

Pengwenlong Gu

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2018-ENST-0011

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité - Informatique et Réseaux

présentée et soutenue publiquement par

Pengwenlong GU

le 21 February 2018

Détection des Comportements Malveillants dans les réseaux véhiculaires

Directeur de thèse: **Ahmed SERHROUCHNI**

Co-directeur de thèse: **Cunqing HUA**

Composition du Jury

Bijan JABBARI, Professeur, George Mason University	Rapporteur
Pascal LORENZ, Professeur, Université de Haute-Alsace	Rapporteur
Guy PUJOLLE , Professeur, UPMC - Sorbonne Universités	Examinateur
Pascale MINET, Faculty, Inria-EVA	Examinateur
Ken CHEN, Professeur, Institut GALILEE – Université Paris 13	Examinateur
Lyes KHOUKHI, Maître de conférences HDR, Université de Technologie de Troyes	Examinateur
Rida KHATOUN, Maître de conférences, Télécom ParisTech	Encadrant
Ahmed SERHROUCHNI, Professeur, Télécom ParisTech	Directeurs de Thèse
Cunqing HUA, Professeur, Shanghai Jiao Tong University	Co-directeurs de Thèse

THÈSE

TELECOM ParisTech

école de l'Institut Mines-Télécom - membre de ParisTech

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

Misbehaviour Detection in Vehicular Networks

Pengwenlong GU

©

Le Département Informatique et Réseaux (INFRES)
LTCI, Télécom ParisTech, Université Paris-Saclay
46 Rue Barrault, Paris CEDEX 13, 75634, France

This thesis is set in Computer Modern 11pt,
with the LaTeX Documentation System

October 2017

Acknowledgments

I would like to take advantage of this opportunity to thank all those who supported me during my thesis. Foremost, I would like to express my sincere gratitude to my advisors Prof. Cunqing HUA and Prof. Ahmed SERHROUCHNI for the continuous support of my Ph.D study and research, for their patience, motivation, enthusiasm, and immense knowledge. My sincere thanks also goes to my supervisor Dr. Rida KHATOUN for his endless support, appreciable help and technical discussions.

Besides my supervisors, I would like to thank the jury members: Prof. Bijan JABBARI, Prof. Pascal LORENZ, Prof. Guy PUJOLLE, Prof. Ken CHEN, Dr. Pascale MINET and Dr. Lyes KHOUKHI, for their interest in my work and their insightful comments.

I am also deeply thankful to my schoolmates and friends at TELECOM ParisTech: Dr. Hao CAI, Dr. Jinxin DU, Jianan DUAN, Dr. Heming HUANG, Longguang LI, Jean-Philippe MONTEUUIS, Dr. Mengying REN, Dr. Mengdi SONG, Marion VASSEUR, Dr. You WANG, Dr. Jun ZHANG and Dr. Yimeng ZHAO for all the fun we have had in the last three years.

Last but not the least, I would like to thank my family: my parents Jian GU and Xiaoqing PENG, my grand-parents Yizhang PENG and Yuping WANG and my girlfriend Xiaomeng WANG for their spiritual support throughout my life.

Abstract

The integration of motor vehicles and information technology has become increasingly popular with the evolution of roadway safety, traffic grid efficiency, and infotainment requirements. To this end, several standardization efforts have been devoted to establish network architecture and communication protocols for vehicular networks. In the USA, the Wireless Access in Vehicular Environments (WAVE) standards suite is based on multiple cooperating standards. In Europe, the ETSI TC ITS working group defines standards for ITS systems.

For vehicular networks, the safety and security of the network architecture and protocols are of vital importance, which have been the central theme of the standardization in both USA and Europe. In general, the provided security services are based on three major mechanisms: Encryption algorithms, Public Key Infrastructure (PKI) and Pseudonymous. These security services provide basic protection for the privacy of users and integrity of messages in vehicular environments. However, several critical issues, such as identity management, message traceability and availability still exist in vehicular networks.

In this thesis, we focus on two major security issues: Sybil attack and radio frequency (RF) jamming attacks. Ranging from theoretical modelling and analysis, to practical algorithm design and optimisation.

Sybil attacks can cause damage in both Networking layer and Application layer in vehicular networks. Since the CSMA/CA is implemented in Networking layer, the cooperation among virtual nodes leads to the possibility of using more channel resource than other benign nodes. In Application layer, the virtual nodes also take part in communicating with other ITS stations. Under this circumstance, when a malicious node uses multiple pseudonyms at the same time, the virtual nodes, generated based on the usage of pseudonymous, can help to increase the influence of fake safety messages by broadcasting them to other benign nodes. In this thesis, we focus on the Sybil attack detection in vehicular networks based on the vehicle driving patterns. Relying on beacon information, we designed a data format Driving Pattern Matrix (DPM) to describe vehicle driving pattern within a time period. Thus, three different machine learning methods: Distance based clustering, Support Vector Machine (SVM) and k-nearest neighbours (kNN) are considered.

The main idea is to evaluate the similarity of vehicle driving patterns, then based on the variation of vehicles' driving pattern to distinguish the malicious nodes from the benign ones. The effectiveness of our proposed solutions is evaluated through extensive simulations based on SUMO simulator and MATLAB. The results show that the proposed detection method can achieve a high detection rate with low error rate even under a dynamic traffic environment.

Radio Frequency (RF) Jamming attacks constitute a major threat to the availability of the vehicular networks. In particular, if the control channel is under persistent jamming attacks, the victim vehicles may fail to receive the safety related messages from the Road Side Unit (RSU), which can possibly cause tremendous economic loss and claim human lives. In this thesis, we propose a cooperative relaying scheme to circumvent the control channel jamming problem in the vehicular networks, whereby the vehicles outside of the jamming area serve as relays to help forward the received control channel signal to the victim vehicles through another jamming-free service channel. By combining the signals from all relays using the Selection Combining (SC) or Maximum-Ratio Combining (MRC) methods, the spatial diversity provided by the relays can be effectively exploited to reduce the outage probability of the victim vehicles. Theoretical models are developed to characterize the performance of this cooperative anti-jamming relaying scheme, which take into account both the large-scale path loss and small-scale channel fading between relaying and victim vehicles under different jamming scenarios. We also propose a relay selection scheme to optimize the performance of the victim vehicles under the relay number constraint. Simulation results are provided to show the performance of the proposed cooperative relaying scheme and relay selection algorithm under different network settings.

Thus, we extend the anti-jamming problem into multi-antenna RSU scenarios and propose a two stage anti-jamming scheme for the control channel jamming issue in vehicular networks, which takes advantage of the multi-antenna diversity and spatial diversity provided by the RSU and relay vehicles to improve the transmission reliability of the victim vehicles. Theoretical models are developed to characterize the performance of this cooperative anti-jamming relaying scheme, the anti-jamming problem is modelled as Mixed-integer Programming (MIP) problem and then reformulated as a sequence of convex sub-problems. By iteratively solving this sequence of convex sub-problems, the optimal solution can be found. We also propose a heuristic relay selection algorithm by exploiting the special structure of the problem, which attempts to tranform the worst-caste vehicle into victim each iteration until the maximum link capacity is achieved. Simulation results show that proposed anti-jamming converges quickly (always within 10 iterations) and can effectively improve the network capacity, and the performance of the heuristic relay selection scheme is close to the optimal solution.

Contents

List of Tables

List of Figures

Résumé Détaillé de la Thèse

Introduction

Contexte et motivations

De nos jours, les réseaux véhiculaires sont prometteurs dans un certain nombre de services utiles axés sur le conducteur et le passager, qui comprennent des installations de connexion Internet exploitant une infrastructure disponible de manière à la demande, un système de télépéage et une variété de services multimédias. Il couvre tous les modes de transport et prend en compte tous les éléments du système de transport: le véhicule, l'infrastructure et le conducteur ou l'utilisateur, qui interagissent de manière dynamique. La fonction générale des réseaux véhiculaires est d'améliorer la prise de décision, souvent en temps réel, des contrôleurs de réseau de transport et d'autres utilisateurs, améliorant ainsi le fonctionnement de l'ensemble du système de transport.

Au-delà de tous leurs avantages, les réseaux véhiculaires soulèvent de nouveaux défis en ce qui concerne la sécurité et la protection de la confidentialité. Il est essentiel, par exemple, que les informations vitales ne puissent être ni modifiées ni supprimées par un attaquant et doivent également déterminer la responsabilité des conducteurs tout en préservant leur confidentialité. Il reste plusieurs problèmes critiques dans les réseaux véhiculaires, qui sont listés comme suit:

- **Problème de gestion d'identité**: La protection de la confidentialité, en particulier l'utilisation de pseudonymes, engendre la vulnérabilité de l'attaque Sybil. Après la demande de pseudonymes, les véhicules utiliseraient les pseudonymes comme identités de communication. Chaque pseudonyme valide a sa propre paire de clés de signature. Par conséquent, l'intégrité et la non-répudiation de l'information peuvent être garanties. Dans ce cas, si un véhicule utilise plusieurs pseudonymes ensemble en même temps, chaque pseudonyme est un véhicule individuel dans la vue des autres utilisateurs.

- **Problème de disponibilité**: La disponibilité est toujours vulnérable dans les réseaux, en particulier dans les environnements sans fil. Dans les réseaux véhiculaires, le canal de contrôle joue un rôle important, un grand nombre de données relatives à la sécurité sont

transmises du RSU aux utilisateurs du canal de contrôle. Si le canal de commande subit des attaques de brouillage persistantes, les véhicules victimes risquent de ne pas recevoir les messages liés à la sécurité du RSU, ce qui peut entraîner une perte économique considérable et entraîner des pertes en vies humaines.

Certaines caractéristiques saillantes des réseaux véhiculaires font qu'il n'est pas trivial de traiter ces problèmes avec les systèmes existants. Spécifiquement, pour les attaques Sybil, les noeuds malveillants peuvent facilement obtenir plusieurs identités légales (par exemple, pseudonyme), les méthodes de prévention basées sur une certification approuvée ne fonctionneraient pas dans ce cas; En utilisant des méthodes basées sur le test des ressources, un inconvénient majeur est que les identités Sybil ne peuvent pas être distinguées des utilisateurs normaux; Avec le système de réputation, les noeuds malveillants peuvent encore avoir le temps d'obtenir une bonne réputation car il leur suffit de lancer l'attaque pendant les heures de pointe afin de maximiser les profits générés en obstruant certains véhicules de leurs chemins. Pendant ce temps, la technique de vérification de position pourrait être une solution prometteuse pour détecter et localiser les noeuds Sybil, mais ce type de méthodes est coûteux quand la densité de trafic est élevée, et la motivation des utilisateurs est également discutable.

Pour les attaques jamming de radiofréquence, premièrement, selon le protocole IEEE 802.11p, un seul canal de contrôle est disponible pour la transmission de messages liés à la sécurité dans les réseaux véhiculaires. Il est donc impossible pour le RSU de basculer sur d'autres canaux si le canal de contrôle est bloqué. Deuxièmement, bien que les techniques multi-antennes aient été considérées comme une solution efficace, les antennes spécifiques n'ont pas été largement utilisées sur les véhicules en raison de la mobilité et de la contrainte de puissance des véhicules. Dernier point mais non le moindre, la mobilité des véhicules est limitée par l'espace routier et la densité de la circulation. Il est donc difficile pour les véhicules de s'échapper de la zone de brouillage si l'attaque est lancée par un véhicule voisin. Par conséquent, il reste difficile de résoudre ces problèmes dans les réseaux véhiculaires. Plus de détails sur les travaux connexes se trouve dans le Chapitre 2.

Résumé et Organisation

Motivés par les défis signalés précédemment, nous menons une recherche systématique dans cette thèse sur la détection et la prévention des mauvaises conduites dans les réseaux véhiculaires en nous concentrant sur plusieurs problèmes de recherche représentatifs d'importance fondamentale et pratique. Dans cette thèse, nous abordons plus particulièrement deux problèmes de sécurité majeurs: l'attaque Sybil et les attaques jamming de radiofréquence (RF).

Dans cette thèse, nous adoptons une ligne de recherche et d'exposition allant de la modélisation et l'analyse théoriques à la conception et l'optimisation d'algorithmes pratiques. Fig 1 illustre la structure de notre thèse. Dans la suite de cette section, nous fournissons un aperçu de haut niveau des contributions techniques de notre thèse, qui sont présentées de manière séquentielle au Chapter 2-5.

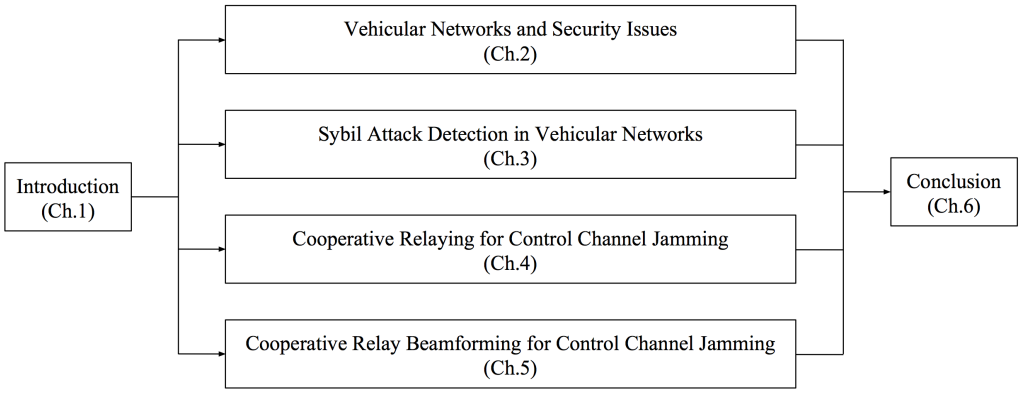


Figure 1: Organisation de la thèse

Réseaux Véhiculaires et Problèmes de Sécurité

Réseaux Véhiculaires

Un réseau de véhicules est un terme générique pour l'application intégrée des technologies de communication, de contrôle et de traitement de l'information au système de transport. Il couvre tous les modes de transport et prend en compte tous les éléments du système de transport : le véhicule, l'infrastructure et le conducteur ou l'utilisateur, qui interagissent de manière dynamique. Sa fonction générale est d'améliorer la prise de décision, souvent en temps réel, par les contrôleurs de réseau de transport et des autres utilisateurs, améliorant ainsi le fonctionnement de l'ensemble du système de transport.

Architecture des Réseaux Véhiculaires

La Figure 2 représente une architecture fondamentale de l'ITS, qui inclut les véhicules, les Road-side Units (RSU) et les réseaux du noyau. Ces systèmes devront soutenir les communications véhicule à infrastructure (V2I) et véhicule à véhicule (V2V).

Standards de Réseaux Véhiculaires

Standards DSRC/WAVE L'architecture de communication IEEE WAVE repose exclusivement sur le standard IEEE 802.11p. Dans la littérature, souvent DSRC (Dedicated Short Range Com-

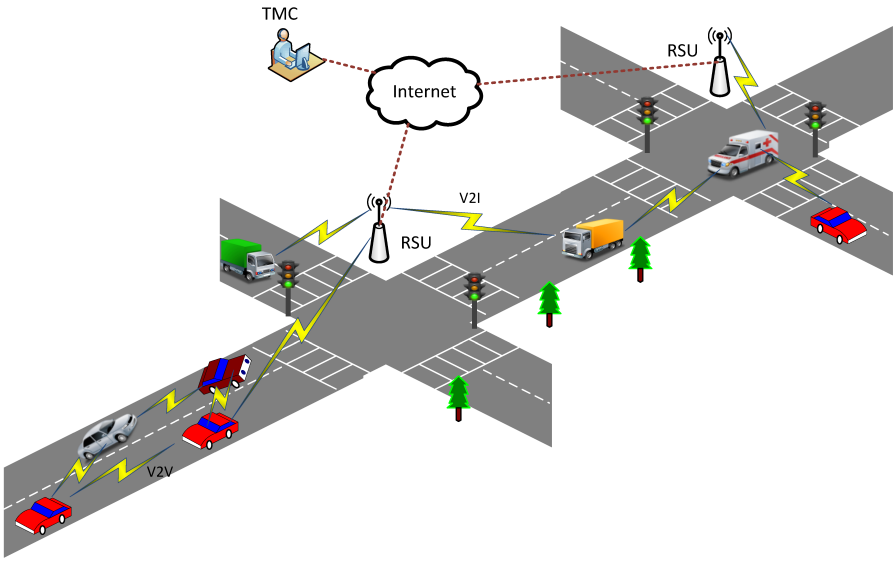


Figure 2: L'architecture des réseaux véhiculaires.

munications), WAVE (Wireless Access in Vehicular Environments) ou même IEEE 802.11p sont utilisés pour désigner l'ensemble de la pile de protocoles de Standards traitant des VANETs.

Standards ETSI L'ensemble de standards ETSI TC ITS est organisé en cinq groupes de travail. Le groupe de travail 1 définit l'ensemble de base des exigences de l'application et les services, le groupe de travail 2 fournit la spécification d'architecture et résout les problèmes multicouches, le groupe de travail 3 développe les protocoles de réseau et de transport pour ITS G5, le groupe de travail 4 analyse les médias du standard IEEE 802.11p et le groupe de travail 5 travaille sur le système de sécurité.

Exigence de Sécurité

L'exigence de sécurité dans les réseaux véhiculaires peuvent être décrites de manière générale en utilisant les propriétés suivantes :

- **Authenticité**: les entités et les messages doivent être authentifiés.
- **Intégrité**: garantit qu'un message n'a pas été modifié entre le moment où il a été envoyé et celui où il a été reçu, car le message reçu doit correspondre au message envoyé.
- **Non-répudiation**: garantit que, après l'authentification et le contrôle d'intégrité, l'entité ne peut pas nier avoir participé à un événement de communication.
- **Confidentialité**: veille à ce que les informations confidentielles et sensibles soient bien cryptées lors des procédures de communication.

- **Disponibilité**: le réseau et les applications doivent rester opérationnels même en présence de défauts ou de conditions malveillants.

- **Privé et Anonymity**: la plupart des problèmes de privé sont liés à la position et aux identifiants dans les réseaux véhiculaires.

Cependant, il est impossible de satisfaire toutes les exigences. Il existe toujours un compromis entre efficacité, sécurité et confidentialité du système lors des procédures de communication [4].

Système de Défense

Il est généralement admis dans ces deux ensembles de standards que les services de sécurité implémentés dans les réseaux véhiculaires reposent sur un système qui contient: des mécanismes de cryptographie, des mécanismes de génération de certificats et des mécanismes de gestion des certificats. En outre, certains mécanismes de protection de la confidentialité peuvent également être utilisés dans ce système, y compris anonymes et pseudonymes.

Analyse de Vulnérabilité

Ces services peuvent protéger la confidentialité des stations ITS, l'authenticité et l'intégrité des messages dans les environnements de communication véhiculaire. Cependant, il existe également plusieurs problèmes de sécurité dans les réseaux véhiculaires. Dans notre travail, nous nous concentrons sur deux problèmes majeurs: le problème de protection de la confidentialité peut mener aux attaques Sybil et le problème de disponibilité rend les réseaux véhiculaires vulnérables aux attaques jamming.

Attaques Sybil

L'attaque de Sybil a été mentionnée pour la première fois en 2002. Dans [5] Douceur a mentionné que, dans un système peer-to-peer sans autorité de confiance logiquement centrale pour garantir une correspondance biunivoque entre l'entité et l'identité, il est possible pour une entité inconnue de présenter plus de une identité. Dans les réseaux véhiculaires, les attaques Sybil peuvent endommager la couche de réseau et la couche d'application.

Comme la CSMA / CA est implémentée dans la couche de réseau, la coopération entre les nœuds virtuels offre la possibilité d'utiliser plus de ressources de canal que les autres nœuds légitimes. Dans la couche d'pplication, les nœuds virtuels participent également à la communication avec d'autres stations ITS. Dans ce cas, lorsqu'un nœud malveillant utilise plusieurs pseudonymes en même temps, les nœuds virtuels, générés à partir de l'utilisation de pseudonymes,

peuvent aider à accroître l'influence des faux messages de sécurité en les diffusant vers d'autres nœuds légitimes.

Les méthodes de détection des attaques Sybil peuvent être divisées en méthodes basées sur des clés sécurisées, en méthodes basées sur le test des ressources, en méthodes basées sur la réputation et en méthodes basées sur la position. Les méthodes existantes de détection des attaques Sybil sont brièvement présentées dans le Tableau 1.

Attaques Jamming de Radiofréquence

En raison de la nature ouverte et partagée du médium sans fil, les réseaux de véhicules sont vulnérables aux attaques jamming de radiofréquence puisqu'un attaquant peut facilement émettre un signal d'interférence pour empêcher un accès légitime au médium ou perturber la réception du signal. Un attaquant peut utiliser différentes stratégies de brouillage en exploitant les vulnérabilités des protocoles de couches PHY et MAC, qui appartiennent généralement à deux catégories: les stratégies de brouillage monocanal et les stratégies de brouillage multicanal.

De nombreuses contre-mesures ont été proposées pour résoudre les problèmes de brouillage de radiofréquence sous différents angles techniques. Par exemple, dans la couche physique, les techniques de saut de fréquence, telles que Frequency-Hopping Spread Spectrum (FHSS) [15], Direct Sequence Spread Spectrum (DSSS) [15], et Hybrid FHSS/DSSS [15] ont été largement adoptés pour éviter les interférences de brouillage en commutant rapidement entre les fréquences ou en répartissant le signal dans une bande beaucoup plus large afin de permettre une plus grande résistance aux interférences involontaires et intentionnelles. Cependant, le principal inconvénient de ces solutions est qu'elles nécessitent une bande passante beaucoup plus large que le signal original. Certaines techniques innovantes, telles que les technologies Ultra Wideband (UWB), la polarisation d'antenne et les techniques multi-antennes [16, 17]sont également proposées pour résoudre le problème de brouillage. En particulier, la technique multi-antenne est une solution anti-jamming prometteuse, car elle permet d'éviter les interférences provenant de sources indésirables. Il peut également améliorer l'efficacité et la fiabilité de la transmission en exploitant le gain de diversité et le gain d'antenne [18]. Certaines contre-mesures tentent également de fournir une protection proactive / réactive contre les attaques lors de la conception des protocoles de couche MAC (par exemple, Carrier-Sense Multiple Access with Collision Avoidance (CSMA / CA), Time Division Multiple Access (TDMA), etc.), ou d'utiliser une stratégie de channel-hopping pour résoudre le problème de brouillage multicanal en tenant compte du comportement contradictoire du brouilleur [19, 20]. Certaines solutions ont été proposées pour les attaques par brouillage dans les réseaux multi-sauts, qui tentent de rediriger le trafic autour de la zone brouillée, ou se retirent

Table 1: Méthodes de Détection des Attaques Sybil

Méthodes	Références	Advantage	Inconvénient
Clés sécurisées	[6, 7, 8]	Pouvoir empêcher la génération de fausses identités	La génération et la gestion d'un grand nombre de clés sont coûteuses et il est difficile de mettre en place une autorité centrale à laquelle tous les participants peuvent faire confiance
Test des ressources	[6]	Pas d'overhead supplémentaire	Les identités Sybil peuvent être difficilement distinguées et localisées
Système de réputation	[9]	Faible overhead et faible taux d'erreur	Les nœuds malveillants peuvent toujours obtenir une bonne réputation si la période d'attaque n'est que d'un faible pourcentage de temps
Vérification de position	[10, 11, 12, 13, 14]	Taux de détection élevé	L'estimation de la localisation basée sur RSSI n'est pas assez précise

du brouilleur en tirant parti de la mobilité des nœuds [21, 22, 23, 24]. Les schémas anti-jamming existants pour les réseaux sans fil sont brièvement présentés dans le Tableau 2.

Table 2: Schémas Anti-jamming pour les Réseaux Sans Fil

Méthodes	Références	Advantage	Inconvénient
Spread Spectrum (SS)	[15, 25, 26, 27]	Résiste aux interférences de signaux de brouillage	Non disponible pour un canal à fréquence fixe
Techniques multi-antennes	[28, 29, 30, 31, 32]	Pouvoir éviter le signal d'interférence des sources indésirables	Les multi-antennes ne sont pas largement utilisées sur les véhicules
Système coopératif	[33, 34, 35, 36, 37, 38]	Fournir une diversité spatiale et d'antenne et pouvoir améliorer la robustesse des communications	La conception du système coopératif est difficile

Détection d'Attaque Sybil dans les Réseaux Véhiculaires

Considérant que les nœud virtuels doivent éviter les positions capturées par les véhicules légitimes, leurs formes de conduite deviennent erratiques, en particulier dans les environnements de trafic dynamiques. Dans ce chapitre, nous examinons la possibilité de détecter les attaques Sybil en fonction de la variation de leurs habitudes de conduite en utilisant une méthode du machine learning. En règle générale, le machine learning est un processus dans lequel un ensemble de paramètres de seuil est formé pour classifier un comportement inconnu [39]. Dans cette thèse, trois méthodes principales sont considérées dans notre travail: Distance based clustering, Support Vector Machine (SVM) et k-Nearest Neighbours (kNN).

Modèle d'Attaque

Sur la base de l'analyse des méthodes de défense et des messages, nous pouvons confirmer qu'il est possible qu'un véhicule utilise simultanément plusieurs pseudonymes pour lancer des attaques Sybil.

Description de la Forme de Conduite du Véhicule

Dans notre algorithme, la Driving Pattern Matrix (DPM) doit d'abord être construite. La forme de conduite d?un véhicule à un moment donné est décrite en utilisant un vecteur de cinq éléments. Par exemple, $\vec{V_n} = (x_{n1}, x_{n2}, x_{n3}, x_{n4}, x_{n5})$ represents the driving pattern of a vehicle at time t_1.

Distance Based Clustering

La procédure de détection peut être décrite comme suit:

- Nous notons que $\lambda_{i1}^{ma} = max\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\}$ et $\lambda_{i2}^{ma} = max\{\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\} - \{\lambda_{i1}^{ma}\}\}$ les deux plus grandes valeurs propres de chaque DPM.

- Pour les n véhicules, une matrice S de n lignes et 2 colonnes est définie comme: $S = (\lambda_{ij}^{ma})_{1 \le i \le n, 1 \le j \le 2}$.

- Sélectionner la valeur médiane de chaque colonne, obtenir un vecteur $\overrightarrow{med} = (x, y)$ (la valeur médiane est choisie au lieu de la valeur moyenne en raison de l'instabilité des valeurs propres des nœuds Sybil.).

- Calculer la distance de Mahalanobis entre chaque vecteur ligne de la matrice S et le vecteur $\overrightarrow{med}$.

- Critère de décision: Vecteur $\vec{r_i}$ est la i-ème ligne de la matrice S, qui représente les formes de conduite du i-ème véhicule. Si la distance de Mahalanobis entre $\vec{r_i}$ et $\overrightarrow{med}$ est supérieure à une valeur de seuil sélectionnée α, où $d(\vec{r_i}, \overrightarrow{med}) \ge \alpha$, le véhicule v_i est considéré comme malveillant.

Résultats Expérimentaux

Résultats de Mesure de Similarité Plusieurs résultats de mesure de similarité des différentes formes de conduite sont détaillés dans la Fig. 3 sous différentes densités de véhicules. Les deux plus grandes valeurs propres sont choisies pour représenter principalement la forme de conduite d'un véhicule.

Taux de Détection Comme le montre la Fig. 4, nous avons choisi la courbe ROC (Receiver Operating Characteristic) pour représenter le taux de détection de notre méthode, qui est considérée comme un moyen complet et visuellement attrayant pour résumer la précision de la détection. On peut observer que notre méthode de détection peut atteindre un TPR élevé avec un faible FPR. Dans tous les cas, plus de 90% des nœuds Sybil peuvent être détectés par notre méthode de

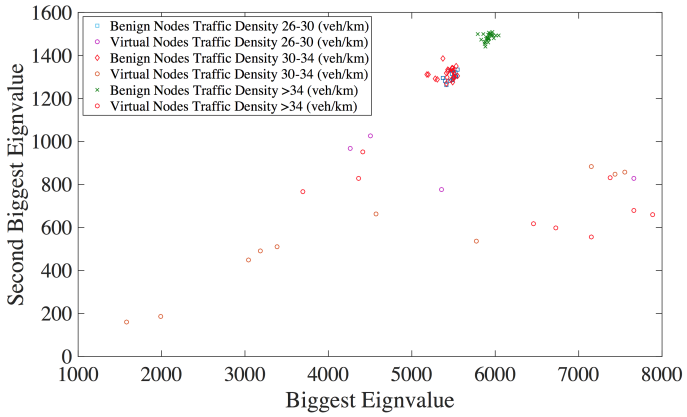


Figure 3: Les deux premières grandes valeurs propres de DPM peuvent être utilisés pour représenter principalement la forme de conduite d'un véhicule

détection. Autrement, le taux de détection a une corrélation positive avec la densité de trafic, ce qui correspond bien à notre hypothèse. Parce que la densité de trafic élevée peut limiter la distance moyenne entre les véhicules, ce qui rend leur comportement de conduite plus similaire. D'autre part, la distance de sécurité limitée limite également les choix du nœud malveillant lorsqu'il génère des nœuds virtuels.

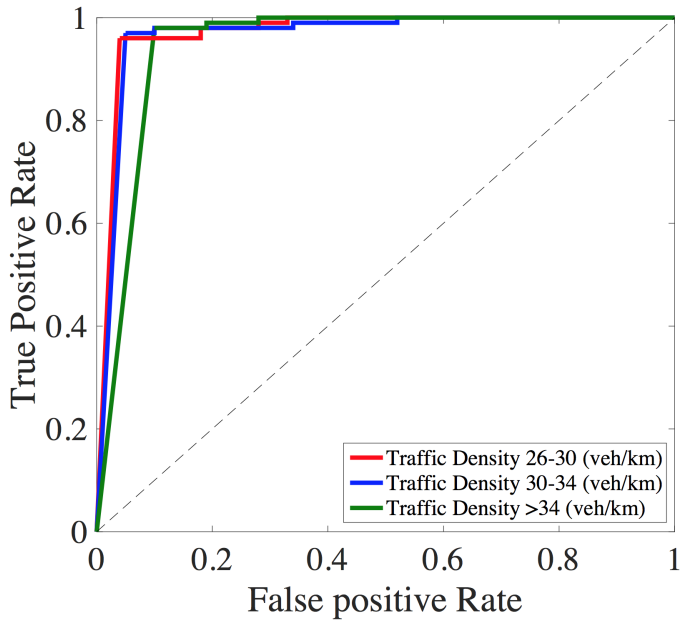


Figure 4: Taux de détection des nœuds Sybil sous différentes densités de trafic

Détection d'attaque Sybil Basée sur Support Vector Machine (SVM)

Dans cette section, nous présentons une méthode qui représente principalement les comprtements de conduite des véhicules en utilisant les valeurs propres de leur matrice de conduite, ainsi que la procédure de classification basée sur plusieurs classificateurs SVM. Une brève description de

l'algorithme est illustrée à la Fig. 5.

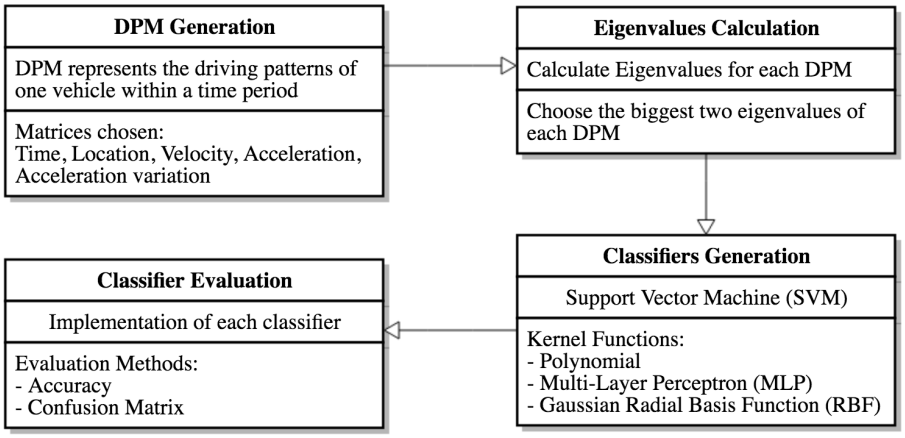


Figure 5: Schéma de l'algorithme

Classification

Nous notons que $\mathcal{V} = \{\vec{v}_i, u_i\}_{1 \leq i \leq n}$ un ensemble avec n véhicules où le vecteur $\vec{v}_i = (\lambda_{i1}^{ma}, \lambda_{i2}^{ma})$ représente les formes de conduite du véhicule v_i, et $u_i \in \{-1, 1\}$ donne l'étiquette de $\vec{v}_i$. L'étiquette -1 signifie que $\vec{v}_i$ est malveillant et sinon $\vec{v}_i$ est légitime.

Normalement, si les données sont linéaires, un hyperplan séparateur peut être utilisé pour diviser les données. Cependant, en raison du caractère erratique du mouvement des nœuds virtuels, il est fréquent que les données soient loin d'être linéaires et que les jeux de données ne soient pas séparables linéairement. Dans ce cas, le jeu de données $\mathcal{V}$ doit être projeté dans l'espace d'entités où le nouvel jeu de données $\mathcal{V}'$ est séparable linéairement:

$$\begin{aligned} &\psi : \mathcal{V} \longrightarrow \mathcal{V}' \\ &\vec{v} \longrightarrow \psi(\vec{v}) = \begin{pmatrix} \psi_1(\vec{v}) \\ \psi_2(\vec{v}) \end{pmatrix} \end{aligned} \tag{1}$$

Ensuite, nous avons $\mathcal{V}' = \{(\psi(\vec{v}), u_i)\}_{1 \leq i \leq n}$ avec $u_i \in \{-1, 1\}$.

Nous définissons ensuite une fonction du noyau k:

$$k(\vec{v}_i, \vec{v}_j) = < \psi(\vec{v}_i), \psi(\vec{v}_j) > \tag{2}$$

Où $< \cdot, \cdot >$ st le produit scalaire entre deux vecteurs.

Dans ce travail, trois fonctions du noyau sont prises en compte: Polynomial, Gaussian Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP).

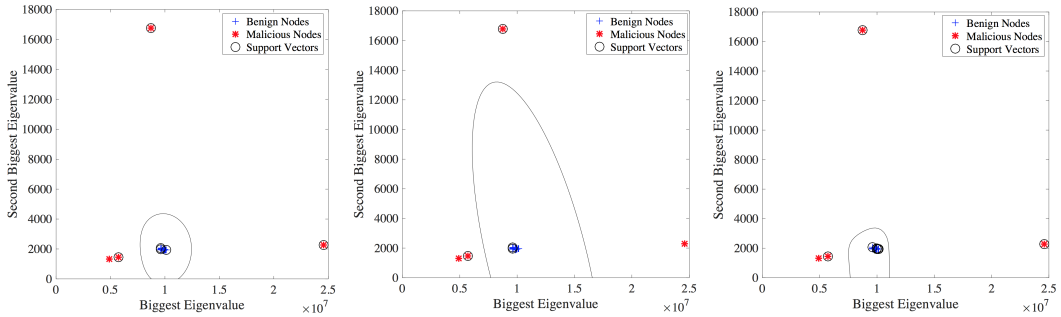
- Polynomial: $k(\vec{v}_i, \vec{v}_j) = (< \vec{v}_i, \vec{v}_j > + h)^d$ où h est une valeur constante.

- RBF: $k(\vec{v}_i, \vec{v}_j) = exp(-\frac{||\vec{v}_i - \vec{v}_j||^2}{2\sigma^2})$
- MLP: $k(\vec{v}_i, \vec{v}_j) = tanh(\rho < \vec{v}_i, \vec{v}_j > +\varrho)$

Leur performance sera évaluée dans la section suivante en fonction des résultats de la simulation.

Résultats Expérimentaux

Caractéristiques des Nœuds Virtuels Comme illustré sur la Fig. 6, les formes de conduite des noeuds légitimes ont une similarité évidente et celles des noeuds Sybil montrent un caractère erratique.



(a) RBF Sigma=0.5 (b) Polynomial d=2 (c) MLP

Figure 6: Résultats de la classification des groupes d'apprentissage avec les trois fonctions du noyau

Précision de la Classification Nous avons lancé 15 attaques Sybil sous 3 densités de trafic différentes. Le premier groupe est choisi comme groupe d'apprentissage et les 14 autres groupes sont des groupes d'essai. Comme illustré dans le Tableau 3, la précision de la classification dans tous les groupes d'essai et la Fig. 6 montrent les résultats de la classification des groupes d'apprentissage. Dans ce travail, les trois fonctions du noyau sont implémentées avec des paramètres différents. De manière générale, en raison du fait que les nœuds légitimes présentent une forte similitude dans leurs conportements de conduite, les classificateurs qui couvrent moins de surface atteignent des précisions plus élevées.

Taux d'Erreur de Classification Comme le montre le Tableau 4, le résumé des matrices de confusion de ces trois classificateurs. D'une manière générale, ces classificateurs ont tous atteint un taux de détection élevé avec un taux d'erreur faible. Quand on rentre plus en détail, on remarque que le FNR du classifieur polynomial est d'environ 8%, mais que le FPR est faible. Comme nous l'avons vu plus haut sur la Fig. 6, les classificateurs polynôme couvert plus de surface que les deux

Table 3: Précision de la Classification des Groupes d'Essai

Densité du Trafic	26-30	26-30	26-30	26-30	30-34	30-34	30-34	30-34	30-34	> 34	> 34	> 34	> 34	> 34
rbf σ = 0.1	0.917	0.956	0.945	1	0.909	0.941	0.961	0.933	0.963	0.961	0.912	1	0.965	1
rbf σ = 0.3	0.917	1	0.945	1	1	0.971	1	1	0.926	1	0.912	1	1	1
rbf σ = 0.5	0.958	0.870	0.946	1	1	0.971	1	1	0.889	1	0.882	1	1	1
rbf σ = 0.7	0.958	0.870	0.919	1	1	0.971	1	1	0.889	0.962	0.882	1	1	1
rbf σ = 0.9	0.958	0.782	0.919	1	1	0.971	1	0.933	0.889	0.884	0.882	0.971	1	0.972
Polynomial d = 2	0.958	0.783	0.919	1	1	0.912	0.923	0.933	0.889	0.846	0.882	0.971	1	0.972
Polynomial d = 4	0.958	0.782	0.892	0.971	0.939	0.853	0.923	0.933	0.889	0.846	0.882	0.943	0.966	0.944
mlp [-1 1]	0.958	0.826	0.973	1	1	0.971	1	1	0.889	1	0.882	1	1	1
mlp [-2 2]	0.958	0.826	1	1	1	0.971	1	1	0.889	1	0.882	1	1	1

autres classificateurs. Dans ce cas, les bénins légitimes peuvent difficilement être signalés comme malveillants, mais les nœuds malveillants pourraient être déclarés comme légitimes.

Table 4: Matrice de Confusion

	TP	TN	FP	FN	TPR	FPR	FNR
rbf σ = 0.3	325	94	2	8	97.6%	2%	2.4%
Poly d = 2	326	74	1	28	92.1%	1.3%	7.9 %
MLP [-2 2]	326	90	1	12	96.5%	1.1%	3.6%

Détection d'attaque Sybil Basée sur k-Nearest Neighbours (kNN)

Dans cette section, nous présentons une méthode qui représente principalement les formes de conduite des véhicules en utilisant les valeurs propres de leur matrice de conduite, ainsi que la procédure de classification basée sur les classificateurs k-Nearest Neighbours classifiers.

Classification

Nous notons que les deux vecteurs $\vec{v_i}$, $\vec{v_j}$ où $\vec{v_i} = (\lambda_{i1}, \lambda_{i2})$, $\vec{v_j} = (\lambda_{j,1}, \lambda_{j,2})$ représentent les formes de conduite des deux véhicules v_i and v_j. La différence entre leurs formes de conduite est définie par la distance de Minkowski entre ces deux vecteurs:

$$d(\vec{v_i}, \vec{v_j}) = [\sum_{s=1}^{2} |\vec{v}_{is} - \vec{v}_{js}|^q]^{\frac{1}{q}} = [|\lambda_{i1} - \lambda_{j1}|^q + |\lambda_{i2} - \lambda_{j2}|^q]^{\frac{1}{q}} \quad (3)$$

La distance de Minkowski est généralement utilisée avec q étant égal à 1 ou 2, connus sous le nom de distance de Manhattan et distance Euclidienne.

Pour classer un véhicule arrivant v_e, sous notons un groupe de formation avec n véhicules comme $K = \{(\vec{v_i}, y_i)i = 1, 2, ..., n\}$ où $\vec{v_i} = (\lambda_{i1}, \lambda_{i2})$ and $y_i \in \{-1, 1\}$ donne l'étiquette de $\vec{v_i}$. L'étiquette -1 signifie que $\vec{v_i}$ est malveillant et $\vec{v_i}$ est légitime.

La distance entre le véhicule arrivant v_e et son voisin le plus proche peut être représentée par $argmin\{d(\vec{v_e}, \vec{v_i}), 1 \leq i \leq n\}$ Et l'étiquette de v_e dépend des étiquettes de ses k voisins les plus proches ($k = 1, 3, 5, ...$).

Optimisation

Dans cet article, nous proposons deux méthodes pour limiter la complexité temporelle de l'algorithme kNN. L'une est méthode Sans Mémoire: le système n'ajoute pas l'objet classé dans l'ensemble d'apprentissage, ainsi, la complexité du temps n'augmente pas. L'autre méthode est basée sur la similarité des formes de conduite des utilisateurs légitimes. Un vecteur $\vec{v_m} = (\lambda_{m1}, \lambda_{m2})$ peut être utilisé pour représenter tous les n véhicules légitimes dans le groupe d'apprentissage, où $\lambda_{m1} = \sum_{i=1}^{n} \lambda_{i1}$, $\lambda_{m2} = \sum_{i=1}^{n} \lambda_{i2}$. Une fois le $(n + 1)$ème véhicule classé comme légitime,

After the $(n+1)$th vehicle is classified as benign, le vecteur $\vec{v_m}$ peut être mis à jour en calculant les valeurs moyennes arithmétiques pondérées des λ_{m1} et λ_{m2}, où $\lambda'_{m1} = (n\lambda_{m1} + \lambda_{(n+1)1})/(n + 1)$ et $\lambda'_{m2} = (n\lambda_{m2} + \lambda_{(n+1)2})/(n + 1)$. Sinon, ce véhicule sera directement ajouté à l'ensemble d'apprentissage.

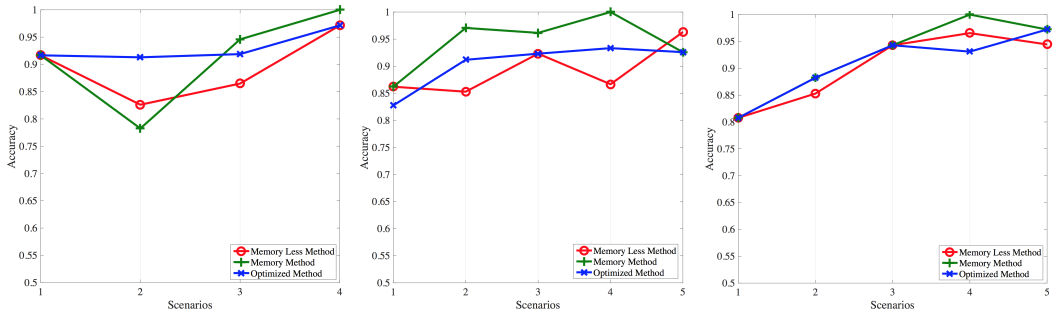
L'efficacité de ces deux méthodes sera évaluée dans la section suivante en fonction des résultats de la simulation.

Résultats Expérimentaux

Précision de la Classification La Fig. 7 montre les résultats de l'évaluation de la précision de la classification.

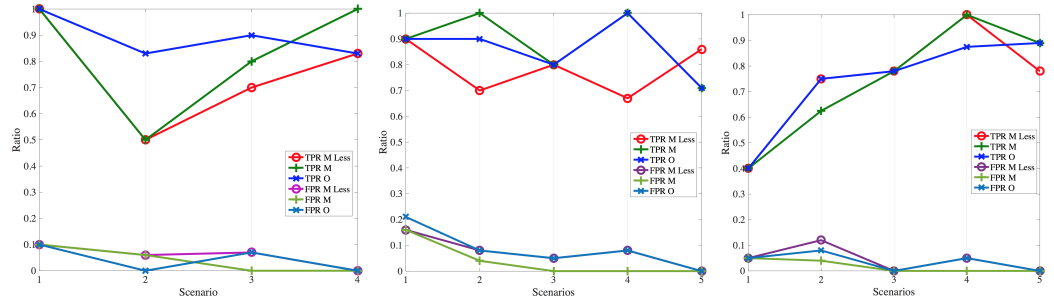
De manière générale, ces trois méthodes peuvent atteindre une précision de classification satisfaisante, respectivement 80%. De manière plus détaillée, les performances de la méthode Sans Mémoire ne sont pas stables par rapport aux deux autres méthodes. La méthode Mémoire présente la meilleure précision de classification. Dans plusieurs scénarios, elle a atteint une précision de classification de 100%. Et pour notre méthode optimisée, ses performances sont plus stables que la méthode Sans Mémoire. Son point faible pourrait être le contrôle des erreurs, qui sera évalué dans la sous-section suivante.

Taux d'Erreur de Classification Comme le montre la Fig. 8, on peut constater que les performances de détection de la méthode Sans Mémoire ne sont pas stables, car les décisions sont



(a) Densité 26-30 véhicules/km (b) Densité 30-34 véhicules/km (c) Densité >34 véhicules/km

Figure 7: Précisions de la classification des groupes d'essai en trois densités de trafic différentes



(a) Densité 26-30 véhicules/km (b) Densité 30-34 véhicules/km (c) Densité >34 véhicules/km

Figure 8: Taux d'erreur de classification des groupes d'essai évalués à l'aide de TPR et de FPR

prises uniquement en fonction du nombre limité de données d'apprentissage. On peut également observer que la méthode Mémoire et notre méthode optimisée fonctionnent bien en détection ainsi qu'en contrôle des erreurs. En particulier les performances de contrôle des erreurs de la méthode Mémoire, qui peuvent atteindre un taux d'erreur de 0 dans certains scénarios. Pour notre méthode optimisée, le point faible potentiel pourrait être sa performance en matière de contrôle d'erreur, car dans cette méthode, un ensemble de données était remplacé par sa valeur moyenne. Cependant, sur la base des résultats de la simulation, nous pouvons constater que notre méthode optimisée a également donné de meilleurs résultats que les méthodes Sans Mémoire en contrôle d'erreur.

Relais Coopératif pour le Brouillage du Canal de Contrôle

Dans ce chapitre, nous nous concentrons sur la conception de contremesures pour le problème de brouillage du canal de contrôle dans les réseaux véhiculaires, ce qui est d'une importance vitale pour la sécurité des communications I2V. À cet effet, nous proposons d'adopter les techniques de relais coopératif pour résoudre le problème de brouillage du canal de contrôle dans les réseaux véhiculaires, qui repose sur l'idée que les véhicules situés à l'extérieur de la zone de brouillage peuvent servir de relais pour aider à transmettre le signal du canal de contrôle aux véhicules vic-

times par d'autres canaux de service sans brouillage. De cette manière, un schéma multi-antenne virtuel peut être formé de plusieurs noeuds de relais afin de desservir de manière coopérative les véhicules victimes. La fiabilité de la transmission peut ainsi être améliorée en exploitant la diversité spatiale de ces noeuds de relais. Nous analysons les performances de ce schéma de relais coopératif anti-brouillage dans différents scénarios de brouillage et concevons un algorithme de sélection de relais pour optimiser les performances des pires véhicules victimes. Les résultats de simulation sont fournies pour démontrer la performance du schéma proposé.

Modèles de Réseau

Étant donné l'emplacement $L(v_m)$ du brouilleur (soit stationnaire ou en mouvement),les plages de $\mathbb{V}$ et de $\mathbb{R}$ peuvent être déterminées comme suit. Que P_R et P_J se réfèrent respectivement à la puissance d'émission de RSU et du brouilleur, puis un véhicule v_x à l'emplacement x est supposé être victime si son SINR (signal-to-interference-plus-noise ratio) est inférieur au seuil prescrit η:

$$SINR = \frac{P_R x^{-\alpha}}{P_J ||x - m||^{-\alpha} + \sigma^2} < \eta, \tag{4}$$

où α est l'exposant de path loss, et σ^2 est la puissance de bruit. Notez que le bruit peut être ignoré par rapport au signal d'interférence puissant émis par le brouilleur. Ainsi, la plage victime pour $\mathbb{V}$ et la plage de relais pour $\mathbb{R}$ peuvent être obtenues comme suit:

$$\mathbb{V} = \{v_x | \frac{m}{1+a} < x < \frac{m}{1-a}\} \tag{5}$$

$$\mathbb{R} = \{v_x | -r \leq x \leq \frac{m}{1+a} \cup \frac{m}{1-a} \leq x \leq r\} \tag{6}$$

où $a = (\frac{\eta P_J}{P_R})^{1/\alpha}$.

Les emplacements des véhicules peuvent être modélisés comme un PPP (Poisson Point Process) homogène avec l'intensitéλ [40, 41]. Ainsi, le nombre de véhicules dans une plage donnée suit une distribution de Poisson avec le taux λ.

Définissons $\gamma_{ij} = d_{ij}^{\alpha}/P_v$ comme la path loss normalisée entre chaque paire de nœud de relais $v_i \in \mathbb{R}$ et de nœud victime $v_j \in \mathbb{V}$, qui est un PPP non homogène avec l'intensité $\lambda(x) = \lambda \frac{2}{\alpha} P_v^{2/\alpha} x^{2/\alpha - 1}$ selon la théorie mapping [42].

Analyse de Probabilité d'Interruption

Nous supposons que la stratégie Decode-and-Forward (DF) est adoptée par les nœuds de relais pour transmettre le message reçu aux véhicules victimes. Soient s et t représentent le signal émis

et le signal reçu respectivement. Ensuite, pour chaque véhicule victime $v_j \in \mathbb{V}$, le signal reçu d'un véhicule relais $v_i \in \mathbb{R}$ peut être donné par:

$$t_{ij} = \sqrt{P_v} h_{ij} \|d_{ij}\|^{-\alpha/2} s + n_0, \tag{7}$$

où P_v est la puissance de transmission du véhicule, $\|d_{ij}\|^{-\alpha/2}$ et h_{ij} sont la path loss et le fading multi-chemin entre les nœuds v_i et v_j. $n_0$est le bruit au noeud v_j, qui est supposé être une variable aléatoire complexe gaussienne de moyenne nulle avec une variance unitaire σ^2.

Les signaux reçus de tous les relais peuvent être décodés avec le schéma MRC (Maximum Ratio Combining) ou les schémas SC (Selection Combining), et les SNRs (Signal-to-Noise Ratios) obtenus au nœud v_j peuvent être donnés par

$$\text{MRC:} \quad \Gamma_j^m = \frac{\sum\limits_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}}{\sigma^2} \tag{8}$$

et

$$\text{SC:} \quad \Gamma_j^s = \frac{\max\limits_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}}{\sigma^2} \tag{9}$$

respectivement

En supposant que les canaux entre les relais et la victime suivent le modèle de fading de Rayleigh et soient indépendants entre les différents relais, puis $\sum_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}$ est distribué de manière exponentielle avec la moyenne $\sum_{v_i \in \mathbb{R}} P_v \|d_{ij}\|^{-\alpha}$ car h_{ij}s sont mutuellement indépendants, et puis nous avons $\sum_{v_i \in \mathbb{R}} P_v \|d_{ij}\|^{-\alpha} = 1/\sum\limits_{v_i \in \mathbb{R}} \gamma_{ij}^{-1}$, où $\gamma_{ij} = \|d_{ij}\|^{\alpha}/P_v$ est la path loss normalisée telle que définie dans la section précédente.

La performance de ce schéma de relais coopératif peut être caractérisée par la probabilité d'interruption P_o. Pour le schéma SC, la probabilité d'interruption d'un véhicule victime $v_j \in \mathbb{V}$ est donné par:

$$\begin{aligned} P_o^{SC} &= \mathbb{P}(\Gamma_j^s < \theta) \\ &= \int\limits_{\gamma_{ij}} \prod_{v_i \in \mathbb{R}} (1 - e^{-\theta \sigma^2 \gamma_{ij}}) f(\gamma_{ij}) d\gamma_{ij}, \end{aligned} \tag{10}$$

où $f(\gamma_{ij})$ est la distribution commune pour tous les γ_{ij}s, qui est donné par [43]:

$$f(\gamma_{ij}, \ldots, \gamma_{nj}) = e^{-\lambda P_v^{2/\alpha} \gamma_{nj}^{2/\alpha}} (\lambda \frac{2}{\alpha} P_v^{2/\alpha})^n \prod_{i=1}^{n} \gamma_{ij}^{2/\alpha - 1} \tag{11}$$

for $(v_1, \ldots, v_n) \in \mathbb{R}$.

De même, pour le schéma MRC, la probabilité d'interruption au véhicule victime v_j peut être

donnée par:

$$\begin{aligned} P_o^{MRC} &= \mathbb{P}(\Gamma_j^m < \theta) \\ &= \int_{\gamma_{ij}} (1 - \exp(-\frac{\theta\sigma^2}{\sum\limits_{v_i \in \mathbb{R}} \gamma_{ij}^{-1}})) f(\gamma_{ij}) d\gamma_{ij}. \end{aligned} \tag{12}$$

Définissons $Q(\gamma_{ij})$ comme suit:

$$Q(\gamma_{ij}) = \begin{cases} \prod\limits_{v_i \in \mathbb{R}} (1 - e^{-\theta\sigma^2 \gamma_{ij}}), & SC \\ 1 - \exp(-\frac{\theta\sigma^2}{\sum\limits_{v_i \in \mathbb{R}} \gamma_{ij}^{-1}}), & MRC, \end{cases} \tag{13}$$

Ensuite, les probabilités d'interruption atteintes par ces deux schémas peuvent être exprimées sous une forme unifiée:

$$P_o = \int_{\gamma_{ij}} Q(\gamma_{ij}) f(\gamma_{ij}) d\gamma_{ij}. \tag{14}$$

Problème de Sélection de Relais Anti-jamming

En pratique, la CSI (channel state information) entre les véhicules peut être obtenue par des mesures en ligne ou par un enregistrement de canal. Par conséquent, le gain de canal $H_{ij} = \|h_{ij}\|^2 \|d_{ij}\|^{-\alpha}$ entre le nœud relais v_i et le véhicule victime v_j peut être supposé être connu, Et donc le SNR réalisable sur un véhicule victime v_j avec un sous-ensemble sélectionné de nœuds relais $R_s \subseteq \mathbb{R}$ peut être donné par:

$$\text{SNR}_j = \sum_{v_i \in \mathbb{S}} \hat{P}_i H_{ij} \tag{15}$$

où $\hat{P}_i = P_i/\sigma^2$ est la puissance d'émission normalisée du nœud relais v_i.

Dans ces conditions, le principal défi consiste à sélectionner un sous-ensemble de nœuds relais $\mathbb{S}$ dans l'ensemble de relais candidat $\mathbb{R}$, de sorte que le SNR du pire véhicule victime est maximisé, ce qui peut être formellement défini comme un problème MMRS (Max-Min Relay Selection) comme suit:

$$\begin{aligned} \max \quad & \min_{v_j \in \mathbb{V}} \sum_{v_i \in \mathbb{R}} \hat{P}_i H_{ij} \\ s.t. \quad & |R_s| \leq K, \end{aligned} \tag{16}$$

où la contrainte spécifie qu'au plus K nœuds relais peuvent être sélectionnés, ce qui est basé sur l'hypothèse que les transmissions de tous les relais sélectionnés sont planifiées avec la méthode TDMA (time division multiple access).

Le problème MMRS peut être formulé comme un problème de linear integer programming (LIP) comme suit:

$$\begin{aligned} \max \quad & \min_{v_j \in \mathbb{V}} \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \leq K, \qquad \forall v_j \in \mathbb{V}, \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{17}$$

où x_i est une variable d'indicateur pour indiquer si le noeud v_i est sélectionné comme un relais ou non.

En introduisant un seuil auxiliaire θ, (17) peut être transformé en la forme équivalente suivante:

$$\begin{aligned} \max \quad & \theta, \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \geq \theta, \forall v_j \in \mathbb{V}, \\ & \sum_{v_i \in \mathbb{R}} x_i \leq K, \forall v_i \in \mathbb{R}. \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{18}$$

Pour un seuil donné $\hat{\theta}$, (18) dégénère en un problème de vérification de faisabilité, c'est-à-dire de vérifier si un sous-ensemble de K relais peut être trouvé de sorte que les SNR de tous les véhicules victimes soient au-dessus du seuil θ. Cette vérification de faisabilité peut être réalisée en résolvant un problème MRS (minimum relay selection) comme suit:

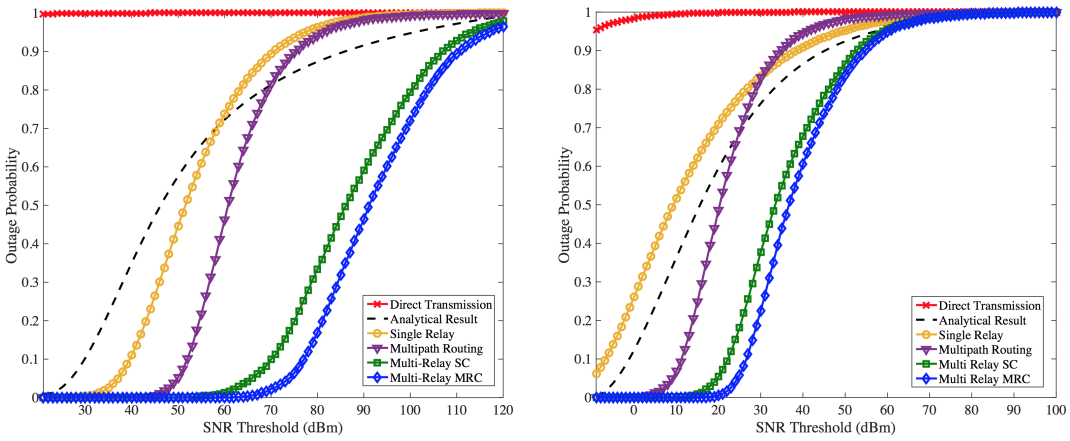
$$\begin{aligned} \min \quad & \sum_{v_i \in \mathbb{R}} x_i, \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \geq \hat{\theta}, \forall v_j \in \mathbb{V}, \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{19}$$

qui tente de trouver le nombre minimum de relais lorsque les SNRs de toutes les victimes sont satisfaits. Si le nombre de relais obtenu à partir de ce problème MRS est supérieur à K, puis (18) est infaisable pour le donnée $\hat{\theta}$, ce qui suggère que $\hat{\theta}$ est trop grand, le seuil doit donc être diminué, sinon nous pouvons augmenter θ pour obtenir un meilleur SNR pour toutes les victimes.

Évaluation de la Performance

Évaluation de l'Instantané

Dans la Fig. 9, nous montrons la fonction de distribution cumulative (cumulative distribution function, CDF) des probabilités d'interruption obtenues par les schémas de relais coopératifs proposés



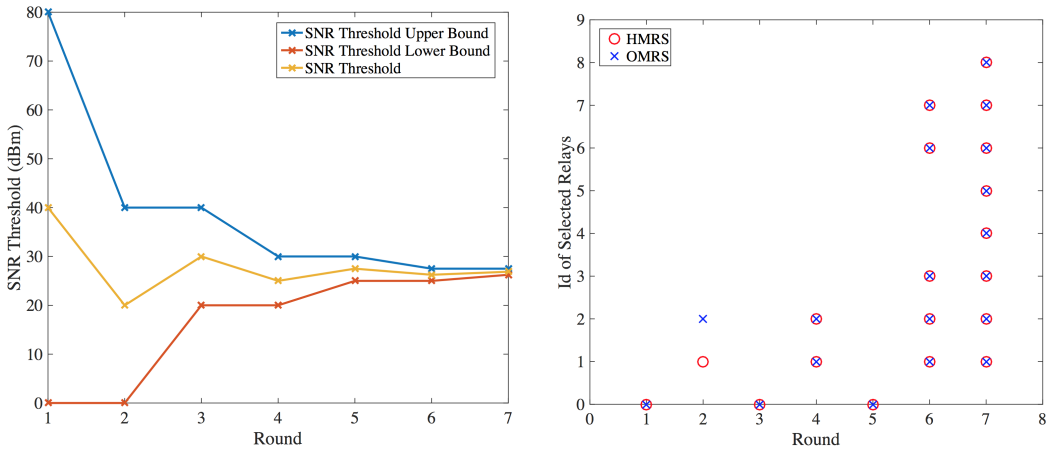
(a) Exponent de path loss $\alpha = 2$ (b) Exponent de path loss $\alpha = 4$

Figure 9: Distributions de probabilité d'interruption pour différents seuils de SNR

sous différents seuils de SNR, qui sont basés sur la moyenne de 100 instantanés des emplacements des véhicules et du brouilleur. On peut constater qu'en utilisant le schéma DT, la probabilité de panne dépasse toujours 95%, c'est-à-dire que plus de 95% des signaux envoyés par le RSU ne peuvent pas être décodés par les véhicules victimes, ce qui confirme l'impact significatif des attaques de jamming. Dans le cas d'un scénario à relais unique, nous supposons qu'un relais unique est distribué uniformément dans la plage de $[-r, \frac{m}{1+a}]$. Il est considéré comme une valeur moyenne parmi tous les scénarios de relais unique, et le résultat de la simulation correspondent bien avec le résultat d'analyse. Le protocole de routage à trajets multiples est conçu sur la base du schéma proposé dans [34], où plusieurs routes de bout en bout sont sélectionnés en fonction de leur histoire de la disponibilité et l'un d'entre elles est choisie pour livrer des paquets à un moment donné. Ainsi, le PDR de chaque lien dans ce chemin peut être mis à jour. On peut constater qu'après l'implémentation du protocole de routage par trajets multiples, la probabilité d'interruption est considérablement réduite et est inférieure à la valeur moyenne de tous les scénarios de relais unique, ce qui suggère qu'en tirant parti du gain de diversité proposé par les nœuds voisins, le protocole de trajets multiples peut prendre de bonnes décisions de routage, ce qui contribue à améliorer les performances anti-jamming. Toutefois, dans le cas où la décision de routage est prise sur la base de l'historique de disponibilité du chemin et que l'état du canal de chaque lien n'est pas surveillé en temps réel. Par conséquent, le chemin sélectionné par ce protocole de routage de couche MAC n'est pas toujours le plus approprié. En utilisant le schéma SC, le nœud de relais avec le meilleur canal vers les véhicules victimes est sélectionné, de sorte que les performances sont considérablement améliorées par rapport aux scénarios à relais unique. Enfin, on peut constater que la méthode MRC surpasse tous les autres schémas dans toutes les conditions, car la contribution de tous les nœuds relais est pleinement exploitée.

Évaluation des Schémas de Sélection de Relais

Dans la Fig. 10(a), nous montrons les mises à jour de θ, θ_L et θ_U au cours de la procédure de bissection de l'Algorithme 1. On peut observer que le θ optimal peut être trouvé après 7 tours, ce qui est beaucoup plus rapide qu'une méthode de recherche linéaire naïve, et les résultats sont les mêmes pour les schémas HMRS et OMRS. Pour vérifier les détails de ces deux schémas, nous montrons les relais sélectionnés par ces deux schémas à chaque itération dans la Fig. 10(b)(Notez que le problème de MRS est infaisable aux 1er, 3e et 5e Itérations, donc le R_s retourné est vide). On peut voir que les relais sélectionnés par ces schémas peuvent être différents.



(a) Convergence de HMRS et OMRS

(b) Relais sélectionnés par HMRS et OMRS

Figure 10: Comparaison des HMRS et OMRS (Contrainte de numéro de relais $K = 8$).

Dans la Fig. 11, nous montrons la probabilité d'interruption obtenue par ces algorithmes de sélection de relais. On peut constater qu'au fur et à mesure que le nombre de nœuds de relais augmente, la probabilité de panne diminue progressivement et le schéma HMRS surpasse celui des deux autres schémas dans toutes les conditions. En particulier, l'écart de performance est le plus important pour les scénarios à relais unique. Ainsi, lorsque le nombre de nœuds de relais dépasse un certain nombre, l'amélioration des performances devient marginale, ce qui est partiellement dû à la topologie particulière des réseaux véhiculaires, c'est-à-dire que plus les nœuds de relais sont éloignés des nœuds victimes, moins leur contribution faite au SNR combiné. Par conséquent, les nœuds de relais doivent être choisis judicieusement en fonction du SNR requis. Nous pouvons voir que lorsque le nombre de relais sélectionnés est supérieur à 5, les performances du schéma ORS sont très proches de celles du schéma HMRS car les relais sélectionnés sont fondamentalement similaires dans les deux méthodes. Plus de détails sont montrés dans la Fig 12 pour les probabilités d'interruption atteintes par ces systèmes avec un seuil de SNR de 15 dBm.

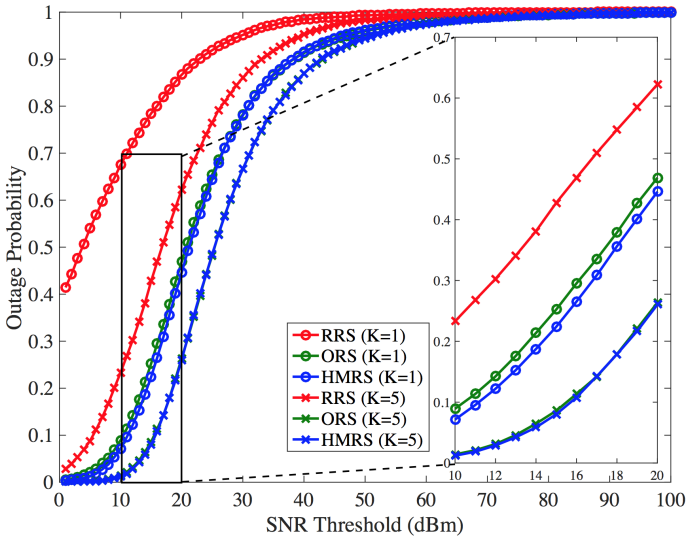


Figure 11: Distributions de probabilité d'interruption pour différents seuils de SNR

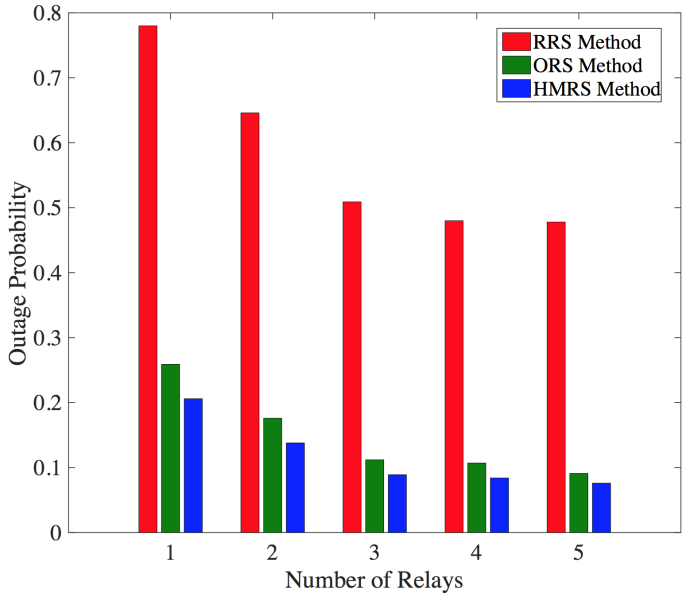


Figure 12: Distributions de probabilité d'interruption pour différents nombres de nœuds de relais (θ = 15).

Beamforming Coopératif de anti-jamming

Dans ce chapitre, nous examinons le scénario dans lequel le RSU est équipé de plusieurs antennes pouvant servir simultanément à plusieurs groupes de véhicules à l'aide de la technique de beamforming multi-groupe et multi-cast [44]. Cependant, en raison des interférences provoquées par le brouilleur, tous les véhicules ne peuvent pas décoder les signaux souhaités du RSU. En guise de solution, nous proposons un schéma anti-brouillage en deux étapes. Les véhicules qui ont décodé avec succès le signal reçu au premier étage seront sélectionnés en tant que relais pour desservir les véhicules victimes en coopération dans le deuxième étage en utilisant les techniques de beam-

forming coordonné sur un canal de service sans brouillage. En tirant parti du gain multi-antenne fourni par le RSU et de la diversité spatiale fournie par les véhicules relais, la fiabilité de transmission de tous les véhicules peut être considérablement améliorée sous la menace d'attaques de jamming.

Modèle de Système

Modèle de Réseau

Dans cet article, nous nous concentrons sur un segment de la route couvert par un RSU spécifique, qui est équipé de L antennes (Fig. 13). Un brouilleur à une seule antenne et plusieurs véhicules légitimes à une seule antenne sont situés dans la zone de transmission du RSU. Sans perte de généralité, le RSU est supposé être situé à l'origine avec une plage de transmission de r mètres, et donc le brouilleur $v_m$et les véhicules sont répartis dans la plage de $[-r, r]$.

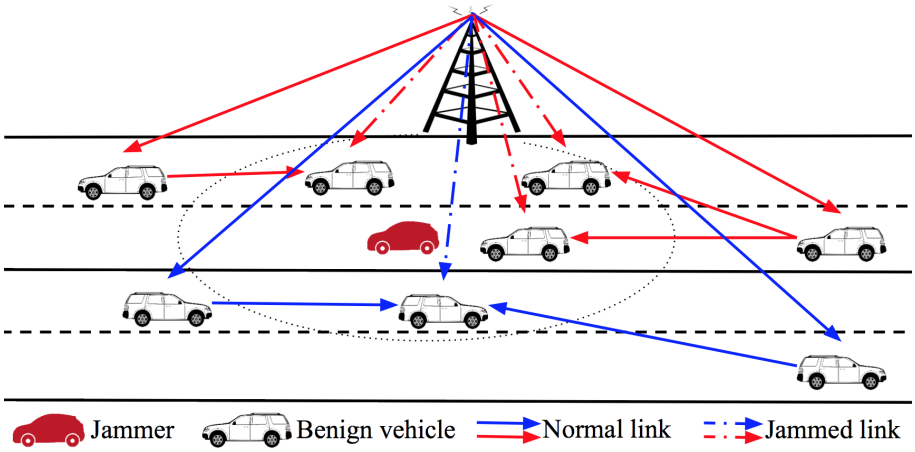


Figure 13: Réseau véhiculaire dans la zone de couverture d'un RSU à antennes multiples

Modèle de Signal

Selon la procédure de relayage précitée, le système anti-jamming peut être divisé en deux étapes. Dans la première étape, le RSU transmet simultanément les signaux de contrôle de tous les groupes en utilisant la technique de beamforming multi-groupe et multi-cast [44]. Puis le signal reçu au véhicule $v_{i,k}$ (denoted as vehicle k in the i-th group) peut être donné comme:

$$y_{i,k} = \mathbf{u}_i \mathbf{h}_k^H s_i + \sum_{p \neq i}^{n} \mathbf{u}_p \mathbf{h}_k^H s_p + \sqrt{P_m} J_k s_m + n_k, \tag{20}$$

où le premier terme est le signal désiré par tous les véhicules du groupe i. Le deuxième terme correspond aux signaux d'interference provenant d'autres groupes. Le troisième terme est l'interférence provoquée par le brouilleur, où s_m et P_m sont respectivement le signal émis par le brouilleur et sa

puissance d'émission, et J_k est le gain de canal entre le brouilleur et le véhicule k. $n_k$il est le bruit additif blanc gaussien au véhicule k, où $n_k \sim \mathcal{CN}(0, \sigma_k^2)$.

En supposant que les signaux de contrôle, le signal de brouillage et le bruit soient statistiquement indépendants, le SINR (signal-to-interference-plus-noise ratio) sur le véhicule $v_{i,k}$ peut être obtenu à partir de (5.1) comme:

$$\mathrm{SINR}_{i,k} = \frac{\mid \mathbf{u}_i \mathbf{h}_k^H \mid^2}{\sum_{p \neq i}^{n} \mid \mathbf{u}_p \mathbf{h}_k^H \mid^2 + P_m \mid J_k \mid^2 + \sigma_k^2}. \tag{21}$$

Le débit de données réalisable $r_{i,k}$ au véhicule $v_{i,k}$ peut être donné par:

$$r_{i,k} = B \log(1 + \mathrm{SINR}_{i,k}), \tag{22}$$

où B est la largeur de bande de canal en hertz.

Lors de la première étape, certains véhicules peuvent ne pas décoder le signal reçu du RSU, qui est déclaré victime. Ensuite, lors de la deuxième étape, l'ensemble des véhicules qui ont décodé les signaux avec succès sera sélectionné comme relais pour retransmettre les signaux aux victimes du même groupe par le biais d'un canal de service sans brouillage en utilisant la stratégie DF (decode-and-forward). Puis le signal reçu sur un véhicule victime $v_{i,k}$ peut être représenté comme:

$$\tilde{y}_{i,k} = \mathbf{w}_i \mathbf{g}_k^H s_i + \sum_{p \neq i}^{n} \mathbf{w}_p \mathbf{g}_k^H s_p + n_k. \tag{23}$$

Semblable à (5.2), le SINR sur un véhicule victime $v_{i,k}$ peut être donné par:

$$\widetilde{\mathrm{SINR}}_{i,k} = \frac{\mid \mathbf{w}_i \mathbf{g}_k^H \mid^2}{\sum_{p \neq i}^{n} \mid \mathbf{w}_p \mathbf{g}_k^H \mid^2 + \sigma_k^2}, \tag{24}$$

et le débit de données réalisable $\tilde{r}_{i,k}$ sur un véhicule victime$v_{i,k}$ peut être donné comme:

$$\tilde{r}_{i,k} = B \log(1 + \widetilde{\mathrm{SINR}}_{i,k}). \tag{25}$$

De même, $\tilde{r}_{i,k}$ doit être supérieur au seuil γ pour que le véhicule victime puisse décoder le signal reçu des relais avec succès.

Problème de Beamforming d'Anti-jamming Cooperative

Du point de vue d'anti-jamming, il est souhaitable de maximiser le débit de données minimum de tous les véhicules sous l'interférence de brouilleurs dans le canal de contrôle, qui peut être formellement déclaré en tant que problème de conception de beamforming :

$$\max_{\{\mathbf{u}_i\},\{\mathbf{w}_i\}} \quad \min_{i\in\{1\dots n\}} \{\min\{\min_{k\in\mathcal{R}_i}\{r_{i,k}\}, \min_{k\in\mathcal{V}_i}\{\tilde{r}_{i,k}\}\}\} \tag{26}$$

$$s.t. \quad \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \le P_0, \quad \forall i, \tag{26a}$$

$$\|\mathbf{w}_{ik}\|_2^2 \le P_{i,k}, \qquad \forall i, \forall k \in \mathcal{R}_i, \tag{26b}$$

$$\mathcal{R}_i \cap \mathcal{V}_i = \emptyset, \qquad \forall i, \tag{26c}$$

où (26a) et (26b) correspondent aux contraintes de puissance du RSU et des relais, respectivement. P_0 et $P_{i,k}$ représentent les bilans de puissance d'émission du RSU et du véhicule $v_{i,k}$. (26c) précise que chaque véhicule ne peut pas être un relais ou une victime en même temps

Ainsi, le problème (26) peut être reformulé comme suit:

$$\max \quad \gamma \tag{27}$$

$$s.t. \quad \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \le P_0, \qquad \forall i, \tag{27a}$$

$$\|\mathbf{w}_{ik}\|_2^2 \le x_{i,k} P_{i,k}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{27b}$$

$$\gamma - r_{i,k} \le M_1(1 - x_{i,k}), \quad \forall i, \forall k \in \mathcal{N}_i, \tag{27c:1}$$

$$\gamma - \tilde{r}_{i,k} \le M_2 x_{i,k}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{27c:2}$$

$$x_{i,k} \in \{0, 1\}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{27d}$$

qui est un problème MINLP (mixed-integer non-linear programming), which is a mixed-integer non-linear programming (MINLP) problem, et la solution optimale est difficile à obtenir en temps polynomial. Dans ce qui suit, nous proposerons quelques schémas d'approximation et de relaxation pour transformer ce problème en une forme manipulable et ainsi le résoudre efficacement.

Approximation du problème et Relaxation

Approximation de ℓ_0-Norme Lissée

Dans (27), $x_{i,k}$ indique l'identité du véhicule $v_{i,k}$ et spécifie que la contrainte (27c:1) devrait s'appliquer si le véhicule $v_{i,k}$ est sélectionné comme relais (i.e., $x_{i,k} = 1$); sinon (i.e., $\neg x_{i,k}$), (27c:2) devrait être satisfait puisqu'il s'agit d'une victime.

En solution, on peut approcher la ℓ_0-norme discontinue de $\|\mathbf{w}_{i,k}\|_2^2$ avec une fonction continue, lisse et concave comme suit [45]:

$$x_{i,k} \approx f_\theta(\|\mathbf{w}_{ik}\|_2^2) = 1 - \exp(-\frac{\|\mathbf{w}_{ik}\|_2^2}{\theta}), \tag{28}$$

où $\theta > 0$ est un paramètre pour contrôler la lisse de l'approximation. En général, un θ plus petit conduit à un meilleur résultat d'approximation et un θ plus grand conduit normalement à une approximation plus lisse. Plus de détails sur le réglage de θ peuvent être trouvés dans [46].

Détente Semi-définitive

Pour remédier à la non-concavité de $r_{i,k}$ et de $\tilde{r}_{i,k}$, nous adoptons la méthode de relaxation semi-définie (semi-definite relaxation, SDR) proposée dans [47].

De cette façon $r_{i,k}$ peut être représenté comme:

$$\begin{aligned} r_{i,k} =& B\log(1+\frac{\mathrm{tr}(\mathbf{U}_i\mathbf{h}_k\mathbf{h}_k^H)}{\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{U}_p\mathbf{h}_k\mathbf{h}_k^H)+P_m\mid J_k\mid^2+\sigma_k^2}) \\ =& B\log(\sum_{i=1}^{n}\mathrm{tr}(\mathbf{U}_i\mathbf{h}_k\mathbf{h}_k^H)+P_m\mid J_k\mid^2+\sigma_k^2)-B\log(\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{U}_p\mathbf{h}_k\mathbf{h}_k^H)+P_m\mid J_k\mid^2+\sigma_k^2) \\ =& y_k(\mathbf{U}_i)-z_k(\mathbf{U}_p), \end{aligned} \tag{29}$$

où $y_k(\cdot)$ et $z_k(\cdot)$ sont concaves par rapport aux matrices $\mathbf{U}_i$ et $\mathbf{U}_p$.

De même, $\tilde{r}_{i,k}$ peut être représenté comme:

$$\begin{aligned} \tilde{r}_{i,k} =& B\log(1+\frac{\mathrm{tr}(\mathbf{W}_i\mathbf{g}_k\mathbf{g}_k^H)}{\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{W}_p\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2}) \\ =& B\log(\sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_i\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2)-B\log(\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{W}_p\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2) \\ =& \tilde{y}_k(\mathbf{W}_i)-\tilde{z}_k(\mathbf{W}_p), \end{aligned} \tag{30}$$

où $\tilde{y}_k(\cdot)$ et $\tilde{z}_k(\cdot)$ sont concaves par rapport aux matrices $\mathbf{W}_i$ et $\mathbf{W}_p$.

Procédure Convexe-concave

Ce type de problème DC peut être résolu en utilisant la procédure convexe-concave (convex-concave procedure, CCP) [48]. L'idée de base est d'écrire chaque fonction non linéaire comme la somme d'une fonction convexe et concave, puis de convexifier le problème en remplaçant la partie concave par leurs extensions de Taylor du premier ordre, puis de résoudre successivement une séquence de sous-problèmes convexes. Plus précisément, la CCP commence par un point initialx_0, et à chaque itération t résout un sous-problème convexe:

$$\begin{aligned} \max \quad & \gamma \\ s.t. \quad & \gamma-[z_i(x^{(t)})+\triangledown z_i(x^{(t)})^T(x-x^{(t)})]+y_i(x)\leq 0, \quad \forall i, \end{aligned} \tag{31}$$

où $x^{(t)}$ la solution optimale obtenue à l'itération précédente.

De cette façon, le problème (27) peut être transformé en une séquence de sous-problèmes convexes comme suit:

$$\max \quad \gamma \tag{32}$$

$$s.t. \quad \mathrm{tr}(\mathbf{U}_i) - P_0 \leq 0, \quad \forall i, \tag{32a}$$

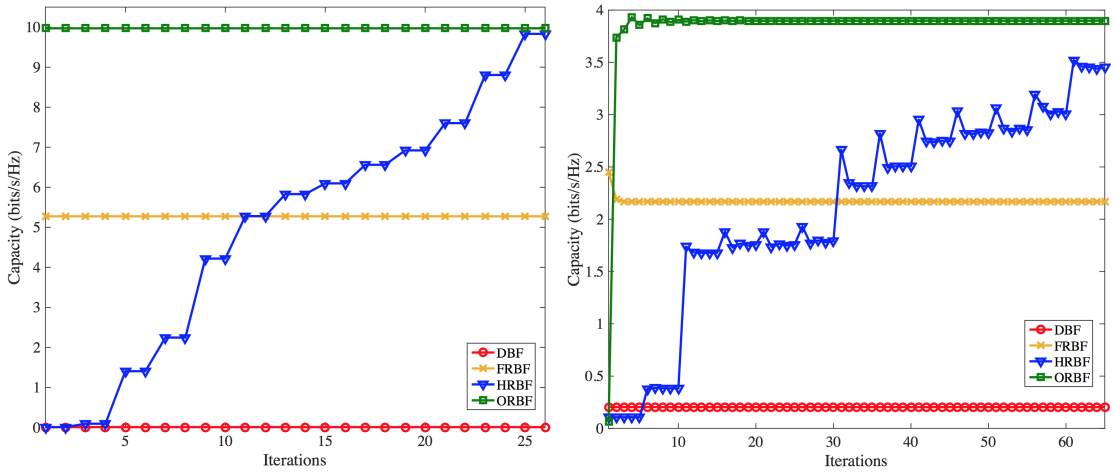
$$\mathbf{W}_{i,k} - (1 - \exp(-\frac{\mathbf{W}_{i,k}}{\theta}))P_{i,k} \leq 0, \qquad \forall i, \forall k \in \mathcal{N}_i \tag{32b}$$

$$\gamma - h_k(\mathbf{U}_i) + Z_k(\mathbf{U}_p|\tilde{\mathbf{U}}_p) - M_1 D_\theta(\mathbf{W}_{i,k}|\tilde{\mathbf{W}}_{i,k}) \leq 0, \quad \forall i, p, \forall k \in \mathcal{N}_i, \tag{32c:1}$$

$$\gamma - h_k(\mathbf{W}_i) + \tilde{Z}_k(\mathbf{W}_p|\tilde{\mathbf{W}}_p)(1 - \exp(-\frac{\mathbf{W}_{i,k}}{\theta})) \leq 0, \quad \forall i, p, \forall k \in \mathcal{N}_i, \tag{32c:2}$$

$$\mathbf{U}_i \succeq 0 \quad and \quad \mathbf{W}_i \geq 0, \quad \forall i. \tag{32d}$$

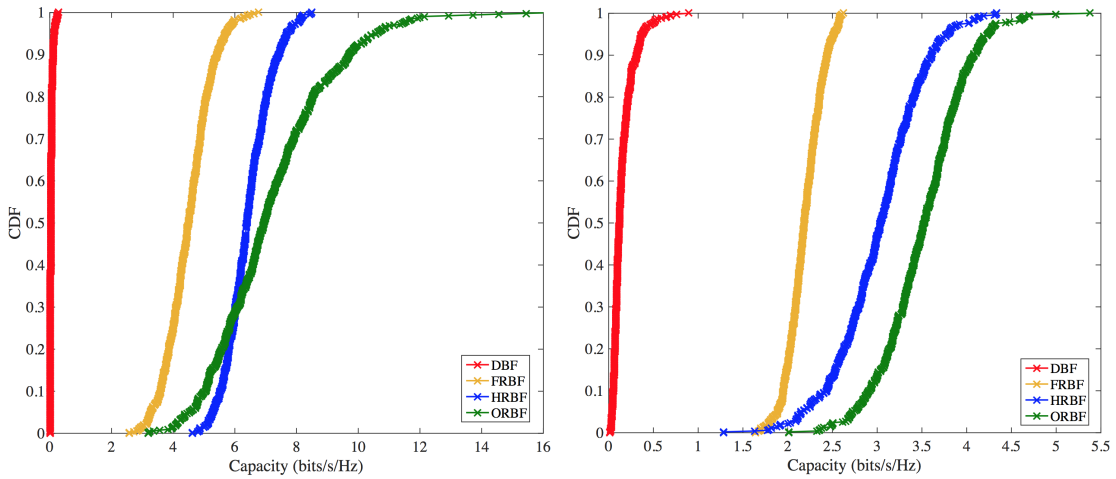
Résultats de simulation



(a) Scénario à groupe unique ($n = 1, N = 15$) (b) Scénario à groupes multiples ($n = 2, N_1 = N_2 = 8$)

Figure 14: Convergence de quatre algorithmes différents dans différents scénarios

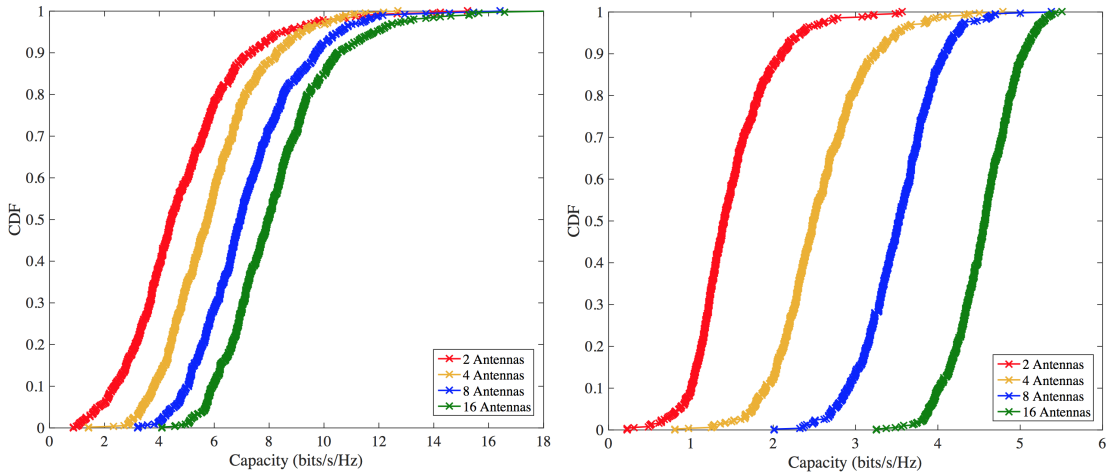
Dans la Fig. 14, Nous démontrons le comportement de convergence de ces quatre schémas différents dans les scénarios à groupe unique et à groupe multiple. On peut constater que, dans les deux scénarios, à l'exception du schéma HRBF, tous les autres schémas convergent rapidement vers le point stationnaire. De plus, la capacité du réseau obtenu est nettement améliorée en utilisant le schéma ORBF. Le schéma HRBF a également de bonnes performances par rapport aux schéma de DBF et FRBF. Cependant, la détermination du jeu de relais approprié prend plus de temps. On peut constater que le relais de bottleneck d'étranglement est placé de manière séquentielle dans l'ensemble des victimes, ce qui permet d'augmenter progressivement la capacité du réseau au fur et à mesure que le nombre de nœuds de relais diminue (13 véhicules sur 15 dans le scénario à groupe unique et 12 sur 16 dans le scénario à groupes multiples).



(a) Scénario à groupe unique (b) Scénario à groupes multiples $n = 2$

Figure 15: CDF de la capacité réseau atteinte sous différents statuts

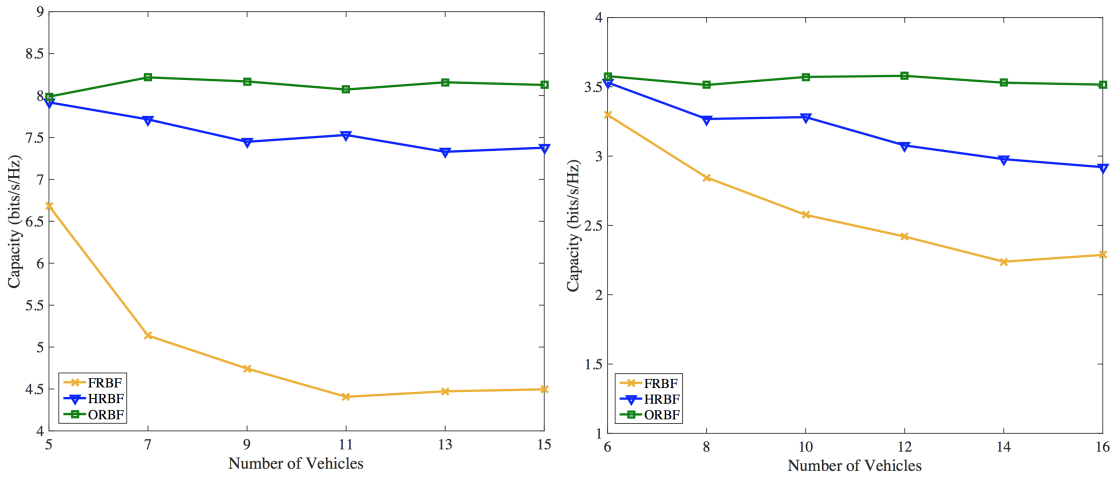
Dans la Fig. 15, nous montrons la fonction de distribution cumulative (cumulative distribution function, CDF) de la capacité du réseau réalisé par ces quatre régimes dans différentes conditions de canal, qui sont obtenus sur la base de 500 scénarios différents. On peut constater que, dans les deux scénarios, le schéma ORBF proposé surpasse les trois autres, et que le schéma HRBF peut atteindre des performances sous-optimales par rapport au schéma ORBF, en particulier dans le scénario à groupe unique. En particulier, on peut constater que, selon le schéma DBF, la capacité du réseau est inférieure à 1 bit/s/Hz, ce qui démontre l'impact significatif de l'attaque de jamming dans le canal de contrôle.



(a) Scénario à groupe unique (b) Scénario à groupes multiples $n = 2$

Figure 16: CDF de la capacité de réseau réalisée par le schéma ORBF avec différents nombres d'antennes

Dans la Fig. 16, Nous traçons la CDF de la capacité de réseau réalisée par le système ORBF avec les différents nombres d'antennes, chaque courbe est obtenue sur la base de 500 scénarios différents. On peut clairement constater que dans les scénarios à groupe unique et à groupe multiples, la capacité de réseau obtenue augmente avec le nombre d'antennes, en particulier dans les scénarios à groupes multiples, ce qui est raisonnable, car plus le nombre d'antennes est élevé, plus la diversité peut être exploitée comme une contre-mesure des attaques de jamming.



(a) Scénario à groupe unique

(b) Scénario à groupes multiples $n = 2$

Figure 17: Capacité moyenne du réseau par rapport au nombre de véhicules

Enfin, nous étudions la capacité du réseau avec un nombre différent de véhicules. Nous considérons trois schémas différents. Le nombre d'antennes est fixé à $L = 8$, et le nombre de véhicules varie de 5 à 15 dans les scénarios à groupe unique, de 6 à 16 dans les scénarios à groupes multiples. Dans la Fig. 17, chaque point est la moyenne des résultats de 100 simulations. On peut constater que la capacité du réseau réalisée par l'ORBF reste stable dans différentes conditions. Cependant, à mesure que le nombre de véhicules augmente, la capacité de réseau moyenne atteinte par les schémas FRBF et HRBF est utilisée progressivement. La raison en est que lorsque le nombre de véhicules est faible, il est possible pour les schémas FRBF et HRBF de sélectionner les mêmes relais que pour le schéma ORBF. Cependant, ces deux schémas ne parviennent pas à trouver les relais optimaux avec l'augmentation du nombre de véhicules, les écarts de performances sont donc considérablement augmentés.

Conclusions et Travaux Futurs

Résumé de Thèse

Dans cette thèse, nous nous sommes concentrés sur deux problèmes majeurs de la couche PHY et de la couche d'application: les attaques jamming de radiofréquence et les attaques Sybil.

En particulier, nous avons adopté trois méthodes différentes de machine learning pour la détection des nœuds Sybil: Distance based clustering, Support Vector Machine (SVM) et k-nearest neighbours (kNN). Basé sur la base de la variation entre les véhicules légitimes et les nœuds Sybil dans leurs formes de conduite, les nœuds virtuels inexistants peuvent être détectés.

Pour les attaques jamming de radiofréquence, nous nous sommes concentrés sur la conception de contremesures pour le problème de brouillage du canal de contrôle dans les réseaux véhiculaires, ce qui est d'une importance vitale pour la sécurité des communications I2V. Ainsi, nous avons étendu les problèmes de jamming dans les scénarios RSU multi-antennes, dans lesquels la RSU peut desservir plusieurs groupes de véhicules simultanément à l'aide de la technique de beamforming multi-groupe et multi-cast . En guise de solution, nous proposons un système anti-jamming en deux étapes. Les véhicules qui ont décodé avec succès le signal reçu dans la première étape seront sélectionnés en tant que relais pour desservir en coopération les véhicules victimes dans la deuxième étape en utilisant les techniques coordonnées de beamforming sur une canal de service sans blocage.

Travaux Futurs

Dans cette thèse, nous nous sommes concentrés sur l'anti-jamming des transmissions downlink (de la RSU aux véhicules), il serait donc intéressant de considérer les transmissions uplink où la RSU est la cible des attaques de jamming. En même temps, sous examinerons la conception anti-jamming de beamforming sous l'hypothèse d'informations de station de canal partielles, ce qui peut réduire considérablement l'overhead pour la mesure du canal ou pour le précodage de beamforming.

Chapter 1

Introduction

1.1 Background and Motivations

Vehicle communications are becoming increasingly popular, propelled by navigation safety requirements and by the investments of car manufacturers and Public Transport Authorities. Vehicular networking has significant potential to enable diverse application associated with traffic safety, traffic efficiency and infotainment.

These networks are attracting considerable attention from the research community as well as the automotive industry. High interest for these networks is also shown from governmental authorities and standardization organizations. To establish a large-scale cooperative network, several standardization efforts has been provided focusing on vehicular communication architecture and protocol standards and can be geographically grouped. In USA, the Dedicated Short-Range Communications (DSRC) standards suite [1] is based on multiple cooperating standards mainly developed by the IEEE. In Europe, the ETSI TC ITS working group defines standards for Intelligent Transport Systems (ITS) systems [2]. Moreover, the Car-to-Car Communication Consortium (C2C-CC) [49] has been initiated in Europe by car manufacturers and automotive OEMs (original equipment manufacturers), with the main objective of increasing road traffic safety and efficiency by means of inter-vehicle communication. From the government side, the project SCOOP@F [50] is announced by the French Minister of Transport in the year of 2014, in order to improving road safety, optimizing traffic management and efficiency, optimizing infrastructure management costs and developing new services and making vehicles fit for the future.

Nowadays, vehicular networks are promising in a number of useful driver and passenger oriented services, which include Internet connections facility exploiting an available infrastructure in an “on-demand” fashion, electronic tolling system, and a variety of multimedia services. It covers all modes of transport and considers all elements of the transportation system: the vehicle,

the infrastructure, and the driver or user, interacting together dynamically. The overall function of vehicular networks is to improve decision making, often in real time, by transport network controllers and other users, thereby improving the operation of the entire transport system.

Beyond all their benefits, the vehicular networks raise new challenges regarding the security and the privacy protection. As mentioned previously, vehicular networks is expected to propose four major types of services: safety messages exchange, internet connection, location based services and user privacy protection. Safety in these networks is crucial because that it affects the life of humans. A statistic that was made by the European Commission shows that in 2014, more than 25 000 people died on the roads of the European Union, and for every death on Europe's roads there are an estimated 4 permanently disabling injuries such as damage to the brain or spinal cord, 8 serious injuries and 50 minor injuries.

It is essential like that the vital information cannot be modified or deleted by an attacker and must be also determine the responsibility of drivers while maintaining their privacy. In vehicular networks, the safety and security of the network architecture and protocols are of vital importance, which have been the central theme of the standardization in both USA and Europe. In general, the provided security services are based on three major mechanisms: Encryption algorithms, Public Key Infrastructure (PKI) and Pseudonymous. These security services provide basic protection for the privacy of users and integrity of messages in vehicular networks. However, there remains several critical issues in vehicular networks, which are listed as follows:

- **Identity management issue**: The privacy protection especially the using of pseudonyms cause the vulnerability to the Sybil attack. After the request of pseudonyms, vehicles would use the pseudonyms as communication identities. Each pseudonym valid has its own key pair of signature. Therefore the integrity and non-repudiation of information can be ensured. In this case, while one vehicle using several pseudonyms together at the same time, each pseudonym is an individual vehicle in the view of other users.

- **Availability issue**: Availability is always vulnerable in networks, especially in wireless environments. In vehicular networks, the control channel plays an important role, a large number of safety-related are transmitted from the RSU to users in the control channel. If the control channel is under persistent jamming attacks, the victim vehicles may fail to receive the safety related messages from the RSU, which can possibly cause tremendous economic loss and claim human lives.

Some salient features of the vehicular networks making it nontrivial to address these issues with existing schemes. Specifically, for Sybil attacks, the proposed detection methods can be divided into secure key based methods [6, 7, 8], resource testing based methods [6], reputation based

methods [9] and position based methods [10, 11, 12, 13, 14]. However, in vehicular networks, malicious nodes can easily get several legal identities (e.g., Pseudonymous), then the trusted certification based prevention methods would not work in this case; Using resource testing based methods, one major drawback is the Sybil identities cannot be distinguished from normal users; With the reputation system, malicious nodes can still have time to get good reputations since they only need to launch the attack during rush hour in order to maximize the profit that is earned by obstructing some vehicles from their paths. Meanwhile, the position verification technique could be a promising solution to detect and locate the Sybil nodes, but this type of methods are costly while the traffic density is high, and the motivation of users is also questionable.

For RF jamming attacks, the anti-jamming schemes proposed for the traditional wireless networks are designed majorly using spread spectrum (SS) communications [15, 25, 26], multi-antenna techniques [16, 31, 32] and cooperative scheme [34, 35, 36]. However, in vehicular networks, firstly, according to the IEEE 802.11p protocol, only one control channel is available for the transmissions of safety related messages in vehicular networks, and thus it is impossible for the RSU to switch to other channels if the control channel is jammed. Secondly, although multi-antenna techniques have been deemed as an effective anti-jamming solution, but the specific antennas have not been widely used on the vehicles due to the mobility and power constraint of vehicles. Last but not the least, the mobility of vehicles is limited by road space and road traffic density, and thus it is difficult for the vehicles to escape out of the jamming area if the attack is launched by a neighbouring vehicle. Therefore, it remains a challenging issue to address these issues in vehicular networks. More detail about the related work can be found in Chapter 2.

1.2 Thesis Overview and Organization

Motivated by the challenges pointed out previously, we conduct a systematic research in this thesis on misbehaviour detection and prevention in vehicular networks by focusing on several representative research problems of both fundamental and practical importance. Specifically, we address two major security issues in this thesis: Sybil attack and radio frequency (RF) jamming attacks. Ranging from theoretical modelling and analysis, to practical algorithm design and optimisation.

In our thesis, we adopt a research and exposition line from theoretical modelling and analysis to practical algorithm design and optimisation. Fig 1.1 illustrates the structure of our thesis. In the remainder of this section, we provide a high-level overview of the technical contributions of our thesis, which are presented sequentially in Chapter 2-5.

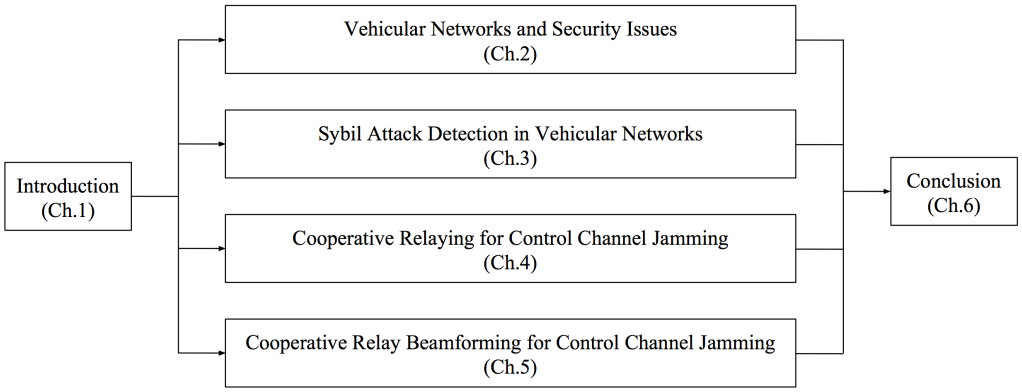


Figure 1.1: Thesis organization

Chapter 2: Vehicular Networks and Security Issues In this chapter, we briefly introduce the standard sets proposed by the IEEE and the ETSI, compare their differences on the architecture, applications and security services. Thus, based on the widely used defending systems, combined with the security requirements of vehicular networks, we confirm that the Sybil attack and the RF jamming attack are the two major threats to the vehicular networks. Afterwards, based on the schemes proposed in the related work, we discuss carefully the suitable solutions in vehicular communication environments.

Chapter 3: Sybil attack detection in vehicular networks In this chapter , we focus on the Sybil attack detection in vehicular networks based on the vehicle driving patterns. Relying on beacon information, we designed a data format Driving Pattern Matrix (DPM) to describe vehicle driving pattern within a time period. Thus, three different machine learning methods: Distance based clustering, Support Vector Machine (SVM) and k-nearest neighbours (kNN) are considered. The main idea is to evaluate the similarity of vehicle driving patterns, then based on the variation of vehicles' driving pattern to distinguish the malicious nodes from the benign ones.

Chapter 4: Cooperative relaying for control channel jamming issues In this chapter, we propose a cooperative relaying scheme to circumvent the control channel jamming problem in the vehicular networks, whereby the vehicles outside of the jamming area serve as relays to help forward the received control channel signal to the victim vehicles through another jamming-free service channel. By combining the signals from all relays using the Selection Combining (SC) or Maximum-Ratio Combining (MRC) methods, the spatial diversity provided by the relays can be effectively exploited to reduce the outage probability of the victim vehicles. Theoretical models are developed to characterize the performance of this cooperative anti-jamming relaying scheme, which take into account both the large-scale path loss and small- scale channel fading between

relaying and victim vehicles under different jamming scenarios. We also propose a relay selection scheme to optimize the performance of the victim vehicles under the relay number constraint.

Chapter 5: Cooperative Relay Beamforming for Control Channel Jamming in Vehicular Networks In this chapter, we extend the anti-jamming problem into multi-antenna RSU scenarios and propose a two stage anti-jamming scheme, whereby the vehicles who have successfully decoded the signal received in the first stage will be selected as relays to cooperatively serve the victim vehicles in the second stage using the coordinated beamforming techniques over a jamming-free service channel. By taking advantage of the multi-antenna gain provided the RSU and spatial diversity provided by the relay vehicles, the transmission reliability of all vehicles can be significantly improved under the threat of jamming attacks. The anti-jamming beamformer design problem is formulated as a Mixed-integer Nonlinear Programming (MINLP) problem, which is intractable in general. We address this challenging problem by reformulating it as a sequence of convex sub-problems using the semi-definite relaxation (SDR) and convex-concave procedure (CCP) methods, and then propose an iterative algorithm to find the approximate solution for the convex sub-problems.

Finally, Chapter 6 concludes the thesis with the summary of the overall results and the perspective for the future research.

Chapter 2

Vehicular Networks and Security Issues

In this chapter, we introduce the general architecture of a vehicular networks and the standard sets proposed by the IEEE and the ETSI, compare their differences on the architecture, applications and message format. Afterwards, we discuss the security issues in vehicular networks via analyse the security requirements, existing defending systems and mechanisms. Thus, we confirm that the Sybil attack and the RF jamming attack are the two major threats to the vehicular networks. Afterwards, based on the schemes proposed in the related work, we discuss the suitable Sybil detection and anti-jamming schemes in vehicular communication environments.

2.1 Vehicular Networks

A vehicular network is a generic term for the integrated application of communications, control and information processing technologies to the transportation system. It covers all modes of transport and considers all elements of the transportation system the vehicle, the infrastructure, and the driver or user, interacting together dynamically. Its overall function is to improve decision making, often in real time, by transport network controllers and other users, thereby improving the operation of the entire transport system.

2.1.1 Vehicular Networks Architecture

As shown in Figure 2.1, is a fundamental architecture of ITS, which includes vehicles, Road-side Unit (RSU) and backbone networks, such systems will need to support both vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communications.

In vehicular networks, each vehicle takes on the role of sender, receiver, and router to broadcast information to the vehicular network or transportation agency, which then uses the information to ensure safe, free-flow of traffic. For communication to occur between vehicles and Road Side

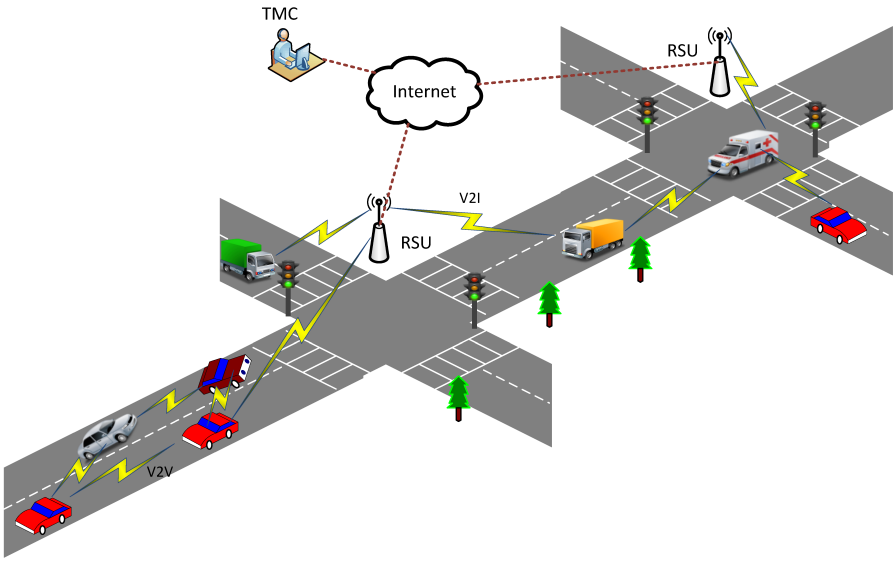


Figure 2.1: The architecture of vehicular networks.

Units (RSUs), vehicles must be equipped with some sort of radio interface or On Board Unit (OBU) that enables short-range wireless ad-hoc networks to be formed. Fixed RSUs, which are connected to the backbone network, must be in place to facilitate communication. The number and distribution of roadside units is dependent on the communication protocol is to be used.

To establish a cooperative system, several standardization efforts has been provided focusing on the communication architecture and protocol standards. In USA, the Dedicated Short-Range Communications (DSRC) standards suite is based on multiple cooperating standards mainly developed by the IEEE. In Europe, the ETSI TC ITS working group defines standards for ITS systems. Furthermore, two main ITS organizations emphasizing the ITS related cooperative work can be identified: CAR 2 CAR Communication Consortium and Safety pilot. More details are given in the following subsections.

2.1.2 Vehicular Networks Standards

DSRC/WAVE Standards The IEEE WAVE communication architecture is based exclusively on the top of IEEE 802.11p. In the literature, often DSRC (Dedicated Short Range Communications), WAVE (Wireless Access in Vehicular Environments) or even IEEE 802.11p are used to designate the entire protocol stack of standards dealing with VANETs. Full-use WAVE standards are in the process of being published as illustrated in Figure 2.2. It is made up of two sub-standards.

For a maximum of interoperability and for the purpose of standardization of frequencies with which the VANETs work, the U.S. government represented by the FCC (Federal Communication Commission) attributed the band 5850 to 5925 GHz (75 MHz band wide) [51]. This band is known as Dedicated Short Range Communications (DSRC). The use of the DSRC band is not subject to

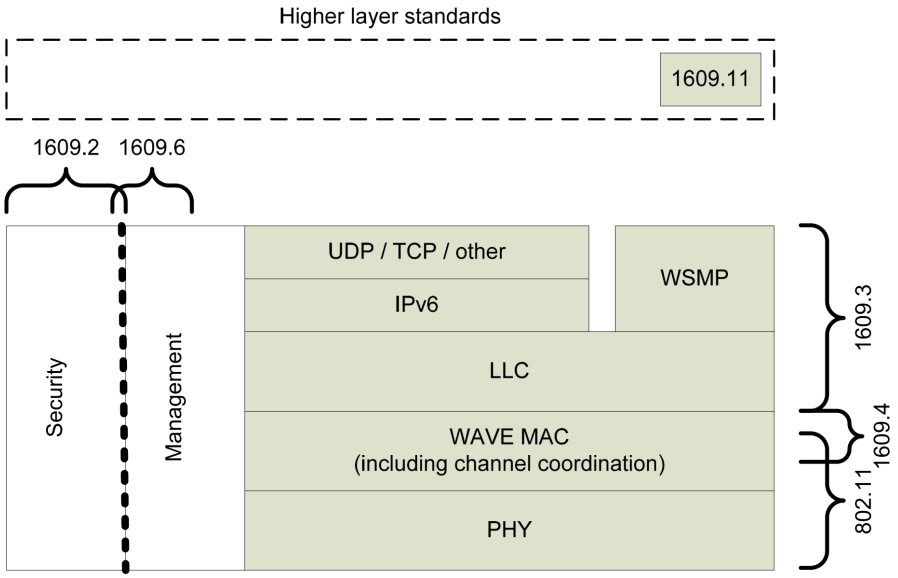


Figure 2.2: DSRC/WAVE Standards [1].

a license, but rather to strict rules of use.

The WAVE IEEE 1609 family (Standard for Wireless Access in Vehicular Environments) defines an architecture and a complementary set of standardized protocols, services and interfaces that allow all WAVE stations to operate in a VANET environment and establish Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications. The WAVE architecture defines also the security of exchanged messages. WAVE Standards form together the basis for the implementation of a wide set of applications in the transportation domain, they include vehicles safety, automatic tolls, improved navigation, traffic management and many other applications [51].

ETSI Standards The ETSI TC ITS is organized on five working groups. WG1 develops the basic set of application requirements and services, WG2 provides the architecture specification and addresses the cross layer issues, WG3 provides network and transport protocols for ITS G5, WG4 provides the media investigation of IEEE 802.11p, and WG5 works with the security system.

As shown in Figure 2.3, the ETSI ITS-S communication architecture is described in ETSI standard EN 302 665 [2]. It consists of four layers: access, networking/transport, facilities and applications.

After detailed the communication architectures of IEEE DSRC/WAVE and ETSI standards, a brief comparison is shown in Table 2.1. We can find out that in the architecture of IEEE, the lower layers (PHY and MAC) are rely on the protocol IEEE 802.11p. This protocol is the only choice and can be used for scenarios where the focus is on short range communication. It is very suitable for safety messages as it uses the concept of the dedicated CCH through which which urgent traffic can be prioritized. And in the access layer of ETSI standard, different technologies are combined

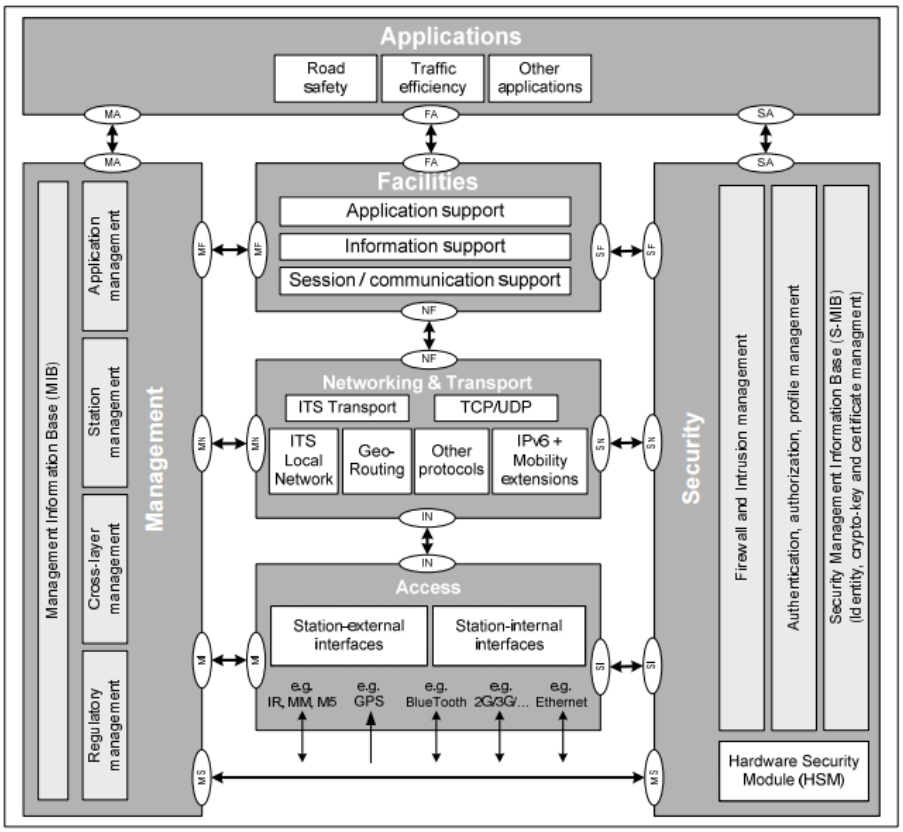


Figure 2.3: ETSI Standards [2].

(802.11p,3G,4G,satellite). Therefore, the implementation with interface handover is a tough job, which may impact the communication efficiency.

Table 2.1: Comparison between DSRC/WAVE and ETSI

	DSRC/WAVE		ETSI	
Application	Safety or Non-Safety		Road Safety, Traffic Efficiency and Others	
Transport	TCP/UDP	WSMP	TCP/UDP	BTP
Networking	IP		IP	GeoNetworking
LLC	IEEE 802.2		IEEE 802.2	
Range	802.11p WBSS		802.11p (enable outside the context of a BSS)	
Channel	SCH	CCH/SCH	SCH	CCH/SCH
PHY	5.9 GHz		ITS-G5	

Furthermore, the comparison messages are presented in Table 2.2. In this table we can find out that both architectures support TCP/IP protocol stacks and the IP packets can be well used for multi-media contents. Otherwise, specific network and transport protocols are defined in each architecture(WSMP in WAVE, BTP & GeoNetworking in ETSI), which have privilege to transmit messages using CCH, in order to improve the safety and efficiency of road traffic.

Table 2.2: Message Delivery in Both Architectures

	DSRC/WAVE		ETSI	
Message	IP Packet	WSM WSA	IP Packet	CAM DENM IP
Size	1500 bytes		1500 bytes	
Transport Networking	TCP/UDP IP	WSMP	TCP/UDP IP	BTP GeoNetworking
Address	IPv6	PSID	IPv4/IPv6	ITS-ID
Channel	SCH	CCH SCH	SCH	CCH SCH

2.2 Security Requirement

As one type of wireless access network, vehicular networks is vulnerable to different kinds of attacks. Therefore, the security is an important topic for both research and industrial development. The security requirement in vehicular networks can be roughly described by using the following properties:

- **Authenticity**: both entities and messages need be authenticated for the reason first to prevent external attacks. Secondly, to ensure the received message is came from a reliable resource, which is not fabricated by a fake node.

- **Integrity**: ensures that a message was not altered between the moment it was sent and received as the received message must match the message sent. Then the receiver is able to corroborate the sender's identity during the transmission. Integrity protects against the unauthorized creation, destruction or alteration of data.

- **Non-repudiation**: ensure that after the authentication and integrity checking, the entity cannot deny having participated in a communication event.

- **Confidentiality**: ensure that the confidential and sensitive information is well encrypted during communication procedures. External nodes are not able to understand confidential information. However, confidentiality is not always necessary in vehicular networks, most safety-related messages like beacons do not contain sensitive information.

- **Availability**: the network and applications should remain operational even in the presence of faults or malicious conditions. This requires not only secure but also fault-tolerant design, resilience to depletion attacks, as well as survivable protocols, which resume their normal operations after faulty participants removed [52]. Furthermore, under some special condi-

tions like traffic congestion, network should also be able to change some parameters like beacon frequency in order to ensure the availability of channel resource.

- **Privacy and Anonymity**: most privacy issues are related to position and identifiers in vehicular networks. Therefore, in order to protect the position and identity privacy, it should be impossible for any observer to learn if a specific vehicle has transmitted or will transmit a message, and it should be also impossible to link any two or more messages of the same vehicle [53].

However, it is impossible to satisfy every requirement. For example, in order to confirm the authenticity and integrity of beacon message, a public key certificate and a signature of sender are always encapsulated in transmission frame. On the other hand, these security mechanisms lead to a significant consumption of communication bandwidth and processing power [54], which affect the availability of the network. Otherwise, there exist always a trade off among system efficiency, security, and privacy during communication procedures [4].

2.3 Security Services

Vehicular communication security has drawn much attention in recent years, many security mechanisms have been proposed and several of them are normalized and recommended by IEEE and ETSI. In this section, we detail the security services recommended by IEEE and ETSI.

2.3.1 IEEE 1609.2 Standard

Security services recommended by IEEE are specified in IEEE 1609.2 standard [3]. It defines secure message formats and processing for use by Wireless Access in Vehicular Environments (WAVE) devices, including methods to secure WAVE management messages and methods to secure application messages.

Generally speaking, the security services defined in IEEE 1609.2 standard manage everything dealing with the encryption and decryption of secured data, certificate generation and validation, as well as signature generation and validation, all of which are supported by elliptic curve cryptographic (ECC) methods. It also describes administrative functions necessary to support the core security functions.

For more details, the following mechanisms are defined in this standard:

- Two types of entities: Certificate authority entities (CA entities) and end entities.
- Public key infrastructure architecture.

- Digital signature using Elliptic curve cryptosystems (ECC).
- Asymmetric encryption with ECC.
- Purely symmetric scheme AES-CCM.

A general model for security processing is illustrated in Figure 2.4 wherein security services are invoked by a secure communications entity (SCE) and return their output to the same SCE.

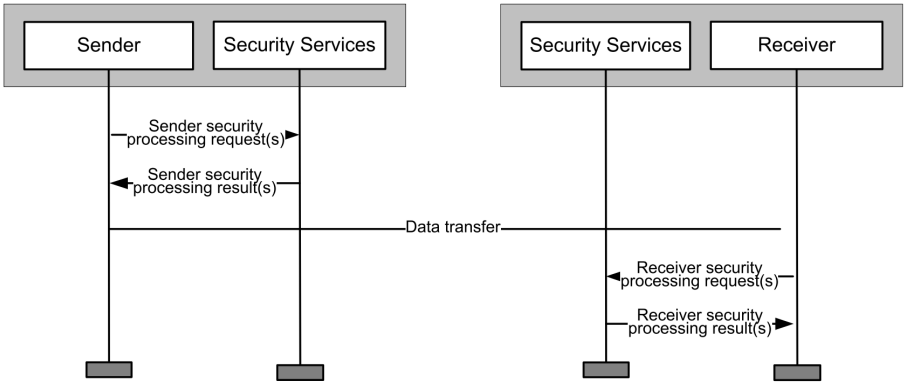


Figure 2.4: Process flow for use of IEEE 1609.2 security services. [3]

2.3.2 ETSI Security Services

Not like IEEE, in ETSI standards security services is not defined as sub-functions in management plane [2].

In ETSI TS 102 941 standard [55], the trust and privacy management for ITS communications are specified. It identifies trust establishment and privacy management required to support security in ITS environment and the relationships that exist between the entities themselves and the elements of the ITS reference architecture. The document starts by presenting ITS authority hierarchy, which is a PKI composed of an Enrolment Authority (EA), Authorization Authority (AA) and a Root CA, and used for distribution and maintenance of trust relationships between ITS stations and authorities or other ITS stations.

The ETSI TS 103 097 [56] standard specifies different security headers and formats to ensure the interoperability of the different elements and security information that are being exchanged between the ITS stations for security purposes. The main security header is the SecuredMessage structure, which specifies how to encode a generic security message, which is itself encapsulated inside a GeoNetworking packet. In this standard, a new certificate format is defined and also how to encode different information required by each type of certificate.

Furthermore, some European projects also follow the security rules defined by ETSI. However, they changed the name of several entities. For example, in the project $SCOOP@F$ [50], they use

the name Long Term Certificate Authority (LTCA) instead of Enrolment Authority (EA), and Pseudonym Certificate (PCA) instead of Authorization Authority (AA) in ETSI standards.

2.3.3 Defending System

Based on the analysis of security services proposed by IEEE and ETSI, it can be noticed that although defined separately, it is commonly agreed in both standard sets that security services implemented in vehicular networks rely on a system contains: cryptography mechanisms, certificate generate mechanisms and certificate management mechanisms. Moreover, some privacy protection mechanisms can also be utilised in this system including anonymous and pseudonymous.

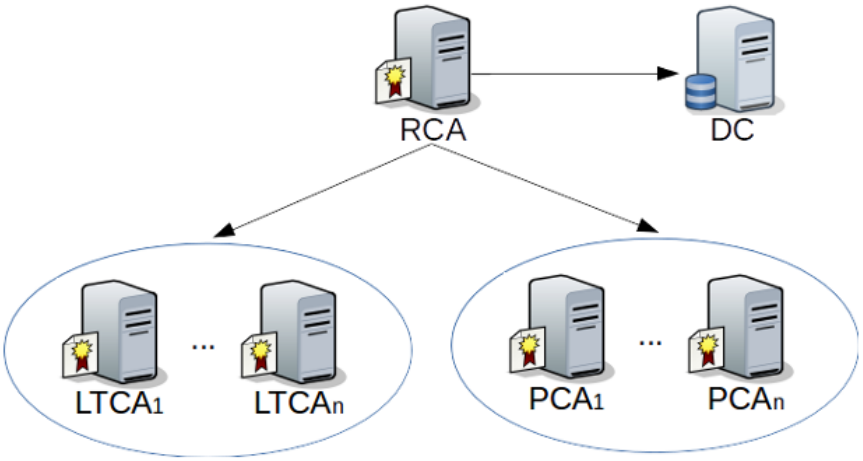


Figure 2.5: Defending system in vehicular networks.

The architecture of a defending system in vehicular networks is illustrated in Figure 2.5. In this system, there are three types of authorities: Root CA, Long Term Certificate Authority (LTCA) and Pseudonym Certificate Authority (PCA). The Root CA issues certificates for LTCA and PCA. It also defines and controls policies among all subordinate certificate issuers. The LTCA issues for each ITS-Station a Long Term certificate that is valid for a long period. This Long Term certificate is only used to identify and authenticate the ITS station (ITS-S) within the PKI, and never used in V2X communication for privacy reasons. It also enables ITS-S to request pseudonym certificates. The PCA issues a short lifetime certificates called Pseudonym certificate, which are used in V2X communications. The PCA guarantees privacy of requesting ITS Stations since it is technically and operationally separated from the LTCA, which is the only authority that knows the real identity of the ITS-S.

In vehicular networks, each vehicle has a 16 bits Unique Identity (UID) to to the authentication by communicating with the Long Term Certificate Authority (LTCA). The LTCA will send him back a long term certificate which contains the permission of this vehicle and the types of

messages it can generate after the authentication is successfully done. Afterwards, this vehicle can communicate with other ITS stations. During the communication procedure, the privacy protection is extremely important. In most cases, instead of using vehicles' real identities, the using of pseudonyms are considered can prevent the trace for vehicle's location history and other information like situation due to the unlinkability between two pseudonyms.

Generally speaking, with the help of hybrid cryptographic scheme and the PKI, the communication procedures can be considered secure and all the players in vehicular networking have to follow the rules. In this case, confidentiality, integrity, availability and authenticity are considered satisfied. Furthermore, with digital signature the non-repudiation requirement can also be satisfied.

The defending system in vehicular networks has gained widespread attention in recent years. In [57], the authors listed several popular pseudonym schemes and analysed their advantages and disadvantages the same time. In [58] a distributed pseudonym scheme is provided, they divided an urban into many small cells and distributed the pseudonyms to the distributed servers in each cell. For information plausibility check, using sensors to detect the real situation and installation of reputation systems are widely used in researches. In a reputation system, the reliability of an information depends on the reputation score of the transmitter. In most of the cases a threshold is set and any score greater than the threshold is thought to be reliable. However, the threshold need to be chosen carefully. In [59], a reputation-based announcement scheme is proposed, in which the reputation score of a player is based on the feedbacks of other players in a certain period.

2.4 Vulnerability Analysis

These services can protect the privacy of ITS stations, the authenticity and integrity of messages in vehicular communication environments. However, there exist also several security issues in vehicular networks. In our work, we concentrate on two major issues: privacy protection issue may leads to the Sybil attacks and the availability issue makes vehicular networks vulnerable to the jamming attacks.

2.4.1 Sybil attacks

The Sybil attack was firstly mentioned in the year of 2002. In [5] Douceur mentioned that in peer-to-peer system with no logically central, trusted authority to vouch for a one-to-one correspondence between entity and identity, it is possible for an unfamiliar entity to present more than one identity. In vehicular networks, Sybil attacks can cause damage in both Networking layer and Application layer.

Since the CSMA/CA is implemented in Networking layer, the cooperation among virtual nodes leads to the possibility of using more channel resource than other benign nodes. In Application layer, the virtual nodes also take part in communicating with other ITS stations. Under this circumstance, when a malicious node uses multiple pseudonyms at the same time, the virtual nodes, generated based on the usage of pseudonymous, can help to increase the influence of fake safety messages by broadcasting them to other benign nodes. In addition, several proposed driving safety and traffic efficiency services are based on voting scheme [60]. With the help of virtual nodes, the malicious node can easily take advantage in voting.

In the past a few years, many algorithms on detecting and defending the Sybil attack have been proposed in various ways in different networking environments, especially in wireless scenarios. The proposed Sybil detection methods can be divided into secure key based methods, resource testing based methods, reputation based methods and position based methods.

Secure Key-based Techniques The secure key-based techniques rely on using trusted certification in order to establish trust between entities. Several works based on wireless sensor networks [6], mobile ad-hoc [7] networks and also in vehicular networks [8] have been published in past a few years. This type of system needs a trusted third party or a centralized authority. The authority only provides valid keys to the honest nodes. However, the main drawbacks of this type solution are: firstly, the generation and management of a huge number of keys is costly. Therefore, it may cause a performance bottleneck in large-scale systems; secondly, it is hard to construct an central authority, which can be trusted by all participants. Furthermore, in vehicular scenarios, malicious nodes can easily get several legal identities (e.g., Pseudonymous), then the trusted certification based prevention methods would not work in this case.

Resource Testing-based Techniques The goal of resource testing is to attempt to determine if a number of identities possess fewer resources than would be expected if they were independent [6]. One example in wireless environment is the collision rate in MAC layer. In 802.11 wireless networks, CSMA/CA is used as multiple access method. The transmitters retransmit data after a back off prior when collision occurred in transmission channel. In this case, one hypothesis is that due to the cooperativeness among the Sybil identities configured on the same physical device, the local collision rate would be lower than in normal cases. However, one of the main drawbacks is that this abnormal state can be easily discovered by the monitoring system using some statistical methods, but the Sybil identities cannot be distinguished from normal users.

Reputation System In recent years, reputation systems have received a significant amount of attention as a solution for mitigating the affects of malicious nodes in peer-to-peer systems and also in ad-hoc networks or Online Social Networks(OSN). In [9], the reputation functions have been evaluated into symmetric and asymmetric function. The results show that symmetric functions cannot distinguish normal users from malicious while nodes can improve their own reputation score. And asymmetric functions employs the notion of transitive trust, which force malicious nodes to build up trust before launching attacks. However, based on the permitted services in vehicular networks, like the traffic condition announcement service, malicious nodes only need to launch the attack during rush hour in order to maximize the profit that is earned by obstructing some vehicles from their paths. In this case, they still have time to get good reputations.

Position Verification In ad-hoc networks, position verification is considered as a promising approach for the detection of Sybil attacks in recent years. In this approach, networks verify the physical position of each node. The Sybil nodes are expected to be detected by using this approach because they are at the same position where the malicious node generates them.

Several methods have been proposed in the past few years. Most of them are based on the beaconing mechanism in vehicular communication. In [10, 11, 12], vehicles estimate the position of their neighbours based on the Received Signal Strength Indication (RSSI), compare it with the geographic position they claimed in beacons and calculate the mean square error. If the mean square error is greater than a threshold, the transmitter is considered as a Sybil node. In [13], the detection is improved by using multiple node observers in order to get better detection rate. And in [14] the detection system is developed based on an information-thoeretic framework. Some base stations are implemented in this method and the Received Signal Strength are collected by the base stations.

As we have seen above, position verification relies on witness and verification by the neighbour nodes. In this case, every node is required to estimate neighbours' position by using RSSI or other information, and compare it with the position that neighbours announced. However, this type of methods are costly while the traffic density is high, and the motivation of users is also questionable.

Furthermore, in [61] a footprint detection scheme is proposed which can reconstruct the trajectory of a vehicle based on the signed message it received from different RSUs it passes by. And in [62] Rabieh *et al.* also proposed a scheme to detect virtual nodes based on their claimed location. A challenge packet is sent to the vehicle's claimed location by using directional antenna from RSUs. The exist Sybil attacks detection methods are briefly presented in Table 2.3.

Table 2.3: Sybil Attacks Detection Methods

Method	References	Advantage	Disadvantage
Secure Key-based	[6, 7, 8]	Can prevent the generation of fake identities	The generation and management of a huge number of keys is costly and it is hard to construct an central authority, which can be trusted by all participants
Resource Testing-based	[6]	No add extra overhead	Sybil identities can be hardly distinguished and located
Reputation System	[9]	Low extra overhead and low error rate	Malicious nodes can still get good reputations if the attack period is only a small percentage of time
Position Verification	[10, 11, 12, 13, 14]	High detection rate	The location estimation based on RSSI is not accurate enough

2.4.2 Radio Frequency Jamming Attacks

As mentioned in the previous section, in general, the vehicular networks are consisted of vehicles, Roadside Units (RSU) and backbone networks, supporting both vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communications. Several standardization efforts have been devoted to establish network architecture and communication protocols for the VNETs. In Europe, the ETSI TC ITS working group defines standards for Intelligent Transport System (ITS) [2], and in the USA, the Wireless Access in Vehicular Environments (WAVE) standards [1] are defined based on multiple cooperating standards. Although the upper layer protocols of these two standards are defined differently, the IEEE 802.11p protocols are supported in lower layer of both standard-sets [63]. The PHY layer of IEEE 802.11p is identical to the OFDM-based IEEE 802.11a protocol. However, in order to compensate for the increased delay spread in outdoor vehicular environments, IEEE 802.11p operates with reduced 10 MHz channel spacing within the 75 MHz bandwidth in the

5.9 GHz band, where a single control channel (CCH) and six service channels (SCH) are defined for V2I or V2V communications. Although not being explicitly specified in IEEE 1609.4 standard, it is commonly accepted that the control channel has high priority in VNETs, and should be mainly used for the transmissions of safety related messages [64] or management information [65], while other services can be carried out over the service channels.

Due to the open and shared nature of wireless medium, vehicular networks are vulnerable to RF jamming attacks since an attacker can easily emit an interference signal to prevent legitimate access to the medium or disrupt the reception of signal. An attacker may utilize different jamming strategies by exploiting the vulnerabilities of PHY and MAC layer protocols, which in general fall into two categories: single channel jamming strategies and multi-channel jamming strategies. For example, spot jamming is powerful in single frequency jamming attack, which can contentiously block a specific channel or frequency, while sweep jamming is capable of blocking several channels by rapidly shifting from one frequency to another, and barrage jamming is able to jam multiple frequencies simultaneously with larger transmit power budget. The attacker can either permanently transmit random electromagnetic signal not following any specific protocols (i.e., constant jammer), or inject regular packets without any gaps between transmissions to block the transmissions of other legitimate users (i.e., deceptive jammer). It is also possible for an attacker to launch some intelligent jamming attacks by exploiting the semantics of higher protocols layers, such as CTS, ACK, and DATA corruption jamming for IEEE 802.11 protocols, or by flooding the routing requests. These RF jamming attacks can lead to disastrous consequence to the VNETs since safety related messages cannot be delivered to the vehicles timely if the control channel is blocked.

Many countermeasures have been proposed to address the RF jammming issues from different technical perspectives. For example, in the physical layer, frequency hopping techniques, such as Frequency-Hopping Spread Spectrum (FHSS) [15], Direct Sequence Spread Spectrum (DSSS) [15], and Hybrid FHSS/DSSS [15] have been widely adopted to avoid the jamming interference by rapidly switching between frequencies, or spreading the signal in a much wider band to allow for greater resistance to unintentional and intentional interference. However, the main drawback of these solutions is that they require much wider bandwidth than the original signal. Some novel techniques, such as Ultra Wideband (UWB) technologies, antenna polarization, and multi-antenna techniques [16, 17] are also proposed to address the jamming problem. In particular, multi-antenna technique is a promising anti-jamming solution since it can avoid the interference signal from unwanted sources. It also can improve the transmission efficiency and reliability by exploiting the diversity gain and antenna gain [18]. There are also some countermeasures that try to

provide a proactive/reactive protection against attacks in the design of MAC layer protocols (e.g., carrier-sense multiple access with collision avoidance (CSMA/CA), time division multiple access (TDMA), etc.), or use some channel hopping strategy to address the multi-channel jamming issue by accounting for the adversarial behaviour of the jammer [19, 20]. Some solutions have been proposed for the jamming attacks in multi-hop networks, which attempt to reroute traffic around the jammed area, or retreat from the jammer leveraging the mobility of the nodes [21, 22, 23, 24].

Firstly we discuss the existing anti-jamming solutions for wireless networks. Thus, based on the characteristics of vehicular networks, we concentrate on whether these methods are applicable in vehicular communication environments, and give the possible solutions for anti-jamming in vehicular networks, especially for the control channel jamming issues.

The anti-jamming schemes in wireless communication environment can be roughly divided into three major categories: Spread Spectrum (SS) communications, multi-antenna interference cancellation and cooperative communication schemes.

Spread Spectrum (SS) communications With the help of signal processing techniques, countermeasures are proposed based on the use of Spread Spectrum (SS) communications which refer to signal structuring techniques that employ direct sequence, frequency hopping or a combination of these. An anti-jamming scheme is proposed in [25], based on the timing channel techniques. The objective is to create a low bit-rate overlay that exists on top of the conventional physical/link-layers and that would survive in the presence of a persistent broadband interferer. In [26], a Randomized Differential DSSS (RD-DSSS) scheme for DSSS-based broadcast communication is proposed. RD-DSSS encodes each bit of data using the correlation of unpredictable spreading codes, which relies completely on publicly known spreading codes. In order to defeat reactive jamming attacks, RD-DSSS uses multiple spreading code sequences to spread each message and rearranges the spread output before transmitting it. In [27], Liu *et. al.* proposed a randomized distributed channel establishment and maintenance scheme to allow nodes to establish a new control channel using frequency hopping to address the control channel issue in wireless networks, which is applicable to networks with static or dynamically allocated spectrum. These techniques can help to resist the interference of jamming signals.

Multi-antenna Techniques More recently, antennas techniques are also considered can be used in anti-jamming since they can avoid the interference from unwanted source and improve the robustness of the communications. Becker *et. al.* considered to deal with the jamming issue in wireless networks via beamformer design, and two genetically evolved anti-jamming beamforming array are introduced in [28] and [29] respectively. These genetic algorithms (GAs) use

SINR as its fitness function. For example, the authors in [30] propose a Technology Independent Multi-Output (TIMO) scheme, which exploits the MIMO capability to remove the interference signal from a single constant jammer. In [31], Shen *et. al.* propose a Multi-Channel Ratio (MCR) Decoding scheme, which can recover the desired signal from both constant and reactive jamming interference by leveraging the multi-channel characteristics of the MIMO systems. In [32], Yan *et al.* adopt the MIMO interference cancellation (IC) and transmit precoding techniques to counteract the reactive jamming attacks. Two novel mechanisms, namely, iterative channel tracking and signal enhancing rotation, are proposed to effectively sustain acceptable throughput under reactive jamming attacks. In [66, 67, 68], several hybrid security schemes consisting of both multi-antenna techniques and cooperative relaying are proposed in order to address the eavesdropping issues in wireless communications. By taking advantage of the diversity gain provided by these two schemes, some intermediate nodes help relay signals to the legitimate destination using distributed beamforming, while the remaining nodes attempt to jam the eavesdropper, which protect the data transmissions in both phases.

In Chapter 5, we also consider the case whereby the RSU is equipped with multiple antennas to transmit common information to multiple vehicles. This problem can be formulated as a multicast beamforming problem, which is NP-hard. In order to achieve favourable performance-complexity trade-offs, some approximation and reformulation efforts have been made in the past decade. In [47], the NP-hard multicast beamformer design problem is approximated to a convex optimization problem by using semidefinite relaxation (SDR) techniques. In [44], Luo *et al.* extended the multicast beamforming problem into multi-group scenarios and show that Lagrangian relaxation coupled with suitable randomization/cochannel multicast power control yields computationally efficient high-quality approximate solutions. In [69], a multicast beamforming problem with non-convex constraint is reformulated by using the convex-concave procedure (CCP), whereby the non-convex constraint is replaced by its first order Taylor expansion, then a sequence of convex sub-problems are solved iteratively until convergence. Based on these relaxation and reformulation techniques, several works have been done in the past a few years. In [70], Tao *et al.* studied the coordinated multicast beamfoming problem in multi-cell cellular networks, and two beamformer design problems, namely quality-of-service (QoS) beamforming and the max-min SINR beamforming are considered. In [71] and [72], beamformer design under per-antenna power constraints and antenna selection problem in multicast multi-group beamforming are studied respectively.

Cooperative Scheme For fixed frequency channel scenarios, several novel schemes are proposed to address the RF jamming issues via cooperative schemes. Specifically, in [33], Cagalj

et al. proposed a reactive anti-jamming scheme for wireless sensor networks using wormholes. The basic idea is that jammed nodes can use channel diversity, to establish communication with another user outside the jammed area. This work suggested that by the cooperative scheme can be used in fixed frequency channel networks. Based on this ideology, Mustafa *et al.* proposed a Jamming-resilient multipath routing protocol [34], where a source node selects multiple different paths for reaching the destination in advance. When one of the paths fails, other working paths will be used to deliver packets and thereby maintain end-to-end availability. Moreover, in [35], Zheng *et al.* studied the performance of cooperative communications under jamming attacks. They implemented both half-duplex (HD) and full-duplex (FD) operation modes in a simple three nodes scenario, where a relay helps forward the signal from the source to the destination. In [36], a cooperative anti-jamming scheme is proposed to optimize the fairness constrained network throughput in the presence of jammers, wherein users cooperate at two levels: regulate their channel access probabilities so that the victims gain a higher share of channel utilization and form a virtual channel array to increase the link capacity. In [37], Li *et al.* investigated the anti-jamming power control of transmitters in a cooperative wireless network attacked by a smart jammer with the capability to sense the ongoing transmission power before making a jamming decision. This problem was formulated as a Stackelberg game, and the Stackelberg equilibrium for this game is derived based on analysis of the optimal strategy for both sides. Moreover, cooperative schemes have also been considered in the context of anti-eavesdropping. For example, relay nodes can be used to exploit the PHY layer properties of wireless channels in order to support a secured end-to-end transmission in the presence of eavesdroppers [73]. In [74], cooperative relaying is used to improve the secrecy capacity in wireless networks by choosing appropriate relay nodes with ”strong” transmission link to intended destination node and ”weak” link to eavesdropper.

For VNETs, in [75, 76], the performance of 802.11p-based vehicular communications is evaluated in the presence of RF jamming attacks, which confirmed that the vehicular networks are extremely vulnerable to the RF jamming attacks. However, the proposed anti-jamming solutions in normal wireless networks may not be adapted in VNETs. For example, neither the spectrum spread (SS) techniques nor the multi-antennas techniques are suitable for control channel jamming issues in vehicular networks since there is only one fixed frequency control channel is available and the multi-antennas are not widely used on the vehicles. In [77], a real-time Medium-Access Control-based (MAC-based) detection method is proposed, which can more accurately distinguish the causes of failed transmissions, such as contention collisions, interferences, and jamming attacks in the VNETs. Unfortunately, no precise anti-jamming solution is given in their work. In [38], a cooperative scheme for medium access control (MAC) in VNETs is proposed, where

the neighbouring nodes can corporately utilizing unreserved time slots to retransmit the packets which failed to reach the target receiver due to a poor channel condition. This cooperative scheme can also be used in VNETs anti-jamming. In this way, we consider to address the control channel jamming issues by adopting a cooperative scheme. Specifically, in our anti-jamming scheme design, we consider also the time varying wireless channel and cooperative scheme in PHY layer (e.g. SC or MRC). After detailed theoretical analysis, we can get the close form expression of the outage probability for both schemes. Thus, we propose a relay selection method by adopting the TDMA scheme and optimise the effectiveness of our proposed anti-jamming scheme. To the best of our knowledge, no similar method has been proposed for jamming issues in vehicular networks.

The exist anti-jamming schemes for wireless networks are briefly presented in Table 2.4.

Table 2.4: Anti-jamming Schemes for Wireless Networks

Method	References	Advantage	Disadvantage
Spread Spectrum (SS)	[15, 25, 26, 27]	Resist the interference of jamming signals	Not available for fixed frequency channel.
Multi-antennas techniques	[28, 29, 30, 31, 32]	Can avoid the interference signal from unwanted sources	Multi-antennas are not widely used on the vehicles
Cooperative Scheme	[33, 34, 35, 36, 37, 38]	Provide spatial and antenna diversity can enhance the robustness of communications	Cooperative scheme design is difficult

Chapter 3

Sybil Attack Detection in Vehicular Networks

Considering virtual nodes have to avoid the positions that are captured by the benign vehicles, their driving patterns become erratic, especially in dynamic traffic environments. In this chapter, we consider the possibility to detect Sybil attacks based on the variation of their driving patterns using some machine learning method. Generally speaking, machine learning is a process in which a set of threshold parameters is trained to classify an unknown behaviour [39]. In this thesis, three major methods are considered in our work: Distance based clustering, Support Vector Machine (SVM) and k-nearest neighbours (kNN).

Distance based clustering is a kind of standard hierarchical clustering methods [78, 79], which can handle data with numeric and categorical values using complex similarity measures. Specifically, we choose the Mahalanobis distance to measure the variances of driving components between users in our work since the variance of users' driving component in different directions are different.

Support Vector Machine (SVM) [80, 81] is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy. It leverages a flexible representation of the class boundaries and implements automatic complexity control to reduce overfitting [82]. Furthermore, it often has good generalization performance and the same algorithm solves a variety of problems with little tuning, which makes SVM suitable for dynamic environment [83].

k-Nearest Neighbours (kNN) is proposed in the year of 1965 [84]. It is a non-parametric method and frequently used as a powerful machine learning algorithm in solving classification and clustering problems. The misclassification rate of the kNN rule approaches the optimal Bayes error rate asymptotically as k increases, and is particularly effective when the probability distributions of the feature variables are not known [85].

3.1 Attack Model

In this section, we discuss the possibility and strategies to launch attacks where defending system is implemented. With the implementation of a system made up of PKI and pseudonymous, the confidentiality, integrity and non-repudiation can be well protected.

The Architecture of PKI and the Pseudonym Certificate (PC) generation procedure are illustrated in Fig. 3.1. There are three types of authorities in PKI: Root CA (RCA), Long Term Certificate Authority (LTCA) and Pseudonym Certificate Authority (PCA). There functions are listed as follow:

- Root CA: issues certificates for LTCA and PCA. It also defines and controls policies among all subordinate certificate issuers.
- LTCA: issues for each ITS-Station a Long Term Certificate (LTC) that is valid for a long period, which enables ITS-S to request pseudonym certificates.
- PCA: issues short lifetime PCs, which are used in V2X communications. The PCA guarantees privacy of requesting ITS-Ss since it is technically and operationally separated from the LTCA, which is the only authority that knows the real identity of the ITS-S.

3.1.1 Vulnerability Analysis

In this PKI system [55], each vehicle has a Unique Identity (UID) to do the authentication with the LTCA. Each authenticated vehicle can get a LTC which contains the service permission of this vehicle. This procedure has to be done by the vehicle manufacturers. Once the LTC is generated, it cannot be modified any more.

Afterwards, this vehicle can communicate with other ITS stations. During the communication procedure, in order to protect users' privacy, instead of using vehicles' UIDs, vehicles are recommended to use their pseudonyms in communicating due to the unlinkability between any two pseudonyms.

One vehicle can request several pseudonyms, this procedure is also illustrated in Fig. 3.1. The PCA generates and stores one certificate for each delivered pseudonym which contains the pseudonym along with its authorized public key and the key's lifetime. Vehicles are suggested periodically change their pseudonyms and one pseudonym has to be changed when the lifetime of its key is up to zero.

However, the privacy protection especially the using of pseudonyms are vulnerable to the Sybil attack. While one vehicle is using several pseudonyms together during the same time period, each

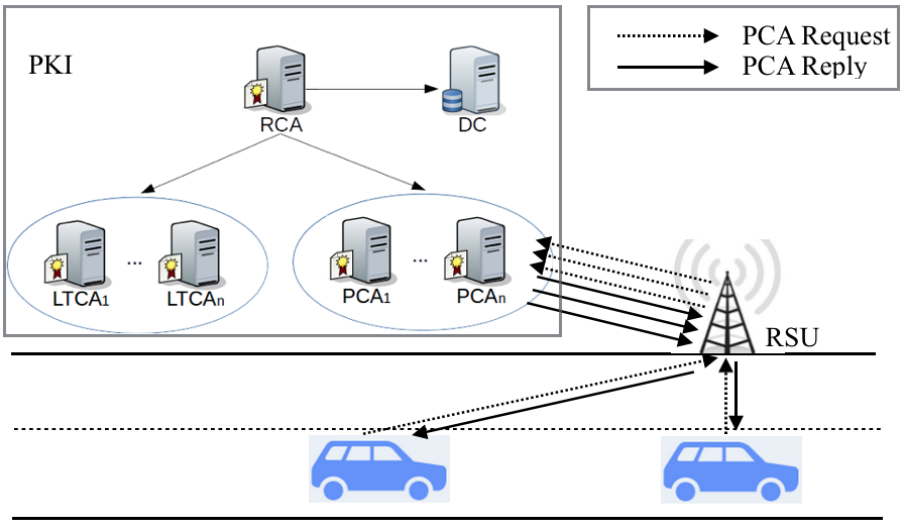


Figure 3.1: PKI architecture and PC generation procedure

pseudonym is an individual vehicle in the view of other ITS stations (include OBUs and RSUs) because each valid pseudonym has its own key pair for signing.

Messages exchanges between ITS stations are normalized in ETSI standards [86, 87], vehicles communicate their own condition by using Cooperative Awareness Message (CAM) and report road condition by using Decentralized Environmental Notification Message (DENM). Additionally, the public information are not encrypted in order to decrease the communication complexity and the unnecessary overhead.

Based on the analysis of defending methods and messages implemented, we can confirm that it is possible for one vehicle to use different pseudonyms together during the same time period to launch Sybil attacks.

3.1.2 Attack Strategies

Sybil attacks in vehicular communication environment can be generally divided into two procedures: virtual nodes generation procedure and launch attack procedure. Which is illustrated in Fig. 3.2. Where V_m is malicious node, V_b are begin nodes and V_v are virtual nodes created by the malicious node by using its own pseudonyms. Based on the standards of ETSI, CAMs can be used in virtual nodes generation procedure in order to let the virtual nodes known by other ITS stations and the launch attack procedure depends on using DENM to report fake road traffic condition to the RSU.

The detection method proposed in this article focus on virtual nodes generation procedure, Our major target is to detect the virtual nodes and eliminate them from the system by revoke the pseudonyms they are using.

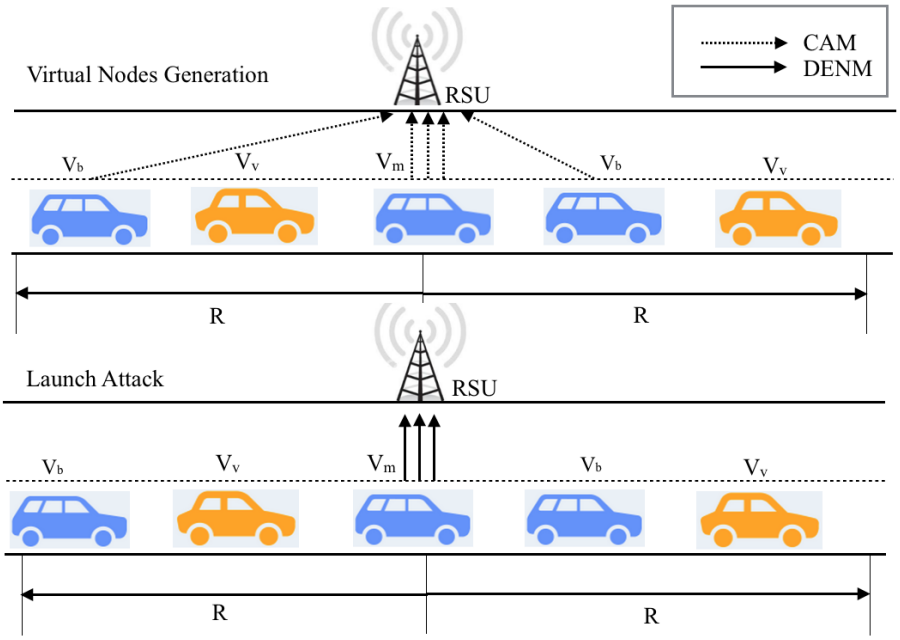


Figure 3.2: Attack model of Sybil nodes

For attack strategy [88], malicious nodes can be divided into rational and irrational attackers based on their profiles. Rational attackers have a specific target and irrational ones do not seek a specific outcome. In the scenarios that we describe, only rational attackers are being taken into consideration since they are more dangerous and more predictable. One assumption is that the rational Sybil attackers only launch attacks when traffic density is high. They report fake traffic condition information with the help of virtual nodes created by themselves, in order to mislead other ITS stations into making wrong decisions. In this way, the traffic density would be potentially decreased after the next intersection, and the attackers can reduce their travel time.

In [89], Hao *et al.* proposed two Sybil attack strategies: "Regular Attack"and "Smarter Attack". In Regular Attacks, malicious node broadcasts beacons for the virtual vehicles using the regular power. In Smarter Attacks, malicious node may reduce his communication range to make the virtual vehicles' behaviour looks reasonable. We only consider the V2I communication, therefore, regular attack strategy is chosen.

Under these circumstances, the strategy of malicious nodes would be upon the received CAMs, calculate reasonable location for virtual nodes in next time slot, then forge CAMs for the virtual nodes and broadcast them within its communication range. More details of virtual node generation procedure and CAM forge are presented in Algorithm 1.

Algorithm 1 Algorithm for virtual nodes generation

Input: CAMs from neighbours

Output: CAMs for virtual nodes

Initialisation :

$c \leftarrow 0$

Do the calculation lane by lane (k lanes in total)

for $l = 1$ to k **do**

  Select CAMs from a certain traffic lane

  Note their location and velocity in two arrays

  Get array $P[n]$ and $V[n]$

  For all vehicles in this traffic lane

  for $v = 1$ to n **do**

    Calculate the distance between each two adjacent vehicles n and $n-1$

    $d_n = P_{n-1} - P_n$

    Calculate average velocity of this traffic lane

    $\bar{v} = \sum_{j=1}^{n} V[j]/n$

    The number of virtual nodes can be set between vehicle $n-1$ *and vehicle* n

    $num = \lfloor \frac{d_n}{2\bar{v}} \rfloor$

    Set velocity and position for the virtual vehicles $R[n]$

    for $i = 1$ to num **do**

      $R_i = P_{n-1} + \bar{v} - 2nv_m$

    end for

    Push virtual nodes' location and their velocity into a matrix $N[i, j]$

    $temp \leftarrow num$

    for $m = c + 1$ to $c + temp$ **do**

      $N[m, 0] = P[c + temp - m]$

      $N[m, 1] = \bar{v}$

      $N[m, 2] = k$

    end for

    $c \leftarrow c + num$

  end for

end for

return $N[i, j]$

3.2 Vehicle Driving Pattern Description

In our algorithm, the Driving Pattern Matrix (DPM) need be constructed first. The driving pattern of a vehicle at certain time is described by using a five elements vector. For example, $\vec{V_n} = (x_{n1}, x_{n2}, x_{n3}, x_{n4}, x_{n5})$ represents the driving pattern of a vehicle at time t_1. Where

- x_{n1} represents time at time t_n
- x_{n2} represents location of vehicle at time t_n
- x_{n3} represents velocity of vehicle at time t_n
- x_{n4} represents acceleration of vehicle at time t_n
- x_{n5} represents vehicle's acceleration variation between time t_{n-1} and time t_n

Therefore, the DPM of a vehicle within time period (t_1, t_n) can be represented as:

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\ x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} & x_{n5} \end{bmatrix} \tag{3.1}$$

3.2.1 Metrics and Definition

We define first one theorem that will be used in our detection algorithm:

Theorem 1: We note $\mathcal{M}_n(\mathbb{R})$ the set of square matrix with real entries. If a matrix $M \in \mathcal{M}_n(\mathbb{R})$ is a symmetric matrix, there exists an orthogonal matrix C where has

$$C^{-1}MC = C^T MC = D \tag{3.2}$$

In this case, D is a diagonal matrix with real entries, and the main diagonal entries are the eigenvalues of matrix M.

3.2.2 Reference Matrix

In our scenarios, for the DPM V_i of each vehicle v_i within time period (t_1, t_n), we construct matrix $V_i^T \cdot V_i$, where exists:

$$(V_i^T \cdot V_i)^T = V_i^T \cdot (V_i^T)^T = V_i^T \cdot V_i \tag{3.3}$$

Therefore, the matrix $V_i^T \cdot V_i$ is symmetric. And as presented in **Theorem 1**, there exists an orthogonal matrix C has:

$$C^{-1}(V_i^T \cdot V_i)C = C^T(V_i^T \cdot V_i)C = diag(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}) \tag{3.4}$$

Where $(V_i^T \cdot V_i)u_j = \lambda_{ij}u_j (1 \leq j \leq 5)$. $\lambda_{ij}(respu_{ij})$ are the eigenvalue (resp eigenvector) of matrix $V_i^T \cdot V_i$ which represent the characteristics of the original DPM.

3.3 Distance Based Clustering

In this section, we present a method that mainly represent vehicles' driving patterns by using eigenvalues of their driving pattern matrix and the similarity measurement is based the Mahalanobis distance between vectors made up by their eigenvalues.

Definition 1: The statistical distance or Mahalanobis distance between two points $x = (x_1, \cdots, x_p)^T$ and $y = (y_1, \cdots, y_p)^T$ in the p-dimensional space $\mathbb{R}^1$ is defined as:

$$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)} \tag{3.5}$$

Where Σ is the covariance matrix.

The Mahalanobis distance can also be defined as a dissimilarity measure between two random vectors $\vec{x}$ and $\vec{y}$ of the same distribution with the covariance matrix Σ. If the covariance matrix is identity matrix, the Mahalanobis distance reduces to the Euclidean distance.

Eigenvalues of a certain matrix are a group of non-negative real numbers, and we imagine that a positive correlation exists between the similarity of matrices and the similarity their eigenvalues. Therefore, the Mahalanobis distance is chosen to measure the similarity of road users' eigenvalues of their DPMs in order to separate Sybil nodes from benign ones. Additionally, for a certain matrix, its eigenvalues decrease rapidly. Therefore, in our detection method the Mahalanobis distance is chosen instead of the Euclidean distance due to it accounts for the fact that the variances in each direction are different and otherwise, it accounts for the covariance between variables. The detection procedure can be described as follow:

- We note $\lambda_{i1}^{ma} = max\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\}$ and $\lambda_{i2}^{ma} = max\{\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\} - \{\lambda_{i1}^{ma}\}\}$ the two biggest eigenvalues values of each DPM.
- For n vehicles, a n rows and 2 columns matrix S is defined as: $S = (\lambda_{ij}^{ma})_{1 \leq i \leq n, 1 \leq j \leq 2}$.
- Select the median value of each column, get a vector $\overrightarrow{med} = (x, y)$ (median value is chosen instead of mean value due to the instability of Sybil nodes' eigenvalues.).
- Calculate the Mahalanobis distance between each row vector of matrix S and the vector $\overrightarrow{med}$.
- Decision criteria: Vector $\vec{r_i}$ is the i-th row of matrix S, who represents the driving patterns of the i-th vehicle. If the Mahalanobis distance between $\vec{r_i}$ and $\overrightarrow{med}$ is greater than a selected threshold value α, where $d(\vec{r_i}, \overrightarrow{med}) \geq \alpha$, the vehicle v_i is considered as malicious.

More details of the proposed Sybil detection method are presented in Algorithm 2.

Algorithm 2 Algorithm for virtual nodes detection

Input: CAMs from vehicles

Output: ID of malicious nodes

for $v = 1$ to n **do**
  Select the concerned n vehicles
  for $t = 1$ to 60 **do**
    Construct the DPM for each vehicle within 60 seconds
    Get $m[v]$
  end for
  Calculate the eigenvector of $m[v]$
  Get $\vec{D_v} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$
  Get $\vec{E_v} = (\lambda_1, \lambda_2)$
  The two biggest eigenvalues
end for
for $v = 1$ to n **do**
  Get $\vec{M} = (\lambda_{1m}, \lambda_{2m})$
  λ_{1m} *and* λ_{1m} *are median values of each group* λ_1 *and* λ_1
end for
for $v = 1$ to n **do**
  Calculate the Mahalanobis distance between $\vec{E_v}$ and $\vec{M}$
  Get D_v
  with a given threshold value th
  if $D_v > th$ **then**
    return v
    Output the IDs of malicious nodes
  end if
end for

3.3.1 Experimental Results

In the following simulations, SUMO simulator is chosen to simulate traffic flows in urban scenario. In [41], Zhuang *et al.* clarified traffic flow condition for different vehicle density. For rational Sybil attackers, they can probably benefit from attacks (Save travel time) in Near-capacity conditions. Therefore, we choose the vehicle density between 26 and 42 vehicles per kilometre in

our simulations. The Sybil nodes generation procedure follows the algorithm defined in Section III. For the detection method, we choose a window size of 60 seconds and vehicles within 10 second before and after the target vehicle. The detection algorithm is implemented in Matlab. More details are presented in Table 3.1:

Table 3.1: Parameters Used in Simulations

Parameter	Value
Simulation Scenario	Urban Scenario
Simulation Time	300s
Window Distance	1 km
Street Width	2 Lanes
Vehicle Velocity	40 - 60 km/h
Number of Vehicles Simulated	350 - 800
Communication Range	300m
Beacon Frequency	1 Hz

Number of Virtual Nodes For Sybil attacks, the attack strength is dependent on the virtual nodes that can be created by the malicious node. As illustrated in Fig. 3.3, t can be observed that the number of virtual nodes during one attack is dependent on the vehicle density in traffic lanes (e.g. If the traffic density is between 26 and 30 vehicles per kilometre, up to 14 virtual node can be generated by one malicious node.). This phenomena fits well our hypothesis of rational attacker. They always try to set virtual nodes to reasonable positions and avoid positions that are captured by other benign road users. If a position is captured by two different vehicles at the same time, it can be recognized as traffic accident in RSU's point of view.

Similarity Measurement Results **Definition 2:** The empirical cumulative distribution function F_n associated to an sample $X_1, \cdots, X_n$ is the function defined by:

$$\forall x \in \mathbb{R}, F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{X_i \leq x} = \begin{cases} 0 & \text{if } x < x_1^* \\ \dfrac{i}{n} & \text{if } x_i^* \leq x < x_{i+1}^* \\ 1 & \text{if } x \geq x_n^* \end{cases} \tag{3.6}$$

Where $X_1^*, \cdots, X_n^*$ are the ordered statistics associated to $X_1, \cdots, X_n$

Several similarity measurement results of different road users' driving patterns are detailed in Fig. 3.4 under different vehicle density. The two biggest eigenvalues are chosen to mainly represent the driving pattern of one vehicle.

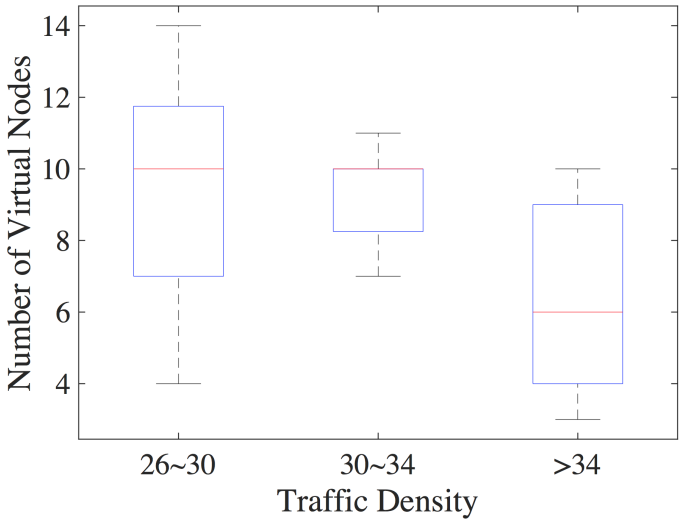


Figure 3.3: Number of virtual nodes in one attack

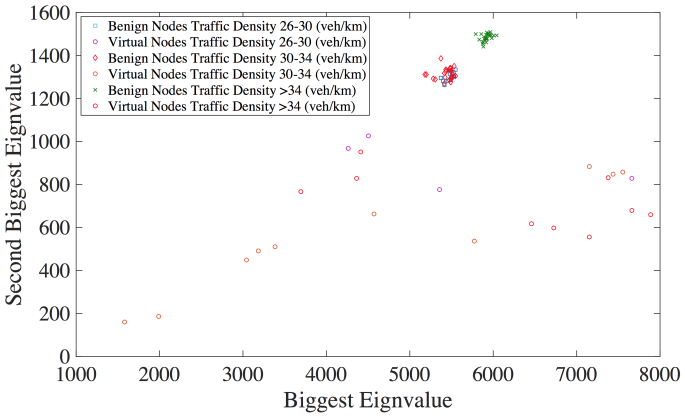


Figure 3.4: The first two biggest eigenvalues of DPM can be used to mainly represent the driving pattern of one vehicle

The driving patterns of benign nodes have obvious similarity and the driving patterns of Sybil nodes show erraticness. This result is caused by two reasons:

- The variation of virtual nodes' lifetime: We measured the lifetime for more than 200 virtual nodes, the CDF (eqn.6) is illustrated in Fig. 3.5. We can find out that more than 70% of the virtual nodes do not have the same length of lifetime as the benign ones (60 seconds).

- The erraticness of virtual nodes' movement: Virtual nodes should not be set to the positions that are captured by the benign ones.

Both these issues cannot be figured out by Sybil nodes, because they cannot control the components of other road users. They can only adjust their own strategies and make their driving patterns as reasonable as possible. Our detection system is directed against this weak point of Sybil nodes, making it possible to separate virtual nodes from benign ones.

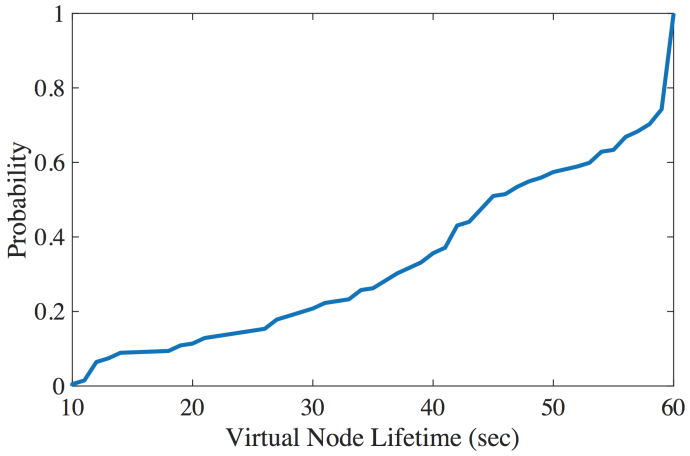


Figure 3.5: Lifetime of virtual nodes

Detection Rate **Definition 3:** We note that the Dirac delta function defined in $\mathbb{R}$ as follow:

$$\delta(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ +\infty & \text{if } t = 0 \end{cases} \tag{3.7}$$

Such as: $\int_{-\infty}^{+\infty} f(y)\, dy = 1$

Definition 4: We note that X is a discrete random variable has n observations $x_1, \cdots, x_n$ with probabilities $p_1, \cdots, p_n$. Then, a probability density function (PDF) f can be defined as follow:

$$f(t) = \sum_{i=1}^{n} p_i \delta(t - x_i) \tag{3.8}$$

The proposed detection mechanism can be recognised as a minimum distance classifier. After the Mahalanobis distance measurement, the PDF (eqn.8) of the Mahalanobis distance that each vehicle to the cluster centre is illustrated in Fig. 3.6.

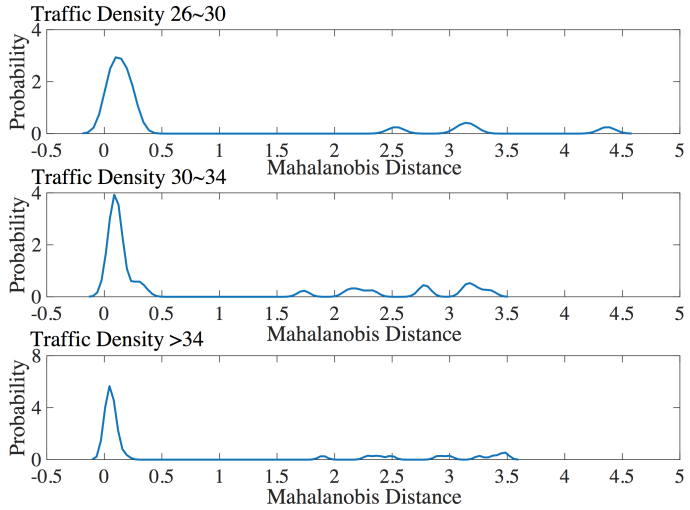


Figure 3.6: PDF of Mahalanobis distance

As shown above, in any case the benign nodes are the majority part in cluster, and there exists obvious similarity among their driving patterns. Therefore, the Mahalanobis distance from a benign node to the cluster centre is considered much smaller than from a virtual node to the same place, and the detection procedure is to set a threshold value to separate Sybil nodes from benign ones. In these scenarios, this threshold value can be set between 0.5 and 1.5.

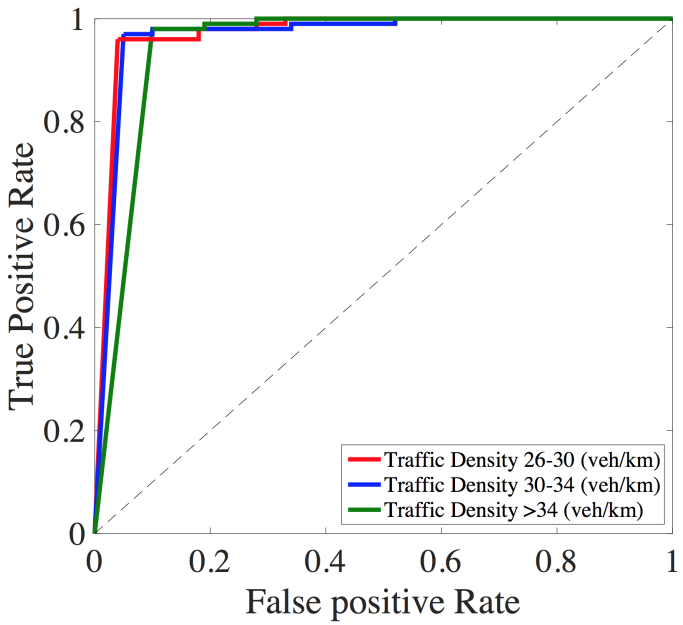


Figure 3.7: Detection rate of Sybil nodes under different traffic density

As illustrated in Fig. 3.7, we choose Receiver Operating Characteristic (ROC) curve to represent the detection rate of our method, which is considered as a comprehensive and visually attractive way to summarise the accuracy of detection. Each point in ROC space correspond two metrics: True Positive Rate (TPR) which means the well detected ratio, and False Positive Rate (FPR) which represents the percentage of benign nodes that are reported as malicious.

It can be observed from the figure that our detection method can reach a high TPR with a low FPR. In all cases, more than 90% of Sybil nodes can be detected by our detection method. Otherwise, the detection rate has a positive correlation with traffic lane traffic density, which corresponds well our hypothesis. Because high traffic density can limit average headway distance between vehicles, which makes their driving pattern more similar. On the other hand, limited headway distance also limits the choices for malicious node when it generate virtual nodes.

3.3.2 Conclusion

In this paper, an efficient Sybil attack detection method based on vehicle driving pattern in urban scenario was introduced. This method was developed to detect the erraticness of Sybil nodes' driving pattern, which includes their lifetime and their unusual movement. In this method, vehicle

driving patterns are mainly represented using the eigenvalues of its DPM in this proposed detection method. The Mahalanobis distance is then used to measure the similarity of their driving patterns. Simulation results show that our detection method can reach a high detection rate with a low error rate. In all cases, more than 90% of Sybil nodes can be detected by the proposed detection method, and the error rate can be controlled under 10% at the same time.

3.4 Support Vector Machine (SVM) Based Sybil Attack Detection

In this section, we present a method that mainly represent vehicles' driving patterns by using eigenvalues of their driving pattern matrix and the classification procedure based on several SVM classifiers. A brief description of the proposed algorithm is illustrated in Fig. 3.8.

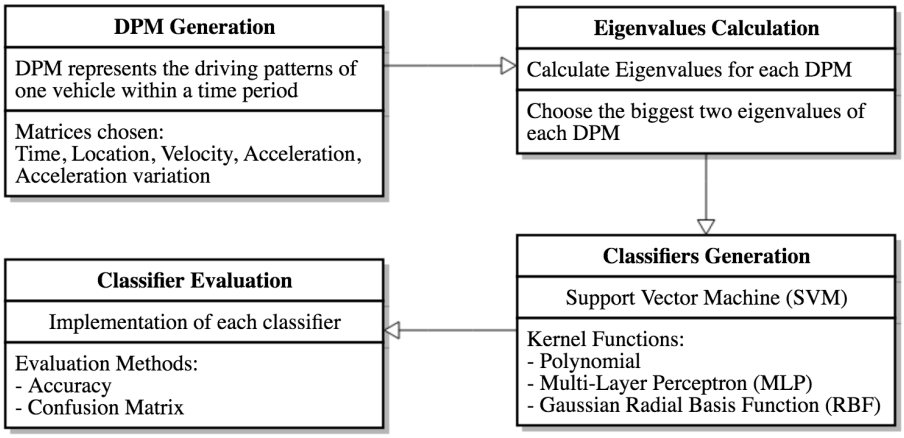


Figure 3.8: Diagram of the proposed algorithm

3.4.1 Classification

Due to the reality that the eigenvalues decrease quickly, we can sort them in decrease order and take the top k ($k < 5$) instead of all 5 eigenvalues to represent the characteristics of the original DPM in classification. In this work, the two biggest eigenvalues are chosen. We note $\lambda_{i1}^{ma} = max\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\}$ and $\lambda_{i2}^{ma} = max\{\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\} - \{\lambda_{i1}^{ma}\}\}$ the two biggest eigenvalues values of each DPM.

We note $\mathcal{V} = \{\vec{v}_i, u_i\}_{1 \leq i \leq n}$ a set with n vehicles where vector $\vec{v}_i = (\lambda_{i1}^{ma}, \lambda_{i2}^{ma})$ represents the driving patterns of vehicle v_i, and $u_i \in \{-1, 1\}$ gives the label of $\vec{v}_i$. The label -1 means $\vec{v}_i$ is malicious and otherwise $\vec{v}_i$ is benign.

Normally if data is linear, a separating hyperplane can be used to divide the data. However, due to the erraticness of virtual nodes' movement, it is often the case that the data is far from linear and the datasets are not linearly separable. Under this circumstance, the dataset $\mathcal{V}$ needs to

be projected into feature space where the new dataset $\mathcal{V}'$ is linearly separable:

$$\begin{aligned} \psi : \mathcal{V} &\longrightarrow \mathcal{V}' \\ \vec{v} &\longrightarrow \psi(\vec{v}) = \begin{pmatrix} \psi_1(\vec{v}) \\ \psi_2(\vec{v}) \end{pmatrix} \end{aligned} \tag{3.9}$$

Then, we have $\mathcal{V}' = \{(\psi(\vec{v}), u_i)\}_{1 \leq i \leq n}$ with $u_i \in \{-1, 1\}$.

We then define a kernel function k:

$$k(\vec{v}_i, \vec{v}_j) = < \psi(\vec{v}_i), \psi(\vec{v}_j) > \tag{3.10}$$

Where $< \cdot, \cdot >$ is the scalar product between two vectors. And the objective is to find out the suitable classifier:

$$f_{\vec{w},b}(\vec{v}) = k(\vec{w}, \vec{v}) + b \tag{3.11}$$

This classifier depends on parameters $\vec{w}, b, e$ to minimize:

$$\frac{1}{2} k(\vec{w}, \vec{w}) + c \sum_l e_l \tag{3.12}$$

Under the following constraints:

$$\begin{cases} u_l[k(\vec{w}, \vec{v}_l) + b] \geqslant 1 - e_l, \forall (\vec{v}_l, u_l) \in V \\ e_l \geqslant 0, \forall l \end{cases} \tag{3.13}$$

Its Lagrange multiplier is:

$$\begin{aligned} L(\vec{w}, b, \vec{e}, \vec{\alpha}, \vec{y}) &= \frac{1}{2} k(\vec{w}, \vec{w}) + c \sum_l e_l - \sum_l \alpha_l [y_l (k(\vec{w}, \vec{v}_l) + b) + e_l - 1] - \sum_l y_l e_l \\ &= \frac{1}{2} k(\vec{w}, \vec{w}) + \sum_l e_l (c - \alpha_l - y_l) + \sum_l \alpha_l - \sum_l \alpha_l y_l (k(\vec{w}, \vec{v}_l) + b) \end{aligned} \tag{3.14}$$

Where the Karush Kuhn Tucker conditions must be satisfied:

$$\begin{cases} \forall l; \alpha_l, y_l, e_l \geqslant 0 \\ \forall l; n_l[k(\vec{w}, \vec{v}_l) + b)] \geqslant 1 - e_l \\ \forall l; y_l \cdot e_l = 0 \\ \forall l; \alpha_l [y_l (k(\vec{w}, \vec{v}_l) + b) + e_l - 1] = 0 \end{cases} \tag{3.15}$$

Under this circumstance, the question above can be considered as to maximize the following function:

$$\sum_{l} \alpha_l - \frac{1}{2} \sum_{k} \sum_{l} \alpha_k \alpha_l u_k u_l k(\vec{v}_k, \vec{v}_l) \tag{3.16}$$

Where $\forall l, \sum_{l} \alpha_l [u_l = 0$ and $0 \leq \alpha_l \leq c$.

Therefore, the classifier can be considered as:

$$f(\vec{v}) = \sum_{l} \alpha_l u_l f_{\vec{w},b}(\vec{v}) = k(\vec{v}_l, \vec{v}) + b \tag{3.17}$$

In this work, three kernel functions are taken into consideration: Polynomial, Gaussian Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP).

- Polynomial: $k(\vec{v}_i, \vec{v}_j) = (< \vec{v}_i, \vec{v}_j > + h)^d$ where h is a constant value.
- RBF: $k(\vec{v}_i, \vec{v}_j) = exp(-\frac{\|\vec{v}_i - \vec{v}_j\|^2}{2\sigma^2})$
- MLP: $k(\vec{v}_i, \vec{v}_j) = tanh(\rho < \vec{v}_i, \vec{v}_j > + \varrho)$

Their performance will be evaluated in the next section based on the simulation results.

3.4.2 Experimental Results

In the following simulations, SUMO simulator is chosen to simulate traffic flows in urban scenario, where parameters are set based on real-time urban traffic. The detection method is implemented on the RSU, and vehicles are demanded to periodically communicate their driving patterns with the RSU via CAM message. we choose a window size of 60 seconds and vehicles within 10 second before and after the target vehicle. More details are presented in Table 3.2:

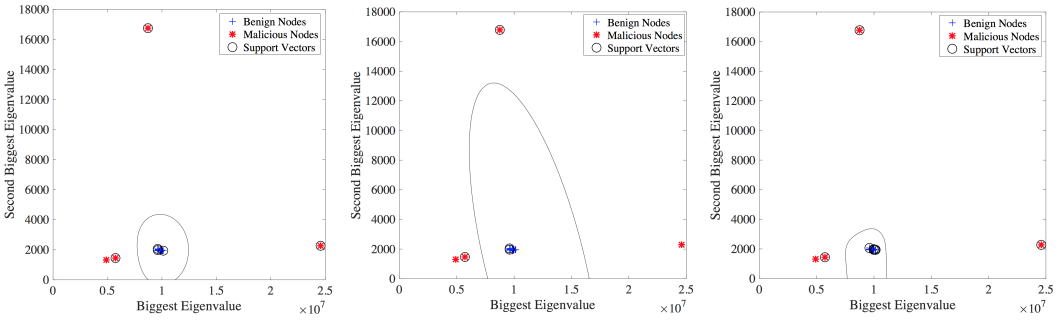
Table 3.2: Parameters Used in Simulations

Parameter	Value
Simulation Scenario	Urban Scenario
Simulation Time	300s
Window Distance	1 km
Street Width	2 Lanes
Vehicle Velocity	40 - 60 km/h
Number of Vehicles Simulated	350 - 800
Communication Range	300m

Virtual Nodes' Characteristics As illustrated in Fig. 3.9, the driving patterns of benign nodes have obvious similarity and the driving patterns of Sybil nodes show erraticness. This result is caused by two reasons:

- The variation of virtual nodes' lifetime: We measured the lifetime for more than 200 virtual nodes, the CDF is illustrated in Fig. 3.5. We can find out that more than 70% of the virtual nodes do not have the same length of lifetime as the benign ones (60 seconds).
- The erraticness of virtual nodes' movement: Virtual nodes should not be set to the positions that are captured by the benign ones.

Both these issues cannot be figured out by Sybil nodes, because they cannot control the components of other road users. They can only adjust their own strategies and make their driving patterns as reasonable as possible. Our detection system is directed against this weak point of Sybil nodes, making it possible to separate virtual nodes from benign ones.



(a) RBF Sigma=0.5 (b) Polynomial d=2 (c) MLP

Figure 3.9: Training group classification results with all three kernel functions

Classification Accuracy We launched 15 times Sybil attacks under 3 different traffic densities. The first group is chosen as the training group and the other 14 groups are testing groups. As illustrated in Table 3.3, the classification accuracy in all testing groups, and Fig. 3.9 shows the training group classification results. In this work, all three kernel functions are implemented with different parameters. Generally speaking, due to the reality that benign nodes show strong similarity in their driving patterns, the classifiers which cover less surface reach higher accuracies.

In more detail, as illustrated in Fig. 3.10, the performance of RBF classifier with different *sigma* values, which can be noticed is that with the increase of the *sigma* value, from 0.1 to 0.9, the classifier reaches its best accuracy at the point $sigma = 0.3$, then its performance decreases with the increase of the *sigma* value. In SVM, a very small value of *sigma* means a large margin is necessary which may leads to misclassified training group. On the other hand, with a very large

Table 3.3: Testing Groups Classification Accuracy

Traffic Density	26-30	26-30	26-30	26-30	30-34	30-34	30-34	30-34	30-34	>34	>34	>34	>34	>34
rbf σ = 0.1	0.917	0.956	0.945	1	0.909	0.941	0.961	0.933	0.963	0.961	0.912	1	0.965	1
rbf σ = 0.3	0.917	1	0.945	1	1	0.971	1	1	0.926	1	0.912	1	1	1
rbf σ = 0.5	0.958	0.870	0.946	1	1	0.971	1	1	0.889	1	0.882	1	1	1
rbf σ = 0.7	0.958	0.870	0.919	1	1	0.971	1	1	0.889	0.962	0.882	1	1	1
rbf σ = 0.9	0.958	0.782	0.919	1	1	0.971	1	0.933	0.889	0.884	0.882	0.971	1	0.972
Polynomial d = 2	0.958	0.783	0.919	1	1	0.912	0.923	0.933	0.889	0.846	0.882	0.971	1	0.972
Polynomial d = 4	0.958	0.782	0.892	0.971	0.939	0.853	0.923	0.933	0.889	0.846	0.882	0.943	0.966	0.944
mlp [-1 1]	0.958	0.826	0.973	1	1	0.971	1	1	0.889	1	0.882	1	1	1
mlp [-2 2]	0.958	0.826	1	1	1	0.971	1	1	0.889	1	0.882	1	1	1

sigma value, the training group can be well classified, however, the classifier would not reach high classification accuracy in testing groups.

As shown in Fig. 3.11, testing groups classification accuracy with different traffic density. As we expected, when the traffic density is high (> 34 vehicles per kilometre), the classifiers reach higher classification accuracy. That because when the traffic density is high, the average distance between two adjacent vehicles are small, their driving patterns would show stronger similarity compare to the low traffic density scenarios. Otherwise, when the distance between vehicles is small, the erraticness of virtual nodes' driving patterns would also be obvious. Because they should not be set to the positions that are captured by the benign ones.

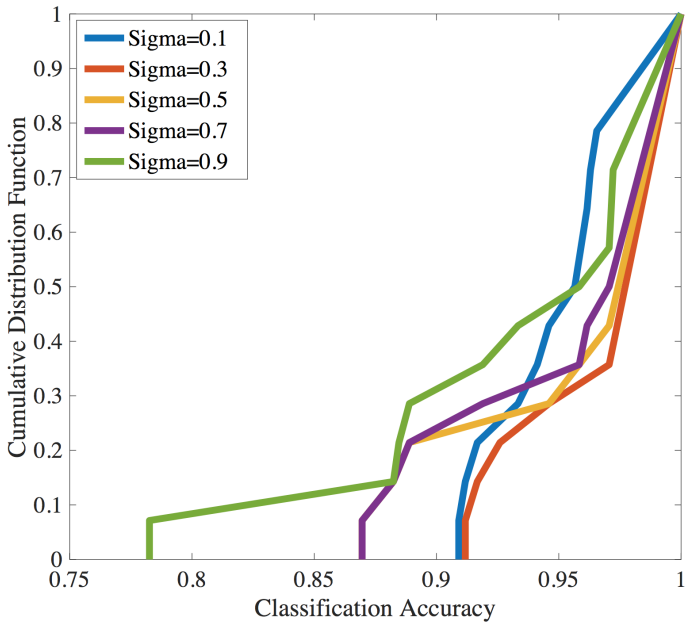


Figure 3.10: RBF Classifier with *sigma* values from 0.1 to 0.9.

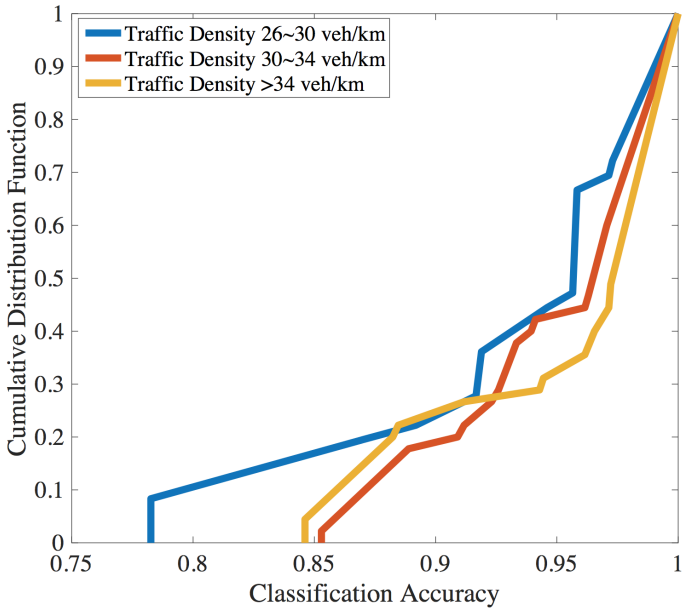


Figure 3.11: Testing groups classification accuracy in three different traffic densities.

Table 3.4: Confusion Matrix

	TP	TN	FP	FN	TPR	FPR	FNR
rbf $\sigma = 0.3$	325	94	2	8	97.6%	2%	2.4%
Poly $d = 2$	326	74	1	28	92.1%	1.3%	7.9 %
MLP [-2 2]	326	90	1	12	96.5%	1.1%	3.6%

3.4.3 Classification Error Rate

In this work, the confusion matrix is also chosen to evaluate the performance of classifiers, which is considered as a comprehensive and visually attractive way to summarise the error rate. Normally a confusion reports the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). The error rate of a classifier corresponds three metrics: True Positive Rate (TPR) which means the well detected ratio, False Positive Rate (FPR) which represents the percentage of benign nodes that are reported as malicious and False Negative Rate (FNR) which means the percentage of malicious nodes are reported as benign.

As shown in Table 3.4 the summary of these three classifiers' confusion matrices. Generally speaking, these classifiers all reached high detection rate with low error rate. When we go into more detail, which can be noticed is that the FNR of polynomial classifier is about 8%, but the FPR is low. As we have seen above in Fig. 3.9, the polynomial classifiers covered more surface than the other two classifiers. In this case, benign nodes can be hardly reported as malicious, but malicious nodes could be reported as benign.

3.4.4 Conclusion

In this paper, a vehicle driving pattern similarity measurement method in near capacity traffic scenario was introduced. This method was developed to measure the benign vehicles' similarity in driving patterns and detect the variation between benign vehicles and Sybil nodes in their driving patterns. This variation can be reflected in their Driving Pattern Matrices. Vehicle driving patterns are mainly represented using the eigenvalues of its DPM in this proposed detection method. The SVM methods are then used to classify the vehicles and distinguish the virtual nodes from benign ones. Simulation results show that in all events, the majority benign vehicles have similar driving patterns, and the Sybil nodes show erraticness in their driving patterns. All the three kernel functions are reached high detection rate with low error rate.

3.5 k-Nearest Neighbours (kNN) Based Sybil Attack Detection

In this section, we present a method that mainly represent vehicles' driving patterns by using eigenvalues of their driving pattern matrix and the classification procedure based on k-Nearest Neighbours classifiers.

3.5.1 Classification

Due to the reality that the eigenvalues decrease quickly, we can sort them in decrease order and take the top k ($k < 5$) instead of all 5 eigenvalues to represent the characteristics of the original DPM in classification. In this work, the two biggest eigenvalues are chosen. We note $\lambda_{i1}^{ma} = max\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\}$ and $\lambda_{i2}^{ma} = max\{\{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\} - \{\lambda_{i1}^{ma}\}\}$ the two biggest eigenvalues values of each DPM. Under this circumstance, the driving patterns of one vehicle within a time period can be represented as a point on a two-dimensional surface. The two axes are the two biggest eigenvalues of its DPM.

We note two vectors $\vec{v_i}$, $\vec{v_j}$ where $\vec{v_i} = (\lambda_{i1}, \lambda_{i2})$, $\vec{v_j} = (\lambda_{j,1}, \lambda_{j,2})$ represent the driving patterns of two vehicles v_i and v_j. The difference between their driving patterns is defined by the Minkowski distance between these two vectors:

$$d(\vec{v_i}, \vec{v_j}) = [\sum_{s=1}^{2} |\vec{v}_{is} - \vec{v}_{js}|^q]^{\frac{1}{q}} = [|\lambda_{i1} - \lambda_{j1}|^q + |\lambda_{i2} - \lambda_{j2}|^q]^{\frac{1}{q}} \tag{3.18}$$

Minkowski distance is typically used with q being 1 or 2, which are known as the Manhattan distance and Euclidean distance.

In order to classify an arriving vehicle v_e, we note a training group with n vehicles as $K = \{(\vec{v_i}, y_i)i = 1, 2, ..., n\}$ where $\vec{v_i} = (\lambda_{i1}, \lambda_{i2})$ and $y_i \in \{-1, 1\}$ gives the label of $\vec{v_i}$. The label -1 means $\vec{v_i}$ is malicious and otherwise $\vec{v_i}$ is benign.

The distance between the arriving vehicle v_e and its nearest neighbour can be represented as $argmin\{d(\vec{v}_e, \vec{v}_i), 1 \leq i \leq n\}$ And the label of v_e depends on the labels of its k nearest neighbours ($k = 1, 3, 5, ...$).

3.5.2 Optimization

kNN is an non parametric lazy learning algorithm. Therefore, all training data is needed during the testing phase. Under this circumstance, the increasing time complexity becomes an issue. The runtime of a traditional kNN algorithm can be represented as $O(nd + kn)$ when k is fixed. If the classifier learns from the classified data (add them to the training group), the runtime will increase linearly.

In this paper, we propose two methods to limit the time complexity of kNN algorithm. One is using memory less method: the system does not add the classified object into the training set, thus, the time complexity does not increase. The other method is based on the benign users' similarity in their driving patterns. A vector $\vec{v}_m = (\lambda_{m1}, \lambda_{m2})$ can be used to represent all n benign vehicles in training group, where $\lambda_{m1} = \sum_{i=1}^{n} \lambda_{i1}$, $\lambda_{m2} = \sum_{i=1}^{n} \lambda_{i2}$. After the $(n + 1)$th vehicle is classified as benign, the vector $\vec{v}_m$ can be updated by calculating the weighted arithmetic mean values of λ_{m1} and λ_{m2}, where $\lambda'_{m1} = (n\lambda_{m1} + \lambda_{(n+1)1})/(n + 1)$ and $\lambda'_{m2} = (n\lambda_{m2} + \lambda_{(n+1)2})/(n + 1)$. Otherwise, this vehicle will be added into the training set directly.

The effectiveness of these two methods will be evaluated in next section based on simulation results.

3.5.3 Experimental Results

In the following simulations, we still consider the Near-capacity conditions. Therefore, we choose the vehicle density between 26 and 42 vehicles per kilometre in our simulations. For the detection method, we choose a window size of 60 seconds and vehicles within 10 second before and after the target vehicle. The detection algorithm is implemented in Matlab. More details are presented in Table 3.5:

3.5.4 Number of Neighbours

As illustrated in Fig. 3.12, un example of vehicle driving patterns description based on the two biggest eigenvalues of its DPM. It can be observed that the driving patterns of benign vehicles show strong similarity. However, the driving patterns of malicious nodes are erratic. Under this circumstance, it is possible to separate malicious nodes from the benign ones by using classification method.

Table 3.5: Parameters Used in Simulations

Parameter	Value
Simulation Scenario	Urban Scenario
Simulation Time	300s
Window Distance	1 km
Street Width	2 Lanes
Vehicle Velocity	40 - 60 km/h
Number of Vehicles Simulated	350 - 800
Communication Range	300m
Beacon Frequency	1 Hz

To test the kNN method with different number of neighbours, the simple Memory Less method is implemented. In this case, the first scenario is chosen as the training set, the rest 14 scenarios are used as testing sets. And the k value is set as $k \in \{1, 3, 5\}$. As illustrated Fig. 3.13 the classification result. Generally speaking, in all these 14 groups of test, the nearest neighbour method ($k = 1$) reached the best performance, the mean detection rate is around 80%. Therefore, the nearest neighbour method will be implemented in following simulations.

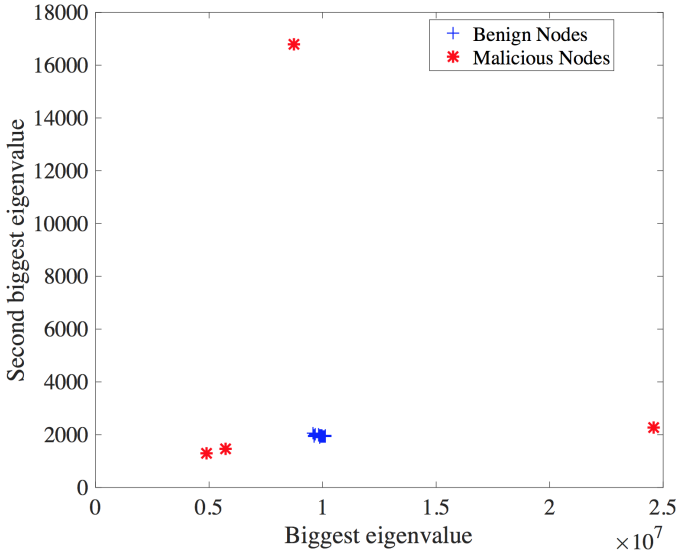


Figure 3.12: The first two biggest eigenvalues of DPM can be used to mainly represent the driving pattern of one vehicle

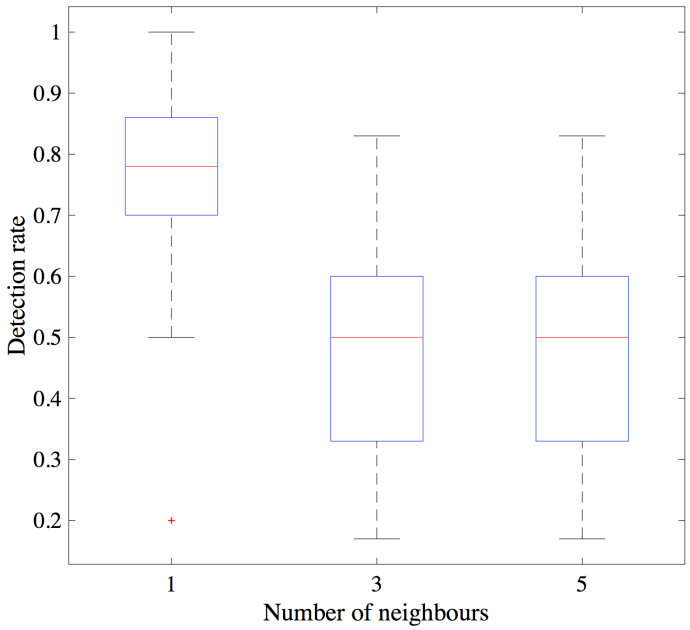


Figure 3.13: Memory Less method detection ratio with different number of k values

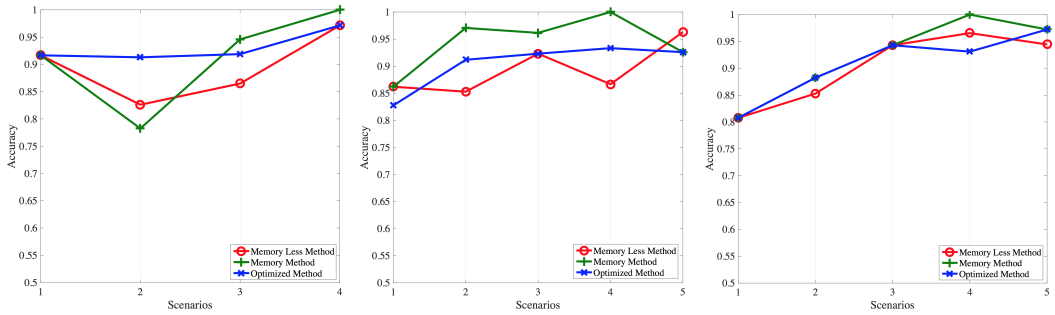
3.5.5 Classification Accuracy

We launched 15 times Sybil attacks under 3 different traffic densities. For the Memory Less method, the first group is chosen as the training group and the other 14 groups are testing groups. For both Memory method and our optimized method, scenarios with the same traffic density are measured together. Both methods learn from the previous scenarios the driving patterns of different vehicles in order to improve their classification accuracy in following scenarios. As illustrated in Fig. 3.14, the results of classification accuracy evaluation.

Generally speaking, all these three methods can reach a good classification accuracy, 80% respectively. In more detail, the performance of Memory Less method is not stable compare to the other two methods. The Memory method performs the best in classification accuracy, in several scenarios, it reached 100% classification accuracy. And for our optimized method, its performance is more stable than the Memory Less method. Its weak point could be in error control, which will be evaluated in next subsection.

3.5.6 Classification Error Rate

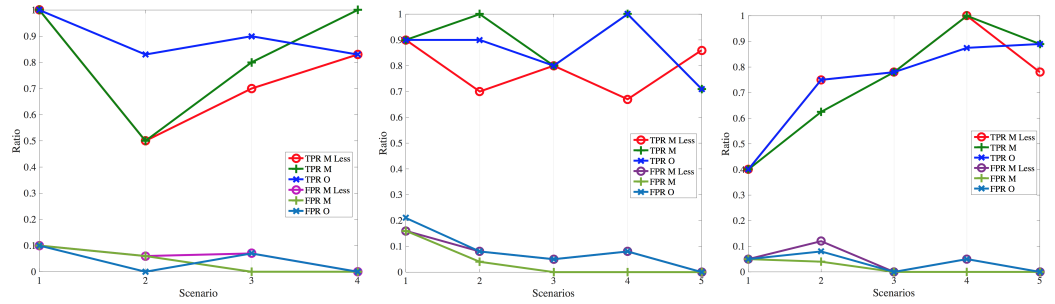
In this work, the confusion matrix is also chosen to evaluate the performance of classifiers, which is considered as a comprehensive and visually attractive way to summarise the error rate. Normally a confusion reports the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). The error rate of a classifier corresponds two metrics: True Positive Rate (TPR) which means the well detected ratio and False Positive Rate (FPR) which represents



(a) Density 26-30 vehicles/km (b) Density 30-34 vehicles/km (c) Density >34 vehicles/km

Figure 3.14: Testing groups classification accuracy in three different traffic densities.

the percentage of benign nodes that are reported as malicious. The pair of TPR and FPR gives a partial picture of a classifier's performance, where TPR represents the effectiveness of a classifier and FPR reflects its performance in error control.



(a) Density 26-30 vehicles/km (b) Density 30-34 vehicles/km (c) Density >34 vehicles/km

Figure 3.15: Testing groups classification error rate evaluated by using TPR and FPR

As illustrated in Fig. 3.15, it can be observed that the detection performance of Memory Less method is not stable, because the decisions are made only based on the limited number of training data. It can also be observed that both the Memory method and our optimized method perform well in detection as well as in error control. Especially the Memory method's error control performance, which can reach 0 error rate in some scenarios. For our optimized method, the potential weak point could be its performance in error control, because in this method, a set of data are replaced by their mean value. However, based on the simulation results, we can find out that our optimized method performed also better than the Memory Less methods in error control.

3.5.7 Conclusion

In this paper, a vehicle driving pattern similarity measurement method in near capacity traffic scenario was introduced. This method was developed to measure the benign vehicles' similarity in driving patterns, and detect the variation between benign vehicles and Sybil nodes in their driving

patterns. This variation can be reflected in their Driving Pattern Matrices. Vehicle driving patterns are mainly represented using the eigenvalues of its DPM in this proposed detection method. The kNN classification methods are then used to classify the vehicles, distinguish the virtual nodes from benign ones. However, one major drawback of kNN methods is its high runtime complexity. Two different methods are proposed in this paper to address this issue in kNN methods. Simulation results show that our proposed method can reach a high detection ratio with a good performance in error control.

Chapter 4

Cooperative Relaying for Control Channel Jamming in Vehicular Networks

In this chapter, we focus on the design of countermeasure for the control channel jamming issue in vehicular networks, which is of vital importance to the safety of I2V communications. There are salient features for this problem, making it nontrivial to address with existing anti-jamming techniques. Firstly, according to the IEEE 802.11p protocol, only one control channel is available for the transmissions of safety related messages in vehicular networks, and thus it is impossible for the RSU to switch to other channels if the control channel is jammed. Secondly, although multi-antenna techniques have been deemed as an effective anti-jamming solution, but the specific antennas have not been widely used on the vehicles due to the mobility and power constraint of vehicles. Last but not the least, the mobility of vehicles is limited by road space and road traffic density, and thus it is difficult for the vehicles to escape out of the jamming area if the attack is launched by a neighbouring vehicle.

To this end, we propose to adopt the cooperative relaying techniques to address the control channel jamming problem in vehicular networks, which is based on the idea that the vehicles outside of the jamming area can serve as relays to help forward the control channel signal to the victim vehicles through other the jamming-free service channels. In this way, a virtual multi-antenna system can be formed by multiple relays nodes to cooperatively serve the victim vehicles, and thus the transmission reliability can be effectively improved by exploiting the spatial diversity of these relay nodes. We analyse the performance of this cooperative anti-jamming relaying scheme under different jamming scenarios, and design a relay selection algorithm to maximize the performance of the worst victim vehicles. Extensive simulation results are provided to demonstrate the performance of the proposed scheme.

In summary, the main contributions of this paper are as follows:

- A cooperative anti-jamming relaying scheme is proposed for the control channel jamming issue in vehicular networks, which takes advantage of the spatial diversity provided by the vehicles outside the jamming area to improve the transmission reliability of the victim vehicles.

- Theoretical model are developed to characterize the outage probability of the cooperative relaying scheme under different jamming scenarios, which accounts for the large scale path loss and small-scale channel fading between vehicles.

- A max-min relay selection problem is formulated, which attempts to maximize the signal-to-noise ratio (SNR) of the worst victim vehicle under the relay number constraint. A heuristic relay selection algorithm is designed by exploiting the special structure of the problem, which is shown to be close to the performance of the optimal solution through simulations.

The rest of this chapter is organized as follows. In Section 4.1, we introduce the system model and discuss different jamming scenarios based on the location of the jammer. Theoretical models are provided in Section 4.2 to analyse the outage probability of the cooperative relaying scheme, and the problem of relay selection is studied in Section 4.3. In Section 4.4, we provide extensive simulation results to validate the theoretical results and evaluate the performance of the proposed schemes. Finally, conclusions are drawn for this paper in Section 4.5.

4.1 Network Models

We consider a vehicular network consisting of RSUs and vehicles moving along a typical multi-lane road in the urban area. We focus on a segment covered by a single RSU, which is located at the origin with a coverage area of $[-r, r]$. We assume that a jammer v_m is located at $L(v_m) = m \in (0, r)$[1]. Based on the locations of the RSU and the jammer, vehicles within the coverage area of the RSU can be classified into two categories: one are the victim vehicles located within the jamming area of the jammer (Dark areas in Fig. 4.1 and Fig. 4.2), which are denoted as $\mathbb{V}$, the others are the vehicles within the coverage area of the RSU but outside of the jamming area (Light-coloured areas in Fig. 4.1 and Fig. 4.2), which are denoted as $\mathbb{R}$ and can be selected as relay nodes for the victim vehicles.

We consider two different kinds of jammers in this paper. One is the road side jammer, which can be considered as a fake RSU who launches jamming attack towards the control channel, and the victims are the vehicles passing through the jamming area, whereby a portion of the management messages from the RSU to vehicles will be blocked. The other is the moving jammer that

[1]The case of $m \in (-r, 0)$ is symmetric and can be analysed similarly

moves along with the road traffic, whose objective is to prevent the neighboring vehicles from receiving messages from the RSUs along their travel routes. The symbols used in this paper are listed in Table 4.1.

Table 4.1: List of Notations

Symbol	Description
P_v	Transmit power of vehicle
P_J	Transmit power of jammer
h_{ij}	Channel vector between v_i and v_j
α	Path loss exponent
r_{ij}	Received signal at vehicle v_j from vehicle v_i
d_{ij}	Euclidean distance between vehicle v_j and vehicle v_i
Γ_j	Received SNR at vehicle v_j
γ_{ij}	Normalized path loss between vehicle v_j and vehicle v_i
n_o	Received noise signal
σ^2	Received noise power
η	Received signal SINR threshold
θ	Received signal SNR threshold

4.1.1 Signal Model

Given the location $L(v_m)$ of the jammer (either road side jammer or moving jammer), the ranges for $\mathbb{V}$ and $\mathbb{R}$ can be determined as follows. Let P_R and P_J denote the transmit power of the RSU and the jammer respectively, then a vehicle v_x at location x is assumed to be a victim if its received signal-to-interference-plus-noise ratio (SINR) is smaller than a prescribed threshold η, that is:

$$SINR = \frac{P_R x^{-\alpha}}{P_J ||x - m||^{-\alpha} + \sigma^2} < \eta, \tag{4.1}$$

where α is the path loss exponent, and σ^2 is the noise power. Note that the noise can be ignored comparing with the strong interference signal from the jammer, and thus the victim range for $\mathbb{V}$ and relay range for $\mathbb{R}$ can be obtained as follows:

$$\mathbb{V} = \{v_x | \frac{m}{1+a} < x < \frac{m}{1-a}\} \tag{4.2}$$

$$\mathbb{R} = \{v_x | -r \leq x \leq \frac{m}{1+a} \cup \frac{m}{1-a} \leq x \leq r\} \tag{4.3}$$

where $a = (\frac{\eta P_J}{P_R})^{1/\alpha}$.

Proof: In this analysis, we assume that the received signal at a vehicle located at x can be decoded if and only the SINR exceeds a prescribed threshold value η, that is,

$$SINR = \frac{P_R x^{-\alpha}}{P_J |x-m|^{-\alpha} + \sigma^2} \geq \eta.$$

By ignoring the noise, we have:

$$\frac{|x-m|}{x} \geq (\frac{\eta P_J}{P_R})^{1/\alpha}.$$

Therefore, the benign vehicles can be classified into two categories: Relay candidate (R) or Victim (V) based on their relative positions to the jammer. We assume that a vehicle cannot be located at the same position as the jammer, i.e., $m \neq x$. For benign vehicles located at the left-hand side of the jammer, we have $-r \leq x < m$, so

$$\frac{x-m}{x} \geq (\frac{\eta P_J}{P_R})^{1/\alpha},$$

which yields

$$x \geq \frac{m}{1-(\frac{\eta P_J}{P_R})^{1/\alpha}}.$$

For benign vehicles located at the right-hand side of the jammer, we have $m < x \leq r$, so

$$\frac{m-x}{x} \geq (\frac{\eta P_J}{P_R})^{1/\alpha},$$

which yields

$$x \leq \frac{m}{1+(\frac{\eta P_J}{P_R})^{1/\alpha}}.$$

4.1.2 Locations of Vehicles

The locations of vehicles can be modelled as a homogeneous Poisson Point Process (PPP) with intensity λ[40][41], and thus the number of vehicles within a certain range follows a Poisson distribution with rate λ. Let us define $\gamma_{ij} = d_{ij}^{\alpha}/P_v$ as the normalized path loss between each pair of relay node $v_i \in \mathbb{R}$ and victim node $v_j \in \mathbb{V}$, which is an non-homogeneous PPP with intensity $\lambda(x) = \lambda \frac{2}{\alpha} P_v^{2/\alpha} x^{2/\alpha - 1}$ according to the mapping theory [42].

Based on the location of the jammer, we have two possible jamming scenarios:

4.1.2.1 Scenario 1: $r(1-a) < m < r$

In this scenario, as illustrated in Fig. 4.1, the jamming range is $(\frac{m}{1+a}, r)$, and the relaying range is $[-r, \frac{m}{1+a}]$. Therefore, the density functions for $\mathbb{R}$ and $\mathbb{V}$ can be obtained as follows:

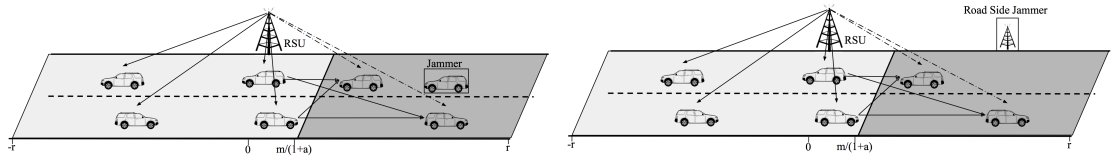
$$\Lambda_R(-r, \frac{m}{1+a}) = \int_{-r}^{m/(1+a)} \lambda(x)dx = \lambda P_v^{\frac{2}{\alpha}}((\frac{m}{1+a})^{\frac{2}{\alpha}} - r^{\frac{2}{\alpha}}), \tag{4.4}$$

$$\Lambda_V(\frac{m}{1+a}, r) = \int_{m/(1+a)}^{r} \lambda(x)dx = \lambda P_v^{\frac{2}{\alpha}}(r^{\frac{2}{\alpha}} - (\frac{m}{1+a})^{\frac{2}{\alpha}}). \tag{4.5}$$

The probabilities of having k relay nodes and l victims are given by:

$$\mathbb{P}[N_R(-r, \frac{m}{1+a}) = k] = \frac{[\Lambda_R(-r, \frac{m}{1+a})]^k}{k!} e^{-\Lambda_R(-r, \frac{m}{1+a})}, \tag{4.6}$$

$$\mathbb{P}[N_V(\frac{m}{1+a}, r) = l] = \frac{[\Lambda_V(\frac{m}{1+a}, r)]^l}{l!} e^{-\Lambda_V(\frac{m}{1+a}, r)}. \tag{4.7}$$



(a) Moving jammer (b) Road side jammer

Figure 4.1: Scenario 1: the jammer is located at $r(1-a) < m < r$.

4.1.2.2 Scenario 2: $0 < m < r(1-a)$

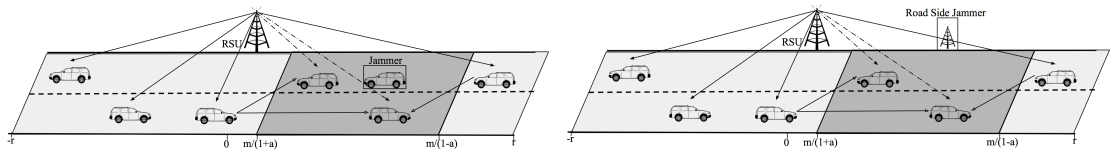
In this scenario, as illustrated in Fig. 4.2, the jamming range is $(\frac{m}{1+a}, \frac{m}{1-a})$, and the relay nodes are separately located within $[-r, \frac{m}{1+a}]$ and $[\frac{m}{1-a}, r]$. Therefore, the density functions for $\mathbb{R}$ and $\mathbb{V}$ can be obtained as follows:

$$\Lambda_R(-r, \frac{m}{1+a}) = \int_{-r}^{m/(1+a)} \lambda(x)dx = \lambda P_v^{\frac{2}{\alpha}}((\frac{m}{1+a})^{\frac{2}{\alpha}} - r^{\frac{2}{\alpha}}) \tag{4.8}$$

$$\Lambda_R(\frac{m}{1-a}, r) = \int_{m/(1-a)}^{r} \lambda(x)dx = \lambda P_v^{\frac{2}{\alpha}}(r^{\frac{2}{\alpha}} - (\frac{m}{1-a})^{\frac{2}{\alpha}}) \tag{4.9}$$

$$\Lambda_V(\frac{m}{1+a}, \frac{m}{1-a}) = \int_{m/(1+a)}^{m/(1-a)} \lambda(x)dx = \lambda P_v^{\frac{2}{\alpha}}((\frac{m}{1-a})^{\frac{2}{\alpha}} - (\frac{m}{1+a})^{\frac{2}{\alpha}}). \tag{4.10}$$

The probabilities of having k candidate relay nodes and l victims can be obtained similar to (4.6) and (4.7) respectively.



(a) Moving jammer

(b) Road side jammer

Figure 4.2: Scenario 2: the jammer is located at $0 < m < r(1-a)$

4.2 Outage Probability Analysis

As discussed in previous section, the victim vehicles within the jamming area cannot decode the control channel messages from the RSU, and they may not establish V2V communication links directly with other vehicles outside the jamming area if the jamming range is larger than the transmission range of the vehicles. As a solution, we propose to adopt the cooperative relaying techniques such that the vehicles outside the jamming area can serve as the relay nodes to forward the control channel signal to the victim vehicles via another jamming-free service channel.

Specifically, we assume the decode-and-forward (DF) strategy is adopted by the relay nodes to forward the received message to the victim vehicles. Let s and t denote the transmitted signal and the received signal respectively. Then, for each victim vehicle $v_j \in \mathbb{V}$, the received signal from a relay vehicle $v_i \in \mathbb{R}$ can be given by:

$$t_{ij} = \sqrt{P_v} h_{ij} \|d_{ij}\|^{-\alpha/2} s + n_0, \tag{4.11}$$

where P_v denotes the transmit power of vehicle, $\|d_{ij}\|^{-\alpha/2}$ and h_{ij} are the path loss and multi-path fading between nodes v_i and v_j. n_0 is the noise at node v_j, which is assumed to be zero-mean complex Gaussian random variable with unit variance σ^2.

The received signals from all relays can be decoded with the maximum ratio combining (MRC) scheme or the selection combining (SC) schemes, and the achieved signal-to-noise ratios (SNRs) at node v_j can be given by

$$\text{MRC:} \quad \Gamma_j^m = \frac{\sum\limits_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}}{\sigma^2} \tag{4.12}$$

and

$$\text{SC:} \quad \Gamma_j^s = \frac{\max\limits_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}}{\sigma^2} \tag{4.13}$$

respectively.

Assuming the channels between relays and victim follow Rayleigh fading model, and are independent between different relays, then $\sum_{v_i \in \mathbb{R}} P_v |h_{ij}|^2 \|d_{ij}\|^{-\alpha}$ is exponentially distributed with

mean $\sum_{v_i\in\mathbb{R}} P_v\|d_{ij}\|^{-\alpha}$ since h_{ij}s are mutually independent, and thus we have $\sum_{v_i\in\mathbb{R}} P_v\|d_{ij}\|^{-\alpha} = 1/\sum_{v_i\in\mathbb{R}} \gamma_{ij}^{-1}$, where $\gamma_{ij} = \|d_{ij}\|^{\alpha}/P_v$ is the normalized path loss as defined in previous section.

The performance of this cooperative relaying scheme can be characterized with the outage probability P_o. For the SC scheme, the outage probability of a victim $v_j \in \mathbb{V}$ is given by:

$$\begin{aligned} P_o^{SC} &= \mathbb{P}(\Gamma_j^s < \theta) \\ &= \mathbb{P}(\gamma_{ij} < \theta | \forall v_i \in \mathbb{R}) \\ &= \mathbb{E}_\gamma(\prod_{v_i\in\mathbb{R}}(1 - e^{-\theta\sigma^2\gamma_{ij}})) \\ &= \int_{\gamma_{ij}} \prod_{v_i\in\mathbb{R}}(1 - e^{-\theta\sigma^2\gamma_{ij}}) f(\gamma_{ij}) d\gamma_{ij}, \end{aligned} \tag{4.14}$$

where $f(\gamma_{ij})$ denotes the joint distribution for all γ_{ij}s, which is given by [43]:

$$f(\gamma_{ij},\ldots,\gamma_{nj}) = e^{-\lambda P_v^{2/\alpha}\gamma_{nj}^{2/\alpha}}(\lambda\frac{2}{\alpha}P_v^{2/\alpha})^n\prod_{i=1}^{n}\gamma_{ij}^{2/\alpha-1} \tag{4.15}$$

for $(v_1,\ldots,v_n) \in \mathbb{R}$.

Proof: The counting process of relay nodes is considered as a non-homogeneous Poisson process, the density is given by $\Lambda_n = \lambda(P\gamma_n)^{2/\alpha}$. Therefore, the probability with no relay nodes within the range of $(0,\gamma_1)$ is given by:

$$\mathbb{P}(0,\gamma_1) = e^{-\Lambda_1} = e^{-\lambda P^{2/\alpha}\gamma_1^{2/\alpha}}. \tag{4.16}$$

The probability with at least one relay within the range $(\gamma_1, \gamma_1+\Delta\gamma_1)$ is

$$\mathbb{P}(>0,(\gamma_1,\gamma_1+\Delta\gamma_1)) = 1 - e^{-\lambda P^{2/\alpha}(\gamma_1+\Delta\gamma_1-\gamma_1)^{2/\alpha}}. \tag{4.17}$$

Let $u = d\Lambda_1/dx = \lambda\frac{2}{\alpha}P^{2/\alpha}\gamma_1^{2/\alpha-1}$,

$$\lim_{\Delta\gamma_1\to 0} 1 - e^{-\lambda P^{2/\alpha}(\gamma_1+\Delta\gamma_1-\gamma_1)^{2/\alpha}} = -u = \lambda\frac{2}{\alpha}P^{2/\alpha}\gamma_1^{2/\alpha-1}. \tag{4.18}$$

Therefore, the probability that the nearest relay is located at γ_1 is

$$\mathbb{P}(\gamma_1) = e^{-\lambda P^{2/\alpha}\gamma_1^{2/\alpha}}\lambda\frac{2}{\alpha}P^{2/\alpha}\gamma_1^{2/\alpha-1}. \tag{4.19}$$

Thus, the probability that the second nearest relay is located at γ_2 given that the nearest relay is located at γ_1 can be obtained by

$$\mathbb{P}(0,(\gamma_1,\gamma_2)) = e^{-\lambda P^{2/\alpha}(\gamma_2-\gamma_1)^{2/\alpha}}, \tag{4.20}$$

$$\mathbb{P}(>0,(\gamma_2,\gamma_2+\Delta\gamma_2)) = \lambda\frac{2}{\alpha}P^{2/\alpha}\gamma_2^{2/\alpha-1}, \tag{4.21}$$

$$\mathbb{P}(\gamma_2|\gamma_1) = e^{-\lambda P^{\frac{2}{\alpha}}(\gamma_2-\gamma_1)^{\frac{2}{\alpha}}}\lambda\frac{2}{\alpha}P^{\frac{2}{\alpha}}\gamma_2^{\frac{2}{\alpha}-1}, \tag{4.22}$$

$$\begin{aligned} \mathbb{P}(\gamma_1,\gamma_2) &= \mathbb{P}(\gamma_1)\mathbb{P}(\gamma_2|\gamma_1) \\ &= e^{-\lambda P^{2/\alpha}(\gamma_2)^{\frac{2}{\alpha}}}(\lambda\frac{2}{\alpha}P^{\frac{2}{\alpha}})^2\prod_{i=1}^{2}\gamma_i^{\frac{2}{\alpha}-1}. \end{aligned} \tag{4.23}$$

For the third nearest relay, we have:

$$\mathbb{P}(0,(\gamma_2,\gamma_3)) = e^{-\lambda P^{2/\alpha}(\gamma_3-\gamma_2)^{2/\alpha}}, \tag{4.24}$$

$$\mathbb{P}(>0,(\gamma_3,\gamma_3+\Delta\gamma_3)) = \lambda\frac{2}{\alpha}P^{2/\alpha}\gamma_3^{2/\alpha-1}, \tag{4.25}$$

$$\mathbb{P}(\gamma_3|\gamma_1,\gamma_2) = e^{-\lambda P^{\frac{2}{\alpha}}(\gamma_3-\gamma_2)^{\frac{2}{\alpha}}}\lambda\frac{2}{\alpha}P^{\frac{2}{\alpha}}\gamma_3^{\frac{2}{\alpha}-1}, \tag{4.26}$$

$$\mathbb{P}(\gamma_1,\gamma_2,\gamma_3) = \mathbb{P}(\gamma_3|\gamma_1,\gamma_2)\mathbb{P}(\gamma_1,\gamma_2) \tag{4.27}$$

$$= e^{-\lambda P^{\frac{2}{\alpha}}(\gamma_3)^{\frac{2}{\alpha}}}(\lambda\frac{2}{\alpha}P^{\frac{2}{\alpha}})^3\prod_{i=1}^{3}\gamma_i^{\frac{2}{\alpha}-1}. \tag{4.28}$$

Following this procedure, it is easy to obtain the general result for n relays:

$$\mathbb{P}(\gamma_1,\gamma_2,\ldots,\gamma_n) = e^{-\lambda P^{2/\alpha}(\gamma_n)^{2/\alpha}}(\lambda\frac{2}{\alpha}P^{2/\alpha})^n\prod_{i=1}^{n}\gamma_i^{2/\alpha-1}. \tag{4.29}$$

Similarly, for the MRC scheme, the outage probability at victim v_j can be given as:

$$\begin{aligned} P_o^{MRC} &= \mathbb{P}(\Gamma_j^m < \theta) \\ &= \mathbb{P}(\sum_{v_i\in\mathbb{R}} P_v|h_{ij}||d_{ij}||^{-\alpha/2}|^2 < \theta\sigma^2) \\ &= \mathbb{E}_\gamma(1-\exp(-\frac{\theta\sigma^2}{\sum_{v_i\in\mathbb{R}}\gamma_{ij}^{-1}})) \\ &= \int_{\gamma_{ij}}(1-\exp(-\frac{\theta\sigma^2}{\sum_{v_i\in\mathbb{R}}\gamma_{ij}^{-1}}))f(\gamma_{ij})d\gamma_{ij}. \end{aligned} \tag{4.30}$$

Let us define $Q(\gamma_{ij})$ as follows:

$$Q(\gamma_{ij}) = \begin{cases} \prod_{v_i\in\mathbb{R}}(1-e^{-\theta\sigma^2\gamma_{ij}}), & SC \\ 1-\exp(-\frac{\theta\sigma^2}{\sum_{v_i\in\mathbb{R}}\gamma_{ij}^{-1}}), & MRC, \end{cases} \tag{4.31}$$

then the outage probabilities achieved by these two schemes can be expressed in a unified form:

$$P_o = \int_{\gamma_{ij}} Q(\gamma_{ij})f(\gamma_{ij})d\gamma_{ij}. \tag{4.32}$$

In the following, we focus on the derivation of the outage probability P_o according to the two jamming scenarios discussed in previous section.

4.2.1 Scenario 1: $r(1-a) < m < r$

In this scenario, let $L(v_j) = x$ denote the location of a specific victim v_j. Based on the characteristics of the homogeneous PPP, x is uniformly distributed within $(\frac{m}{1+a}, r)$ with a density function $g(x) = (1+a)/(r+ra-m)$, and the relays are located within the range of $[-r, \frac{m}{1+a}]$. Thus, it can

be obtained that $d_{ij} \in [x - \frac{m}{1+a}, x + r], \forall i \in \mathbb{R}$. Since $x - \frac{m}{1+a} > 0$ and $x + r > 0$, the range of γ_{ij} can be given as:

$$\gamma_{ij} \in \left[\frac{(x - \frac{m}{1+a})^{\alpha}}{P_v}, \frac{(x+r)^{\alpha}}{P_v} \right] \tag{4.33}$$

Under these conditions, for each victim $v_j \in \mathbb{V}$, we have:

$$\begin{aligned} P_o &= \sum_k \int_x \int_{\gamma_{ij}} Q(\gamma_{ij}) f(\gamma_{ij}) d\gamma_{ij} g(x) dx \mathbb{P}[N(k)] \\ &= \sum_k \int_{\frac{m}{1+a}}^{r} \int_{l_m}^{l_n} Q(\gamma_{ij}) e^{-\lambda P_v^{2/\alpha} \gamma_{kj}^{2/\alpha}} (\lambda \frac{2}{\alpha} P_v^{2/\alpha})^k \prod_{i=1}^{k} \gamma_{ij}^{2/\alpha - 1} d\gamma_{ij} (\frac{1+a}{r + ra - m}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = k], \end{aligned} \tag{4.34}$$

where $l_m = \frac{(x - \frac{m}{1+a})^{\alpha}}{P_v}$ and $l_n = \frac{(x+r)^{\alpha}}{P_v}$.

Unfortunately, the close-form expression of (4.34) is difficult to obtain. Therefore, we resort to the derivation of the outage probability in the worst-case scenario. Specifically, for both the MRC and SC schemes, we have $\mathbb{E}_\gamma[(1 - e^{-\theta\sigma^2 \gamma_{1j}})] > \mathbb{E}_\gamma[(1 - e^{-\theta\sigma^2 \gamma_{1j}})(1 - e^{-\theta\sigma^2 \gamma_{2j}})] > \ldots > \mathbb{E}_\gamma(\prod_{i=1}^{k}(1 - e^{-\theta\sigma^2 \gamma_{ij}}))$ and $\mathbb{E}_\gamma[(1 - \exp(-\frac{\theta\sigma^2}{\gamma_{ij}^{-1}}))] > \mathbb{E}_\gamma[(1 - \exp(-\frac{\theta\sigma^2}{\gamma_{1j}^{-1} + \gamma_{2j}^{-1}}))] > \ldots > \mathbb{E}_\gamma[(1 - \exp(-\frac{\theta\sigma^2}{\sum_{i=1}^{k} \gamma_{ij}^{-1}}))]$ Therefore, the outage probability with $K (K \geq 1)$ can be upper bounded by the single relay scenario (K=1). In this case, $Q(\gamma_{ij}) = 1 - e^{-\theta\sigma^2 \gamma}$, so we have:

$$\begin{aligned} P_o &= \int_x \int_{\gamma} (1 - e^{-\theta\sigma^2 \gamma}) f(\gamma) d\gamma g(x) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= \int_{\frac{m}{1+a}}^{r} [1 - \int_{l_m}^{l_n} \exp(-(\theta\sigma^2 \gamma + \lambda P_v^{2/\alpha} \gamma^{2/\alpha})) (\lambda \frac{2}{\alpha} P_v^{2/\alpha}) \gamma^{2/\alpha - 1} d\gamma] (\frac{1+a}{r + ra - m}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1]. \end{aligned} \tag{4.35}$$

In practice, the path loss exponent α is normally between 2 and 5, so we consider two typical cases of $\alpha = 2$ and $\alpha = 4$ to further derive the outage probability for the single relay scenario.

$\alpha = 2$: In this case, P_o can be further simplified as:

$$\begin{aligned} P_o &= \int_{\frac{m}{1+a}}^{r} [1 - \int_{\frac{(x - \frac{m}{1+a})^2}{P_v}}^{\frac{(x+r)^2}{P_v}} \exp(-(\theta\sigma^2 + \lambda P_v)\gamma) \lambda P_v d\gamma] (\frac{1+a}{r + ra - m}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= \int_{\frac{m}{1+a}}^{r} [1 + \frac{\lambda P_v}{\mu} (\exp(-\frac{\mu(x+r)^2}{P_v}) - \exp(-\frac{\mu(x - \frac{m}{1+a})^2}{P_v}))] (\frac{1+a}{r + ra - m}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1], \end{aligned} \tag{4.36}$$

where $\mu = \theta\sigma^2 + \lambda P_v$.

Let $\kappa = \mu/P_v$, $y = x + r$ and $z = x - \frac{m}{1+a}$, then (4.36) can be approximated as follows:

$$P_o = [1 + \frac{\lambda\sqrt{\pi}(\text{erf}(\sqrt{\kappa}2r) - \text{erf}(\sqrt{\kappa}(\frac{m}{1+a} + r)))}{2\sqrt{\kappa}} - \frac{\lambda\sqrt{\pi}\,\text{erf}(\sqrt{\kappa}(r - \frac{m}{1+a}))}{2\sqrt{\kappa}}](\frac{1+a}{r + ra - m})\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1]. \tag{4.37}$$

$\alpha = 4$**:** In this case, P_o can be obtained as follows:

$$\begin{aligned} P_o &= \int_{\frac{m}{1+a}}^{r}[1 - \int_{\frac{(x-\frac{m}{1+a})^4}{P_v}}^{\frac{(x+r)^4}{P_v}} \exp(-\theta\sigma^2\gamma - \lambda P_v^{1/2}\gamma^{1/2})\frac{\lambda P_v^{1/2}}{2}\gamma^{-1/2}d\gamma](\frac{1+a}{r + ra - m})dx\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= \int_{\frac{m}{1+a}}^{r}[1 - \int_{\frac{(x-\frac{m}{1+a})^2}{P_v^{1/2}}}^{\frac{(x+r)^2}{P_v^{1/2}}} \lambda P_v^{1/2}\exp(-(\theta\sigma^2 y^2 - \lambda P_v^{1/2}y))dy](\frac{1+a}{r + ra - m})dx\mathbb{P}[N(-r, \frac{m}{1+a}) = 1] \\ &= \int_{\frac{m}{1+a}}^{r}(\frac{1+a}{r + ra - m})[1 - \frac{\sqrt{\pi}\lambda P_v^{1/2}\exp(\frac{\lambda^2 P_v}{4\theta\sigma^2})}{2\theta\sigma^2}(\text{erf}(\frac{2\theta\sigma^2\frac{(x+r)^2}{P_v^{1/2}} + \lambda P_v^{1/2}}{2\sqrt{\theta}\sigma}) \\ &\qquad - \text{erf}(\frac{2\theta\sigma^2\frac{(x-\frac{m}{1+a})^2}{P_v^{1/2}} + \lambda P_v^{1/2}}{2\sqrt{\theta}\sigma}))]dx\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1]. \end{aligned} \tag{4.38}$$

Let $\phi = \theta\sigma^2$, $\varphi = P_v^{1/2}$, $y = (x + r)^2$ and $z = (x - \frac{m}{1+a})^2$, then (4.38) can be approximated as:

$$\begin{aligned} P_o = [1 - (&\frac{e^{-\frac{(\lambda\varphi)^2}{4\phi}}(\sqrt{\pi}\lambda\varphi e^{(\frac{(\lambda\varphi)^2}{4\phi})}\,\text{erf}(\frac{(\lambda\varphi}{2\sqrt{\phi}}) + 2\sqrt{\phi})}{2\sqrt{\pi}\phi/\varphi} \\ &+ \frac{e^{-\frac{(\frac{2\phi}{\varphi}y'' + \lambda\varphi)^2}{4\phi}}(\sqrt{\pi}\psi_{y''}e^{\frac{\phi}{\varphi^2}y''^2 + \lambda y''}\,\text{erf}(\frac{(\frac{2\phi}{\varphi}y' + \lambda\varphi}{2\sqrt{\phi}}) + 2\sqrt{\phi})}{2\sqrt{\pi}\phi/\varphi} \\ &- \frac{e^{-\frac{(\frac{2\phi}{\varphi}y' + \lambda\varphi)^2}{4\phi}}(\sqrt{\pi}\psi_{y'}e^{\frac{\phi}{\varphi^2}y'^2 + \lambda y'}\,\text{erf}(\frac{(\frac{2\phi}{\varphi}y' + \lambda\varphi}{2\sqrt{\phi}}) + 2\sqrt{\phi})}{2\sqrt{\pi}\phi/\varphi} \\ &- \frac{e^{-\frac{(\frac{2\phi}{\varphi}z'' + \lambda\varphi)^2}{4\phi}}(\sqrt{\pi}\psi_{z''}e^{\frac{\phi}{\varphi^2}z''^2 + \lambda z''}\,\text{erf}(\frac{(\frac{2\phi}{\varphi}z'' + \lambda\varphi}{2\sqrt{\phi}}) + 2\sqrt{\phi})}{2\sqrt{\pi}\phi/\varphi}) \\ &\frac{\lambda\varphi\sqrt{\pi}e^{\frac{(\lambda\varphi)^2}{4\phi}}}{2\sqrt{\phi}}](\frac{1+a}{r + ra - m})\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1], \end{aligned} \tag{4.39}$$

where $y' = (\frac{m}{1+a} + r)^2$, $y'' = (2r)^2$, $z'' = (r - \frac{m}{1+a})^2$ and $\psi_{y'} = \frac{2\phi}{\varphi}e^{\frac{(\lambda\varphi)^2}{4\phi}}y' + \varphi e^{\frac{(\lambda\varphi)^2}{4\phi}}$, $\psi_{y''} = \frac{2\phi}{\varphi}e^{\frac{(\lambda\varphi)^2}{4\phi}}y'' + \varphi e^{\frac{(\lambda\varphi)^2}{4\phi}}$, $\psi_{z''} = \frac{2\phi}{\varphi}e^{\frac{(\lambda\varphi)^2}{4\phi}}z'' + \varphi e^{\frac{(\lambda\varphi)^2}{4\phi}}$.

4.2.2 Scenario 2: $0 < m < r(1-a)$

In this scenario, x is uniformly distributed within $(\frac{m}{1+a}, \frac{m}{1-a})$ with a density function of $g(x) = (1-a^2)/2am$. The relay candidates v_i are located within the range $[-r, \frac{m}{1+a}]$ and relay candidates $v_{i'}$ are located within the range $[\frac{m}{1-a}, r]$. Thus, it can be obtained that $d_{ij} \in [x - \frac{m}{1+a}, x + r]$ and $d_{i'j} \in [\frac{m}{1+a} - x, r - x]$.

Similar to previous case, since $x - \frac{m}{1+a} > 0$, $x + r > 0$ and $\frac{m}{1-a} - x > 0$, $r - x > 0$, the range of $\gamma_{v_i v_j}$ can be given as:

$$\gamma_{ij} \in [\frac{(x - \frac{m}{1+a})^\alpha}{P_v}, \frac{(x+r)^\alpha}{P_v}] \tag{4.40}$$

$$\gamma_{i'j} \in [\frac{(\frac{m}{1-a} - x)^\alpha}{P_v}, \frac{(r-x)^\alpha}{P_v}] \tag{4.41}$$

Therefore, in both schemes, for each victim $v_j \in \mathbb{V}$, we have:

$$\begin{aligned} P_o = & \sum_u \int_x \int_{\gamma_{ij}} t(\gamma_{ij}) f(\gamma_{ij}) d\gamma_{ij} g(x) dx \mathbb{P}[N(u)] + \sum_w \int_x \int_{\gamma_{i'j}} t(\gamma_{i'j}) d\gamma_{i'j} f(x) dx \mathbb{P}[N(w)] \\ = & \sum_u \int_{\frac{m}{1+a}}^{r} \int_{l_m}^{l_n} t(\gamma_{ij}) e^{-\lambda P_v^{2/\alpha} \gamma_{uj}^{2/\alpha}} (\lambda \frac{2}{\alpha} P_v^{2/\alpha})^u \prod_{i=1}^{u} \gamma_{ij}^{2/\alpha - 1} d\gamma_{ij} (\frac{1-a^2}{2am}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = u] \\ & + \sum_w \int_{\frac{m}{1+a}}^{r} \int_{l_{m'}}^{l_{n'}} t(\gamma_{i'j}) e^{-\lambda P_v^{2/\alpha} \gamma_{wj}^{2/\alpha}} (\lambda \frac{2}{\alpha} P_v^{2/\alpha})^w \prod_{i'=1}^{w} \gamma_{i'j}^{2/\alpha - 1} d\gamma_{i'j} (\frac{1-a^2}{2am}) dx \mathbb{P}[N_R(\frac{m}{1-a}, r) = w], \end{aligned} \tag{4.42}$$

where $l_m = \frac{(x - \frac{m}{1+a})^\alpha}{P_v}$, $l_n = \frac{(x+r)^\alpha}{P_v}$, $l_{m'} = \frac{(\frac{m}{1-a} - x)^\alpha}{P_v}$ and $l_{n'} = \frac{(r-x)^\alpha}{P_v}$.

For single relay case, the relay node can be located within any of these two areas. If v_i is in the range of $[-r, \frac{m}{1+a}]$, we have:

$$\begin{aligned} P_o = & \int_x \int_\gamma (1 - e^{-\theta \sigma^2 \gamma}) f(\gamma) d\gamma g(x) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ = & \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{l_m}^{l_n} \exp(-(\theta \sigma^2 \gamma + \lambda P_v^{2/\alpha} \gamma^{2/\alpha})) (\lambda \frac{2}{\alpha} P_v^{2/\alpha}) \gamma^{2/\alpha - 1} d\gamma] \\ & (\frac{1-a^2}{2am}) dx \mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1]. \end{aligned} \tag{4.43}$$

Otherwise if relay node $v_{i'}$ is in the range of $[\frac{m}{1-a}, r]$, we have:

$$\begin{aligned} P'_o &= \int_x \int_\gamma [1 - e^{-\theta\sigma^2\gamma}] f(\gamma) d\gamma f(x) dx \mathbb{P}[N_R(\frac{m}{1-a}, r) = 1] \\ &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{l_{m'}}^{l_{n'}} \exp(-(\theta\sigma^2\gamma + \lambda P_v^{2/\alpha}\gamma^{2/\alpha}))(\lambda\frac{2}{\alpha}P_v^{2/\alpha})\gamma^{2/\alpha-1}d\gamma] \\ &\qquad (\frac{1-a^2}{2am})dx\mathbb{P}[N_R(\frac{m}{1-a}, r) = 1]. \end{aligned} \tag{4.44}$$

$\alpha = 2$**:** In this case, P_o and P'_o can be obtained similar to (4.42).

$$\begin{aligned} P_o &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{\frac{(x-\frac{m}{1+a})^2}{P_v}}^{\frac{(x+r)^2}{P_v}} \exp(-(\theta\sigma^2 + \lambda P_v)\gamma)\lambda P_v d\gamma](\frac{1-a^2}{2am})dx\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 + \frac{\lambda P_v}{\mu}(\exp(-\frac{\mu(x+r)^2}{P_v}) - \exp(-\frac{\mu(x-\frac{m}{1+a})^2}{P_v}))](\frac{1-a^2}{2am})dx\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= [1 + \frac{\lambda\sqrt{\pi}(\mathrm{erf}(\sqrt{\kappa}(\frac{m}{1-a} + r)) - \mathrm{erf}(\sqrt{\kappa}(\frac{m}{1+a} + r)))}{2\sqrt{\kappa}} \\ &\qquad - \frac{\lambda\sqrt{\pi}\,\mathrm{erf}(\sqrt{\kappa}(\frac{m}{1-a} - \frac{m}{1+a}))}{2\sqrt{\kappa}}](\frac{1-a^2}{2am})\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1], \end{aligned} \tag{4.45}$$

$$\begin{aligned} P'_o &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{\frac{(\frac{m}{1-a}-x)^2}{P_v}}^{\frac{(r-x)^2}{P_v}} \exp(-(\theta\sigma^2 + \lambda P_v)\gamma)\lambda P_v d\gamma](\frac{1-a^2}{2am})dx\mathbb{P}[N_R(\frac{m}{1-a}, r) = 1]. \\ &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 + \frac{\lambda P_v}{\mu}(\exp(-\frac{\mu(r-x)^2}{P_v}) - \exp(-\frac{\mu(\frac{m}{1-a}-x)^2}{P_v}))](\frac{1-a^2}{2am})dx\mathbb{P}[N_R(-r, \frac{m}{1+a}) = 1] \\ &= [1 + \frac{\lambda\sqrt{\pi}(\mathrm{erf}(\sqrt{\kappa}(r - \frac{m}{1+a}) - \mathrm{erf}(\sqrt{\kappa}(r - \frac{m}{1-a})))}{2\sqrt{\kappa}} \\ &\qquad - \frac{\lambda\sqrt{\pi}\,\mathrm{erf}(\sqrt{\kappa}(\frac{m}{1-a} - \frac{m}{1+a}))}{2\sqrt{\kappa}}](\frac{1-a^2}{2am})\mathbb{P}[N_R(\frac{m}{1-a}, r) = 1], \end{aligned} \tag{4.46}$$

$\alpha = 4$**:** In this case, P_o and P'_o also can be obtained similar to (4.42).

$$\begin{aligned} P_o &= \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{\frac{(x-\frac{m}{1+a})^4}{P_v}}^{\frac{(x+r)^4}{P_v}} \exp(-\theta\sigma^2\gamma - \lambda P_v^{1/2}\gamma^{1/2})\frac{\lambda P_v^{1/2}}{2}\gamma^{-1/2}d\gamma] \\ &\qquad (\frac{1-a^2}{2am})dx\mathbb{P}[N(-r, \frac{m}{1+a}) = 1] \end{aligned} \tag{4.47}$$

$$P'_o = \int_{\frac{m}{1+a}}^{\frac{m}{1-a}} [1 - \int_{(\frac{m}{1-a}-x)^4/P_v}^{(r-x)^4/P_v} \exp(-\theta\sigma^2\gamma - \lambda P_v^{1/2}\gamma^{1/2}) \frac{\lambda P_v^{1/2}}{2} \gamma^{-1/2} d\gamma] \\ (\frac{1-a^2}{2am}) dx \mathbb{P}[N(\frac{m}{1-a}, r) = 1]. \tag{4.48}$$

4.3 Anti-jamming Relay Selection Problem

In previous section, we analyse the outage probability for the cooperative anti-jamming relaying scheme under the conditions of random distributions of vehicles and channel fading. In practice, the channel state information (CSI) between vehicles can be obtained through on-line measurements or channel overhearing. Therefore, the channel gain $H_{ij} = \|h_{ij}\|^2 \|d_{ij}\|^{-\alpha}$ between relay node v_i and victim vehicle v_j can be assumed to be known, and thus the achievable SNR at a victim vehicle v_j with a selected subset of relay nodes $R_s \subseteq \mathbb{R}$ can be given by:

$$\text{SNR}_j = \sum_{v_i \in \mathbb{S}} \hat{P}_i H_{ij} \tag{4.49}$$

where $\hat{P}_i = P_i/\sigma^2$ is the normalized transmit power of relay node v_i.

Under this condition, the major challenge is to select a subset of relay nodes $\mathbb{S}$ from the candidate relay set $\mathbb{R}$, such that the SNR of the worst victim vehicle is maximized, which can be formally defined as a Max-Min Relay Selection (MMRS) problem as follows:

$$\begin{aligned} \max \quad & \min_{v_j \in \mathbb{V}} \sum_{v_i \in \mathbb{R}} \hat{P}_i H_{ij} \\ s.t. \quad & |R_s| \leq K, \end{aligned} \tag{4.50}$$

where the constraint specifies that at most K relay nodes can be selected, which is based on the assumption that the transmissions of all selected relays are scheduled with the time division multiple access (TDMA) method. As a result, the setting of parameter K involves the trade-off between transmission reliability and decoding delay. Specifically, a smaller outage probability can be achieved with larger value of K since the diversity of more relays can be exploited, but it will incur larger decoding delay since the signals of all relays have to be received to decode the symbol using the MRC or SC scheme.

The MMRS problem can be formulated as a linear integer programming (LIP) problem as

follows:

$$\begin{aligned} \max \quad & \min_{v_j \in \mathbb{V}} \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \leq K, \qquad \forall v_j \in \mathbb{V}, \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{4.51}$$

where x_i is an indicator variable to denote whether node v_i is selected as relay or not.

By introducing an auxiliary threshold θ, (4.51) can be transformed to the following equivalent form:

$$\begin{aligned} \max \quad & \theta, \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \geq \theta, \forall v_j \in \mathbb{V}, \\ & \sum_{v_i \in \mathbb{R}} x_i \leq K, \forall v_i \in \mathbb{R}. \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{4.52}$$

For a given threshold $\hat{\theta}$, (4.52) degenerates to a feasibility-checking problem, that is, to check if a subset of K relays can be found such that the SNRs of all victim vehicles are above the threshold θ. This feasibility-checking problem can be achieved by solving a minimum relay selection (MRS) problem as follows:

$$\begin{aligned} \min \quad & \sum_{v_i \in \mathbb{R}} x_i, \\ s.t. \quad & \sum_{v_i \in \mathbb{R}} x_i \hat{P}_i H_{ij} \geq \hat{\theta}, \forall v_j \in \mathbb{V}, \\ & x_i \in \{0,1\}, \qquad \forall v_i \in \mathbb{R}, \end{aligned} \tag{4.53}$$

which attempts to find the minimum number of relays such that the SNRs of all victims are satisfied. If the obtained number of relays from this MRS problem is larger than K, then (4.52) is infeasible for the given $\hat{\theta}$, which suggests that $\hat{\theta}$ is too large, so the threshold should be decreased, otherwise we can increase θ to achieve better SNR for all victims.

Based on this property, we propose a bisection algorithm to search the optimal θ for the transformed MMRS problem in (4.52). As shown in Algorithm 3, the algorithm proceeds in rounds with the initial lower and upper thresholds θ_L and θ_U for θ. In each iteration, θ is set to $(\theta_L + \theta_U)/2$, and the MRS problem is solved using Algorithm 4 to check the feasibility of problem (4.53) under

this θ. If the problem is feasible, the lower threshold θ_L is increased to θ, otherwise the upper threshold θ_U is reduced to θ. This procedure is repeated until convergence (e.g., $(\theta_U - \theta_L)/2$ is less than a certain threshold Δ).

Algorithm 3 MMRS Algorithm

Inputs: θ_L, θ_U and threshold Δ.

Outputs: Selected relay set R_s;

while $(\theta_U - \theta_L)/2 > \Delta$ **do**
  $\theta \leftarrow (\theta_U + \theta_L)/2$;
  Find relay set R_s with Algorithm 4;
  if $R_s \neq \emptyset$ **then**
    $\theta_L \leftarrow \theta$;
  else
    $\theta_U \leftarrow \theta$;
  end if
end while
return R_s;

The MRS problem in (4.53) belongs to the family of cooperative covering problems (CCP), which is known to be significantly harder to find the exact solution than the standard set covering problem[90]. Therefore, we propose a heuristic algorithm as shown in Algorithm 4. The basic idea is to incrementally select a relay from the candidate set that yields the maximum SNR increment to the worst victim node. This process is repeated until the SNRs of all victims are above θ (which indicates (4.52) is feasible), or the number of selected relays is above K (which indicates (4.52) is infeasible).

The complexity of these two algorithms can be analysed as follows. For Algorithm 1, the optimal value of θ can be found with $O(\log_2 \frac{(\theta_U-\theta_L)}{\Delta})$ iterations. In each iteration, Algorithm 2 is invoked to solve the MRS algorithm. For the case of m candidate relays and n victims, the worst case time complexity to select at most k relays is given as $O(kmn)$. Therefore, the overall complexity of these two algorithms is $O(\log_2 \frac{(\theta_U-\theta_L)}{\Delta} kmn)$.

4.4 Performance Evaluation

In this section, we provide simulation results to evaluate the performance of the proposed cooperative anti-jamming relaying scheme and relay selection algorithm. In the simulations, the Urban

Algorithm 4 MRS algorithm

Input: Candidate relay set $\mathbb{R}$ and victim set $\mathbb{V}$, normalized transmit power $\hat{P}_i, \forall i \in \mathbb{R}$, channel gain matrix H_{ij}, SNR threshold θ, and relay number constraint K.

Output: Selected relay set R_s;

$R_s \leftarrow \emptyset$;
2: $S_j \leftarrow 0, \forall j \in \mathbb{V}$;
$SNR_{min} \leftarrow 0$;
4: $n \leftarrow 0$;
while $SNR_{min} < \theta$ and $n < K$ **do**
6: **for each** $i \in \mathbb{R} \setminus R_s$ **do**
$S_{min}[i] \leftarrow \min_{j \in \mathbb{V}}(S_j + \hat{P}_i H_{ij})$;
8: **end for**
$i' \leftarrow \arg\max_{i \in \mathbb{R}} S_{min}[i]$;
10: $R_s \leftarrow R_s \cup i'$;
for each $j \in \mathbb{V}$ **do**
12: $S_j \leftarrow S_j + \hat{P}_{i'} H_{i'j}$;
end for
14: $SNR_{min} \leftarrow \min_{j \in \mathbb{V}} S_j$;
$n \leftarrow n + 1$;
16: **end while**
if $SNR_{min} < \theta$ **then**
18: $R_s \leftarrow \emptyset$;
end if
20: **return** R_s;

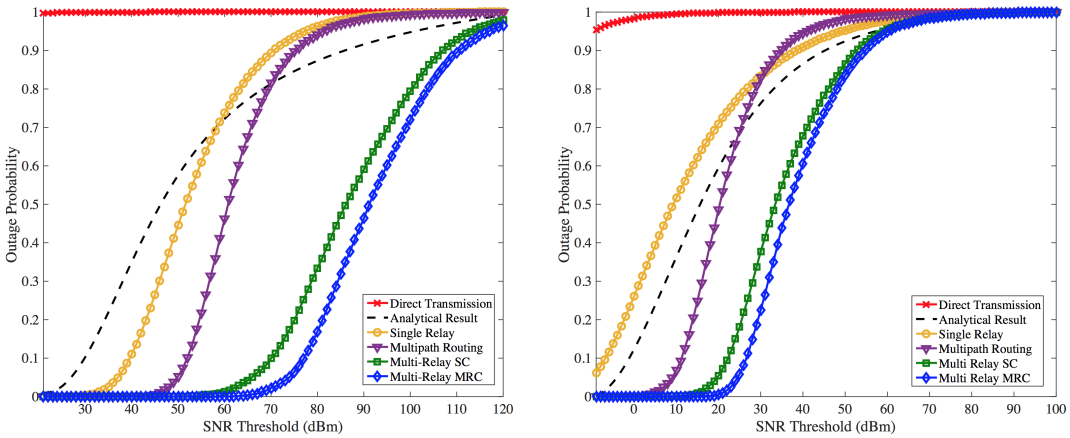
MObility (SUMO) simulator is used to generate traffic flows for urban scenario, where parameters are set based on real-time urban traffic. These traffic flow data are used as inputs to obtain the simulation results using Matlab. The parameter settings in the simulation are summarized in Table 4.2:

4.4.1 Snapshot Evaluation

In Fig. 4.3, we show the cumulative distribution function (CDF) of the outage probabilities achieved by the proposed cooperative relaying schemes under different SNR thresholds, which are obtained based on the average of 100 snapshots of the locations of the vehicles and jammer. Note that from the theoretical analysis, it can be seen that the outage probability is similar regardless of location

Table 4.2: Parameters Used in Simulations

Parameter	Value
Simulation Time	300s
RSU coverage	1 km
Street Width	2 Lanes
Vehicle Velocity	40 - 60 km/h
Number of Vehicles	350 - 800
Fading model	Rayleigh
Transit power (RSU, Vehicles)	150mw and 100mw
Noise power	-95 dBm



(a) Path loss expoent $\alpha = 2$ (b) Path loss expoent $\alpha = 4$

Figure 4.3: Outage probability distributions for different SNR thresholds

of the relay nodes within the range of $[-r, \frac{m}{1+a}]$ or $[\frac{m}{1-a}, r]$. Therefore, we only consider the scenarios where the relay nodes are located within the range of $[-r, \frac{m}{1+a}]$, and the path loss exponent (α) is set to 2 and 4 respectively. From this figure, it can be observed that using the DT scheme, the outage probability always exceed 95%, that is, more than 95% signals sent by the RSU cannot be decoded by the victim vehicles, which confirms the significant impact of the jamming attacks. For the single relay scenario, we assume a single relay is uniformly distributed within the range of $[-r, \frac{m}{1+a}]$. It is considered as a mean value among all single relay scenarios, and the simulation result match well with the analytical result. The multipath routing protocol is designed based on the scheme proposed in [34], where multiple end-to-end routes are selected based on their availability history and one of them is chosen to deliver packets at a given time. Thus, the PDR of each link in this path updated. It can be observed that after implemented the multipath routing pro-

tocol, the outage probability is dramatically reduced and its achieved outage probability is lower than the mean value of all single relay scenarios, which suggest that by taking advantage of the diversity gain proposed by the neighbouring nodes, the multipath protocol can make fairly well routing decisions. It helps improving the anti-jamming performance. However, in the case the routing decision is made based on the path's availability history, and the channel state of each link is not monitored in real time. Therefore the path selected by this MAC layer routing protocol is not always the most suitable one. Using the SC scheme, the relay node with the best channel to the victim vehicles is selected, so the performance is significantly improved than the single relay scenarios. Finally, it can be seen that the MRC method outperforms all other schemes under all conditions since the contribution of all relay nodes are fully exploited.

4.4.2 Continuous Jamming Process Evaluation

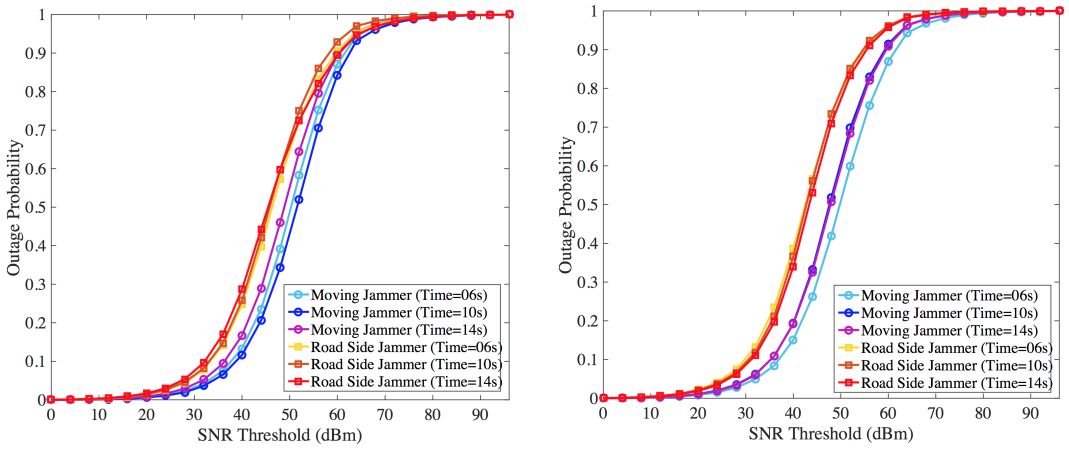
In this part, the jamming attacks are modelled as a continuous process, and both the road side jammer and moving jammer are considered in the simulations. We still consider the worst case scenarios in the simulations. That is, the road side jammer is located at the position where its jamming range reaches the maximum value ($m = 350$ in our case), while for the moving jammer, it launches attack once it enters the coverage area of the RSU. We consider three different lengths of time duration (6s, 10s, 14s) and transmit power levels (100mw, 125mw and 150mw).

In Fig. 4.4 and Fig. 4.5, we show the outage probability under different jamming power levels and time durations. In general, the road side jammer causes greater damage than the moving jammer since it can always reach its desire jamming range. For more details, for different time periods, in the face of the moving jammer, the performance variance of the proposed cooperative relaying scheme is relatively larger than the case of road side jammer since the interference range of the moving jammer changes over time with its movement. It also can be seen that the overall performance of the proposed scheme decreases with larger jamming power levels. However, even if the transmit power of the jammer is increased to the same level of the RSU (150mw), our proposed scheme can still guarantee a small outage probability for the low SNR regime (less than 20dBm).

4.4.3 Evaluation of Relay Selection Schemes

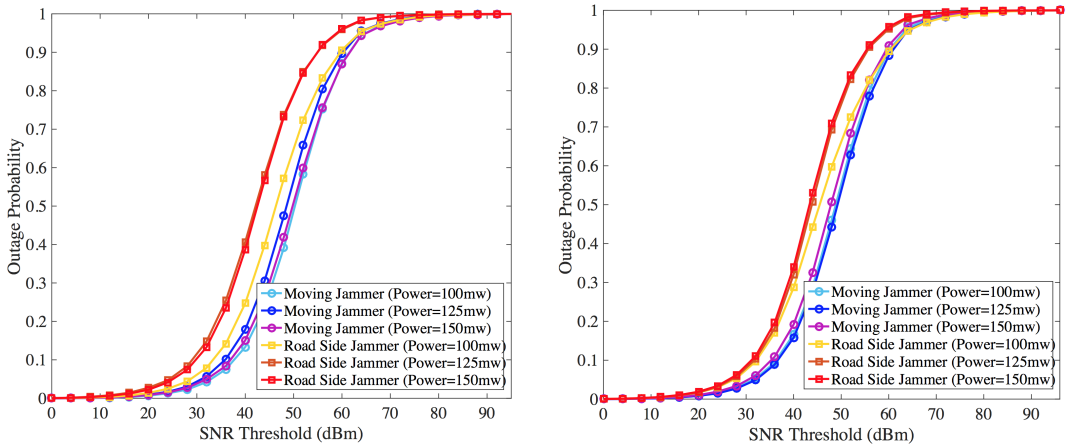
In these simulations, we compare the performance of four different relay selection schemes for the multiple relay scenarios:

- Random Relay Selection (RRS): In this scheme, K relays are randomly selected from the available candidate relays;



(a) Jamming power $P = 100mw$ (b) Jamming power $P = 150mw$

Figure 4.4: Outage probability under different jamming power levels



(a) Duration $t = 6s$ (b) Duration $t = 14s$

Figure 4.5: Outage probability for different jamming durations

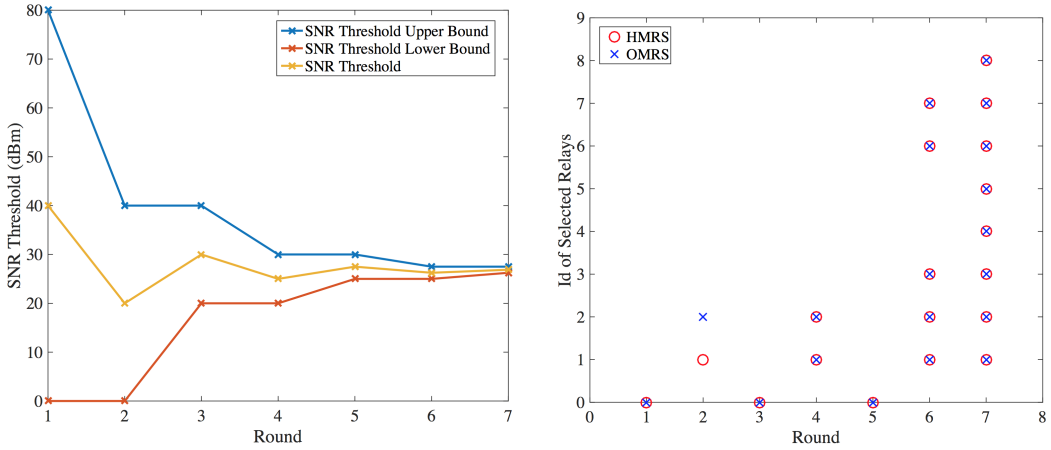
- Ordering Relay Selection (ORS): In this scheme, the relay selection procedure is designed based on the Geocast routing protocols in vehicular networks [91, 92], where the candidate relays are sorted in the decreasing order of their distances to the victim nodes, and then K nearest relays are incrementally selected into the relay set;

- Heuristic Max-min Relay Selection (HMRS): This is our proposed scheme that selects K relays using Algorithm 3 and 4;

- Optimal Max-min Relay Selection (OMRS): Which is similar to HMRS scheme except that the minimum relay set is found with the integer linear programming solver (intlinprog) in Matlab.

Firstly, we compare the performance of our proposed HMRS scheme with the OMRS scheme.

In Fig. 4.6(a), we show the updates of θ, θ_L and θ_U during the bisection procedure of Algorithm 1. It can be observed the optimal θ can be found after 7 rounds, which is much faster than a naive linear search method, and the results are the same for both the HMRS and OMRS schemes. To check the details of these two schemes, we show the selected relays by these two schemes in each iteration in Fig. 4.6(b)(Note that the MRS problem is infeasible in the 1st, 3rd and 5th iterations, so the returned R_s is empty). It can be seen that the relays selected by these schemes can be different. For example, in the 2nd iteration, the HMRS scheme selects relays 1, but the OMRS scheme selects relays 2. But after the convergence, both schemes select the same relays (e.g., in the 7th iteration). Therefore, in the following simulations, we only consider three methods: RRS, ORS and HMRS.



(a) Convergence of HMRS and OMRS.

(b) Relays selected by HMRS and OMRS.

Figure 4.6: Comparison of HMRS and OMRS (Relay number constraint $K = 8$).

In Fig. 4.7, we show the outage probability achieved by these relay selection algorithms. It can be seen that as the number of relay nodes increases, the outage probability is decreasing gradually, and our proposed HMRS scheme outperforms other two schemes under all conditions. In particular, the performance gap is the largest for the single relay scenarios. Thus, after the number of relay nodes exceeds a certain number, the performance improvement becomes marginal, which is partially due to the special topology of the vehicular networks, that is, the further away the relay nodes from the victim nodes, the less contribution to the combined SNR. Therefore, the relay nodes should be selected judiciously according to the required SNR. We can see that when the number of selected relays is beyond 5, the performance of the ORS scheme is very close to the HMRS scheme because the selected relays are basically similar under both methods. More details are shown in Fig 4.8 for the outage probabilities achieved by these schemes with the SNR threshold of 15dBm.

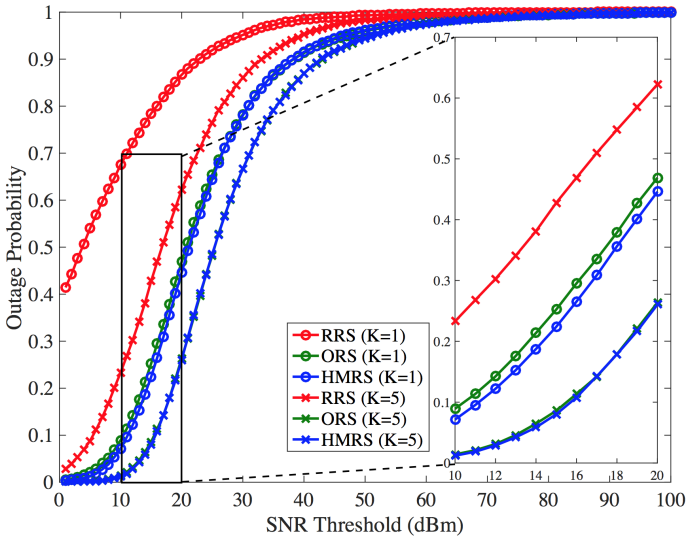


Figure 4.7: Outage probability distributions for different SNR thresholds.

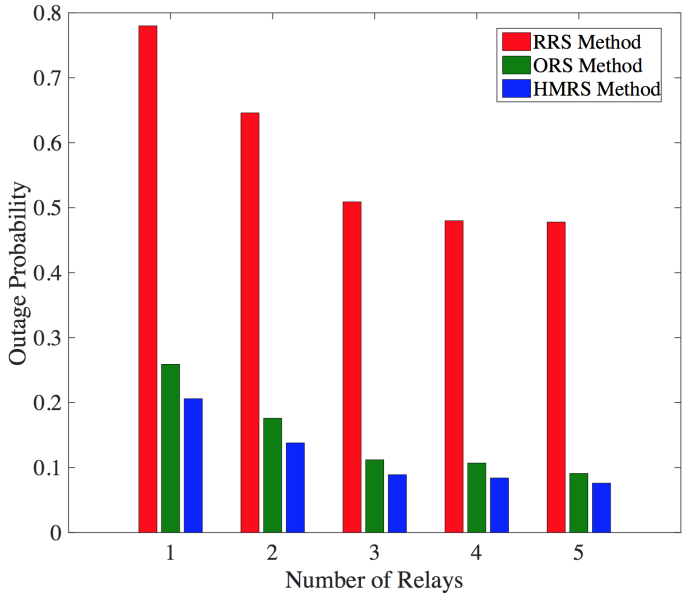


Figure 4.8: Outage probability distributions for different number of relay nodes (θ = 15).

4.5 Conclusion

In this chapter, a cooperative anti-jamming relaying scheme was presented for the control channel jamming problem in the VNETs. The outage probability for this cooperative relaying scheme was analysed based on the PPP model of the distribution of vehicles under different jamming scenarios, and the close-form expressions are derived for the single relay case. A max-min relay selection problem was formulated to maximize the SNR of the worst victim vehicle under the relay number constraint. A heuristic algorithm was designed to solve this nontrivial problem based on the bisection method. The performance of the proposed schemes were evaluated and compared with other schemes through simulations, which demonstrate that the proposed cooperative relaying

schemes can effectively reduce the outage probability, and the performance of the heuristic relay selection algorithm is close to the optimal solution.

For future work, we will consider the relay selection problem under the assumption of partial channel station information, which can significantly reduce the overhead for channel measurement. It is also interesting to extend the problem to the multi-jammer cases, and multi-antenna techniques can be incorporated with the cooperative relaying scheme to achieve better anti-jamming performance.

Chapter 5

Cooperative anti-jamming Beamforming for Multi-antenna Vehicular Networks

Due to the importance of the control channel in I2V communications, we focus on the design of anti-jamming schemes to address the control channel jamming issue in vehicular networks in this paper. Specifically, we consider the scenario where the RSU is equipped with multiple antennas, who can serve multiple groups of vehicles simultaneously using the multi-group multicast beamforming technique[44]. However, due to the interference from the jammer, not all vehicles can decode the desired signals from the RSU. As a solution, we propose a two stage anti-jamming scheme, whereby the vehicles who have successfully decoded the signal received in the first stage will be selected as relays to cooperatively serve the victim vehicles in the second stage using the coordinated beamforming techniques over a jamming-free service channel. By taking advantage of the multi-antenna gain provided the RSU and spatial diversity provided by the relay vehicles, the transmission reliability of all vehicles can be significantly improved under the threat of jamming attacks.

In summary, the main contributions of this paper are as follows:

- We propose a two-stage cooperative anti-jamming beamforming scheme for the control channel jamming issue in vehicular networks, which is modelled as a MINLP problem to characterize the selection of relays and beamformer design problem in the unified framework;
- We adopt the relaxation and approximation schemes based on the SDR and CCP methods to reformulate the MINLP problem with a sequence of convex sub-problems, which can be solved efficiently using an iterative procedure;
- We provide extensive simulation results to show the convergence as well as the significant

performance gain of the proposed scheme comparing with other benchmark schemes.

The rest of this chapter is organized as follows. We introduce the system model and formulated the cooperative anti-jamming beamformer design problem in Section 5.1 and 5.2 respectively, and some relaxation and approximation schemes are presented in Section 5.3. The performance of the proposed schemes are evaluated with simulation results in Section 5.4. Finally, conclusions are drawn for this paper in Section 5.5.

Notations: Uppercase bold letters denote matrices, whereas lowercase bold letters stand for column vectors. Let $\mathbb{C}$ represents the complex space and $(\cdot)^H$ represent the Hermitian (conjugate) transpose. The complex Gaussian distribution can be represented as $\mathcal{CN}(\cdot,\cdot)$. Let $\mathrm{tr}(\cdot)$ and $\mathrm{rank}(\cdot)$ denote the trace operator and the rank of a matrix respectively. The absolute of a scalar or the determinant of a square matrix is described as $|\cdot|$ and $\|\cdot\|_2$ stands for the Euclidean norm of a vector or matrix. For a square matrix $\mathbf{W}$, $\mathbf{W} \succeq \mathbf{0}$ means that $\mathbf{W}$ is positive semi-definite.

5.1 System Model

5.1.1 Network Model

In general, a vehicular network in the urban area consists of a set of RSUs deployed at the road sides communicating with the vehicles moving along the multi-lane road. In this paper, we focus on a segment of the road covered by a specific RSU, which is equipped with L antennas (Fig. 5.1). A single-antenna jammer and multiple single-antenna benign vehicles are located in the transmission range of the RSU. Without loss of generality, the RSU is assumed to be located at the origin with a transmission range of r meters, and thus the jammer v_m and vehicles are distributed within the range of $[-r, r]$.

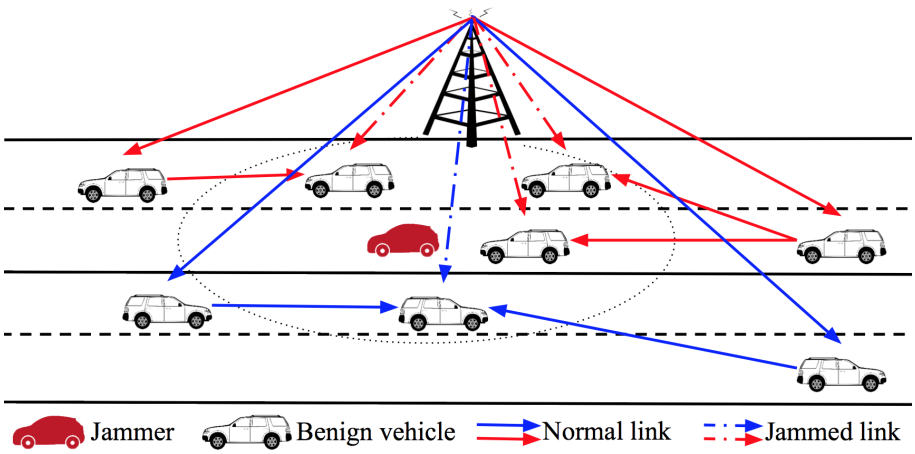


Figure 5.1: Vehicular network within the coverage area of a multi-antenna RSU.

In practice, due to the difference of speed limits or directions at different road lanes, the vehicles moving along different lanes may experience different road situations. Therefore, we assume that vehicles can be divided into n groups. The vehicles within the same group are expected to receive the same control messages from the RSU, and the control channel is shared by all n groups.

5.1.2 Anti-jamming Scheme

We assume that the vehicular network adopts a multi-channel protocol as standardized in IEEE 802.11p and 1609.4 protocols. The channel coordination scheme is illustrated in Fig. 5.2, whereby each synchronization interval consists of a control channel (CCH) interval and a service channel (SCH) interval. Based on this channel coordination scheme, we assume that the CCH is reserved for the I2V downlink transmission and the SCH is for the V2V communications and the V2I uplink transmission. Thus, in the CCH interval, all vehicles sense the control channel and receive control messages from the RSU, then in the SCH interval, each vehicle selects one of the six service channels to communicate with the RSU or other vehicles.

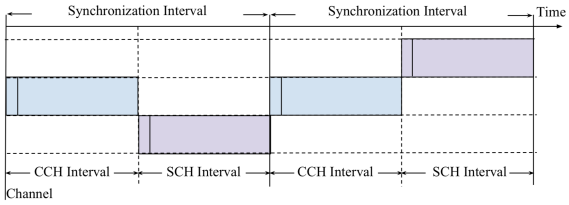


Figure 5.2: Channel coordination in vehicular networks.

In this paper, we consider the constant jamming scenario, whereby the jammer continuously emits interference signals in the control channel, which may result in the failure of decoding of the control channel signals at some nearby victim vehicles. Due to the fixed frequency of channels in vehicular networks, we believe that constant jamming may lead to a greater damage to the availability of the system, since the system cannot defend against the jamming attacks via retransmission. As a solution, we propose a two-stage anti-jamming mechanism based on the standardized multi-channel coordination scheme, which is motivated by the fact a single jammer may not block all vehicles within the cover range of the RSU in the CCH interval due to path loss and channel fading. Therefore, in the following SCH interval, those vehicles who have successfully decoded the control channel signals can serve as relays to cooperatively forward the signal to the victims belonging to the same group through a jamming free service channel using the coordinated beamforming technique. Note that the topology of the vehicular network is more stable compared with other mobile networks since the mobility of vehicles is limited by road space and road traffic density. Therefore, the topological relationship between the candidate relaying vehi-

cles and victim vehicles will remain unchanged within the same synchronization interval, so the two stages in this cooperative anti-jamming scheme can be considered jointly.

Let $\mathcal{V}_i$ and $\mathcal{R}_i$ denote the set of victim vehicles and candidate relaying vehicles in the i-th group respectively, then the set of vehicles in group i is given by $\mathcal{N}_i = \mathcal{V}_i + \mathcal{R}_i$. The symbols used in this paper are listed in Table 5.1.

Table 5.1: List of Notations

Symbol	Description
P_0	Transmit power of the RSU
$P_{i,k}$	Transmit power of vehicle k in group i
P_m	Transmit power of jammer
$\mathbf{u}_i$	beamformer from the RSU to vehicles in group i
$\mathbf{w}_i$	beamformer from all $N_i - 1$ vehicles to victim $v_{i,k}$ in group i
$\mathbf{h}_k$	Channel vector between the RSU and $v_{i,k}$
$\mathbf{g}_k$	Channel vector from all $N_i - 1$ vehicles to $v_{i,k}$ in group i
J_k	Channel gain between the jammer and $v_{i,k}$
$y_{i,k}$	Received signal at $v_{i,k}$ from the RSU
$\tilde{y}_{i,k}$	Received signal at $v_{i,k}$ from all $N_i - 1$ vehicles in group i
s_i	Signal to vehicles in group i
s_m	Jamming signal
n_k	Received noise signal at $v_{i,k}$
σ_k^2	Received noise power at $v_{i,k}$
$r_{i,k}$	Achievable data rate in the first stage
$\tilde{r}_{i,k}$	Achievable data rate in the second stage

5.1.3 Signal Model

According to the aforementioned relaying procedure, the anti-jamming scheme can be divided into two stages. In the first stage, the RSU transmits the control signals of all groups simultaneously using multi-group multicasting beamforming technique[44]. We assume that channels between the RSU and vehicles as well as between vehicles follow Rayleigh fading model. In addition, we assume that the channel state information (CSI) can be obtained using channel feedback or channel estimation schemes, such as those proposed in [93, 94]. Let $\mathbf{h}_k = [h_{1k}, h_{2k}, \ldots, h_{Lk}] \in \mathbb{C}^L$ denote the $L \times 1$ complex channel vector from L transmit antennas of the RSU to vehicle $k \in \mathcal{N}_i$,

which is assumed to remain constant within each time slot[1], but will vary from one time slot to another. $\mathbf{u_i} = [u_{1i}, u_{2i}, \ldots, u_{Li}] \in \mathbb{C}^L$ denote the beamforming weight vector of the antennas of the RSU for the ith group, and s_i is the symbol of group i. Then the signal received at vehicle $v_{i,k}$ (denoted as vehicle k in the i-th group) can be given as:

$$y_{i,k} = \mathbf{u}_i \mathbf{h}_k^H s_i + \sum_{p \neq i}^{n} \mathbf{u}_p \mathbf{h}_k^H s_p + \sqrt{P_m} J_k s_m + n_k, \tag{5.1}$$

where the first term is the signal desired by all vehicles in group i, s_i is the normalized signal symbol for group i, i.e., $\mathbb{E}[|s_i|^2] = 1$. The second term corresponds to the interference signals from other groups. The third term denotes the interference caused by the jammer, where s_m and P_m are the signal emitted by the jammer and its transmit power respectively, and J_k is the channel gain between the jammer and vehicle k. n_k is the additive white Gaussian noise at vehicle k, where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$.

Assuming the control signals, jamming signal and noise are statistically independent, then the signal-to-interference-plus-noise ratio (SINR) at vehicle $v_{i,k}$ can be obtained from (5.1) as:

$$\mathrm{SINR}_{i,k} = \frac{| \mathbf{u}_i \mathbf{h}_k^H |^2}{\sum_{p \neq i}^{n} | \mathbf{u}_p \mathbf{h}_k^H |^2 + P_m | J_k |^2 + \sigma_k^2}. \tag{5.2}$$

The achievable data rate $r_{i,k}$ at vehicle $v_{i,k}$ can be given by:

$$r_{i,k} = B \log(1 + \mathrm{SINR}_{i,k}), \tag{5.3}$$

where B is the channel bandwidth in hertz.

In the first stages, some vehicles may fail to decode the received signal from the RSU, which are denoted as victims. Then in the second stage, the set of vehicles who have successfully decoded the signals will be selected as relays to re-transmit the signals to the victims of the same group through a jamming-free service channel using the decode-and-forward (DF) strategy. Note that according to the DF strategy, a vehicle $v_{i,k}$ will be selected as relay if and only if it can successfully decode the signal from the RSU, or equivalently, its achievable data rate $r_{i,k}$ is larger than a threshold γ.

To exploit the spatial diversity of different relays and mitigate the co-channel interference from the relays of other groups, the coordinated beamforming technique is adopted by the relays in the re-transmission of the signals. Specifically, let $\mathbf{w}_{ik}$ denote the beamformer of the antenna of relay vehicle $v_{i,k}$, then $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{iN_i}] \in \mathbb{C}^{N_i}$ is the aggregate transmit beamforming weight vector of all relays in group i[2]. Let $\mathbf{g}_k$ denote the channel vector from all N_i vehicles to the k-th

[1] we assume the duration of each time slot is smaller than the channel coherence time.

[2] The transmit beamformer weight vector of the victim vehicles can be assumed to be zero as to be discussed in Section 5.3.1, which simplifies the presentation of the problem.

vehicle, that is, $\mathbf{g}_k = [\mathbf{g}_{1k}, \mathbf{g}_{2k}, \ldots, \mathbf{g}_{N_i k}] \in \mathbb{C}^{N_i}$. Then the received signal at a victim vehicle $v_{i,k}$ can by represented as:

$$\tilde{y}_{i,k} = \mathbf{w}_i \mathbf{g}_k^H s_i + \sum_{p \neq i}^{n} \mathbf{w}_p \mathbf{g}_k^H s_p + n_k. \tag{5.4}$$

Similar to (5.2), The SINR at a victim vehicle $v_{i,k}$ can be given by:

$$\widetilde{\mathrm{SINR}}_{i,k} = \frac{| \mathbf{w}_i \mathbf{g}_k^H |^2}{\sum_{p \neq i}^{n} | \mathbf{w}_p \mathbf{g}_k^H |^2 + \sigma_k^2}, \tag{5.5}$$

and the achievable data rate $\tilde{r}_{i,k}$ at a victim vehicle $v_{i,k}$ can be given as:

$$\tilde{r}_{i,k} = B \log(1 + \widetilde{\mathrm{SINR}}_{i,k}). \tag{5.6}$$

Similarly, $\tilde{r}_{i,k}$ should be larger than the threshold γ so that the victim vehicle can decode the signal received from the relays successfully.

5.2 Cooperative Anti-jamming Relay Beamforming Problem

According to the discussion in the previous section, although the relay set $\mathcal{R}_i$ and victim set $\mathcal{V}_i$ are not identified explicitly for all groups, it is specified that the achievable rate for either relay or victim vehicle should be larger than a prescribed threshold. Therefore, from anti-jamming's perspective, it is desirable to maximize the minimum achievable data rate of all vehicles under the interference of jammer in the control channel, which can be formally stated as the following beamformer design problem:

$$\max_{\{\mathbf{u}_i\},\{\mathbf{w}_i\}} \quad \min_{i \in \{1 \ldots n\}} \{\min\{\min_{k \in \mathcal{R}_i}\{r_{i,k}\}, \min_{k \in \mathcal{V}_i}\{\tilde{r}_{i,k}\}\}\} \tag{5.7}$$

$$s.t. \quad \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \leq P_0, \quad \forall i, \tag{5.7a}$$

$$\|\mathbf{w}_{ik}\|_2^2 \leq P_{i,k}, \quad \forall i, \forall k \in \mathcal{R}_i, \tag{5.7b}$$

$$\mathcal{R}_i \cap \mathcal{V}_i = \emptyset, \quad \forall i, \tag{5.7c}$$

where (5.7a) and (5.7b) correspond to the power constraints of the RSU and relays respectively. P_0 and $P_{i,k}$ represents the transmit power budgets of the RSU and vehicle $v_{i,k}$. (5.7c) specifies that each vehicle cannot be a relay or a victim simultaneously.

By introducing an auxiliary threshold γ for $r_{i,k}$ and $\tilde{r}_{i,k}$, problem (5.7) can be transformed to

the following equivalent form:

$$\begin{aligned}
\max \quad & \gamma && && (5.8)\\
s.t. \quad & \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \le P_0, && \forall i, && (5.8a)\\
& \|\mathbf{w}_{ik}\|_2^2 \le P_{i,k}, && \forall i, \forall k \in \mathcal{R}_i, && (5.8b)\\
& r_{i,k} \ge \gamma, && \forall i, \forall k \in \mathcal{R}_i, && (5.8c{:}1)\\
& \tilde{r}_{i,k} \ge \gamma, && \forall i, \forall k \in \mathcal{V}_i, && (5.8c{:}2)\\
& \mathcal{R}_i \cap \mathcal{V}_i = \emptyset, && \forall i && (5.8d)
\end{aligned}$$

where constraints (5.8c:1) implies that for a vehicle $v_{i,k}$ to be qualified as a relay, it has to satisfy the minimum rate constraint. Otherwise, if it is a victim, then constraint (5.8c:2) should be satisfied.

However, since each vehicle can only be a relay or a victim, so (5.8c:1) and (5.8c:2) should not be active simultaneously. To this end, we can introduce a binary indicator variable $x_{i,k} \in \{0, 1\}$ for each vehicle $v_{i,k}$, where $x_{i,k} = 1$ specifies that vehicle $v_{i,k}$ is selected as a relay (which requires $r_{i,k} \ge \gamma$), while $x_{i,k} = 0$ implies that vehicle $v_{i,k}$ is a victim (which requires $\tilde{r}_{i,k} \ge \gamma$). In this way, constraints (5.8c:1), (5.8c:2) and (5.8d) can be unified as a disjunctive set, which consists of two disjunctions separated by the or ($\bigvee$) operator and negation ($\neg$) operator:

$$\begin{bmatrix} x_{i,k} \\ r_{i,k} \ge \gamma \end{bmatrix} \bigvee \begin{bmatrix} \neg x_{i,k} \\ \tilde{r}_{i,k} \ge \gamma \end{bmatrix} \tag{5.9}$$

which specifies that if $x_{i,k} = 1$, the first inequality applies, otherwise (i.e., $\neg x_{i,k}$), the second inequality should be satisfied. Using (5.9), only one of the constraints in (5.8c:1) and (5.8c:2) will be activated based on the value of $x_{i,k}$.

This disjunctive set can be relaxed using the big-M formulation [95, 96] and transformed into the following constraints:

$$\begin{aligned}
& \gamma - r_{i,k} \le M_1(1 - x_{i,k}), && \forall i, \forall k \in \mathcal{N}_i,\\
& \gamma - \tilde{r}_{i,k} \le M_2 x_{i,k}, && \forall i, \forall k \in \mathcal{N}_i, && (5.10)
\end{aligned}$$

where M_1 and M_2 are two sufficiently large parameters, which can be given by the maximum achievable rates in both stages:

$$\begin{aligned}
M_1 &= B\log(1 + \frac{P_0 \lambda_{\max}[\mathbf{h}_k^H \mathbf{h}_k]}{n_k^2}) && \forall i, \forall k \in \mathcal{N}_i,\\
M_2 &= B\log(1 + \frac{P_{i,k} \lambda_{\max}[\mathbf{g}_k^H \mathbf{g}_k]}{n_k^2}) && \forall i, \forall k \in \mathcal{N}_i, && (5.11)
\end{aligned}$$

where $\lambda_{\max}[\,\cdot\,]$ denotes the largest eigenvalue of a matrix.

By substituting constraints (5.8c:1), (5.8c:2) and (5.8d) with (5.10), problem (5.8) can be reformulated as follows:

$$\max \quad \gamma \tag{5.12}$$

$$s.t. \quad \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \leq P_0, \qquad \forall i, \tag{5.12a}$$

$$\|\mathbf{w}_{ik}\|_2^2 \leq x_{i,k} P_{i,k}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{5.12b}$$

$$\gamma - r_{i,k} \leq M_1(1 - x_{i,k}), \quad \forall i, \forall k \in \mathcal{N}_i, \tag{5.12c:1}$$

$$\gamma - \tilde{r}_{i,k} \leq M_2 x_{i,k}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{5.12c:2}$$

$$x_{i,k} \in \{0, 1\}, \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{5.12d}$$

which is a mixed-integer non-linear programming (MINLP) problem, and the optimal solution is difficult to obtain in polynomial time. In the following, we will propose some approximation and relaxation schemes to transform this problem into a tractable form, and thus it can be solved efficiently.

5.3 Problem Approximation and Relaxation

5.3.1 Smoothed ℓ_0-Norm Approximation

In (5.12), $x_{i,k}$ indicates the identity of vehicle $v_{i,k}$, and specifies that the constrain (5.12c:1) should apply if vehicle $v_{i,k}$ is selected as relay (i.e., $x_{i,k} = 1$); otherwise (i.e., $\neg x_{i,k}$), (5.12c:2) should be satisfied since it is a victim.

On the other hand, if vehicle $v_{i,k}$ is a victim, then its transmit beamformer $\mathbf{w}_{ik}$ should be zero since it will not participate in the cooperative relaying in the second stage (it is a receiver instead). Therefore, $x_{i,k}$ can be connected with $\mathbf{w}_{ik}$ as follows:

$$x_{i,k} = \begin{cases} 0, & \|\mathbf{w}_{ik}\|_2^2 = 0, \quad \forall k \in \mathcal{N}_i, \\ 1, & \|\mathbf{w}_{ik}\|_2^2 > 0, \quad \forall k \in \mathcal{N}_i, \end{cases} \tag{5.13}$$

which can be further represented as the ℓ_0-norm of $\|\mathbf{w}_{i,k}\|_2^2$ as follows:

$$x_{i,k} = \left\| \|\mathbf{w}_{ik}\|_2^2 \right\|_0. \tag{5.14}$$

Unfortunately, the ℓ_0-norm is a non-convex discontinuous function. As a solution, we can approximate the discontinuous ℓ_0-norm of $\|\mathbf{w}_{i,k}\|_2^2$ with a continues smooth concave function as follows [45]:

$$x_{i,k} \approx f_\theta(\|\mathbf{w}_{ik}\|_2^2) = 1 - \exp(-\frac{\|\mathbf{w}_{ik}\|_2^2}{\theta}), \tag{5.15}$$

where $\theta > 0$ is a parameter to control the smoothness of approximation. In general, a smaller θ leads to a better approximation result, and a larger θ normally lead to a smoother approximation. More details about the setting of θ can be found in[46].

In this way, problem (5.12) can be approximated as follows using the smoothing function:

$$\max \quad \gamma \tag{5.16}$$

$$s.t. \quad \sum_{i=1}^{n} \|\mathbf{u}_i\|_2^2 \leq P_R, \qquad \forall i, \tag{5.16a}$$

$$\|\mathbf{w}_{ik}\|_2^2 \leq (1 - \exp(-\frac{\|\mathbf{w}_{ik}\|_2^2}{\theta}))P_{i,k}, \quad \forall i, \forall k \in \mathcal{N}_i, \tag{5.16b}$$

$$\gamma - r_{i,k} \leq M_1 \exp(-\frac{\|\mathbf{w}_{ik}\|_2^2}{\theta}), \qquad \forall i, \forall k \in \mathcal{N}_i, \tag{5.16c:1}$$

$$\gamma - \tilde{r}_{i,k} \leq M_2(1 - \exp(-\frac{\|\mathbf{w}_{ik}\|_2^2}{\theta})), \quad \forall i, \forall k \in \mathcal{N}_i. \tag{5.16c:2}$$

The remaining challenges of (5.16) are due to the non-concave achievable rates $r_{i,k}$ and $\tilde{r}_{i,k}$. In the following two subsections, we introduce two different approximation techniques to address this problem.

5.3.2 Semi-definite Relaxation

To address the non-concavity of $r_{i,k}$ and $\tilde{r}_{i,k}$, we adopt the semi-definite relaxation (SDR) method proposed in [47]. Specifically, we introduce two set of matrices based on $\mathbf{u}_i$ and $\mathbf{w}_k$ as follows: $\{\mathbf{U}_i := \mathbf{u}_i^H \mathbf{u}_i\}_{i=1}^n$ and $\{\mathbf{W}_i := \mathbf{w}_i^H \mathbf{w}_i\}_{i=1}^n$, where $\mathbf{U}_i$ and $\mathbf{W}_i$ are $L \times L$ and $N_i \times N_i$ complex matrices respectively. Note that $\mid \mathbf{u}_i \mathbf{h}_k^H \mid^2 = \mathrm{tr}(\mathbf{U}_i \mathbf{h}_k^H \mathbf{h}_k)$ and $\mid \mathbf{w}_i \mathbf{g}_k^H \mid^2 = \mathrm{tr}(\mathbf{W}_i \mathbf{g}_k^H \mathbf{g}_k)$, if and only if $\mathbf{U}_i \succeq 0$, $\mathrm{rank}(\mathbf{U}_i) = 1$ and $\mathbf{W}_i \succeq 0$, $\mathrm{rank}(\mathbf{W}_i) = 1$, where $\mathrm{tr}(\cdot)$ is the trace operator. In particular, let $\mathbf{W}_{i,k}$ denote the kth item in the main diagonal of $\mathbf{W}_i$, then $\mathbf{W}_{i,k} = \|\mathbf{w}_{ik}\|_2^2$.

In this way, $r_{i,k}$ can be represented as:

$$\begin{aligned} r_{i,k} =& B \log(1 + \frac{\mathrm{tr}(\mathbf{U}_i \mathbf{h}_k \mathbf{h}_k^H)}{\sum_{p \neq i}^{n} \mathrm{tr}(\mathbf{U}_p \mathbf{h}_k \mathbf{h}_k^H) + P_m \mid J_k \mid^2 + \sigma_k^2}) \\ =& B \log(\sum_{i=1}^{n} \mathrm{tr}(\mathbf{U}_i \mathbf{h}_k \mathbf{h}_k^H) + P_m \mid J_k \mid^2 + \sigma_k^2) - B \log(\sum_{p \neq i}^{n} \mathrm{tr}(\mathbf{U}_p \mathbf{h}_k \mathbf{h}_k^H) + P_m \mid J_k \mid^2 + \sigma_k^2) \\ =& y_k(\mathbf{U}_i) - z_k(\mathbf{U}_p), \end{aligned} \tag{5.17}$$

where $y_k(\cdot)$ and $z_k(\cdot)$ are concave with respect to matrices $\mathbf{U}_i$ and $\mathbf{U}_p$.

Similarly, $\tilde{r}_{i,k}$ can be represented as:

$$\begin{aligned}\tilde{r}_{i,k} =& B\log(1+\frac{\mathrm{tr}(\mathbf{W}_i\mathbf{g}_k\mathbf{g}_k^H)}{\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{W}_p\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2})\\ =& B\log(\sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_i\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2)-B\log(\sum_{p\neq i}^{n}\mathrm{tr}(\mathbf{W}_p\mathbf{g}_k\mathbf{g}_k^H)+\sigma_k^2)\\ =& \tilde{y}_k(\mathbf{W}_i)-\tilde{z}_k(\mathbf{W}_p),\end{aligned} \tag{5.18}$$

where $\tilde{y}_k(\cdot)$ and $\tilde{z}_k(\cdot)$ are concave with respect to matrices $\mathbf{W}_i$ and $\mathbf{W}_p$.

Dropping the non-convex rank-one constraints for $\mathbf{U}_i$ and $\mathbf{W}_i$, problem (5.12) can be relaxed as follows:

$$\max \quad \gamma \tag{5.19}$$

$$s.t. \quad \mathrm{tr}(\mathbf{U}_i)-P_0\leq 0, \quad \forall i, \tag{5.19a}$$

$$\mathbf{W}_{i,k}-(1-\exp(-\frac{\mathbf{W}_{i,k}}{\theta}))P_{i,k}\leq 0, \qquad \forall i,\forall k\in\mathcal{N}_i \tag{5.19b}$$

$$\gamma-y_k(\mathbf{U}_i)+z_k(\mathbf{U}_p)-M_1\exp(-\frac{\mathbf{W}_{i,k}}{\theta})\leq 0, \quad \forall i,\forall k\in\mathcal{N}_i, \tag{5.19c:1}$$

$$\gamma-\tilde{y}_k(\mathbf{W}_i)+\tilde{z}_k(\mathbf{W}_p)-M_2(1-\exp(-\frac{\mathbf{W}_{i,k}}{\theta}))\leq 0, \quad \forall i,\forall k\in\mathcal{N}_i, \tag{5.19c:2}$$

$$\mathbf{U}_i\succeq 0 \quad and \quad \mathbf{W}_i\succeq 0, \quad \forall i. \tag{5.19d}$$

5.3.3 Convex-concave Procedure

Note that (5.19c:1) and (5.19c:2) are the difference of convex (DC) constraints. In general, an optimization problem with DC constraints can be stated as follows:

$$\begin{aligned}\max \quad & \gamma \\ s.t. \quad & \gamma-z_i(x)+y_i(x)\leq 0, \quad i=1,\ldots,n,\end{aligned} \tag{5.20}$$

where y_i and z_i are all convex functions.

This kind of DC problem can be solved using the convex-concave procedure (CCP) [48]. The basic idea is to write each non-linear function as the sum of a convex and concave function, then convexify the problem by replacing the concave part by their first order Taylor expansions, then solve a sequence of convex sub-problems successively. Specifically, the CCP starts with an initial point x_0, and in each iteration t solves a convex sub-problem:

$$\begin{aligned}\max \quad & \gamma \\ s.t. \quad & \gamma-[z_i(x^{(t)})+\nabla z_i(x^{(t)})^T(x-x^{(t)})]+y_i(x)\leq 0, \quad \forall i,\end{aligned} \tag{5.21}$$

where $x^{(t)}$ is the optimal solution obtained in the previous iteration.

In the following, we discuss the solution of problem (5.19) based on CCP scheme under two different kinds of scenarios: single group ($n=1$) and multi-group ($n>1$).

Single-Group Scenarios In this case, there is no interference term in the denominator of the SINR, and thus problem (5.19) can be simplified as follows:

$$\max \quad \gamma \tag{5.22}$$

$$s.t. \quad \mathrm{tr}(\mathbf{U}) - P_0 \le 0, \tag{5.22a}$$

$$\mathbf{W}_k - (1 - \exp(-\frac{\mathbf{W}_k}{\theta}))P_k \le 0, \quad \forall k \in \mathcal{N}, \tag{5.22b}$$

$$\gamma - B\log(1 + \frac{\mathrm{tr}(\mathbf{U}\mathbf{h}_k\mathbf{h}_k^H)}{P_m \mid J_k \mid^2 + \sigma_k^2}) - M_1 \exp(-\frac{\mathbf{W}_k}{\theta}) \le 0, \quad \forall k \in \mathcal{N}, \tag{5.22c:1}$$

$$\gamma - B\log(1 + \frac{\mathrm{tr}(\mathbf{W}\mathbf{g}_k\mathbf{g}_k^H)}{\sigma_k^2}) - M_2(1 - \exp(-\frac{\mathbf{W}_k}{\theta})) \le 0, \quad \forall k \in \mathcal{N}, \tag{5.22c:2}$$

$$\mathbf{U} \succeq 0 \quad \textit{and} \quad \mathbf{W} \succeq 0. \tag{5.22d}$$

where $\mathbf{W}_{i,k}$ in problem (5.19) is replaced with $\mathbf{W}_k$ since there is only one group.

Let $d_\theta(\mathbf{W}_k) = \exp(-\frac{\mathbf{W}_k}{\theta})$, then its first order Taylor expansion can be given by:

$$D_\theta(\mathbf{W}_k|\tilde{\mathbf{W}}_k) = \exp(-\frac{\tilde{\mathbf{W}}_k}{\theta}) - 1/\theta \exp(-\frac{\tilde{\mathbf{W}}_k}{\theta})(\mathbf{W}_k - \tilde{\mathbf{W}}_k), \tag{5.23}$$

where $\tilde{\mathbf{W}}_k$ is an initial point in each iteration. Thus, problem (5.22) can be transformed into a sequence of convex sub-problems as follows:

$$\max \quad \gamma \tag{5.24}$$

$$s.t. \quad \mathrm{tr}(\mathbf{U}) - P_0 \le 0, \tag{5.24a}$$

$$\mathbf{W}_k - (1 - \exp(-\frac{\mathbf{W}_k}{\theta}))P_k \le 0, \quad \forall k \in \mathcal{N}, \tag{5.24b}$$

$$\gamma - B\log(1 + \frac{\mathrm{tr}(\mathbf{U}\mathbf{h}_k\mathbf{h}_k^H)}{P_m \mid J_k \mid^2 + \sigma_k^2}) - M_1 D_\theta(\mathbf{W}_k|\tilde{\mathbf{W}}_k) \le 0, \quad \forall k \in \mathcal{N}, \tag{5.24c:1}$$

$$\gamma - B\log(1 + \frac{\mathrm{tr}(\mathbf{W}\mathbf{g}_k\mathbf{g}_k^H)}{\sigma_k^2}) - M_2(1 - \exp(-\frac{\mathbf{W}_k}{\theta})) \le 0, \quad \forall k \in \mathcal{N}, \tag{5.24c:2}$$

$$\mathbf{U} \succeq 0 \quad \textit{and} \quad \mathbf{W} \succeq 0. \tag{5.24d}$$

Multi-Group Scenarios In this case, both $z_k(\mathbf{U}_p)$ and $\tilde{z}_k(\mathbf{W}_p)$ in (5.19c:1) and (5.19c:2) are concave functions. Their first order Taylor expansions can be represented as:

$$\begin{aligned} Z_k(\mathbf{U}_p|\tilde{\mathbf{U}}_p) = \; & B\log(\sum_{p\neq i}^{n} \mathrm{tr}(\tilde{\mathbf{U}}_p\mathbf{h}_k\mathbf{h}_k^H) + P_m \mid J_k \mid^2 + \sigma_k^2) \\ & + \frac{B\sum_{p\neq i}^{n}(\mathrm{tr}(\mathbf{U}_p\mathbf{h}_k\mathbf{h}_k^H) - \mathrm{tr}(\tilde{\mathbf{U}}_p\mathbf{h}_k\mathbf{h}_k^H))}{(\sum_{p\neq i}^{n}\mathrm{tr}(\tilde{\mathbf{U}}_p\mathbf{h}_k\mathbf{h}_k^H) + P_m \mid J_k \mid^2 + \sigma_k^2)\ln 2} \end{aligned} \tag{5.25}$$

and

$$\begin{aligned} \tilde{Z}_k(\mathbf{W}_p|\tilde{\mathbf{W}}_p) = \; & B\log(\sum_{p\neq i}^{n} \mathrm{tr}(\tilde{\mathbf{W}}_p\mathbf{g}_k\mathbf{g}_k^H) + \sigma_k^2) \\ & + \frac{B\sum_{p\neq i}^{n}(\mathrm{tr}(\mathbf{W}_p\mathbf{g}_k\mathbf{g}_k^H) - \mathrm{tr}(\tilde{\mathbf{W}}_p\mathbf{g}_k\mathbf{g}_k^H))}{(\sum_{p\neq i}^{n}\mathrm{tr}(\tilde{\mathbf{W}}_p\mathbf{g}_k\mathbf{g}_k^H) + \sigma_k^2)\ln 2} \end{aligned} \tag{5.26}$$

respectively, where $\tilde{\mathbf{U}}_p$ and $\tilde{\mathbf{W}}_p$ are the initial points in each iteration.

In this way, problem (5.19) can be transformed into a sequence of convex sub-problems as follows:

$$\max \quad \gamma \tag{5.27}$$

$$s.t. \quad \mathrm{tr}(\mathbf{U}_i) - P_0 \le 0, \quad \forall i, \tag{5.27a}$$

$$\mathbf{W}_{i,k} - (1 - \exp(-\frac{\mathbf{W}_{i,k}}{\theta}))P_{i,k} \le 0, \qquad \forall i, \forall k \in \mathcal{N}_i \tag{5.27b}$$

$$\gamma - h_k(\mathbf{U}_i) + Z_k(\mathbf{U}_p|\tilde{\mathbf{U}}_p) - M_1 D_\theta(\mathbf{W}_{i,k}|\tilde{\mathbf{W}}_{i,k}) \le 0, \quad \forall i, p, \forall k \in \mathcal{N}_i, \tag{5.27c:1}$$

$$\gamma - h_k(\mathbf{W}_i) + \tilde{Z}_k(\mathbf{W}_p|\tilde{\mathbf{W}}_p)(1 - \exp(-\frac{\mathbf{W}_{i,k}}{\theta})) \le 0, \quad \forall i, p, \forall k \in \mathcal{N}_i, \tag{5.27c:2}$$

$$\mathbf{U}_i \succeq 0 \quad and \quad \mathbf{W}_i \succeq 0, \quad \forall i. \tag{5.27d}$$

The CCP algorithm is a powerful heuristic method to find the local optimal solution for the DC problem by solving a sequence of sub-problems. The optimal solution in each iteration is always a feasible point of the original problem, which guarantees the convergence of CCP as proven in [97][98].

5.3.4 Randomization Method

Recalling the non-convex rank-one constraints are ignored in (5.19) using the SDR scheme. Due to this relaxation, the rank-one property may not be held by the matrices $\{\mathbf{U}_i^*\}_{i=1}^n$ and $\{\mathbf{W}_i^*\}_{i=1}^n$ obtained from problem (5.24) or (5.27). If they are rank one, then their principal components is the optimal solution to the original problem, the optimal beamforming weight vector $\{\mathbf{u}_i\}_{i=1}^n$ and $\{\mathbf{w}_i\}_{i=1}^n$ can be obtained directly by adopting the eigenvalue decomposition (i.e., $\mathbf{u}_i = \sqrt{\lambda_i}\mathbf{u}_i^*$ and $\mathbf{w}_i = \sqrt{\mu_i}\mathbf{w}_i^*$, where $\mathbf{U}_i^* = \lambda_i \mathbf{u}_i^* \mathbf{u}_i^{*H}$ and $\mathbf{W}_i^* = \mu_i \mathbf{w}_i^* \mathbf{w}_i^{*H}$). Otherwise, an approximate to the original problem can be found using randomization techniques [44, 99]. The basic idea is to generate beamforming vectors $\{\mathbf{u}_i\}_{i=1}^n$ and $\{\mathbf{w}_i\}_{i=1}^n$ from the optimum solution matrices $\{\mathbf{U}_i^*\}_{i=1}^n$ and $\{\mathbf{W}_i^*\}_{i=1}^n$ and choose the one can satisfy the transmit power constraints and maximize the link rate γ. Specifically, let $\mathbf{u}_i^c$ and $\mathbf{w}_i^c$ denote a candidate beamforming vector for the i-th group, and the eigen-decomposition of the optimal matrices are given by $\mathbf{U}_i^* = \mathbf{X}_i \mathbf{\Sigma}_i \mathbf{X}_i^H$ and $\mathbf{W}_i^* = \mathbf{Y}_i \mathbf{\Psi}_i \mathbf{Y}_i^H$ respectively. Thus, $\mathbf{u}_i^c$ and $\mathbf{w}_i^c$ can be given by $\mathbf{u}_i^c = \mathbf{X}_i \mathbf{\Sigma}_i^{1/2} \hat{\mathbf{u}}^c$ and $\mathbf{w}_i^c = \mathbf{Y}_i \mathbf{\Psi}_i^{1/2} \hat{\mathbf{w}}^c$, where $\hat{\mathbf{u}}^c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\hat{\mathbf{w}}^c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Therefore, it can be guaranteed that $\mathbb{E}[\mathbf{u}_i^c (\mathbf{u}_i^c)^H] = \mathbf{U}_i^*$ and $\mathbb{E}[\mathbf{w}_i^c (\mathbf{w}_i^c)^H] = \mathbf{W}_i^*$.

5.3.5 Algorithm Design and Complexity Analysis

By incorporating the aforementioned relaxation and approximation schemes, we design a CCP-based Optimal Relay Beamforming (ORBF) algorithm as shown in Algorithm 5, which proceeds

iteratively to solve the convex sub-problems based on the CCP scheme. Finally, the beamformers $\{\mathbf{u}_i\}$ and $\{\mathbf{w}_i\}$ can be found based on the obtained matrices $\{\mathbf{U}_i^*\}$ and $\{\mathbf{W}_i^*\}$ using the EVD or randomization method.

Algorithm 5 CCP-based ORBF Algorithm

Inputs: P_0, $P_{i,k}$, $\mathbf{h}_k$, $\mathbf{g}_k$, $\forall i, k \in \mathcal{N}_i$;

Outputs: γ, $\mathbf{u}_i$, $\mathbf{w}_i$, $\forall i$;

Initialization: Feasible initial matrices $\{\tilde{\mathbf{U}}_i\}$ and $\{\tilde{\mathbf{W}}_i\}$, $\forall i$;

repeat

Calculate $D_\theta(\mathbf{W}_{i,k}|\tilde{\mathbf{W}}_{i,k})$, $Z_k(\mathbf{U}_p|\tilde{\mathbf{U}}_p)$ and $\tilde{Z}_k(\mathbf{W}_p|\tilde{\mathbf{W}}_p)$ according to (5.23), (5.25) and (5.26), and obtain sub-problem (5.27) at points $\{\tilde{\mathbf{U}}_p\}$, $\{\tilde{\mathbf{W}}_{i,k}\}$ and $\{\tilde{\mathbf{W}}_p\}$;

Solve sub-problem (5.27) and obtain $\{\mathbf{U}_i^*\}$ and $\{\mathbf{W}_i^*\}$;

Let $\{\tilde{\mathbf{U}}_i\} \leftarrow \{\mathbf{U}_i^*\}$ and $\{\tilde{\mathbf{W}}_i\} \leftarrow \{\mathbf{W}_i^*\}$;

until convergence

if $\text{rank}(\mathbf{U}_i^*) = 1, \forall i$ and $\text{rank}(\mathbf{W}_i^*) = 1, \forall i$ **then**

Obtain $\{\mathbf{u}_i^*\}$ and $\{\mathbf{w}_i^*\}$ via the eigen-decomposition method;

else

Obtain $\{\mathbf{u}_i^*\}$ and $\{\mathbf{w}_i^*\}$ via the randomization method;

end if

Let $\mathbf{u}_i \leftarrow \mathbf{u}_i^*$ and $\mathbf{w}_i \leftarrow \mathbf{w}_i^*$.

return γ, $\mathbf{u}_i$, $\mathbf{w}_i$.

The main computational complexity of the proposed algorithm comes from the complexity of solving the convex sub-problem in each iteration. For a n-group scenario, the complexity of solving convex sub-problem (5.27) is the polynomial time about the problem size, which can be given as $O(\sum_{i=1}^{n}[L^2N_i + R_i^2(N_i - R_i)])$, where R_i denotes the number of vehicles in group i serving as relays in the second stage, which should be no more than N_i. Moreover, as the number of groups increases, solving this problem requires introducing more variables, which makes the problem size larger and more complicated to solve.

5.4 Simulation Results

In previous section, we solve the original optimization problem in (5.8) using the disjunctions and Big-M reformulation techniques. The binary variables are relaxed into continuous variable using the smooth function, which may affect the accuracy of the optimization results and increases the complexity of the algorithm. In this section, we provide simulation results to evaluate the per-

formance of the proposed ORBF scheme. In the simulations, the Urban MObility (SUMO) [100] simulator is used to generate traffic flows for urban scenario, where parameters are set based on real-time urban traffic. These traffic flow data are used as inputs to obtain the simulation results using CVX in Matlab. CVX is a modeling framework for disciplined convex programming, which turns Matlab into a modeling language, allowing constraints and objectives to be specified using standard Matlab expression syntax [101, 102]. In addition, Rayleigh fading model is adopted in our simulations. Each fading parameter is a product of two parts, path loss $d^{-\alpha/2}$ and multi-path fading h , where the multi-path fading h is a zero mean complex Gaussian variable, d denotes the distance between the RSU and one vehicle or the distance between two vehicles for the I2V and V2V communications respectively. And α denotes the path loss exponent in urban area. The parameter settings in the simulation are summarized in Table 5.2:

Table 5.2: Parameters Used in Simulations

Parameter	Value
Simulation Environment	Urban
RSU coverage	1 km
Street Width	2 Lanes
Vehicle Velocity	40 - 60 km/h
Number of Vehicles	5-20
Fading model	Rayleigh
Transit power (RSU, Vehicles)	60mw and 20mw
Noise power	-95 dBm

In order to make a fair comparison, we also propose a heuristic relay beamforming (HRBF) algorithm for problem (5.8) as a benchmark scheme. The basic idea is to incrementally update the identity of each vehicle (as relay or victim). Specifically, at beginning, we solve a multi-group multicast beamformer design problem assuming all vehicles receive data directly from the RSU. Based on the obtained results, the vehicle with the worst SINR is classified as a victim, while the rest vehicles are identified as relays. Then given the identity of each vehicle $v_{i,k}$ (i.e, $x_{i,k}$ is fixed),

the original problem in (5.8) can be formulated as follows:

$$\begin{aligned}
\max \quad & \gamma && && (5.28)\\
s.t. \quad & \|\mathbf{u}_i\|_2^2 \le P_0, && \forall i, && (5.28a)\\
& \|\mathbf{w}_{i,k}\|_2^2 \le x_{i,k} P_{i,k}, && \forall i,k, && (5.28b)\\
& r_{i,k} \ge x_{i,k}\gamma, && \forall i,k, && (5.28c:1)\\
& \tilde{r}_{i,k} \ge (1 - x_{i,k})\gamma, && \forall i,k, && (5.28c:2)
\end{aligned}$$

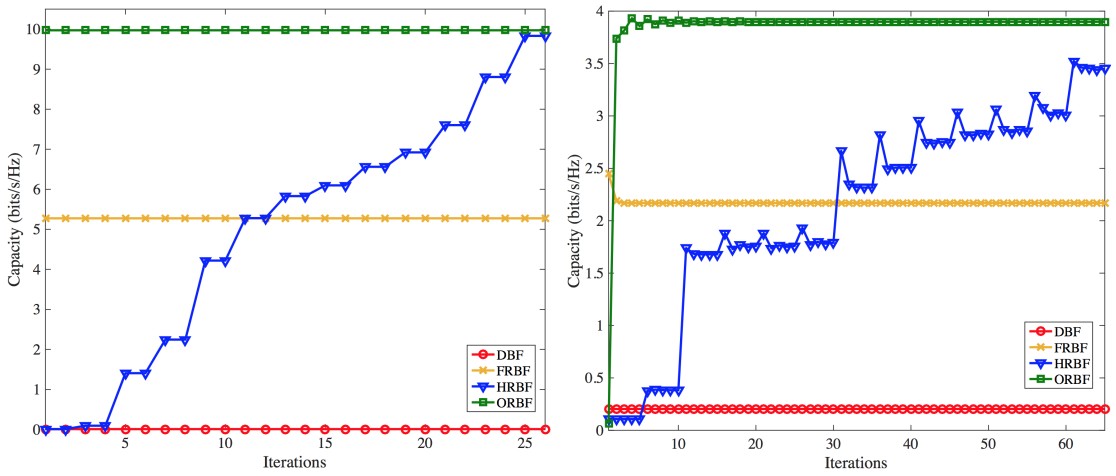
The non-convexity of $r_{i,k}$ and $\tilde{r}_{i,k}$ in this problem can be addressed using the SDR and CCP schemes as proposed in previous section. By solving this optimization problem, the relay vehicle with the worst SINR is selected into the victim set of the corresponding group, and problem (5.28) is solved again based on the updated identifies of all vehicles. This process is repeated until the objective function cannot be improved.

In addition to the ORBF and HRBF schemes, we also consider the following anti-jamming schemes in the simulations:

- Direct Beamforming (DBF): In this scheme, the RSU transmits the signal directly to all vehicles of all groups simultaneously using the multi-group multicast beamforming technique;
- Fixed Relay Beamforming (FRBF): This scheme is an extension of the anti-jamming scheme presented in our previous work [103] to the multi-antenna RSU scenarios, whereby a vehicle is classified as a victim if its received SINR from the RSU is smaller than a prescribed threshold, and thus the set of relays are determined in advance based on the transmit power and channel conditions between RSU, jammer and vehicles.

We consider both single-group and multi-group scenarios in the simulations. Due to high computational complexity of the multi-group scenarios, we only provide results for the case of two groups. Note that our proposed anti-jamming scheme is designed to take advantage of multi-antenna and multi-user diversity provided by the RSU and nearby vehicles in two stage, so in simulations we firstly fix the number of antennas at the RSU and the number of vehicles to investigate the effectiveness of these four schemes under different channel conditions in terms of the convergence rate and network capacity. We then study the performance of these scheme under different network settings.

In Fig. 5.3, we demonstrate the convergence behaviour of these four different schemes in the single-group and multi-group scenarios. It can be seen that in both scenarios, except for the HRBF scheme, all other schemes converge quickly to the stationary point. Additionally, it can be observed that the convergence curves in the multi-group scenario are not monotonic, this because



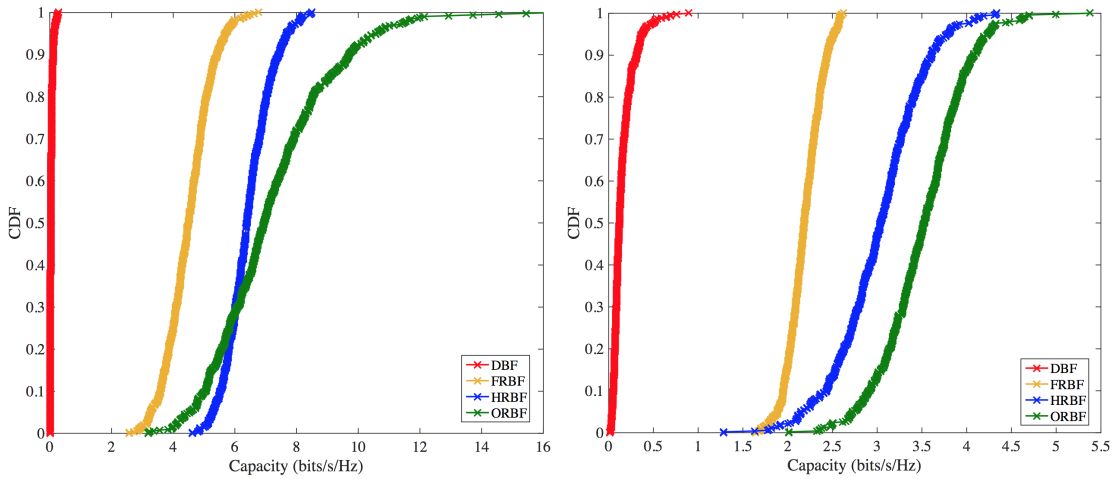
(a) Single-group Scenario ($n = 1, N = 15$) (b) Multi-group Scenario ($n = 2, N_1 = N_2 = 8$)

Figure 5.3: Convergence of four different algorithms in different scenarios.

in the CCP procedure, we relaxed the the original function using its first-order Taylor expansion and solve it iteratively. In this process, the solutions obtained in the previous iterations do not satisfy the conditions of the equation, but the whole process can gradually converge. Comparing the computational complexity of the four algorithms, for FRBF, HRBF and ORBF, as mentioned in Section V, the main computational complexity comes from the complexity of solving the convex sub-problem in each iteration. Compared with the ORBF, FRBF converges faster because the simulator (CXV in our case) introduces fewer variables when solving the fixed relay problems. Thus, the problem size is smaller.

Moreover, the trade-off between complexity and performance can also be observed. Compared with the DBF and FRBF scenarios, the achieved network capacity is significantly improved using the proposed ORBF scheme. The HRBF scheme has also a better performance when compared with the DBF and FRBF schemes. However, it takes longer time to determine the suitable relay set. It can be seen since the bottleneck relay is put into the victim set sequentially, and thus the network capacity increases gradually as the number of relay nodes decreases. Eventually, most of the vehicles are selected as victims (13 out of 15 vehicles in the single-group scenario, and 12 out of 16 vehicles in the multi-group scenario).

In Fig. 5.4, we show the cumulative distribution function (CDF) of the network capacity achieved by these four schemes under different channel conditions and in Table. 5.3 their mean values, which are obtained based on 500 different scenarios. It can be seen that in both scenarios, the proposed ORBF scheme outperforms all other three schemes, and the HRBF scheme can achieve suboptimal performance comparing with the ORBF scheme, especially in the single-group scenario. In particular, it can be observed that using the DBF scheme, the network capacity drops



(a) Single-group Scenario (b) Multi-group Scenario $n = 2$

Figure 5.4: CDF of the achieved network capacity under different channel status.

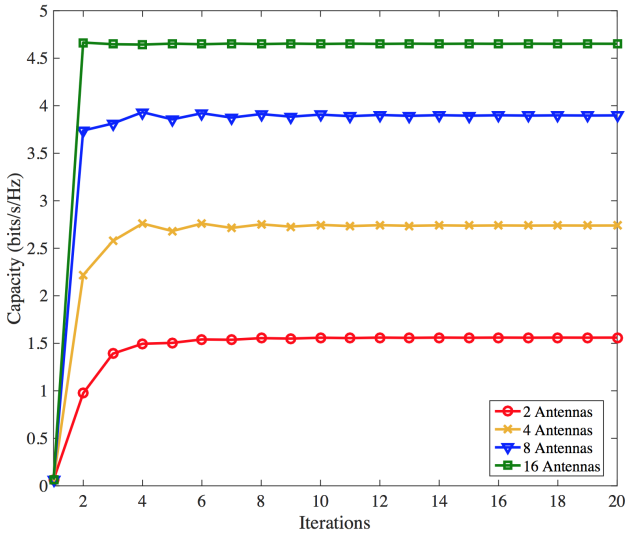


Figure 5.5: Convergence behaviour of the ORBF scheme under two-group scenario.

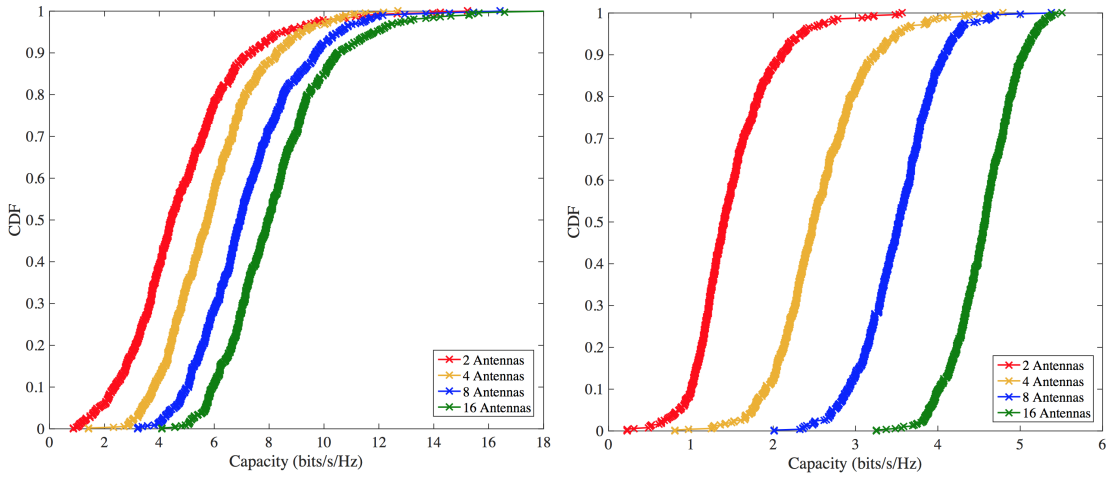
below 1 bit/s/Hz, which demonstrates the significant impact of the jamming attack in the control channel.

Table 5.3: Mean Network Capacity with Different Scheme

	DBF	FRBF	HRBF	ORBF
Single-group	0.0486	4.4960	5.7068	7.2228
Multi-group ($n = 2$)	0.1542	2.1896	2.9204	3.5156

We investigate the impact of the number of antennas at the RSU on the convergence of the proposed ORBF scheme. As illustrated in Fig. 5.5, the ORBF scheme can converge within about 10 iterations, which suggests that the convergence behaviour of the proposed scheme does not

change too much under different the number of antennas.



(a) Single-group Scenario

(b) Multi-group Scenario $n = 2$

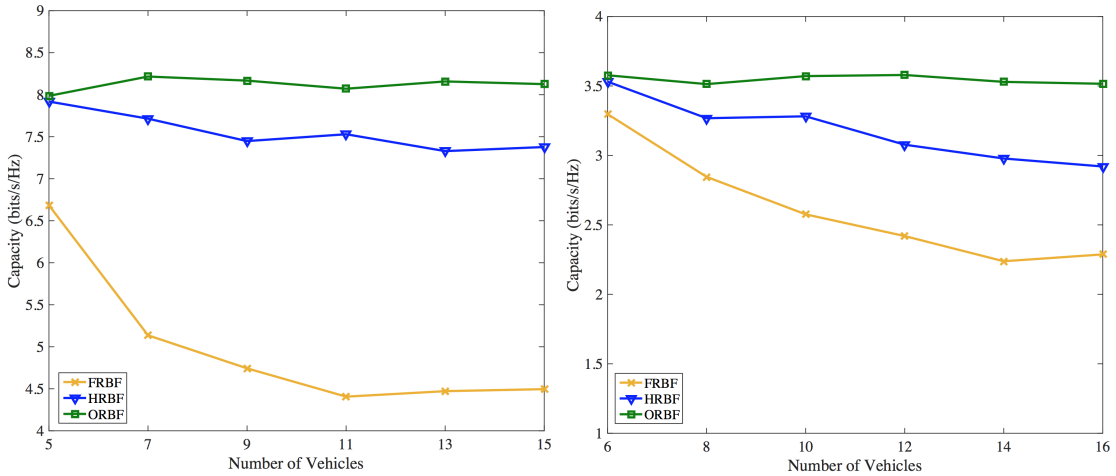
Figure 5.6: The CDF of the network capacity achieved by the ORBF scheme with different number of antennas.

In Fig. 5.6, we plot the CDF of the network capacity achieved by the ORBF scheme with different number of antennas and in Table. 5.4 their mean values, wherein each curve is obtained based on 500 different scenarios. It can be clearly observed that in both singe-group and multi-group scenarios, the achieved network capacity increases with the number of antennas, especially in multi-group scenarios, which is reasonable since the more antennas, the more diversity can be exploited as a countermeasure of the jamming attacks.

Table 5.4: Mean Network Capacity with Different Number of Antennas

	2 Antennas	4 Antennas	8 Antennas	16 Antennas
Single-group	4.7772	5.8969	7.2228	8.1773
Multi-group ($n = 2$)	1.4852	2.5498	3.5156	4.4698

Finally, we investigate the network capacity with different number of vehicles. We consider three different schemes. The number of antennas is fixed as $L = 8$, and the number of vehicles varies from 5 to 15 in the single-group scenarios, from 6 to 16 in the multi-group scenarios. In Fig. 5.7, each point is the average results of 100 simulations. It can be observed that for the network capacity achieved by the ORBF remains stable under different conditions, and the capacity achieved by the DBF remains under 0.5 bits/s/Hz due to the tremendous impact of the jamming signal in the control channel. Meanwhile, as the number of vehicles increases, the average network capacity achieved by the FRBF and HRBF schemes decreases gradually. The reason is when the number



(a) Single-group Scenario

(b) Multi-group Scenario $n = 2$

Figure 5.7: Average network capacity vs. number of vehicles.

of vehicles is small, it is possible for the FRBF and HRBF schemes to select the same relays as the ORBF scheme. However, these two schemes fail to find the optimal relays with the increase of the number of vehicles, so the performance gaps are increased significantly.

5.5 Conclusion

In this chapter, a two stage cooperative relay beamforming scheme was presented for the control channel jamming problem in vehicular networks, which takes advantage of multi-antenna and spatial diversity provided by the RSU and relaying vehicles in two stages respectively. This problem is formulated as a MINLP problem to unify the selection of relays, as well as the beaformer design for the RSU and relays. In order to solve this intractable problem, we adopted the SDR and CCP methods to approximate it with a sequence of convex sub-problems, which can be solved efficiently using the CCP based algorithm. Simulation results show that the proposed scheme is guaranteed to converge to the stationary point quickly, and the network capacity achieved by our scheme is significantly improved comparing with other benchmark schemes.

For future work, we will consider the anti-jamming beamforming design problem under the assumption of partial channel station information, which can significantly reduce the overhead for channel measurement. This work is focused on the downlink transmissions (from the RSU to vehicles), it would be interesting to consider the uplink transmissions whereby the RSU is the target of the jamming attacks.

Chapter 6

Conclusion and Future Work

6.1 Thesis Summary

This thesis has been dedicated to addressing the misbehaviour detection problem in vehicular networks. Specifically, we focus on two major issues in PHY layer and application layer respectively: Radio Frequency (RF) Jamming attacks and Sybil attacks.

Specifically, we adopted three different machine learning methods including Distance based clustering, Support Vector Machine (SVM) and k-nearest neighbours (kNN) in Sybil nodes detection. Based on variation between benign vehicles and Sybil nodes in their driving patterns, the non-existent virtual nodes can be detected.

For RF jamming attacks, we focused on the design of countermeasure for the control channel jamming issue in vehicular networks, which is of vital importance to the safety of I2V communications. We proposed to adopt the cooperative relaying techniques to address the control channel jamming problem in vehicular networks, which is based on the idea that the vehicles outside of the jamming area can serve as relays to help forward the control channel signal to the victim vehicles through other the jamming-free service channels. In this way, a virtual multi-antenna system can be formed by multiple relays nodes to cooperatively serve the victim vehicles, and thus the transmission reliability can be effectively improved by exploiting the spatial diversity of these relay nodes.

Thus, we extended the jamming issues in multi-antenna RSU scenarios, where the RSU can serve multiple groups of vehicles simultaneously using the multi-group multicast beamforming technique. As a solution, we propose a two stage anti-jamming scheme, whereby the vehicles who have successfully decoded the signal received in the first stage will be selected as relays to cooperatively serve the victim vehicles in the second stage using the coordinated beamforming techniques over a jamming-free service channel. By taking advantage of the multi-antenna gain

provided the RSU and spatial diversity provided by the relay vehicles, the transmission reliability of all vehicles can be significantly improved under the threat of jamming attacks.

6.2 Future Work

In this thesis, we focused on the downlink transmissions (from the RSU to vehicles) anti-jamming. Thus, for future work, it would be interesting to consider the uplink transmissions whereby the RSU is the target of the jamming attacks. Meanwhile, we will consider the anti-jamming beamforming design problem under the assumption of partial channel station information, which can significantly reduce the overhead for channel measurement, or precoding beamforming problem.

Uplink Anti-jamming

According to the IEEE 802.11p standard, six services channels are available around 5.9 GHz frequency, and multi-channel operations are specified by IEEE 1609 standards, whereby vehicles can communicate with the RSU using any of these six channels, but are only allowed to operate in one of them at each time. If a jammer launches a jamming attack over the service channels, a portion of the reports or situation updates from vehicles to the RSU will be blocked. As a consequence, the RSU does not have full knowledge of the vehicles within its coverage area, and the following updates from the RSU to vehicles will be incomplete.

In this case, channel switching (or hopping) is an effective way to circumvent the jamming issue since multiple service channels are available. On the other hand, since the jammer cannot block all six service channels simultaneously, it may adopt different jamming strategies to adapt its jamming channel so as to maximize its utility (or damage) to different channels.

Admission Control and Precoding Beamforming

If the traffic density is high, for example, when the number of vehicles is much larger than the number of antennas equipped on the RSU, the efficiency of beamforming would drop dramatically since the cochannel interference (CCI) between users cannot be totally eliminated. In this way, it is necessary to adapt a more dynamic admission control scheme, wherein the users with near-orthogonal channels are grouped in the same group, and different groups are served in different timeslots according to the TDMA scheme.

Meanwhile, in these scenarios, transmit beamforming and receive combining could be costly. Thus, a predetermined codebook is considered can well reduce the complexity of beamforming design where the receiver only sends the label of the best beamforming vector in a predetermined codebook to the transmitter [104, 105].

Bibliography

[1] "IEEE Guide for Wireless Access in Vehicular Environments (WAVE) - Architecture," *IEEE Std 1609.0-2013*, pp. 1–78, March 2014.

[2] "Intelligent Transport Systems (ITS);Communications Architecture," *ETSI EN 302 665 V1.1.1*, pp. 1–44, Spt 2010.

[3] "IEEE Standard for Wireless Access in Vehicular Environments Security Services for Applications and Management Messages," *IEEE Std 1609.2-2013 (Revision of IEEE Std 1609.2-2006)*, pp. 1–289, April 2013.

[4] Q. Wu, J. Domingo-Ferrer, and U. Gonzalez-Nicolas, "Balanced trustworthiness, safety, and privacy in vehicle-to-vehicle communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 559–573, Feb 2010.

[5] J. R. Douceur, "The Sybil Attack," in *Proceedings of 1st International Workshop on Peer-to-Peer Systems (IPTPS)*, 2002.

[6] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil attack in sensor networks: analysis defenses," in *Information Processing in Sensor Networks, Third International Symposium on*, April 2004.

[7] S. Capkun, L. Buttyan, and J. P. Hubaux, "Self-organized public-key management for mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 1, pp. 52–64, Jan 2003.

[8] T. Zhou, R. R. Choudhury, P. Ning, and K. Chakrabarty, "P2DAP -; Sybil Attacks Detection in Vehicular Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 3, pp. 582–594, 2011.

[9] A. Cheng and E. Friedman, "Sybilproof Reputation Mechanisms," in *Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-peer Systems*. New York, NY, USA: ACM, 2005, pp. 128–132.

[10] B. Xiao, B. Yu, and C. Gao, "Detection and Localization of Sybil Nodes in VANETs," in *Proceedings of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks*, ser. DIWANS '06. New York, NY, USA: ACM, 2006, pp. 1–8.

[11] B. Yu, C.-Z. Xu, and B. Xiao, "Detecting Sybil attacks in VANETs," *Journal of Parallel and Distributed Computing*, vol. 73, no. 6, 2013.

[12] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and Localizing Identity-Based Attacks in Wireless and Sensor Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, Jun 2010.

[13] C. Piro, C. Shields, and B. N. Levine, "Detecting the Sybil Attack in Mobile Ad hoc Networks," in *Securecomm and Workshops*, Aug 2006.

[14] S. Yan, R. Malaney, I. Nevat, and G. W. Peters, "Optimal Information-Theoretic Wireless Location Verification," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 7, pp. 3410–3422, Sept 2014.

[15] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of Spread-Spectrum Communications - A Tutorial," *IEEE Transactions on Communications*, vol. 30, no. 5, pp. 855–884, May 1982.

[16] R. Ramanathan, "On the Performance of Ad Hoc Networks With Beamforming Antennas," in *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'01)*, Long Beach, California USA, October 2001.

[17] G. Noubir, "On Connectivity in Ad Hoc Networks under Jamming Using Directional Antennas and Mobility," in *Wired/Wireless Internet Communications: Second International Conference, WWIC 2004, Frankfurt (Oder), Germany, February 4-6, 2004. Proceedings*, Berlin, Heidelberg, 2004, pp. 186–200.

[18] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications - a comprehensive literature survey," *IEEE Communications Surveys Tutorials*, vol. 11, no. 2, pp. 87–105, Second 2009.

[19] Q. Wang, S.-P. Sheng, J. Abernethy, and M. Liu, "Jamming Defense Against a Resource-Replenishing Adversary in Multi-channel Wireless Systems," in *2014 12th International*

Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), May 2014, pp. 210–217.

[20] E. V. Belmega and A. Chorti, "Protecting Secret Key Generation Systems Against Jamming: Energy Harvesting and Channel Hopping Approaches," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2611–2626, Nov 2017.

[21] K. Pelechrinis, M. Iliofotou, and S. V. Krishnamurthy, "Denial of Service Attacks in Wireless Networks: The Case of Jammers," *IEEE Communications Surveys Tutorials*, vol. 13, no. 2, Second 2011.

[22] M. Li, I. Koutsopoulos, and R. Poovendran, "Optimal Jamming Attacks and Network Defense Policies in Wireless Sensor Networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 1307–1315.

[23] W. Xu, T. Wood, W. Trappe, and Y. Zhang, "Channel Surfing and Spatial Retreats: Defenses Against Wireless Denial of Service," in *Proceedings of the 3rd ACM Workshop on Wireless Security*, ser. WiSe '04. New York, NY, USA: ACM, 2004, pp. 80–89.

[24] A. D. Wood, J. A. Stankovic, and S. H. Son, "JAM: A Jammed-Area Mapping Service for Sensor Networks," in *RTSS 2003. 24th IEEE Real-Time Systems Symposium, 2003*, Dec 2003, pp. 286–297.

[25] W. Xu, W. Trappe, and Y. Zhang, "Anti-jamming Timing Channels for Wireless Networks," in *Proceedings of the First ACM Conference on Wireless Network Security*, ser. WiSec '08. New York, NY, USA: ACM, 2008, pp. 203–213. [Online]. Available: http://doi.acm.org/10.1145/1352533.1352567

[26] Y. Liu, P. Ning, H. Dai, and A. Liu, "Randomized Differential DSSS: Jamming-Resistant Wireless Broadcast Communication," in *2010 Proceedings IEEE INFOCOM*, March 2010, pp. 1–9.

[27] S. Liu, L. Lazos, and M. Krunz, "Thwarting Control-Channel Jamming Attacks from Inside Jammers," *IEEE Transactions on Mobile Computing*, vol. 11, no. 9, pp. 1545–1558, Sept 2012.

[28] J. M. Becker, J. D. Lohn, and D. S. Linden, "An anti-jamming beamformer optimized using evolvable hardware," in *2011 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS 2011)*, Nov 2011, pp. 1–5.

[29] J. Becker, J. D. Lohn, and D. Linden, "An in-situ optimized anti-jamming beamformer for mobile signals," in *Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation*, July 2012, pp. 1–2.

[30] S. Gollakota, F. Adib, D. Katabi, and S. Seshan, "Clearing the RF Smog: Making 802.11N Robust to Cross-technology Interference," in *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11. New York, NY, USA: ACM, 2011, pp. 170–181.

[31] W. Shen, P. Ning, X. He, H. Dai, and Y. Liu, "MCR Decoding: A MIMO approach for defending against wireless jamming attacks," in *2014 IEEE Conference on Communications and Network Security*, Oct 2014, pp. 133–138.

[32] Q. Yan, H. Zeng, T. Jiang, M. Li, W. Lou, and Y. T. Hou, "MIMO-based jamming resilient communication in wireless networks," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 2697–2706.

[33] M. Cagalj, S. Capkun, and J. p. Hubaux, "Wormhole-Based Antijamming Techniques in Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 1, pp. 100–114, Jan 2007.

[34] H. Mustafa, X. Zhang, Z. Liu, W. Xu, and A. Perrig, "Jamming-Resilient Multipath Routing," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 852–864, Nov 2012.

[35] G. Zheng, E. A. Jorswieck, and B. Ottersten, "Cooperative Communications against Jamming with Half-Duplex and Full-Duplex Relaying," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[36] L. Zhang, Z. Guan, and T. Melodia, "Cooperative anti-jamming for infrastructure-less wireless networks with stochastic relaying," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 549–557.

[37] Y. Li, L. Xiao, J. Liu, and Y. Tang, "Power control Stackelberg game in cooperative anti-jamming communications," in *The 2014 5th International Conference on Game Theory for Networks*, Nov 2014, pp. 1–6.

[38] S. Bharati and W. Zhuang, "CAH-MAC: Cooperative ADHOC MAC for Vehicular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 470–479, September 2013.

[39] J. F. C. Joseph, B. S. Lee, A. Das, and B. C. Seet, "Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FDA," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 2, pp. 233–245, March 2011.

[40] H. Wu, R. Fujimoto, and G. Riley, "Analytical models for information propagation in vehicle-to-vehicle networks," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 6, Sept 2004, pp. 4548–4552 Vol. 6.

[41] K. Abboud and W. Zhuang, "Stochastic Analysis of a Single-Hop Communication Link in Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2297–2307, Oct 2014.

[42] M. Haenggi, *Stochastic Geometry for Wireless Networks*, 1st ed. New York, NY, USA: Cambridge University Press, 2012.

[43] G. Nigam, P. Minero, and M. Haenggi, "Coordinated Multipoint Joint Transmission in Heterogeneous Networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4134–4146, Nov 2014.

[44] E. Karipidis, N. D. Sidiropoulos, and Z. Q. Luo, "Quality of Service and Max-Min Fair Transmit Beamforming to Multiple Cochannel Multicast Groups," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, March 2008.

[45] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-Centric Sparse Multicast Beamforming for Cache-Enabled Cloud RAN," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6118–6131, Sept 2016.

[46] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed *ell*^{0} Norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, Jan 2009.

[47] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit Beamforming for Physical-Layer Multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, June 2006.

[48] A. L. Yuille and A. Rangarajan, "The Concave-Convex Procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.

[49] The mission and objectives of the CAR 2 CAR Communication Consortium. [Online]. Available: https://www.car-2-car.org/index.php?id=5

[50] French Minister of Transport SCOOP Project. [Online]. Available: http://www.scoop.developpement-durable.gouv.fr/en/

[51] M. N. Mejri, J. Ben-Othman, and M. Hamdi, "Survey on VANET security challenges and possible cryptographic solutions ," *Vehicular Communications*, vol. 1, no. 2, pp. 53 – 66, 2014.

[52] R. G. Engoulou, M. Bellaiche, S. Pierre, and A. Quintero, "VANET Security Surveys ," *Computer Communications*, vol. 44, pp. 1 – 13, 2014.

[53] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, and J. P. Hubaux, "Secure vehicular communication systems: design and architecture," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 100–109, November 2008.

[54] E. Schoch and F. Kargl, "On the efficiency of secure beaconing in vanets," in *Proceedings of the Third ACM Conference on Wireless Network Security*, ser. WiSec '10. New York, NY, USA: ACM, 2010, pp. 111–116.

[55] "Intelligent Transport Systems (ITS); Security; Trust and Privacy Management," *ETSI TS 102 941 V1.1.1*, pp. 1–30, Juin 2012.

[56] "Intelligent Transport Systems (ITS); Security; Security header and certificate formats," *ETSI TS 103 097 V1.1.1*, pp. 1–33, April 2013.

[57] J. Petit, F. Schaub, M. Feiri, and F. Kargl, "Pseudonym Schemes in Vehicular Networks: A Survey," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 1, pp. 228–255, Firstquarter 2015.

[58] G. Yan, S. Olariu, J. Wang, and S. Arif, "Towards Providing Scalable and Robust Privacy in Vehicular Networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 7, pp. 1896–1906, July 2014.

[59] Q. Li, A. Malip, K. Martin, S.-L. Ng, and J. Zhang, "A Reputation-Based Announcement Scheme for VANETs," *Vehicular Technology, IEEE Transactions on*, vol. 61, no. 9, pp. 4095–4108, Nov 2012.

[60] B. Ostermaier, F. Dotzer, and M. Strassberger, "Enhancing the Security of Local Danger-Warnings in VANETs - A Simulative Analysis of Voting Schemes," in *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*, April 2007, pp. 422–431.

[61] S. Chang, Y. Qi, H. Zhu, J. Zhao, and X. Shen, "Footprint: Detecting Sybil Attacks in Urban Vehicular Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 6, pp. 1103–1114, June 2012.

[62] K. Rabieh, M. M. E. Mahmoud, T. N. Guo, and M. Younis, "Cross-layer scheme for detecting large-scale colluding Sybil attack in VANETs," in *2015 IEEE International Conference on Communications (ICC)*.

[63] "IEEE Standard for Information technology–Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," *IEEE Std 802.11p-2010 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11n-2009, and IEEE Std 802.11w-2009)*, pp. 1–51, July 2010.

[64] X. Chen, H. H. Refai, and X. Ma, "A Quantitative Approach to Evaluate DSRC Highway Inter-Vehicle Safety Communication," in *IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference*, Nov 2007, pp. 151–155.

[65] C. Campolo and A. Molinaro, "Multichannel communications in vehicular Ad Hoc networks: a survey," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 158–169, May 2013.

[66] H. M. Wang, M. Luo, Q. Yin, and X. G. Xia, "Hybrid Cooperative Beamforming and Jamming for Physical-Layer Security of Two-Way Relay Networks," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2007–2020, Dec 2013.

[67] F. Zhu, F. Gao, M. Yao, and H. Zou, "Joint Information- and Jamming-Beamforming for Physical Layer Security With Full Duplex Base Station," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6391–6401, Dec 2014.

[68] H. M. Wang, F. Liu, and M. Yang, "Joint Cooperative Beamforming, Jamming, and Power Allocation to Secure AF Relay Systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4893–4898, Oct 2015.

[69] M. Beko, "Efficient Convex Optimization for Beamforming in Cognitive Radio Multicast Transmission," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 1895–1899.

[70] Z. Xiang, M. Tao, and X. Wang, "Coordinated Multicast Beamforming in Multicell Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 12–21, January 2013.

[71] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted Fair Multicast Multigroup Beamforming Under Per-antenna Power Constraints," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5132–5142, Oct 2014.

[72] O. Tervo, L. N. Tran, H. Pennanen, S. Chatzinotas, M. Juntti, and B. Ottersten, "Energy-efficient coordinated multi-cell multi-group multicast beamforming with antenna selection," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017, pp. 1209–1214.

[73] L. J. Rodriguez, N. H. Tran, T. Q. Duong, T. Le-Ngoc, M. Elkashlan, and S. Shetty, "Physical layer security in wireless cooperative relay networks: state of the art and beyond," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 32–39, Dec 2015.

[74] G. Chen, Z. Tian, Y. Gong, Z. Chen, and J. A. Chambers, "Max-Ratio Relay Selection in Secure Buffer-Aided Cooperative Wireless Networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 719–729, April 2014.

[75] O. Punal, C. Pereira, A. Aguiar, and J. Gross, "Experimental Characterization and Modeling of RF Jamming Attacks on VANETs," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 2, Feb 2015.

[76] Y. O. Basciftci, F. Chen, J. Weston, R. Burton, and C. E. Koksal, "How Vulnerable Is Vehicular Communication to Physical Layer Jamming Attacks?" in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Sept 2015, pp. 1–5.

[77] A. Benslimane and H. Nguyen-Minh, "Jamming Attack Model and Detection Method for Beacons Under Multichannel Operation in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6475–6488, July 2017.

[78] C. Ding and X. He, "Cluster merging and splitting in hierarchical clustering algorithms," in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 2002, pp. 139–146.

[79] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, "Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering," in *Proceedings of the 2008 ACM*

Conference on Recommender Systems, ser. RecSys '08. New York, NY, USA: ACM, 2008, pp. 259–266. [Online]. Available: http://doi.acm.org/10.1145/1454008.1454048

[80] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ser. COLT '92. New York, NY, USA: ACM, 1992, pp. 144–152.

[81] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, Jun. 1998.

[82] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and Localization of Multiple Spoofing Attackers in Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 1, Jan 2013.

[83] W. Hu, J. Gao, Y. Wang, O. Wu, and S. Maybank, "Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection," *IEEE Transactions on Cybernetics*, vol. 44, Jan 2014.

[84] D. O. Loftsgaarden and C. P. Quesenberry, "A Nonparametric Estimate of a Multivariate Density Function," *Ann. Math. Statist.*, vol. 36, no. 3, pp. 1049–1051, 06 1965.

[85] K. Fukunaga and L. Hostetler, "K-nearest-neighbor Bayes-risk Estimation," *IEEE Trans. Inf. Theor.*, vol. 21, no. 3, pp. 285–293, Sep. 2006.

[86] "Intelligent Transport Systems (ITS);Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service," *ETSI EN 302 637-2 V1.3.1*, pp. 1–44, Spt 2014.

[87] "Intelligent Transport Systems (ITS);Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service," *ETSI EN 302 637-3 V1.2.1*, Spt 2014.

[88] R. G. Engoulou, M. Bellaïche, S. Pierre, and A. Quintero, "VANET security surveys ," *Computer Communications*, vol. 44, pp. 1 – 13, 2014.

[89] Y. Hao, J. Tang, and Y. Cheng, "Cooperative Sybil Attack Detection for Position Based Applications in Privacy Preserved VANETs," in *Global Telecommunications Conference, 2011 IEEE*, Dec 2011, pp. 1–5.

[90] O. Berman, Z. Drezner, and D. Krass, "Cooperative cover location problems: The planar case," *IIE Transactions*, vol. 42, no. 3, pp. 232–246, 2009.

[91] F. Chiti, R. Fantacci, and G. Rigazzi, "A mobility driven joint clustering and relay selection for IEEE 802.11p/WAVE vehicular networks," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 348–353.

[92] P. Salvo, F. Cuomo, A. Baiocchi, and I. Rubin, "Probabilistic relay selection in timer-based dissemination protocols for VANETs," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 2725–2730.

[93] T. A. Thomas and F. W. Vook, "Method for Obtaining Full Channel State Information for RF Beamforming," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 3496–3500.

[94] R. A. Stoica, S. Severi, and G. T. F. de Abreu, "Learning the Vehicular Channel Through the Self-Organization of Frequencies," in *2015 IEEE Vehicular Networking Conference (VNC)*, Dec 2015, pp. 68–75.

[95] C. Hua and R. Zheng, "Robust Topology Engineering in Multiradio Multichannel Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 492–503, March 2012.

[96] F. Trespalacios and I. Grossmann, "Improved Big-M reformulation for generalized disjunctive programs," vol. 76, 05 2015.

[97] G. R. Lanckriet and B. K. Sriperumbudur, "On the Convergence of the Concave-Convex Procedure," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds., 2009, pp. 1759–1767.

[98] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, Jun 2016.

[99] S. Zhang and Y. Huang, "Complex Quadratic Optimization and Semidefinite Programming," *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 871–890, 2006.

[100] D. Krajzewicz, "Traffic Simulation with SUMO – Simulation of Urban Mobility," in *Fundamentals of Traffic Simulation*. New York, NY: Springer New York, 2010, pp. 269–293.

[101] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014.

[102] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[103] P. Gu, C. Hua, R. Khatoun, Y. Wu, and A. Serhrouchni, "Cooperative Anti-Jamming Relaying for Control Channel Jamming in Vehicular Networks," in *2017 IEEE Global Communications Conference: Ad Hoc and Sensor Networks (Globecom2017 AHSN)*, Singapore, Singapore, Dec. 2017.

[104] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, Oct 2003.

[105] S. Park, W. Seo, S. Choi, and D. Hong, "A Beamforming Codebook Restriction for Cross-Tier Interference Coordination in Two-Tier Femtocell Networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1651–1663, May 2011.

Détection des Comportements Malveillants dans les réseaux véhiculaires

Pengwenlong GU

ABSTRACT: For vehicular networks, the safety and security of the network architecture and protocols are of vital importance, which have been the central theme of the standardization in both USA and Europe. In general, the provided security services are based on three major mechanisms: Encryption algorithms, Public Key Infrastructure (PKI) and Pseudonymous. These security services provide basic protection for the privacy of users and integrity of messages in vehicular environments. However, several critical issues, such as identity management, message traceability and availability still exist in vehicular networks. In this thesis, we focus on two major security issues: Sybil attack and radio frequency (RF) jamming attacks. Ranging from theoretical modelling and analysis, to practical algorithm design and optimisation.

Specifically, we focus on the Sybil attack detection in vehicular networks based on the vehicle driving patterns. Relying on beacon information, we designed a data format Driving Pattern Matrix (DPM) to describe vehicle driving pattern within a time period. Thus, three different machine learning methods: Distance based clustering, Support Vector Machine (SVM) and k-nearest neighbours (kNN) are considered. For RF Jamming attacks, we propose a cooperative relaying scheme to circumvent the control channel jamming problem in the vehicular networks, whereby the vehicles outside of the jamming area serve as relays to help forward the received control channel signal to the victim vehicles through another jamming-free service channel. Thus, we extend the anti-jamming problem into multi-antenna RSU scenarios and propose a two stage anti-jamming scheme for the control channel jamming issue in vehicular networks, which takes advantage of the multi-antenna diversity and spatial diversity provided by the RSU and relay vehicles to improve the transmission reliability of the victim vehicles.

KEY-WORDS: Vehicular networks, Sybil attacks, Machine learning, Radio Frequency(RF) jamming, Cooperative relaying, Cooperative beamforming

TELECOM ParisTech

EDITE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY