EFFICIENT SECURE DATA CLUSTERING IN VEHICULAR AD HOC NETWORKS

A Thesis submitted in fulfillment of the requirement for the award of the degree of

DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By

Rasmeet Singh Bali
(Registration No: 951203002)

Under the guidance of

Dr. Neeraj Kumar
Associate Professor
Computer Science & Engineering
Thapar University,
Patiala (Punjab), India

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY, PATIALA – 147004

OCTOBER 2016

CERTIFICATE

I, Rasmeet Singh Bali, Regn.No. 951203002, hereby declares that the thesis entitled **"Efficient Secure Data Clustering in Vehicular Ad Hoc Networks"** submitted to the Department of Computer Science & Engineering at Thapar University, Patiala, Punjab, India is an authenticated record of my own work for the award of the degree of "Doctor of Philosophy" under the supervision of Dr. Neeraj Kumar. This report has not been submitted to any other institution for award of any degree.

Rasmeet Singh Bali
951203002

Place: Patiala
Date: 14-10-2016

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Approved by:

Dr. Neeraj Kumar,
(Supervisior)
Associate Professor,
Department of CSE,
Thapar University,Patiala.

Abstract

Over the last few years, Vehicular Ad Hoc Networks (VANETs) have emerged as a new class of efficient information dissemination technology among communities of users mainly because of their wide range of applications in different domains such as safety applications, healthcare, data dissemination and online entertainment. Vehicles in VANETs act as intelligent machines to provide various resources to the end users with/without the aid of the existing infrastructure. The development of dedicated standards used for communication such as DSRC and WAVE have resulted in designing an Intelligent Transport System (ITS) which offers a number of promising solutions for efficient traffic management to increase passenger safety. However, future ITS solutions may require enhanced and robust message delivery solutions due to the high mobility and varying density of vehicles on the road.

The availability of limited communication resources due to high mobility and uneven distribution of vehicles in VANETs makes it difficult for the vehicles to maintain end-to-end connection for timely delivery of messages. Clustering can be one of the viable solutions to solve the aforementioned problem to have a better network throughput. It is grouping of the vehicles based upon metrics such as density, velocity and geographical locations of vehicles. But, there are number of challenges that needs to be addressed for designing an efficient solution for clustering. Most of the existing solutions reported in the literature use a combination of parameters to depict driver behavior for an optimized message delivery. But in these solutions, due to large number of nodes and lack of routers, a flat network scheme may cause serious scalability and hidden terminal problems. Vehicles require additional infrastructure like a GPS, transceivers, Lane Detection System, Digital maps, RSU's, odometer etc. for cluster formation and maintenance. These facts motivate us to analyze and develop new clustering techniques to improve the network stability with reliable communication among vehicles.

To address the issue of efficient resource utilization, so that available resources could be utilized for other network management applications, a predictive clustering scheme has been designed. Future mobility of vehicles on the road is predicted by a novel future mobility prediction algorithm which assists in clustering and determining the member vehicles within a cluster. The proposed algorithms estimate the clustering duration to determine the number of vehicles in a cluster based on an average predictive variation algorithm. These algorithms have been extensively studied using simulation by varying the number of vehicles and cluster durations in comparison to existing schemes. The predictive clustering scheme has been further improved

by incorporating learning automata into the prediction process. A Predictive Clustering Algorithm using Learning Automata (PCALA) is proposed in the designed solution. The learning automata stationed on vehicles are used to estimate future positions of the vehicles more accurately. The actions of the automata are rewarded or penalized based upon their current prediction accuracy and their previous actions. Extensive simulations are performed to evaluate the performance of the proposed scheme with respect to various metrics. Results obtained confirm the effectiveness of the PCALA in comparison to predictive clustering scheme.

Due to dynamically changing topology, wireless medium, and lack of centralized monitoring points, information related to vehicular applications can be altered or misused. These security breaches can lead to disastrous results such as loss of life or financial frauds. Therefore, some security mechanisms need to be implemented in VANETs for enhancing security. With this objective, a Cloud Based Distributive Intrusion Detection System (DIDS) is proposed for detecting attacks in VANETs. To secure communication among vehicles, a standard HMAC based cryptography technique is used. The viability of the proposed scheme is measured through simulations and the test results show its adaptability in real-time environment.

Internet-enabled devices have the capabilities of computing as well as communication to provide ease to a number of applications for the end users. The significance of this type of environment can be further enhanced by inclusion of vehicular clustering. However, security is one of the major concerns as devices communicate with one another using different protocols which are susceptible to various types of attacks. To address these issues, we propose novel secure clustering algorithms for efficient data dissemination between vehicles. A new trust metric based on dynamically varying transmission is defined for trust computation among the different devices which is evaluated both at local, and global levels. This trust metric is used to establish the current security level of vehicles and is the key parameter for creating secure clusters. Algorithms for performing secure clustering and trust establishment are designed in the proposed scheme. The performance of the proposed scheme is evaluated with respect to different evaluation metrics in various network scenarios. The results obtained clearly depict satisfactory performance of the proposed scheme in vehicular environment.

ACKNOWLEDGMENTS

First and foremost, thanks to almighty God for all his blessings without which nothing of my work would have been possible. I am highly indebted to Prof. (Dr.) Prakash Gopalan, Director, Thapar University, Patiala, Prof. (Dr.) R. S. Kaler, Deputy Director, and Dr. Susheel Mittal, Deputy Director, Thapar University, Patiala, for providing me the opportunity to pursue my course work and research. My deep regards to Dr. O.P. Pandey, Dean of Research and Sponsored Projects for his enormous help in completion of my research work. This thesis would have not been possible without the enabled guidance and keen interest of my supervisor Dr. Neeraj Kumar, Associate Professor, Department of Computer Science and Engineering, Thapar University, Patiala. The extensive discussions with him have always made me to stay on the right track. He has always been patient and cooperative whenever his guidance and expertise was needed. He helped me not only by sparing his valuable time, but also analytically reviewing my experimental setup, publications, reports and presentation from time to time. The constructive criticism and motivation by my supervisor has groomed me to my present shape. I would also like to express my gratitude to Dr. Maninder Singh, Head, Computer Science and Engineering Department, Thapar University, Patiala for providing me the continuous feedback and motivation throughout my work. I am highly thankful to my doctoral committee members Dr. Anil Kumar Verma, Dr. V.P. Singh, and Prof. (Dr.) Rajesh Khanna for their constructive suggestions and ensuring the progress of my research work at correct pace. Their critical analysis and encouraging support brought me to this stage of my course. The helped rendered by them is greatly accredited. My deep regards to Dr. O.P. Pandey, Dean of Research and Sponsored Projects for his enormous help in completion of my course work. I will go amiss if I forget to thank Dr. Prateek Bhatia for his valuable suggestions throughout the program. I wish to thank all the faculty and staff members of Computer Science and Engineering Department of Thapar University, Patiala for their enormous support.Lastly, I would like to pay my gratitude to my parents, wife, daughter, brother and to all friends and colleagues for love and encouragement all through my life. They have always helped me at every twirl of my life when I am in a need. Words cannot truly express my deepest gratitude and appreciation to all of the above. I am sorry, if I have forgotten someone and I cannot thank everyone enough for the involvement they have shown and the willingness they have expressed to take on the completion of tasks beyond their comfort zones.

Contents

Certificate ii

Abstract iii

Acknowledgment v

List of Figures ix

List of Tables xi

List of Important Abbreviations xii

List of Symbols and meanings used in equations xvi

1 Introduction 1
- 1.1 Infrastructure Networks . . . 1
- 1.2 Infrastructure-less Networks . . . 2
  - 1.2.1 Wireless Sensor Networks . . . 2
  - 1.2.2 Mobile Ad hoc Networks . . . 2
  - 1.2.3 Vehicular Ad hoc Networks . . . 3
- 1.3 Introduction to Vehicular Networks . . . 4
- 1.4 Vehicular Network Architecture . . . 5
- 1.5 Wireless Technologies in VANETs . . . 10
- 1.6 Characteristics of VANETs . . . 12
- 1.7 Applications of VANETs . . . 13
- 1.8 Fundamentals of Clustering . . . 17
- 1.9 Security Challenges in VANETs . . . 19
  - 1.9.1 Threats to Authenticity . . . 20
  - 1.9.2 Threats to Confidentiality . . . 21
  - 1.9.3 Threats to Availability . . . 22
- 1.10 Security Requirements in VANETs . . . 23
- 1.11 Security Proposals for VANETs . . . 24
- 1.12 Principles of Connected Dominating Set . . . 26
- 1.13 Research Gaps . . . 27
- 1.14 Research Objectives . . . 28
- 1.15 Thesis Organization . . . 29

2 Literature Review **30**
2.1 Challenges in Clustering . . . 31
2.2 Taxonomy of Clustering in VANETs . . . 33
2.2.1 Predictive Clustering . . . 33
2.2.2 Backbone Clustering . . . 39
2.2.3 MAC based Clustering . . . 41
2.2.4 Traditional Clustering . . . 46
2.2.5 Active Clustering . . . 47
2.2.6 Passive Clustering . . . 54
2.2.7 Hybrid Clustering . . . 56
2.2.8 Secure Clustering . . . 60
2.2.9 Discussion on Secure Clustering protocols . . . 61
2.3 Generic Security Protocols in VANETs . . . 61
2.4 Comparative Analysis and Discussion . . . 69
2.5 Conclusion . . . 78

3 Predictive Clustering **79**
3.1 Predictive Clustering Approach . . . 81
3.1.1 Network Model . . . 82
3.1.2 Movement Direction Estimation . . . 82
3.1.3 Beacon Messages Scheduling . . . 84
3.1.4 Future Position Estimation . . . 85
3.2 Proposed Predictive Clustering Approach . . . 86
3.2.1 Prediction Expertise Computation . . . 87
3.2.2 Average Prediction Variation . . . 88
3.2.3 Prediction of Future Vehicle Position . . . 90
3.2.4 CH Election Process . . . 91
3.2.5 Complexity Analysis of Predictive Clustering Approach . . . 92
3.3 Predictive Clustering Approach using Learning Automata . . . 94
3.3.1 Learning Automata for Predictive Clustering . . . 94
3.3.2 System Model for LA based Predictive Clustering . . . 96
3.4 Proposed LA Based Predictive Clustering Approach . . . 100
3.4.1 Future Vehicular Position Estimation . . . 100
3.4.2 LA Based Future Position Prediction Scheme . . . 103
3.4.3 CH Election Process . . . 107
3.5 Performance Evaluation . . . 108
3.5.1 Simulation Settings . . . 108
3.6 Results and Discussions . . . 109
3.6.1 Predictive Clustering Scheme . . . 109
3.6.2 LA based Predictive Clustering . . . 112
3.7 Conclusion . . . 117

4 Intelligent Cluster Based Intrusion Detection **118**
4.1 Distributed Intrusion Detection System . . . 120
4.1.1 Attack Model . . . 121
4.1.2 Network Architecture and Assumptions . . . 122
4.2 Proposed DIDS Scheme . . . 124
4.2.1 Leadership Formation . . . 124

4.2.2 Cryptographic Mechanism . . . 128
4.2.3 Intrusion Detection Mechanism . . . 129
4.3 Simulation Settings for DIDS . . . 136
4.3.1 Performance evaluation metrics . . . 136
4.3.2 Simulation environment . . . 136
4.4 Results and Discussions . . . 138
4.5 Conclusion . . . 140

5 Secure Clustering 142
5.1 Secure Clustering Approach . . . 142
5.2 System Model for Secure Clustering . . . 145
5.2.1 On-Board Broadcasting Unit . . . 146
5.2.2 Vehicular Cloud based Estimation . . . 147
5.2.3 Principle of Vehicular Clustering . . . 147
5.2.4 Attack Model . . . 148
5.3 Proposed Clustering Scheme . . . 149
5.3.1 Cluster Head Election . . . 154
5.3.2 Identification of Neighboring Vehicles . . . 155
5.3.3 Vehicular Parameters Estimation . . . 155
5.3.4 CH Election using CDS Algorithm . . . 157
5.3.5 Secure Data Dissemination . . . 160
5.4 Results and Discussion . . . 164
5.4.1 Cluster Head Selection . . . 165
5.4.2 Cluster Stability . . . 166
5.4.3 Cluster Transmission Efficiency . . . 168
5.4.4 Cluster Security . . . 169
5.5 Conclusion . . . 171

6 Conclusion and Future Scope 172
6.1 Conclusion . . . 172
6.2 Future Research Scope . . . 174
6.2.1 Secure Clustering . . . 174
6.2.2 Future Security Enhancements . . . 175

List of Publications 193

List of Figures

1.1 The main system components of vehicular network architecture. 6
1.2 Basic communication patterns in VANETs. 7
1.3 A network of sensors and actuators in a modern vehicle. 8
1.4 Different wireless technologies applicable in VANETs. 11
1.5 Communication services utilized for Internet enabled VANETs. 14
1.6 The structure and communication model of a basic cluster. 18
1.7 A connected dominating set structure in a graph. 27

2.1 Taxonomy of existing clustering approaches for VANETs. 33
2.2 Classification of Traditional Clustering. 47

3.1 Basic vehicular network model for cluster formation. 81
3.2 Vehicular road scenario at a time instant T 82
3.3 Vehicular road scenario at time instant $T + t$ 83
3.4 Estimation of movement direction of vehicles. 84
3.5 Prediction of future position of vehicles. 85
3.6 State diagram depicting prediction expertise of nodes. 88
3.7 Basic operations of a Learning Automaton 95
3.8 Prediction of future position of vehicles. 101
3.9 Operation of LA based predictive clustering where all three states can act as initial states 103
3.10 Simulated road map of Chandigarh City for predictive clustering. 109
3.11 (a) Variation in packet delivery ratio with increasing number of vehicles (b) Variation in End-to-End delay with increasing number of vehicles (c) Variation in throughput with increasing number of vehicles (d) Variation in probability of message transmission with increasing number of vehicles in predictive clustering scheme. 110
3.12 (a) Variation in packet delivery ratio with increasing cluster duration (b) Variation in End-to-End delay with increasing cluster duration (c) Variation in throughput with increasing cluster duration (d) Variation in probability of message transmission with increasing cluster duration in predictive clustering scheme. 111
3.13 Road Map of Chandigarh City used in Simulation of LA based Predictive clustering. 112

3.14 (a) Comparison of prediction accuracy with respect to prediction interval (b) Comparison of average prediction accuracy with respect to prediction interval. (c) Variation in prediction accuracy with respect to number of vehicles (d) Comparison of average prediction accuracy against prediction intervals in LA based predictive clustering scheme. 114
3.15 (a) Average cluster head duration as a function of vehicle speed (b) Percentage of vehicles as cluster heads with respect to vehicle speed in LA based predictive clustering scheme. 115
3.16 (a) Comparison of End-to-End Latency and prediction intervals. (b) Variation of PDR with respect to prediction intervals in LA based predictive clustering scheme. 116

4.1 Network model used in distributive intrusion detection system. 122
4.2 Framework used in distributive intrusion detection system. 123
4.3 Movement of vehicles in Chandigarh city for DIDS (a) Movement of vehicles at round about with traffic lights (b) Vehicles at entry point of the city. . 137
4.4 (a) Message Loss Rate when τ is considered. (b) Disseminated Messages when τ is considered in DIDS protocol. 139
4.5 (a) False Positive Rate when τ is considered. (b) Detection Rate when τ is considered in DIDS protocol. 140
4.6 (a) Message Loss Rate when τ is not considered. (b) Disseminated Messages when τ is not considered. (c) False Positive Rate when τ is not considered. (d) Detection Rate when τ is not considered in DIDS protocol. 141

5.1 Vehicular cloud environment for cluster based information dissemination . 144
5.2 Communication Framework model for Secure Clustering scheme. 146
5.3 State transition diagram based on trust values in secure clustering. 152
5.4 (a) Road side scenario.and (b) Equivalent clustering scenario used in secure clustering. 153
5.5 Road map depicting simulation scenario used in secure clustering scheme . 164
5.6 (a) Number of CHs with respect to varying clustering interval (b) Percentage of CH nodes with respect to total number of vehicles in secure clustering scheme. 166
5.7 (a) Percentage of cluster head duration with respect to vehicle velocity (b) Average cluster head duration with respect to velocity in secure clustering scheme. 167
5.8 Comparision of cluster head duration with respect to number of nodes with and without vehicular cloud in secure clustering scheme 168
5.9 (a) Packet delivery ratio as a function of clustering interval (b) End-to-End latency with respect to prediction interval in secure clustering scheme. . . 169
5.10 (a) Number of secure vehicles as function of clustering interval (b) Percentage lifetime of secure vehicles as a function of simulation time in secure clustering scheme. 170

List of Tables

1.1 Relative comparison between MANETs and VANETs. 3
1.2 Description of various application areas for VANETs. 16

2.1 Relative comparison of Position based clustering protocols. 36
2.2 Relative comparison of Destination based clustering protocols. 38
2.3 Relative comparison of Lane based clustering protocols. 39
2.4 Relative comparison of K-hop clustering protocols. 41
2.5 Relative comparison of IEEE 802.11 MAC based clustering protocols. . . . 43
2.6 Relative comparison of TDMA based clustering protocols. 45
2.7 Relative comparison of SDMA based Clustering protocols. 46
2.8 Relative comparison of Beacon based clustering protocols. 48
2.9 Relative comparison of Mobility based clustering protocols. 51
2.10 Relative comparison of Density based clustering protocols. 52
2.11 Relative comparison of Dynamic clustering protocols. 55
2.12 Relative comparison of Passive clustering protocols. 56
2.13 Relative comparison of Intelligence based clustering protocols. 57
2.14 Relative comparison of Cooperative De-Centralized clustering protocols. . . 58
2.15 Relative comparison of Driver Behavior based clustering protocols. 59
2.16 Relative comparison of Secure clustering protocols. 61
2.17 Relative comparison of existing clustering protocols in VANETs 77

3.1 Key symbols used in predictive clustering protocols and their meanings. . . 86
3.2 Simulation parameters required for LA based predictive clustering scheme. 113

4.1 Simulation parameters used in DIDS scheme. 134

5.1 Complexity of various algorithms used in secure clustering scheme. 164
5.2 Simulation parameters required for secure clustering protocol. 165
5.3 Comparative analysis of secure clustering scheme with existing cluster based scheme. 171

LIST OF IMPORTANT ABBREVIATIONS

Acronyms	Meanings
ISM	Industrial, Scientific and Medical
U-NII	Unlicensed-National Information Infrastructure
WSNs	Wireless Sensor Networks
MANETs	Mobile Ad hoc Networks
VANETs	Vehicular Ad hoc Networks
DSRC	Dedicated Short Range Communication
WAVE	Wireless Access in Vehicular Environment
V2V	Vehicle to Vehicle
V2R	Vehicle to RoadSide
V2I	Vehicle to Infrastructure
ITS	Intelligent Transport System
VII	Vehicle Infrastructure Integration
USDOT	U.S Department of Transportation
ASV	Advanced Safety Vehicle
AHS	Advanced Cruise-Assist Highway System
VICS	Vehicle Information and Communication System
DSSS	Driving Safety Support Systems
C2C	Car 2 Car
C2CCC	Car 2 Car Communication Consortium
AU	Application Unit
OBU	On-Board Unit
RSU	Road Side Unit
PDA	Personal Digital Assistant
I2V	Infrastructure to Vehicle
CC	Communication Consortium
EDR	Event Data Recorder
GPS	Global Positioning System
TPD	Tamper Proof Device
GSM	Global System for Mobile
2G	Second-Generation Wireless Telephone
2.5G	Second and a Half Generation
ETSI	European Telecommunications Standards Institute
GPRS	General Packet Radio Service
EDGE	Enhanced Data rates for GSM Evolution
UMTS	Universal Mobile Telecommunication System
HSPA	High Speed Down link Packet Access

CDMA	Code Division Multiple Access
WLAN	Wireless Local Area Network
Wi-fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
QoS	Quality of Service
VoIP	Voice over Internet Protocol
MIMO	Multiple Input and Multiple Output
OFDM	Orthogonal Frequency Division Multiplexing
FCC	Federal Communication Commission
SCH	Service Channels
CH	Cluster Head
CM	Cluster Member
DoS	Denial of Service
ID	Identity Description
CA	Certification Authority
MITM	Man-In-The-Middle
TPM	Trusted Platform Modules
PKI	Public Key Infrastructure
ECC	Elliptic Curve Cryptography
CRL	Certificate Revocation List
RC^2RL	Revocation using Compressed CRL
CDS	Connected Dominating Set
P2P	Peer-to-Peer
CG	Cluster-Gateway
Km	Kilometer
NEW-ALM	New Aggregate Local Mobility
ALM	Aggregate Local Mobility
MC-DRIVE	Modified Clustering based on Direction in Vehicular Environment
TTL	Time-To-Live
DMAC	Distributed and Mobility-Adaptive Clustering
LICA	Location Improvement with Cluster Analysis
RSS	Received Signal Strength
AMACAD	Adaptable Mobility-Aware Clustering Algorithm based on Destination
CBLR	Cluster Based Location Routing
CCT	Cluster Connect Time
TE	Transmission Efficiency
BDA	Broadcasting based Distributed Algorithm
CHL	Cluster Head Level
CCA	Criticality-based Clustering Algorithm
HCA	Hierarchical Clustering Algorithm
TDMA	Time Division Multiple Access
MAC	Medium Access Control
CB-MMAC	Clustering with Medium Access Control
DBA-MAC	Dynamic Backbone Assisted Medium Access Control
DCF	Distributed Co-ordination Function
VeSOMAC	Vehicular Self-Organized Medium Access Control
VeMAC	Multichannel Medium Access Control protocol for VANETs

CBMAC	Cluster Based Medium Access Control
TC-MAC	TDMA Cluster-based Medium Access Control
SDMA	Space Division Multiple Access
CGP	Clustered Gathering Protocol
CTS	Clear to Send
CDGP	Clustered Data Gathering Protocol
RSS	Received Signal Strength
LORA-CBF	Routing Algorithm with Cluster-Based Flooding
F	Force
SBCA	Stability Based Clustering Algorithm
APROVE	Affinity Propagation for Vehicular Networks
CDMA	Code Division Multiple Access
D-CUT	Distributed Construct Underlying Topology
DBC	Density Based Clustering
VWCA	Vehicular Weighted Clustering Algorithm
PassCAR	Passive Clustering Aided Routing
ALCA	Agent Learning–based Clustering Algorithm
COIN	Clustering for Open Inter vehicle communication Networks
VDDZ	VANET Dynamic Demilitarized Zone
SAV	Security Architecture for VANETs
P2DAP	Privacy Preserving Detection of Abuses of Pseudonyms
DMV	Department of Motor Vehicle
ECDSA	Elliptic Curve Digital Signature Algorithm
MDS	Misbehavior Detection Scheme
PCN	Post-Crash Notification
C2X	Car-to-X
ICS	ITS Central Station
IRS	ITS Roadside Station
IVS	ITS Vehicle Station
CCU	Communication and Control Unit
CAM	Cooperative Awareness Message
DENM	Decentralized Environment Notification Message
REP	Random Encryption Period
EDR	Efficient Decentralized Revocation
MAAC	Message Authentication Acceleration
HMAC	Hashed Message Authentication Code
MDS	Misbehavior Detection System
EWM	Event Warning Messages
PoW	Proof-of-Work
EWC	Event Warning Certificate
QoS-OLSR	Quality of Service Optimized Link State Routing
MPR	MultiPoint Relay
VPKI	Vehicular Public Key Infrastructure
AAA	Authentication, Authorization and Accountability
LTCA	Long-Term Certification Authority
PCA	Pseudonym Certification Authority
RA	Resolution Authority

SKCD	Secure Scheme based on Clustering and Key Distribution
LIDS	Local Intrusion Detection System
GIDS	Global Intrusion Detection System
GDS	Global Decision System
PPS	Payment and Punishment based Scheme
E2ED	End-to-End Delay
WDMA	Wavelength Division Multiple Access
B	Beginner State
N	Novice State
P	Predictive State
L	Learned State
G	Graph
V	Vertices
E	Edges
LA	Learning Automata
L_{R-I}	Linear Reward Inaction
PCALA	Predictive Clustering Algorithm using Learning Automata
A	Set of Actions of Lerning Automata
Q	Set of Automaton of Lerning Automata
R	Set of Reactions of Lerning Automata
T	Reinforcement Scheme of Lerning Automata
D	Set of Average Reward Values of Lerning Automata
SUMO	Simulation of Urban Mobility
PDR	Packet Delivery Ratio
VWCA	Vehicular Clustering based on the Weighted Clustering Algorithm
EEPC	Energy Efficient Predictive Clustering
VCC	Vehicular Cloud Computing
IoT	Internet of Things
V2S	Vehicle-to-Sensor
RFID	Radio Frequency Identification
MDS	Misbehavior Detection Systems
MITM	Man-In-The-Middle
IDS	Intrusion Detection System
VM	Virtual Machine
SOA	Software Oriented Architecture
NT	Normal Time
UR	Update Round
DCT	Data Collection Time
ARV	Aggregate Relative Velocity
LE	IDS Leader
CN	Cluster Member for IDS
CID	Cluster-Head ID
SK_i	Leadership Secret Key
PK_i	Leadership Public Key
PV_i	Cluster Private Key
PB_i	Cluster Public Key
$dSign_i$	Digital Signatures

SHA1	Secure Hash Algorithm1
Req_Join	Joining Request Message
Req_Leave	Leaving Request Message
NS-2	Network Simulator-2
SUMO	Simulation of Urban Mobility
VID	Vehicular Identifier
OBB	On Board Broadcasting
VCE	Vehicular Cloud-based Estimation
U	Set of Unvisited Nodes
C	Set of Covered Nodes
D	Set of Dominating Nodes
ECIES	Elliptic Curve Integrated Encryption Scheme
MsgAC	Message Authentication Code
KDF	Key Derivation Function

List of Important Symbols and their Meanings

Symbols	Meaning
$TH_{distance}$	Threshold Value of Distance
$F_{v,2}$	Metric for Clustering
T_{speed}	Threshold value of Speed
T_m	Trust Metric
t	Short time prediction interval
T	Future position estimation time interval
N, n	Number of vehicles
W	Width of road section
$d_{A,t}$	Distance of Vehicle A from a reference vehicle at time t
$D_{A,t}$	Position of vehicle A at time t
W_J	Weight of vehicle-Group
$x_{A,t}$	x-coordinate of vehicle A at time t
$y_{B,t}$	y-coordinate of vehicle A at time t
$predpos_T$	Predicted position of vehicle at time T
$actualpos_T$	Actual position of vehicle at time T
v_T	Speed of vehicle at time T
p_i	Probability of vehicle i to participate in CH election
$\in$	Average Prediction Variation Coefficient
v_{mean}	Mean velocity of all vehicles
k	Number of time intervals in vehicle prediction
d_T	Predicted Final Position
σ	Cluster Formation Parameter
λ_T	Future Mobility Factor
$\vec{P}$	Prediction Accuracy Vector
E_{it}	Enhanced Value of Prediction Accuracy
U_{it}	Unchanged Value of Prediction Accuracy
R_{it}	Reduced Value of Prediction Accuracy
λ_R	Learning Scheme modification parameters
τ	Threshold value of tolerance to malicious activities
L	Number of Leaders in a Cluster
δ	Degree of Connectivity
C_{ave}	Average Degree of Connectivity
M_{tot}	Total number of successful messages received by vehicle i at time instant t
D	Average Density of vehicles
$V_Y^{rel}(X)$	Current relative velocity of a vehicle with respect to its neighbors

(V_i)	Variance of Relative velocity
S_T	Predefined Threshold for Leadership Election
Ω_k	Resultant value of k^{th} node for Leadership Election
n^{rew}	Number of Rewards
n^{pen}	Number of Penalties
$selectCH$	Select CH Message
C_data_i	Summary of Collected Data
$suspect\_list_i$	List of Suspect Nodes
mal_list_i	List of Malicious Nodes
$final\_mal\_list$	Final List of Malicious Nodes
ξ	Ratio of number of rewards and penalties
th_r	Threshold value for IDS based Learning Automata
new_view	Computed view of Malicious Nodes
(Req_Join)	Joining Request Message
(Req_Leave)	Leaving Request Message
$(W_{i,t})$	Trustworthiness Weights values of vehicles
(p_t)	Number of Packets Transmitted by a Vehicle
(p_r)	Number of Packets Received by a Vehicle
$\eta_{i,t}$	Activity Profile for a Vehicle
$\eta_{avg,t}$	Average Value of Activity Profile
$\delta_{s,t}$	Dynamic Threshold Value for Secure State
$\delta_{u,t}$	Dynamic Threshold Value for UnSecure State
$q_{i,t}$	Current State of Vehicle
$q_{i+1,t}$	Next State of Vehicle
$W_{\mu,t}$	Mean value of Trustworthiness Weights
$W_{\sigma,t}$	Variance of Trustworthiness Weights
D_{avg}	Average Distance of a Vehicle from its neighbors
d_{ij}	Distance between Vehicles
Δ	Average Distance Parameter
W_{ini}	Initial Value of Trust Weight
p	Field Order of Elliptic Curve
a,b	Elements specifying the Elliptic Curve
G	Base Point of Elliptic Curve
h	Cofactor of Elliptic Curve
m	Prime Number prime which is the order of G
P_i	User Pseudo Identity
V_i	Vehicle Pseudo identity
P_{kr}	ECC Private Key
P_{ku}	ECC Public Key
C	Cipher Text
M	Plain Text

Chapter 1

Introduction

During recent decades, rapid advancements have been observed in the field of wireless communication technologies. The wireless communications provide an easy access to the users for services and information regardless of their geographical positions through the use of microwave spectrum for communication. The various frequency bands used for wireless communication are 2.4 GHz Industrial, Scientific and Medical (ISM) band with bandwidth around 83 MHz and 5 GHz Unlicensed-National Information Infrastructure (U-NII) band with bandwidth around 300 MHz [1]. The communication range of wireless network depends on the factors like emission power, frequency, types of antenna and data rate. The allocation of frequency bands are done by different laws in various countries. For example, 14 channels are allocated for 2.4 GHz range out of which only 11 channels are used in America whereas all 14 channels are used in Japan. These laws regulate the frequency utilization and transmission power accross different locations. The ability to self-manage, self-maintain, self-organize and self-configure are some of the key attractive features of wireless systems [2]. However, with an exponential increase in wide range of applications, these networks need to adapt to different heterogeneous environments to increase overall network performance. Wireless networks are broadly classified into two categories as:

- Infrastructure networks
- Infrastructure-less networks

1.1 Infrastructure Networks

These networks consist of fixed and wired gateways for communication where mobile hosts communicate with a fixed base station within its communication range. While communicating, a mobile node can move geographically but the base stations are fixed.

During movement when a mobile node is out of range of one base station, it temporarily halts its communication and restarts it after it connects with some other base station.

1.2 Infrastructure-less Networks

Infrastructure-less or ad hoc networks are defined as a network of nodes which communicate with each other without any predefined infrastructure. These networks function without any pre-installed infrastructure and the communication between the nodes is through a dynamic wireless link. According to the situations these networks are created and managed in an ad hoc manner generally for a short duration. All the nodes in these networks behave as routers to participate in route discovery and maintenance. These networks are used in emergency situations like search-rescue operations, information sharing etc. The increasing demand of wireless communication has resulted in the need of such self organizing and self managing networks which works without the interference of any centralized infrastructure. Wireless sensor networks, Mobile ad hoc networks and vehicular ad hoc networks are examples of infrastructure less networks.[3].

1.2.1 Wireless Sensor Networks

Wireless Sensor Networks (WSN's) are distributed networks which consist of multiple autonomous sensor nodes deployed in specific areas for monitoring purposes like sensing the environmental conditions and transferring aggregated information to the base stations or the central entity known as sink nodes [4]. The sensor nodes being self organizable and fault tolerant can provide periodic and event based data in an effective manner and are deployed in various application areas such as disaster monitoring, forest fire detection, underwater surveillance, military, disaster relief management etc. Sensor nodes being battery operated have lower power capabilities and limited energy. The challenges faced by wireless sensor networks include restricted processing power, limited available energy and energy consumption. It is impossible to replace or recharge the sensor nodes in certain conditions when there is inaccessibility of sensor nodes. Therefore, a lot of attention from researchers has been focused on increasing network lifetime by managing the energy and reducing the power consumption for WSN's.

1.2.2 Mobile Ad hoc Networks

Mobile Ad hoc Networks (MANETs) [1] are a kind of wireless communication networks in which messages are exchanged between nodes without using any access point or infrastructure. When two nodes are not in range of each other then they can forward packets through intermediate nodes. The nodes connected in such networks are free to join and

leave the network. In MANET, each nodes in the network acts both as router and host and hence must possess the capability to forward packets to other nodes. The nodes in transmission range of each other communicate directly but messages are transferred through multi-hop communication if the distance between the transmitted nodes is more than transmission range. Some of the key attributes of MANETs are dynamic network topology, self-organizing type of network and fluctuating link capacity [5]. The major challenges faced by MANETs are link bandwidth variations, power management and mobility of nodes leading to topological changes. The nodes could be mobile laptops and personal digital assistants which facilitates users by providing many applications like print sharing, file transferring, video streaming and voice conferencing. They are also used in military battlefield and crisis management applications.

1.2.3 Vehicular Ad hoc Networks

Vehicular Ad hoc Networks (VANETs) [6] are a subclass of MANETs in which nodes are vehicles that move with different speeds and communicate with each other using either homogeneous or heterogeneous wireless technology. Vehicles are equipped with transmission devices which provide short range wireless connectivity. Vehicular communication based applications are gaining importance during the past few years. VANET being a wireless ad hoc network provides the communication among the vehicles and roadside infrastructure. The similarity between MANET and VANET is characterized by the movement and self organization of nodes and low variable bandwidth, infrastructure less and short range connectivity. However in a VANET, node mobility is higher than in MANET so routing protocols being used in MANETs are not used for VANETs [5]. The topology created by vehicles is significantly non-uniformly distributed and usually dynamic. Table 1.1 shows the comparisons between MANET and VANETin terms of transmission characteristics and other network parameters.

Table 1.1: Relative comparison between MANETs and VANETs.

Parameters	MANET	VANET
Production Cost	Cheap	Expensive
Mobility	Low	High
Range	Up to 100 m	Up to 500 m
Node density	Sparse	Dense and frequently variable
Reliability	Medium	High
Node lifetime	Depends on power resource	Depends on lifetime of vehicle
Topology changes	Very Slow	Highly Dynamic
Multi-hop routing	Available	Weakly available
Moving pattern of nodes	Unplanned	Consistent
Position Acquisition	Using ultrasonic	Using GPS, Radar

1.3 Introduction to Vehicular Networks

Communication technologies are gradually becoming an essential part of our lives and they also provide a number of opportunities that may be utilized to meet our daily needs. VANETs have also been attracting a lot of attention from both industry and research community in recent years, as a valuable communication technology. With an increase in number of private vehicles on the road problems such as, traffic jams and accidents are also increasing. Due to these factors, the number of fatalities due to vehicular traffic are also increasing. Accidents impose danger and serious problems to our society. Thus one of the key objectives for VANETs is to make the journey safe on road and comfortable for vehicle users on road as well as fulfilling their communication requirements while they are traveling on the road. VANETs use cars as a mobile node for creation of a mobile network which allows communication between road transport vehicles for promoting passenger safety on the road. The cars in VANETs act as wireless routers and can also be connected to neighboring cars that are in range of approximate 100 to 300 meters [7]. The cars drop out of the network when they fall out of the signal range and other cars join in to create a communication mobile network.

VANETs allow communication between vehicles wirelessly with the help of Dedicated Short Range Communication (DSRC) Protocol. It is also known by Wireless access in Vehicular Environment (WAVE) [6]. Vehicle-to-Vehicle communication (V2V) is a kind of communication in which vehicles can communicate directly where as Vehicle-to- RoadSide (V2R) or Vehicle-to-Infrastructure(V2I) communication allows vehicles to communicate with infrastructure units installed near roadside [6]. Wireless communication allows vehicle to share different kind of information like safety, accident prevention and traffic jams. It is also used to disseminate information like location of nearby petrol filling station, resorts etc. It can provide safety message to warn drivers about critical condition such as accidents. The topology of VANETs depends upon the density of vehicle at a particular instant. If population of vehicles are dense then vehicles can communicate directly to other vehicle by using multi-hop communication. But if the number of vehicles are less or sparse environment then vehicle can store message for a given time-stamp and send it using opportunistic routing.

VANETs provide different type of opportunities for enhancing the capabilities of wireless communication. They allow educators to develop various tools, protocols and applications. It also provides different challenges as well as opportunity for researchers in this field. In 2003, Intelligent Transport System (ITS) World Congress, the Vehicle Infrastructure Integration (VII) [8] was initiated by the U.S Department of Transportation (USDOT) and VII Association was formed in 2005 that helped in designing, testing and evaluation of vehicular network. In Japan, advanced ITS solutions, named as Ad-

vanced Safety Vehicle (ASV) (1991) , Advanced Cruise-Assist Highway System (AHS) (1996), Vehicle Information and Communication System (VICS) and Driving Safety Support Systems (DSSS) were initiated (2002) [9]. In USA, the nation wide deployment of ITS-Safety 2010 project was done in 2010. In Europe, Six European car manufacturers, namely DaimlerChrysler, BMW, Fiat, Renault, Audi and Volkswagen launched a Car 2 Car (C2C) Communication Consortium (C2CCC) [10].

In 1999, 75 MHz of DSRC spectrum bands with bandwidth range between 5.850 to 5.925 GHz were allocated for V2V and V2I communication in North America by U.S. Federal Communication Commission. Alternatively in Japan, the frequency bands having range from 5.770 to 5.850 GHz were assigned for DSRC. In Europe, the frequency bands from 5875 to 5905 GHz were considered for road safety and an additional frequency band of 20 MHz for future extension and 5855-5875 MHz band was made available for non-safety applications. The DSRC spectrum is standardized as IEEE 802.11p. The working group of IEEE 1609 [11] has standardized the DSRC communication stack. The IEEE 1609 working group introduced four standards designed for WAVE: (i) IEEE P1609.1 which is known as WAVE Resource Manager outlines data read/write protocol for application platform [11], (ii) IEEE 1609.2, i.e., WAVE Security Services outlines security, authenticity, confidentiality, anonymity and privacy [11], (iii) IEEE 1609.3, i.e., WAVE Networking Services outlines network layer and transport layer services such as addressing and routing in secure and reliable data interchange [11], and (iv) IEEE P1609.4, i.e., WAVE Multichannel Operation Management helps to provide coordination and management of DSRC frequency band and also manages lower level usage of frequency bands [11].

1.4 Vehicular Network Architecture

Wireless communication between Vehicles and Road Side infrastructure is done using WAVE protocol as discussed above. The main components of VANETs are Application Unit (AU), On-board Unit (OBU) and Roadside Unit (RSU) [12]. The RSU provides service to user and vehicles use that services through OBU. Vehicles contain a set of sensors and OBU so that they can collect and process the messages. RSU can further establish connection with Internet so that it can provide various services to the users. Figure 1.1 shows the essential components of VANETs to support vehicular communication for exchanging information.[13].

Figure 1.1 shows the essential components of VANETs to support vehicular communication. AU uses communication facility of OBU for exchanging information with Road Side Unit [13]. These components are explained in brief as follows:

- **On Board Unit**: The device which can be used to wirelessly communicate with

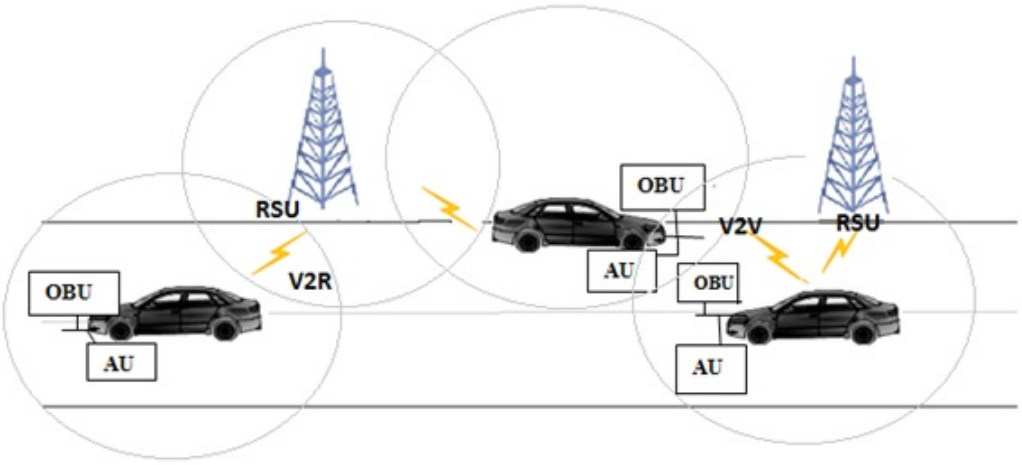


Figure 1.1: The main system components of vehicular network architecture.

other vehicles and RSU is called OBU. It is based on WAVE protocol. It is equipped with processor having good computing capabilities, user interface, memory used to store and retrieve information and short range wireless equipment to communicate wirelessly using 802.11p protocol [14]. It also supports various wireless technologies such as Wi-Fi/IEEE 802.11 a/b/g/n which are used for communication and other applications. The OBU connects with other OBU's or RSU's using 802.11p protocol using the available channel for communication. IP mobility, geographical routing, wireless radio access, reliable message transfer, congestion control, ad hoc network, and data security are the main functions which are supported by the OBU [12].

- **Application Unit**: A vehicle can use AU device for mobility verification and it enhances the communication capabilities of OBU. The AU may be a devoted communication equipment or Personal Digital Assistant (PDA) which is used for running applications through existing networks such as Internet with or without V2I communication. When AU is a separate component then it is connected with OBU through either a wired or wireless connection.

- **Road Side Unit**: RSUs are physical network devices deployed along the roadsides or in dedicated locations like road junctions, gas stations, restaurants or parking spaces. It is a device with WAVE standard. The RSU is based on radio short range communication i.e. 802.11p technology and is used to communicate with infrastructure network like Internet and other owned networks. The functionality of RSU involves V2I or Infrastructure to Vehicle (I2V) communication as well as Internet connectivity. The RSUs may be interconnected with each other and also further connected with Internet to provide various services such as multimedia data sharing and variety of web services [12]. According to Communication Consortium (CC), various functions and procedures associated with the RSU are:

– Extending the communication range of the network is re-distributing the information to other OBU's and sending information to other RSU's and to forward it further to other OBU's.

– Providing Internet connectivity to OBU's.

– Running safety applications like low bridge warning, accident warning, using I2V and acting as a dedicated information source.

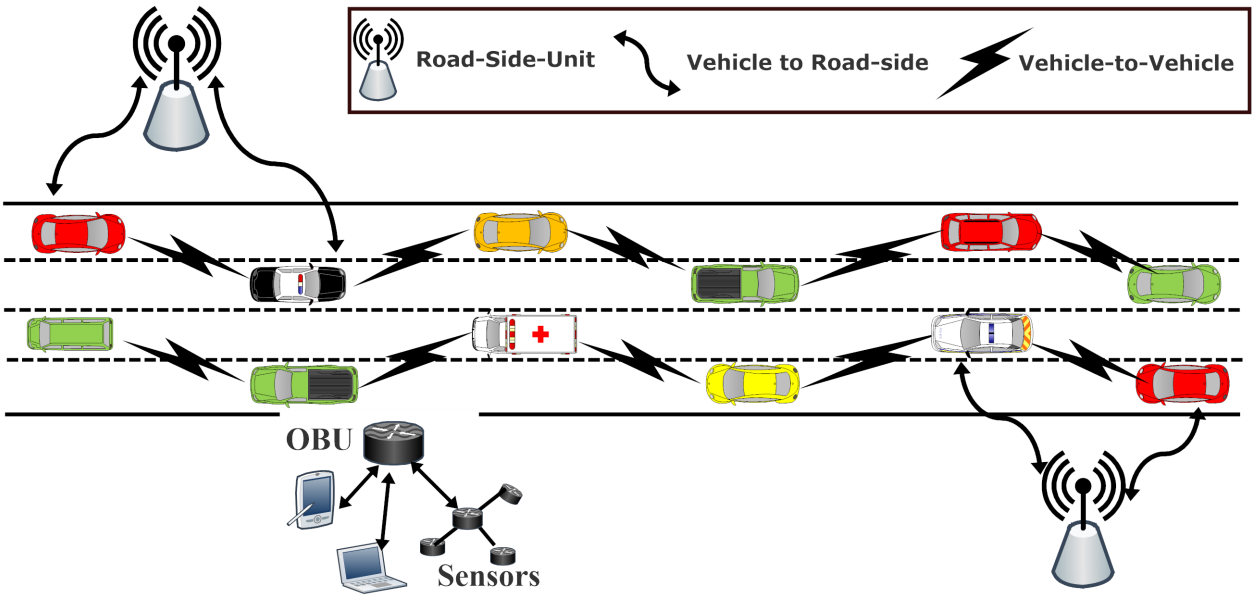


Figure 1.2: Basic communication patterns in VANETs.

Figure 1.2 shows the two basic types of communication patterns for VANETs. When vehicles communicate with RSU, it is called as V2R communication and the communication which occurs between vehicles is termed as V2V communication [15]. These two types of communication are generally prevalent for data dissemination and are explained in detail as below:

- **Vehicle-to-Vehicle Communication**

  V2V communication [16] uses multi-hop broadcast or multicast for transmitting information related to traffic like dynamic route scheduling or collision related emergency messages over the multiple hops to the receivers. Intelligent broadcasting and naive broadcasting are two ways by which message forwarding in vehicle-to-vehicle communications takes place. In naive broadcasting vehicles at regular time interval broadcast the messages to the vehicle behind it. Large number of broadcast messages increases the risk of message collisions which may lead to lower delivery rate and more delivery time. In intelligent broadcasting the message broadcasting,

is limited for emergency events which addresses the issues of naive broadcasting. Vehicle moving back is responsible to broadcast the messages to the rest of vehicles. Vehicle acts upon the first message it receives, if it receives same message from more than one sources.

- **Vehicle-to-Roadside Communication**

  The V2R communication [16] involves single hop broadcasting of messages between the RSU's and the vehicles which are in range of RSU's. High bandwidth link communication is provided between the vehicles and roadsides units. RSU's placed at regular interval ensures high data rates for maintaining heavy traffic. For example, RSU's determine appropriate speed limit based upon traffic conditions and broadcast the speed limit to the vehicles in its range and if the speed limit is violated, the auditory or visual warning broadcast would be delivered to the vehicle to slow down the speed of vehicles moving along that road.

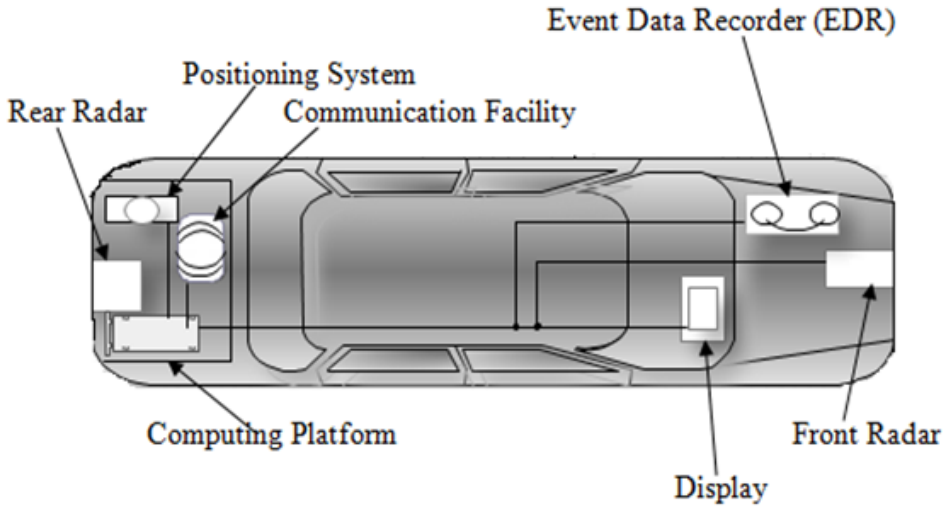


Figure 1.3: A network of sensors and actuators in a modern vehicle.

In VANET, each vehicle is composed of a variety of sensors, actuators and computing devices as shown in Figure 1.3. The vehicle has a computing device used for computations, front and rear sensor for sensing various activities, short range wireless interface which enables the communication and Event Data Recorder (EDR) for recording events of the vehicles. Global Positioning System (GPS) provides vehicle's location, direction, speed and the current time [17]. OBU of the vehicle connect the vehicle via DSRC radios to RSUs that are in the transmission range. Each vehicles own information like speed, route, driver's identity, keys, trip details etc. is stored in Tamper Proof Device (TPD).

The functions performed by OBU's include network congestion control, data security, wireless radio access, reliable message transfer, IP mobility support, geographical ad hoc routing and others [18]. RSU is a network device dedicated at fixed intervals that helps in providing Internet connectivity to OBUs, running safety applications like low bridge warning, accident warning and redistributes the information to other OBU's and to the other RSU's to forward it further to other OBU's.

Intelligent Tranportation system (ITS) has been developed for improving the efficiency and safety in transportation system [19]. The ITS architecture provides a framework for emerging mobile services and applications in vehicles and transportation infrastructure. The main applications of ITS focus on improvement in providing support for public safety and collision avoidance [18]. The communication in VANETs can be initiated from the RSU or the vehicle. The communication that can be either single-hop or multi-hop could be one way or two way communication for increasing the received messages on time. [20]. Vehicles periodically broadcast beacon messages which consists of information like vehicles position and speed. This information is then utilized for a number of service applications in VANETs. The main communication patterns [21] for VANETs that are applied for communication are discussed below:

- *Unicast*: Data packets are unicasted to the specific destinations over multiple hops from the sender. These data packets consists of the coordinates of the destination. These packets are forwarded to their neighbors by relaying nodes. The communication may consist of routing messages over multiple hops towards the destination or only a single hop. The destination of packets is either a remote destination region or a single node.

- *Single-hop broadcast*: Data packets are broadcast continuously or periodically to all the neighbors where communication is in single hop. Continuous updation of the information linked to vehicle such as its position or speed is shared among all the neighboring nodes is done as a single hop broadcast. For example, if an accident occurs, then the beaconing mechanism starts by an external trigger. The communication is unidirectional if each and every node is beaconing.

- *Multi-hop Broadcast*: Messages are disseminated by the relay nodes throughout the target area. These messages are not continuously sent but they are triggered during external events. These are unidirectional and need low latency messages for informing addressed vehicles as they are event based messages.

The technique of spreading or broadcasting the information in an ad hoc network is known as dissemination [22]. Whenever vehicle encounters any message which is broadcasted

by other vehicles in the network it forwards it after updating the stored information. This broadcasting of information can either occur in all the directions or in a particular direction. There are three cases considered for dissemination mechanism in which mobility of vehicles occurs in the opposite direction, same direction or in both the directions. some of the data dissemination techniques [23] are classified as follows:

- *V2R Dissemination*: This dissemination is push and pull based [23]. The pull based data dissemination, is a request and respond based model in which vehicles query information about their location and get the response. But, data is delivered from vehicle or RSU to other vehicles in push based data dissemination.

- *V2V Dissemination*: In this technique, either flooding or relaying approaches are followed. In this relaying, relay node is selected which forwards the data to the next hop and so on. This helps to reduce the network congestion. In flooding, every node participates to flood the time sensitive data to all the vehicles.

- *Geographical Dissemination*: The message is transmitted to the closest node towards the destination direction till it reaches to the destination. This is usually used in continuous changing topology.

- *Cluster Based Dissemination*: Nodes are grouped in set of clusters in which cluster head gathers the data and then it sends it further to neighboring cluster heads which helps in reducing broadcast storm problem [23] and also helps to attain high delivery ratio.

- *Opportunistic Dissemination*: In this technique, each vehicle stores the information and then forwards it to the neighboring vehicles that it encounter till the destination is not reached.

1.5 Wireless Technologies in VANETs

A large number of mobile radio communication technologies such as cellular telephone systems and wireless networks are available which can be used by vehicles to communicate with other vehicles. These technologies are used to improve road safety, avoid jams and provide comfort to passengers. Figure 1.4 shows the usage of various existing technologies in VANETs. The most widely used wireless technologies [12] are described as follows:

Cellular systems: Cellular system is defined as the technology to reuse the narrow frequency available for communication. Global System for Mobile (GSM) [24] communication is the standard used for cellular communication system. It can facilitate a data rate up to 9.6 Kbps. It is also known by other name as Second-Generation Wireless telephone

(2G). It uses two frequency bands, 890-915 MHz frequency band for uplink where 935-960 MHz frequency band for downlink. The frequency of given band is equally distributed into channels. Each channel is of 200 KHz.

General Packet Radio Service: It is also known by another name as Second and a Half Generation (2.5G) cellular wireless technology and it is an evolution to GSM. European Telecommunications Standards Institute (ETSI) [12] [24] provides a standard called as General Packet Radio Service (GPRS). It allows a data rate of equal to 170 Kbps for Internet. There is another standard named as Enhanced Data rates for GSM Evolution (EDGE) [12] which is known as 2.75G. EDGE is evolved version of GPRS. EDGE offers maximum data rate of 384 Kbps. This brings an evolution to the 3G. The Universal Mobile Telecommunication System (UMTS) and its next version the High-Speed Down link Packet Access (HSDPA) offers a data rate of around 2 Mbps [12]. Cellular system based on Code Division Multiple Access (CDMA) 2000 standard offers a data rate of 3 Mbps for down link usage and also it offers 1.8 Mbps for up link respectively.

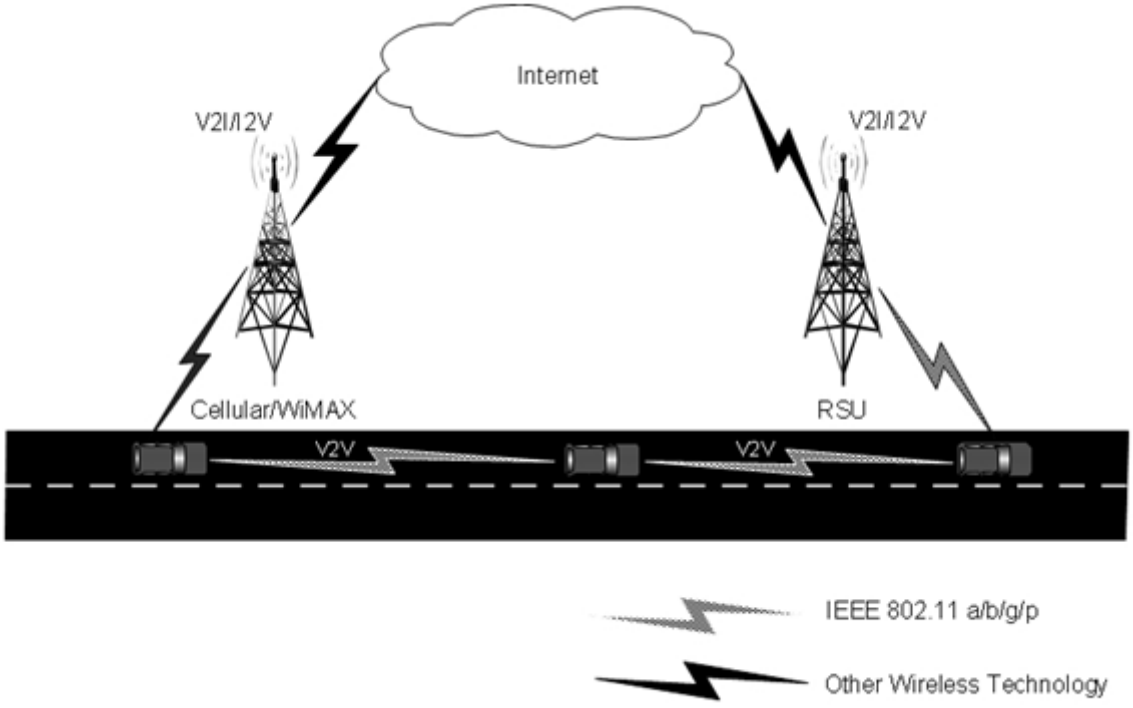


Figure 1.4: Different wireless technologies applicable in VANETs.

Wireless Local Area Networks: The Wireless Local Area Network(WLAN) or Wireless Fidelity (Wi-Fi)[14] standards are used to provide wireless access for communication. They are used for enabling V2V communication or V2I communication. The standard which is used for providing wireless connectivity is defined as the family of IEEE 802.11 standard [14]. IEEE 802.11a has the capability to works at 5 GHz bandwidth and provides 54 Mbps data rate for usage. The other IEEE standard which is widely implemented is

named as IEEE 802.11g. It facilitates the same data rate for communication and offers the same range as of IEEE 802.11a. It is also capable of being operational at 2.4 GHz frequency. IEEE 802.11b is based on 2.4 GHz frequency band and capable of producing a data rate around 11 Mbps for usage. The 802.11g and 802.11b [14] are widely used in VANET and have an operational communication range of 45 meters. The 802.11 standards are widely used for mobile computing devices like smartphones, laptop, and various other wireless devices.

Mobile-Worldwide Interoperability for Microwave Access: IEEE 802.16 or Worldwide Interoperability for Microwave Access has been available since 2004. It was modified by researchers to give Mobile Worldwide Interoperability for Microwave Access (WiMAX) or IEEE 802.16e [25]. It provides wide range of communication and Quality of Service (QoS). It also offers high data rate which makes WiMAX suitable for Voice over Internet Protocol (VoIP), multimedia, video and other applications. With the help of Mobile-WiMAX, data rate up to 35 Mbps using Multiple Input and Multiple Output (MIMO) can be generated with Orthogonal Frequency Division Multiplexing (OFDM). It also covers 15kms of transmission range.

Dedicated Short Range Communication: DSRC technology is also known as 802.11p protocol which has been modified from 802.11a protocol. It provides low overhead operation in 5.850 GHz to 5.925 GHz frequency bandwidth. The DSRC protocol family includes IEEE 802.11p [11] [25] has been standardized by IEEE 1609 working group and is known as WAVE. It has a fixed 75 MHz licensed spectrum allocated by researchers in US Federal Communication Commission (FCC). It is used for both V2V and V2I communication in the United States. The 75 MHz spectrum is further divided into seven channels where each channel has 10 MHz capacity. In this spectrum, 172 is starting channel number and 184 is ending channel number. Control Channel 178 is used for safety applications, while channels number 172 and 184 are reserved for safety application in VANETs. Service Channels are used for safety as well as non-safety applications. WAVE supports moving vehicles at highest speed of 200 Kmph. Its communication range is 300 m to 1000 m and data rate is above 27 Mbps [11] [25].

1.6 Characteristics of VANETs

The characteristics of a VANET are unique as compared to other infrastructure-less networks. Due to its distinctive properties, it provides various opportunities for increasing network performance but also encounters many challenges [26]. Some of the salient characteristics of VANETs [27] are described as follows:

- *Dynamic Topology*: In VANET, vehicle's movement is limited to road patterns. Mobility patterns of the vehicles can be predicted by utilizing the roadway geometry

with a reasonable accuracy. Even though the vehicles mobility pattern can be predicted but the relatively high speed of vehicles in VANETs results in frequent changes being encountered in its topology.

- *Frequent Network Disconnection*: Highly dynamic topology in VANETs leads to change in link connectivity as well as network structure. This results in rapid changes in the density of vehicles on road. Vehicular density has an effect on network performance as varying network density results in disconnections that vary with density.

- *Different Traffic Scenarios*: The communication environments in which VANETs operate are based on highway or city traffic scenarios. The mobility of vehicles is more complex in city scenario due to higher vehicle and road density as compared to the highway scenario. The communication overheads are also more in city scenario due to more vehicular density whereas in highway scenario, the communication overhead is less.

- *Large Scale network*: The size of network in VANETs depends upon the number of road links. In urban scenario, the road network is very dense hence scalability of VANET is more pronounced due to larger vehicular density.

- *Adequate Storage*: The vehicular nodes have comparatively large storage capacity instead of limited storage available in small hand held devices that are used as a node in other types of ad hoc networks.

- *No Power Constraints*: The battery in vehicles provide continuous supply to the OBU for processing where as RSUs can also be connected to a continuous power supply so there is no power constraint issues in VANETs as compared to limited battery power in some other types of ad hoc networks.

- *Density of the Network*: The network density depends upon number of vehicles on the road under consideration. The density of network is high in urban scenario but it is generally less in case of rural areas. Traffic jams may also impact the vehicular density.

1.7 Applications of VANETs

VANETs offer a wide range of applications like cooperative traffic monitoring, preventing collisions, traffic flow control etc [27]. VANETs can also enhance a vehicles potential to broadcast warnings of environmental hazards existing traffic and road conditions, vehicular congestion or emergency braking and entertainment information to other vehicles.

Once a vehicle receives such messages that indicate events such a road closure, accident or traffic jam ahead, a driver can take advance actions for safely of passengers. Thus VANETs can be used to provide assistance that helps to make roads safer. Various types of assistance that can be provided to users from various types of applications through VANETs [28] are as follows:

- *Applications for Commercial Assistance*: VANET can be used to provide commercial services such as Internet access, streaming audio and video etc. Some of the commercial applications are categorized as follows:
  - *Cooperative assistance*: This is used for distribution of critical data to the vehicles for example in broadcasting the warning messages of accidents.
  - *Cooperative awareness*: This is used in extending driver assistance by transmitting messages that indicate dangerous road conditions, adverse weather conditions or temporary obstacles on roads with help of on board sensors on some vehicles.
  - *Cooperative maneuvering*: This is used in exchanging information such as relative position and dynamics between vehicles or handling situations related to cooperative driving such as indicating, lane change or lane merge assistance.

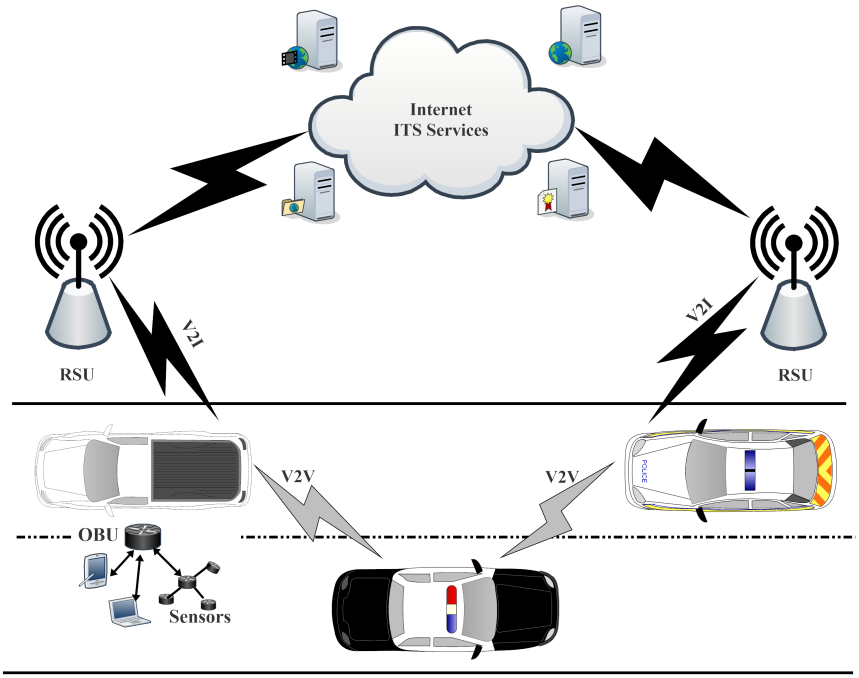


Figure 1.5: Communication services utilized for Internet enabled VANETs.

- *Application Convenience and Internet Connectivity Assistance*: The OBUs can communicate with infrastructure points and with other vehicles to provide the driver

with services like traffic flow information, automatic toll payment, parking availability information etc by using the available infrastructure network such as Internet as shown in Figure 1.5. There convenience and personalized applications include following types of assistance:

- *Car-to-Car*: This type of assistance is achieved through exchange of information between car users so as to provide real time exchange of information such as multimedia or other data files.
- *Car to Mobile devices*: In this category, those applications are included which provide assistance based on communication among car and mobile devices such as laptop, mobile phone etc.
- *Car to Office or home*: Communication which exchange information between the car and private network including home or office are also an important type of assistance that can be provided through VANETs.
- *Car to Enterprise*: This includes communication among the car and companies and includes application such as identification of parking areas, restaurants, gas stations, etc that provides effective road services.

- *Applications for Safe Navigation Assistance* The VANETs are used to identify conditions that may potentially creates danger to the driver's safety by avoiding collisions, offering warnings related to state of roads and intersections and reducing pile up of vehicles after an accident [29]. The messages can reach the destination vehicles either in a few numbers of hops or in a single hop. There are two kinds of safety messages that are generally transmitted, i.e., periodic messages and event-driven messages [16]. Periodic messages are sent for detecting non-safe situations like providing information related to the position, speed, direction, or some other information about a vehicle to the surrounding vehicles. The event-driven messages are those messages which are generated on demand, for example sending alert messages, when front car applies the brake. For all such situations, rapid diffusion of information is vital to avoid an accident. Some examples of applications that provide assistance for safe navigation are as follows:
  - *Slow or Stopped Vehicle Advisory*: When an alert is broadcast by some vehicle that suddenly stops or slows down on the road thereby enabling the recipients to take necessary action.
  - *Road Hazard Condition Notification*: A vehicle on observing a hazardous condition on the road like road slipperiness broadcasts an alert warning to other vehicles about this condition.

- *Post-Crash Notification*: A vehicle involved in a crash or some problem broadcasts an alert message to other neighboring vehicles to inform about its status and its location enabling the vehicle behind it to take possible action.
- *Road Feature Notification*: A vehicle on observing a road feature, like sharp curve or a hill that does not have a road sign broadcasts an alert informing the other vehicles that would help the driver to take corrective action.
- *Collision Avoidance*: These services help in saving human lives and prevent injuries. For example, when vehicle reduces its speed significantly after observing or experiencing an accident, it broadcasts its location to its neighboring vehicles and other receivers relays the message further which helps drivers far behind to get an alarm signal before they face the accident so they can take suitable actions.
- *Cooperative Driving*: One of the important factor that causes accidents is the lack of cooperation between drivers.These services can reduce the probability of life-endangering accidents. Therefore by transmitting more information about conflicts and thereby increasing cooperation among drivers, we can prevent many accidents. In such applications the driver's play a leading role by sharing information like turn conflict warning, curve warning, violation warning, lane merging warning etc.thereby providing a cooperative driving environment.

Table 1.2: Description of various application areas for VANETs.

Application Type	Classification	Sub-Classification
Safety	Safety-critical application	V2V Based
		V2V/I2V Based
Safety	Safety-related application	V2V Based
		V2V/I2V Based
Non-Safety	Roadside service finder	
	Payment services	
	Infotainment	
	Traffic optimization	

- *Applications for improving safety*: Safety applications have the ability to reduce traffic accidents and to improve general safety. These can be further categorized as safety-critical and safety-related applications [30].
  - *Safety-critical*: These applications are used in the case of hazardous situations like collisions. They include situations where the danger is high or is imminent. Such applications can access the communication channel with highest priority

and transmit messages. In this case, latency and reliability of messages has an important role in realizing the goal of enhancing safety of passenger.

- *Safety-related*:These include safety applications where the danger is either low but still foreseeable such as broadcasting curve speed warning messages or elevated in situations where a message informing about work zone warning needs to be transmitted. In safety-related applications, the latency requirements are not as important as in safety-critical.

- *Assistance for Non-safety applications*: These applications provide traffic information and enhance driving comfort. These services access the channels in the communication system, except the control channel. They access the channel in a low priority mode as compared to safety applications. Non-safety applications include applications for:

  - *Traffic optimization*: This helps vehicles in data collection and transmitting the traffic condition information for the vehicular network. Here the vehicles can detect conditions such as whether any vehicle is violating the speed limit of the lane or the count of neighboring vehicles on a particular lane is abnormally high that may lead to congestion on road. This allows vehicles that are approaching the congestion point to have enough time to choose their alternate routes, by avoiding such hotspots on roads. Various applications of VANETs are also depicted in Table 1.2 that provides a summarized description of the application areas where VANETs can be used [30] [31]. These applications can be further classified into, following categories:
  - *Infotainment*: This helps vehicles in providing Internet access, media downloading, instant messaging etc.
  - *Payment services*: This helps vehicles to facilitate electronic toll collection, parking management etc.
  - *Roadside service finder*: This helps in finding nearest fuel station, restaurants etc. This involves communication of vehicles with road side infrastructure and the associated database.

1.8 Fundamentals of Clustering

Clustering is the concept which is used in wireless communication from past many years [32, 33]. Grouping together of the mobile, sensor or the vehicular nodes located in a geographical area based upon certain rules is known as clustering. The clustering helps to make network more scalable and robust. The use of clustering algorithms helps in forming

kind of virtual groups termed as clusters. The main entities of clustering are Cluster Head (CH) and Cluster Members (CMs). Each cluster generally has one designated head or the leader known as the CH which is selected by other member nodes or CMs. CH becomes the leader for other nodes in the cluster and is responsible for cluster management, relaying information between the nodes within the clusters or with the other clusters. Generally, each CM could be elected as CH but in few priority [34] schemes nodes with additional functionality have higher probability to become CH. These CMs send their information to CHs at regular time intervals.

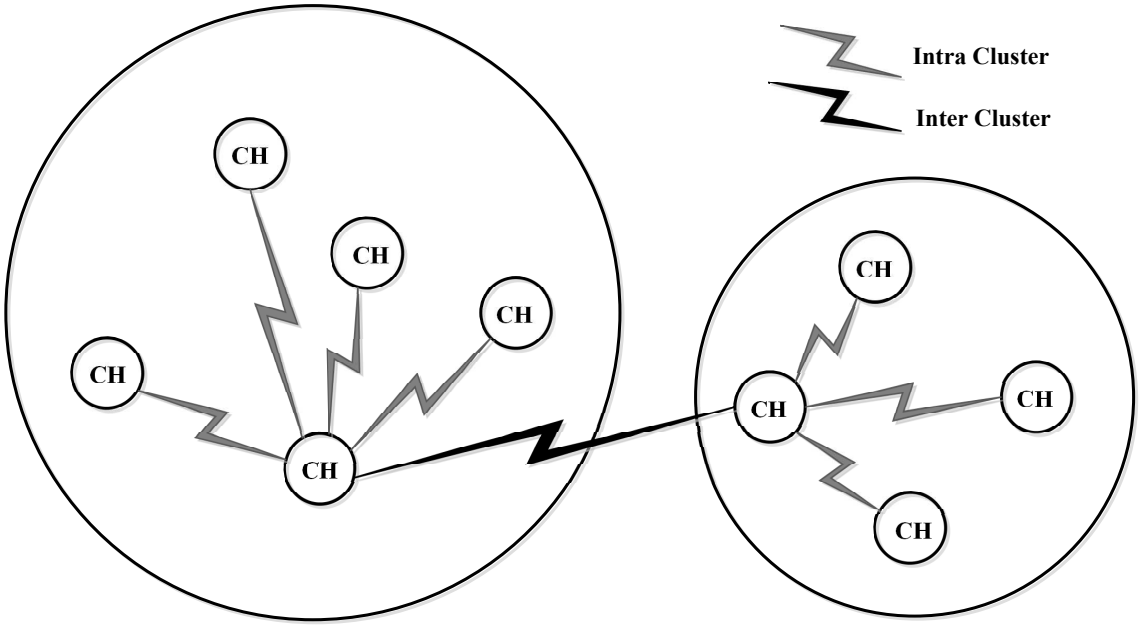


Figure 1.6: The structure and communication model of a basic cluster.

There are two types of communication in clustering which are inter-cluster and intra-cluster communication. As shown in Figure 1.6, a CM when communicating within the cluster is called as intra-cluster communication [35]. Some clustering techniques require a border node or a gateway node to help connecting clusters together. A CH collects the data from its CMs and communicates with other CHs and gateways which is called as inter-cluster communication to deliver data to destination.

Clustering enhances effective broadcasting and relaying of messages. Clustering reduces the signaling overhead because the links among the vehicular nodes within the same cluster are more stable. Clustering improves the consumption of scarce resources, for example bandwidth and increases the efficiency in data delivery. In large scale complex and distributed network, the clustering process is helpful in network management by dividing the network into smaller manageable segments. The advantages of using clustering schemes are: decreasing the number of messages transferring within the network, reducing congestion in V2R or V2V communication, increasing scalability of the network by creating smaller network segments, decreasing contention and hidden station prob-

lems, providing quality of service in routing. Along with these, benefits like dealing with dynamic topological and density changes are also important for VANET scenario.

Cluster size is not same for all the clusters and the variation depends upon the transmission range of the wireless communication device. Most of the clustering algorithms need to consider the cluster stability as a measure of performance. It is regarded as an important goal which every cluster algorithm tries to achieve [32]. Cluster stability can be defined as the number of times the CHs are changing and the CM associating with its CH.

Clustering schemes help in simplifying routing, efficient resource allocation, network management and making network more stable in terms of each node in the cluster. CHs assist in facilitating inter-cluster and intra-cluster transmission hence improving the network capacity and increasing the spatial channel reuse. In VANET, a good clustering algorithm should not only have low cluster maintenance overhead but also it provides stability during dynamic topological changes.

1.9 Security Challenges in VANETs

Due to rapid increase of modern wireless technology, VANETs may become ubiquitous in future. This will provide a serious challenge in instigating various security threats. As VANET is a subclass of MANET hence it inherits all the security issues of MANET. The malicious users can modify message which would be fatal to other vehicular users. Security mechanisms are required so that efficient delivery of message can take place along with providing guarantee about protection of personal information of given vehicles and passengers. It is important to provide security otherwise any vulnerability in the network can have disastrous consequences.

VANETs offer a lot of benefits to the society but there are number of security challenges to the research community also . In VANETs, the vehicle can perform routing operation with the help of various cooperative nodes within the network without using trust based infrastructure. Due to open nature of wireless communication medium, dynamic change in topology and lack of centralized monitoring medium, there are number of attacks that can be implemented in VANETs. Some of these attacks are Denial of Service (DoS), sending bogus information, spoofing, sybil attack, privacy attack, data trust attack, replay attack, physical tampering, brute force attack, black hole attack, traffic analysis, illusion attack etc [36]-[37].

The various types of attacks can be classified on the basis of nature, type of adversaries, target and impact on the robustness of the application. In a given network there are two type of attackers namely insider and outsider [38]. The insider attackers are authentic users that launch attack from inside the network and these users have full knowledge of

the given network and they exploit its normal functioning. On the other hand, outsider attackers are those which are not caused by authentic users of the network but they try to capture data or gain access using unfair means or illegally. The outsider attackers can launch various kinds of attacks if they are successful in getting unauthorized access to the network. There are various types of attackers such as malicious or rational and active or passive attacks, where each attacker generates an attack for some specific purpose [38]. The reliability of information shared between two vehicles depends on data authenticity. Also malicious vehicle can impersonate other vehicle hence there is need to build network of trust and ignoring information from untrusted senders.

The attacks in VANETs can be categorized into three different types such as threat to authenticity, threat to confidentiality and threat to availability. In the next section, the main threats posed by authenticity, confidentiality and availability in various situations and their consequences are discussed.

1.9.1 Threats to Authenticity

Authentication is a fundamental security requirements. The authenticity provides a mechanism to protect authentic vehicles from insider and outsider attackers[39]. It involves signing the sender messages before transmitting and verifying the signature after receiving these messages. The threat to authenticity is caused by attackers that infiltrates the network with a number of false identities hence disrupting the normal functioning of the network. These attackers launch different kinds of attacks like message alteration, message tampering, Sybil attack, location spoofing, privacy attack and certificate replication. In Masquerading attack, an attacker joins the VANET by using functional on-board unit. If the attacker acquires the number or some other Identity description (ID) of legitimate vehicles in the network then they contribute together to generate a variety of attacks such as false message generation, black hole nodes formation and others [36] [40].

In replay attack, the attacker uses the earlier received message and transmit it again and again in the network. It is basically used to poison the route table of other nodes. In order to avoid replay attack, there must be source of accurate time inside the message so that it can be used to compare recently received message with previous messages [36]. However, the WAVE protocol is protected against replay attack.

In Message Tampering attack, an attacker can alter the content of the message that are exchanged during either V2V or V2I communication. These message are considered as a threat to authenticity as they involve counterfeit response and falsify transaction application requests of the user. The legitimate vehicles or insider attacker can masquerade the network thereby producing false messages [40].

In Message Fabrication, message alteration or message suppression, an attacker can either alter the application so that it cannot send or receive application beacons or phys-

ically deactivate the inter-vehicle communication. An attack where attackers attempt to damage the given network by duplicating the identity of vehicle through other vehicles in the network is called Certificate and key replication attack. This attack would be done by an attacker to make the traffic regulatory authorities ineffective and also precludes from identification of vehicle during hit and run event [36] [40]. It was allied to broadcast tampering attack.

The GPS system maintains the position table and identities of the vehicles along with their geographical position. Various routing protocols in VANETs require position information for efficient working. Location privacy is also an important issue for VANETs. In GPS spoofing attack, an attacker generates false reading in the GPS devices thereby providing fake position information of the attacker to other vehicles. This attack is possible with the help of GPS satellite simulator with high signal strength as compared to genuine satellite. Hence, it sends false location messages to the vehicles in VANETs.

In Tunnelling attack, an attacker misuses the temporary loss of positioning information because the vehicles enter the tunnel and the attacker can inject false data into the on-board unit of the vehicle before it gets the trustworthy positioning information from the satellite when it comes outside the tunnel. In Position attack, an attacker can modify the position packet, generate bogus position packet and also drop position packets. In such types of attacks, the line of sight of the vehicle is blocked by some means. Position attack can also send multiple fake position of the attacker at the same time and disrupt the network working [36].

Sybil attack involves counterfeiting multiple attacks on the system. In this attack, an attacker transmit multiple message to other vehicles but each message consist of different false source identity so that originator is unknown and create illusion in network. Due to this attack, traffic jams may occur and vehicles would need to take alternate route. Also it make delivery of safety message difficult [37]-[40].

1.9.2 Threats to Confidentiality

In VANETs, messages are exchanges between different vehicles using open wireless medium. The confidentiality of messages is vulnerable when illegitimate collection of messages are done by the attacker using eavesdropping. Confidentiality can also be compromised by the collection of location information by the attacker which is available through broadcasts messages. Eavesdropping occurs when insider or outsider attackers gather the information of various road users without any knowledge about the users and then apply the gathered information at an occasion when they are not present. Anonymity and Location privacy are becoming significant concerns for nodes or vehicles in VANETs. Location privacy comprises of protecting or confusing the user's exact location. When users make continuous requests then it should not be possible to obtain the identity of given user so

that anonymity of users can be preserved [36]-[40].

1.9.3 Threats to Availability

The malicious vehicle or attacker can intimidate the V2V and V2R communication and this may result in unavailability of information to destined vehicle. The information is not available due to variety of attacks such as Denial of Service attack, Malware, Spamming, Broadcast Tampering and Black Hole attack [41] .

Denial of Service attacks in VANETs may be launched by different attackers such as insider or outsider attacking vehicle. The DoS attack attempts to diminish the network availability by carrying out flooding with duplicate packets or by jamming the functionality of the given channel. This type of attack blocks the delivery of any valuable information between two vehicles. Hence the network will become unavailable to authentic users. Due to such kind of attack there is may be some kind of loss in the network or even some catastrophe. In such scenario, the attacker can divide the network without negotiating the cryptography schemes[36].

A number of malwares like worms, Trojans and viruses also have the capability to effect the proper operation of the VANETs. The introduction of malware attacks can be done by rouge insider. It can be introduced when the on-board unit and roadsides units of the network take software and firmware update from insiders. Spamming attacks occur through introduction of spam messages into VANETs and results in enhanced the risk of transmission latency of the network. The spamming attack is comparatively difficult to handle or control as there is absence of any infrastructure and Certification Authorities (CA) to deal with these attacks.

In Broadcast Tampering attack, the intruder can generate fake message or disrupt the legitimate traffic warning. Such kind of attack causes accident by injecting false message. An insider attacker can initiate such kinds of attacks as they have access to cryptographic key that is used to decrypt the encrypted message whereas outsider attacker has to first obtain the parameters of encrypted message [40].

Black hole attack is a kind of DoS attack where some nodes does not forward messages to network or deliberately drops messages. When the given node starts dropping messages then route from that vehicle is impoverished. So these kind of nodes are called black hole nodes. In such kind of attack the attacking node sends route reply message with minimum cost to the sender node. Then, it drops all the packet instead of forwarding. Another variation of black hole attack is grey hole attack. In this type of attack the node can transmit only some packet and drop other. Such kind of attacks do not route all the safety message to destined node and hence create problems of availability in VANETs[36].

1.10 Security Requirements in VANETs

In order to provide countermeasure against different security attacks, a large number of requirements need to be satisfied. The messages transmitted in VANETs may belong to different classes and pattern based on the services. The broadcast message is mainly for safety services and unicast message is for non-safety services. The security mechanism must be different for different set of services. The following parameters define the list of security requirements available for VANET environments [42].

Availability: The communication network is less reliable if there is network overloading and channel jamming. The condition becomes worse when communication layer in the network is not consistent. Security mechanisms such as continuous monitoring and reputation based system will be useful for mitigating non-availability risk. It can be done by finding non-cooperative node by using cognitive radio technique or by channel exploration [42].

Message Integrity: The message integrity requirement does not allow change in the content of the message. This is used for both safety and non-safety messages as it protects against message modification. The authenticity and integrity cannot be separated otherwise it is difficult to find correct message's sender if message is modified by someone.

Confidentiality: Confidentiality requirements means protecting the message so that unauthorized parties cannot disclose it. ITS services are meant for physical safety of passengers on vehicles. Even if a vehicle that is not legitimate is involved in accident then the legitimate vehicle may be affected. So, the main goal of maintaining confidentiality services is to deliver safety messages to all the users in VANET. However, for certain applications, the receiving vehicles should process the message only if confidentiality has been guaranteed. Therefore, the messages in VANETs need to be signed and verified using source authentication [42].

Source Authentication: In safety application, the illegitimate vehicles must be prevented from generating safety messages as these vehicles do not have any proof of authenticity . Only those broad casted messages can be considered which are signed by authenticated user. This may be one of the basic requirements for vehicular communication services as these message are broad casted by a number of vehicles. Authentication also helps to distinguish between legitimate and illegitimate vehicle in the VANET.

Mutual Authentication, Authorization, and Access Control: In order to avoid Man-In-The-Middle (MITM) attack, mutual authentication is considered as an important mechanism. In commercial or non-safety services, there must be mutual authentication between client and service provider so that both are authenticated. A simple technique to carry out mutual authentication is symmetric key sharing between nodes in the VANETs. It may be designed as a plug and play service, thereby requires less processing and communication overhead. In order to handle a large scale network such as VANETs, public key cryptography is used as compared to symmetric key. This is due to the fact that with symmetric key, an attacker can compromise any node in the network. It then breaks the security and threatens the whole network. It can further break confidentiality of messages and impersonate the other vehicles [42].

Nonrepudiation: Vehicles injecting malicious message in the network must be identified reliably so that they can be isolated from the network. By incorporating digital signature, the vehicular network can help in providing nonrepudiation property.

Privacy Protection: As wireless technology becomes pervasive where the users are worried about individual's anonymity and non-traceability. However, guaranteeing anonymity and non-traceability may require enforcing non repudiation. The privacy and security must be applied to both safety and non-safety applications. Network operators and governmental authorities use legitimate process for traceability. But, if an attacker knows who is sending message, what one is sending, which application is one using and where he is going etc, then it can have severe consequences and hamper the whole security of the VANET [43]. However, non-traceability is one of the most challenging requirements to achieve in VANET as attackers can manipulate the sensitive information.

1.11 Security Proposals for VANETs

In recent years a large number of security techniques are proposed by researcher for VANETs. In the next section, the various techniques and solutions to remedy the security problem in VANETs are discussed in brief [37].
Security Hardware: The vehicle is equipped with a large number of hardware components. Two types of hardware components are used to achieve security in VANET, based on Event Data Tracker or Tamper Proof Devices. The EDR which is used to track the occurrence of specific event on road. It can be considered as similar to black box in an aeroplane. EDR is assumed to be tamper-proof storage device which can record all emergency event such as speed, direction, time, position and rate of acceleration. It can also be used to store critical event. The second type of hardware component is TPD which is

used to collect information and also encrypts/decrypts the messages. TPD has the ability to verify and sign the messages. It can be independent component and separated from external environment. It has its clock and chargeable battery so that it can be synchronized with trusted authority and RSU.

The Trusted Platform Module (TPM) are also used as an alternative to TPD due to their lower cost. TPD basically provides keys for secure communication and implementation of different primitives for cryptography. TPM is also capable of defense against various kind of attacks [37].

Vehicular Public Key Infrastructure: In VANET, Public Key Infrastructure (PKI) involves district, state and national registration authorities. These authorities work as a Centralized Certification Authorities. The CA is responsible for issuing certificates and private key pairs to various nodes in VANET. The OBU is a tamper proof device in which private key of vehicle is stored. The OBU has message receiving and cryptographic processing capabilities. The sender's vehicle signs safety message with its own private key and adds certificate to it for verification. The receiving vehicle can verify the received message by using public key available in the certificate. The receiver must need public key of CA in order to use certificate. The signing and verifying of the safety message causes a lot of overhead. The overhead of such system can be reduced by using faster encryption schemes such as Elliptic Curve Cryptography (ECC). The emergency messages such as accident warning, crash in VANETs needs to be signed. On the other hand, non-safety messages such as entertainment, parking slot availability can be transmitted without any encryption scheme. It is done because cryptography process involves a lot of overhead and computation cost due to keys acquiring, encryption and decryption process [37].

Group Signature: The group signature is a special type of digital signature. In this scheme, vehicles build a group area consisting of number of vehicles. The vehicles in the group can then receive group public key and session key from group manager. The public and session keys for each vehicle must be updated and transmitted after certain fixed time interval. Using these keys, the vehicles can sign the message without revealing their identities. But, this technique creates additional overhead because of frequent changing and transmission of signature and keys in the network. The main problem with this technique is that it is difficult to form group in dynamic environment. Other problem is that how vehicles communicate internally without forfeiting security needs.

Authentication: The authentication of the sender's message is also a fundamental requirement in vehicular communication. The sender signs the legitimate message be-

fore transmitting and receiver verifies the signature of the received message. The digital signature scheme is based on private/public key pair based cryptosystem and one way hash algorithm. The hash algorithm is applied to message before transmitting and it is encrypted using a private key. The resultant signature is transmitted along with the message and then receiver verifies the message using public key and hash algorithm for checking its validity [42, 44].

Certificate Revocation List: PKI implementation in vehicular network is a challenging task. In VANET, PKI does not revoke any vehicle which is sending false message. If some malfunction or malicious vehicle is found in VANET then its certificate must be revoked by the trusted authority. It is done because all the legitimate vehicles must forward actual messages and exhibit cooperative behaviour so that there trust factor in the network can be maintained. But, distribution and updating of revocation list which indicates malicious vehicles in VANET involves a lot of overhead. IEEE 1609.2 involves Certificate Revocation List (CRL) for all revoked certificates of malicious vehicles. However, CRL implementation depends on central trust authority. Revocation using Compressed CRL (RC^2RL) divides the CRL so that it can be distributed more easily. The CRL is made of self-verifiable parts. Short lived certificate is proposed by the IEEE 1609.2 standard so that they can automatically updated after certain time period [45]

Detection/Correction of Malicious Data: In VANET, an authenticate user can transmit message with false information. Such kind of users launch insider attack. These vehicles generate authenticated message using tamper proof device. For such kind of attacks, there is need of mechanism to verify sender along with legitimation of data received. Thus, a security mechanism should also verify data along with vehicle position. There is a plausibility for checking the message along with vehicle position verification. Another issue is to determine the time duration for which messages need to be stored and decision taken for detection of malicious senders and messages. Some trust based metrics are required for creating and collecting the above information [45].

1.12 Principles of Connected Dominating Set

In an undirected graph $G = (V, E)$ the dominating set S is the subset $S \subseteq V$ such that each node in V is adjacent to at least one node in S. The dominating set algorithm is used to construct a building block for solving problems where the network needs to be partioned into smaller clusters for routing by creating effective backbone using the dominating sets. Connected Dominating Set (CDS), C of the graph$G$ is the dominating set of G which induces the connected sub graph of G. The CDS nodes are the dominator and

the other nodes are dominatees [46] [47] .

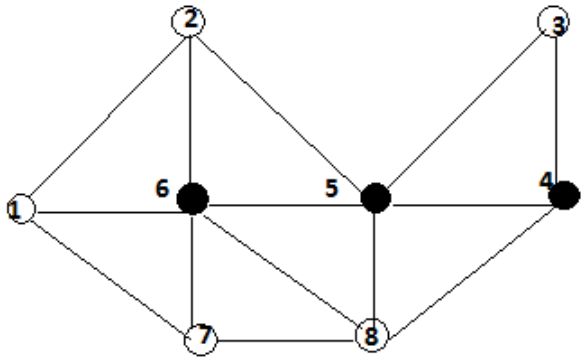


Figure 1.7: A connected dominating set structure in a graph.

Figure 1.7 shows that a set of fully shaded nodes forms CDS as there is path among the nodes in CDS which does not use the nodes which are not part of CDS. So, these darker nodes are dominators and others are dominatees. Many CDS algorithms have a coloring mechanism for effectively describing a CDS. All the nodes are initially colored white then the dominatees are colored grey whereas the dominators are colored black [48]. Routing issues are minimized if the main routing tasks are restricted to the dominator nodes. Therefore, if a source node need to deliver the packets to some destination node, then this packet is forwarded to the dominator which then forwards the packet to dominator neighbor of the destination and the neighbor dominator forwards the packet to the destination node. CDS forms a communication backbone in which the nodes not belonging to this set communicate by transferring the messages through neighbors of the sets. For constructing virtual backbone for the wireless networks, CDS has been used widely [49], [50].

1.13 Research Gaps

Clustering has been investigated by the research community from different perspectives in many applications used in VANETs. In order to have low communication cost and delay during message exchange between vehicles, there is a requirement of a clustering protocol that allows vehicles to communicate efficiently in this environment. Several significant works are carried by researchers in this direction but still there is scope for further improvements in the existing solutions. Hence, based on the above discussions, following gaps are identified for secure clustering in VANETs.

- In high mobility environments, clusters become unstable as clustering/de-clustering is constantly executed with high frequency. A predictive method for minimizing the number of CH changes needs to be developed by using some predictive approach that adapts to the existing traffic environment.

- A fully distributed algorithm for cluster size estimation on the basis of density estimation over both highway and urban scenarios at different densities and area sizes is required. The effect of background traffic on the efficiency of existing clustering schemes also needs to be investigated.

- Most of the schemes have frequent cluster membership which changes with time with changes in vehicles mobility, and density. This may create a bottleneck for efficient clustering in VANETs. Hence, a high stability clustering protocol based on the characteristic such as-mobility and density of vehicles need to be investigated further.

- Techniques for clustering should include an efficient mechanism for preserving security at various levels in VANETs. This includes the design and implementation of new secure key distribution and management scheme at various levels.

- A clustering scheme based on trust management and security of cluster member nodes to increase the reliability should also be investigated further. The proposed scheme should provide the concept of security groups defined at different levels in VANET environment.

1.14 Research Objectives

From the above discussion and gap identified, following are the objectives of the thesis.

- To review and analyze various secure data clustering schemes used for VANETs.

- To propose an efficient secure clustering scheme by considering the parameters such as-varying density, direction, speed, and mobility of vehicles. The proposed clustering scheme enhances the security and reliability of messages transmission between different vehicles.

- Design and implement the secure data clustering scheme.

- Test and validation of the proposed scheme in different scenarios with respect to various metrics.

1.15 Thesis Organization

The thesis is organized as follows

Chapter 2, provides an analysis about various challenges and existing solutions about the existing clustering algorithms and security schemes in VANETs. A complete taxonomy on clustering in VANETs has been provided based upon various parameters. A detailed discussion is also provided for each category of clustering and security aspects in VANETs which includes various challenges, existing solutions and future directions.

Chapter 3, introduces the predictive clustering schemes in detail. We first describe the architectural model that prediction based model uses and then provides efficient algorithms for future mobility predictions and average variations of vehicles on the road. Algorithms estimate the clustering duration and determine the number of vehicles in a cluster. To further increase the reliability of communication between different nodes, a novel Learning Automata based hybrid clustering scheme for vehicles is also proposed. The proposed solution estimates future position of the vehicles more accurately. An algorithm called as Predictive Clustering Algorithm using Learning Automata is also proposed where the actions of Learning Automata are rewarded or penalized based upon their current prediction accuracy and their previous actions where they have predicted the future positions. Various parameters of these clustering schemes are listed along with their evaluation parameters.

Chapter 4, discusses the novel clustering model to prevent the information security attacks that result in alteration or misuse of information. Clustering-as-a-Service security model based on Distributive Intrusion Detection System is proposed for detecting attacks in VANETs. The proposed scheme utilizes vehicular clusters and a cloud based architecture that enhance its capabilities for intrusion detection and thus can be utilized for efficient transmission of heterogeneous multimedia data. The security of data from malicious activities is ensured by utilizing a standard cryptographic technique.

Chapter 5, presents a clustering scheme in a realistic VANET environment by using data dissemination based communication model. A design of novel secure clustering scheme for efficient data dissemination between different devices in Vehicular environment has also been discussed . A new trust metric based on dynamically varying transmission characteristics of vehicles which is evaluated both at local, and global levels is defined for trust computation among different devices. This trust metric is used to establish the current security level of vehicles and is the key parameter for creating secure clusters. Algorithms for secure clustering and trust establishment are also designed in the proposed scheme.

Chapter 6, provides the conclusion and also discusses the limitations and future work.

Chapter 2

Literature Review

Vehicular Ad Hoc Networks(VANETs) have emerged as a new class of efficient information dissemination among communities of users mainly because of their wide range of applications in different domains such as ITS, safety applications, online entertainment during the mobility of the vehicles etc. Vehicles in VANETs behave as intelligent machines and thereby can provide various resources to the end users with/without the aid of the existing infrastructure. But due to the high mobility and sparse distribution of the vehicles on the road, it is a challenging task to route the messages to their final destination. To address this issue, clustering has been widely used in various existing proposals in literature.[1]

Clustering is a mechanism of grouping of vehicles based upon predefined metrics such as density, velocity, geographical locations of the vehicles etc. In this chapter, various challenges and existing solutions used for clustering in VANETs are reviewed based upon a large number of parameters. Based upon this categorization, a detailed discussion is provided for each category of clustering. Also, a comprehensive analysis of all the existing proposals in literature is provided with respect to vehicular topology, additional infrastructure requirements, road scenario, node mobility, data handled, and relative direction, density of the nodes, relative speed, communication mode, and communication overhead.

VANETs consist of Vehicles/Mobile nodes communicating with each other over wireless links with/without existing infrastructure [12] [51]. Vehicles have the capability to communicate directly with other vehicles in Peer-to-Peer (P2P) manner or indirectly using the existing infrastructure alongside the road. Vehicles and roadside infrastructure need to be equipped with dedicated hardware for providing safety and security to the passengers. Additionally, standardization of wireless communication technology is also required for providing entertainment to the passengers. Therefore research on VANETs

[1]The content of this chapter has been taken from :

- Rasmeet S. Bali, Neeraj Kumar, Joel JPC Rodrigues, "Clustering in Vehicular Ad hoc Networks: Taxonomy, Challenges and Solutions". Vehicular Communications, Elsevier, Volume 1, Number 3, pages 134-152, 2014.

has been receiving a lot of attention in last few years, both on the algorithmic aspects as well as on standardization of IEEE 802.11p WAVE and IEEE 1609 standards. In a clustering scheme, the mobile nodes are divided into a number of virtual groups based on predefined rules. These virtual groups are called clusters. In a cluster structure, mobile nodes may be assigned a different status or function, such as CH , Cluster-Gateway(CG), or cluster-member. CH normally serves as a local coordinator for its cluster, by performing intra-cluster transmission arrangement, data forwarding etc. A CG is a non-CH node with inter-cluster links, so it can access neighboring clusters and forward the information between clusters. A CM is usually called as an ordinary node, which is a non-CH node without any inter-cluster links.

The notion of cluster organization has been used for MANETs in a number of issues such as routing, security and QoS. [12] However, due to the characteristics of VANETs such as high speed, variable density of the nodes, clustering schemes which are proposed for conventional MANETs may not be suitable for VANETs. Due to the time taken for cluster formation and maintaining a cluster structure, clustering requires additional control overhead. Thus, a good clustering algorithm should not only focus on forming minimum number of clusters but also dynamically maintains the cluster structure without increasing a high communication overhead over the network. A clustering also allows the formation of a virtual communication backbone that supports efficient data delivery in VANETs to improve the consumption of scarce resource such as bandwidth.

A low cost clustering method should be able to partition a VANET in a short time with little overhead of control message broadcasting. Hence, VANETs must follow a tight set of constraints as compared with MANETs and therefore require specialized clustering scheme. A clustering algorithm should also be distributed, with no central coordinator. The algorithm should also handle the locality property, i.e., single topology change should have as local impact as possible on the cluster topology and should be able to detect and react to the topology changes. Because of the high degree of mobility in VANETs, a clustering algorithm should have fast convergence and reduced overheads to minimize the time lost in the clustering process.

2.1 Challenges in Clustering

As VANETs have been used in various applications with the goal to provide safety and comfort to passengers, hence there is a requirement of optimized solutions for clustering in VANETs. Also, due to large number of nodes and lack of routers, a flat network scheme may cause serious scalability and hidden terminal problems. In VANETs a possible solution to above problems is the use of an efficient clustering algorithm. However, there are a number of challenges that need well designed solutions for clustering of vehicles. Some

of these challenges are high mobility of the vehicles, sparse connectivity in some regions and security. Thus, based on these observations, we have categorized existing vehicular clustering techniques into various subcategories. This standardization of subcategories of clustering, helps us to provide a comparative analysis of all the reviewed clustering protocols.

The existing clustering schemes have large and varied nature of clustering parameters and therefore it is difficult to consider a single parameter for evaluating their performance. To accommodate this diversity, all the parameters were analyzed and then synthesized into six standard categories. These six categories are classified based on parameters that primarily impact vehicular movement, characterize efficiency of the clustering technique and constrain the network performance. Vehicluar movement is affected by two categories that are 'vehicle density' and 'vehicle speed' whereas next three categories named as 'cluster stability', 'cluster convergence' and 'cluster connect time' define efficiency of the clustering technique. The last category, called 'transmission efficiency' impresses the network performance.These categories are defined as follows,

- Vehicle density designates the average number of vehicles defined in the terms of vehicles per kilometer (km) or vehicles per lane [34]. For urban scenarios, vehicle density has a higher value as compared to highways.

- Vehicle speed is the range of speeds considered for simulation by a particular protocol in terms of m/sec or km/hr. A speed range that varies realistically indicates better adaptability.

- Cluster stability is the average life time of a cluster. A high value of cluster stability indicates a better clustering protocol.

- Cluster connect time refers to percentage time duration that a vehicle stays connected to a single cluster. A high value of cluster connect time indicates the higher suitability of a protocol for clustering.

- Cluster Convergence refers to the duration required for all the nodes to join a cluster at the initiation of a clustering scheme. The suitability of a clustering scheme for VANETs is more when it exhibits low clustering convergence.

- Transmission efficiency is described as the average number of messages or packets that are transmitted or received by a cluster member during a time duration. High transmission efficiency shows that a clustering scheme is more effective in data dissemination.

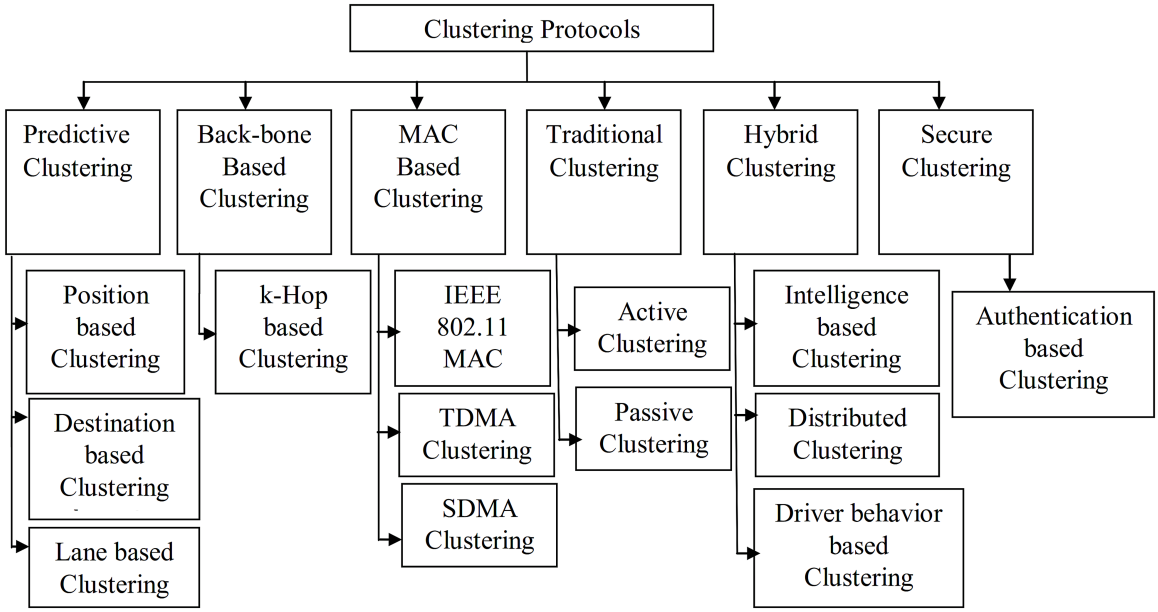


Figure 2.1: Taxonomy of existing clustering approaches for VANETs.

2.2 Taxonomy of Clustering in VANETs

For efficient communication among the nodes in the network, stable clustering is required. In this direction, many researchers have used different techniques to create cluster among the nodes. Some of these techniques consist of the use of signal strength received, position of the node from the cluster head, velocity of the nodes, direction and destination of node. Keeping in view of the above issues, a detailed taxonomy of various clustering algorithms is described in Figure 2.1 as follows,

2.2.1 Predictive Clustering

In predictive clustering, the cluster structure is determined by the current geographic position of vehicles and its future behavior. This vehicle traffic information helps to associate priorities which were used to assist in cluster formation. The future position and the intended destinations of vehicles have been used in the literature to form clusters in VANETs. Some of these protocols are further classified into position based and destination based clustering and are described as follows.

2.2.1.1 Position based Clustering

Position based clustering is a technique of forming clusters on the basis of geographic position of the vehicle and CH. Salhi *et al.* [51] proposed a position based clustering algorithm called as New Aggregate Local Mobility (NEW-ALM) algorithm, which is an improvement to the existing Aggregate Local Mobility (ALM) algorithm. The cluster

structure is determined by the geographic position of the vehicle and the CH is elected based on priorities associated with each vehicle. A hash function based on the estimated travel time is used to generate this priority for the vehicle. The stability of the system is improved by electing the vehicles having a longer trip as the CHs. Though this solution gives a stable cluster structure but its performance has not been tested in sparse and jammed traffic conditions which are very frequent in dense urban scenarios.

Wang *et al.* [52] proposed another position based clustering algorithm. It is a cross layer algorithm based on hierarchical and geographical data collection and dissemination mechanism. The cluster formation in this protocol is based on the division of road segments. However, this protocol incurs more overheads for V2V and V2I communication. Its performance is also dependent on availability of roadside infrastructure. Maslekar *et al.* [53] proposed a new CH election policy for direction based clustering algorithm called as Modified Clustering based on Direction in Vehicular Environment (MC-DRIVE). The primary functioning of MC-DRIVE is based on the parameter $TH_{distance}$. This value yields an optimal value of the cluster and is dependent on the speed and the radio range of the vehicles approaching the intersection. The proposed clustering algorithm is able to maintain the stability of the cluster in terms of the number of nodes within a cluster. This also helps to achieve better accuracy in density estimation which is the basic parameter for estimating $TH_{distance}$. It is also observed that the accuracy can be further improved by reducing the radio range up to a predefined threshold value. However, any further reduction in the radio range leads to an increase in the number of CHs that results an increase in overhead of the system.

Wolny [54] optimized existing Distributed and Mobility-Adaptive Clustering (DMAC) algorithm presented in [55] by estimating road traffic mobility more efficiently. The main idea for Modified-DMAC was to increase the cluster stability by avoiding re-clustering when groups of vehicles move in different directions. The algorithm is based on periodical transmission of status message and it also forms k-cluster so that nodes can be k-hops away from CH. This is achieved by introducing Time-To-Live(TTL) parameter in messages sent by the nodes. Modified-DMAC also introduced a method for estimating the connection time, that was denoted as freshness in DMAC between two moving nodes. By periodic computation of freshness value, re-clustering was avoided when two nodes are within the connection range for a short period of time which provides an increased cluster stability. Although Modified-DMAC increases the algorithm overhead but it reduces the number of cluster changes thereby increasing the stability of cluster formation. Its performance has also not been evaluated in jammed traffic conditions which are very frequent in dense urban traffic scenarios.

Fan *et al.* [56] proposed a clustering scheme (DCA) where a utility based cluster formation technique is used by extending the concept of spatial dependency which was

initially proposed in [57]. The utility function is defined in terms of position and velocity of the vehicle. The threshold is computed based on the previously available traffic statistics. A status message is periodically sent by all the vehicles. After receiving this information, each vehicle chooses its CH based on the results produced by the utility function. The node with the highest value is chosen as the CH, by taking into considerations the specific characteristics of VANETs. However it still applies many fixed weights and the parameters like fixed cluster formation interval which implies a synchronous formation of clusters. This scheme fails to adapt to traffic dynamics and is also not effective for frequent cluster re-organization.

Wang *et al.* [58] proposed a Cluster Establishment algorithm, which uses the Relative Angle (CERA) to obtain the distance from a potential CH to its neighbor cluster associate nodes. The relative angle is measured between the movement direction of the current vehicle and its closest neighbor so as to minimize the number of CHs. The algorithm also limits the number of vehicles in each cluster by electing cluster associate node on the basis of the distance between potential CM and their neighboring CH. This clustering scheme also estimates whether or not a particular vehicle has any neighbor or not by gathering relative angle information and distance. This position information is then used to establish the clusters and their cluster members.

Hassanabadi *et al.* [59] proposed Affinity Propagation based Algorithm (APA) for clustering in VANETs in a distributed manner. The vehicles are grouped into clusters based upon minimum relative velocity and minimum distance between a CH and its members. The CHs are elected periodically using affinity propagation. Every node transmits the responsibility and availability messages to its neighbors and then makes an independent decision on clustering. The affinity propagation has an aim to maximize similarities which represents the suitability of the vehicles for clustering. The similarity function of any vehicle pair is determined based upon euclidean distance between the current position and the future position of the vehicular node and the CH is selected based on the minimum value of this similarity.

2.2.1.2 Discussion on Position based Clustering Protocols

In recent times, many position based clustering protocols have been proposed by considering various characteristics. Among these proposals, the protocols based on the vehicles positions are more suitable to VANETs due to their resilience to handling the variation in relative positions of vehicles [60]. Table 2.1 provides a relative comparison of these protocols with respect to key parameters that influence position based clustering. Since the above clustering protocols primarily rely on the position of the vehicle, so the range of values for vehicular density and vehicle speed exhibit a variation for all the protocols as shown in Table 2.1. However, the value of cluster convergence rate is low even if vehicle

density and cluster dynamics increase which leads to a better cluster stability for these schemes. The variation in cluster size also affects performance in terms of mean cluster diameter and dismiss threshold for position based clustering. The value of transmission efficiency which effects packet delivery ratio is also average in these proposals. Therefore, from Table 2.1, it can be concluded that cluster convergence and cluster connect time need further analysis to improve the overall efficiency of these clustering schemes.

Table 2.1: Relative comparison of Position based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
NEW-ALM [51]	H	H	H	H	L	H
PPC [52]	L	H	L	L	L	H
MC-DRIVE [53]	L	M	L	L	L	H
DCA [56]	L	M	H	L	L	H
CERA [58]	L	H	L	M	L	M
APA [59]	L	H	L	M	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.1.3 Destination based Clustering

Destination based clustering techniques take into account the current location, speed, relative and final destination of vehicle for cluster formation. The destination is known in prior using the navigation systems. Various proposals in this category are described as follows.

Farhan *et al.* [61] proposed an algorithm for improving the accuracy of GPS devices called Location Improvement with Cluster Analysis (LICA). Vehicles are able to collect real-time data and relay the information to other vehicles, guiding the drivers to reach the destination safely and efficiently. To measure distance, time-of-arrival and Received Signal Strength (RSS) techniques are used. LICA uses a modified tri-alteration technique in which multiple measurements can be taken and average value can then be used as the final distance measurement resulting in a set of possible refined x-y coordinates on which a cluster analysis is applied. Thus it allows more weight to be given to computed data which results an improvement in nodes location estimation. By using accurate distance measurements, the location error is reduced in LICA and thereby helps in achieving higher performance.

Tian *et al.* [62] presented a clustering method based on a vehicles position and moving direction. The clustering method is based on Euclidean distance which uses the position information as well as the moving direction to divide the vehicles into clusters. Each vehicle broadcasts beacon message that include its ID, latitude, longitude, direction and

time to the whole network. The receiving vehicle first checks the beacons hop count value and if the number of hops is larger than the maximum value, it discards this beacon. Then sender vehicle updates its topology table by calculating the distance between the vehicles. The CHs are generated by selecting the vehicle with minimum distance parameter as the CH. The remaining vehicles are then divided into clusters.

The Adaptable Mobility-Aware Clustering Algorithm based on Destination (AMACAD) [63] is based on final destination of vehicles to enhance the clustering stability. It operates in a distributed way with the final destination, relative destination, speed and current location of a vehicle as parameters to calculate a metric called $F_{v,2}$ by exchanging messages with its neighbors. This manages to improve the lifetime of the cluster and thus decreases the number of CH changes. It implements an efficient message dissemination mechanism to respond in real time and avoid global re-clustering. The algorithm is based on certain assumptions like the destination of each vehicle is known and the routing is geographic based. The minimum value of $F_{v,2}$ is the selection criterion used by a vehicle to join a cluster. The Region Group Mobility model proposed in [64] was also modified to make it suitable for VANETs. In AMACAD, the authors evaluated how the variation of the transmission range and speed affects the AMACAD performance. The algorithm works well when average speed of vehicles is almost constant which makes this scheme more efficient in urban areas.

Santos *et al.* [65] proposed Cluster Based Location Routing (CBLR) algorithm to choose CHs in VANETs. This algorithm is based on the regular transmission of beacons, which are used to distribute the state of the vehicles. According to the states of the sending vehicle, the nearby vehicle then choose the appropriate state. To cope with the changes in the topology, each node maintains a neighbor table, in which it lists the nodes with which it can exchange information. The update of this table is done according to the received beacon messages.

2.2.1.4 Discussion on Destination based Clustering Protocols

Table 2.2 compares the destination based clustering schemes. In order to keep the clustering process stable, the frequency of cluster changes should be minimized because a vehicle only leaves a cluster when it encounters a CH whose destination is more similar as compared to destination of current CH. Thus exploiting the vehicular behavior by taking into account the final destinations of vehicles enhances the cluster stability and improves the Transmission Efficiency in message delivery. It also results in higher Cluster Connect Time (CCT) as the probability of a vehicle leaving a cluster is generally low due to similarity in their destinations. However, in case the number of vehicles in a cluster becomes large, message broadcast results in high transmission overhead which impacts Transmission Efficiency (TE). The impact of vehicle density and vehicle speed on clustering is

also not too significant as their values generally are more dependent on characteristics such as existing traffic conditions and road scenario for destination based clustering protocols. Thus, these protocols can be integrated with algorithms that minimize message retransmissions to further improve their efficiency.

Table 2.2: Relative comparison of Destination based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
LICA [61]	M	M	M	M	M	H
EUCLIDEAN DISTANCE [62]	H	H	M	H	M	H
AMACAD [63]	L	H	M	M	M	H
CBLR [65]	L	M	M	M	M	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.1.5 Lane based Clustering

Lane based clustering creates cluster structures based on estimation of the lanes on which the vehicles are moving. Some of the main proposals in this category are explained as follows.

Fan *et al.* [66] proposed Broadcasting based Distributed Algorithm (BDA) aimed at stabilizing the existing clusters. This scheme requires the vehicles to have only single hop neighbor knowledge. This approach attempts to improve the performance of classical clustering algorithms by making them aware of the vehicle's movement. However, all nodes attempt to re-evaluate their conditions by computing utility values periodically which may cause traffic overhead and therefore consume more bandwidth. BDA gives maximum priority to leadership duration for cluster formation, which is difficult to compute and may result in large overhead by a node during cluster formation.

Almalag *et al.* [67] presented a lane-based clustering algorithm based on the traffic flow of vehicles. The proposed algorithm is based on the assumption that each vehicle knows its exact lane on the road through some lane detection system and in depth digital street map that includes lane information. It also uses GPS combined with wheel odometer for lane detection of a vehicle. The authors use the same general idea as the utility algorithm in [66], but apply a different set of rules. Each vehicle computes and broadcasts its Cluster Head Level (CHL) along with its speed and other parameters. The vehicle with the highest CHL will be selected as the CH. CHL is determined on the basis of network connectivity level of vehicles and average velocity of traffic flow.

2.2.1.6 Discussion on Lane based Clustering Protocols

Lane based clustering algorithms use the availability of lane information to create clusters. Table 2.3 illustrates that the above two schemes have low number of CH changes, so the cluster stability is improved. The transmission overhead of these schemes is also reasonable on account of small number of retransmissions of broadcast messages since re-clustering is performed only at lane intersections. These schemes also have an improved transmission efficiency due to better broadcasting reachability and good CH lifetime as the vehicles in the same lane move with constant relative speed which results in highly stable cluster dynamics. These schemes also have a small delay that demonstrates their usefulness for maintaining the cluster even for high mobility of vehicles. The CCT for these clustering schemes is also reasonable. The observed values of vehicle characteristics such as density and speed is on the lower scale since these protocols are adaptive for urban environment due to the constraint of vehicle traveling in the same lane.

Table 2.3: Relative comparison of Lane based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
BDA [66]	M	H	M	M	L	H
Lane based Clustering [67]	L	H	L	M	L	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.2 Backbone Clustering

Backbone based clustering technique is based on forming a backbone for cluster communication. The backbone then manages the network and performs function such as communication and assists in CH election. Various backbone based clustering techniques are classified as follows.

2.2.2.1 K-hop clustering

In multi hop or k-hop clustering, cluster structure is controlled by the hop distance. The distances between a CH and the members of a cluster are within a predetermined maximum number of hops which can be one or more than one. Some of the research proposals in this category are explained as follows.

Zhang *et al.* [68] proposed a multi-hop clustering scheme based on the mobility metric for representing N-hop mobility. A vehicle is allowed to broadcast beacon message periodically and it calculates relative mobility based upon two consecutive beacon messages received from the same node in N hop distance. Each vehicle node then calculates the

aggregate mobility value, which is the sum of relative mobility values into weight value for all the neighboring nodes in N-hops. The vehicle nodes then broadcast their aggregate mobility value in the N-hop neighborhood and the vehicle with smallest aggregate mobility value is selected as the CH. The vehicles join a cluster if they receive the beacons broadcast from the CH node. When a vehicle node receives multiple beacon messages, then it selects the CH which is closest one in terms of number of hops. If several CHs have the same hops, then the vehicle node joins the cluster which has the lowest relative mobility.

Zhang *et al.* [69] proposed a novel K-hop clustering approach that takes into account the highest connectivity, vehicle mobility and host ID to select CH. The proposed clustering approach modifies max-min K-hop heuristic approach defined in [70] for cluster formation by considering highest connectivity in terms of signal strength and vehicle mobility. This scheme is also capable of dynamically adjusting the period of broadcasting location information according to vehicle velocity in order to suppress transmission overheads. Moreover, the distance-based converge-cast is deployed to collect all memberships within the cluster, including the members located on the cluster border. Another feature of this approach is its ability to enhance cluster stability due to vehicle activation and deactivation by considering the radio link expiration time and the number of vehicles connected to a CH. The cluster-based topology discovery scheme proposed in this approach also utilizes the advantage of K-hop cluster architecture to improve the network topology scalability. It improves the network topology stability with a capability to tolerate false routes and balance traffic loads by considering the inter-cluster link expiration time. By taking into account the factor of vehicle mobility, it reduces the overhead and the latency caused by route path recovery.

Wei *et al.* [71] proposed a robust Criticality-based Clustering Algorithm (CCA) for VANETs that employed the metric called as 'Network Criticality' to perform the process of building clusters. Network Criticality is motivated from the definition of random walk between graphs. It is a global measure on a graph which quantifies the robustness of a network graph to the environmental changes, such as traffic shifts, topology modifications, and changes in the origin and destination for the traffic. The CCA algorithm improves the lifetime of clusters, and it provides a more stable structure, especially for multi-hop VANETs.

Dror *et al.* [72] proposed a distributed randomized two-hop Hierarchical Clustering Algorithm (HCA) which was influenced by the work presented in [73]. HCA forms Time-Division Multiple Access(TDMA) like synchronized clusters. In order to reduce the number of collisions by simultaneous transmissions in the same cluster, transmissions are only allowed on the assigned slots by the CH. The algorithm differs from other clustering algorithms for VANETs as it is capable of creating clusters with a larger span from the

CH. It does not require knowledge of the vehicles' locations. The algorithm handles the channel access and does not assume any lower layer connectivity. Even though HCA forms few redundant clusters, the formed clusters are much more stable and robust to topological changes caused by vehicles. However, the mobility pattern influences the algorithm's behavior and has a great impact on the cluster stability. Nevertheless, HCA also suffers the difficulties in terms of inter cluster interferences which causes redundant cluster changes and message loss due to message collisions.

2.2.2.2 Discussion on K-hop Clustering Protocols

Multi-hop clustering algorithms shown in Table 2.4 utilize the advantages of K-hop cluster architecture to improve clustering efficiency. It is evident from Table 2.4 that K-hop clustering schemes, have better cluster stability as well as low cluster dynamics. This can be attributed to the reduced variation in CH and CM lifetime. Thus, K-hop clustering schemes can provide improved and reliable performance for VANETs, especially for large multi-hop wireless networks when number of vehicles increases in the network. However, the impact of vehicle speed and behaviour of vehicle density also needs further analysis as its effect has not been investigated in detail in these protocols. Although these protocol have better cluster convergence, but they suffer from inter cluster interference which needs to be analyzed for improving transmission efficiency and reducing transmission overhead. The increased transmission efficiency and hierarchical clustering structure results in larger span as compared to single-hop cluster spans that result in good cluster convergence and large cluster connect time.

Table 2.4: Relative comparison of K-hop clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
N-HOP USING REL. MOB. [68]	M	M	L	L	M	H
K-HOP [69]	M	M	L	M	M	H
CCA [71]	L	H	L	L	H	H
HCA [72]	L	M	M	M	H	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.3 MAC based Clustering

Several Medium Access Control (MAC) based clustering techniques have been proposed for cluster formation in VANETs. These techniques use IEEE 802.11 MAC protocol to generate clusters. Some of popular MAC based protocols are discussed as follows.

2.2.3.1 IEEE 802.11 MAC based Clustering

Su *et al.* [74] proposed a cluster based Multichannel communication scheme that integrates Clustering with MAC protocols (CB-MMAC). The proposed scheme consists of three core protocols called Cluster Configuration Protocol that groups all vehicles in the same direction into clusters. The Inter-cluster Communication Protocol which dictates the transmissions of real-time safety messages and non-real-time traffics among clusters over two separate IEEE 802.11 MAC-based channels respectively and the Intra-cluster Coordination and Communication Protocol that employs Multichannel MAC algorithms for each CH vehicle to collect and deliver safety messages from/to CMS using the upstream TDMA/downstream-broadcast method and allocating available data channels to cluster member vehicles for non-real-time traffics. The proposed scheme requires the use of two transceivers where one is used for delay sensitive communication within the cluster, while the other is used for inter-cluster data transfer.

Bononi *et al.* [75] proposed a cross-layered clustering scheme for fast propagation of broadcast messages which is called as Dynamic Backbone Assisted MAC (DBA-MAC) scheme that may be considered an extension of the MAC scheme described in [76]. A dynamic virtual backbone infrastructure is established through a distributed proactive technique. The backbone formation process considers the current distance among candidate backbone vehicles and the estimated lifetime of the wireless connection among neighboring backbone member. The scheme has been eveluated with DBA-MAC that has been shown to be compliant with IEEE 802.11 Distributed Co-ordination Function (DCF) systems, and its performance is comparable in terms of reliability, and overhead reduction as compared to simple 802.11 MAC flooding scheme and fast broadcast protocol proposed [74] protocols.

2.2.3.2 Discussion on IEEE 802.11 MAC based Clustering Protocols

MAC based protocols exhibit increased percentage collisions and average message delivery delay that results in lower transmission efficiency and high transmission overhead. These overheads occur due to an increased contention because of increase in number of vehicles or vehicle speed. Message delivery delay is mainly caused by mobility and sparse distribution of vehicles. It directly impacts the application design and deployment for VANETs. Liu *et al.* [77] identified message delivery distance and density of vehicles as two main factors for such behavior based on a bidirectional vehicle traffic model. The consideration of bi-directional traffic also effects cluster connect time and results in lower cluster convergence. However, in these protocols, the incurred overhead for delivering safety messages is decreased by reducing channel contentions for achieving timely and reliable delivery of safety messages. Due to low average relative speed among CHs, the overall impact of

variation of vehicle speed on these clustering schemes is also low as shown in Table 2.5.

Table 2.5: Relative comparison of IEEE 802.11 MAC based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
CB-MMAC [74]	M	H	H	L	L	H
DBA-MAC [75]	M	L	L	L	L	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.3.3 TDMA based Clustering

The process of assigning time slots using TDMA technique can be done using clustering where slots are assigned to CMS for data transmission. Some of the proposals in this category are described as follows.

Biswas *et al.* [78] proposed Vehicular Self-Organized MAC (VeSOMAC) protocol based on a self-configuring TDMA slot reservation protocol which is capable of inter-vehicle message delivery with short and deterministic delay bounds. To achieve the shortest delay, vehicles determine their TDMA time slot based on their location and movement on the road. The TDMA slot assignment is designed to be in the same sequential order with respect to the vehicles physical location. The process of assigning time slots is performed without using infrastructure or virtual schedulers. However, the assumption taken in this scheme about forwarding messages without processing time or propagation delay is unrealistic. This is because if the message needs to be delivered from the tail to the head of the platoon, it will need a time frame for each hop.

Omar *et al.* [79] proposed a Multichannel MAC protocol for VANETs (VeMAC), that aims to reduce interference between vehicles and reduce transmission collisions caused by vehicle mobility. VeMAC is based on TDMA scheme for inter-vehicle communication. All vehicles moving in both directions and RSUs are assigned time slots in the same TDMA time frame. Also, VeMAC is designed based on one control channel and multiple service channels in the network in compliance with DSRC/WAVE standards. VeMAC assumes that there are two transceivers on each vehicle and that all vehicles are time synchronized using GPS. The first transceiver is assigned to the control channel, while the second transceiver is assigned to the service channels.

Gunter *et al.* [80] proposed Cluster Based Medium Access Control Protocol (CBMAC), where the CH takes on a managerial role and facilitates intra-cluster communication by providing a TDMA schedule to its cluster members. The CBMAC protocol uses an adoption of CBLR protocol proposed in [63] for cluster formation. Unlike CBLR which is based on regular transmission of status messages, the frequency for sending these messages

in CBMAC depends on the state of node. In this scheme the CH takes the responsibility to assign bandwidth to the member of the cluster which reduces the packet collisions due to IEEE 802.11 and also improves QoS. Although CBMAC minimizes the hidden station problem and provides better scalability, it depends on the CH every time a new TDMA frame starts, which will lead to increased communication overhead. CBMAC also demonstrated that the probability of a node for becoming CH for a short period of time is higher than the probability when the period is long. Traffic Density also amplifies this affect as increase in traffic density increases the number of neighboring CHs that are in transmission range. However in CBMAC neighboring CHs are shown to operate together for a certain amount of time rather than immediately changing their state. This scheme also has some scaling issues due to CHs exchanging their local schema with conflicting CHs.

Almalag *et al.* [81] proposed a new TDMA Cluster-based Medium Access Control (TC-MAC) that can be used for intra-cluster communications in VANETs. This protocol integrates the centralized approach of cluster management and TDMA slot reservation scheme. The created clusters allowed vehicles to send and receive non-safety messages without any impact on the reliability of sending and receiving safety messages, even if the traffic density is high. The authors also changed the concept of having two intervals by providing vehicles with capability of listening to the control channel and the service channels during the same cycle. TC-MAC also decreases collisions and packet drops in the channel, as well as provide fairness in sharing the wireless medium and minimizing the effect of hidden terminals. TC-MAC delivers non-safety messages within reasonable time constraints, as well as meeting the requirements of minimum latency in case of safety messages.

2.2.3.4 Discussion on TDMA based Clustering Protocols

The access to the medium within a cluster in these protocols is based on TDMA which is primarily used for optimizing communication. These clustering protocols reduce intra cluster collisions as well as packet loss compared to traditional clustering protocols and thus provide fairness in sharing the wireless medium for VANETs. Table 2.6 shows that TDMA algorithms have relatively smaller delay of multi-hop safety messages as compared to other clustering schemes. Thus they provide better transmission efficiency for cluster maintenance which improves the overall throughput of both inter-cluster and intra-cluster communication. These protocols also exhibit high transmission overhead due to extra cost of channel assignment for TDMA slots. However by using some optimization techniques the performance of these protocols can be further enhanced. Calafate*et al.* [82] proposed a scheme that minimizes content delivery time by seeking optimal packet size for content delivery. Thus TDMA based clustering schemes have capable transmission characteristics,

but their behavior needs further analysis on traditional vehicle characteristics such as vehicle speed and density. Cluster stability is also low for these protocols. Although cluster connect time is comparatively reasonable, but high clustering convergence due to TDMA time slot is a serious bottleneck in implementing these protocols in VANETs.

Table 2.6: Relative comparison of TDMA based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
VeMAC [79]	L	L	M	M	H	M
CBMAC [80]	L	M	M	M	M	H
TC-MAC [81]	L	L	M	H	M	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.3.5 SDMA based Clustering

In Space Division Multiple Access (SDMA) based protocols, the road is subdivided into fixed length segments, and a segment is divided into a fixed number of blocks. Each block is assigned a timeslot representing the allowed time for a vehicle to transmit data. SDMA is known to have better performance in a dense network where practically all slots are used. But, the performance decreases proportionally with the density. Hence, in sparse networks, SDMA gives poor performance.

Salhi *et al.* [83] proposed a protocol for hybrid vehicular architecture, called Clustered Gathering Protocol (CGP). The protocol is designed to provide real-time data related to different characteristics of vehicles such as its speed, location etc. to base station by choosing the closest node at the end of the segment as CH can increase the time duration in cluster formation by repeatedly running of the election procedure. This also results in increased delay in message propagation in CGP. The use of IEEE 802.11 DCF can prevent the establishment of possible communications between vehicles in neighboring segments even after receiving a Clear to Send (CTS) packet sent by the neighboring CH. Another drawback of CGP is that it doesn't define any retransmission mechanism to deal with the reception of erroneous data.

Chang *et al.* [84] [85] proposed different dynamic cluster based vehicle to vehicle protocols using SDMA. The protocol was called Traffic Gather. This protocol inherits the main drawback of the channel allocation while using a static medium access technique in wirless networks. Thus, in the case of sparse density, many allowed slots are not utilized. Although the reliability of SDMA increases in the case of dense network, use of flooding technique may cause a broadcast storm problem even without using a mechanism of retransmission.

Table 2.7: Relative comparison of SDMA based Clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
TrafficGather [84]	H	L	L	M	M	L
CDGP [86]	M	L	L	L	M	H

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

Brik *et al.* [86] proposed a new data collection protocol for vehicular environments called Clustered Data Gathering Protocol (CDGP). The use of Dynamic Clustering technique in hybrid architecture based on DYnamic-SDMA in the data collection phase and retransmission mechanism to deal with erroneous data are some key characteristics of CDGP. It avoids collision problems by implementing a centralized, dynamic medium access technique, and enhancing the reliability by the integration of retransmission mechanism.

2.2.3.6 Discussion on SDMA based Clustering protocols

Table 2.7 shows that SDMA based clustering protocols show average clustering convergence and transmission efficiency. This is due to the data collection time and number of time slots increasing linearly at approximately constant rate for the discussed protocols. The SDMA mechanism also effects clustering overhead in terms of packet delivery ratio. The vehicle density also influences the message transmission time and results in lower cluster stability. SDMA based schemes also have larger cluster connect time due to re-clustering frequency being high. The throughput and number of transmissions also have an impact on cluster dynamics. The cluster connect time and vehicle density affects the performance of SDMA-based clustering schemes. The vehicle density will also degrade the message transmission time for these protocols. However, SDMA based clustering protocols can be used for providing V2I wireless communication for improving the ITS infrastructure. Initial investment costs could discourage the deployment of a ubiquitous roadside infrastructure to support on-the-road networks and their absence implies discontinuous coverage and short-lived connectivity [87]. The scheme proposed by Salhi *et al.* [83], has not been discussed in Table 2.7 as it does not specify any simulation results.

2.2.4 Traditional Clustering

This section discusses the Traditional Clustering techniques used in VANETs. These techniques are subdivided in to active and passive clustering based upon the role of nodes in VANET. Figure 2.2 shows the subcategory of each of these techniques in VANETs.

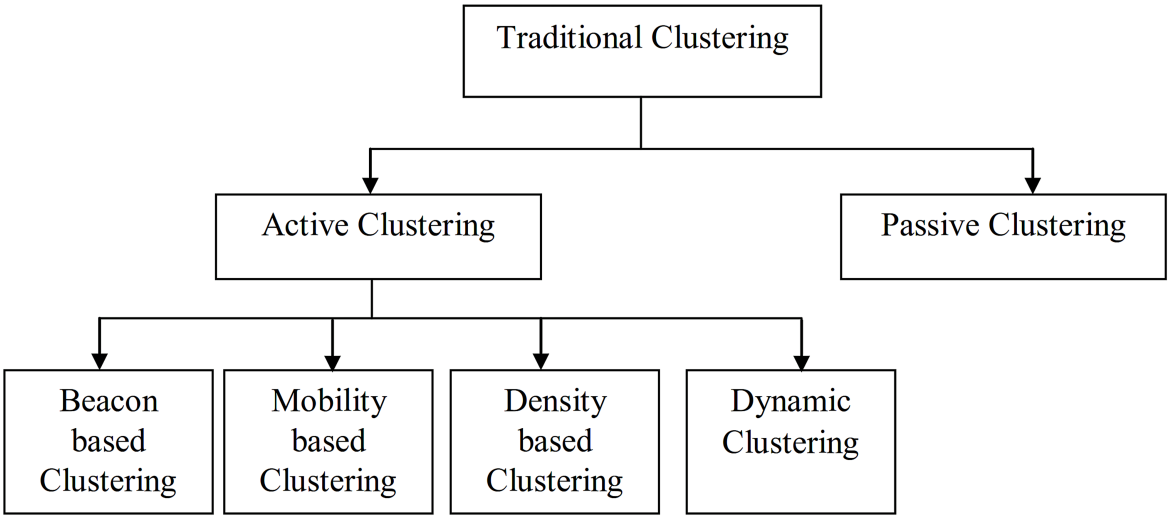


Figure 2.2: Classification of Traditional Clustering.

2.2.5 Active Clustering

In case of active clustering protocols, there are continuous updates of the clustering information and routing table for route discovery after a periodic interval of time. They generally initiate clustering process through flooding which generates the routing overhead. The various active clustering protocols are described as follows.

2.2.5.1 Beacon based Clustering

In Beacon based clustering, clusters are formed based on some vehicular or network parameter detected by beacons or hello messages by the receiving vehicle. Little *et al.* [88] proposed a beacon based clustering model, which is an extension of the algorithm proposed in [[51], [57]]. In this approach, the clusters are formed based on mobility metric and the signal power detected at the receiving vehicles on the same directed pathway. Received Signal Strength (RSS) value is used as a criteria to assign weights to the nodes and based on this weights the CH is elected. Using this method, the proposed protocol helps in forming stable clusters. However, it does not consider the occurrence of losses in the wireless channel. In practical scenario, effects of multipath fading are bound to affect the cluster formation method and thus the stability.

Teshima *et al.* [89] proposed an active clustering scheme that combines the traditional Epidemic routing with autonomous clustering Scheme proposed in [40]. They have considered a complete cluster as a single virtual node, and only the CH stores data packets. Whenever a cluster, encounters a new neighboring cluster, the CH forwards data packet to the CH of the neighboring cluster. The data packets are forwarded to the destination node based on the hierarchical tree with the CH at the root. The CH constructs the CH-based tree that contains the list of CMs to manage its cluster. The proposed scheme

is more efficient in terms of data storage since it store data packets only in the CH and all the nodes do not have to store data packets, which result in reduced consumed packet buffer.

Santos *et al.* [90] presented a reactive location based routing algorithm that uses cluster-based flooding for VANETs called Location Routing Algorithm with Cluster-Based Flooding (LORA-CBF). This clustering approach is based on regular beacon transmissions which advertise the state of the vehicle. Each vehicle can be a CH, gateway or cluster member. The CH maintains information about its member and gateways packets. Based on the state of the neighboring vehicles, a vehicle can select its own state. A CH only considers a change of its state if it receives a message from another CH. A CH receiving a hello message from another CH remains in the same state if it has more CMs on its cluster than the sender. This simple criterion however, favours larger clusters and does not take into account the mobility of the cluster members, how cohesive the smaller cluster is, or if the clusters are moving in opposite directions. Also, with large neighborhoods, clusters have the tendency to grow which in turn overload their CHs.

2.2.5.2 Discussion on Beacon based Clustering Protocols

Beacon based clustering protocols have more transmission overhead especially due to increase in number of vehicles and hop-counts. The effect of transmission efficiency in terms of volume of consumed packet buffer and end-to-end delay is high but decreases gradually with an increase in size of the network in terms of vehicle density. This indicates an efficient message delivery at low traffic but the packet delivery ratio starts degrading as hop count or number of vehicles increases. The periodic transmission of beacons helps in cluster convergence but it also affects the throughput of vehicular network, especially at higher traffic density and ultimately the transmission overhead as shown in Table 2.8. The protocols can be modified by using some quota based protocols like TTL Based Routing as described in [91] that restricts the maximum number of copies of a message in the network as well as enhances the chance of message delivery. This assists in improving cluster dynamics that has relatively lower values for all the three schemes discussed above. The impact of vehicle speed also needs to be analyzed in more detail for these protocols.

Table 2.8: Relative comparison of Beacon based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
DPP [88]	L	M	L	L	M	L
ER-AC [89]	H	L	L	M	M	L
LORA-CBF [90]	L	L	L	L	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.5.3 Mobility based Clustering

Maglaras *et al.* [92] proposed a distributed clustering algorithm which forms stable clusters based on force directed algorithms. Every node applies to its neighbors a Force (F) according to their distance and their velocities. Vehicles that move to the same direction or towards each other apply positive forces while vehicles moving away apply negative forces. According to the current state of the node and the relation of its F to neighbors F, every node takes decisions about clustering formation, cluster maintenance and role assignment. This work also proposes a mobility metric based on forces applied between nodes according to their current and their future position and their relative mobility.

Ahizoune *et al.* [93] proposed a Stability Based Clustering Algorithm (SBCA), that reduces the communication overhead that is caused by the cluster formation and maintenance, as well as to increase the lifetime of the cluster. SBCA makes use of mobility, number of neighbors, and CH duration in order to provide a more stable architecture. The nodes remain associated with a given cluster and not with any CH as is the case with most existing clustering approaches. When one CH is no longer in the cluster, another CH takes over. Thus, the cluster structure does not change but only the node playing the role of CH is re-elected. This allows for stable cluster architecture, with low overhead and better performance. SBCA has shown improved cluster residence time, for each node, reducing the overhead and thus improving the performance and reliability of a VANET.

Souza *et al.* [94] presented a beacon-based clustering algorithm for prolonging the cluster lifetime in VANETs by using a new aggregate local mobility criterion to decide upon cluster re-organisation. A node's Aggregate Local Mobility is the variance of the relative mobility over all neighbors. Two nodes moving closer together result in a negative mobility. A lower variance means less mobility of the node in relation to its neighbors. The intuition behind this scheme is that a node with less variance relative to its surroundings is a better and more stable choice for CH. The proposed clustering algorithm displays better performance in terms of stability. However since the nodes are highly dynamic in nature the position of the nodes change very fast and hence may induce an computational overhead in calculating the weight associated with the nodes.

An algorithm called, Affinity Propagation for Vehicular networks (APROVE), [95] which is a distributed mobility-based clustering scheme that forms cluster with low relative velocity between CMs also gives stable clusters by improving CH duration, CM duration, and reducing rate of CH change. The proposed clustering technique uses the fundamental idea of Affinity Propagation proposed by [96] which has been shown to produce clusters in much less time, and with much less error than previous techniques [48]. The proposed algorithm is validated by comparing it to the mobility-based ad-hoc clustering scheme, MOBIC [90] and it shows better performance.

Kayics *et al.* [97] addressed mobility by first classifying nodes into speed groups such

that nodes will only join a CH of similar velocity. Code Division Multiple Access (CDMA) scheme used to assign orthogonal codes to the previously identified vehicular nodes. Every vehicle knows the speed group to which it belongs prior to cluster formation phase. The data packet header transmitted by the sending node is modified by adding speed group information to it. If a major speed change occurs and the node moves at a speed out of its clustering group interval during a threshold value T_{speed} , it updates its speed and clustering group information and the node seeks another cluster to join. In this way, the time until a CM leaves the communication range of its CH is extended which increases the life span of the cluster. Consequently, node transition rate between clusters also decreases.

Chiti *et al.* [98] proposed a Mobility Driven Joint Clustering Scheme (MDJCS) for gathering and disseminating messages among a group of vehicles. The proposed scheme selects the CHs based upon degree of connectivity of each vehicle within an estimated coherence time interval. The maintenance of clustering scheme is provided by selecting relay nodes which are then responsible of providing connectivity and allowing CHs to stay in communication with the CMs. Relay nodes hold the cluster related information which is periodically sent by the CHs. Relay nodes belonging to different clusters provide opportunity for inter cluster communication to establish the connection and exchange the data by relaying. The proposed clustering scheme demonstrates better cluster convergence and less transmission overheads.

2.2.5.4 Discussion on Mobility based Clustering Protocols

Mobility based clustering protocols minimize relative mobility as well as distance of each CH to its cluster members and thereby attempts to improve the cluster convergence and cluster dynamics. Table 2.9 indicates that the discussed protocols also display better cluster stability that can be attributed to reduced values of clustering overhead and average number of CH changes thereby creating lesser and more stable clusters. This improved stability helps in improving the performance as well as reliability of VANETs. This also makes mobility based clustering protocols more suitable for environments like VANETs which have dynamic behavior and where mobility can be represented efficiently. However vehicle density and vehicle speed which are the predominant factors affecting mobility need to be investigated in more detail for all these protocols. It can also be concluded from Table 2.9 that the packet delivery ratio for the discussed protocols has a lower value which indicates a decreased transmission efficiency and increased transmission overhead.

Table 2.9: Relative comparison of Mobility based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
SP-CL [92]	L	H	M	L	L	H
SBCA [93]	M	M	M	L	L	H
APROVE [95]	L	H	M	M	L	M
MDJCS [98]	M	M	L	H	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.5.5 Density based Clustering

Yair *et al.* [99] proposed an iterative algorithm named as Distributed Construct Underlying Topology (D-CUT) in which each node discovers and maintains a geographically optimal clustering for the current network configuration. D-CUT algorithm partitions the network into geographically optimized clusters. The protocol is applied in two phases. In first phase, beacons in the same cluster are aggregated by a CH in a synchronized manner. In the second phase, the CH disseminates a compressed aggregated beacon of its own cluster to its adjacent clusters. The vehicles produce a snapshot of the surrounding vehicle map, and update the clustering solution according to the changes in the network configuration. All neighboring vehicles share matching partitioned vehicle maps producing the same new partitioned map for each vehicle in the network. The algorithm updates the partitioning according to the most recent topological changes thus maintaining the geographically optimized clusters.

Kuklinski *et al.* [100] proposed a multi-level cluster algorithm called the Density based Clustering (DBC) based on several factors like connectivity level, link quality, relative node position prediction of a nodes position in future and node reputation. This algorithm has three phases. In the first phase, a node estimates its connectivity level defined as number of connection which is used to discover density of local neighborhood of a node. Every node counts the number of received acknowledgments to find the number of active links. This information determines whether a node belongs to the dense or sparse parts of networks by comparing the connectivity level against a threshold value. The aim of the second phase is to select stable links from all the current links. This selection is made on some prediction about future, but it also takes into account the past knowledge of speed and direction of vehicle. This is the basis for estimating the links quality. In this evaluation a vehicle also uses signal-to-noise ratio of the link. In the last phase communication history is used to determine nodes reputation before it becomes a CM. The effects of multipath fading are also taken into account in this density based clustering algorithm.

2.2.5.6 Discussion on Density based Clustering Protocols

Density based clustering protocols allow strong connections between CMs and low variation in number of CH changes results in improved cluster stability. The data in Table 2.10 indicates that cluster stability is high for density based protocols irrespective of number of vehicles in a cluster. The density information helps in improved awareness in each vehicle about the composition of its cluster and provides strong connections between cluster members for creating a more reliable clustering topology that provides comparable cluster convergence. However further studies need to be undertaken for improving transmission efficiency and decreasing transmission overheads. Dependence only on density limits the cluster connect time and cluster dynamics. The vehicle speed also needs to be further investigated as it has a direct impact on density of the vehicles.

Table 2.10: Relative comparison of Density based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
D-CUT [99]	L	L	L	L	M	L
DBC [100]	H	H	L	L	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.5.7 Dynamic Clustering

VANETs have relatively more dynamic nature as compared to MANETs which results in fast change in the network topology. The design and implementation of an efficient and scalable algorithm for information dissemination in VANETs is a major issue that should be tackled. Indeed, in this dynamic environment, an increasing number of redundant broadcast messages will increase resource utilization, which would indirectly affect the network performance [101]. Dynamic clustering technique forms cluster structure based on node dynamics like mobility patterns, velocity and density.

Kakkasageri *et al.* [102] developed a multi agent based dynamic clustering scheme for VANETs. The scheme comprises of heavy-weight static and light-weight mobile agents and forms a moving dynamic cluster on a lane between two intersections by considering parameters such as vehicle speed, direction, connectivity degree to other vehicles and mobility pattern. Initially, cluster members are identified based on vehicle's relative speed and direction for dynamic clustering. CH is selected among the cluster members based on stability metric derived from connectivity degree, average speed and time to leave the road intersection. It consists of a set of static and mobile agents. The relative speed difference among neighboring vehicles is the main parameter used for cluster formation. The neighbour vehicles travelling in the same direction on a lane are only considered.

The cluster member with highest stability metric is considered as the CH. This scheme also has certain limitations like assumption that all vehicles need to have relatively strong computational resources, capability of authenticating and validation of vehicles during dynamic clustering process, which is not practical for current vehicular networks.

Daeinabi *et al.* [103] proposed a novel clustering algorithm, vehicular clustering based on the Vehicular Weighted Clustering Algorithm (VWCA) that takes into consideration the number of neighbors based on dynamic transmission range, the direction of vehicles, the entropy model proposed in [104], and the distrust value parameters. These parameters can increase stability and connectivity and can reduce overhead in network. VWCA works with an Adaptive Allocation of Transmission Range technique, where hello messages and density of traffic around vehicles are used to adaptively adjust the transmission range among them. VWCA uses distrust value in the weighted sum operation. The distrust value has been obtained from proposed Monitoring Malicious Vehicle algorithm. Using distrust value, vehicles that have lower distrust value than their neighbors are elected as CHs. Therefore, CHs are more trusty vehicles than other vehicles in the network. The VWCA technique mainly focuses on improving the CH duration, membership duration and security. Using VWCA, communication overheads required for joining to a new cluster in network decreases because the membership duration for each vehicle has increased. In addition, using the entropy term in the weighted sum operation, VWCA can reduce the number of overheads created by high speed vehicles. Furthermore, VWCA can increase network connectivity when electing CHs.

Kakkasageri and Manvi [105] proposed a Dynamic Clustering Scheme for VANETs (DCSV) based on degree of connectivity and vehicle mobility. The proposed scheme uses the relative speed, mobility traces and connectivity of vehicle among neighboring vehicles to form dynamic and stable clusters. The scheme is divided in two phases that are called identification of CMs and dynamic cluster formation with CH selection. During CM identification, the relative speed difference is the key factor to identify the members of the clusters. A vehicle periodically broadcast its speed to the vehicles within its transmission range. Vehicles having small value of inter-vehicle distance and within each others communication range are designated as neighbors. The vehicles travelling in opposite direction and the vehicles with higher relative speed in same direction are then eliminated and remaining vehicles are selected as the CMs. During dynamic cluster formation and CH selection, information propagation agent collects the vehicle information like vehicle ID, location and connected vehicles and transmits this to manager agent. The manager agent then uses its knowledge base to construct a cluster information table and determine the CH based upon the connectivity with other neighboring vehicles.

Louazani *et al.* [106] proposed a clustering based algorithm for connectivity maintenance in VANETs. The clustering scheme extended the conventional AODV protocol

for clustering maintenance in VANETs (AODV-CV), wherein a metric termed as average speed factor is used to differentiate the traveling directions of neighboring vehicles. The average speed factor is used to differentiate the vehicles traveling in opposite directions and to reduce periodic connection and disconnection thereby reducing overload on network on account of control messages. Initially every vehicle broadcasts hello messages and also sets its status as CM. It then updates its status as CH, if it does not receive any hello messages within some specified period. The vehicle computes average speed of its neighbors by the formula,

$$Average\ Speed = \sum_{i \epsilon N_L} speed_i/(N_L) \tag{2.1}$$

Vehicles obtain the speed of their neighboring vehicles through the received hello messages and then verify their own actual speed. A vehicle having closest speed to the average speed is declared as CH. The CH then broadcasts CH-packet containing the Cluster Head ID and one hop neighbor list. A vehicle receiving more than one CH message changes its status to gateway and other vehicles become CMs. The AODV-CV scheme depicts superior cluster maintenance and reduced clustering overhead. However the effectiveness of this scheme depends on requirements for higher computational resources for determining vehicular characteristics.

2.2.5.8 Discussion on Dynamic Clustering Protocols

Table 2.11 shows the comparison of Dynamic clustering protocols. The cluster dynamics has an average value that can be attributed to cluster lifetime decreasing by variable rates. The vehicle density also has a negative effect on cluster stability. Initially, the stability is relatively high but decrease with network load. The transmission efficiency has a comparable value with other clustering techniques which can be attributed to moderate range of values of control overhead and percentage connectivity. Since these schemes also have permissible overhead and connectivity, these provide more flexibility and adaptability and can be considered as a good add-on to existing clustering schemes. However the impact of vehicle speed needs to be investigated for realistic vehicle scenario and impact of these clustering protocols on cluster connect time also needs further analysis.

2.2.6 Passive Clustering

Passive clustering is a clustering mechanism that passively constructs a cluster structure [71] [107]. At any instant, a node in a cluster possesses an external or internal state. In passive clustering each vehicle can lower the control overhead due to packet flooding by the use of on-going data packets instead of extra explicit control packets to construct and

Table 2.11: Relative comparison of Dynamic clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
MULTI AGENT CLUSTERING [102]	L	M	L	M	L	L
VWCA [103]	M	L	L	M	L	L
DCSV [105]	L	M	M	L	M	M
AODV-CV [106]	M	L	M	M	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

maintain the clusters. When a node receives data packets, it may change its cluster state based on the state information piggybacked in on-going data packets. This reduces the number of explicit control packets. Thus, passive clustering mechanism generates significantly less overhead for cluster maintenance than the traditional cluster-based technique because the nodes do not maintain cluster information all the time.

Wang *et al.* [104] proposed a Passive Clustering Aided Routing protocol for VANETs (PassCAR) that refines the passive clustering mechanism proposed in [107] whose main goal was to construct a reliable and stable cluster structure for enhancing the routing performance in VANETs. The proposed mechanism also includes the route discovery, route establishment, and data transmission phases. The main idea behind PassCAR was to select suitable nodes to become CHs or gateways, which then forward route request packets during the route discovery phase. PassCAR assesses the suitability of nodes using a multi-metric election strategy. This strategy considers link reliability, link stability, and link sustainability as the main factors and quantifies them using the metrics of node degree, expected transmission count, and link lifetime, respectively. Each CH or gateway candidate self-evaluates its qualification for CH or gateway based on a priority derived from a weighted combination of the proposed metrics. Thus, PassCAR designs an efficient passive clustering based mechanism that operates at the logical link control sub-layer, and the proposed mechanism can easily be associated with any routing protocol to support stable, reliable, and permanent data delivery.

2.2.6.1 Discussion on Passive Clustering Protocols

The number of clusters constructed using Passive Clustering as shown in Table 2.12 remains steady and low for varying vehicular concentration that indicates medium cluster stability. This can be attributed to the consideration of node degree as a key parameter in this protocol. The achieved transmission efficiency is also comparable with other clustering protocols. However due to consideration of link quality, passive clustering has high overhead for cluster formation and maintenance especially in urban environment where

obstacles have an effect on link quality. This degrades the cluster connect time. Passive clustering also displays lower cluster convergence due to the use of information being piggybacked in ongoing data packets.

Table 2.12: Relative comparison of Passive clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
PASSCAR [104]	M	L	H	M	L	L

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.7 Hybrid Clustering

Hybrid clustering techniques combine two or more existing techniques such as use of artificial intelligence, fuzzy logic etc. Following are the schemes in this category of clustering.

2.2.7.1 Intelligence based Clustering

Hafeez *et al.* [108] proposed a distributed and dynamic CH selection criteria to organize the network into clusters. CH is elected based on stability criteria which reflect the relative movement between adjacent vehicles. The vehicle's acceleration was used to predict its speed and position in future. However the decision to accelerate, to retard or to stay on the same speed depends on many factors such as distance between the vehicle and its front neighbour, the relative speed between them, the road conditions and the driver's behaviour. Since the drivers' behaviors and how they estimate the inter distance and other factors are subjective, so triangular fuzzier was used to deal with this uncertainty using the fuzzy logic inference system. The proposed scheme can achieves stable cluster topology which makes it more suitable for implementation in VANETs. However the distributed processing overhead results in decreased message transmission effiency.

Kumar *et al.* [109] proposed an Agent Learning–based Clustering Algorithm (ALCA). Agents are able to learn from the environment in which they are operating and perform the task of CH selection. The proposed approach consists of selection of CH keeping in view the direction of mobility and density of nodes. The direction of mobility of the nodes was calculated by the agent in an interactive manner. Agents learn from the direction of motion of the vehicle and traffic flow across different junctions of the road. Agents were deployed at different road junctions for monitoring the activities of the vehicles. Agents perform their action, and accordingly, their action is rewarded or penalized in unit steps. The density of the vehicles and average speed were used for dividing the time into different zones. These zones were then used for collecting the information about

vehicles that is used as input to the agents for clustering. Learning rate is also defined for the agents to take adaptive decisions. For each action performed by the agents, the corresponding action is rewarded or penalized, and value of the learning parameter is incremented or decremented. This process continues until the maximum value is reached. The performance of the proposed scheme is evaluated by varying the number of agents with various parameters. The results obtained show that the proposed scheme can be used in the future applications in VANETs.

Wang *et al.* [110] proposed an Analytical Model for Clustering (AMC) design for VANET. In this approach the clustering process was based on modelling an unsaturated VANET cluster with a Markov chain through introduction of an idle state. The wireless channel fading and vehicle mobility are integrated by explicitly deriving the joint distribution of inter-vehicle distances. A vehicle which maintains an active connection with RSU is elected as CH. The CH then broadcasts a beacon message to vehicles within its communication range, which reply with a request to join message along with their identification information. This proposed model provides design and management of clustering the vehicles for maintaining acceptable communication performance.

2.2.7.2 Discussion on Intelligence based Clustering Protocols

Table 2.13 shows that all the intelligence based clustering protocols have good cluster stability which is due to large CH duration and cluster member duration. The values of these parameters also improve as the vehicle density increases. These protocols generate reasonably stable clusters but they also cause large transmission overhead which reduces the packet delivery ratio resulting in reduced transmission efficiency for the discussed protocols. Since hybrid techniques or heuristics are employed for cluster formation the additional overhead results in high cluster dynamics. Thus, these protocols can be a good alternative for use in future vehicular networks or for those networks that implement a specific application like security, multimedia applications etc.

Table 2.13: Relative comparison of Intelligence based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
FUZZY BASED CH ALGO [108]	M	H	L	M	L	H
ALCA [109]	H	H	L	M	L	M
AMC [110]	M	H	M	M	L	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.7.3 Cooperative De-Centralized Clustering

Cooperative vehicular systems are being investigated to design innovative ITS solutions for road traffic management and safety. Through various wireless technologies, cooperative systems can support novel decentralized strategies for ubiquitous and cost effective traffic monitoring system [111]. QuickSilver [112] is a light weight distributed clustering protocol that integrates traditional source routing protocol for intra cluster node centric communication and the construction of a multichannel link for contention free inter cluster data centric communication. It is a system architecture that provides efficient use of available resources to guarantee that no harmful competition takes place for the channel bandwidth. QuickSilver employs lightweight-clustering where clusters form and behave in an uncoordinated manner without requiring a cluster ID and there are no CHs. QuickSilver utilizes two radio interfaces that allows vehicles to maintain their intra cluster connectivity and at the same time look for inter cluster contact opportunities. Cluster formation and maintenance was done by building a cluster list of neighbor at each node.

2.2.7.4 Discussion on Cooperative De-Centralized Clustering Protocols

Table 2.14 shows that these clustering protocols have low cluster stability and average cluster connect time. This is due to the fact that average number of inter-cluster links that are active when vehicles are in contact initially increases as the overlapping region for a vehicle increases and then it shows a corresponding decrease as vehicles move away from each other. Transmission efficiency also has comparable value for these protocols. The effect of channel assignment on different node densities is represented in the form of number of links that shows an increase with the number of channels. This indicates the effectiveness of the protocols for inter-cluster communication as shown in Table 2.14. However high transmission overhead and lower vehicle density results in reduced effectiveness for intra-cluster communication in spite of decentralized clustering schemes considering realistic vehicular speed conditions.

Table 2.14: Relative comparison of Cooperative De-Centralized clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
QuickSilver [112]	L	L	H	M	H	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.7.5 Driver behaviour based Hybrid Clustering

Vehicles nowadays are provided with a variety of sensors capable of gathering information from their surroundings. In near future, these vehicles will also be capable of sharing all the harvested information, with the surrounding environment and among nearby vehicles over smart wireless links. They will also be able to connect with emergency services in case of accidents [113]. Blum *et al.* [114] proposed a Clustering for Open Inter vehicle communication Networks (COIN) algorithm. In COIN, CH election is based on vehicular dynamics and driver intentions that are the inputs for clustering instead of any conventional parameter like vehicle ID, relative mobility or some other parameters that are used in classical clustering methods. Further COIN attempts to preserve CH for a longer duration and uses mobility information for clustering.

Cheng *et al.* [115] proposed an innovative car-society clustered network based on salient classification scheme. The proposed scheme has the same interest and operate in the same communication range. The aim of the proposed approach is to increase the lifetime of the interest group, and to increase throughput in V2V environments. The proposed scheme develops an interesting ontology of Cellular Automata clustering by using Zone of Interest for mobi-cast communications in VANET environments.

2.2.7.6 Discussion on Driver behaviour based Hybrid Clustering Clustering protocols

Table 2.15 shows that Driver Intention based Hybrid clustering protocols improve the effectiveness of clustering in terms of cluster lifetime of same interest groups of users that shows good cluster stability. The transmission efficiency for these protocols is also comparable with other clustering schemes but these may generate lower values of vehicular speed and densities. This results in satisfactory transmission overhead and cluster dynamics which may however not happen if realistic vehicular conditions are considered. Clustering Convergence and Cluster Connect Time also needs further investigations. Although these protocols provide adequate stability in terms of lifetime of same interest groups, they also need to be further analyzed on the basis of several other parameters for considering their suitability in vehicular environments.

Table 2.15: Relative comparison of Driver Behavior based clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
Cellular Automata Clustering [115]	L	H	L	H	L	L
COIN [114]	L	M	L	M	L	L

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.2.8 Secure Clustering

VANETs can support applications and services for safety and comfort for the on board passengers and assist in improving the efficiency of the road transportation network. However, several serious challenges remain to be solved before efficient and secure VANET technology becomes available. One of these challenges is an efficient authentication of messages using cryptographic techniques [116]. Solutions for secure clustering in VANETs require efficient clustering algorithms in terms of complexity, scalability, availability and reach ability. Several algorithms have been proposed in the literature based on Public Key Infrastructure check for enabling communications security in vehicular environments. These are based on a trusted third party certification authorities which is responsible for certifying the public keys of vehicles. Several research schemes have been proposed for distributing the responsibility of the CA's among a set of nodes in the network, using mobility as metric to elect the vehicles that will assume the role of CA.

Blum *et al.* [117] used a PKI with virtual infrastructure where a set of elected CHs are responsible for disseminating messages after digitally signing them. This scheme is valid only for the attack called intelligent collisions. However, a PKI in VANETs must cope with different attacks. Raya *et al.* [118] proposed a distributed PKI for VANETs managed by many CA, each corresponding to a particular region. The different CAs have to be cross-certified so that vehicles from different regions can authenticate each other's CA. It require that each vehicle store the public keys of all CAs whose certificates were needed to be verified. A location-based approach to form a cluster was used where the area was divided into small zones or cells that form clusters. A vehicle automatically know to which group it belongs by comparing its GPS position to a preloaded dissection of the area map into cells. But its CH is dynamically determined as the vehicle closest to the center of the cell. The disadvantage of this proposal was the non-availability of the CA in case of a break in the connectivity.

Sivagurunathan *et al.* [119] proposed a self-organized key management system based on clustering. In this model, the network was divided into number of clusters based on the concept that any user can sign any other public key. The set of signatures formed a network of trusted relationships. However, the drawback of this self-organized approach stemmed from the assumption that trust is transitive and therefore the system became more vulnerable to the intrusion of malicious vehicles.

Gazdar *et al.* [120] proposed an efficient dynamic architecture of PKI for VANETs based on a trust model. Each elected vehicle was the CA in its cluster. The proposed clustering algorithm was based on a Trust Metric (T_m) which defined the trust level of a vehicle and was a continuous value in the interval [0-1] and mobility metric which was the relative velocity of a vehicle related to its vicinity. Vehicles started with T_m=0.1 and had to prove their good behaviour and good cooperation to increase their T_m through a

hierarchical monitoring process that supervised the behaviour of nodes at the MAC and network layers as described in [102]. A vehicle with T_m=1 was defined as a confident vehicle. Each node, with a high trust level T_m, monitored its neighbors with lower trust levels. Whenever a vehicle became a member in a given cluster, it auto generates a short term pair of keys and then requested certification from its CA. These locally generated pairs of short term keys and their correspondent certificates were assumed to stay valid as long the signing CA kept serving its cluster. The authors also used a new approach called the VANET Dynamic Demilitarized Zone (VDDZ). The role of the VDDZ was to prevent unknown vehicles from directly communicating with CA vehicles, thus shielding CAs from malicious nodes. Whenever a vehicle V sent a JOIN request to some CA in a given cluster, the VDDZ had to intercept and authenticate this request.

2.2.9 Discussion on Secure Clustering protocols

Secure clustering is an approach to increase the channel transmission efficiency and to decrease delay in authentication of a vehicle or a message. These protocols also provide higher level of security which provides data correctness and higher number of message broadcasts that increases the transmission overhead. Table 2.16 shows that secure clustering approach improves parameters like number of vehicles in a cluster, and cluster lifetime which indicates comparative performance of these protocols in terms of cluster connect time and clustering convergence. However, the impact of vehicle density and speed in cluster stability when authentication is also incorporated requires further analysis for future research work.

Table 2.16: Relative comparison of Secure clustering protocols.

Protocol	Veh. Density	Cluster Stability	Veh. Speed	Trans. Eff.	Cluster Convergence	Cluster Connect Time
ESA [118]	L	M	M	M	M	M
SAV [121]	L	M	L	M	M	M

H-High; M-Medium; L-Low ; Veh.- Vehicle; Trans.- Transmission ; Eff.- Efficiency;

2.3 Generic Security Protocols in VANETs

VANETs support different kind of services and application. These services help to provide convenience to the driver on road. But, wireless communication medium for VANET is not secure. There are security threats to onboard units and roadside units by attackers. In recent years, number of security framework and models have been proposed for VANETs by various researchers.

Plossl *et al.* [121] proposed Security Architecture for VANETs (SAV) with reference to various security needs. It included different application for VANETs such as telematics messages and warnings, alarm signals and value added services. It also outlined main security and privacy requirement such as integrity, confidentiality and availability. In the end, security architecture for VANET was proposed. The architecture included three layers named basic security level, single hop security and multi hop security. The given security architecture deployed positioning and timing, beaconing services, pseudo anonymity, public key infrastructure and intrusion detection system. It helped to prevent attacks and also allowed vehicles to communicate securely. But it did not allow any kind of revocation and detection mechanism for malicious vehicle and confidentiality was not maintained.

Zhou *et al.* [122] proposed technique for detecting Sybil attack on the basis of cryptography. The proposed technique was named as Privacy Preserving Detection of Abuses of Pseudonyms (P2DAP). The vehicles communicate via multi-hop manner and it was monitored by RSU in passive mode. The RSU installed in the VANET were further connected with Department of Motor Vehicle (DMV). The DMV provided vehicles with set of pseudonyms. These pseudonyms were used for preserving vehicle privacy. In order to avoid Sybil attack, the pseudonym were carefully hashed and stored in DMV and RSU. The RSU detects Sybil attack, if the calculated hash value of overhead pseudonym came from the same pool. RSU also sends the suspected pseudonym and hashed value to DMV to crosscheck whether pseudonym was assigned to the original vehicle or not. Also Sybil attacks were further detected based on the different pseudonyms that arrive from the same pool and identified in a small time span. The pseudonyms were used in this framework for preserving privacy of vehicles and remained trusted if RSU and DMV were secure and tamper proof.

Yan *et al.* [123] proposed a novel solution to prevent from large number of attack in VANETs that were position-related attacks and were based on forging positions. Sybil attack was based on forging identities and the combination of the both attacks. The proposed work aimed to provide secure topology information and secure network for applications. The technique to prevent position attack was built using "Seeing is believing" axiom. In proposed technique, on-board radar was used as virtual eye of the node or vehicle. Because of uncertain radar transmission range signal reception was limited on the road and a vehicle could only perceive nearby vehicles and get the information of vehicle's GPS coordinates only from the received message. A vehicle authenticated the real position of its neighbor vehicle by comparing the observation of the radar with received message and segregated malicious vehicles to attain local security. When the bogus vehicle was created by some malicious vehicle, the author used on-board radars in order to prevent discrepancies of Sybil attack. The vehicle used three type of data in order to

verify remote vehicle. The given vehicle collected radar detections, reports and oncoming traffic reports from neighbors. After collecting data, it determined similarities between data. Each set of collected data had diverse weight. The collected data was accepted if the similarities were more than the threshold value, else it was disregarded and not considered. If radar was working, more weight-age had to be given to radar collected data. When radar did not work then weight-age was given to neighbor collected data. The history of vehicle movement was created with the help of these accepted data. The collected data then helped in making decision that whether the newly received data would be valid or not. The given technique was based on isolation of malicious vehicle by the vehicle that received invalid message and the malicious vehicle's information was not communicated to other vehicles.

Petit *et al.* [124] discussed about Elliptic Curve Digital Signature Algorithm (ECDSA) algorithm for VANETs. Intelligent transport system allowed secure communication between V2V so that any catastrophe should be avoided. The existing security mechanism that were used for authenticating broadcast V2V messages had overhead in terms of computation, reliability and communication delays. The ECDSA algorithm was based on IEEE 1609.2 standard that has been proposed for vehicular ad hoc networks. This algorithm is the key protocol for providing authentication mechanism. The paper evaluated the computation cost and overhead of ECDSA algorithm. The authors also checked the influence of given algorithm used for authentication on VANET performance.

Ghosh *et al.* [125] proposed scheme for VANETs named as Misbehavior Detection Scheme (MDS) for Post-Crash Notification (PCN) application. The PCN application notified the driver about the accidental vehicle. The PCN alert was generated by vehicle close to crash. The proposed system assumed that the position information of vehicle was always true even if there was false alert. The proposed system involved calculating the deviation or difference among the driver's reporting behavior with the predictable behavior on crash. If the measured deviation or difference was greater than certain threshold then the alert was false and message would be rejected. The position of alert reporting vehicle would be sensed periodically when the alert was generated and given vehicle passed through reported crash site for the validation of given alert. The expected trajectory of the vehicle is calculated using the initial lane of the vehicle during alert generation time and lane where the accident had been take place. The difference between the actual trajectory and expected trajectory followed by the vehicle was calculated. If deviation between two trajectories was greater than certain threshold then the alert was deemed to be false. This scheme had drawback because it was based on assumption that the vehicle would always send correct position information.

Bissmeyer *et al.* [126] proposed a simTD security architecture. It involved utilization of security and privacy in field operational tests. Its main aim was to support ITS

used for Car-to-X (C2X). The simTD architecture involved three communication parties' names as ITS Central Station (ICS), ITS Roadside Station (IRS), and ITS Vehicle Station (IVS). ITS vehicle station included Communication and Control Unit (CCU) and AU. The CCU handled all communication from physical to network layer and also ran low latency application. The CCU was also responsible for all cryptographic operation while AU hosted simTD applications. The simTD messages were divided into two subtype's named as Cooperative Awareness Messages (CAMs) and Decentralized Environment Notification Messages (DENMs). The CAMs message also included location information while DENMs were related to local events and used for safety application. The simTD security architecture was based on private public key pair and included corresponding certificate from certification authority. The certificates were used to sign messages called as pseudonyms. The simTD security architecture referred to security standards of IEEE 1609.2 provided by WAVE. In simTD the sign and verify process were shifted to the network layer but in IEEE 1609.2 signing of messages were done at application layer. The simTD used RSA algorithm for cryptography where IEEE 1609.2 used ECDSA. The Security Daemon running on the CCU was responsible for message authenticity, integration and confidentiality. The security architecture also included plausibility checks for message contents so that faulty vehicle and RSU were detected. The proposed architecture provided privacy by using pseudonyms provided by CA. The main problem with the proposed system was the verification of message after signing took lot of time hence it degrade the performance of the system.

Hortelano *et al.* [127] improved the effectiveness of watchdog module for detecting intrusion in VANET environment. The watchdog worked as primary component for finding out intrusion in given system. The watchdog worked in threefold manner. Firstly, the watchdog component was defined independently so that it could work independently with any ad hoc protocol. Secondly, it possessed high detection coverage but it latency must be low. Third, it guaranteed that there would be reduction in number of false positive and negative rates. The main problem with this system was that it could not check any duplicity of given packet.

Wasef *et al.* [128] proposed complementary PKI to secure VANETs. The author discussed various infirmity in PKI and proposed number of mechanism to improve it. The author proposed location privacy using Random Encryption Periods (REPs). In this group communication, there was an encrypted communication zone where unrevoked vehicles could decrypt the communication by using group key and therefore outsider could not decrypt the message. In order to complement the lengthy revocation list supplied by trust authority, the author proposed Efficient Decentralized Revocation (EDR) protocol. The EDR revoked misbehaving vehicles by group of neighboring vehicles. It was based on secret sharing, in this a given master key was mathematically divided into number

of shadows or simple parts. These shadows were distributed to all the vehicles. Each vehicle used its own shadow or simple part of master key to compute revocation share and sent its share to the revocation coordinator. The revocation coordinator combined all the share to generate the revocation list of misbehaving vehicles. The author had also proposed Message Authentication Acceleration (MAAC) protocol. It used Hashed Message Authentication Code (HMAC) to create hash code of the message. In this protocol, the group key shared between unrevoked vehicles was used to generate hash code. When vehicle broadcast message then it append HMAC to it. In the end author proposed mitigation from DoS attack. The proposed technique involved collecting invalid signature up to some time span and then append HMAC message to all outgoing signed message. However, the author did not discuss about the generation of the group key, the location privacy and also protection against insider attacks were not addressed.

Stübing *et al.* [129] proposed model named verifying mobility data under privacy consideration in C2X communication. The proposed system used PKI for integrity and message authentication. The simTD certification authority generated certificates to sign messages. The vehicles in simTD broadcast their current position, speed and direction in CAMs. In order to avoid tracking each vehicle, the scheme spontaneously changed its identifiers called pseudonyms. The vehicles in simTD network used store and forward technique. Kalman filter-based approach was used to predict future mobility information of given vehicles. The Kalman filter approach generated optimal prediction after fixed interval based on two phases that were called prediction and correction phases. To avoid pseudonym change problem, the authors proposed a tracking algorithm that was used to assign each vehicle with a permanent identifier. The permanent identifier with vehicle ID was permanently assigned to application unit and hence privacy was reserved. The problem with given system was that it assumed tamper proof devices which were very costly.

Ruj *et al.* [130] proposed a model based on data centric intrusion detection in vehicular ad hoc network. The authors discussed various limitation of Misbehavior Detection System (MDS). The vehicles may send wrong information due to selfish reasons. The data centric misbehaviour detection system focussed on detection of false information for identifying source vehicle. The proposed system worked with two types of messages that are alert messages and beacon messages. The vehicle autonomously checks the accuracy of received alert or information by comparing it using List of Invalid Events (LIE). When the received alert was true, the MDS further verified that whether the alert was true to false. The expected movement pattern of sender vehicle decided about the correctness of the given alert. If any anomaly was found then received alert was treated as false. The periodic beacon messages were used to determine the distance between the sending and receiving nodes whenever any event has occurred. If there was any anomaly found during

distance between the sending and receiving vehicles then the sending vehicle was marked as misbehavior and the given incident would be reported to the CA. The CA imposed the fine on misbehavior and then revoked its certificate. However, the scheme performance degraded when number of invalid event that might occur became large and so it would be difficult to maintain and store LIE. Also there would be problem if sender moved out of the communication range of receiver while validating some event.

Gazdar *et al.* [131] proposed a secure and distributed PKI for VANET. It was dependent on hybrid trust model. The hybrid trust model was defined for evaluating the trust metric of each vehicle to check the behavior and legitimacy of transmit message. The given model was based on monitoring system meeting out two important aspects that are cooperation of various vehicles and also the legitimacy of the broadcast data by the vehicle. When monitoring vehicle received a warning message based on urgent events it first evaluated the cooperation rate of the source vehicle. Then it used fuzzy based system to monitor and filter out malicious vehicle. Based on the outcome of the observing process, it updated the trust metric of the sender vehicle. The trust metric value always lie in the range 0 and 1. If the trust metric value become 1 then a given vehicle was considered as trusted and malicious if its value reaches 0. The trusted vehicles were allowed to form cluster which further selected CA based on low mobility and high trust. If the vehicle was more cooperative then its trust metric value must be high. The certified authority shared private public key pairs to all the member of given cluster. The legitimacy of given message was determined by the certified authority. Only those messages were legitimate that were signed by private key of member vehicle of the cluster.

Palomar *et al.* [132] developed a model for thwarting false event propagation in vehicular network. The proposed mechanism allowed sender to send Event Warning Messages (EWM) along with some evidence called Proof-of-Work (PoW). PoW involved a non-negligible computational cost and had to be good enough so that dishonest vehicle were discouraged to flood the same message. The inter vehicle warning message was comprised of two extra fields that were named, Event Warning Certificate (EWC) and PoW evidence. The PoW calculation worked on the basis of challenge response concept. The proposed mechanism had a verifier and a prover as two parties. Before giving any access to the resources or offering any specific service, the verifier asked a simple test or puzzle to the prover for solving. The successful completion of test confirmed that the proving node could be referred to as a legal individual in the vehicular network. After successful accomplishment of the test, the verifier would allow prover to get a POW certificate. RSU could work online or offline to transmit EWM and penalizing misbehavior to improve road security.

Daeinabi *et al.* [133] proposed system that detected malicious vehicles in VANET environment. The vehicles in VANETs exchanged messages but some attackers damage

the message. So the authors consider security while transmitting the message. The proposed system was named as Detection of malicious vehicles. The network was divided into clusters where each cluster consisted of CH and one spare CH. The spare CH vehicle had trust level next to but less than cluster head vehicle. The system helped to find out malicious vehicle that drop or duplicate packet by monitoring of traffic. The monitoring was done by verifier nodes that had high trust value. If verifier detect any malicious vehicle then it increase the distrust value of that vehicle and report it to CA.

Wahab *et al.* [134] discussed about the problem of clustering in VANETs. The clustering algorithm was based on Quality of Service Optimized Link State Routing (QoS-OLSR) protocol. Several mobility based algorithms for clustering ignore the quality of service based algorithm. The QoS based algorithms were important for safety and emergency services but they ignored high speed mobility. The proposed work considered both quality of service and high mobility constraints. It helped to provide stable cluster while maintaining quality of service. The proposed framework used three algorithm called as build stable cluster using high mobility metrics. The second algorithm involved Ant Colony Optimization technique for the selection of MultiPoint Relay (MPR) by CH and third was based on MPR recovery algorithm to provide alternative routes while ensuring network connectivity during link failure. During CH selection it used truth telling and cheating prevention mechanism by encrypting QoS parameters. However, the proposed mechanism does not give details about encryption mechanism of the packets.

Alexiou *et al.* [135] proposed Vehicular Public Key Infrastructure (VPKI) architecture. The proposed architecture supported Authentication, Authorization and Accountability (AAA) without exposing vehicle identity. In this system, the vehicle used private key to sign packets and used corresponding certificate to transmit the packet. The vehicle used secure channel to achieve confidentiality. The architecture involved three trusted CAs named as Long-Term Certification Authority (LTCA), Pseudonym Certification Authority (PCA) and Resolution Authority (RA), The LTCA was responsible for sharing long term certificate and tickets to the vehicles. The PCA provided each vehicle with short lived pseudonyms so that it would prevent them from tracking their location. The RA violated the pseudonym certificate if any adversaries was allowed access. The tickets were used to guarantee uplink ability between the vehicle so as to request for pseudonym. The certification revocation list were prepared by RA in order to revoke certificates and tickets and make it publicly available so that vehicles could use them. The given system was based on ECC-256 keys for both infrastructure and vehicles certificates. The ticket size was 498 bytes. The system did not suggest time period for change of pseudonyms and how to save ticket if some adversary take place.

Daeinabi *et al.* [136] proposed scheme that offered security in VANETs environment. It provided security because VANET communication was vulnerable to different kind of

attacks. The attacker exploited other vehicles in VANET by sending bogus message. The authors proposed advanced Secure scheme based on Clustering and Key Distribution (SCKD) while considering members and CH in VANET. The system was based on symmetric cryptography, hashed based authentication code and proxy signature. It helped to improve authentication, non-repudiation, confidentiality and integrity. But the computation time for proxy signature validation, key generation and distribution might affect the VANET system throughput.

Mondal *et al.* [137] proposed system for VANET in which there was detection and revocation of misbehaving vehicles. It was based on hierarchical or tree based detection system in which vehicles were under the coverage of base station and these base stations were within the coverage of CA. The vehicles detected misbehaving vehicles and created certificate revocation list and further sent it to parent or within coverage base station. The base station again built certificate revocation list after getting lists from attached vehicles and sent this list to CA. Then final certificate revocation list was built by certifying authority after receiving all the revocation lists from base stations within coverage area. The CA then broadcast this list to all vehicles. The main problem with this work was that there would be requirement of number of base station and CA to maintain this system and also computation cost for this system was very high.

Sedjelmaci *et al.* [138] proposed an Accurate and Lightweight Intrusion Detection system framework for VANETs. The main aim of this framework was to protect network against the malicious vehicle attack. It supported VANET's characteristics such as high speed vehicle and rapid topology change. The given framework used secure clustering algorithm. The clustering algorithm was built considering vehicles mobility and network vulnerability. The CH was selected based on nodes mobility and trust level of various vehicle. In order to reduce broadcast storm and communication overhead, cluster selection procedure was implemented by the author. The given intrusion detection system had three agents that were called Local Intrusion Detection System (LIDS) used by CMs. Global Intrusion Detection System (GIDS) run on CH, Global Decision System (GDS) run at road side unit. The LIDS used rule based policy in order to find intrusion detection with high accuracy where as GIDS was governed by rules based detection and anomaly detection based on Support Vector Machine. The anomaly detection phase depended on learning algorithm to handle new attack and model normal behavior. The GDS had ability to aggregate the reputation of each vehicle and generated new trust level list for each vehicle. Based on trust level, the RSU filtered the list of malicious vehicle. It broadcasted only fraction of blacklist and suspected list that had probability to pass RSU's communication range based on mobility prediction to avoid high overhead.

Jesudoss *et al.* [139] proposed a Payment and Punishment based Scheme (PPS). The proposed model encouraged nodes to participate in cluster formation and also monitor the

exchange of information between nodes or clusters. In this, vehicles support truth telling and the CH was selected on basis of highest number of resources such as connectivity, low speed etc. The vehicles were analyzed on the basis of their truth telling by offering incentive so that they could not provide any wrong information. Vehicles got highest reputation based on their honesty in data election, forwarding and monitoring process. The model also consisted of Extended Dempster-Shafer model so that malicious vehicles might be isolated from network through monitoring by watchdogs. However, the proposed scheme did not include any kind of security related information such as encryption and decryption of packet.

2.4 Comparative Analysis and Discussion

This section provides a detailed comparison and analysis of various techniques used for clustering in VANETs. Table 2.17 provides a comparative analysis of various protocols discussed above by considering number of parameters used for clustering in VANETs. These parameters cover most of the key areas that need to be considered for designing a clustering protocol. The detailed analysis of all the main clustering protocols with respect to various parameters is listed in Table 2.17. The parameters selected for discussion of clustering are as follows:

- Key clustering parameter
- Type of topology handled
- Additional infrastructure requirements
- Road scenario handled
- Types of messages handled
- Relative direction of vehicles in a cluster
- Density of vehicles
- Velocity of vehicles,
- Communication mode employed
- additional overhead incurred

VANETs face many research challenges in multiple areas, from privacy and anonymity to the detection and eviction of misbehaving nodes and many others in between. Multiple solutions have been proposed to address those issues [140].

With respect to the basic parameter used for clustering, we have identified distance between vehicles, direction of vehicle, final destination, vehicle speed, mobility pattern, and duration of a vehicle within a segment, position within a segment, medium contention and number of hops for message transmission as some of the other parameters. Most of the existing clustering protocols use a combination of one parameter that depicts vehicle behaviour and another parameter for optimizing message delivery to generate clusters. Topology indicates the structure of vehicles after the completion of cluster formation. The topology of a cluster can be divided into hierarchical or flat structure. Due to their inherent structure and domain requirements, a large majority of reviewed clustering protocols have tree like hierarchical topology.

The vehicles in a network may require some additional infrastructure like a GPS, transceivers, Lane Detection System, Digital maps, RSU's, odometer etc. for cluster formation and maintenance. This requirement results in some extra hardware modification to be performed on the network which may result in more efficient clustering but may not be possible to implement depending upon the local environment. Thus only a few protocols have shown any need of additional infrastructure and most of the protocols rely on available communication network infrastructure for clustering. The road scenario handled by a protocol is also an important issue and it is used in simulations for estimating parameters like density and speed of vehicles. The clustering protocols reviewed have been classified to be simulated under either highway or urban road traffic conditions and very few clustering protocols have considered both the road scenarios. Mobility is another aspect that indicates whether the vehicles mobility during cluster formation has been considered or not. Various proposals have been considered based upon different mobility models for vehicles.

Type of data handled by a cluster is an issue that is used to identify the application area for a clustering protocol. Since safety message broadcast with very small latency has been considered as one the main objective of VANET, most of the proposals have attempted to optimize them. Some recent protocols have considered other messages like non real time message transmissions also. Relative direction of a vehicle specifies the direction of vehicles that participate in cluster formation. Since the movement of vehicles in a cluster could be constrained by road condition, so the direction of vehicles can be in the same direction called uni-directional movement or in opposite direction known as bidirectional traffic. Majority of the reviewed clustering protocols have considered uni-directional vehicular movement for cluster formation since it increases the stability of clusters. This is due to the reason that vehicles in a cluster can be in contact for a larger time interval in case of unidirectional clustering with each other as compared to relatively small contact time for bi-directional clustering.

Density is another important aspect which is used to estimate the size of the cluster in VANETs. Most of the exiting protocols have considered this aspect where density varies from low, medium to high values. However, there are no standard values for these levels of densities and different protocols have considered different ranges varying from a minimum value of 2 vehicles/km/lane to a high value of approximately 60 vehicles/km/lane. The issue of vehicle speed is also important and it defines the speed range of vehicles in a cluster. Different clustering protocols have used various values of speeds and it is an important parameter for verifying the performance of the protocols.

Communication mode is also an important issue in which vehicles use the available communication infrastructure. The communication mode between vehicles can be broadcast or its variant like unicast or multicast, store and carry or any other. Most of the existing protocols use broadcasting as the communication mode because of it being supported in most of the underlying channel access protocols. The use the wireless media for transferring the data between a CH and CMs also results in some communication cost during clustering. However, due to the lack of communication reliability in VANETs, it is necessary to implement fault-tolerant techniques during the discovery of service providers in VANETs [[141][142][143][144][145]].

Communication overhead is used to estimate the coverage and connectivity of the nodes which is used for cluster maintenance by the CH. This is an important factor to be investigated because some of the CMs may be located at the corners of the cluster and have poor connectivity with the CHs so these nodes may not be able to communicate with their respective CHs. The communication overhead has been characterized into three levels identified as low, medium and high [[146][147]]. Since vehicular networks do not have any problem regarding battery life unlike MANETs, the communication overhead in most of the protocols has been found to be medium. In most of the clustering protocols, metrics such as delay, cluster lifetime, cluster head duration, number of packets transferred in unit time, throughput etc. are considered to compute the effectiveness of the any proposed scheme. Table 2.17 provides a detailed categorization of all the existing proposals based upon the above mentioned issues.

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
CBLR (2003)	-	Hier.	GPS Scenario	Circular	-	-	-	-	B	M
COIN (2003)	Vehicular Dynamics	-	-	-	-	-	-	-	-	L
LORA CBF(2005)	Location	-	-	-	-	-	-	-	B	-
Inf.Prop. (2005)	Signal Strength	-	-	-	-	-	-	-	-	-
ESA (2006)	Secure Message Aggregation, group comm.	Hier.	GPS	U	Safety	Bi	2-5	15	B	H
DISCA (2007)	Direction, Leadership	Hier.	Hgh, U	-	-	-	-	-	B	-
CBMMAC (2007)	Multi Channel	Hier. Hier.	Trn.	Hgh. Hgh	Real,Non	- Real time	12,24,40	20-50	B	L
CBMAC (2007)	TDMA	-	U	-	-	- -	2,4,7	-	B B	H H
DBA-MAC (2007)	Distributed, Dynamic	Hier.	GPS	Hgh	Safety	Bi	25,50,7	20-30	B	-

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
VeSOMAC (2007)	TDMA	-	-	Hgh.	-	-	-	-	-	-
Inter-Veh Comm(2007)	Distance	-	-	U	-	Bi	-	-	B	H
MDMAC (2008)	Mobility	K-Clusters	-	U	Safety	Bi	-	11-31	B	M
TrafficGather (2008)	SDMA	-	-	-	-	-	-	-	-	-
Pos. based (2009)	Position	Hier.	-	-	Safety	-	-	-	-	M
APROVE (2009)	Affinity Propogation	-	GPS	Hgh	-	Uni	-	15,25,35 40,50	B	H
CGP (2009)	SDMA	-	-	-	-	-	-	-	B	-
DBC (2009)	Density	Hier.	-	-	-	-	-	-	B	M
SCA (2009)	Distr. PKI based on trust mgmt .	-	-	Hgh	Safety, Non-safety	Bi.	50	10-40	Multi.	H

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
ALM (2010)	Local Mobility	-	-	U	-	Bi.	Y 25,30	10,15,20	-	M
EDC (2010)	Euclidean Distance	Hier.	GPS	Hgh Y	Safety	Uni	10-60	15,25	B	M
NEW ALM (2010)	veh. position within segment	-	-	-	-	-	-	-	-	M
LBC (2010)	Direction of traffic flow	- -	Odometer, LIDAR	U	-	Uni	60	11-33	B	L
ER-AC (2011)	Autonomous Clustering	Hier.	-	U	-	-	-	10-15	Store and Forward	H
MC-DRIVE (2011)	Direction of traffic flow	-	GPS, Digital Maps	-	-	Uni	Low, H	8.5-14	B	M
DCA (2011)	Spatial Dependency	Hier.	GPS	-	-	Uni.	-	10-20	B	-

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
Rel. Mob. (2011)	Aggregate Mobility	Hier.	-	Hgh, U	-	-	-	10-35	B	H
D-CUT (2011)	location, density	Hier.	GPS	-	-	-	-	-	-	-
LICA (2011)	Location Improvement	-	GPS	-	-	-	-	-	B	-
VWCA (2011)	Direction algorithm	Hier.	-	Hgh.	-	-	-	20-35	B	L
AMACAD (2011)	Final Dest.	-	GPS	U	-	-	0 to 5 per 100 m^2	5-20	Store and forward	M
Fuzzy based (2012)	Ave. Speed Difference	-	GPS	Hgh	Safety	Uni	-	20-30	-	-
VeMAC (2012)	TDMA	-	GPS, Trn.	-	-	Bi	-	-	-	M
CCA (2012)	Rel. Velocity	- -	- -	U -	- -	Bi -	Constant -	20-35 -	B -	M
SBCA (2012)	Mobility Mobility	-	-	Hgh.	-	Uni	Low-High	25-35	-	L

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
SP-CL (2012)	Relative Force	-	-	Hgh.	-	-	-	20-45	-	L
K-hop (2012)	Max. Connectivity	Hier.	GPS	-	-	-	-	-	B	L
Quick-Silver (2012)	Sequence number of messages	Flat	-	U	-	-	-	-	Unic., Multi	-
CDGP (2012)	SDMA	-	GPS, Dig. Maps	Hgh.	-	Uni.	-	15- 50	B	L
TC-MAC (2012)	TDMA TDMA	-	-	Hgh.	Safety, Non-safety	-	5,12, 21,50	-	Multi	M
ALCA (2012)	traffic flow, mobility	-	-	-	-	-	-	15-25 15-25	-	H H
PassCAR (2013)	Passive clustering	Hier.	-	U, Hgh.	-	-	10-20	5-20	B	-
HCA (2013)	Random, k-hop	Hier.	-	U	Safety, Warning	-	Low-High	15-30	-	L
SBCA (2012)	Mobility	-	-	Hgh.	-	Uni	Low-High	25-35	-	L

Protocol (Year)	Parameter	Topology	Additional Infr.	RoadSide Scenario	Data Handled	Relative Dir. of Vehicle	Vehicle Density (V/km)	Rel. Vel. of Vehicle m/s	Comm. Mode	Cluster Overhead
CERA (2014)	Vehicle Position	Hier.	-	Hgh.	-	Uni	Low-High	-	-	M
APA (2014)	Affinity Propagation	Hier.	-	U, Hgh.	Safety Non-Safety	Bi	Low-High	-	B	L
MDJCS (2014)	Vehicle Mobility	Flat	-	U, Hgh.	-	Uni	Low-High	-	B	M
DCSV (2014)	Degree of Connectivity	Hier.	-	Hgh.	-	-	High	-	-	L
AODV-CV (2014)	Average Speed	Hier.	-	U	Safety, Non-Safety		High	-	B	L
AMC (2015)	Channel Fading & Mobility	Hier.	RSUs	-	Safety Non-Safety	Uni	High	-	B	L

Table 2.17: Relative comparison of existing clustering protocols in VANETs

Abbreviations Used:-*Hier-Hierarchial; U-Urban ; Hgh-Highway; Uni-Unidirectional; Bi-Bidirectional; B-Broadcast; L-Low; M-Medium; H-High; Multi-Multicast; Unic- Unicast;*

2.5 Conclusion

Vehicular Ad Hoc Networks are being used in wide areas of applications in recent times. Clustering of vehicles has been investigated by the research community from different perspective in many of the applications used in VANETs. But, it has been a challenging task to perform clustering due to the dynamic nature of nodes in VANETs. This chapter provides a complete taxonomy with challenges, constraints and solutions on clustering in VANETs based upon various parameters. Also, a detailed discussion with comparative analysis is provided for each categorization of clustering which includes various challenges, existing solutions and future directions. Each section is described with various clustering techniques and their advantages/disadvantages over the others. The analysis provided for various existing proposals also provides parameters to select one of the proposals with respect to its merits over the others.

Chapter 3

Predictive Clustering

VANETs have emerged as a new powerful technology with an aim to provide safety and comfort to the passengers during their mobility on the road. As vehicles are constrained with respect to the available computation and storage resources, so a lot of overheads are incurred for managing these networks. Also, due to high speed and varying density of vehicles on the road, it is a challenging task to maintain Quality of Service with respect to various metrics such as throughput, energy emission and End-to-End Delay (E2ED). Hence, an intelligent approach is required to optimize the various operations in this environment that should provide minimum overheads. To address these issues, this chapter describes a predictive clustering approach for vehicles in VANETs. Efficient algorithms for future mobility predictions and average variations of vehicles on the road are proposed. The algorithms estimate the clustering duration and determine the number of vehicles in a cluster. The performance of the designed algorithms is studied using extensive simulations by varying the number of vehicles and cluster durations in comparison to benchmarked scheme in the literature. [1]

Vehicles may communicate with each other either through V2V or V2I communication to generate safety alerts and other messages to the passengers. Most of the ITS applications require an up-to-date and real time knowledge of the vehicle position [150]. From the past few decades, there has been growing interest of the research community in developing algorithms that provide efficient and convenient driving conditions through VANETs. Due to vehicles' high mobility, VANETs suffer from the rapid network topology changes, frequent link ruptures and variable vehicle density, which results in variable

[1] The content of this chapter has been taken from :

- Rasmeet S. Bali, Neeraj Kumar, Joel JPC Rodrigues, "An Efficient Energy-Aware Predictive Clustering Approach for Vehicular Ad hoc Networks", International Journal of Communication Systems, John Wiley & Sons, Volume 10, pages 1099-1131, 2015.
- Rasmeet S. Bali, Neeraj Kumar, "Learning Automata-assisted Predictive Clustering Approach for Vehicular Cyber-Physical System", Computers & Electrical Engineering, Elsevier, doi 2015.

network connectivity. So, to keep the connectivity at high rate among the vehicles, these are grouped together based upon some predefined criteria to create clusters.

Clustering in VANETs is of crucial importance for addressing the scalability problems. The performance of communication protocols is largely influenced by the existence of vehicles in the neighborhood [[109],[151]]. Clustering is used to limit channel contention, provide fair channel access, increase the network capacity by the spatial reuse of the network resources and effectively control the network topology. But, the main challenge in clustering is the overhead introduced to elect the CHs and to maintain the membership in a highly dynamic and fast changing environment such as VANETs.

Usually, vehicles can communicate with the neighboring vehicles within the range of 300 to 500 meters. Moreover, unlike traditional MANETs, VANETs do not have any battery power constraints, as vehicles can recharge their batteries while running. In VANET, a vehicle consists of components like front and rear sensors, OBU, Global Positioning System and Omni directional antennas. Sensors are used to sense traffic or road condition, while OBU consist of necessary software that interpret input received from the sensor or antennas and display it in the form of messages on the screen. Antennas are used to transmit and receive messages to and from other vehicles. Vehicles also communicate with infrastructure units along the road such as toll collection booths, traffic lights, petrol stations etc. and RSU's. The RSU's are geographically fixed units and interconnected by a backbone network and can be used for providing always-on internet access in vehicles as shown in Figure 3.1. Thus the V2I communication utilizes the RSU due to their additional capability of data storage and processing as compared to an OBU. As soon as any vehicle moves into communication range of any RSU it will request for access and then utilize the services of that RSU. Overlapping ranges between adjacent RSU's result in handoff mechanisms to avoid load imbalances in the network. Since the proposed clustering scheme primarily relies on V2V communication, issues arising due to V2I communication are not considered in this work.

In recent years, there has been a lot of attention devoted to research in the area of clustering for VANETs [[94],[142]]. The focus in this research area has moved towards increasing the efficiency for cluster formation so that available resources could be utilized for user satisfaction level in traffic management and safety applications [[152],[153],[154]]. The developed clustering algorithm should also be distributed, with no central coordinator, to be used in clustering. The algorithm should also handle topology changes with minimized local impact on the cluster topology and should also be able to detect and react to topology changes. Because of the high degree of mobility in VANETs, a clustering algorithm should converge fast and should have a reduced overhead to minimize the time lost in the clustering process. In past [155], a number of parameters has been considered for performing clustering in VANETs such as relative mobility, position, lane of the vehi-

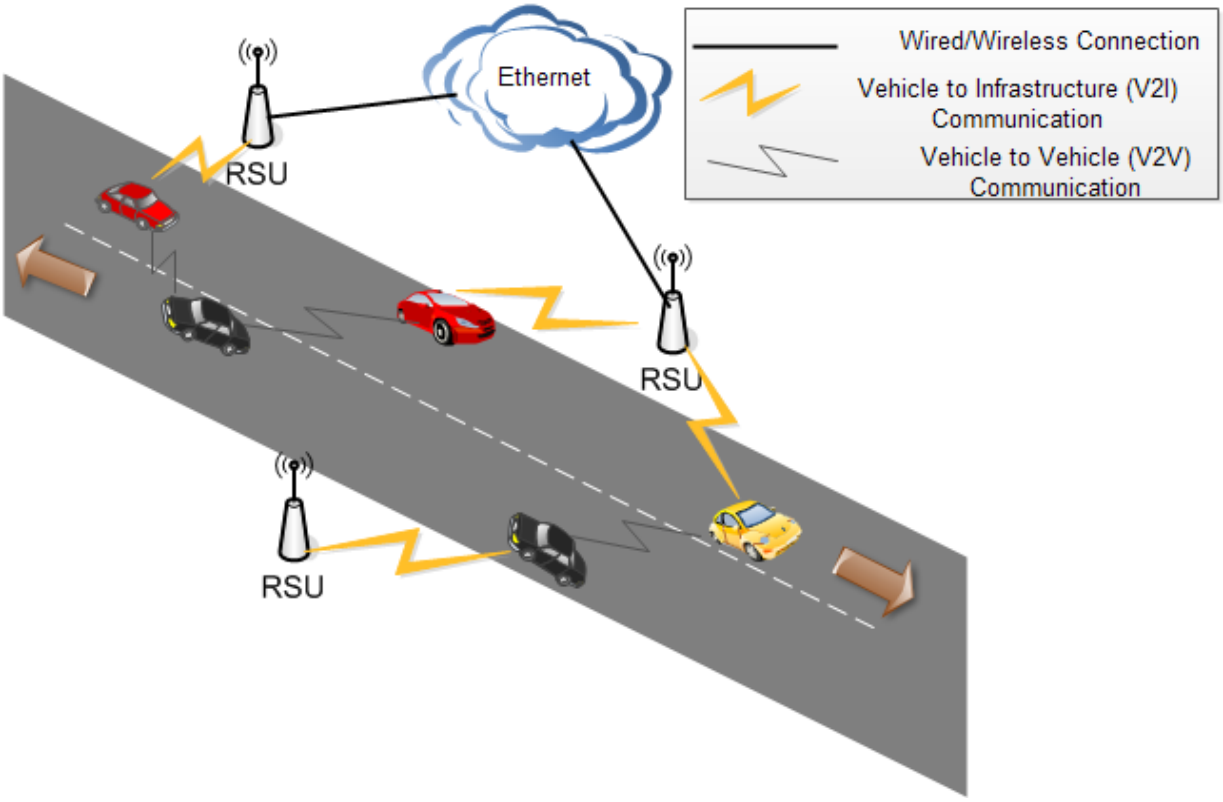


Figure 3.1: Basic vehicular network model for cluster formation.

cle, and distance. Also, some additional network based characteristics are also considered such as Space Division Medium Access, Time Division Multiple Access and Wavelength Division Multiple Access (WDMA) to increase the reliability of message transmission in VANETs. However, clustering based on prediction based technique has not been explored to its full potential in VANETs. Predictive clustering schemes are based on predicting the future position of vehicles by considering its present and future position and then this predicted future position is used to generate clusters among the participating vehicles. The clusters are generated by selecting some vehicles as CHs based on their relative Prediction accuracy. The vehicles with minimum value of mobility are selected as CHs and remaining nodes are made CMs. For effective data dissemination, vehicles that are in contact with multiple clusters are designated as Gateway Nodes.

3.1 Predictive Clustering Approach

This section introduces the problem of predictive clustering and network model used, along with the assumptions used in proposed predictive clustering scheme.

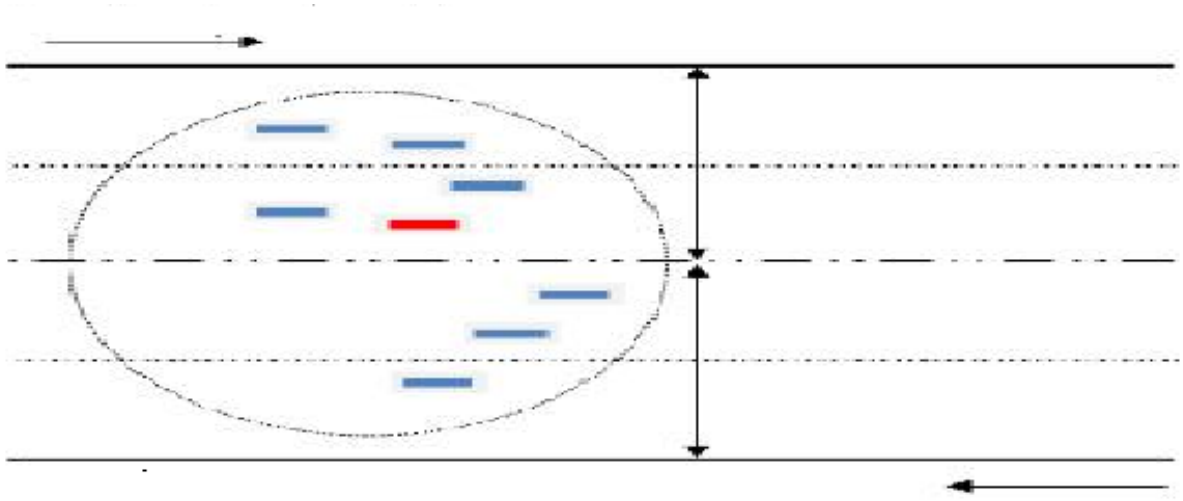


Figure 3.2: Vehicular road scenario at a time instant T

3.1.1 Network Model

In developing the prediction based clustering scheme, it is assumed that all the vehicles are equipped with OBU devices. The vehicles are also equipped with speed and distance measuring devices such as GPS and Digital Odometer. These devices are commonly available on vehicles for recognition of movement and travelled distance. Two consecutive vehicles in the network are assumed to be connected, if the inter-vehicle distance between them is smaller than transmission range of those vehicles. The path or section of road on which vehicles move is also assumed to be straight with only orthogonal intersections. Each section of road is assumed to be of width W and is separated into bi-directional carriage ways for movement of traffic. Each carriageway is assumed to be further partitioned into two or three lanes as shown in Figure 3.2 and 3.3. The road is divided into lanes and vehicles moving on adjacent lanes in same direction are within communication range of neighboring vehicles for sufficiently long interval of time due to their low relative velocity. The identification of movement direction for vehicles is attained automatically as vehicles moving in one direction will be out of communication range of vehicles that are going in opposite direction after a very short period of time as they have comparatively high relative velocity. The proposed predictive approach considers this time interval by selecting short time prediction interval (t) to be sufficiently large so that vehicles on opposite lanes are not considered for clustering

3.1.2 Movement Direction Estimation

Each vehicle broadcasts a periodic beacon or *hello* message containing the information such as vehicle identification, present coordinates, the road direction (e.g., n for R_r or

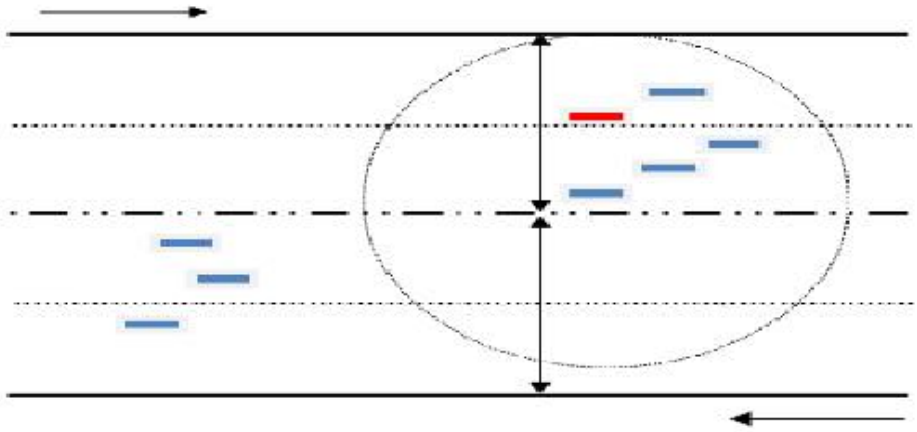


Figure 3.3: Vehicular road scenario at time instant $T + t$

s for R_l)and the time stamp. In Figure 3.4, a vehicle R located at some position on road receives beacon messages from vehicles A and B with a time stamp t and estimates its distance to A and B, i.e., $d_{A,t}$ and $d_{B,t}$, respectively, by using the ranging technique. On the basis of the vehicle position information and the distance estimates, two possible positions of both vehicles, i.e., $D_{A,t}$ $(x_{A,t}$, $y_{A,t})$ and $D_{B,t}$ $(x_{B,t}$, $y_{B,t})$, are obtained corresponding to the intersection points of two vehicles. The equations of the distance between two are expressed as,

$$(x - x_A)^2 + (y - y_A)^2 = d_{A,t}^2 \tag{3.1}$$

$$(x - x_B)^2 + (y - y_B)^2 = d_{B,t}^2 \tag{3.2}$$

However, given just one set of beacon messages, the receiving vehicle cannot be sure whether the vehicles are traveling in same or opposite direction. Thus, R waits until it receives the next set of broadcast messages from A and B and then re-estimates its distances $d_{A,t+1}$ and $d_{B,t+1}$, respectively. On the basis of this new information, R obtains two possible position estimates, i.e., $D_{A,t+1}$ and $D_{B,t+1}$, corresponding to the two consecutive messages received by a vehicle, which are used to estimate the movement direction of these vehicles

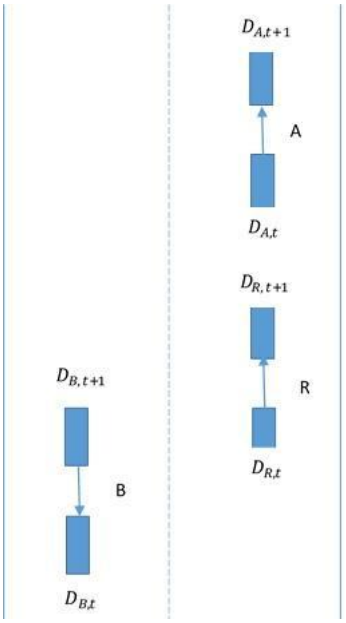


Figure 3.4: Estimation of movement direction of vehicles.

$$(x - x_A)^2 + (y - y_A)^2 = d^2_{A,t+1} \tag{3.3}$$

$$(x - x_B)^2 + (y - y_B)^2 = d^2_{B,t+1} \tag{3.4}$$

The vehicles then re-compute their position with respect to the reference vehicle. This calculation of the distance between vehicles is used as the basis for computing the vehicle characteristics such as speed, and acceleration which is then used in future for clustering and message dissemination.

3.1.3 Beacon Messages Scheduling

Due to large number of beacon messages being transmitted by vehicles and to implement the predictive mobility in an efficient manner, a two stage message scheduling has been used. These stages use two different time-intervals for scheduling beacon messages that are used for short term and actual prediction purposes. While time is divided into *'t'* discrete time-intervals for short term prediction, the actual future position estimation is done by considering *'T'* as a time-intervals, where $t \leq T$.

Since *'T'* is a time duration after which the CHs need to implement the prediction process, so there are *'k'* intervals of *'t'* after which this process is repeated. Thus, following

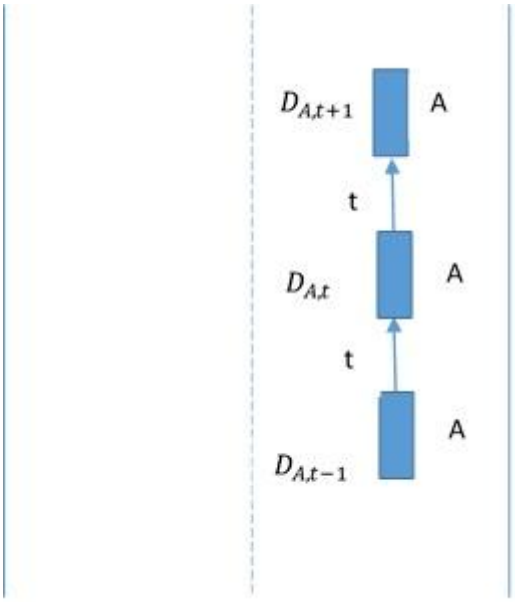


Figure 3.5: Prediction of future position of vehicles.

holds,

$$T = k * t \tag{3.5}$$

3.1.4 Future Position Estimation

To estimate the future position of vehicles in the proposed scheme, periodic beacon messages transmissions are utilized. The time period for these messages is either *'t'* or *'T'* depending on whether the estimation is for short term or an actual position prediction. In Figure 3.5, a vehicle A located at position $D_{A,t}$ is moving with speed $v_{A,t}$. The predicted position of A at time *'t+1'*, given by $D_{A,t+1}$ can be computed as follows,

$$D_{A,t+1} = v_{A,t-1} * t + (a_{A,t} * t^2)/2 \tag{3.6}$$

where $v_{A,t-1}$ is the speed of the vehicle at time *'t-1'* and $a_{A,t}$ is its acceleration. The acceleration can be defined in terms of current and past velocity as,

$$a_{A,t} = (v_{A,t} - v_{A,t-1})/t \tag{3.7}$$

Substituting this value of $V_{A,t-1}$, the calculation of $D_{A,t+1}$ can be further simplified as,

$$D_{A,t+1} = (v_{A,t-1} + v_{A,t}/2) * t \tag{3.8}$$

Table 3.1: Key symbols used in predictive clustering protocols and their meanings.

Symbols	Meaning
W	Width of road section
MN	Member Node
CH	Cluster Head
B	Beginner node
N	Novice Node
P	Predictive Node
L	Learned Node
G(V,E)	Unit Disk Graph
$D_{A,t}$	Position of Vehicle A at time t
W_J	Weight of Vehicle-Group
$x_{A,t}$	x-coordinate of Vehicle A at time t
$y_{B,t}$	y-coordinate of Vehicle A at time t
$predpos_T$	Predicted position of vehicle at time T
$actualpos_T$	Actual position of vehicle at time T
v_T	Speed of vehicle at time T
p_i	Probability of vehicle i
$\in$	Average Prediction Variation Coefficient
v_{mean}	Mean velocity of all Vehicles
k	Number of time intervals in vehicle prediction
d_T	Predicted Final Position
σ	Cluster Formation Parameter
λ_T	Future Mobility Factor

The key objective of the proposed approach is to achieve accurate prediction about future position of vehicles that will assist in effective cluster formation in VANETs. The objective function for the clustering scheme is to minimize the deviation between the predicted position and actual position of all vehicles in the network thereby reducing communication overhead as well as improving the overall efficiency of message dissemination in VANETs,

$$\min\left(pred_{pos} - actual_{pos}\right) \tag{3.9}$$

The symbols used in the algorithm and their meaning are described in Table 3.1.

3.2 Proposed Predictive Clustering Approach

Considering the fact that clustering based on mobility prediction of the future positions of vehicles has not been explored to its full potential, in this paper, we propose a scalable

clustering protocol for VANETs based on the future position of a vehicle. Our approach is motivated by the fact that due to constrained mobility patterns of the nodes in VANETs, prediction can be performed with relatively high degree of accuracy. This prior cluster formation will eventually help in providing increased availability of resources as well as efficient utilization of network bandwidth for supporting higher QoS. To achieve the balance between the two conflicting and essential requirements of clustering, i.e., up-to-date information and minimal overhead in clustering process, estimation of the future position of a vehicle can be considered as a viable clustering scheme. Thus, a scalable predictive scheme is proposed that considers the historical behavior of a vehicle to create clusters. The constrained and regular movement of vehicles on a road is one of the key parameters that has been considered in our proposed scheme. The scheme is described in details as follows:

3.2.1 Prediction Expertise Computation

In this scheme, each vehicle initially forms a unit disk graph *G (V,E)*, where the set V represents the set of vertices and set E represents the set of edges of the Graph. The graph is formed by considering the current position of all its neighboring vehicles as vertice's. The proposed scheme uses *G* at discrete time interval for predictive cluster formation. At any time instant, each vertex can be in one of the four states *Beginner(B), Novice(N), Predictive(P)* and *Learned (L)*. All vertices are initially assigned the *Beginner* state. These states indicate the prediction expertise of the vehicle within the network. A node in *B* state has not yet initiated its prediction process, so it cannot be a considered as a reliable source. *N* state indicates that the particular node has achieved the elementary expertise for generating the future node positions based on previous values but its prediction accuracy is average. The *Novice* and *Active* nodes generally exist in the CM state. A node in *P* state generates reasonable future positions, but has not yet assumed the role of CH and does not take part in CH election process. Nodes in *L* state are those vehicles that form a virtual backbone for the created cluster. They can also initiate and are used to decide the CH for the predicted cluster.

Figure 3.6 shows the transition diagram of nodes into various states. Initially, a node starts in *B* state and does not participate in CH election process and only gathers information to be used for predicting its future position. After few time cycles, it will start predicting its position in *B* and moves to *N*. A node in *N* state moves to *B* state, if it is unable to predict its position due to some constraints. The node remains in *N* state until its relative position in *G* and its prediction accuracy crosses the threshold value. At this stage, it moves into P state, and subsequently into *L* State when it becomes a CH. Generally, a *L* node is elected as the CH and/or Gateway nodes. The number of nodes that can act as *L* nodes depends on vehicular density. These nodes are also assigned

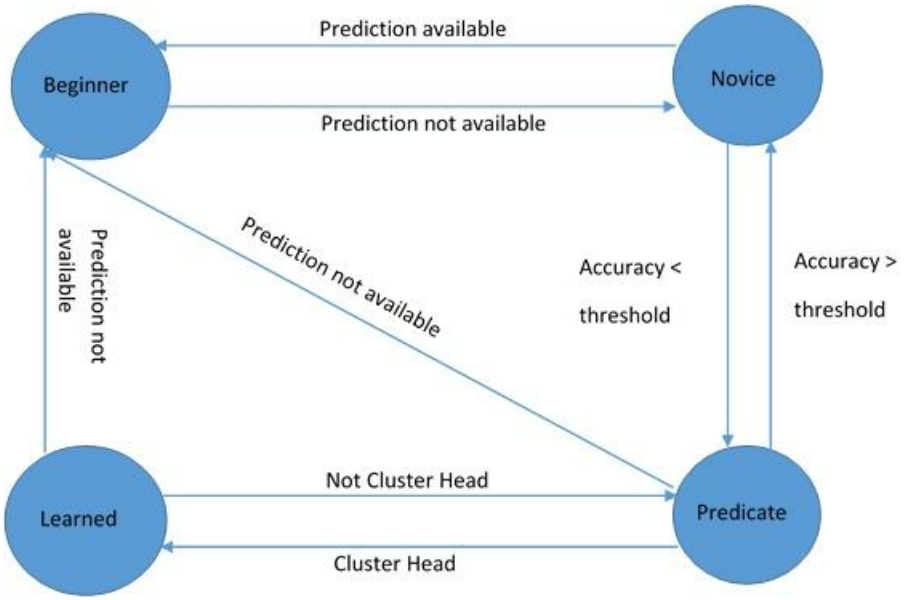


Figure 3.6: State diagram depicting prediction expertise of nodes.

weights that are used to generate the aggregate predicted position of each vehicle. The *Learned* Nodes are assigned maximum weights where as the *Beginner* nodes are given minimum value. Initially, the vehicle is designated as B, and all of them start with equal probability to participate in CH election such that following holds,

$$\sum_{i=1}^{\infty} p_i = 1 \tag{3.10}$$

where p_i denotes the probability of vehicle *'i'* and *'n'* is the total number of vehicles in the network. Initially,

$$p_i = 1/n, \forall i = 1, 2, ...n \tag{3.11}$$

If any vehicle enters the network, it is also initially assigned the probability p_i

3.2.2 Average Prediction Variation

After a small and discrete time interval of *'t'*, each vehicle predicts its future position for the next *'(n-1)'* time instants. This is done by considering the previous and current speed of vehicles as follows,

$$predpos_{(k+1)} = actualpos_{(k)} + t * (v_{(k-1)} + (v_k - v_{(k-1)})/2), \forall k = 1, 2, ...n - 1 \tag{3.12}$$

where $predpos_{(k+1)}$ is the predicted position at instant *(k+1)*, $v_{(k-1)}$ and v_k are the speeds of that vehicle at time *(k-1)* and k respectively. At time instant *k+1* , the predicted position of each vehicle($predpos_{(k+1)}$) is now compared with actual position of that vehicle ($actualpos_{(k+1)}$) and the difference of these values is averaged with previously computed differences over a number of intervals to generate *average prediction variation*($\in_{k+1}$). Thus, we have the following:

$$\in_{(k+1)}= (\sum_{i=1}^{k} \in_i +(predpos_{(k+1)} - actualpos_{(k+1)}))/2 \tag{3.13}$$

The procedure to compute $\in$ is described in Algorithm 1

Algorithm 1 Average Prediction Variation

Inputs: Previous and Present position of Vehicle,
Actual position of all vehicles during previous prediction, (pos_{actual}),
Speeds of Vehicle during Previous time periods.
Outputs: Average Prediction Variation ($\in_i$)
Assumption: Time is divided into k discrete intervals with all vehicles having synchronized clocks

for $(vehicle = 1; vehicle \leq M)$ **do**
  $\in_{i,0}= 0$;
  $flag_i = 0$;

  $actualpos_{(t)}$= *Previous known position of vehicle i* ;
  v_0= *previously computed speed of vehicle i*;
  v_1= *current speed of vehicle i*;
  for $(t = 1; t \leq k)$ **do**
    $actualpos_{(t)}$= *last calculated position of i*;
    $v_{t-1} = v_t$;
    Find current speed of vehicle i v_t ;
    $predpos_t = actualpos_t + t * (v_{t-1} + (v_t - v_{t-1})/2)$;
    if $(t \geq 2)$ **then**
      $flag_i = 1$;
    end if
    if $(flag_i \neq 0)$ **then**
      $\in_t= (\sum_{i=1}^{t-1} \in_i +(predpos_t - actualpos_t))/2$;
    end if
  end for
end for
END

3.2.3 Prediction of Future Vehicle Position

The generation of future position of vehicles is based on prediction of relative position prediction of each vehicle which is represented in terms of the four categories as discussed above and its $\in$ value. The proposed model derives probabilistic prediction of a vehicle's mobility by using the variation from the computed value of its historical predicted mobility pattern. The repeated computation of $\in$ for relatively smaller time-intervals helps to improve the accuracy of prediction by generating a more realistic value of variation coefficient to be used for predicting the future position for each vehicle. This value can then be used in predicting the future position of vehicles for the time instant at which predictive re-clustering is being done. Thus, each vehicle computes Future Mobility Factor (λ) after T interval of time as follows,

$$\lambda_K = W_J * \in_{k+1} \tag{3.14}$$

where W_J is the weight of the node that depends on the category to which the vehicle currently belongs. The values of W_J are 0.1, 0.2, 0.3 and 0.4 for nodes belonging to B, N, P or L categories respectively. Initially, all the nodes are designated as B nodes and thus the weight has a negligible impact on the value of λ. However, as the prediction accuracy of a vehicle improves, it also has a corresponding impact on values of λ. Thus, each node computes its λ coefficient and broadcasts this value to the current CH and the other L nodes in the cluster. The CHs also receive average velocity of neighboring vehicles over the same time period in the periodic *hello* messages. The CHs then calculate the aggregate velocity of all the vehicles in the cluster as follows,

$$v_{mean} = 1/M \sum_{j=1}^{M} avg\_v_j \tag{3.15}$$

where avg_v_j is the average velocity of j^{th} vehicle over $n*t$ small time intervals and v_{mean} is the average velocity of all M vehicles in the neighborhood of a CH. Now, the CHs compute the general predicted position range of the vehicles at a time instant *'(K+1)'* by considering the previous actual position and the computed values of v_{mean}, and the difference in previous prediction as follows,

$$predpos_{(K+1)} = actualpos_{(K)} + (T * v_{mean} + T * (v_k - v_{(k-1)}))/2 \tag{3.16}$$

The CH now utilizes the computed value of $predpos_{(K+1)}$ and λ to generate the precise position of each vehicle. The predicted position of vehicle thus lies in the range that

depends on $predpos_{(K+1)}$ as well as λ_K and is defined as the range between the following,

$$predpos_{(K+1)} - \lambda_K \; to \; predpos_{(K+1)} + \lambda_K \tag{3.17}$$

The final predicted position of vehicles is now determined by considering the accumulated past prediction accuracy in terms of λ and the difference of actual and predicted position of that vehicle during the previous iteration as follows,

$$Finalpredpos_{K+1} = predpos_{(K+1)} + \lambda_K * (predpos_{(K)} - actualpos_{(K)}) \tag{3.18}$$

The CHs use the predicted values of vehicles to generate the graph G. The generated graph is used for CH election and cluster formation at time instant *(K+1)*. The procedure to compute the values of λ and G are described in Algorithm 2.

Algorithm 2 Future Position Prediction

Inputs: *Average Prediction Variation Coefficient of each vehicle i* $(\in_i)$,
Average velocity of each vehicle i $(v_{i,agg})$,
Actual previous position of each vehicle i $(actualpos_T)$,
Predicted previous Position of each vehicle i $(predpos_T)$,
State Information of each vehicle i
Outputs: $(final - pred - posi)$,
Unit Disk Graph V(G,E) with each Node Signifying the Predicted position of each vehicle
Assumption: *After K=t*k discrete time intervals, with all vehicles having synchronized clocks*

for $(vehicle = 1; vehicle \leq M)$ **do**
  $\lambda_i = W_i * \in_i$;
  Broadcast Beacon message containg λ_i *to current CH*;
end for
CH Waits till all messages are not received ;
$v_{mean} = 1/M \sum_{i=1}^{M} avg\_v_i$;
for $(vehicle = 1; vehicle \leq M)$ **do**
  $predpos_{T+1} = actualpos_T + (T * v_{mean} + T * (v_T - v_{T-1}))/2$;
  $Finalpredpos_{K+1} = predpos_{(K+1)} + \lambda_K * (predpos_{(K)} - actualpos_{(K)})$;
end for
Construct G(V,E) using the Final Predicted Position of each Vehicle;
END

3.2.4 CH Election Process

After determining the position of each vehicle, the decision about the next CH to be selected is made depending on whether the current CH is available or not. In case the

cluster does not contain a CH, the member nodes elect the CH in a distributed manner by constructing a dominating set from G. However, instead of using traditional graph based metrics, the proposed scheme uses a cluster formation parameter σ for creating the dominating set. Initially, the node with greatest σ values are elected as CHs and their neighboring nodes then act as Dominate nodes. However, if CH exists, then current CHs will decide the next CH after receiving σ values from all the MN as follows:

$$\sigma_i = \lambda_K * W_i * (Predpos_{(i)} - Actualpos_{(i)}), \forall i \tag{3.19}$$

The CH also periodically broadcasts *hello* messages that contain its cluster information as well as its future position. When vehicles that are not part of any cluster receive these messages, it sends a JOIN message to the respective CH. The CH then includes this vehicle as a MN and sends a JOIN_CLUSTER message. The CH and MN also periodically broadcast *hello* messages. While the messages from the MN are only disseminated within the cluster, messages from CHs are also allowed to be transmitted to neighboring CHs through gateway nodes. The CHs also use the predicted position to determine the approximate time till which a particular MN will remain in its cluster. If after this time, the CH does not receive *'j'* consecutive messages from that MN, it is assumed that the vehicle has left the cluster.

Once G is available, the weights of all the nodes are also modified and then divided into four categories using a threshold based parameter. The nodes selected as dominating nodes in G are considered as L nodes. The nodes that are in single hop neighbor of L nodes are defined as P nodes. The remaining nodes are designated as N Nodes. A node that has not joined any cluster becomes a B. The procedure to compute σ is described in Algorithm 3.

3.2.5 Complexity Analysis of Predictive Clustering Approach

To estimate the overall complexity of proposed algorithms, both computation and communication overheads are computed in this section. While computational complexity determines the functional complexity of the proposed algorithms, communication overhead between vehicles also needs to be considered. In Algorithm 1, the inner *for loop* from *steps 9 to 19* executes for *'k'* number of times. The *outer for loop* executes once for every vehicle taking part in clustering *'n'*. Thus, the computation complexity of Prediction Variation algorithm is *O(nk)*. Since no beacon messages are exchanged, the algorithm has no communication overhead.

In algorithm 2, each vehicle predicts its future position on the road once and assuming *'n'* nodes, the complexity of the algorithm is *O(n)*. Since all the vehicles except the existing CHs broadcast beacon messages, the communication overhead is of order of n.

Algorithm 3 CH and MN Selection

Inputs: *Unit Disk Graph V(G,E) with each Node Signifying the Predicted position of each vehicle,*
λ_K *for each Vehicle,*
Current Weight of each Vehicle,
Predicted and Actual Position of each Vehicle.
Outputs: *Information about CH and MN vehicles in the network,*
Modified Weights of each Vehicle.
Assumption: *Each vehicle broadcasts periodic BEACON messages every t time units with all vehicles having synchronized clocks.*

for $(v = 1; v \leq M)$ **do**
  $\sigma_v = \lambda_K * W_v * (Predpos_v - Actualpos_v)$;
  if $(v == CH || v == MN)$ **then**
    Broadcast a message containing σ_v;

  else
    Broadcast JOIN Message;
    if (JOIN-CLUSTER message received) **then**
      Broadcast σ_v *message to CH* ;

      $Weight_v = 0.1$;
      $Status_v = Beginner$;
    end if
  end if
end for
Sort THE σ_v's in ascending order
Select CH and MN nodes using σ_v *as a parameter through Weekly Connected Dominating Set Algorithm*;
for $(v = 1; v \leq M)$ **do**
  if *(v is a CH)* **then**
    $Weight_v = 0.4$;
    $Status_v = Learned$;
  end if
  if *(v is a Dominatee)* **then**
    $Weight_v = 0.3$;
    $Status_v = Predictive$;
  else
    $Weight_v = 0.2$;
    $Status_v = Novice$;
  end if
end for
END

Algorithm 3, performs CH election process using Dominating Set procedure. Since the order of these algorithms is *O(nlogn)*, our proposed scheme also has the same computational complexity, i.e.,*O(nlogn)*. Since periodic beacon messages *'m'* are to be broadcast by vehicles, the communication overhead for algorithm 3 is of the order of *mn*.

Thus, the total computational complexity and communication algorithm for the proposed scheme is given as,

$$Computation\ Complexity= O(kn)+O(n)+O(nlogn) \tag{3.20}$$

Since k is very small compared to n, we can consider k to be a constant. Thus, the computational Complexity of proposed scheme can be approximated as,

$$\begin{aligned} &Overall\ Computation\ Complexity= O(nlogn) \\ &Overall\ Communication\ Overhead= n(1+m) \end{aligned} \tag{3.21}$$

3.3 Predictive Clustering Approach using Learning Automata

Vehicular systems are one of the most popular wireless systems of the modern era due to their abilities to disseminate the information to the other nodes well on time. To increase the reliability of communication between different nodes, clustering has been widely used in these systems. But due to high velocity and constant topological changes, clustering is also one of the most difficult tasks to be performed in this environment. In this section, we propose a novel Learning Automata (LA) based hybrid clustering scheme for vehicles. We have improved our existing solution described in previous sections, in which future mobility prediction was not taken in to account to its full potential. In the current solution, LA stationed on the vehicles are used to estimate future positions of the vehicles more accurately. A Predictive Clustering Algorithm using Learning Automata (PCALA) is proposed in the current solution. The actions of the LA are rewarded or penalized based upon their current prediction accuracy and their previous actions. Through extensive simulations, the enhanced approach has been found to be more effective as compared to previous approach.

3.3.1 Learning Automata for Predictive Clustering

A Learning automaton is an executable code that interacts with a random environment to perform its actions. Figure 3.10 describes the basic structure of an automaton.

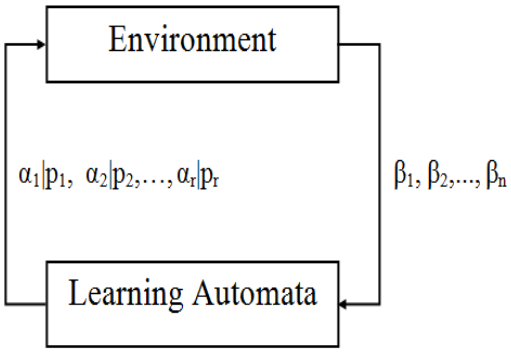


Figure 3.7: Basic operations of a Learning Automaton

The output of the automaton called the *action* is input to the environment and the output of the environment denoted as the reaction is input to the LA. The automaton is defined by (A, Q, R, T) and the environment by (A, R, D), where,

- A defines the set of all actions and

$$A = \{\alpha_1, \alpha_2, ..., \alpha_r\} \tag{3.22}$$

  where r is the number of possible actions available to the LA.
  At time t the action is denoted by $A(t)$.

- $R = \{\beta_1, \beta_2, ..., \beta_m\}$ describes the set of reactions from the environment, where m is the number of possible outputs for the environment. Here $\beta(t)$ indicates the reaction to the automaton at instant t where,

$$\beta(t)\ \epsilon\ \mathbb{R}\ \forall\ t \tag{3.23}$$

  for a given LA $\alpha_i,\ \beta_j\ \epsilon\ \mathbb{R}, for\ i = 1\ to\ r,\ j = 1\ to\ m.$ $\mathbb{R}$ is the set of real numbers.

- $D = d_1, d_2, ..., d_r$ is the set of average reward values, where,

$$d_t(k) = \mathrm{E}\ [\ \beta(k)\ |\ \alpha(k) = A(t)] \tag{3.24}$$

  The nature of d_t's provide an indication about whether the environment is stationary or not.

- Q is the state of the automaton defined by,

$$Q(k) = [p(k), d(k)] \tag{3.25}$$

where, p(k) is also called the probability vector and,

$$p(k) = [\ p_1, p_2,\ \ldots\ ,\ p_r] \tag{3.26}$$

such that

$$\sum_{k=1}^{r} p_t(k) = 1, \quad \forall\ k \tag{3.27}$$

Also,

$$\hat{d}(k) = [\hat{d}_1, \hat{d}_2, ..., \hat{d}_r] \tag{3.28}$$

is the vector of estimates of the average reward values at some time instant.

- T is the reinforcement scheme which is used to update the current state of the LA

$$Q(k+1) = \mathrm{T}\ [\ Q(k),\ \alpha(k),\ \beta(k)] \tag{3.29}$$

The LA operates by selecting an action at instant t. Let $\alpha(t)$ be the selected action from the set of actions A. Generally this action is selected based on a probability updating algorithm or a reinforcement scheme. This action is then provided as input to stochastic environment which responds with a reaction $\beta(t)$. Next, the automation computes $\alpha(t+1)$ by using the probability updating scheme. This cycle is then repeated for subsequent iterations. The key objective of the LA based scheme is to identify the action that provides optimum reward and for proposed clustering schemes. The objective function is stated as follows,

$$\Delta_t = min\ \{d_t|\ \forall i\ \epsilon\ \mathrm{N}\} \tag{3.30}$$

where, Δ_t is the reward value that provides optimal actions and d_t is the selected value of reward at time t and N is the total number of occurrences for the distributed LA.

3.3.2 System Model for LA based Predictive Clustering

In the proposed clustering scheme, the future predicted position which is a measure of prediction accuracy for vehicles is improved by using LA. Every periodic prediction is controlled and monitored by the LA that takes the decision regarding the final predicted position of vehicles, based on current traffic conditions and relative positions of vehicles within the cluster. Initially, each vehicle computes its future position using the approach proposed in [156]. This predicted position along with the current prediction accuracy is then used by the LA on board the vehicles to further improve it's prediction accuracy.

We use the L_{RI} based LA in the proposed scheme. In this scheme, after selection of an action a_i at time instant t, the reinforcement scheme is then applied. The updating rules for the L_{RI} scheme are defined as follows,

$$p_i(t+1) = 1 - \sum_{i \neq 1} \lambda_r p_i(t) \qquad if \;\; \alpha_i \;\; is \;\; chosen \;\; and \;\; \beta = 0 \tag{3.31}$$

$$p_i(t+1) = \lambda_r p_i(t) \qquad if \;\; \alpha_i \;\; is \;\; chosen \;\; and \;\; \beta = 0 \tag{3.32}$$

$$p_i(t+1) = p_i(t) \qquad if \;\; \alpha_i, \; \alpha_j \;\; is \;\; chosen \;\; and \;\; \beta = 1 \tag{3.33}$$

where $p_i(t)$ is the probability of choosing action a_i at time t and satisfies $(0 < p_i(t) < 1)$. The sum of $p_i(t)$ is unity. In the above equations, λ_r is the parameter of the scheme and it satisfies $(0 < \lambda_r < 1)$. Typically, λ_r is close to unity. The LA also considers the density and current relative speed by taking into account the acceleration and decelaration of a vehicle with respect to it's neighboring vehicles to take the decision regarding its future position. The following steps describe the working of PCALA approach.

Step 1: Initialize the value of decision variables for representing the prediction accuracy and the future predicted position of the vehicles.

$$\Delta_t = (d_1, d_2, ..., d_m) \tag{3.34}$$

The value of Δ is dependent on vehicle density and the relative speed of the vehicle.

Step 2: Compute the current optimum value of prediction accuracy from actual and predicted positions of the vehicle. This value is denoted by $thresh(t)$ and its initial value denoted by $P_{Pred,0}$ equals the prediction accuracy i.e.,

$$thresh(0) = P_{Pred,0} \tag{3.35}$$

Step 3: Using the above initialized values, the probability vector at t^{th} time interval is represented by a $m \times N$ matrix where each value is represented as $P_{i,j}(t)$. Initially, we have the following,

$$p_{i,j}(0) = \frac{1}{N} \; for \; i = 1, 2, ..., m, \;\; j = 1, 2, ..., N \tag{3.36}$$

where, $p_{i,j}$ represents the probability that action d_i is selected by the j^{th} vehicle at time t.

Step 4: Generate the set of possible actions, using the value of respective probability

distribution and from this set select some actions randomly based on the value of $P_{m,d(t)}$.

Step 5: Use the value of $thresh(t)$ to determine the next state and response of the automation using the prediction based probability distribution accuracy vector.

Step 6: Update the probability distribution vector by increasing the probability if the new prediction values result in more accurate predictions. However, if the new predictions are inferior, then the probability is not modified since the LA is based on L_{RI} model i.e.,following holds,

$$P_{i,j}(t+1) = \begin{cases} P_{i,j}(t) & \text{if } \beta(t) = 0 \\ P_{i,j}(t) + \Delta_t & \text{if } \beta(t) = 1 \end{cases} \tag{3.37}$$

Step 7: Modify the value of $thresh(t)$ and then take a decision regarding whether the terminating condition for the algorithm is satisfied or not. These values are computed using the following equation,

$$thresh(t+1) = \begin{cases} P_{i,j}(t+1) & \text{if } \beta(t) = 0 \\ thresh(t) & \text{if } \beta(t) = 1 \end{cases} \tag{3.38}$$

The above equations compute the new probabilities and if the probability is more that the threshold value then the algorithm is terminated. Otherwise above steps from 2-7 are repeated until the maximum number of iteration is reached.

Theorem 1 The predicted future positions of the vehicle and its actual value is bounded by an upper bound,

Proof: Consider the process of vehicle prediction by using the Markov decision process for vehicular movements. Thus we have the following,

$$d_{t+1} = f(d_t, a_t, u_t) \quad for \quad t = 0, 1, 2, ...N-1 \tag{3.39}$$

where, d_t is a random variable based on reward values that ranges over a set of states D. D is assumed to be countably infinite set, where d_t indicates the state at stage t, a_t is the action taken from finite set of actions A at stage i and u_t is a random adjustment value at i^{th} stage ranging between [0, 1] which is uniformly selected.

The value of u_t is dependent on the reinforcement scheme Q and it represents the uncertainty in the system and it is independent of the other vehicular parameters. So we have the following,

$$Q \rightarrow [0, 1]$$

Thus, the next state function f can now be stated as follows,

$$f : D \times A \times [0,1] \rightarrow D \tag{3.40}$$

Let P denote the set of all possible non stationary Markovian policies for determining the predicted positions. So we have,

$$p_{t+1}|p_t : D \rightarrow A \qquad t = 0, 1, ...N-1 \tag{3.41}$$

We need to determine the value of prediction decision function (Δ) for the current position d at time instant t. Therefore, we can say that the following condition holds,

$$\Delta_t^*(d) = sup_{p\epsilon P} E\left[\sum_{j=t}^{N-1} F(d_i, p_i(d_i), u_i)\Big| d_{i-1} = d\right] \tag{3.42}$$

where,

$$d \ \epsilon \ D \ \ and \ \ u = (u_t, u_{t+1}, ...u_{N-1}) \ \text{ such that } \ u_j \backsim U(0,1)$$

$$j = i, ...N-1$$

F is a bounded by non negative prediction accuracy function and is defined as follows,

$$F : D \times A \times [0,1] \rightarrow R^+ \tag{3.43}$$

$$\text{also} \qquad \Delta_N(d) = 0 \ \ \forall \ \ d\epsilon D \tag{3.44}$$

$$d_t = f(d_{t-1}, p_{t-1}(s_{t-1}), u_{t-1}) \tag{3.45}$$

represents the time variant value of random variable that defines its state at time instant t.

Initially, we compute the value of $\Delta_0^*(d_0)$ and then obtain an optimal value of $p, (p^*)$ that achieve the best prediction.

$$F_{max} = sup_{d\epsilon D, a\epsilon\Delta} \ F_T(d, a, u) \tag{3.46}$$

Since $F_{max} < \infty$ and since we have assumed that the all action are possible for vehicles at any time interval so, we get the following,

$$V_i^*(d) = max_{a\epsilon A}\Delta_t^*(d)(d, a) \tag{3.47}$$

Now, Δ_t^* can be written in recursive form as follows,

$$\Delta_t^*(d,a) = E_w[R(d,a,u) + V_{t+1}^*(f(d,a,u))] \tag{3.48}$$

$$\text{where} \quad V_t^* = max_{a\epsilon A}\Delta_t^*(d,a)$$

The automoton samples the actions with respect to the current probability distribution and the value of V^* at every iteration.
If a given action $a(t)$ is sampled then its count variable $N_{a(t)}(d_0)$ is also incremented. So, the environment response can be obtained as follows,

$$R(d_t, a(t), u_t) + V_t^*(f(d_t, a(t), u_t)) \tag{3.49}$$

Thus, the environment response at any stage (t) is given as follows,

$$\Delta_t^*(d_t,a) = \frac{1}{N_{a(t)}^0(d_0)} \sum_{j=0}^{t} R(d_t, a(t), u_t) + \hat{V}_t^*(f(d_0, a(0), u_t)) \tag{3.50}$$

The above process is repeated until we do not obtain feasible values of prediction. Since each computation of Δ_i helps in improving the overall prediction accuracy, so we conclude that the proposed scheme converges to an optimal value after finite number of iterations. Let K be the maximum number of stages after which the scheme converges to an optimal value. Therefore, we say that the complexity of the proposed scheme is bounded by the number of stages K.

(Hence, Proved)

3.4 Proposed LA Based Predictive Clustering Approach

The proposed PCALA scheme considers the predicted positions of a vehicle and takes into account the restricted mobility patterns of vehicles. By forming a cluster based on a vehicle's future positions on the road, PCALA results in efficient usage of network resources. Following sections describe the proposed scheme in detail.

3.4.1 Future Vehicular Position Estimation

In PCALA, every vehicle first estimates its future position by considering its present position and the average velocity. Unlike most of existing schemes where periodic beacon messages are required to be transmitted, PCALA utilizes the computing capability of

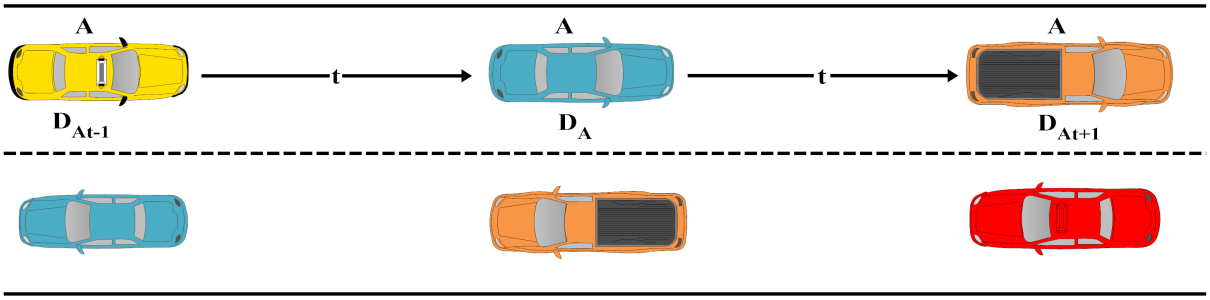


Figure 3.8: Prediction of future position of vehicles.

each vehicle to generate its own future position. Since a vehicle only computes its own predicted position, message overheads are reduced. The prediction is performed after fixed periodic intervals denoted by t. The predicted position can be determined based on computation procedure used in [156].

As shown in Figure 3.11, a vehicle is at its current position of $D_{A,t}$ and is assumed to be traveling with current speed $v_{A,t}$. The predicted position of A at a future time instant $t+1$ is represented as $D_{A,t+1}$ and can be determined based on computation procedure used in [156] as follows,

$$D_{A,t+1} = (v_{A,t-1} + v_{A,t}/2) * t \tag{3.51}$$

After computing its future position, an on board distributed automaton further optimizes the prediction value. The vehicular density and average relative speed of the vehicles are used by automaton for further improving the accuracy of predicted values. The main objective of the automaton is to minimize the difference between the predicted position and it's obtained actual position, i.e.,

$$F_{obj} = \min\left(pred_{pos} - actual_{pos}\right) \tag{3.52}$$

Each vehicle initializes and stores the values of its prior actual positions, average velocity during previous iterations and its current velocity. Then, the vehicle predicts its position for next time interval *(t+1)* using its previous two predictions. The proposed model derives probabilistic prediction of a vehicle's mobility by using the variation from the computed value of its historical predicted mobility pattern. The vehicles compute general predicted position range of the vehicles at a time instant *t+1* by considering the previous actual position and the computed values of v_{mean}, and the difference in previous prediction as follows,

$$predpos_{(t+1)} = actualpos_{(t)} + (t * v_{mean} + T * (v_t - v_{(t-1)}))/2 \tag{3.53}$$

Algorithm 4 Future Position Prediction

Inputs: *Previous predicted and actual position of vehicles, Speed of vehicles .*
Outputs: *Predicted Position of vehicles*
Assumption: *Vehicles are synchronized by common clock and equipped with GPS, N is the total number of vehicles, t is current time interval*

Begin
for $(i = 1; i \leq N; i++)$ **do**
  $flag_i = 0$;
  $actualpos_0$= *Initial known position of vehicle* i ;
  v_0= *Initial speed of vehicle* i;
end for
for $(t = 1; t \leq K; t++)$ **do**
  for $(i = 1; i \leq N; i++)$ **do**
    $actualpos_t$= *last calculated position of vehicle* i;
    Find current speed of vehicle i (v_t) ;
    $pred\_pos_t = actual\_pos_t + t * (v_{t-1} + (v_t - v_{t-1})/2)$;
    if $(t \geq 2)$ **then**
      $flag_i = 1$;
    end if
    if $(flag_i \neq 0)$ **then**

$$avg\_v_i = v_t + \frac{1}{t-1}\sum_{j=1}^{t-1} v_j$$

$$v_{mean} = 1/N \sum_{i=1}^{N} avg\_v_i;$$

      $pred\_pos_{t+1} = actual\_pos_t + (t * v_{mean} + t * (v_t - v_{t-1}))/2$;
      Generate the predicted position of the vehicle on road using GPS
    end if
  end for
  Predict future position by invoking PCALA // Algorithm 2
  for $(i = 1; i \leq N; i++)$ **do**
    Broadcast the computed future position
  end for
end for
End

The final predicted position of vehicles is determined by considering the accumulated past prediction accuracy and the difference of actual and predicted positions of that vehicle during the previous iteration. The predicted values of vehicles are then used to generate its position which becomes the input to LA. Algorithm 3.4 describes the procedure to compute the predicted estimated position.

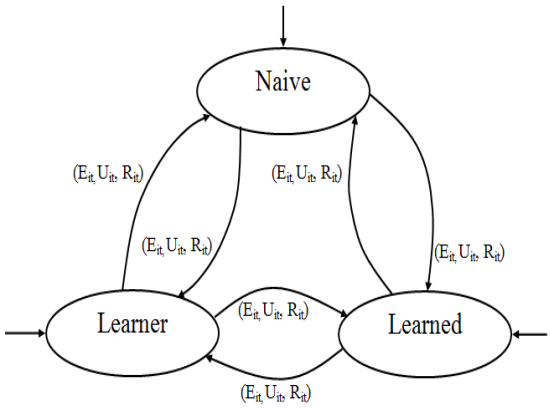


Figure 3.9: Operation of LA based predictive clustering where all three states can act as initial states

3.4.2 LA Based Future Position Prediction Scheme

The proposed LA based clustering scheme is described in algorithm 3.5 and algorithm 3.6. The vehicles first initialize the prediction accuracy vector based on the threshold value of the prediction accuracies. Then they compute their prediction accuracy and the LA modifies these values based on it's current actions. By repeating the above process for k iterations till required accuracy is not achieved, the LA helps to achieve relatively high value of accuracy in prediction.

We use a Prediction Accuracy Vector $\vec{P}$ to represent the prediction probability of action at a particular instant of time. Here $\vec{P}$ consists of three parameters E_{it}, U_{it} *and* R_{it} that denote three possible variations by LA. These variations are based on the difference between the predicted, and actual position for any vehicle.

As shown in Figure 3.12, the value of prediction accuracy can be Enhanced (E_{it}), kept Unchanged (U_{it}), or Reduced (R_{it}) which then affects the prediction accuracy for the vehicle. Figure 3.12 classifies a vehicle into three states named as *Naive or Beginner, Learner or Average and Learned or Expert* prediction state. Based upon its prediction accuracy, a vehicle can belong to any of these three states at the start and then depending upon its current prediction, the vehicle moves from one state to another.

Algorithm 5 ***Predictive Clustering using Learning Automata***

Inputs: *Predicted Accuracy, Total iterations, Threshold, Learning Scheme modification parameters* $\lambda_{R_1}, \lambda_{R_2}$*, Current Time (t), Previous Time (t-1)* ,
Outputs: *Optimized prediction accuracy Assignment Vector*
Assumption: *N is the total number of vehicles*

1: **Begin**
2: Initialize $\lambda_{R_1},\ \lambda_{R_2},\ K$
3: **for** $(i = 1; i \leq N; i++)$ **do**
4: **if** $(CurrentState == Naive)$ **then**
5: $E_{it} = \frac{1}{K}$
6: $U_{it} = \frac{1}{K}$
7: $R_{it} = 0$
8: **end if**
9: **if** $(CurrentState == Learner)$ **then**
10: $E_{it} = 0$
11: $U_{it} = \frac{1}{K}$
12: $R_{it} = \frac{1}{K}$
13: **end if**
14: **if** $(CurrentState == Learned)$ **then**
15: $E_{it} = \frac{1}{K}$
16: $U_{it} = \frac{1}{K}$
17: $R_{it} = \frac{1}{K}$
18: **end if**
19: $Pred\_Acc_{it} = actual\_pos_t - pred\_pos_t$
20: **end for**
21: **for** $(i = 1; i \leq N; i++)$ **do**
22: $Avg\_Pred\_Acc_t = \frac{1}{N}\sum_{i=1}^{N} Pred\_Acc_{it}$
23: **end for**
24: $Threshold = Avg\_Pred\_Acc_t + \frac{1}{t-1}\sum_{i=1}^{t-1} Avg\_Pred\_Acc_t$
25: **for** $(i = 1; i \leq N; i++)$ **do**
26: $Opt\_Pred\_Acc = MAX(Opt\_Pred\_Acc_{it})$
27: **end for**

The procedure for modifying the prediction accuracy, starts with the value of the vector initialized to a random value which is assumed to be in the currently feasible range. Then, the network initiates the learning phase where, Linear Reward Inaction (L_{RI}) based strategy is used for identifying more suitable values by increasing or decreasing the prediction accuracy. To achieve this objective, an automaton raises or lowers the current

28: **while** $(count \leq K)$ && $(Threshold < Reqd\_Pred\_Acc_t)$ **do**
29: **for** $(i = 1; i \leq N; i++)$ **do**
30: $Action\_i = RAND(E_{it}, U_{it}, R_{it})$
31: **if** (Prediction is feasible) **then**
32: **if** $(Action_i == E_{it})$ **then**
33: $Mod\_Pred\_Acc(D_{it}, \lambda_{R_1})$
34: **end if**
35: **if** $(Action_i == U_{it})$ **then**
36: $Mod\_Pred\_Acc(S_{it}, \lambda_{R_1})$
37: **end if**
38: **if** $(Action_i == R_{it})$ **then**
39: $Mod\_Pred\_Acc(D_{it}, \lambda_{R_1})$
40: **end if**
41: **else**
42: Maintain previous prediction accuracy
43: **end if**
44: **if** $(Prediction\ is\ feasible)$ && $(Pred\_Acc_{it} < Opt\_Pred\_Acc)$ **then**
45: **if** $(Action_i == E_{it})$ **then**
46: $Mod\_Pred\_Acc(D_{it}, \lambda_{R_2})$
47: **end if**
48: **if** $(Action_i == U_{it})$ **then**
49: $Mod\_Pred\_Acc(S_{it}, \lambda_{R_2})$
50: **end if**
51: **if** $(Action_i == R_{it})$ **then**
52: $Mod\_Pred\_Acc(D_{it}, \lambda_{R_2})$
53: **end if**
54: **end if**
55: Update State of i using $< E_{it}, U_{it}, R_{it} >$
56: **end for**
57: **end while**
58: **End**

action probability vector associated with every vehicle. The associated probability vector for this assignment is then modified to yield the updated solution and the new probability vector. Following scenarios are considered for updating the value of probability vectors.

Scenario I: This is the scenario when the prediction accuracy index of the node is at the lowest possible value, called the 'naive' state. In such a case, following holds,

$$E_{it} = \frac{1}{K}, U_{it} = \frac{1}{K}, R_{it} = 0 \qquad for\ \ 0 < t < T. \tag{3.54}$$

Scenario II: This is the scenario where the prediction accuracy index of the node is in learner states, i.e., the velocity can be raised or lowered or maintained the same, and

hence, following holds,

$$I_{it} = \frac{1}{K}, U_{it} = \frac{1}{K}, R_{it} = \frac{1}{K} \qquad for \;\; 0 < t < T. \tag{3.55}$$

Scenario III: This is the scenario where the prediction accuracy index of the node is at the highest possible value, n, called the 'learned' state, i.e., that the velocity cannot be raised further. Thus,

$$I_{it} = 0, U_{it} = \frac{1}{K}, R_{it} = \frac{1}{K} \qquad for \;\; 0 < t < T. \tag{3.56}$$

Algorithm 6 ***Mod_Pred_Acc***

Inputs: *Action at time t: Enhance, Unchanged, Reduced*
Learning Scheme modification parameters λ_R,
Current probability vector$< E_{it}, U_{it}, R_{it} >$
Outputs: *Modified probability vector* $< E_{it}, U_{it}, R_{it} >$

Begin
if (Action == Enhance) **then**
$D_{it} = \lambda R * D_{it}$
$S_{it} = \lambda R * S_{it}$
$I_{it} = 1 - (D_{it} + S_{it})$
end if
if (Action = = Unchanged) **then**
$I_{it} = \lambda R * I_{it}$
$D_{it} = \lambda R * D_{it}$
$S_{it} = 1 - (I_{it} + D_{it})$
end if
if (Action = = Reduce) **then**
$I_{it} = \lambda R * I_{it}$
$S_{it} = \lambda R * S_{it}$
$D_{it} = 1 - (I_{it} + S_{it})$
end if
END

The final solution vector comprises of prediction accuracies, that exhibits S_{it} probability values which are closest to the converging value of unity. The closer this value is to unity, higher the level of accuracy in the predicted position of each vehicle. Using this terminology, Algorithm 3.15 and 3.16 describes the initial configuration as well as updating strategies for the proposed prediction scheme.

3.4.3 CH Election Process

Algorithm 7 ***CH Election using CDS***

Inputs: *Predicted position of vehicle*

Outputs: *G(V, E) containing CDS*

Assumptions: *UNVISITED, VISITED, COVERED and CDS are sets of vehicles containing the number of remaining, dominating, dominatee and CH nodes respectively*

Begin
while All vehicles not covered **do**
for $(i = 1; i \leq N; i++)$ **do**
Broadcast computed predicted position by LA
end for
Compute number of neighbors of vehicle denoted by σ_i
$v[i] = \sigma_i$
for $(i = 1; i \leq N; i++)$ **do**
mark all nodes as UNVISITED
set $j = max(v[i])$
Add $v[j]$ to CDS and mark $v[j]$ as VISITED
for $(k = 1; k < N; k++)$ **do**
if $(v[k]\ \epsilon\ CDS)$ **then**
Mark every neighbor of $v[k]$ as COVERED
end if
end for
for $(k = 1; k < N; k++)$ **do**
if $(v[k]\ \epsilon\ COVERED)$ **then**
Count the unvisited neighbors of $v[k]$
end if
end for
Find covered nodes with MAX number of UNVISITED neighbors
Mark this nodes as VISITED and each neighbor as COVERED
Add new COVERED node to CDS
end for
Until (all nodes ϵ COVERED)
end while
End

The proposed algorithm is capable of performing both probability updating and action selection in a small interval of time, regardless of the number of actions. This ability is also an additional feature of our proposed scheme.

Once each vehicle has determined its future position, the decision about the next CH to be selected is done using a centralized computing environment such as RSU's or Vehicular Clouds. The nodes elect the CH by transmitting their predicted position to the centralized computing repository. However, instead of using traditional graph-based metrics, the proposed scheme uses a cluster formation parameter (σ) for creating the dominating set. The number of neighboring nodes is computed based on the transmission range of the vehicles. All the vehicles within the transmission range are assumed to be the neighbors of the vehicle. The value of σ is based on the number of neighboring nodes. Initially, the node with higher σ values are elected as Dominating nodes or CHs and the nodes in their one-hop neighborhood then assume the role of Dominatee nodes. Algorithm 7 describes the key steps performed in the selection of CH using Connected Dominating Sets.

3.5 Performance Evaluation

To evaluate effectiveness of proposed schemes, performance evaluation through extensive simulations has been carried out. The preceding sections provide a detailed analysis and comparative evaluation of predictive clustering schemes.

3.5.1 Simulation Settings

In order to generate realistic vehicle mobility pattern, the microscopic traffic simulator Simulation of Urban Mobility (SUMO) [148] was used for simulation of the proposed scheme. The mobility pattern was generated on the map of Chandigarh city as shown in Figure 3.10. The map consist of bi-directional lanes and number of bi-directional lanes varies between 2 to 4. For simulation, vehicle's acceleration rate of $0.8ms^2$, the deceleration rate of $4.5ms^2$ with the maximum speed of 36 m/s are used. The lane speed was varied from 10 m/s to 30 m/s. However, the entry rate of vehicles in the network vary from 10 vehicles/min to 24 vehicles/min. In order to provoke dynamic changes, vehicles can join and leave the VANET composition from different entry and exit points at different rates. The message transmission range of each vehicle was 250 m to 300 m. The proposed scheme is compared with VWCA [103]. Following parameters are selected for evaluation of the proposed scheme.

- *Packet Delivery Ratio(PDR)*: It is defined as the average value for ratio of successful packets transmitted to the total number of packets sent from source to destination.

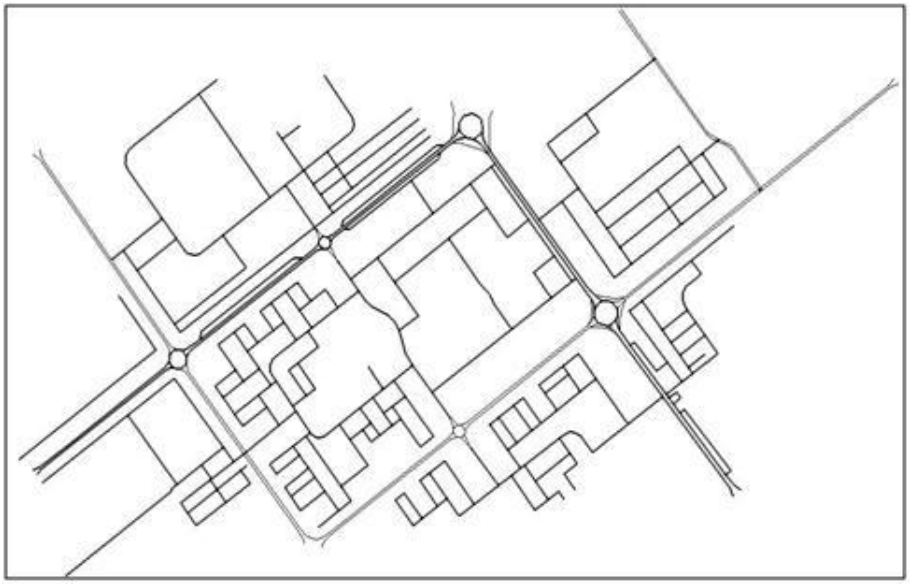


Figure 3.10: Simulated road map of Chandigarh City for predictive clustering.

- *End-to-End delay(E2ED)*: It is defined as the total delay incurred for the message transmission from source to destination.
- *Throughput*: It is defined as the average number of packets transmitted in a unit interval of time from a particular node.
- *Probability of message transmission*: It is defined as the probability of message transmission in presence of constraints such as high velocity and density of vehicles on the road.

3.6 Results and Discussions

3.6.1 Predictive Clustering Scheme

Figure 3.11 (a) shows the variation of PDR with number of nodes. The PDR is very high, when the number of nodes is low. It is due to the reason that the number of messages exchanged is less and there is low burden on network and hence the percentage of dropped packets is also less. Also, when the number of nodes increases, the value of PDR decreases. The decrease in PDR is less in the proposed protocol as compared to VWCA protocol. The PDR is 95% when there are 200 vehicles in the network and reduces to 85% as vehicles increases to 450 in the network. In contrast, the PDR for VWCA is 89% when number of vehicles is 200 and as low as 75% when vehicles are 500 in number. The reason for the improved performance of the proposed scheme is due to lesser number of beacon messages being used in the proposed scheme. Also, future mobility prediction is done in

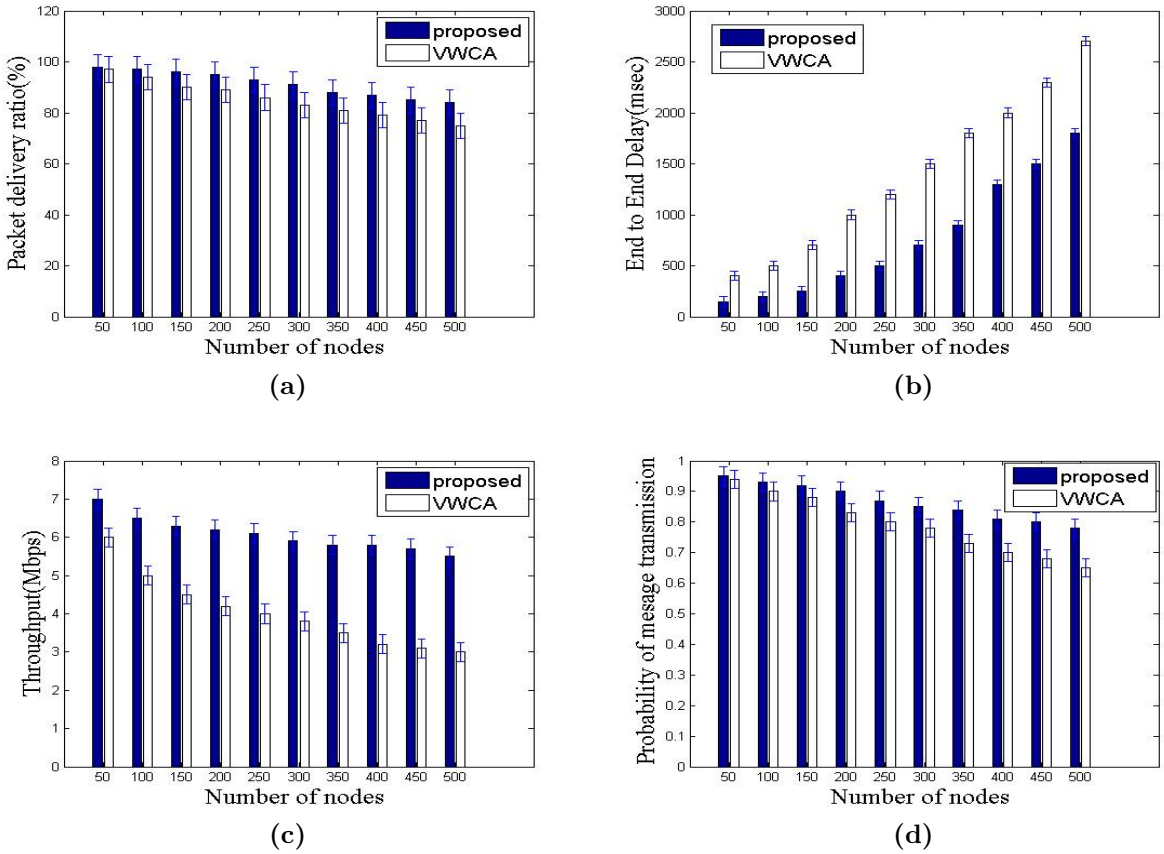


Figure 3.11: (a) Variation in packet delivery ratio with increasing number of vehicles (b) Variation in End-to-End delay with increasing number of vehicles (c) Variation in throughput with increasing number of vehicles (d) Variation in probability of message transmission with increasing number of vehicles in predictive clustering scheme.

the proposed scheme which will guide the vehicles to take adaptive decisions for route selection and data dissemination. This phenomenon results a less overhead generation in the proposed scheme that contributes to its improved performance.

Also, E2ED increases with number of vehicles mainly due to extra burden in network that causes collision and loss in messages. The E2ED of proposed scheme is 150 msec and increases to 900 msec when number of vehicles is 50 and 350 respectively. In contrast, the E2ED is 1800 msec when the number of vehicles is 350 and increases to 2700 msec when the number of vehicles is 500. The variation of E2ED with increasing number of vehicles is increased to 3.11(b). Again, the better performance of the proposed scheme with respect to this metric is attributed to prior prediction of the position of vehicles which helps in improving the message delivery to their destinations. The lesser number of beacon messages from nodes in the network also helps in reducing the E2ED as this type of phenomenon is missing in the conventional scheme

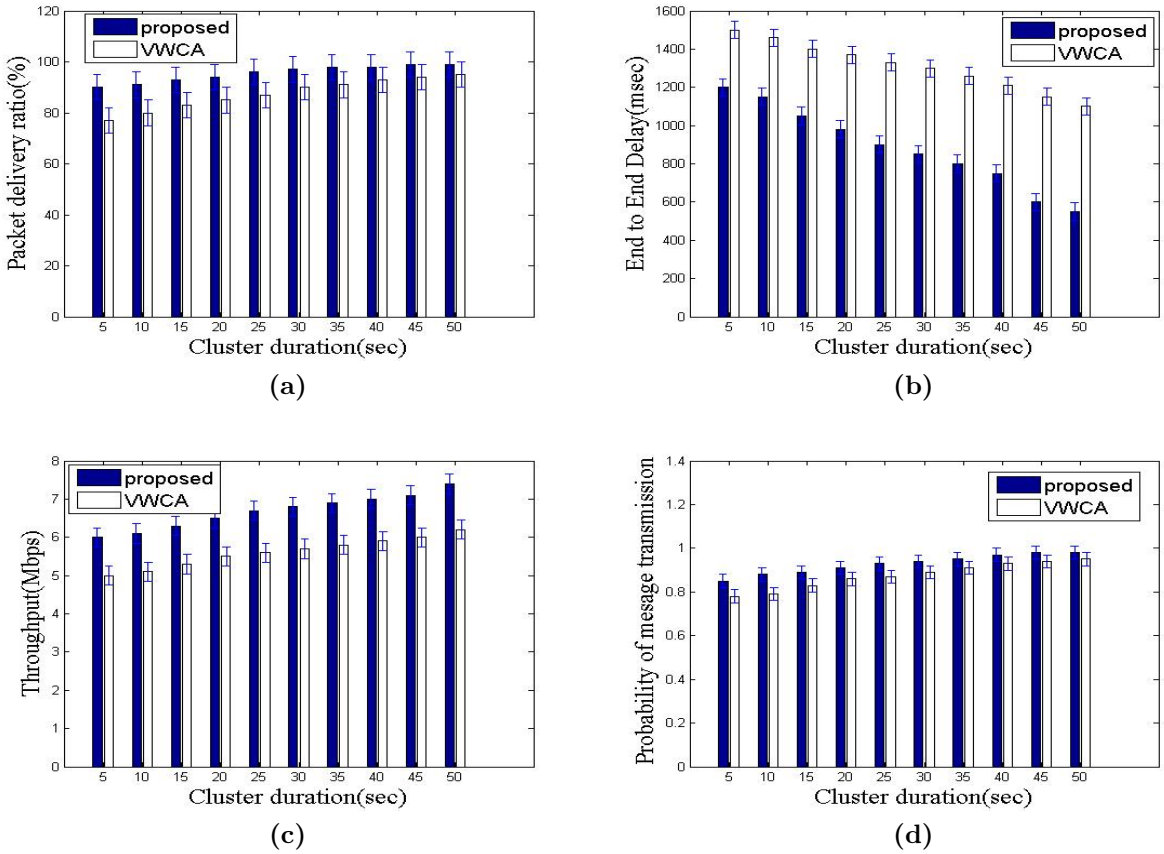


Figure 3.12: (a) Variation in packet delivery ratio with increasing cluster duration (b) Variation in End-to-End delay with increasing cluster duration (c) Variation in throughput with increasing cluster duration (d) Variation in probability of message transmission with increasing cluster duration in predictive clustering scheme.

As shown in figure 3.11(c) and 3.11(d), throughput and probability of message transmission marginally decreases with the increase in number of vehicles. The decrease is marginal in case of proposed scheme, the negative slope is very high for VWCA. Throughput is 5.5 Mbps for proposed scheme while it is as low as 3 Mbps for VWCA when there are 500 vehicles in the network. The probability of message transmission has a difference of 0.13 between two schemes when the number of nodes are 500. Again, due to the less beaconing and adaptive decision making capabilities, it results in improved values of throughput and probability of messages transmissions in the proposed scheme.

As shown in figure 3.12(a), the PDR increases as the cluster duration increases since the time taken to deliver the messages increases when the channel is busy. PDR increases from 90% to 98 % as the cluster duration increases from 5 seconds to 40 seconds.The same factor causes E2ED to decreases which is represented in figure 3.12(b). The throughput and probability of message transmission also increases with increase in cluster duration as shown in figure 3.12(c). The value of throughput increases from 6 Mbps to 7.4 Mbps

when the cluster duration rises from 5 seconds to 50 seconds. Similarly, the increase in probability of message transmission is 0.13 as there is an increase in cluster duration. The relative position of CHs within a cluster is a factor as the CHs are selected based on their position with respect to all the other nodes. This impacts the probability of message transmission and contributes to its improved values.

From the above discussion, it can be concluded that the proposed scheme performs better than conventional scheme due to the less beaconing and adaptive decisions making capabilities. As the overhead generated and message complexity are less in the proposed scheme, so less time is consumed in performing various operations in the network which makes the proposed system more efficient.

3.6.2 LA based Predictive Clustering

To evaluate the performance of the proposed scheme, ns-2 simulator was used. Vehicular mobility traces were generated using SUMO traffic simulator[148]. The network scenario used in the simulation consists of vehicles moving in urban scenario, with the road map based on Chandigarh city as shown in Figure 3.13. The map consists of bidirectional lanes, where the number of lanes in each direction varies between 2 to 4. The parameters

Figure 3.13: Road Map of Chandigarh City used in Simulation of LA based Predictive clustering.

considered for evaluation are delay, throughput, overhead, and packet loss. All simulations were done with different set of vehicles which enter the Chandigarh city from different entry points and the vehicles meet at traffic light junctions and roundabouts. The values of main parameters used in the simulation are summarized in Table 3.2.

Table 3.2: Simulation parameters required for LA based predictive clustering scheme.

Parameters	Value
Number of Vehicles	100,200,300
Transmission range	300 m
Permissible lane speed	10-30 m/s
Number of bi-direction lanes	2,4
Vehicle acceleration range	0.8 m/s^2
Vehicle deceleration range	4.5 m/s^2
MAC protocol	IEEE 802.11p
Simulation Time	500 sec
Prediction interval	10-120 sec

We have evaluated and compared the PCALA approach with our previous existing clustering scheme, the Energy Efficient Predictive Clustering (EEPC) [156]. To evaluate the performance of the proposed clustering algorithm, the following performance metrics were used.

- **Prediction Variance:** It is defined as the difference between the predicted position of a vehicle and subsequent actual position that the vehicle reaches at any instant.
- **Prediction Variance per Second:** It is defined as the prediction accuracy of a node in predicting its position for every unit time interval.
- **Average Prediction Accuracy:** It is defined as the prediction accuracy of the vehicle over a n number of consecutive prediction intervals.
- **End-to-End Message Latency:** It is defined as the total time taken by a packet to reach to its destination using the elected CHs.
- **Average Cluster Head lifetime:** It is defined as the average time period for which a node acts as a CH with respect to the number of CH nodes.
- **Percentage of CHs or CH density:** It is defined as the ratio of number of CHs to the total number of nodes at discrete time interval.
- **Packet Delivery Ratio:** It is defined as the number of successful packets received to the total number of packets transmitted in a unit interval of time.

Figure 3.14(a) represents the variation in prediction variance as a function of the prediction interval. The results indicate that the difference in the value of the prediction variance increases as we increase the prediction interval for both the schemes. This may be attributed to the fact that the higher prediction interval increases the prediction variation due to the vehicular traffic conditions. However, the use of LA appreciably decreases

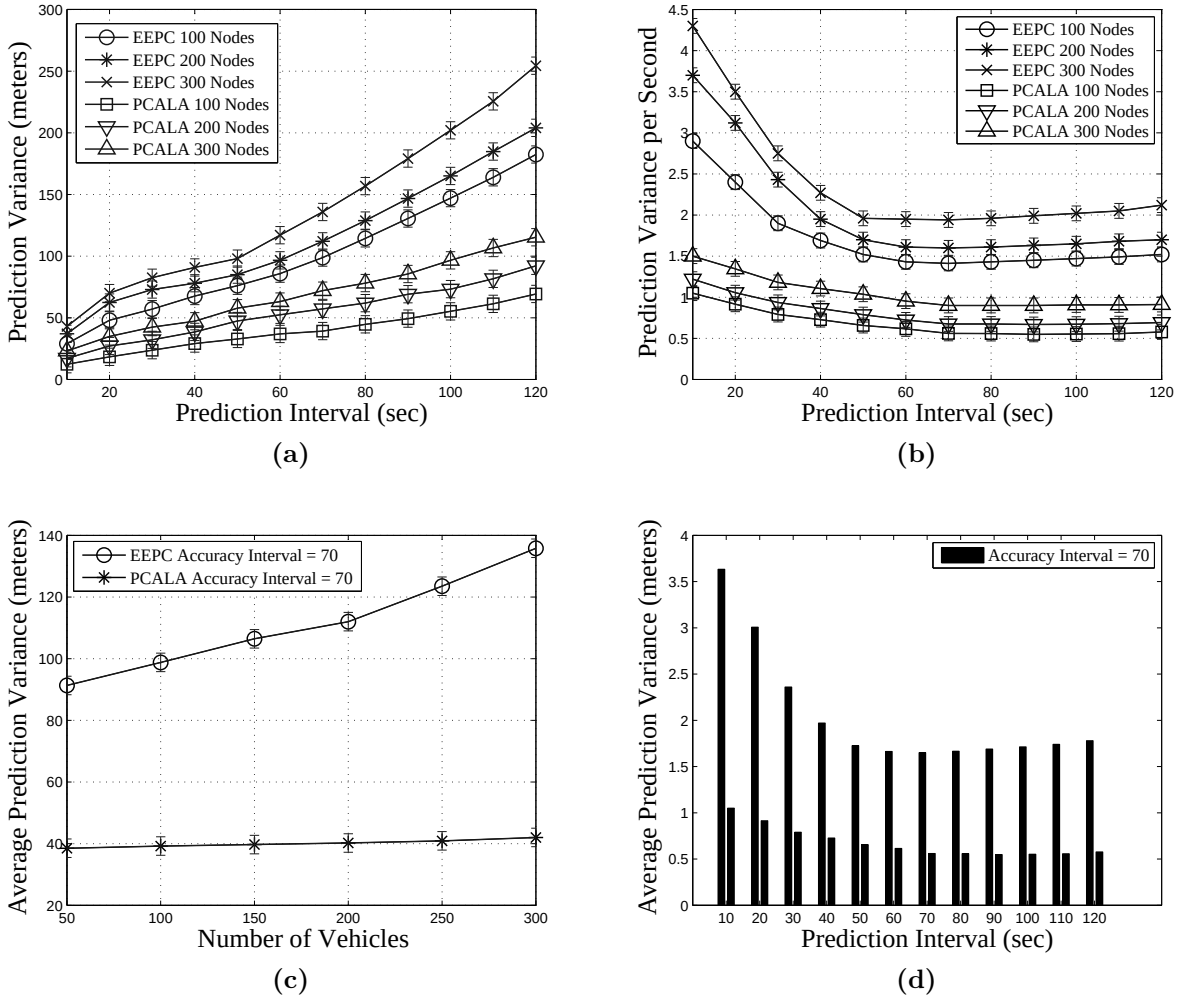


Figure 3.14: (a) Comparison of prediction accuracy with respect to prediction interval (b) Comparison of average prediction accuracy with respect to prediction interval. (c) Variation in prediction accuracy with respect to number of vehicles (d) Comparison of average prediction accuracy against prediction intervals in LA based predictive clustering scheme.

this variation, especially when the prediction interval increases.

Figure 3.14(b) describes the comparative relationship between the average prediction accuracy and the prediction interval for the predictive clustering schemes. The results obtained show that the prediction accuracy ranges between a small value for both the clustering schemes. The value of prediction accuracy reaches its optimum level at 70 seconds. The results also show that the value of prediction accuracy for PCALA is better than the prediction scheme.

Figure 3.14(c) describes the variation between the average prediction accuracy in terms of number of vehicles. The results are plotted for a prediction interval of 70 seconds as it results in optimum values of prediction accuracy for both PCALA and EEPC. The PCALA approach achieves a superior value of prediction accuracy as compared to the traditional predictive clustering scheme based on distance. The average prediction accuracy for PCALA shows a minor variation in the interval of approximately 37 to 43 meters with the number of vehicles whereas for EEPC, this variation is between 90 to 130 meters for the same number of vehicles. This is due to the incorporation of LA into the prediction schemes which results in better and stable prediction.

Figure 3.14(d) compares the average prediction accuracy and prediction intervals for PCALA and existing prediction schemes. Figure 8 shows that the variation in average prediction accuracy for PCALA is better as compared to EEPC. This is attributed to the fact that PCALA considers the existing vehicular conditions for generating a real-time prediction. LA continuously learns from its environment and modifies its action accordingly. This helps to achieve better prediction accuracy.

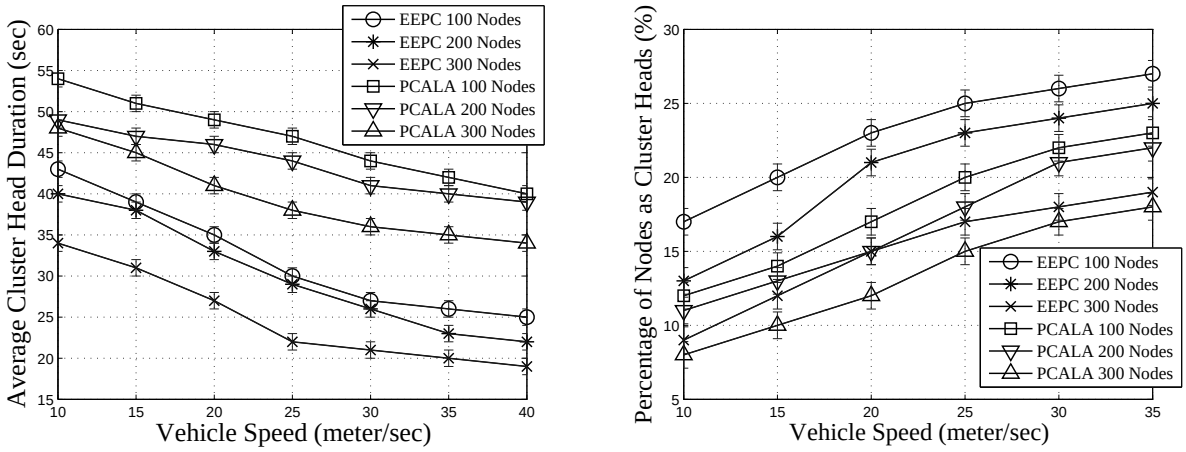


Figure 3.15: (a) Average cluster head duration as a function of vehicle speed (b) Percentage of vehicles as cluster heads with respect to vehicle speed in LA based predictive clustering scheme.

Figure 3.15(a) describes the relationship between average CH duration and vehicle speed. The proposed PCALA algorithm gives better value than conventional predictive algorithm. However, the value of average CH lifetime decreases with an increase in maximum lane speed for both PCALA and EEPC. This is because at high speed, the CH remains in the cluster for a lesser amount of time. As the traffic density increases, average CH lifetime decreases for both the techniques. This is due to the fact that in more dense traffic scenarios, a CH comes in the transmission range of comparatively more CHs. This results in higher cluster formation. The average CH duration in PCALA is

approximately between 35 to 54 seconds whereas, it varies approximately between 19 to 34 seconds in the case of EEPC. Thus, PCALA provides higher cluster stability due to the superior value of CH duration.

As shown in Figure 3.15(b), the percentage of CHs increases with an increase in vehicle speed. The number of CHs increase as the speed of vehicle is increased for both PCALA and EEPC. This pattern occurs because at high speed, the existing clusters break down into small clusters frequently and new clusters are built. However, cluster-density is higher in the low traffic scenario than in high traffic environment and it increases on faster lanes. This occur because, less clusters in the low traffic scenario may deteriorate communication between vehicles. The PCALA clustering mechanism avoids such situation by increasing the number of clusters as compared to EEPC and provides better network connectivity to vehicles.

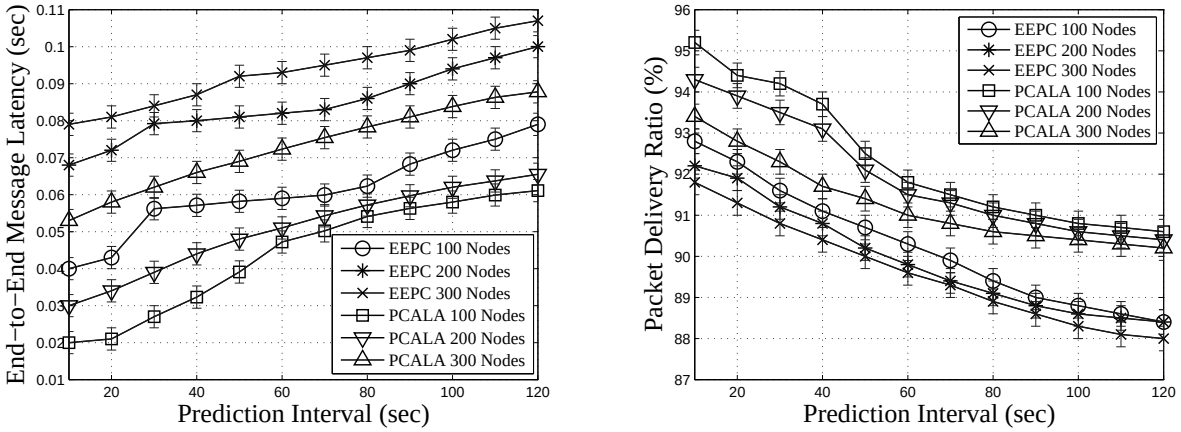


Figure 3.16: (a) Comparison of End-to-End Latency and prediction intervals. (b) Variation of PDR with respect to prediction intervals in LA based predictive clustering scheme.

In Figure 3.15(a), variation in end-to-end latency is compared against the prediction intervals for different number of nodes. The results obtained show that the latency increases linearly with an increase in prediction interval and LA help to achieve lower latency. Although, the end-to-end latency increases with prediction interval for both PCALA and EEPC, but PCALA has marginally lower latency for all the prediction intervals. The value of latency also increases as the number of nodes are increased. Since the packet overheads are reduced in PCALA due to comparatively less beacon messages transmitted, it impacts the value of latency. The EEPC scheme requires periodic transmission of beacon messages for updating the prediction accuracy parameters. However in PCALA, every node performs the prediction based on its own vehicular dynamics. This helps in reducing latency in PCALA.

Figure 3.15(b) describes the variation of PDR in terms of prediction interval. The increased prediction interval causes higher message overhead which decrease the value of PDR as the prediction interval increases. The PDR for both EEPC and PCALA decreases as the prediction accuracy increases. This is due to the fact that increased network traffic is observed as the prediction intervals increases. PDR values also decrease when the number of nodes increase from 100 to 300 for both the schemes. However, the PDR of PCALA is higher and ranges approximately between 91% to 96% as compared to 93% to 88% for EEPC. This is because the reduced packet overhead results in better packet delivery in PCALA as compared to EEPC.

3.7 Conclusion

In this chapter, a predictive clustering scheme for an urban traffic roadside scenario with every road having multiple lanes and several traffic entry and exit points. The proposed protocol addresses the key issues such as high reliabe message transmission and throughput with minimum delay. We have next proposed an average predictive variation algorithm which predicts the average variation of the vehicles on the road to overcome the challenge of small duration of cluster membership of the vehicles on the road. Also, future mobility of the vehicles are estimated so that proactive measures can be taken for reliable message transmission on the road. The proposed scheme was evaluated in different network conditions by varying the cluster size and number of vehicles on the road.

We have also proposed an LA-based hybrid predictive clustering scheme to increase the reliability of the message delivery. LA are assumed to be deployed on the vehicles and future mobility of the vehicles is estimated using which decisions about the clustering are made in the proposed scheme and CHs are elected using connected dominating set. The LA based predictive clustering technique is compared with traditional predictive clustering scheme where the prediction accuracy was not done to its full potential. Based upon the rewards and penalties obtained, future actions to be taken are estimated in the game. The performance of the proposed schemes were evaluated using various metrics where their performance was found satisfactory with respect to the existing scheme. The results obtained also show that these predictive clustering schemes are effective in maintaining high throughput, lower end-to-end delay and high probability of transmissions.

Chapter 4

Intelligent Cluster Based Intrusion Detection

Vehicular Cloud Computing (VCC) has emerged as a promising technology which is being used in wide range of applications. In VCC, vehicles act as the intelligent machines which can be used to collect and transfer the data to local or global sites for storage and computation purposes, as vehicles are having comparatively limited storage and computation power for handling these multimedia files. However, due to the dynamic changes in topology, and lack of centralized monitoring points, this information can be altered or misused. These security breaches can result in disastrous consequences such as-loss of life or financial frauds. Therefore, to address these issues, a Learning Automata-assisted Distributive Intrusion Detection System is designed based on clustering. In the proposed scheme, LA are assumed to be stationed on the vehicles which take clustering decisions intelligently and select one of the members of the group as a CH. The CHs then assist in efficient storage and dissemination of information through a cloud based infrastructure. To secure the proposed scheme from malicious activities, standard cryptographic technique is used in which the automation learns from the environment and takes adaptive decisions for identification of any malicious activity in the network. A reward or a penalty is given by the stochastic environment where an automaton performs its actions so that it updates its action probability vector after getting the reinforcement signal from the environment.[1]

From the past few years, there has been an exponential growth in the usage of Internet by community of users to access various resources from far flung places even on-the-fly. This has become possible due to the evolution of Internet of Things (IoT), in which various

[1]The content of this chapter has been taken from :

- Neeraj Kumar, Jaskaran Preet Singh, Rasmeet S. Bali, Sudip Misra, Sana Ullah "An Intelligent Clustering Scheme for Distributed Intrusion Detection in Vehicular Cloud Computing", Cluster Computing, Springer, Volume 18, Number 3, pages 1263–1283, 2015.

objects/devices are connected to the Internet, and communicate with one another using different protocols and standards to share their data. In this environment, vehicles can be viewed as intelligent objects which are capable of capturing, storing, and processing the collected data from the environment where these objects are deployed.

Objects in IoT environment can also be interconnected with vehicular nodes to form a type of VANET which is used to provide efficient inter-vehicle communication between different objects/vehicles connected to Internet or some other wireless networks. The inter vehicle communication takes place over a range of 300 to 500 meters using IEEE 802.11 protocols, based on the DSRC standard [157]. In Vehicle-to-Sensor(V2S), vehicles communicate with onboard sensors or other sensor deployed on the road using low and high data rates communications such as Bluetooth, Radio Frequency Identification (RFID), and Wi-Fi [[157], [158]].

As vehicles are constrained in terms of resources such as computation and storage so, there is a requirement of some schemes which integrate vehicles to cloud using Internet so as to access various resources from the cloud infrastructure [159]. From the cloud, resources can be accessed at any time even on-the-fly as per the requirements of the end users [160]. Cloud computing can thus be integrated with VANETs for providing on demand, convenient access to shared resources such as-infrastructure, storage, and computation. In this direction, VCC has emerged as a new technology in which vehicles get real time information about the resources and services they can access from the remote sites on-the-fly [161].

As evident from the above, although VCC has many advantages for multimedia-based transmissions, but it also has many challenges, especially with respect to security and privacy as it is hard to maintain a centralized control in this environment due to high velocity of the vehicles [162]. So, it is important that the communication between different vehicles should not be modified or wrong information should not be inserted by malicious persons [[163],[145]]. Current security mechanisms for VANETs use cryptographic techniques provided by IEEE 802.11/p WAVE standard [[158],[164]]. Apart from cryptographic mechanism, certain models proposed in [[165],[166],[167]] use Misbehavior Detection Systems (MDS) for detecting malicious vehicles in VCC. Clustering is also an efficient technique for topology preservation in VCC [[67],[94]].

Keeping focus on all the above discussed issues, a new learning automata-based intelligent clustering is proposed for distributive intrusion detection using VCC. Learning automata are assumed to be deployed on the vehicles by creating clusters among the vehicles. Different from the other existing techniques, the intrusion detection in the proposed scheme is not limited to CH only. But, it is performed intelligently in the backbone network by the cluster leader. The proposed scheme uses cryptographic mechanism to ensure secure communication among all the participants in the network. The crypto-

graphic mechanism uses a cloud-based CA to distribute keys and digital certificates to nodes/vehicles in the network. Learning automata perform various actions after getting the inputs from the environment and update their action probability vector.

The important parameters in the proposed IDS model are number of VANET facilities required and its infrastructure support. This model can be a framework for information exchange which often requires and uses data stored at a cloud provider. In this model vehicles manage and control the user data about the passengers while on the road. However, a common security issue shared by VCC applications are ownership, authentication, authorization, integrity and confidentiality. Ownership of information is defined as creator of information. Establishing the ownership of information is necessary for protection against unauthorized access or misuse of a user's information. Authenticity in general refers to the truthfulness of origins, attributes, commitment and intensions. Integrity means preserving the accuracy and consistency of data. The authentication of information can pose problems such as, Man-In-The-Middle (MITM) attacks and is often implemented with authentication identity. To secure the personal record from the malicious activities, the proposed model uses Intrusion Detection System (IDS) based authentication and integrity mechanism. The proposed scheme is also evaluated using different performance evaluation metrics in VCC environment.

4.1 Distributed Intrusion Detection System

Cloud-based Vehicular networks have emerged as an indispensable technology which can be used for transmitting different types of collected multimedia data from the users using small and long range transmissions. In order to provide cost saving, ease of use and high throughput, the proposed model can be utilized by a number of applications. Thus, VCC inherits a number of benefits compared to traditional cloud environment such as dynamic provisioning, scalability, and ease of integration, and seamless mobility. Moreover, by incorporating Intrusion Detection System into such a system further helps to improve the reliability and security of applications, as the data is maintained by cloud based remote sites. The proposed system model is shown in Figure 4.2 having key components as- Vehicular networks, cloud services, distributed intrusion detection system and users. The vehicles collect and process user data and the distributed intrusion detection system ensures the integrity and authenticity of data. The cloud services are used for processing as well as transmitting the data to the users through vehicles. We assume that cloud servers have sufficient Virtual Machine (VM) resources for providing the required services to the end users. Thus, the proposed system model can be used for sharing data to authorized networks for providing enhanced services to the users and thereby increasing the applicability and impact of traditional systems through increased reachability.

4.1.1 Attack Model

In this section, some standard attacks that are implemented to detect the efficiency of the proposed model are discussed along with how those are defended against in the proposed IDS. In the proposed work, two types of attacks are implemented that represent flooding and black hole attack which can occur for transmitting multimedia data. In flooding attack, a node floods a whole network with large number of packets, either by replaying the same message or by generating the artificial message in very short duration. The receiving node has to drop all the packets as it cannot process all the packets in short duration. In black hole attack, a node impersonates itself as a destination node by forging its identity and drops all the packets that are routed through it. A forged identity is created by generating and distributing a fake routing table. Due to black hole attack, a message will not reach the intended recipient or may reach its destination after considerable delay.

In order to detect black hole attack, the PDR of each node in the cluster is compared with the threshold value. If PDR is greater then a predetermined threshold value, a node is considered as malicious. The threshold is set to the maximum value of PDR for all the nodes in each cluster when there is no attack in the network. Also the number of packets a node can accept at a time is equal to its queue length in our simulations. This queue length must be announced in the cluster whenever a node is ready to receive data.

In flooding attack, the behavior of sender is considered suspicious if the receiver is receiving packets at a rate equal to twice its queue length per second. However, a node can be considered as malicious only if all the packets received during this time are for same destination. The proposed model uses authentication mechanism for preventing a node from forging its identity. Each node is assigned a digital certificate by CA, which is linked to the real identity of the equipment owner. Thus, a node cannot forge its identity because of digital certificate but still if a node drops all the packets that are routed through it then, it is considered as malicious by the detection mechanism. Thus, the proposed intrusion detection model can be considered as an anomaly detection mechanism in VCC environment.

The architecture of Vehicular Clouds built on Vehicular networks consists of intelligent vehicles having learning automata deployed on them which serve as collectors of information and gateway vehicles dedicated for cloud access via existing networks. They coordinate with application servers or make local decisions and then, provide the require information to the users on vehicle. The cloud controllers then schedule the data dissemination tasks on VMs. These VMs return the results and are rented by application service providers. There are situations where the deployment of cloud can be done on application servers. However, the security and confidentiality of this information needs to be preserved by using the IDS.

4.1.2 Network Architecture and Assumptions

In the proposed model, the intrusion detection mechanism is implemented on clusters of vehicles. These vehicles, without disconnection, can communicate with one another. The responsibility of CH is to coordinate the communication of nodes within the nodes of cluster and with nodes of other clusters [67]. In the proposed model, instead of one CH, each cluster consists of a number of leaders that form a backbone known as cluster leadership/leadership. One leader from the leadership plays the role of cluster-head. All

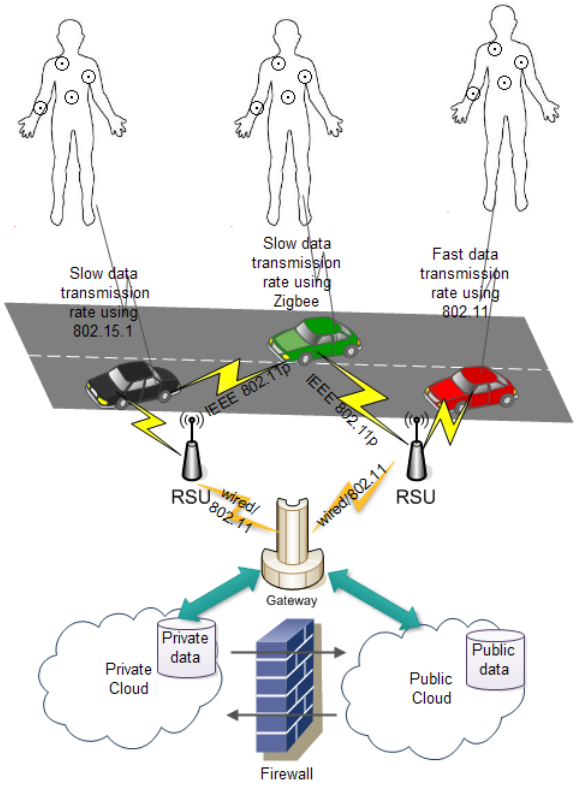


Figure 4.1: Network model used in distributive intrusion detection system.

the other remaining nodes are designated as collector nodes having components such as sensor nodes for collecting the data. Figure 4.2 shows the generalized network model used in the proposed scheme.

Figure 4.2 describes the layered framework of the proposed system that includes application middle ware and physical layer standard. The RSUs and cloud based services on different layers can be integrated to exchange information, share resources and collaborate on the clouds [168]. In this proposed layered architecture, different layers have different purposes. In general, the layers on the bottom provide a foundational support for the layers on the top. Software Oriented Architecture (SOA) will be applied to integrate different information and communication services and connect in-vehicle and out-vehicle applications seamlessly though the vehicular data clouds. SOA allows vehicular application developers to organize, aggregate and package the available data delivery applications into

aggregated application services. Middleware is used to hide the implementation details of underlining technologies and provides support for the integration of specific applications deployed on the vehicular data cloud [[169] [169]] as shown in figure 4.2,

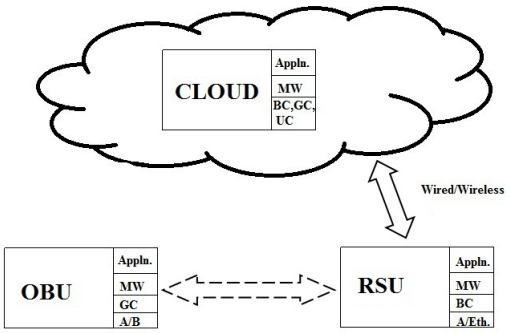


Figure 4.2: Framework used in distributive intrusion detection system.

The vehicles which are moving in same direction and in same lane are grouped in the same cluster. A leader is selected by considering its average relative velocity and degree of connectivity. The selection of CH is based on vehicular dynamics and a leadership node with minimum differential velocity is elected as CH. In order to make the proposed scheme tolerant to malicious activities in the cluster, a threshold value is defined. The threshold value (τ), determines the number of malicious activities that are allowed in the cluster. The value of τ is calculated using Equation 4.6. Here, L is the number of leaders in the cluster.

$$\tau = \frac{(L-1)}{2} \tag{4.1}$$

The collector nodes form the largest part of the cluster and are responsible for collecting and sending data about their neighbors to the leadership. For an intrusion detection system to work efficiently, we assume that the collector nodes should be connected to at least $\tau + 1$ leaders. This restriction is required so that data collected by collector node should reach the maximum number of leaders. The leader nodes are the decision making nodes. The leader nodes analyze data received from collector nodes of the cluster and make their own decisions locally about the malicious nodes. These local decisions are later shared and compared within the leadership of the cluster. The leadership must consist of at least $2\tau + 1$ leaders. If the number of leaders in the leadership decreases below $2\tau + 1$, then the cluster can not exist and all the nodes have to re-organize among themselves in new cluster with new leadership. The same parameter based decision of at least $\tau + 1$ leaders is enough to consider a node as malicious. A CH is selected from among the leader nodes and has the additional responsibility of managing the cluster. The CH maintains the list of nodes in the cluster and their roles. It generates and broadcasts a new cluster

view by adding new nodes and removing routes to those nodes that are malicious or no longer available in the cluster.

The topology of VANETs is highly dynamic which changes frequently due to random joining and leaving of vehicles. Therefore, the security mechanism for such networks should be self-organized and should adjust to dynamic changes in the network by itself. In order to make the proposed model adaptive to such dynamic changes, the concept of time period is introduced. Three types of time periods are introduced in order to help the intrusion detection components synchronization with one other: Normal Time (NT), Update Round (UR) and Data Collection Time (DCT).

During NT period, changes in cluster composition are not considered. Any possible changes in cluster composition are considered during the next NT period. Local decisions related to malicious nodes are taken during this period. Each NT period consists of multiple DCT periods. During DCT period, the collector nodes send the summaries of collected data to cluster leadership. DCT period is assumed to occur n times in each NT period and its time period can be calculated as

$$DCT = \frac{NT}{n} \tag{4.2}$$

At the end of each NT period, the UR period is initiated. During this period, the cluster leadership takes decision related to changes in cluster composition that occurs during last NT period. The decisions regarding malicious nodes as well as joining and leaving nodes are considered during this period. It is during this period that the leaders share their partial decision regarding malicious nodes and remove malicious nodes from the cluster. At the end of each UR period, the CH generates a new cluster view and broadcasts it within the cluster as well as to the nearby clusters.

4.2 Proposed DIDS Scheme

The proposed scheme has following components which are described as follows.

4.2.1 Leadership Formation

The proposed clustering technique uses a distributed Learning Automata based algorithm on the basis of Aggregate Relative Velocity (ARV) and degree of connectivity (δ) of vehicles for finding a set of vehicles that are designated as cluster leadership. This helps in limiting both the number of vehicles as well as number of transmitted messages used in leadership formation. The vehicles in the leadership then take responsibility for CH election and re-election process. The actions of leadership are rewarded or penalized using L_{R-I} model. This process of selection and reselection is done by the automaton deployed

on the vehicles by taking inputs from the stochastic environment.

Each vehicle first detects its neighbors by periodically broadcasting a *hello* message containing attributes such as- node ID, ARV, status, location (latitude and longitude), and velocity, which are necessary for computing ARV. The periodic transmission of *hello* messages takes place after every t seconds. After sensing its neighbors, each node counts its number of neighbors and compute its δ at time instant t. In the first iteration, vehicles that have maximum value of δ form a cluster leadership.

Initially k vehicles are selected as leadership with random probability such that the cardinality of these vehicles is greater than *2N+1*, here N is the predicted number of vehicles. N is determined by counting the neighboring vehicles for each vehicle and then, computing the product of vehicular density and average number of neighbors as follows,

$$N = C_{ave} * D \tag{4.3}$$

$$C_{ave} = \frac{1}{t} | \sum_{i=1}^{t} M_{tot} | \tag{4.4}$$

where, C_{ave} is the average degree of connectivity and M_{tot} is the total number of successful messages received by vehicle i at time instant t and D is average density of vehicles.

The CH is selected from among the leaders in the leadership by Learning Automata-based on its ARV value in an iterative manner. At each stage, the action of selection of a vehicle is rewarded or penalized using L_{R-I} reinforcement scheme depending on degree of connectivity of all the vehicles in the leadership. After each action, the automaton changes its number of possible actions and action probability vector depending upon the cluster leadership. Thus, the proposed scheme selects a superset of weakly connected dominating set as the leadership backbone. This process helps to improve cluster stability with reduced overhead for reclustering.

The automaton uses variance of vehicle's relative velocity over all its neighbors to obtain the value of ARV. A vehicle is chosen as the leader if ARV crosses a dynamic threshold which is the current minimum ARV for the vehicles in the leadership. The ARV metric is preferred over the other parameters because, it reduces the impact of node's mobility on detection accuracy of its malicious behavior in the network. To calculate any node's ARV, its current relative velocity is computed with respect to its neighbors after every t time intervals using Equation 4.10 as follows,

$$V_Y^{rel}(X) = |v_y - v_x| \tag{4.5}$$

where, v_y, and v_x represent the velocity of nodes Y and X. ARV is computed by taking the variance of relative velocity (V_i) of the nodes as V_i-

$$V_i = var(V_Y^{rel}(X_j)) \tag{4.6}$$

So, we have the following,

$$\Omega = \delta_i - V_i \tag{4.7}$$

If the value of Ω reaches a predefined threshold S_T, then a node is selected as IDS Leader (LE) by an automaton; otherwise, it becomes cluster member(CN) for IDS of that of the corresponding cluster. After calculating its Ω each node shares it with other vehicles in the cluster through flooding. After receiving Ω of all the participating nodes, each node calculate the mean of all the values and this resultant value is S_T. In this way each node will know Ω of every other node in the cluster and can calculate its own S_T.

$$S_T = \frac{1}{N}\sum_{k=1}^{N}\Omega_k \tag{4.8}$$

where, N is taken as the total number of nodes in the cluster and Ω_k is the resultant value of k^{th} node. Thus, the proposed algorithm iteratively constructs the cluster leadership and also updates its action probability vectors until it finds an optimal cluster leadership which depends on the value of ARV. The sequence of operations for leadership formation are shown in Algorithm 8 (lines 1-17).

In the next step, all the selected leaders disseminate a *selectCH* message in the cluster (line 18). After comparing the ARV with all the leaders, a leader with minimum ARV is declared as the CH (lines 19-22). If the message sending node j is in leadership, then node i compares its ARV with j (lines 21-22). If ARV of i is smaller than that of j's, then i wins and is declared as the CH; otherwise, it remains as a leader. During this process, each action of the automaton gets a reward or penalty from the environment which increases or decreases the probability of taking next action by $\xi \in (0, 1)$. Also, the number of rewards and penalties n^{rew}, n^{pen} is re-computed till the maximum number of actions K have not been taken by the automaton.

The whole process of leadership formation and CH selection must complete before the expiry of clusteringTime. After the completion of clusteringTime, the CH announces its ID within the cluster (lines 31-35). The leader nodes also announce themselves as leaders to other cluster members and update their cluster-Head ID (CID) to the ID of the new CH (lines 38 to 45). CNs update their CID and also construct new leadership list (Leadership_List).

Algorithm 8 *LeadershipFormation*

Inputs: Neighbor-List, Velocity of node, Location at time t, Node ID, clusteringTime, number of succesful msgs, $\phi, \xi \in (0,1), n^{rew}, n^{pen}, K$
Output: Node status
Assumption: Repeated hello messages

for $(n^{act} \neq K)$ **do**
  for $(clusteringTimehasnotexpired)$ **do**
    for $(allvehicles)$ **do**
      δ_i=$\sum$ Neighbor-List (i) // sum of all nodes in neighbor-list of node i at time t
      $M_t = M_{t-1} + m_i$
    end for
    Select k vehicles for leadership with probability ϕ
    $C_{ave} = \frac{1}{t}|\sum_1^t M_{tot}|$
    $N = C_{ave} * D$
    Change the number of actions and probability vector of Learning Automata as
    $V_i^{rel}(j) = |v_i - v_j|$ // relative velocity of i w.r.t. j at time t
    $V_i = var(V_i^{rel}(X_j))$ // ARV of node i at time t
    $\Omega = \delta_i - V_i$
    Broadcast (Ω_i)
    Compute S_T
    if $(\Omega \geq S_T)$ **then**
      Action of automaton is rewarded as
      $\phi = \phi + \xi$
      $n^{rew} = n^{rew} + 1$
      set STATUS(i)=LE
    else
      $\phi = \phi - \xi$
      $n^{pen} = n^{pen} + 1$
      set STATUS(i) = CN
    end if
    Compute the ratio of $\frac{n^{rew}}{n^{pen}}$
    Disseminate (selectCH$_i$, leadership_knowledge$_i$)
    On receiving *selectCH$_j$* at node i
    if $(STATUS(j) = LE)$ **then**
      if $(V_i < V_j)$ **then**
        set STATUS(i) = CH
      end if
    end if
  end for
  if $(STATUS(i) = CH)$ **then**
    Broadcast (CH_{id})
  end if
  if $(STATUS(i) = LE)$ **then**
    Broadcast (LE_{id})
    $CID = CH_{id}$
  end if
  if $(STATUS(i) = CN)$ **then**
    $CID = CH_{id}$
    $Leader\_List = Leader\_List$ U LE_{ID}
  end if
  Update the ratio of number of rewards and penalties for the next round
end for

If any leader moves out of the cluster then the changes in cluster composition are considered only during UR period. During UR period the CH checks the number of leaders present in the cluster. If the number of leaders in the cluster are less than $2\tau + 1$, the leadership is dissolved and new leadership is elected again using Algorithm-8. However, if the number of leaders are greater than or equal to $2\tau + 1$, than no action is taken by the CH. In this work if some leader leave the cluster and cluster still contains $2\tau + 1$ leader, than the decision of that leader can be taken by other leaders as each collector node is assumed to be connected to atleast $\tau + 1$ leaders.

4.2.2 Cryptographic Mechanism

In this step, the process of how information is secured among the different objects such as vehicles is explained. The collectors use a secure channel to send data to the leaders. These secure channels are built by using a public key based infrastructure (PKI) in the network according to the scheme proposed in [170]. Using this cryptographic technique, the intrusion detection related messages are encrypted and authenticated in both cluster-to-leader and leader-to-leader communication. However, multimedia data sharing in VANETs is more challenging as the devices/vehicles are resource constrained due to their inherent characteristics like short wireless communication range, rapid movements of the vehicles and topological changes and frequent disconnections. So traditional techniques must be modified for designing effective mechanisms for transmission of multimedia data in VCC.

CA is assumed to be a trusted third party in the proposed scheme which possesses information related to all the vehicles in the network. A network can be constrained by a geographical area such as city, town or district and consists of a large number of small sub networks. In the proposed model, the cryptographic mechanism works as follows:

At the end of each UR, the CH sends a request to CA to generate the set of keys for the cluster. After validating the request by reviewing the credentials of all the nodes of the cluster, the CA generates four types of keys for the cluster: n private keys $(PV_1, PV_2 \ldots\ldots, PV_n)$, n public keys $(PB_1 \ldots\ldots, PB_n)$, Leadership Secret Key (SK_l), and Leadership Public Key (PK_l). The public and private key pair (PV_i/PB_i) is distributed among all the nodes in the cluster and can be used in both IDS and non-IDS communications. The PK_l is known to all nodes in the cluster and is used by collectors to authenticate messages from leadership. The SK_l should be known only to leaders and can be used by leaders to authenticate other leaders. In the proposed model, entity authentication is based on digital certificates. CA generates a certificate $(cert_i)$ for every node in the network which includes node's public key.

For generation of digital signatures, the key HMAC algorithm is used. The HMAC algorithm generates a fixed size unique hash code using a secret key. It is important

to verify that no attacker has tampered with the message before processing it. Since there are again some pitfalls in doing this using ad-hoc verification or simple hashing, message authentication code based on HMAC has been used. By concatenating the key and the message, and hashing them together, the probability of finding what it is hash of, given a particular cryptographic hash are thus reduced appreciably. This also makes it relatively difficult to find the key. HMAC also reduces the computational overhead of encrypting and decrypting thereby making it favorable in vehicular environment. In collector-leadership communication, collectors use their private key to generate Digital Signatures ($dSign_i$). However, to authenticate messages from leader nodes, the collector use PK_l. For communication within the leadership, the SK_l which is known only to the leader nodes of the cluster is used so as to stop any malicious node from pretending as the leader node. Therefore, using shared SK_l, a leader node can prove itself as a valid cluster leader. In every new UR, these set of keys are updated by the CA.

To achieve efficient transmission of multimedia data in the proposed scheme, a cloud-based network is split into distinct regions, and the large sized multimedia data is divided into small parts so as to reduce the overhead during various authentication mechanisms. Each vehicle in a cluster is provided access to some portion of this data by usage of Secure Hash Algorithm1 (SHA1) hashing technique on this data. The vehicle records the information about this multimedia data availability in Distributed Hash Tables. Thus, the acquisition of information by vehicles even in case of large sized multimedia data can be made more effective by combination of cloud and vehicular networks. Therefore, a good scalability is achieved by applying this method because message and memory overhead are decreased in comparison with the other protocols.

4.2.3 Intrusion Detection Mechanism

The functioning of the proposed intrusion detection scheme is divided into two separate processes as local and global detection. The process of intrusion detection begins with local detection of malicious nodes by the leaders. During local detection, each leader creates a list of suspicious nodes. Each leader then shares its list of suspicious nodes with other leaders to create a global detection mechanism. Both types of detection are illustrated in the following schemes.

4.2.3.1 Local Detection

The local detection takes place during DCT period. A DCT period is a part of NT period which is repeated multiple times before NT period ends. The local detection mechanism is presented in Algorithm 9 which is an extension from the algorithm given in [171]. The collectors collect data related to their neighbors by overhearing the network traffic.

When the DCT period begins, each automaton residing at collector nodes generates the summary of the collected data represented as C_data_i and sends it to the leadership using Dissemination function as explained in Algorithm 9 (lines 2 to 5). For this action, environment provides a feedback in terms of a reward or a penalty. The automaton at leader validates the message by checking if the message is not old (line 9 by comparing the values of UR and DCT counters received in the message with the current values).

The automaton also validates the received message by verifying the digital signatures using HMAC algorithm. The automaton learns using the parameter ∇ and takes adaptive decisions for intrusion detection in local domain. If the message is valid, the leader analyzes the data and constructs its suspect list $suspect\_list_i$,(lines 13 to 15). The leader analyzes data for two common type of attacks: blackhole and flooding attack. Other types of attack can also be handled similarly. The procedure to analyze data and setting of thresholds to estimate abnormal behavior of a node has been explained in earlier section 4.1. During this time period, any changes in the cluster composition like node's joining and leaving requests are not considered. After the completion of current DCT period, all nodes update their DCT_i counters (line 20).

Algorithm 9 *DCTPeriod* at node i

Inputs: p, Status, UR counter, DCT counter, $data_i$, $\nabla \in (0, 1)$
Output: $suspect\_list_i$
Assumption: Repeated hello messages

1: Set timeout $\leftarrow$ p/3 // p is time period between two DCT's
2: **if** $(STATUS(i) = CN)$ **then**
3: $C\_data_i = GenerateSummary(data_i)$
4: $C\_msg_i = (C\_data_i, UR_i, DCT_i, dSign_i, cert_i)$
5: Disseminate $(C\_msg_i, Leadership\_knowledge_i)$
6: **end if**
7: **if** $(STATUS(i) = LE)$ **then**
8: On receiving (C_msg_j) // j is some collector node
9: **if** $((C\_msg_j.UR = UR_i)$ AND $(C\_msg_j.DCT = DCT_i))$ **then**
10: Authenticate C_msg_j using HMAC
11: Performs learning using ∇
12: $monitor\_buf_i = monitor\_buf_i \bigcup C\_data_j$
13: **if** $((monitor\_buf_i.count \geq \tau + 1)$ AND $(Current\_Time - DCT_i * p \geq timeout))$ **then**
14: $Analyze_i = AnalyzeData(C\_data_j)$
15: **if** $(Analyze_i) = Is\_Suspect$ **then**
16: $suspect\_list_i = suspect\_list_i \bigcup id_n$
17: **end if**
18: **end if**
19: **end if**
20: **end if**
21: $DCT_i = DCT_i + 1$

When all DCT periods are over, all leaders synchronize with one another by sending *sync* messages before the end of current NT period. Leadership synchronization is necessary, so that all the nodes should start the UR round at the same time and participate in global detection mechanism together.

After the completion of local detection mechanism, the global detection mechanism is started. The global detection mechanism consists of UR period, in which the local decisions of the leaders share their information within the leadership and list of final malicious nodes is created.

4.2.3.2 Global Detection Mechanism

The global detection of malicious nodes takes place during the UR period. The detection mechanism is adapted from the algorithm given in [171]. All the changes that have occurred in the previous NT period are implemented in the cluster. During this phase, all the leaders share their list of suspected nodes with one other using Dissemination function (lines 1 to 3). If a node is considered suspicious by $\tau + 1$ or more leaders then, an alarm is generated about that node and it is added to the list of malicious nodes mal_list_i, (lines 6 to 9). Each leader then sends its mal_list_i to the CH (line 10). The CH then generates the final list $final\_mal\_list$ and broadcasts it within the cluster as well as to the neighboring clusters (lines 20 to 23).

After the identification of malicious nodes, an automaton updates the learning rate after analysis of the generated malicious nodes. The value of learning rate is increased or decreased by constant value $\xi \in 0, 1$ as a ratio of number of rewards and penalties given as

$$\xi = \frac{n^{rew}}{n^{pen}} \tag{4.9}$$

This process ir repeated until the value of ξ crosses a predefined threshold value th_r.

The leadership considers the joining and leaving requests of the previous NT period. If the requesting node is not malicious, its request is considered and new cluster view is generated accordingly. Then, the CH generates and broadcasts a new cluster view by excluding the malicious, inactive and leaving nodes from the cluster (lines 24 to 26). At the end, each node updates its UR counter for next UR period and DCT counter is reset to 1 (lines 32 and 33).

Before the completion of UR period, the CH sends the request to update the keys of the cluster. The CH also sends the mal_list_c and new_view to the CA (lines 27 to 28). The CA then revokes the certificate of the malicious nodes and broadcasts this information to all CHs in the network. A node communicates with the other nodes only if it has a key

issued from CA. So, CA also generates the new pair of keys according to the new cluster view.

If the UR period has not finished, and the CH has already left the cluster or does not generates new cluster view within time period *viewTimer* then, leadership elects a new CH immediately using Algorithm 8 (lines 25 to 26) for controlling the abnormal behavior of the Cluster-Head.

4.2.3.3 Initial Update Round (UR_0)

Initially, in the update round period, the system is formulated and intrusion detection is operated to capture various events. During this phase, the formation of clusters takes place according to the Algorithm 8. After the creation of clusters, the leadership is formed and CH is selected. During this phase, the public, private, leadership's public keys and its secret key are generated and distributed among the nodes in the cluster with the assistance of CA. Before the completion of UR_0, the DCT and UR counters are initialized in the cluster. After the completion of UR_0, first NT period starts which consists of multiple DCT periods followed by UR period.

4.2.3.4 Data Dissemination

For collector-to-leader and leader-to-leader communication, a separate dissemination function is presented as proposed in [172]. All intrusion detection messages are distributed using this function.

In this algorithm, a node i sends different messages to those leaders that matches its knowledge about the leadership (lines 2 to 4). On receiving a message, nodes check if the received message is already presented in their buffer. The message is accepted if it is not present in the buffer, and then it sends a message to its own leadership and also delivers it locally (lines 6 to 9). Then the CHs further send this message to cloud infrastructure after discrete amount of time.

4.2.3.5 Node Joining and Leaving

The algorithm for joining and leaving nodes is motivated from [171]. While joining, a node broadcasts its Joining Request (Req_Join) by including its identity and digital certificate. If the receiving node is a leader then, after validating the request, the new node is added to the list of inactive nodes and the request is further disseminated to the nodes up in the hierarchy (lines 2 to 6). If the receiving node is a collector then, it disseminates the received request to its leader (lines 8 to 10). The leader then sends the new cluster view, if the new node is not malicious. Upon receiving new view, the nodes update their status as decided by the leader of the group (lines 12 to 19). Moreover, the new node also

Algorithm 10 *URPeriod*

Inputs: $suspect\_list_i$, *Status, UR counter, DCT counter,* $\nabla \in (0,1)$, *thr*
Outputs: $final_mal_list, new_view, PV_n, PB_n, SK_l, PK_l$
Assumption: Repeated hello messages

if $(STATUS(i) = LE)$ **then**
  $leader\_msg_i \leftarrow (suspect\_list_i, UR_i, dSign_i, cert_i)$
  Disseminate $(leader\_msg_i, Leadership\_knowledge_i)$
  On receiving $leader\_msg_j$
  if $(leader\_msg_j.UR = UR_i)$ **then**
    Authenticate $leader\_msg_j$ using HMAC
    $suspect\_buf_i \leftarrow suspect\_buf_i \bigcup suspect\_list_j$
    if $(suspect\_buf_i.count \geq \tau + 1)$ **then**
      $mal\_list_i \leftarrow identify\_suspect(suspect\_buf_i, UR_i)$
      send mal_list_i to cluster-head
    end if
  end if
  Start *viewTimer*
end if
if $(\frac{n^{rew}}{n^{pen}} > thr)$ **then**
  $\nabla \leftarrow \nabla + \xi$
else
  $\nabla \leftarrow \nabla - \xi$
end if
if $(STATUS(c) = CH)$ **then**
  On receiving mal_list_j
  $final\_mal\_list = final\_mal\_list \bigcup mal\_list_j$
  Broadcast $(final\_mal\_list)$
  $new\_view = Generate\_view\ (inactive\_nodes, leaving\_nodes, new\_nodes, mal\_nodes, UR_c)$
  $CH\_msg_c = (new\_view, UR_c, id_c, dSign_c, cert_c)$
  Broadcast (CH_msg_c)
  $Key\_req\ (new\_view, final\_mal\_list, id_c, dSign_c, cert_c)$
  send Key_req to CA // CA is a cloud based authority
end if
When viewTimer expires at node i
Disseminate $(selectCH_i, leadership\_knowledge_i)$
$UR = UR + 1$
$DCT = 1$

Algorithm 11 Dissemination of information

1: Disseminate $(message_i, Leadership\_knowledge_i)$
2: **for** $(l_j \in Leadership\_knowledge_i)$ **do**
3: send $(DISSEMINATE, message_i)tol_j$
4: **end for**
5: $Onreceiving(DISSEMINATE, message_i)atnodei$
6: **if** $((DISSEMINATE, message_i) \neq Received\_Buffer_i)$ **then**
7: $Received\_Buffer_i \leftarrow Received\_Buffer_i \bigcup (DISSEMINATE, message_i)$
8: Disseminate $(message_i, Leadership\_knowledge_i)$
9: $deliverDisseminate(message_i)$
10: **end if**

Table 4.1: Simulation parameters used in DIDS scheme.

Parameters	Values
Simulation Time	6000 seconds
DCT period	60 seconds
UR period	300 seconds
Number of nodes	100
Average Cluster Duration	19 seconds
Transmission range	250m
MAC	802.11p
Vehicle entry rate	10 v/min, 24 v/min
Maximum Lane Speed	25m/s
Queue Length	10 packets
Routing protocol	AODV
Traffic Type	CBR (UDP)
No. of malicious nodes when τ respected	0, 10, 20, 30, 40
No. of malicious nodes when τ not respected	20, 30, 40
Packet rate	3, 6, 11 packets/second

update routes to the leader nodes as well as to its neighbors as described in (lines 21 to 22). Finally, all the nodes update their UR counter according to the received view (line 23). The procedure for node leaving the cluster is described in lines 24 to 29.

While leaving, a node must send a Leaving Request (*Req_Leave*) to the cluster leader. On receiving the Req_Leave_i, the leader node checks whether the requesting node is malicious (line 27). If the leaving node is not malicious then, it is added to the list, *leaving_nodes* (line 28). If a node joins or leaves the cluster without informing the leadership then, it is considered as a malicious node.

Algorithm 12 Node Joining or Leaving algorithm

Inputs : $NodesID, mal\_list_i, view_c$
Outputs : $inactive\_list_i, leaving\_nodes_i$
Assumption: Repeated hello messages

Broadcast $(Req\_Join_i, id_i, dSign_i, cret_i)$
On receiving $(Req\_Join_j, id_j, cret_j) atnodei$
Authenticate $Req\_Join_j usingHMAC$
if $(STATUS(i) = LE) AND (j \neq mal\_list_i)$ **then**
  $inactive\_list_i \leftarrow inactive\_list_i \bigcup Req\_Join_j$
  Disseminate $(Req\_Join_j, leadership\_knowledge_i)$
else
  if $(STATUS(i) = CN) AND (j \neq mal\_list_i)$ **then**
    Disseminate $(Req\_Join_j, leadership\_knowledge_i)$
  end if
end if
On receiving $view_c$ at node i
Authenticate $view_c$ using HMAC
if $i \in view_c.Leader$ **then**
  set STATUS (i) = LE
else
  if $(i \in view_c.Collector)$ **then**
    $setSTATUS(i) = CN$
  end if
end if
Update routes to *Leadership* according to $view_c$
Update neighbor list according to $view_c$
$UR_i \leftarrow view_c.UR + 1$
If node i wants to leave the cluster
Disseminate $(Req\_Leave_i, leadership\_knowledge_i)$
On receiving $(Req\_Leave_j, id_j, dSign_j, cret_j)$ at node i
if $(STATUS(i) = LE) AND (j \neq mal\_list_i)$ **then**
  $leaving\_nodes_i \leftarrow leaving\_nodes_i \bigcup Req\_Leave_j$
end if

4.3 Simulation Settings for DIDS

In this section, the performance of the proposed scheme is evaluated with respect to various performance evaluation metrics with presence of some of the malicious nodes in dynamic environment.

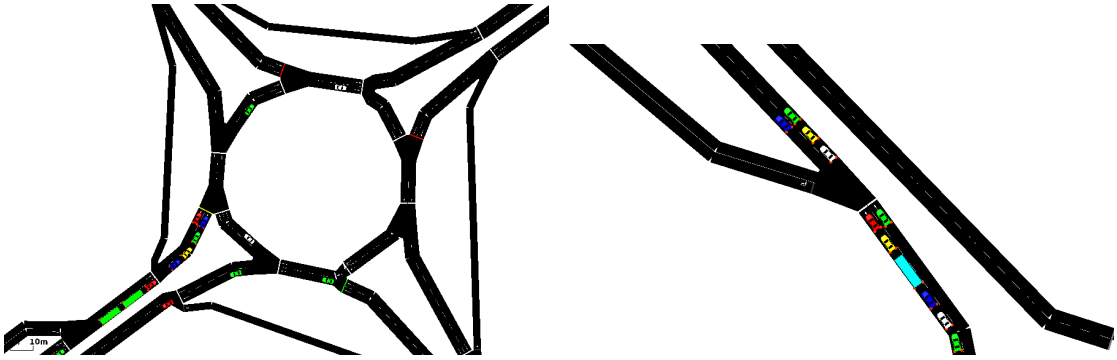
4.3.1 Performance evaluation metrics

To test the performance of the proposed scheme, four metrics i.e. message loss rate, disseminated messages, false positive rate and detection rate are used. These metrics are defined as follows.

- Message Loss Rate- It is the ratio of number of messages dropped to the total number of messages received. High value of this parameter signifies that the attack is highly successful and has also affected the various intrusion detection components.
- Disseminated Messages- It is defined as the percentage of the total successful message transmitted to the total number of disseminate messages. This parameter signifies the impact of attack on intrusion detection components. Low value of this parameter signifies that the intrusion detection components are not working properly.
- Detection Rate- It is defined as the ratio of number of intrusion successfully detected to the total number of intrusion occurred. High value of this parameter indicates a highly efficient intrusion detection system.
- False Positive Rate- It is defined as the ratio of false intrusion observed to the total number of intrusions. False positive occurs when, the intrusion detection generates an alarm without the occurence of the same. This parameter signifies the detection accuracy of the intrusion detection in the network with high value indicates lesser accuracy of the designed intrusion detection system.

4.3.2 Simulation environment

The proposed scheme is implemented using Network Simulator 2 (NS-2). In order to generate realistic vehicle mobility pattern, the microscopic traffic simulator Simulation of Urban Mobility (SUMO) [148] is used. The mobility pattern is generated on the map of Chandigarh city. The network consist of 100 nodes which are grouped in one cluster. In order to provoke dynamic changes, vehicles can join and leave the network from different entry and exit points at different rates. The portion of Chandigarh city in Figure 4.4 (a) and (b) shows the movement of vehicles on the map. Several rounds of simulation tests



(a) (b)

Figure 4.3: Movement of vehicles in Chandigarh city for DIDS (a) Movement of vehicles at round about with traffic lights (b) Vehicles at entry point of the city.

are performed in the proposed scheme. Tests are also performed to test the threshold value τ. To generate malicious behavior, black hole and flooding attacks are performed on the network. In black hole attack, the malicious node drops all the packets that are routed through it and also forges its identity. For flooding attack, the malicious node sends data at the rate of 6 packets per second or more, while non-malicious nodes send data at the rate of 3 packets per second.

To obtain simulation results, two tests are performed, Test-1 and Test-2. Test-1 is conducted with two different scenarios using threshold τ, i.e., number of malicious nodes never crosses number of leaders in the cluster. In Scenario-1, vehicles enter at a rate of 10 vehicles/min with low vehicle density scenario. In Scenario-2, vehicles enter at a rate of 24 vehicles/min with a high vehicle density scenario. In both the scenarios, vehicles leave the network when they reach their destinations. Vehicle destinations are selected dynamically at run time using DFROUTER function present in SUMO. The routes of the vehicles, their entry and destination points are defined using a function DFROUTER present in SUMO [148]. In order to define time at which vehicles enter or exit from the network, the DFROUTER takes three inputs, starting time, exit time and offset. Starting at the designated starting time the DFROUTER inserts vehicles in the network randomly after each offset. The vehicles are also selected randomly by DFROUTER from the given set of vehicles. The exit time is the time at which vehicles reach their destination and leave the network. Both these scenarios are evaluated by varying the number of malicious nodes in the cluster and are presented in Table 1. Test-2 is also conducted with two scenarios but when threshold value τ was not considered. In Test-2, two scenarios of 10 and 24 vehicles/min are considered but without considering τ limit. The number of malicious nodes are increased to 20%, 30% and 40% but, the size of leader is kept at 10%. The simulation parameters are summarized in Table 1.

4.4 Results and Discussions

The results of Test-1 are presented in Figures 4.5(a)-(b) and Figures 4.6(a)-(b). As observed from Figure 4.5(a), the number of messages dropped by nodes in the cluster increases with an increase in malicious activities in the network. When the number of malicious nodes are 40%, the message loss rate increases to 35% in Scenario-1 and to 39.5% in Scenario-2. This occurs mainly due to the combined effect of high node speed and attacks launched by the malicious nodes. However, the high message loss is countered by the high dissemination rate in the cluster, which shows a highly effective distributive mechanism in the proposed scheme. The simulation results closely mirror the performance achieved through analytical observations which prove the effectiveness of our scheme.

As observed in Figure 4.5(b), the dissemination rate decreases with an increase in malicious nodes but, it remains above 90% for both scenarios even in the presence of high malicious activities. The dissemination rate remains unaffected from malicious activities because of the use of connectivity for selecting the leadership. Only the nodes with high connectivity are selected in the leadership. Therefore, a message sent using dissemination function is able to reach at least $\tau + 1$ leaders. Thus, there is an improvement in the dissemination rate. Moreover, the simulation results have comparable values with their corresponding analytical values.

As observed in Figure 4.6(a), the false positive rate increases with an increase in malicious nodes in the network in both the scenarios. In addition, false positive also occurs due to high mobility of nodes in the network. The use of mobility information in forming the leadership helps in reducing the false positive rate as observed in Figure 4.7. As a result, the false positive rate remains below 10% in both the scenarios. The false positive rate in Scenario-1 is comparatively greater than Scenario-2. This is due to the fact that number of connections between nodes are more in Scenario-2 than in Scenario-1. Hence, due to fewer connection, a leader node has lesser data and hence, comparatively higher false positive rate. The analytical values are also similar to the simulated results.

As observed in Figure 4.6(b), the detection rate decrease with an increase in malicious nodes in the cluster in both the scenarios for simulated and analytical values. This is due to the fact that with an increase in malicious nodes, large number of disseminated messages are dropped as observed in the measured disseminated messages but, still due to the distributive and cooperative architecture of the intrusion detection system, the detection rate remains close to 90% even when malicious nodes have reached to 40%. However, the detection rate in Scenario-2 is comparatively greater than that of Scenario-1. This is due to the fact that node density in Scenario-2 is relatively higher with more number of connections. Therefore, large number of nodes are able to detect the malicious node and report it to the leadership. Hence, the chances of false positives are reduced with high

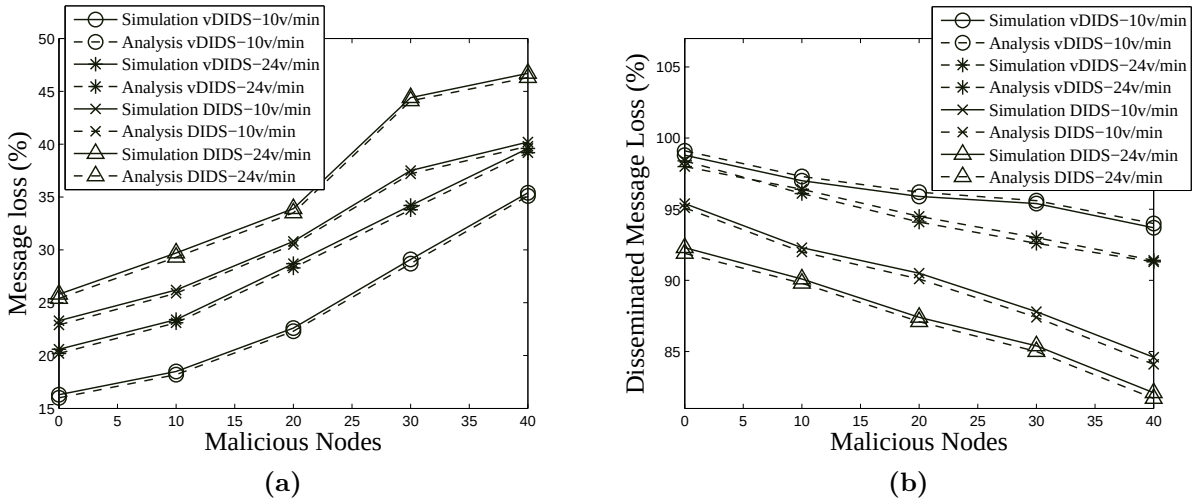


Figure 4.4: (a) Message Loss Rate when τ is considered. (b) Disseminated Messages when τ is considered in DIDS protocol.

detection rate.

The results of Test-2 are presented from Figures 4.7(a) to 4.7(d). In Test-2, the tests were performed for evaluating the efficiency of the proposed scheme when value of τ is surpassed. In this test, high message loss is observed as compared to Test-1. As observed in Figure 4.7(a), the message loss increases to 44% in Scenario-1 and to 47% in Scenario-2. This occurs mainly due to the presence of large number of malicious nodes and lesser leader nodes for both simulated and analytical values. However, the dissemination rate remains close to 85% in both the scenarios as described in Figures 4.7(b), and 4.7(c). This is mainly due to the fact that due to high connectivity in the leadership, the disseminated messages reach almost every leader resulted in attaining high rate of disseminated messages.

As observed in Figure 4.7(d), the malicious nodes also affect the false positive rate in both the scenarios. The false positive rate increases due to an increase in the message loss rate which occurs due to the presence of malicious nodes but still, the false positive rate remains below 15% in both the scenarios. This is due to the fact that the false positive rate due to mobility of nodes is reduced by the leadership formation mechanism. The false positive rate in Scenario-1 is comparatively greater than Scenario-2. This is due to the fact that number of connections between nodes are more in Scenario-2 than in Scenario-1. Hence, due to fewer connections, a leader node analyzes lesser data resulting in higher false positive rate.

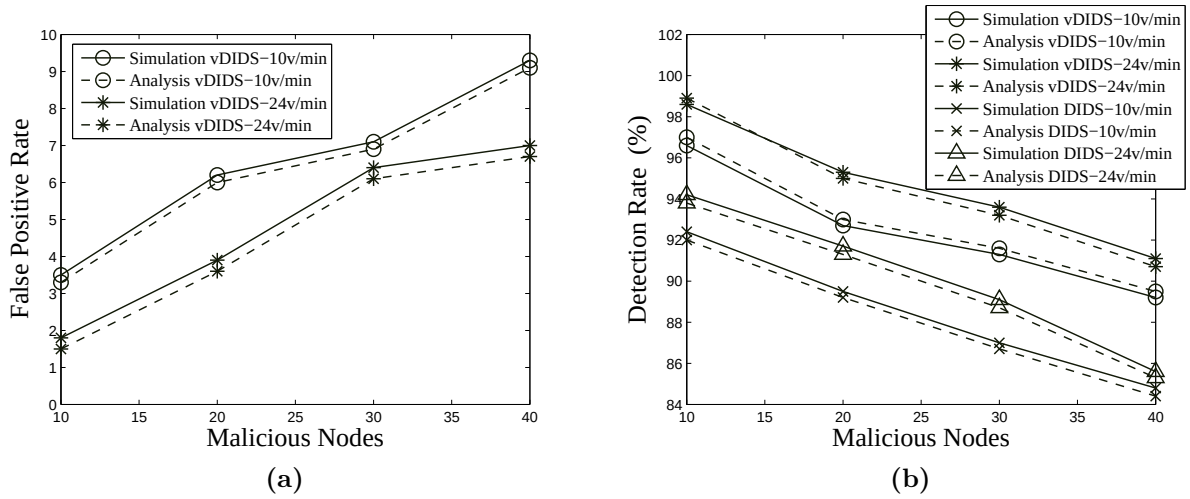


Figure 4.5: (a) False Positive Rate when τ is considered. (b) Detection Rate when τ is considered in DIDS protocol.

An increase in malicious nodes also affects the detection rate of the intrusion detection system but still, the proposed scheme detects 93% to 85% malicious activities in the network as observed in Figure 4.7(d). This is due to the fact that, even in the presence of 40% malicious nodes, the dissemination rate of approximately 85% is obtained. The higher dissemination rate helps the intrusion detection system in taking correct decisions. As a result, the proposed scheme detects malicious activities in almost 85% of the cases. However, the detection rate in Scenario-2 is comparatively greater than that of Scenario-1. This is due to the fact that node density in Scenario-2 is relatively higher having more connections. Therefore, large number of nodes detect the malicious node and report it to their leader. Hence, the chances of false positives are reduced resulting in higher detection rate. The analytical values obtained for above graphs are similar to the simulated results. Thus the proposed scheme gives comparable performance in terms of both simulated and analytical values.

4.5 Conclusion

In this chapter, a distributed IDS model for VANETs is proposed. The proposed model is deployed on clusters and consist of number of distributed algorithms that make the working of IDS mechanism dynamic and distributive. As observed in simulation results the clustering approach used in this work help in improving the performance of the IDS mechanism. Through the clustering process only those nodes are selected for the leadership that can perform their operations for longer duration without leaving the cluster. Unlike other approaches the intrusion detection task is not dependent on a single node, a group of nodes that form a leadership are responsible for IDS operations. The simulation

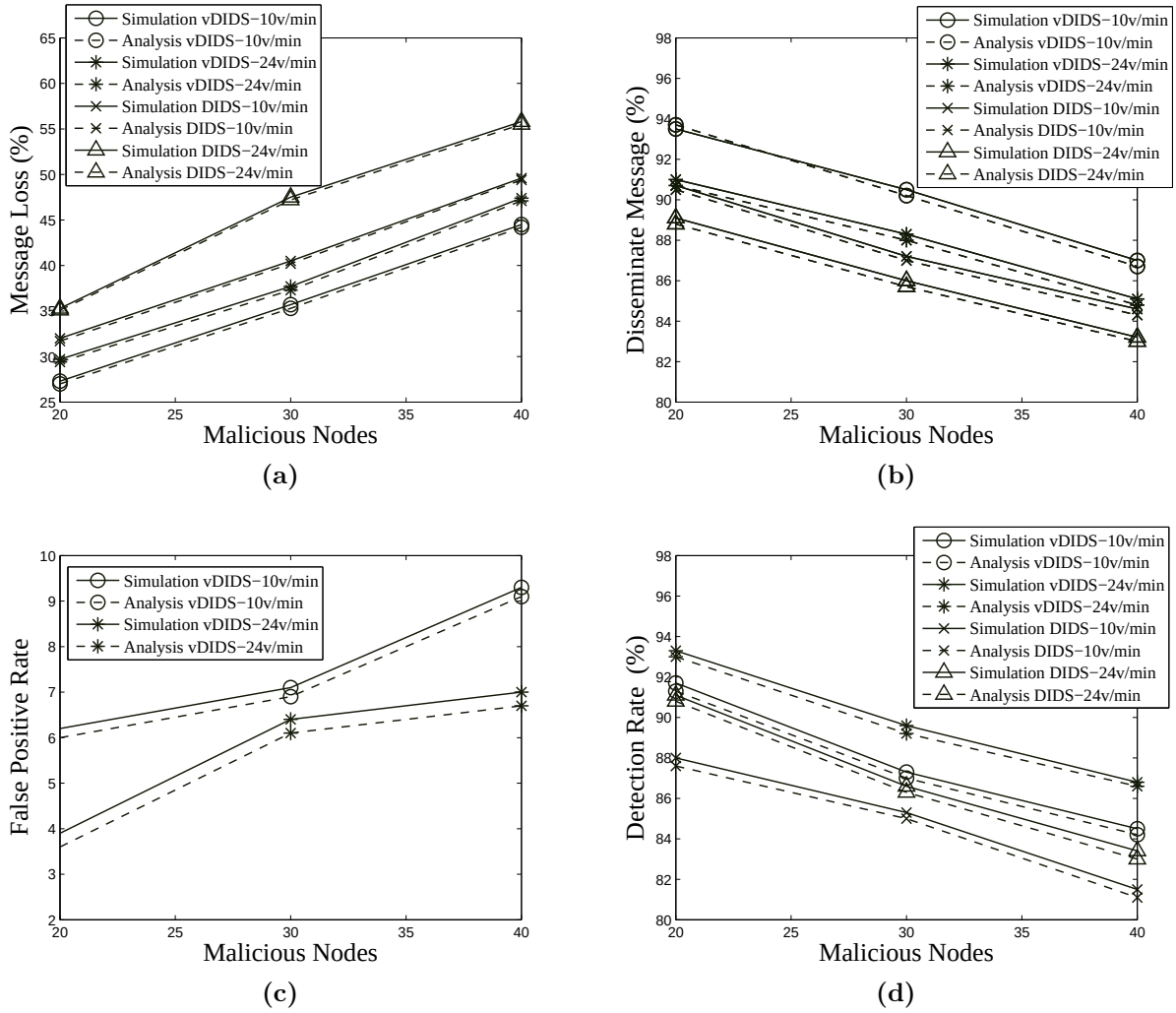


Figure 4.6: (a) Message Loss Rate when τ is not considered. (b) Disseminated Messages when τ is not considered. (c) False Positive Rate when τ is not considered. (d) Detection Rate when τ is not considered in DIDS protocol.

results show that due to the presence of leadership the IDS mechanism continues to work even in the presence of large number of malicious nodes in the network. The proposed mechanism is also adaptable to the dynamic changes in the topology of VANET like nodes joining and leaving the network. Moreover, the proposed IDS model also consider intrusion and fault occurrences in its own components. The simulation tests show that in some scenarios, when the faulty and malicious nodes crosses the τ limit, the IDS continues to operate correctly.

Chapter 5

Secure Clustering

From the past few years, there has been tremendous growth with respect to the usage of Internet-enabled vehicular devices for computing and storage. Computing as well as communication capabilities of vehicles have resulted in facilities to provide the development of a large number of applications for the end users. However, in vehicular environment, security is one of the major concerns as devices within vehicles may communicate with one another using a varieties of protocols which are susceptible to various types of attacks. To address these issues, we propose a novel secure clustering for efficient data dissemination between different devices. A new trust metric based on dynamically varying transmission characteristics of vehicles is defined for trust computation among the different devices which is evaluated both at local, and global levels. This trust metric is used to establish the current security level of vehicles and is the key parameter for creating secure clusters. Algorithms for secure clustering and trust establishment are designed in the proposed scheme and performance of the proposed scheme is evaluated with respect to different evaluation metrics in various network scenarios.[1]

5.1 Secure Clustering Approach

VANET is a seamless service to the onboard passengers such as-safety alerts, ease of driving, and entertainment. As communication among the vehicles is mainly done using wireless communication so there is a need for designing an efficient secure communication among different components in this environment. VANETs provide base for a large number of applications in different domains related to the transportation environment such as

[1]The content of this chapter has been taken from :

- Rasmeet S. Bali, Neeraj Kumar, "Secure Clustering for Efficient Data Dissemination in Vehicular Cyber-Physical Systems", Future Generation Computer Systems, Elsevier, Volume 56, pages 476–492, 2016.

vehicular security and brake warning, an enhanced navigation, and road traffic management [[172],[173]]. This necessitates the requirement for inter-vehicular communication to create a reliable transportation system that allows secure broadcasting and collection of various types of information related to entertainment, and safety.

However, the design and development of various solutions for secure communication in VANETs have various challenges keeping in view the high mobility and varying density of the vehicles on the road [170]. So, it is essential to ensure that the critical information should not be captured by different attackers in the network. Moreover, the system should be able to establish the identity of the drivers and at the same time must protect the privacy of drivers, and passengers. The IEEE 1609.2 standard is focused on available security services for vehicles to authenticate other neighboring vehicles using certificates managed by a centralized CA within a PKI. However, deployment of PKI is difficult in VANETs as these are designed for networks that have a centralized controller, and ubiquitous connectivity.

VANETs on the other hand are characterized by relatively high velocity with constrained mobility pattern which result in frequent changes in network topology [152]. Moreover, deploying PKI in VANET's with an architecture designed for centralized networks has a large number of challenges pertaining to scalability as CAs cannot be made accessible to all vehicles due to their high mobility [174]. Also, having limited CA for the whole network may create a single point of failure while multiple CAs may enhance the overhead of the network. Some existing solutions [[153], [175]] have considered RSU as CA. However, their density need to be sufficiently high so that a vehicle need not have to wait a long period outside the RSU transmission range [153].

To resolve the above challenges and issues of VANETs, Vehicular Cloud Computing has emerged as one of the leading technologies for next generation vehicular networks[175]. VCC is a hybrid technology that provides various solutions for road and traffic management by providing quick response using instant decisions by accessing vehicles resources such as-storage, communication, and computation. VCC takes advantage of cloud-based computing resources for enhancing applications for driver convenience. The main objective of VCC is to provide fast computational services at a comparatively lower cost thus enhancing the applicability of VANETs for applications such as reducing traffic congestion, and accidents on the road.

VCC provides an integration of Wireless Networks, ITS and Cloud Computing for better road safety to achieve secure modern ITS [170]. In modern ITS, the end users can be provided a large number of facilities for their convenience through VCC. By leveraging the communication, storage and computing resources available in the vehicles, the cloud-based infrastructure can result in an efficient resource management at various levels in this environment [[176], [177]]. The underutilized vehicular resources including com-

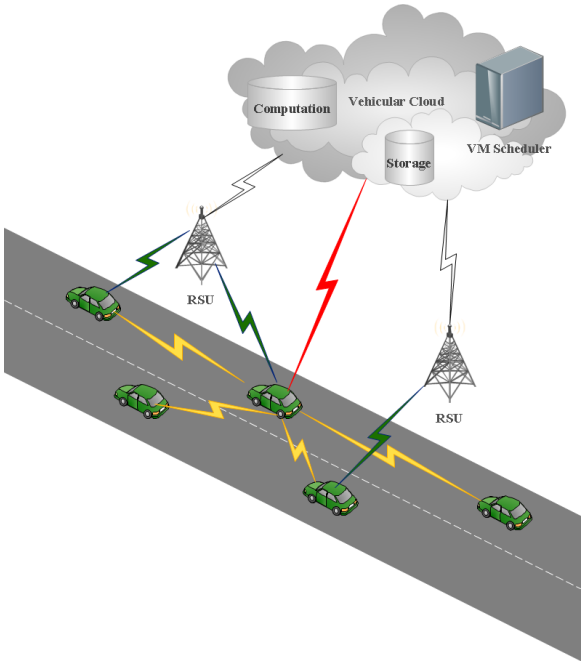


Figure 5.1: Vehicular cloud environment for cluster based information dissemination

puting power, communication and storage facilities can be pooled with other vehicles on the road or rented to customers, similar to the way in which the resources of the present conventional cloud are provided to the end users as described in Figure 5.1.

With the evolution of Internet of Things and need for information retrieval from various objects distributed across the globe, VCC has gained lot of momentum to provide seamless services to the end users [153]. The proposed clustering scheme forms the basis for a large number of applications in VANET's. In this environment, we have studied the communication issues and parameters of vehicular dynamics to support secure V2V cloud-based clustering for efficient communication. Most of the existing proposals in literature depend upon the V2R communication using roadside infrastructure to provide security. Therefore, we propose a distributed and dynamic clustering scheme for VANET's to fulfill the requirement of security in cloud-based infrastructure for generating high security clusters among the vehicles.

The proposed scheme is distributed because the set of elected vehicles in a cluster become CHs having the responsibility of monitoring the neighboring vehicles. The presence of infrastructure on the road is optional and it is dynamic because the election of CHs depends on the topological changes using the parameter based on vehicular mobility. Unlike existing clustering schemes in literature, the election of CHs is based on two parameters based on vehicular dynamics and trustworthiness behavior. The trustworthiness weight of a vehicle depends on its own trust level as compared to average trust levels of all the vehicles within the network.

The proposed scheme uses a vehicular framework for sensing and broadcasting the

current vehicular parameters such as velocity, density, and location of each vehicle on the road. To account for the high speed of vehicles and vehicular mobility patterns, the proposed scheme considers different values of clustering durations for constructing optimized clusters. The clustering duration are varied between 10 to 140 seconds by taking into account varying speed of vehicles on urban roads. The variation in clustering duration helps in attaining optimum cluster stability, using simulations.

The proposed scheme consists of a module for sensing vehicular parameters and a cloud-based module for computing trustworthiness values. The sensing module is deployed with on-board vehicles which provides information about a vehicle's traffic and security characteristics. It broadcasts this information periodically to the vehicular cloud using various short, medium, and long range communication protocols such as- RFID, Bluetooth, WiFi, WIMAX, and Long Term Evolution. The centralized cloud-based computing module helps to evaluate the security behavior of every vehicle in immediate vicinity and then uses the past behavior characteristics of the vehicle to modify its trustworthiness level with respect to other vehicles.

5.2 System Model for Secure Clustering

The key objective of the proposed secure clustering scheme is to create clusters by using a new efficient trust-based weight computation method for VANET's. The proposed scheme creates stable clusters and also provides enhanced security to the clusters. The process of creating secure clusters is formulated as follows.

Each vehicle is assumed to be equipped with a standard radio interface having a standard transmitting range of R units. A vehicle can communicate with all other vehicles in its neighboring region and each vehicle is uniquely identified by using its Vehicular Identifier (VID). The communication framework for the proposed scheme uses a three layered communication model that is used for interaction between the vehicular and cloud environments. Figure 5.2 describes the key functions and protocols used by each layer along with the type of interface (wired or wireless) used by the interfaces between these layers.

To enhance the security of the created clusters, the trustworthiness values received from a centralized cloud based repository is assumed to be available for the vehicles. The proposed clustering scheme uses a framework consisting of On Board Broadcasting (OBB) module and the Vehicular Cloud-based Estimation (VCE) module for creating clusters. Figure 5.2 also shows the pictorial description of this framework which describes the key functionalities and inter-connection of the different modules in the vehicular environment. The OBB module tracks the location, and transmission characteristics of vehicles. It also computes the trustworthiness weight values which are then transmitted to the VCE.
The VCE which is located on a vehicular cloud then computes the trustworthiness weights

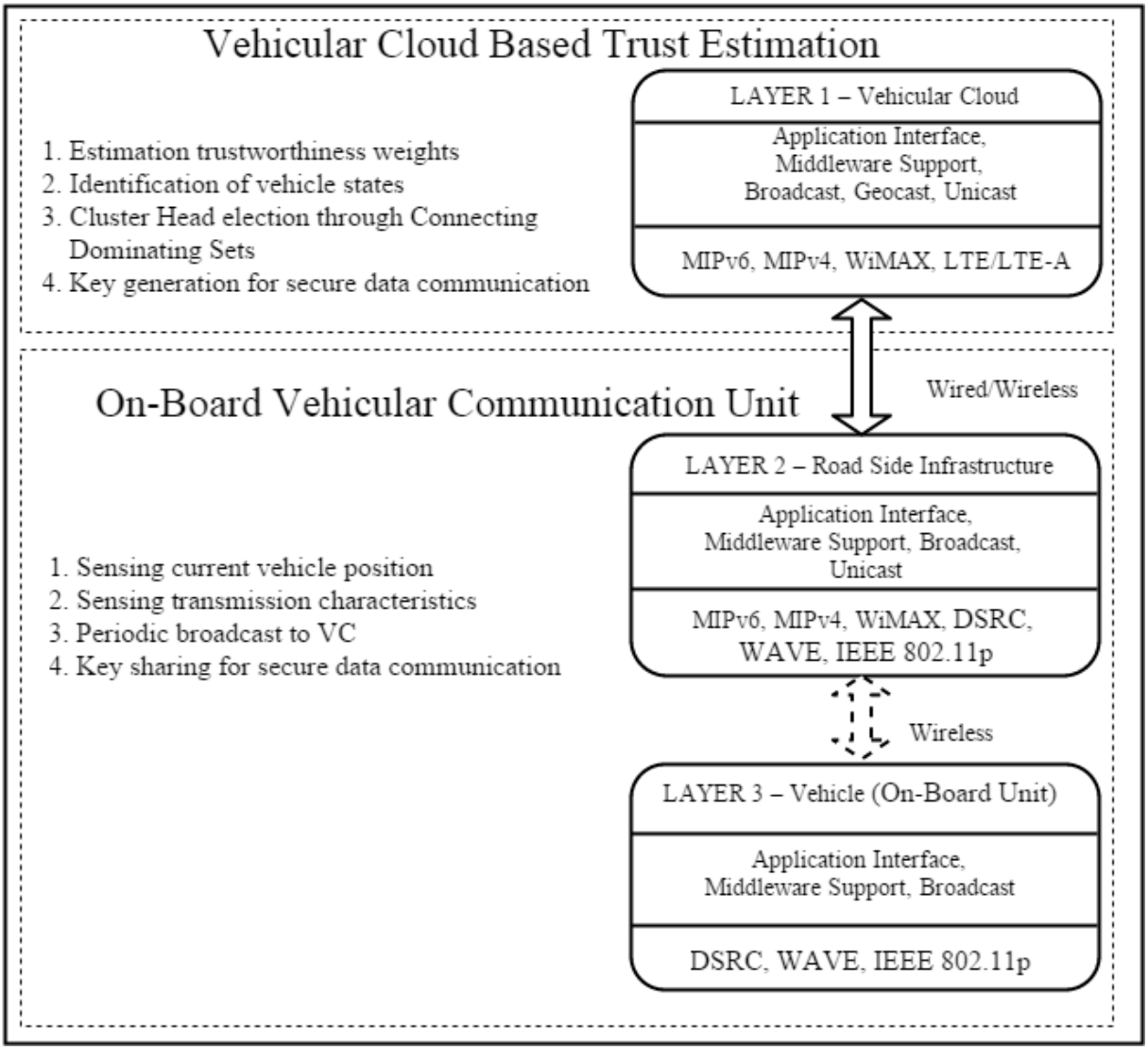


Figure 5.2: Communication Framework model for Secure Clustering scheme.

of all vehicles by considering the received messages from the OBB module.

5.2.1 On-Board Broadcasting Unit

We have considered that all the vehicles use the GPS transmitter for repeatedly estimating the current position of the vehicles along with WAVE, DSRC Radio, an on-board communication unit based on Wi-Fi, WIMAX or existing cellular network. A vehicle continuously tracks its own position and intimates it to the network as well as the Vehicular Cloud by transmitting periodic beacon messages.

5.2.2 Vehicular Cloud based Estimation

The main responsibility of VCE module is the estimation of trustworthiness values for all vehicles. The VCE module receives the periodic beacon messages from all the vehicles and then computes their trustworthiness values. These values are utilized for detecting the security level of every vehicle and then classifying them into states based on the computed trustworthiness values. The categorization of vehicles is done based on two threshold values (δ_s, δ_u). The repeated computation of these values, provides continuous information about the behavior of each vehicle, which then assists in creating CHs that are more secure by observing their past and present security characteristics.

The threshold values δ_s and δ_u are used to categorize vehicles into secure and unsecure states respectively. The VCE also initiates the process of clustering and determines the CHs using connected dominating set technique based on the trustworthiness weight. The vehicular cloud broadcasts this clustering information back to vehicles. The OBB module then initiates the actual re-organization by taking account of the clustering information received from VCE.

5.2.3 Principle of Vehicular Clustering

Let us consider a set N of n vehicles. A vehicular cluster V_c is a partition of N such that , $V_c : 2^N \rightarrow Q$,
where $V_c = c_1, c_2...c_m$ and, c_i is the i^{th} cluster of vehicle and there exists a mapping of each vehicle to a rational number. Hence, following equalities hold for the created clusters.

$$\bigcup_{i\epsilon[1...m]} c_i = V_c \; and$$
$$c_i \cap c_j = \phi \text{ for } i \neq j$$

Thus V_c denotes the set of all the clusters created for the network. Let $V_c(t)$ denote the cluster at time instant t. We have also assumed that the preference for a vehicle to join a particular cluster $V_c(t)$ is decided by its security state and its physical location. Each vehicle also has a value defined as its utility or weight value that depends on its relative trustworthiness with respect to other vehicles. This trustworthiness value is indicated as $W_{i,t}$ for i= 1 to n, $t \leq T$ where, T is the total lifetime of the network and t is the time periodic duration after which re-clustering is performed. Thus,

$$t << T \tag{5.1}$$

Each vehicle periodically generates the value of weight for itself. This value depends on the security level of the vehicle based on its behavior with respect to its role in providing integrity and authentication to the cluster. We have assumed that for the proposed

clustering scheme, a particular cluster does not indicate its collective behavior to the vehicular cloud as a group. Therefore, vehicles cooperate for increasing the cluster stability and reducing the cost of cluster formation by periodically transmitting beacon messages among CHs. Since, the stability of vehicles within a particular cluster affects the stability of other clusters also, the efficiency of clustering may be higher when all vehicles cooperate for constructing clusters as compared to the case where each cluster works in isolation. This necessitates the need for each cluster to be constructed by considering the characteristics of that vehicle as well as its neighborhood by utilizing the values of the combined trustworthiness weight of the whole network as a threshold. The vehicular cloud then creates the optimized cluster by considering the trustworthiness values obtained from these vehicles, based on a centralized clustering scheme.

Since the vehicles are not aware about the clusters, the vehicular cloud helps in aggregating and comparing the values of the vehicles to construct an optimized cluster for a relatively longer duration. This also helps to reduce the transmission overhead as frequent re-clustering may decrease the overall performance of the network. We have also assumed that there is a coordinated change in clusters, and vehicles can not change their cluster asynchronously. This phenomenon helps to reduce the frequent re-clustering and also limits the transmission overhead of messages required for CH selection between vehicles and Vehicular Cloud. Further, the proposed clustering scheme also assumes that all the vehicles are rational and the clusters are formed by giving precedence to current values of trustworthiness with the weight of past behavior gradually decreasing with time. The proposed clustering approach also assumes that the decision regarding which cluster a vehicle belongs to is delegated to the vehicular cloud. However, a vehicle can leave a cluster by broadcasting a message to the vehicular cloud if it has no other vehicle in its vicinity. In this case, when a vehicle has left a cluster it is considered as unattached CH till subsequent cluster re-organization.

5.2.4 Attack Model

Some of the most important threats in VANETs are related to authentication of data and vehicles. We can refer to these vulnerabilities as vehicle identification related vulnerabilities. These attacks are described as a model in which a malicious vehicle may add malicious or incorrect safety messages for other vehicles or attempt to flood the network by launching a Black-Hole attack through some victim vehicles. Our attack model focuses on the analysis and detection of these attacks, by providing information to CHs that then categorize these attacks through the sequence of events. Our system does not use certificates or CAs as they result in restricted access for vehicles and thus may not be able to prevent a vehicle that has valid certificates from attacking. However if the system would make use of certificates, our analysis would be the same, with the exception that

the attacker would have to get access to valid certificates.

Also our attack model employs reactive security mechanisms such as consistency checks through CHs. The basic mechanism for checking the vulnerability of the system relies on number of messages that are sent during particular time duration. The messages are discarded if they contain a time value that is older than a threshold value based on the clustering duration. The position of the sender vehicle is also to be considered as only those messages are considered where the position of the sender is within the receiver's radio range otherwise, the message is ignored.

Some more checks have been added to gradually improve the attacker model. These checks comprise further validations of vehicular characteristics such as the vehicle's velocity and direction, acceleration and heading change. Then verifications can be done to check whether these values are consistent with current traffic situation. Another important check is to validate the movement of the vehicles relative to each other so as check the relative order of vehicles and its movement pattern over large time duration relative to other vehicles. Multi-lane roads also require mechanisms that can distinguish different mobility pattern of vehicles for different lanes.

5.3 Proposed Clustering Scheme

Clustering in VANETs is performed to create a hierarchical structure of vehicles, thereby providing an organized structure for achieving efficient management of the network. A cluster is considered as a connected subgraph of the vehicular network at a particular instant of time that contains CHs and CMs. Generally, a vehicle belongs to a single cluster and cluster is also required to follow some constraints so as to improve the network performance with minimal overheads. We now present the weight-based secure clustering scheme. The proposed technique creates cluster based on values of trustworthiness weights ($W_{i,t}$) of the vehicles.

The values of trustworthiness weights ($W_{i,t}$) at time instant t are measured based on the number of packets transmitted (p_t) and number of packets received (p_r) by each vehicle since the last re-clustering time. These values are then used to compute the Activity Profile ($\eta_{i,t}$) for each vehicle.

$$\eta_{i,t} = (|p_r - p_t)|/(p_r - p_t) \tag{5.2}$$

The computed value of $\eta_{i,t}$ is then forwarded to the vehicular cloud. The first task performed by vehicular cloud is to compute the average value of Activity Profile $\eta_{avg,t}$ for

all the vehicles in the network as follows:

$$\eta_{avg,t} = \frac{1}{n} \sum_{i=1}^{n} \eta_{i,t} \tag{5.3}$$

The above values of $\eta_{i,t}$ and $\eta_{avg,t}$ are used to compute the trustworthiness weight for all the vehicles at time t as follows.

$$W_{i,t} = |\eta_{i,t} \ - \ \eta_{avg,t}| + \frac{1}{t-1} \sum_{j=1}^{t-1} W_{i,j} \tag{5.4}$$

In the proposed model, the trustworthiness property of a vehicle is used to determine its weight value. Based on the computed values of trustworthiness, vehicles are classified into various states. Generally, a vehicle is assumed to be in three different states. These states are designated as *Secure, Insecure* and *Vulnerable* as shown in Figure 5.3. The value of current weight for each vehicle is the key parameter for deciding the next state ($q_{i+1,t}$) of a vehicle. The proposed scheme also uses dynamic threshold values $\delta_{s,t}$ & $\delta_{u,t}$ to categorize the vehicles into states. Thus, at a particular instant of time t, the next state $q_{i+1,t}$ of a vehicle V_i is defined in terms of following inequalities:

$$\textbf{If } (W_{i,t} \geq \delta_{s,t}) \qquad \textbf{then } q_{i+1,t} = Step\_Up\ (q_{i,t}) \tag{5.5}$$

$$\textbf{If } (\delta_{s,t} > W_{i,t} \geq \delta_{u,t}) \quad \textbf{then } q_{i+1,t} = Continue\ (q_{i,t}) \tag{5.6}$$

$$\textbf{If } (W_{i,t} \leq \delta_{u,t}) \qquad \textbf{then } q_{i+1,t} = Step\_Down\ (q_{i,t}) \tag{5.7}$$

To make the transition from the current state to the next state three procedures i.e. Step_up, Continue, and Step_Down have been developed. All the three algorithms take the current state $q_{i,t}$ as the input and determine the next state as $q_{i+1,t}$ as shown in algorithms 1, 2 and 3.

The values of δ_s and δ_u are computed by vehicular cloud as follows.

$$W_{\mu,t} = \frac{1}{n} \sum_{i=1}^{n} W_{i,t} \tag{5.8}$$

where, $W_{\mu,t}$ is the average value of trustworthiness weights for VANET's. The values of $\delta_{u,t}$ and $\delta_{s,t}$ are computed for every clustering duration using the mean and variance of

Algorithm 13 *Step_Up*

Inputs: $q_{i,t}$, $W_{i,t}$
Output: $q_{i+1,t}$

Begin
if ($W_{i,t} \geq \delta_{s,t}$ & $q_{i,t}$ == Secure) **then**
  $q_{i+1,t}$= Secure
end if
if ($W_{i,t} \geq \delta_{s,t}$ & $q_{i,t}$ == Vulnerable) **then**
  $q_{i+1,t}$= Secure
end if
if ($W_{i,t} \geq \delta_{s,t}$ & $q_{i,t}$ == Unsecure) **then**
  $q_{i+1,t}$= Vulnerable
end if
End

Algorithm 14 *Continue*

Inputs: $q_{i,t}$, $W_{i,t}$
Output: $q_{i+1,t}$

Begin
if ($\delta_{s,t} \geq W_{i,t} > \delta_{u,t}$) **then**
  $q_{i+1,t}$= Secure
end if
End

Algorithm 15 *Step_Down*

1
Inputs: $q_{i,t}$, $W_{i,t}$
Output: $q_{i+1,t}$

Begin
if ($W_{i,t} < \delta_{u,t}$ & $q_{i,t}$ == Secure) **then**
  $q_{i+1,t}$= Vulnerable
end if
if ($W_{i,t} < \delta_{u,t}$ & $q_{i,t}$ == Vulnerable) **then**
  $q_{i+1,t}$= Unsecure
else
  $q_{i+1,t}$= Unsecure
end if
End

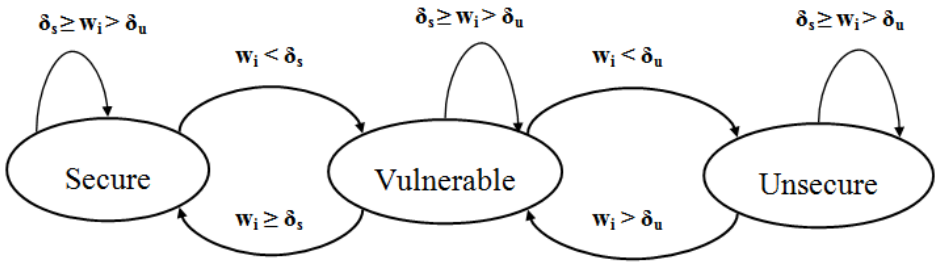


Figure 5.3: State transition diagram based on trust values in secure clustering.

the weight values as follows.

$$W_{\sigma,t} = \sqrt{\Sigma_{i-1}^{n}(W_{i,t} - W_{\mu,t})} \tag{5.9}$$

Now, the difference between the values of $W_{\mu,t}$ and $W_{\sigma,t}$ is used to compute the threshold values $\delta_{s,t}$ and $\delta_{u,t}$, i.e.,

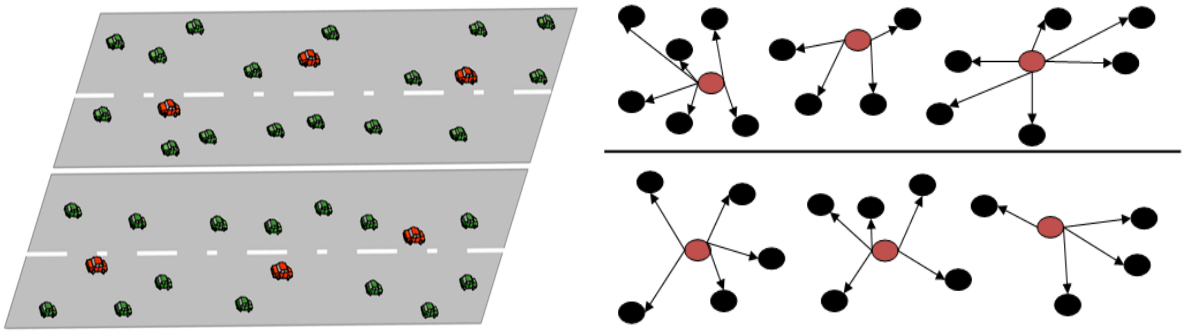
$$\delta_{s,t} = W_{\mu,t} + W_{\sigma,t} \tag{5.10}$$

$$\delta_{u,t} = W_{\mu,t} - W_{\sigma,t} \tag{5.11}$$

After the states of the vehicles are modified, the set of probable CHs are identified by the vehicular cloud. The CHs are elected by considering the degree of connectivity along with the trustworthiness weights of the vehicles. Figure 5.4(a) describes a basic road side scenario with vehicles along the road and its equivalent clustering scenario is shown in Figure 5.4(a). As shown in Figure 5.4(b), CHs are selected based on their trustworthiness values and relative average distance of a vehicle to its neighbors. Vehicles in the network may have different weights but, we have used the tuple $< W_{i,t}, D_{avg}, VID >$ to assign distinct weights for every vehicle. The selection of CH is based on the trustworthiness weight associated with each vehicle. A vehicle with higher possible value of weight is considered to have more suitability to become CH.

Thus, the proposed scheme uses CDS technique for selection of CHs. However, by selecting only nodes that belong to secure state as CHs, the proposed technique ensures better security as well as comparable stability for the generated clusters. A node can be considered to be qualified for being elected as CH iff it is in secure state or no other node in its vicinity is in secure state, i.e.,

$$CH = \{V_i \mid \forall i,\ i\ \epsilon Secure \vee\ N_i^+ = \phi\} \tag{5.12}$$



(a) (b)

Figure 5.4: (a) Road side scenario.and (b) Equivalent clustering scenario used in secure clustering.

where, N_i^+ indicates the secure neighborhood of i.

Thus, the proposed scheme attempts to create clusters in which vehicles with higher security weight act as CHs. The CH election is performed by considering the node with higher weight value among the nodes in secure state along with its relative position within the cluster that is indicated by its degree of connectivity Δ. If a CH leaves a cluster, then, those CMs that detect the absence of CHs intimate other members in its neighborhood about the re-clustering process by broadcasting a cluster re-election request message. However, the process of re-clustering will begin only when vehicular cloud receives re-clustering message from vehicles which are in secure state. This helps to ensure the message validation since the vehicles belonging to secure state exhibit better trustworthiness behavior as compared to other vehicles.

Lemma 5.1: A cluster V_i is a more secure cluster if CHs have higher out-degree value.
Proof: The proof for the above can be deduced from the description of vehicular transition graph which is used for CH selection. As shown in Figure 5.4(c), a vehicular scenario is considered to describe the complete process of clustering through an example of transition graph with various nodes.
In the graph, if the nodes with lesser out degree are selected as CHs then it affects the overall network security as the vehicles with greater trust values transmit messages using CHs having comparatively lower values.
This either compromises the network security or reduces the role of secure nodes in the network functioning. This is not desirable for effective functioning of cluster.
Hence, the proof of the Lemma 5.1 holds.

5.3.1 Cluster Head Election

The vehicular cloud performs CH election by using an algorithm based on CDS protocol. Although, CDS has been used previously for clustering in adhoc networks but, our proposed approach constructs a CDS-based cluster that assigns and updates the trust values dynamically for each vehicle in the network. The proposed scheme gives greater priority to the trust values of vehicles rather than conventional vehicular parameters for cluster formation. The nodes transmit the values of packets transmitted and received during the previous time interval to the vehicular cloud. The vehicular cloud receives the current position and the transmission characteristics of all vehicles in the network. Based on their current state and received messages, the vehicular cloud then updates their trust values by using a centralized computing algorithm.

The key reason for delegating this responsibility on a centralized cloud repository is to help provide a higher degree of security because a malicious user or vehicle would not be able to manipulate the trust values. Since the proposed scheme computes the trustworthiness values by also considering past behavior pattern of the vehicle, accounts for sudden and abnormal spurts or losses of packet from a particular vehicle during particular time interval. Thus, even if any vehicle manipulates its transmission characteristics during some time duration, it would not have too much effect on its weight.

The vehicular cloud uses a model of transition graph to elect CHs. Each node is placed at its corresponding position on the graph by computing its relative position with respect to other nodes. The GPS based position of each vehicle is used as the initial input for position estimation. To account for the variation due to the movement of vehicles, the proposed scheme uses the prediction estimation algorithm [156] for performing short term prediction during the computation of CHs. Figure 5.6 depicts the process of creating the transition graph based on the present value of every vehicle. As shown in Figure 5.4, initially the position of each vehicle on the road is used to identify the vertices's of the transition graph. In the next step shown in Figure 5.5, trustworthiness-based weight values are computed for the vehicles. Then, the directed edges are added to the graph as shown in Figure 5.6 where the direction of the edge is from the node with higher trust value towards the lower value node. In case, the weights are equal, the vehicle with lower VID value is considered to be the source vertex. The CH election process is then initiated and the CHs are elected using CDS technique based on trust value of every vertex as describe in Figure 5.7, and 5.8. The final clusters are then created and clustering information is broadcasted back to the vehicular networks. We have constructed the transition graph for the cluster formation process by building a directed graph $G(V, E)$ where, each vertex V is a vehicle and there is an edge from vehicle V_i to V_j iff $W_i > W_j$. The vehicular cloud constructs a CDS by selecting the nodes having higher weight values as the members of the dominating set.

The proposed CH election approach consists of three main steps described as follows,

- Identification of neighbors for each vehicle present in the network.
- Dynamic updation of trust values for the vehicles at discrete time intervals.
- Implementation of CDS algorithm based on trustworthiness weight for CH election.

5.3.2 Identification of Neighboring Vehicles

Initially, the distance of each vehicle 'V_i' present in the network to every other vehicle 'V_j' is computed and then the average value of distance between all the other vehicles with respect to single vehicle is calculated as follow,

$$D_{avg,i} = \frac{1}{n} \sum_{j=1}^{n} d_{ij} \tag{5.13}$$

where, d_{ij} represents the distance between the vehicle V_i and V_j. D_i is the average distance between the reference vehicle V_i and all other n vehicles in the network. Then, the average distance parameter Δ is computed as follows,

$$\Delta = \frac{1}{n} \sum_{i=1}^{n} D_i \tag{5.14}$$

The value of Δ is the threshold value for selecting the clustering neighborhood of particular vehicle. A vehicle whose distance from a vehicle V_i is less than Δ is considered to be in clustering neighborhood of V_i' by vehicular cloud. The distance D_{ij} of vehicle V_i to vehicle V_j is compared with the threshold value and if D_{ij} is less than the threshold value, then V_j is declared as a neighbor of V_i. This helps the vehicular cloud to compute the cluster neighborhood for a vehicles. Algorithm 16 describes the steps to find the neighboring nodes of a vehicle in the network.

5.3.3 Vehicular Parameters Estimation

The next step is to estimate the vehicular parameters for computing the trust values of vehicles. The CH election is performed only after discrete time intervals by considering the reputation i.e., the trust value of nodes. In trust estimation algorithm, each vehicle is assigned a trust weight. Initially, all the vehicles are assigned a trust weight W_{ini}. The value of $W_{ini,0}$ is dependent on the number of vehicles in the network and is defined as

Algorithm 16 *Vehicle Neighborhood Detection*

Inputs: Neighbor list of the vehicle, Array d[][],sum=0,avg
Output: Set of neighbors of each vehicle (N_i^+)
Assumptions:

Begin
Compute total number of vehicles 'n'
for each vehicle v_i **do**
4: calculate the distance *dist* to each vehicle v_j from v_i vehicle
$d[i][j] = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
end for
for each vehicle v_i **do**
8: **for** each vehicle v_j **do**
sum = sum + d [i][j] //calculate sum of all distances of vehicles
end for
end for
12: Calculate *avg*, average distance of all vehicles
Δ = (sum/n)
for each vehicle v_i **do**
for each vehicle v_j **do**
16: **if** $(dist_{ij} < threshold)$ **then**
set vehicle v_j as neighbor of v_i
add v_j to N_i^+
end if
20: **end for**
end for
End

follows.

$$W_{ini,0} = 1/n \ \forall \ V_i \ \epsilon \ N \tag{5.15}$$

The trustworthiness weight is the key parameter for CH election and cluster formation. A higher value of trustworthiness weight results in greater probability of that vehicle becoming a CH. The packets transmitted and received by each node are communicated to the vehicular cloud. Based on the number of transmitted and received packets along with it's average trustworthiness value during the previous iterations, each node's trustworthiness value is re-computed by the cloud.

Another important aspect involves the behavior of the proposed scheme, when vehicles enter or leave the network. In the case, when a new vehicle enters in the network, it transmits a request for joining some cluster by broadcasting a CLUSTER_JOIN message. This vehicle is initially declared as unattached CH and is assign the initial weight ($W_{ini,0}$). During next re-clustering cycle, these unassigned vehicles will become part of some cluster. Subsequently these vehicles will periodically broadcast their transmission

characteristics and will be assigned trustworthiness values by the vehicular cloud. If any vehicle leaves the network, it will broadcast a CLUSTER_LEAVE message. When the vehicular cloud receives the message, it will remove this vehicle from its repository. Algorithm 17 describes the process of vehicular trustworthiness weight estimation process.

5.3.4 CH Election using CDS Algorithm

The last step of the proposed clustering scheme is the election of CHs with the CDS approach. A clustering technique should provide efficient data delivery and cluster stability. However, creating secure clusters is also an important aspect for any clustering scheme. Clustering algorithms for vehicular environment should be efficient so that minimum number of vehicles becomes CH and all the vehicles should be covered by these elected CHs. We have applied the CDS approach based on trust values for clustering due to its computational efficiency. Another criteria for clustering is to select a minimal set of nodes from a graph by processing all or majority of nodes in the minimum time. Although various techniques are available that satisfy the above criteria, but CDS based algorithm has been preferred due to its suitability in vehicular environment.

All the nodes are first marked as unvisited and a node with maximum neighborhood based on the number of outgoing edges is selected for evaluation. Then, this node is marked as visited and added to set of dominating nodes. In the next step, nodes in the neighborhood of the selected dominating node are identified and considered as visited nodes. The process goes on until all the neighboring nodes of the dominating node have not been added in the set of visited nodes. The above steps are repeated but, the nodes already in dominating set and their neighbors are not considered in subsequent steps.

This process of adding nodes onto set of dominating and visited set of nodes is repeated till every node is either dominating node, or its one hop neighbor. To distinguish between three types of nodes, algorithm 18 uses three different sets that are denoted as *Unvisited (U), Covered (C)* and *Dominating set (D)*. Initially, no node has been traversed and all nodes are entered into set U. When the node with maximum neighborhood is selected, it is put into set D and removed from U. Then all the neighbors of the set which have been traversed are also removed from U and they become member of set C. Now, the node with maximum neighborhood from remaining nodes is selected and entered into set D. All its one hop neighbors are put into set C. This process continues till all nodes are traversed and the set U become empty set. Algorithm 18 depicts the CH election algorithm and describes the steps of proposed clustering technique to select CH's in a given vehicular network.

The vehicles that act rationally will attempt to change its cluster, if it can join a cluster that is more stable and secure. Since the value of weights are computed in a centralized manner, it is possible to create and analyze the clustering model. As discussed

Algorithm 17 *Vehicular Trust Value Estimation*

Inputs: List of all vehicles,
Total Simulation Time 'T'
Cluster Formation Time 't'
Output:Trust Values of Vehicle Cluster
Assumptions: δ is extremely small time duration used for CH selection

Begin
Simulation Time = 0
for each vehicle v_i **do**
  Initialize all vehicle's trust value W_i with random value less than 1
  $W = \sum_{i=1}^{n} W_i$
end for
while (Simulation_Time $\leq$ T) **do**
  for every time interval t **do**
    if (CLUSTER_JOIN || CLUSTER_LEAVE message received) **then**
      Update Vehicle List
    end if
    for each vehicle v_i **do**
      Calculate total number of received and transmitted packets
      Compute present position of vehicle
      Estimate future position of vehicle for $t + \delta$
      Compute $\eta_i = |p_r - p_t / p_r + p_t|$
      Broadcast η_i value & estimated future position
    end for
    while all message not receieved **do**
      wait
    end while
    $T_{avg} = 1/n \sum_{i=1}^{n} T_i$
    for each vehicle v_i **do**
      $W_{i,t} = |T_i - T_{avg}| + 1/(t-1) \sum_{j=1}^{j-1} W_j$
    end for
    Compute next state $q_{i+1,t}$
    Construct vehicular transition graph G
    Initiate CH election using CDS
    Broadcast CH information to vehicles
  end for
  Simulation_Time ++
end while
End

previously, the vehicular cloud builds a transition graph where, the nodes represent the present position of vehicles and the directed edges are represented based on the differences in trustworthiness weight of two vehicles. Thus, an edge from vehicle V_i to its neighboring vehicle V_j exists iff

$$W_i > W_j \tag{5.16}$$

i.e., weight of V_i is greater then weight of V_j.

Algorithm 18 *CH Election CDS*

Inputs: Neighbor list of the vehicle, Array v[]
Sets U, D and C
Output: Cluster Heads, N_i is the neighborhood of v_i
Assumptions:

1: **Begin**
2: **for** each vehicle v_i **do**
3: Set num = number of neighbors of vehicles v_i
4: v[i]=num
5: **end for**
6: Add all the nodes into U, Add all nodes into N_i
7: Set $max_i = max(v[i])$
8: Set $v_i = max_j$
9: **for** each $N_i \ \epsilon \ U$ **do**
10: Add all $v_i \ \epsilon \ N_i \ into \ C$
11: **end for**
12: Add $v_i \ into \ Set \ D$
13: $U = U - v_i$
14: $U = U - N_i$
15: **for** each $v_i \ \epsilon \ U$ **do**
16: Create modified N_i
17: **end for**
18: **End**

The CHs are elected by selecting the vehicle having highest out degree. This results in selection of those vehicles for assuming the role of CH that have comparatively better connectivity along with more trustworthiness. In case two or more vehicles have same out degree, their trustworthiness weight is considered for CH election. If the weights are also same then the VID's are compared and the vehicle with lower VID is then added as CH. Since the number of vehicles in a cluster is limited to a upper bound to account for

network limitations due to mobility and transmition constraints, we have assumed that maximum number of vehicles in a cluster is k. Thus, for each node, there can be almost k outgoing edges from each node.

The transition graph model describes the possible roles that the vehicle can take. If the node has no out-degree in the graph then, it can only become a member node. For other nodes, there can be a maximum of k out going edges. The average value of outgoing edges can thus be taken as $\frac{k}{2}$. The bounds for the number of CHs to be selected from the graph can thus be bounded between $O(\frac{2n}{k})$ to O(n), where n is the total number of vehicles. The probability of an edge being processed for any vehicle by vehicular cloud is defined as $\frac{2}{k}$.

5.3.5 Secure Data Dissemination

The secure data dissemination is performed through a key based encryption/decryption scheme. It is assumed that the any node that has data for transmission will first transfer that message to it's corresponding CH. Each CH acts as a gateway for transferring messages to their neighboring CHs or to vehicles within its own cluster.To provide a secure data transmission, initially the system initialization and registration process is performed followed by a privacy preserving communication process.

System Initialization: All the components of the model such as CM, CH, OBU, RSU and Vehicular Cloud are assumed to be using the same Elliptic Curve domain parameters. These parameters (p, G, a, b, m, h) are generated by Vehicular Cloud using Algorithm 19. The scheme uses Advanced Encryption Standard based encryption and decryption function for the symmetric key sharing encryption and decryption protocol respectively. The CHs and CMs are considered as authorized users of the proposed network.

Registration Process: All the vehicles are registered with the Vehicular Cloud as soon as they transmit their CLUSTER_JOIN message and then these vehicles become part of the network. Each user has a unique pseudo identity P_i and each vehicle has a unique Pseudo identity V_i that is assigned through the vehicular cloud. The CHs store the identities of all the vehicles in its cluster along with pseudo identities in the local database. The user's use their pseudo identities while transmitting any message.

Privacy Preserving Communication Process: In proposed work, the privacy is preserved by using encryption scheme for the messages. The sender and receiver use different algorithms for encryption and decryption. The message transfer between CMs and CHs or between CHs are done using elliptic curve cryptography scheme. In order to preserve privacy and authentication the various components of vehicular network use Elliptic Curve Integrated Encryption Scheme (ECIES). The proposed privacy preserving communication process is now described where both secure intra-cluster communication as well as secure

inter-cluster communication are performed using below mentioned process.

Algorithm 19 *Elliptic Curve Domain Parameter Generation*

Inputs: *Field Order p, Field Representation for* F_p*, Security Level L satisfying* $160 \leq L \leq \lfloor log_2 p \rfloor$ *and* $2^L \geq 4\sqrt{p}$

Outputs: $D = (p, G, a, b, m, h)$

1: **Begin**
2: Select $a, b \ \epsilon \ F_p$
3: Let E be $y^2 = x^3 + ax + b$ and in the case F_p is a prime field.
4: Compute $N = Hash \ (\ E \ (\ F_p \ ) \ )$.
5: Verify that N is divisible by a large prime m satisfying $m \geq 2^L$. If not, then go to step 2.
6: Verify that m does not divide $p^k - 1$ for $1 \leq k \leq 20$. If not, then go to step 2
7: Verify that $m \neq p$. If not, then go to step 2.
8: Set $h \leftarrow N/m$
9: Select an arbitrary point $G^{'} \ \epsilon \ E \ (\ Fp \ )$ and set $G = Hash \ (G^{'})$. Repeat until $G \neq \infty$.
10: Return (p, G, a, b, m, h)
11: **End**

Algorithm 20 *Key Pair Generation*

Inputs: *Elliptic Curve Domain parameters (p, G,a,b, m, h), id,* P_S*,* K_S *, hash function* $h()$*, Concatenation Operation +*

Outputs:*Public key* P_{ku}*, Private key* P_{kr}

Assumptions: *The sender and receiver have respective key pairs i.e. public and private keys and also others public key*

1: **Begin**
2: Select $P_{kr} \ \epsilon \ R \ [\ 1, \ m - 1 \ ]$.
3: Compute $P_{ku} = \ P_{kr}. \ G$.
4: Return (P_{ku}, P_{kr})
5: **End**

Any vehicle that needs to transmit some message initially generates a timestamp T for that message. The vehicle also attaches the pseudo identity of the user to this message. Now the sender transmit's the message to it's current CH. For secure communication between the sender and CH, both source node and CH use elliptic curve key generation algorithm for private/public key pair. They use domain parameters $D = (p, G, a, b, m, h)$ from Algorithm 19 as input and then generate private/ public key pair (P_{kr}, P_{ku}). For key generation Algorithm 20 is used by the sender CMs and CHs. The CMs and CHs communicate using ECIES protocol which involves the predefined parameters that are based on a Message Authentication Code (MsgAC) algorithm, a symmetric key encryption

algorithm, an elliptic curve Diffie-Hellman key establishment protocol for establishing the shared secret key between sender and CH. The symmetric keys are generated using Key Derivation Function(KDF).

The sender node selects a random secret key P_{kr} such that it's value lies between 0 and $m-1$. Then the public key is computed using a point P_{ku} on the elliptic curve such that $P_{ku} = P_{kr}.G$. The vehicles share their public key for with CHs for communication.

ECIES Encryption: In order to encrypt a message M with the receiver public key PKI, the sender uses ECIES encryption Algorithm 21. In this algorithm, sender selects a random number k such that it lies between 1 and m-1, and also generates a point R on the curve using secret key k and the base point G. The sender uses key establishment protocol to generate a shared secret key Z such that it derives from random secret key k and receiver public key P_{ku}. The sender also verifies that Z is not infinity. The sender uses KDF that is already established between sender and receiver to generate the keying data K which is combination of $k1$ and $k2$. The length of K should be Encryption key length i.e. $k1$ and MsgAC key length i.e. $k2$. Encryption key length will be used for symmetric key encryption and MsgAC key length will be used as MsgAC key. $K1$ is used to generate ciphertext C and MsgAC algorithm uses key $k2$ to compute tag T.

Algorithm 21 *Sender Node Encryption*

Inputs: *Domain parameters D = (p,G,a,b,m,h), Public key P_{ku}, Plaintext M*

Outputs:*Public key P_{ku}, Private key P_{kr}*

Assumptions: *The sender and receiver have respective key pairs i.e. public and private keys and also others public key*

1: **Begin**
2: Select $k \epsilon R [1, m-1]$.
3: Compute $R = k.G$ and $Z = h.(k.(P_{ku}))$.
4: **if** $(Z == \infty)$ **then**
5: Go to Step 2.
6: **end if**
7: $(k1, k2) \leftarrow DF(xZ, R)$, where xZ is the x-coordinate of Z.
8: Compute $C = ENC(k1(M))$ and $T = MsgAC(k2(C))$.
9: Return (R, C, T).
10: **End**

ECIES Decryption: In order to decrypt the received message (R, C, T), the receiver follows the Algorithm 22. The receiver node parses the message to recover the values of R, C and T. Then receiver performs the validation of embedded public key R. If the validation of embedded public key R fails then it will reject the message. The receiver uses key establishment protocol to compute the shared secret Z. The shared secret can

be computed using private key P_{kr} of the receiver and embedded public key R. If the value of the Z is infinity then the receiver will also reject the message. The receiver uses KDF that is already established between sender and receiver to generate the keying data K which is combination of $k1$ and $k2$. The length of K should be Encryption key length i.e. $k1$ and MsgAC key length $k2$. Encryption key length will be used for symmetric key decryption process and MsgAC key length will be used as MsgAC key. $K1$ is used to decrypt the ciphertext C into M and MsgAC algorithm uses key $k2$ to compute tag T_1. If tag T_1 is equal to T, then it will accept the ciphertext otherwise reject the ciphertext. We also evaluate the computational complexity in terms of time or key

Algorithm 22 *Receiver Node Decryption*

Inputs: *Domain parameters D = (p,G,a,b,m,h), Private key P_{kr}, Ciphertext (R, C, T)*
Outputs:*Plaintext M or rejection of the Ciphertext.*
Assumptions: *The sender and receiver have respective key pairs i.e. public and private keys and also others public key*

1: **Begin**
2: Perform an embedded public key validation of R.
3: **if** (Validation Fails) **then**
4: return ("Reject the Ciphertext").
5: **end if**
6: Compute $Z = h.\ P_{kr}\ \ R$.
7: **if** $(Z == \infty)$ **then**
8: return ("Reject the Ciphertext").
9: **end if**
10: $(k1, k2) \leftarrow KDF\ (\ xZ, R\ )$, where xZ is the x-coordinate of Z.
11: Compute $T_1 = MsgAC\ (\ k2(\ C\ )\ )$.
12: **if** $(T_1 \neq T)$ **then**
13: return ("Reject the Ciphertext").
14: **end if**
15: Compute m = DEC . (k1. (C)).
16: Return (M).
17: **End**

operations used by the main algorithms in our proposed scheme. In order to evaluate the complexity we have used some parameters that describe the time or cost requirements for the basic steps that have been implemented in our proposed scheme. Table 5.1 depicts these time complexities for our secure clustering based data dissemination scheme. The terms T_{exp}, T_{rnd}, T_{hash}, T_{sym} and T_{mac} describe the number of group operations such as exponentiation , random numbers generation, hashing, symmetric cypher operation and mac operations required for the encryption/decryption scheme respectively.

Table 5.1: Complexity of various algorithms used in secure clustering scheme.

Algorithm	**Complexity**
Vehicle Neighborhood Detection	$\theta(n^2)$
Vehicular Trust Value Estimation	$O((\frac{T}{t}) * n)$
Cluster Head Election	$\log_k(n)$
Sender Node Encryption	$2T_{exp} + T_{rnd} + T_{hash} + T_{sym} + T_{mac}$
Receiver Node Decryption	$T_{exp} + T_{hash} + T_{sym} + T_{mac}$

5.4 Results and Discussion

To evaluate the performance of the proposed clustering protocol, ns-2 [178] simulator is used. Vehicular mobility traces are generated through SUMO [148] traffic simulator. The network scenario used in our simulation consists of vehicles moving in urban scenario with the road map based on Chandigarh city as shown in Figure 5.5. The map consist of bidirectional lanes where the number of lanes in each direction varies between 1 to 3. The vehicular moment for our clustering protocol was based on Manhattan [149] mobility model.

Figure 5.5: Road map depicting simulation scenario used in secure clustering scheme

This division of bidirectional lanes helps to achieve reduced overhead across the network thereby facilitating better V2V communications in a cloud based environment. All simulations were done with a different set of vehicles which move along the roads in

Table 5.2: Simulation parameters required for secure clustering protocol.

Parameters	Value
Number of Vehicles	100-300
Clustering Duration	10-120 sec
Transmission range	300 m
Permissible lane speed	10-30 m/s
Number of bi-direction lanes	2,4
Vehicle acceleration range	0.8 m/s^2
Vehicle deceleration range	4.5 m/s^2
MAC protocol	IEEE 802.11p
Simulation Time	750 sec

Chandigarh city starting from different starting points, with these vehicles converging at traffic light points and round abouts.

The total simulation time has been kept at 750 seconds and new vehicles are allowed to enter the network by using Poisson Distribution till the upper limit of vehicles (100, 200, 300 vehicles) for that simulation instance is not reached. Thereafter, the number of vehicles is kept constant with a new vehicle being allowed to enter, only if some old vehicle leaves the network. This then, helps to create a real time vehicle scenario with vehicles continuously entering and leaving the network.

The values of main parameters used in the simulation are summarized in Table 5.2.

Since we have considered periodic re-clustering for cluster formation, the re-clustering duration is an important parameter that has been used in evaluating the performance of created clusters along with the conventional parameters such as delay, throughput, packet delivery ratio, and clustering overhead that is used to measure the stability of the selected clusters. The proposed scheme has also considered the percentage of secure vehicles and evaluating the security provided through the lifetime clustering scheme. The performance of the proposed clustering algorithm is evaluated based on the following performance metrics:

5.4.1 Cluster Head Selection

Cluster Head Density may be defined as the number CHs selected from secure nodes compared to the total number of nodes in the network. Figure 5.6(a) shows the variation in number of CHs in terms of re-clustering duration. Three different scenarios were considered by considering 100, 200, 300 number of vehicles in the network. Figure 5.6(a)

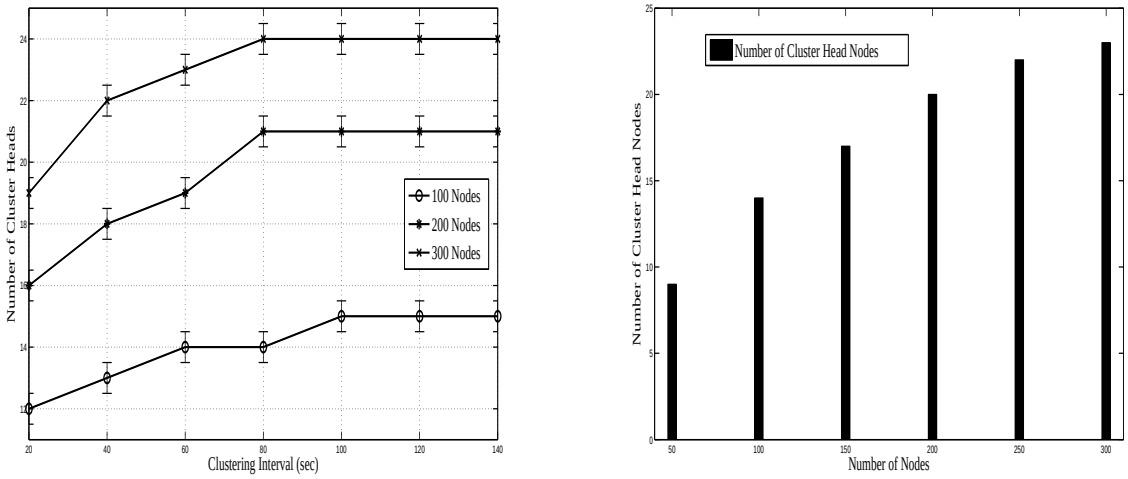


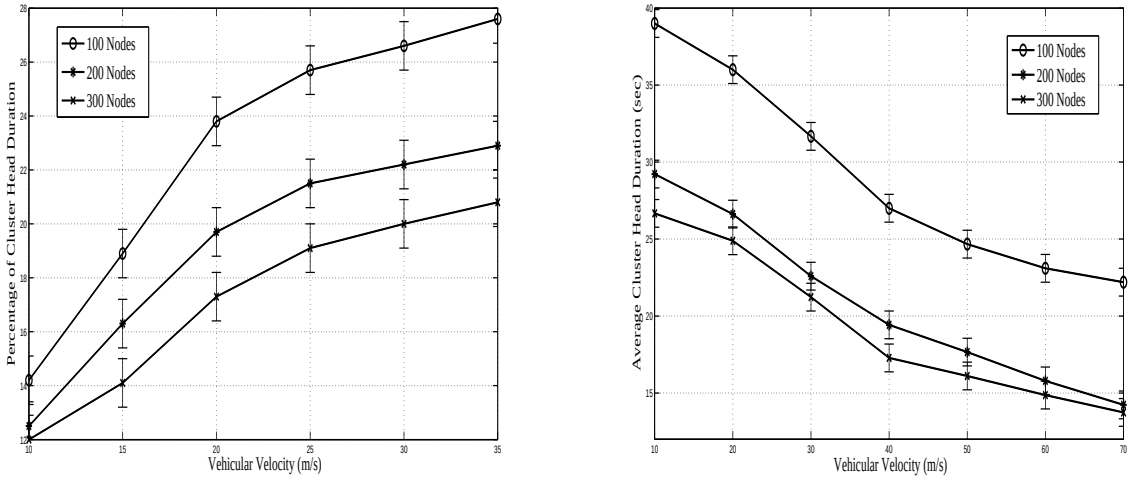
Figure 5.6: (a) Number of CHs with respect to varying clustering interval (b) Percentage of CH nodes with respect to total number of vehicles in secure clustering scheme.

shows that the number of CHs initially increases with an increase in clustering interval. However, after the clustering interval reaches 80 secs, the density of CHs shows only marginal variations and reaches an almost constant values as we further increase the time interval for re-clustering. This could be due to the fact that smaller re-clustering intervals results in clusters being created more frequently. As the prediction interval increases, the average distance between the vehicle also grows and this necessitates the creation of more CH's. Thereafter, there is very small variation in the number of CHs for any further increase in the CH clustering interval, due to the inter-vehicle distance also stabilizing to almost a constant value for all the nodes.

Figure 5.6(b) depicts the relationship between the number of CH vehicles with respect to the total number of vehicles. The number of CH increases with the number of vehicles. Since the CH vehicles are elected based on CDS technique so there are relatively small number of CHs when the total number of vehicles in the network are less. As the total number of vehicles are increased, there is an increase in the number of CH vehicles also. However, the percentage increase in the number of CHs shows lesser growth as the total number of vehicles are increased beyond zero vehicles. This indicates that when total number of vehicles are high, the proposed scheme achieves better CH selection.

5.4.2 Cluster Stability

To measure the effectiveness of the proposed clustering scheme, we have evaluate the stability of the clusters in terms of percentage of the CH lifetime and average CH duration. Average CH duration is defined as the continuous time duration for which a vehicle acts



(a) **(b)**

Figure 5.7: (a) Percentage of cluster head duration with respect to vehicle velocity (b) Average cluster head duration with respect to velocity in secure clustering scheme.

as a CH. Figure 5.7(a) shows the variation in the CH duration in terms of vehicular velocity. At lower speeds, the clusters display better stability and this results in higher average duration of CH. However, as the average vehicular speeds are increased, the vehicles are in contact with their neighboring vehicle for comparatively a smaller duration. This impacts the CH duration in the proposed scheme. The percentage of CH also has comparatively larger value when the number of nodes are less. Thus a higher number of nodes results in higher vehicle density that then provides more stable clusters.

In Figure 5.7(b), average CH duration is varied with respect to total number of vehicles. It is defined as the total percentage of time a vehicle acts as a cluster. When the number of vehicles are less, the vehicular density is also low which increases the percentage or chances of vehicles to be elected as CHs. As the number of nodes are increased, the density also becomes high. The high density also impacts the relative distance between the vehicles in a cluster and consequently more vehicles are within the transmission range of a CH. Thus, the overall percentage of CH vehicles also decreases in the network. Average CH duration also indicates the average duration of clustering that exists between vehicles in the vehicular environment.

As shown in Figure 5.7(b), average CH duration is reduced as we increase the vehicular velocity. This is because at higher speed, there is more frequent re-clustering which affects the cluster stability. Thus, more cluster reformation messages are broadcasted by the vehicles to the vehicular cloud as the vehicle speed is increased. Figure 5.8 also describes the comparative difference in average cluster duration, when clusters are created with and without the use of vehicular clouds.

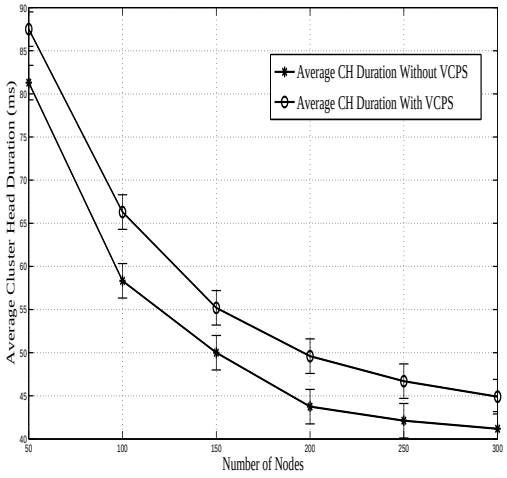


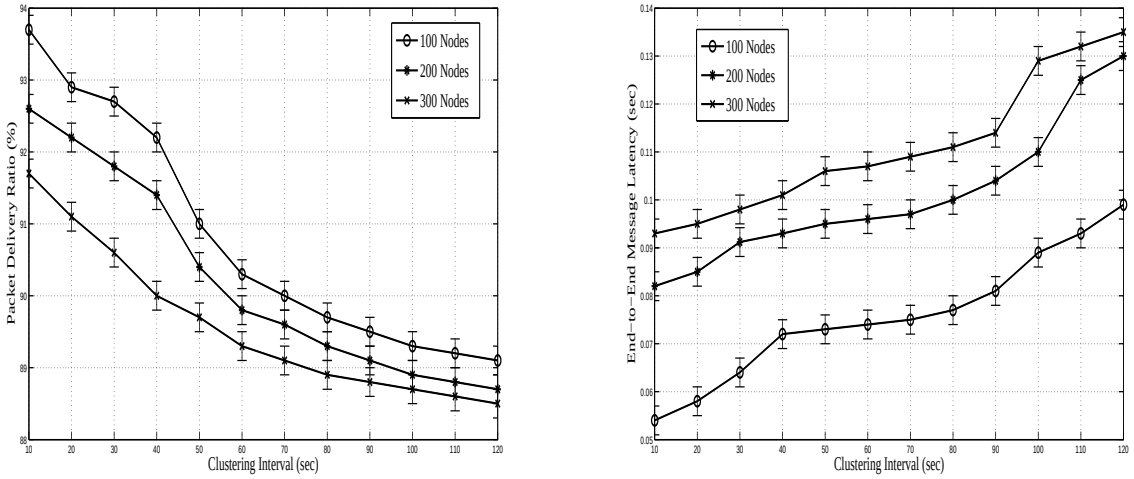
Figure 5.8: Comparision of cluster head duration with respect to number of nodes with and without vehicular cloud in secure clustering scheme

The above results indicate the comparable performance of the proposed scheme with respect to existing vehicular scheme and number of vehicles. However, due to vehicles continuously leaving and joining the network during simulation, the overall cluster stability shows minor variations. Since, the current state of the vehicles depends on the behavior of all the vehicles during previous iterations, the system performance continuously varies depending upon the security state and other transmission characteristics of vehicles in VANET.

5.4.3 Cluster Transmission Efficiency

For evaluating the efficiency of data dissemination for our clustering scheme, PDR and end-to-end message latency are considered. The variation in PDR of the proposed scheme in terms of clustering frequency is depicted in Figure 5.9(a). The value of PDR decreases as the clustering duration is increased. The increased re-clustering interval causes higher message overhead which decreases the value of PDR. The number of nodes also affects the PDR, i.e., the values of PDR decreases as the number of nodes are increased from 100 to 300.

End-to-End Latency is defined as the total time taken by a packet to reach its destination. In Figure 5.8(b), variation in end-to-end latency is compared against the clustering duration for different number of vehicles. Figure 5.9(b) shows that the latency increases linearly as the clustering frequency is decreased. This is because the message overhead is more as clustering interval is increased. Consequently, the high message overhead also increases the value of latency. The value of latency increases marginally for the variation in clustering duration between 10 to 90 secs. Thereafter, the latency grows at a higher rate



(a) **(b)**

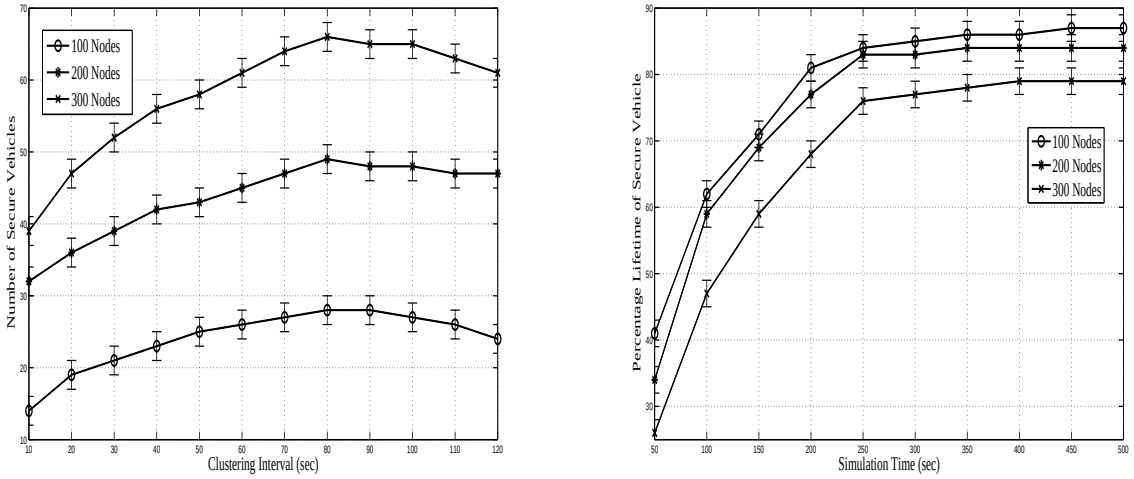
Figure 5.9: (a) Packet delivery ratio as a function of clustering interval (b) End-to-End latency with respect to prediction interval in secure clustering scheme.

for values of cluster duration greater than 90 secs and then reaches an almost constant value.

5.4.4 Cluster Security

To evaluate the performance of clustering scheme in terms of security, we have considered number of secure vehicles and percentage lifetime of secure vehicles with respect to the clustering intervals and overall simulation time respectively. The percentage lifetime of secure vehicle is delivered as the average duration of time, vehicles are in the secure state relative to the lifetime of the CH vehicles. Figure 5.10(a) displays the number of secure vehicles as a function of the clustering interval. The number of vehicles in the secure state initially increase and then reach a maximum value at the clustering duration of 80 secs after which they start decreasing. This is due to the fact that larger clustering intervals help to create a relatively larger pool of secure vehicles. However, after reaching its maxima, any further increase in clustering interval results in creation of an additional network overhead due to higher transmission delays thereby, affecting the duration of vehicles which remain in secure state.

In Figure 5.10(b), the percentage lifetime of secure vehicles is varied with respect to total simulation time. The lifetime of secure vehicles, initially grows at a fast rate for 100, 200 and 300 vehicles and then stabilizes to a constant value. This can be attributed to the fact that the number of secure vehicles are initially small and after every periodic cluster formation this number increases. However, once the vehicle reaches the secure level, the percentage lifetime of secure vehicle also starts varying within a small range. Another



(a) **(b)**

Figure 5.10: (a) Number of secure vehicles as function of clustering interval (b) Percentage lifetime of secure vehicles as a function of simulation time in secure clustering scheme.

reason for this behavior is that the values of trustworthiness weights have been varied in small values and thus once a vehicle enters a secure state, it remains in that state for a larger duration, thereby providing higher lifetime for vehicles in secure state.

We have compared security feasibility of the proposed scheme with SKCD algorithm [136]. This is because both SKCD and the proposed scheme are based on clustering based techniques for enhancing security of VANETs. As shown in Table 5.3, both SKCD and the proposed scheme use symmetric cryptography for key sharing that provides faster encryption and decryption as compared to asymmetric schemes. However since the proposed scheme is based on ECC, key generation is done in asymmetric mode. This provides more efficient and same level of security as provided by keys of smaller size used in other existing schemes.

As the routing is performed by CHs through connected dominating sets, the proposed scheme does not require the sending vehicle to broadcast the message to all vehicles that are within its transmission range. This helps to reduce the communication overheads thereby providing increased efficiency.

Table 5.3: Comparative analysis of secure clustering scheme with existing cluster based scheme.

Scheme	**Proposed**	**[136]**
Transmission Range	Dynamic	Dynamic
Cryptography Scheme in Communication	Symmetric/Asymmetric	Symmetric
Privacy Preservation Scheme Used	ECC	HMAC
Distribution of Keys	Yes	Yes
Authentication Phase	No	No
Use of Certification Authority	No	Yes
Broadcasting of Keys	No	No
Key Generation	Asymmetric	Symmetric
Key Sharing	Symmetric	Symmetric

5.5 Conclusion

VANETs provide a number of seamless services to the end users having high mobility in vehicular environment. During message dissemination across the different vehicles, the information may be leaked which may degrade the performance of the various solutions in this environment. To overcome this difficulty, in this chapter, we have proposed a novel connecting dominant set based trust evaluation and computation algorithms for maintaining trust at various levels. A trust computation metric is defined to compute the trust between different peers in this environment. Algorithms for cluster formation and trust computation among the peers are also implemented in this environment. The performance of the proposed scheme is evaluated using different evaluation metrics in this environment. With respect to the selected metrics, the performance of the proposed scheme is found satisfactory.

Chapter 6

Conclusion and Future Scope

This chapter concludes the research work on secure data Clustering in VANETs presented in this thesis. It also discusses open research problems along with future research scope.

6.1 Conclusion

Due to high velocity of vehicles and constant topological changes, clustering is one of the most difficult tasks to be performed in this environment. In this work, we have proposed vehicular clustering schemes thereby improving the communication capabilities of VANETs. These vehicular clustering and security enhancing schemes based on clustering are developed on mobile nodes in VANETs, with reference to standard vehicle mobility models. An efficient grouping of vehicles into clusters provides improved network stability and also ensures reliable communication through a centralized coordinator within a cluster. This also partitions the network into a hierarchical structure that provides better resource utilization for ITS based applications. However, an unstable cluster may result in frequent cluster reorganization that may undermine the effectiveness of whole network. Following objectives are discussed and analyzed in the thesis.

First, we discuss the classification of clustering schemes for VANETs. After reviewing a wide range of available literature which helped in preparing adetailed analysis from the various research proposals, few gaps in the existing literature where identified which required further investigation. After presenting a review of clustering techniques used in VANETs, we present a predictive clustering scheme, which fulfills most of the requirements for an effective clustering algorithm. We have identified prediction of the vehicles future positions as the main metric to design the new predictive clustering scheme along with a novel secure clustering scheme based on a new trust metric for efficient data dissemination in VANETs.

We have assumed an urban traffic roadside scenario with every road having multiple lanes and several traffic entry and exit points. We have also assumed that the flow of traffic

is unidirectional and arrival and departure of vehicles from the road are modeled using a real traffic scenario. The proposed protocol addressed the issues such as high reliability of message transmission, and high throughput with minimum delay in VANETs. We have then proposed an average predictive variation algorithm which predicts the average variation of the vehicles on the road to overcome the challenge of small duration of cluster membership of the vehicles on the road. Also, future mobility of the vehicles are estimated so that proactive measures can be taken for reliable message transmission on the road. The proposed scheme was evaluated in different network conditions by varying the cluster size and number of vehicles on the road. The results obtained show that the predictive clustering scheme is effective in maintaining high throughput, lower end-to-end delay and high probability of transmissions.

In the second part of this thesis, we proposed an LA-based hybrid predictive clustering scheme to increase the reliability of the message delivery. LA are assumed to be deployed on the vehicles and future mobility of the vehicles is estimated using which decisions about the clustering are made in the proposed scheme and CHs are elected using connected dominating set. LA based predictive clustering technique is compared with existing predictive clustering scheme where the prediction accuracy was not done to its full potential. Based upon the rewards and penalties obtained, future actions to be taken are estimated in the game. The performance of the proposed scheme was evaluated using various metrics where its performance was found satisfactory with respect to the existing scheme.

We also proposed, a cloud based distributed IDS scheme for VANETs. The proposed scheme is deployed on clusters having a number of distributed algorithms which makes the working of IDS as dynamic and distributive. As observed in simulation results, the clustering approach used in this work helps in improving the performance of the proposed IDS scheme. Through the clustering process, only those nodes are selected for the leadership that performs their operations for longer duration without leaving the cluster. Unlike other approaches, the intrusion detection task is not dependent on a single node but, a group of nodes, which forms a leaders that are responsible for IDS operations. The simulation results show that due to the presence of leadership, the IDS mechanism continues to work even in the presence of large number of malicious nodes in the network. The proposed mechanism is also adaptable to the dynamic changes in the topology of VANET like nodes joining and leaving the network. Moreover, the proposed IDS scheme has also considered intrusion and fault occurrences in its own components. The simulation results show that in various scenarios, when the faulty and malicious nodes crosses the τ limit, the IDS continues to operate correctly. The proposed technique was tested by simulating two types of attacks.

Extending this work and focusing on another prominent issue of security, with the

objective of providing seamless and secure service to the end users we have proposed a secure clustering based efficient data dissemination scheme for VANETs. During message dissemination across the different vehicles, the information may be leaked which may degrade the performance of the various solutions in this environment. To overcome this difficulty, connecting dominant set based trust evaluation and computation algorithms for maintaining trust at various levels have been proposed. A trust computation metric is defined to compute the trust between different peers in this environment. Algorithms for cluster formation and trust computation among the peers have also been developed.. The performance of the proposed scheme is evaluated using different evaluation metrics in this environment. With respect to the selected metrics, the performance of the proposed scheme is found superior in comparison to other existing schemes.

The main novelty factors of this research work can be summarized as follows,

- **Predictive Clustering Model** The constrained mobility patterns, high density along with high speed make clustering a challenging task. Most of the existing schemes for clustering in VANETs have focused on divergent vehicular characteristics that affects their performance. Hence, by predicting the position of vehicles by estimating their future mobility, assists in creating clusters which are more stable which in turn, improves the network performance.

- **Secure Vehicular Clustering Framework** : As vehicles that have a high trust values and strong connectivity among them are selected as CHs so, the proposed clustering approach achieves higher efficiency with respect to the data dissemination. For maintaining the connectivity among the vehicles at various levels, Connected Dominating Set is used to construct CHs among the vehicles to provide enhanced cluster stability and higher security.

6.2 Future Research Scope

6.2.1 Secure Clustering

In the predictive clustering scheme, the future position of vehicles is measured based on vehicular dynamics. The accuracy of this process can be improved further by incorporating more precise computational techniques such as fuzzy logic and machine learning algorithms. The stability of the clustering mechanism can also be enhanced by considering hybrid techniques based on intelligent driver behavior and intelligent density estimation techniques.

In the future, we will explore more features with respect to the secure data dissemination of created clusters by including more infrastructure based security mechanisms such as IDS or Group Signature scheme.

6.2.2 Future Security Enhancements

In future, different set of attacks and their variations can be used to check the viability of the proposed IDS scheme. The future work will also focus on using data set of real-time attacks or real network traffic to evaluate the functioning of the proposed IDS.

The clustering technique used in this research has used only relative velocity to select leadership and cluster head. More efficient cryptographic schemes can be used to secure IDS components.

Bibliography

[1] Magnus Frodigh, Per Johansson, and Peter Larsson. Wireless ad hoc networking: the art of networking without a network. *Ericsson Review*, 4(4):249, 2000.

[2] David B Johnson and David A Maltz. Dynamic source routing in ad hoc wireless networks. In *Mobile computing*, pages 153–181. Springer, 1996.

[3] Stephan Olariu and Michele C Weigle. *Vehicular networks: from theory to practice*. Crc Press, 2009.

[4] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

[5] Levente Buttyán and Jean-Pierre Hubaux. Stimulating cooperation in self-organizing mobile ad hoc networks. *Mobile Networks and Applications*, 8(5):579–592, 2003.

[6] Hassnaa Moustafa and Yan Zhang. *Vehicular networks: techniques, standards, and applications*. Auerbach publications, 2009.

[7] Zong Da Chen, HT Kung, and Dario Vlah. Ad hoc relay wireless networks over moving vehicles on highways. In *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 247–250. ACM, 2001.

[8] James A Misener and Steven E Shladover. Path investigations in vehicle-roadside cooperation and safety: A foundation for safety and vehicle-infrastructure integration research. In *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, pages 9–16. IEEE, 2006.

[9] Sed S Saad. Its in japan, a different approach to transportation policy. *World Review of Intermodal Transportation Research*, 1(1):45–54, 2006.

[10] Roberto Baldessari, Bert Bödekker, Matthias Deegener, Andreas Festag, Walter Franz, C Christopher Kellum, Timo Kosch, Andras Kovacs, Massimiliano Lenardi, Cornelius Menig, et al. Car-2-car communication consortium-manifesto. 2007.

[11] Scott E Carpenter. Inter-vehicle communications (ivc): Current standards and. 2013.

[12] Saif Al-Sultan, Moath M Al-Doori, Ali H Al-Bayatti, and Hussien Zedan. A comprehensive survey on vehicular ad hoc network. *Journal of network and computer applications*, 37:380–392, 2014.

[13] Ghassan MT Abdalla, Mosa Ali Abu-Rgheff, and Sidi Mohammed Senouci. Current trends in vehicular ad hoc networks. In *Proceedings of UBIROADS workshop*, 2007.

[14] Brian P Crow, Indra Widjaja, Jeong Geun Kim, and Prescott T Sakai. Ieee 802.11 wireless local area networks. *Communications Magazine, IEEE*, 35(9):116–126, 1997.

[15] Raj Bala and C Rama Krishna. Performance analysis of topology based routing in a vanet. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*, pages 2180–2184. IEEE, 2014.

[16] Saleh Yousefi, Mahmoud Siadat Mousavi, and Mahmood Fathy. Vehicular ad hoc networks (vanets): challenges and perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761–766. IEEE, 2006.

[17] Jijun Yin, Tamer ElBatt, Gavin Yeung, Bo Ryu, Stephen Habermas, Hariharan Krishnan, and Timothy Talty. Performance evaluation of safety applications over dsrc vehicular ad hoc networks. In *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 1–9. ACM, 2004.

[18] Yi Qian and Nader Moayeri. Design of secure and application-oriented vanets. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 2794–2799. IEEE, 2008.

[19] Eiichi Taniguchi and Hiroshi Shimamoto. Intelligent transportation system based dynamic vehicle routing and scheduling with variable travel times. *Transportation Research Part C: Emerging Technologies*, 12(3):235–250, 2004.

[20] Yi Qian and Nader Moayeri. Design of secure and application-oriented vanets. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 2794–2799. IEEE, 2008.

[21] Theodore L Willke, Patcharinee Tientrakool, and Nicholas F Maxemchuk. A survey of inter-vehicle communication protocols and their applications. *Communications Surveys & Tutorials, IEEE*, 11(2):3–20, 2009.

[22] Tamer Nadeem, Pravin Shankar, and Liviu Iftode. A comparative study of data dissemination models for vanets. In *Mobile and Ubiquitous Systems-Workshops, 2006. 3rd Annual International Conference on*, pages 1–10. IEEE, 2006.

[23] Wai Chen, Ratul K Guha, Taek Jin Kwon, John Lee, and Yuan-Ying Hsu. A survey and challenges in routing and data dissemination in vehicular ad hoc networks. *Wireless Communications and Mobile Computing*, 11(7):787–795, 2011.

[24] Moe Rahnema. Overview of the gsm system and protocol architecture. *Communications Magazine, IEEE*, 31(4):92–100, 1993.

[25] R Bhakthavathsalam and Starakjeet Nayak. Operational inferences on vanets in 802.16 e and 802.11 p with improved performance by congestion alert. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pages 467–471. IEEE, 2011.

[26] Yu Wang and Fan Li. Vehicular ad hoc networks. In *Guide to wireless ad hoc networks*, pages 503–525. Springer, 2009.

[27] Yue Liu, Jun Bi, and Ju Yang. Research on vehicular ad hoc networks. In *Control and Decision Conference, 2009. CCDC'09. Chinese*, pages 4430–4435. IEEE, 2009.

[28] Hannes Hartenstein and Kenneth P Laberteaux. A tutorial survey on vehicular ad hoc networks. *Communications Magazine, IEEE*, 46(6):164–171, 2008.

[29] Ming-Fong Tsai, Yung-Cheng Chao, Lien-Wu Chen, Naveen Chilamkurti, and Seungmin Rho. Cooperative emergency braking warning system in vehicular networks. *EURASIP Journal on Wireless Communications and Networking*, 2015(1): 1–14, 2015.

[30] Yasser Toor, Paul Muhlethaler, and Anis Laouiti. Vehicle ad hoc networks: applications and related technical issues. *Communications Surveys & Tutorials, IEEE*, 10(3):74–88, 2008.

[31] Saif Al-Sultan, Moath M Al-Doori, Ali H Al-Bayatti, and Hussien Zedan. A comprehensive survey on vehicular ad hoc network. *Journal of network and computer applications*, 37:380–392, 2014.

[32] A Bruce McDonald and Taieb F Znati. A mobility-based framework for adaptive clustering in wireless ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 17(8):1466–1487, 1999.

[33] Dilip Kumar, Trilok C Aseri, and RB Patel. Eehc: Energy efficient heterogeneous clustered scheme for wireless sensor networks. *Computer Communications*, 32(4): 662–667, 2009.

[34] Rasmeet S Bali, Neeraj Kumar, and Joel JPC Rodrigues. Clustering in vehicular ad hoc networks: taxonomy, challenges and solutions. *Vehicular communications*, 1 (3):134–152, 2014.

[35] Saleha Mubarak AlMheiri and Hend Saeed AlQamzi. Manets and vanets clustering algorithms: a survey. In *GCC Conference and Exhibition (GCCCE), 2015 IEEE 8th*, pages 1–6. IEEE, 2015.

[36] Jesús Téllez Isaac, Sherali Zeadally, and José Sierra Camara. Security attacks and solutions for vehicular ad hoc networks. *Communications, IET*, 4(7):894–903, 2010.

[37] Yeongkwun Kim and Injoo Kim. Security issues in vehicular networks. In *Information Networking (ICOIN), 2013 International Conference on*, pages 468–472. IEEE, 2013.

[38] Maxim Raya and Jean-Pierre Hubaux. The security of vanets. In *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, pages 93–94. ACM, 2005.

[39] Peng Zeng, Zhenfu Cao, K-KR Choo, and Shengbao Wang. On the anonymity of some authentication schemes for wireless communications. *IEEE Communications Letters*, 13(3):170–171, 2009.

[40] Sherali Zeadally, Ray Hunt, Yuh-Shyan Chen, Angela Irwin, and Aamir Hassan. Vehicular ad hoc networks (vanets): status, results, and challenges. *Telecommunication Systems*, 50(4):217–241, 2012.

[41] João AFF Dias, Joel JPC Rodrigues, Feng Xia, and Constandinos X Mavromoustakis. A cooperative watchdog system to detect misbehavior nodes in vehicular delay-tolerant networks. *IEEE Transactions on Industrial Electronics*, 62(12):7929–7937, 2015.

[42] Christian Tchepnda, H Moustafa, H Labiod, and G Bourdon. A panorama on vehicular networks security. In *International Workshop on Interoperable Vehicles (IOV2008) co-located with the Internet of Things 2008 conference*, 2008.

[43] Hassnaa Moustafa and Gilles Bourdon. Vehicular networks deployment view: applications, deployment architectures and security means. *Ubiquitous Computing and Communication Journal, special issue on Ubiquitous Roads*, 2008.

[44] Anand Mudgerikar and Manik Lal Das. Secure multicast using ipsec and multi-party key computation. *International Journal of Internet Technology and Secured Transactions*, 5(2):149–162, 2014.

[45] Elyes Ben Hamida, Hassan Noura, and Wassim Znaidi. Security of cooperative intelligent transport systems: Standards, threats analysis and cryptographic countermeasures. *Electronics*, 4(3):380–423, 2015.

[46] Donghyun Kim, Yiwei Wu, Yingshu Li, Feng Zou, and Ding-Zhu Du. Constructing minimum connected dominating sets with bounded diameters in wireless networks. *Parallel and Distributed Systems, IEEE Transactions on*, 20(2):147–157, 2009.

[47] Sudipto Guha and Samir Khuller. Approximation algorithms for connected dominating sets. *Algorithmica*, 20(4):374–387, 1998.

[48] Javad Akbari Torkestani and Mohammad Reza Meybodi. Finding minimum weight connected dominating set in stochastic graph based on learning automata. *Information Sciences*, 200:57–77, 2012.

[49] Jie Wu, Wei Lou, and Fei Dai. Extended multipoint relays to determine connected dominating sets in manets. *IEEE transactions on computers*, 55(3):334–347, 2006.

[50] Fei Dai and Jie Wu. An extended localized algorithm for connected dominating set formation in ad hoc wireless networks. *IEEE transactions on parallel and distributed systems*, 15(10):908–920, 2004.

[51] Ismail Salhi, Mohamed Cherif, and S Senouci. Data collection in vehicular networks. In *Proc. ASN symposium*, 2007.

[52] Zhigang Wang, Lichuan Liu, MengChu Zhou, and Nirwan Ansari. A position-based clustering technique for ad hoc intervehicle communication. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 38(2):201–208, 2008.

[53] Nitin Maslekar, Mounir Boussedjra, Joseph Mouzna, and Houda Labiod. A stable clustering algorithm for efficiency applications in vanets. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, pages 1188–1193. IEEE, 2011.

[54] Grzegorz Wolny. Modified dmac clustering algorithm for vanets. In *Systems and Networks Communications, 2008. ICSNC'08. 3rd International Conference on*, pages 268–273. IEEE, 2008.

[55] Stefano Basagni. Distributed clustering for ad hoc networks. In *Parallel Architectures, Algorithms, and Networks, 1999.(I-SPAN'99) Proceedings. Fourth InternationalSymposium on*, pages 310–315. IEEE, 1999.

[56] Wei Fan, Yan Shi, Shanzhi Chen, and Longhao Zou. A mobility metrics based dynamic clustering algorithm for vanets. In *Communication Technology and Application (ICCTA 2011), IET International Conference on*, pages 752–756. IET, 2011.

[57] Nitin Maslekar, Joseph Mouzna, Houda Labiod, Manoj Devisetty, and Ming-Chyi Pai. Modified c-drive: Clustering based on direction in vehicular environment. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 845–850. IEEE, 2011.

[58] Wang Xiaonan and Qian Huanyan. Constructing a vanet based on cluster chains. *International Journal of Communication Systems*, 27(11):2497–2517, 2014.

[59] Behnam Hassanabadi, Christine Shea, L Zhang, and Shahrokh Valaee. Clustering in vehicular ad hoc networks using affinity propagation. *Ad Hoc Networks*, 13:535–548, 2014.

[60] António Fonseca and Teresa Vazão. Applicability of position-based routing for vanet in highways and urban environment. *Journal of Network and Computer Applications*, 36(3):961–973, 2013.

[61] Farhan Ahammed, Javid Taheri, and Albert Zomaya. Lica: robust localization using cluster analysis to improve gps coordinates. In *Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications*, pages 39–46. ACM, 2011.

[62] Daxin Tian, Yunpeng Wang, Guangquan Lu, and Guizhen Yu. A vanets routing algorithm based on euclidean distance clustering. In *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, volume 1, pages V1–183. IEEE, 2010.

[63] Mildred M Caballeros Morales, Choong Seon Hong, and Young-Cheol Bang. An adaptable mobility-aware clustering algorithm in vehicular networks. In *Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific*, pages 1–6. IEEE, 2011.

[64] Yan Zhang, Jim Mee Ng, and Chor Ping Low. A distributed group mobility adaptive clustering algorithm for mobile ad hoc networks. *Computer Communications*, 32 (1):189–202, 2009.

[65] RA Santos, RM Edwards, and NL Seed. Inter vehicular data exchange between fast moving road traffic using an ad-hoc cluster-based location routing algorithm and 802.11 b direct sequence spread spectrum radio. In *PostGraduate Networking Conference*, 2003.

[66] Peng Fan. Improving broadcasting performance by clustering with stability for inter-vehicle communication. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 2491–2495. IEEE, 2007.

[67] Mohammad S Almalag and Michele C Weigle. Using traffic flow for cluster formation in vehicular ad-hoc networks. In *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, pages 631–636. IEEE, 2010.

[68] Zhenxia Zhang, Azzedine Boukerche, and Richard Pazzi. A novel multi-hop clustering scheme for vehicular ad-hoc networks. In *Proceedings of the 9th ACM international symposium on Mobility management and wireless access*, pages 19–26. ACM, 2011.

[69] Liren Zhang and Hesham El-Sayed. A novel cluster-based protocol for topology discovery in vehicular ad hoc network. *Procedia Computer Science*, 10:525–534, 2012.

[70] Tracy Camp, Jeff Boleng, and Lucas Wilcox. Location information services in mobile ad hoc networks. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 5, pages 3318–3324. IEEE, 2002.

[71] Weiwei Li, Ali Tizghadam, and Alberto Leon-Garcia. Robust clustering for connected vehicles using local network criticality. In *Communications (ICC), 2012 IEEE International Conference on*, pages 7157–7161. IEEE, 2012.

[72] Efi Dror, Chen Avin, and Zvi Lotker. Fast randomized algorithm for hierarchical clustering in vehicular ad-hoc networks. In *Ad Hoc Networking Workshop (Med-Hoc-Net), 2011 The 10th IFIP Annual Mediterranean*, pages 1–8. IEEE, 2011.

[73] Peng Fan, James G Haran, John Dillenburg, and Peter C Nelson. Cluster-based framework in vehicular ad-hoc networks. In *Ad-hoc, mobile, and wireless networks*, pages 32–42. Springer, 2005.

[74] Hang Su and Xi Zhang. Clustering-based multichannel mac protocols for qos provisionings over vehicular ad hoc networks. *Vehicular Technology, IEEE Transactions on*, 56(6):3309–3323, 2007.

[75] Luciano Bononi and Marco Di Felice. A cross layered mac and clustering scheme for efficient broadcast in vanets. In *Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE International Conference on*, pages 1–8. IEEE, 2007.

[76] Luciano Bononi, Marco Di Felice, Lorenzo Donatiello, Danilo Blasi, Vincenzo Cacace, Luca Casone, and Salvatore Rotolo. Design and performance evaluation of cross layered mac and clustering solutions for wireless ad hoc networks. *Performance Evaluation*, 63(11):1051–1073, 2006.

[77] Yazhi Liu, Jianwei Niu, Jian Ma, Lei Shu, Takahiro Hara, and Wendong Wang. The insights of message delivery delay in vanets with a bidirectional traffic model. *Journal of Network and Computer Applications*, 36(5):1287–1294, 2013.

[78] Subir Biswas, Jelena Mišić, and Vojislav Mišić. Id-based safety message authentication for security and trust in vehicular networks. In *Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference On*, pages 323–331. IEEE, 2011.

[79] Hassan Omar, Weihua Zhuang, and Li Li. Vemac: A tdma-based mac protocol for reliable broadcast in vanets. *Mobile Computing, IEEE Transactions on*, 12(9): 1724–1736, 2013.

[80] Y Gunter, Bernhard Wiegel, and Hans Peter Großmann. Medium access concept for vanets based on clustering. In *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*, pages 2189–2193. IEEE, 2007.

[81] Mohammad S Almalag, Stephan Olariu, and Michele C Weigle. Tdma cluster-based mac for vanets (tc-mac). In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, pages 1–6. IEEE, 2012.

[82] Carlos T Calafate, Giancarlo Fortino, Sascha Fritsch, Janio Monteiro, Juan-Carlos Cano, and Pietro Manzoni. An efficient and robust content delivery solution for ieee 802.11 p vehicular environments. *Journal of Network and Computer Applications*, 35(2):753–762, 2012.

[83] Ismail Salhi, Mohamed Oussama Cherif, and Sidi Mohammed Senouci. A new architecture for data collection in vehicular networks. In *Communications, 2009. ICC'09. IEEE International Conference on*, pages 1–6. IEEE, 2009.

[84] Wang-Rong Chang, Hui-Tang Lin, and Bo-Xuan Chen. Trafficgather: An efficient and scalable data collection protocol for vehicular ad hoc networks. In *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, pages 365–369. IEEE, 2008.

[85] Soheila V Bana and Pravin Varaiya. Space division multiple access (sdma) for robust ad hoc vehicle communication networks. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 962–967. IEEE, 2001.

[86] Bouziane Brik, Nasreddine Lagraa, Mohamed Bachir Yagoubi, and Abderrahmane Lakas. An efficient and robust clustered data gathering protocol (cdgp) for vehicular networks. In *Proceedings of the second ACM international symposium on Design and analysis of intelligent vehicular networks and applications*, pages 69–74. ACM, 2012.

[87] Claudia Campolo, Hector Agustin Cozzetti, Antonella Molinaro, and Riccardo Scopigno. Augmenting vehicle-to-roadside connectivity in multi-channel vehicular ad hoc networks. *Journal of Network and Computer Applications*, 36(5):1275–1286, 2013.

[88] Thomas DC Little and Abhishek Agarwal. An information propagation scheme for vanets. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 155–160. IEEE, 2005.

[89] Satoshi Teshima, Tomoyuki Ohta, Eitaro Kohno, and Yoshiaki Kakuda. A data transfer scheme using autonomous clustering in vanets environment. In *Autonomous Decentralized Systems (ISADS), 2011 10th International Symposium on*, pages 477–482. IEEE, 2011.

[90] Rául Aquino Santos, Arthur Edwards, RM Edwards, and N Luke Seed. Performance evaluation of routing protocols in vehicular ad-hoc networks. *International Journal of Ad Hoc and Ubiquitous Computing*, 1(1-2):80–91, 2005.

[91] Anindya Tahsin Prodhan, Rajkumar Das, Humayun Kabir, and Gholamali C Shoja. Ttl based routing in opportunistic networks. *Journal of Network and Computer Applications*, 34(5):1660–1670, 2011.

[92] Leandros A Maglaras and Dimitrios Katsaros. Distributed clustering in vehicular networks. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, pages 593–599. IEEE, 2012.

[93] Ahmed Ahizoune and Abdelhakim Hafid. A new stability based clustering algorithm (sbca) for vanets. In *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*, pages 843–847. IEEE, 2012.

[94] Evandro Souza, Ioanis Nikolaidis, and Pawel Gburzynski. A new aggregate local mobility (alm) clustering algorithm for vanets. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010.

[95] Christine Shea, Behnam Hassanabadi, and Shahrokh Valaee. Mobility-based clustering in vanets using affinity propagation. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6. IEEE, 2009.

[96] Rituparna Ghosh and Stefano Basagni. Mitigating the impact of node mobility on ad hoc clustering. *Wireless Communications and Mobile Computing*, 8(3):295–308, 2008.

[97] Ömer Kayiş and Tankut Acarman. Clustering formation for inter-vehicle communication. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 636–641. IEEE, 2007.

[98] Francesco Chiti, Romano Fantacci, and Giovanni Rigazzi. A mobility driven joint clustering and relay selection for ieee 802.11 p/wave vehicular networks. In *Communications (ICC), 2014 IEEE International Conference on*, pages 348–353. IEEE, 2014.

[99] AlloucheYair Yair and Michael Segal. Near-optimal, reliable and self-organizing hierarchical topology in vanet. In *Proceedings of the Eighth ACM international workshop on Vehicular inter-networking*, pages 79–80. ACM, 2011.

[100] Slawomir Kukliński and Grzegorz Wolny. Density based clustering algorithm for vanets. In *Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops, 2009. TridentCom 2009. 5th International Conference on*, pages 1–6. IEEE, 2009.

[101] Mohamed Bakhouya, Jaafar Gaber, and Pascal Lorenz. An adaptive approach for information dissemination in vehicular ad hoc networks. *Journal of Network and Computer Applications*, 34(6):1971–1978, 2011.

[102] MS Kakkasageri and SS Manvi. Multiagent driven dynamic clustering of vehicles in vanets. *Journal of Network and Computer Applications*, 35(6):1771–1780, 2012.

[103] Ameneh Daeinabi, Akbar Ghaffar Pour Rahbar, and Ahmad Khademzadeh. Vwca: An efficient clustering algorithm in vehicular ad hoc networks. *Journal of Network and Computer Applications*, 34(1):207–222, 2011.

[104] Sheng-Shih Wang and Yi-Shiun Lin. Passcar: A passive clustering aided routing protocol for vehicular ad hoc networks. *Computer Communications*, 36(2):170–179, 2013.

[105] MS Kakkasageri and SS Manvi. Connectivity and mobility aware dynamic clustering in vanets. *International Journal of Future Computer and Communication*, 3(1):5, 2014.

[106] Ahmed Louazani, Sidi Mohammed Senouci, and Mohammed Abderrahmane Bendaoud. Clustering-based algorithm for connectivity maintenance in vehicular ad-hoc networks. In *Innovations for Community Services (I4CS), 2014 14th International Conference on*, pages 34–38. IEEE, 2014.

[107] Taek Jin Kwon, Mario Gerla, Vijay K Varma, Melbourne Barton, and T Russell Hsing. Efficient flooding with passive clustering-an overhead-free selective forward mechanism for ad hoc/sensor networks. *Proceedings of the IEEE*, 91(8):1210–1220, 2003.

[108] Khalid Abdel Hafeez, Lian Zhao, Zaiyi Liao, and Bobby Ngok-Wah Ma. A fuzzy-logic-based cluster head selection algorithm in vanets. In *Communications (ICC), 2012 IEEE International Conference on*, pages 203–207. IEEE, 2012.

[109] Neeraj Kumar, Naveen Chilamkurti, and Jong Hyuk Park. Alca: agent learning—based clustering algorithm in vehicular ad hoc networks. *Personal and ubiquitous computing*, 17(8):1683–1692, 2013.

[110] Huixian Wang, Ren Ping Liu, Wei Ni, Wei Chen, and Iain B Collings. Vanet modeling and clustering design under practical traffic, channel and mobility conditions. *Communications, IEEE Transactions on*, 63(3):870–881, 2015.

[111] Ramon Bauza and Javier Gozálvez. Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. *Journal of Network and Computer Applications*, 36(5):1295–1307, 2013.

[112] Riccardo Crepaldi, Mehedi Bakht, and Robin Kravets. Quicksilver: application-driven inter-and intra-cluster communication in vanets. In *Proceedings of the third ACM international workshop on Mobile Opportunistic Networks*, pages 69–76. ACM, 2012.

[113] Javier Barrachina, Piedad Garrido, Manuel Fogue, Francisco J Martinez, Juan-Carlos Cano, Carlos T Calafate, and Pietro Manzoni. Veacon: A vehicular accident ontology designed to improve safety on the roads. *Journal of Network and Computer Applications*, 35(6):1891–1900, 2012.

[114] Jeremy Blum, Azim Eskandarian, and Lance Hoffman. Mobility management in ivc networks. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pages 150–155. IEEE, 2003.

[115] Sheng-Tzong Cheng, Gwo-Jiun Horng, and Chih-Lun Chou. Using cellular automata to form car society in vehicular ad hoc networks. *Intelligent Transportation Systems, IEEE Transactions On*, 12(4):1374–1384, 2011.

[116] Bidi Ying, Dimitrios Makrakis, and Hussein T Mouftah. Privacy preserving broadcast message authentication protocol for vanets. *Journal of Network and Computer Applications*, 36(5):1352–1364, 2013.

[117] Jeremy Blum and Azim Eskandarian. The threat of intelligent collisions. *IT professional*, 6(1):24–29, 2004.

[118] Maxim Raya, Adel Aziz, and Jean-Pierre Hubaux. Efficient secure aggregation in vanets. In *Proceedings of the 3rd international workshop on Vehicular ad hoc networks*, pages 67–75. ACM, 2006.

[119] S Sivagurunathan, P Subathra, V Mohan, and N Ramaraj. Authentic vehicular environment using a cluster based key management. *European Journal of Scientific Research ISSN*, 36(2):299–307, 2009.

[120] Tahani Gazdar, Abderrahim Benslimane, and Abdelfettah Belghith. Secure clustering scheme based keys management in vanets. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pages 1–5. IEEE, 2011.

[121] Klaus Plossl, Thomas Nowey, and Christian Mletzko. Towards a security architecture for vehicular ad hoc networks. In *Availability, Reliability and Security, 2006. ARES 2006. The First International Conference on*, pages 8–pp. IEEE, 2006.

[122] Tong Zhou, Romit Roy Choudhury, Peng Ning, and Krishnendu Chakrabarty. Privacy-preserving detection of sybil attacks in vehicular ad hoc networks. In *Mobile and Ubiquitous Systems: Networking & Services, 2007. MobiQuitous 2007. Fourth Annual International Conference on*, pages 1–8. IEEE, 2007.

[123] Gongjun Yan, Stephan Olariu, and Michele C Weigle. Providing vanet security through active position detection. *Computer Communications*, 31(12):2883–2897, 2008.

[124] Jonathan Petit. Analysis of ecdsa authentication processing in vanets. In *New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on*, pages 1–5. IEEE, 2009.

[125] Mainak Ghosh, Anitha Varghese, Arzad A Kherani, and Arobinda Gupta. Distributed misbehavior detection in vanets. In *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pages 1–6. IEEE, 2009.

[126] Norbert Bißmeyer, Hagen Stübing, Manuel Mattheß, Jan Peter Stotz, Julian Schütte, Matthias Gerlach, and Florian Friederici. simtd security architecture: Deployment of a security and privacy architecture in field operational tests. In *7th Embedded Security in Cars Conference (ESCAR), Düsseldorf*, 2009.

[127] Jorge Hortelano, Juan Carlos Ruiz, and Pietro Manzoni. Evaluating the usefulness of watchdogs for intrusion detection in vanets. In *Communications Workshops (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010.

[128] Albert Wasef, Rongxing Lu, Xiaodong Lin, and Xuemin Shen. Complementing public key infrastructure to secure vehicular ad hoc networks [security and privacy in emerging wireless networks]. *Wireless Communications, IEEE*, 17(5):22–28, 2010.

[129] Hagen Stübing, Attila Jaeger, Norbert Bißmeyer, Christoph Schmidt, and Sorin A Huss. Verifying mobility data under privacy considerations in car-to-x communication. In *17th ITS World Congress*, 2010.

[130] Sushmita Ruj, Marcos A Cavenaghi, Zhen Huang, Amiya Nayak, and Ivan Stojmenovic. On data-centric misbehavior detection in vanets. In *Vehicular technology conference (VTC Fall), 2011 IEEE*, pages 1–5. IEEE, 2011.

[131] Tahani Gazdar, Abderrahim Benslimane, Abderrezak Rachedi, and Abdelfettah Belghith. A trust-based architecture for managing certificates in vehicular ad hoc networks. In *Communications and Information Technology (ICCIT), 2012 International Conference on*, pages 180–185. IEEE, 2012.

[132] Esther Palomar, José M de Fuentes, Ana I González-Tablas, and Almudena Alcaide. Hindering false event dissemination in vanets with proof-of-work mechanisms. *Transportation Research Part C: Emerging Technologies*, 23:85–97, 2012.

[133] Ameneh Daeinabi and Akbar Ghaffarpour Rahbar. Detection of malicious vehicles (dmv) through monitoring in vehicular ad-hoc networks. *Multimedia tools and applications*, 66(2):325–338, 2013.

[134] Omar Abdel Wahab, Hadi Otrok, and Azzam Mourad. Vanet qos-olsr: Qos-based clustering protocol for vehicular ad hoc networks. *Computer Communications*, 36 (13):1422–1435, 2013.

[135] Nikolaos Alexiou, Marcello Laganà, Stylianos Gisdakis, Mohammad Khodaei, and Panagiotis Papadimitratos. Vespa: Vehicular security and privacy-preserving architecture. In *Proceedings of the 2nd ACM workshop on Hot topics on wireless network security and privacy*, pages 19–24. ACM, 2013.

[136] Ameneh Daeinabi and Akbar Ghaffarpour Rahbar. An advanced security scheme based on clustering and key distribution in vehicular ad-hoc networks. *Computers & Electrical Engineering*, 40(2):517–529, 2014.

[137] Atanu Mondal and Sulata Mitra. Detection and revocation of misbehaving vehicles from vanet. In *Advanced Computer and Communication Engineering Technology*, pages 767–777. Springer, 2015.

[138] Hichem Sedjelmaci and Sidi Mohammed Senouci. An accurate and efficient collaborative intrusion detection framework to secure vehicular networks. *Computers & Electrical Engineering*, 43:33–47, 2015.

[139] Auxeeliya Jesudoss, SV Kasmir Raja, and Ashraph Sulaiman. Stimulating truth-telling and cooperation among nodes in vanets through payment and punishment scheme. *Ad Hoc Networks*, 24:250–263, 2015.

[140] David Antolino Rivas, José M Barceló-Ordinas, Manel Guerrero Zapata, and Julián D Morillo-Pozo. Security on vanets: Privacy, misbehaving nodes, false information and secure data aggregation. *Journal of Network and Computer Applications*, 34(6):1942–1955, 2011.

[141] Kaouther Abrougui, Azzedine Boukerche, and Hussam Ramadan. Performance evaluation of an efficient fault tolerant service discovery protocol for vehicular networks. *Journal of Network and Computer Applications*, 35(5):1424–1435, 2012.

[142] Peng Fan, Prasad Sistla, and Peter Nelson. Theoretical analysis of a directional stability-based clustering algorithm for vanets. In *Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking*, pages 80–81. ACM, 2008.

[143] You Lu, Biao Zhou, Fei Jia, and Mario Gerla. Group-based secure source authentication protocol for vanets. In *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pages 202–206. IEEE, 2010.

[144] Zaydoun Y Rawashdeh and Syed Masud Mahmud. A novel algorithm to form stable clusters in vehicular ad hoc networks on highways. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):1–13, 2012.

[145] Jesús Téllez Isaac, Sherali Zeadally, and José Sierra Camara. Security attacks and solutions for vehicular ad hoc networks. *Communications, IET*, 4(7):894–903, 2010.

[146] Jinyuan Sun, Chi Zhang, Yanchao Zhang, and Yuguang Fang. An identity-based security system for user privacy in vehicular ad hoc networks. *Parallel and Distributed Systems, IEEE Transactions on*, 21(9):1227–1239, 2010.

[147] Huaqun Guo, Lek Heng Ngoh, Yongdong Wu, Lian Hwa Liow, Choon Hwee Kwek, Feng Tao, and Jun Jie Ang. Embedded info-security solutions for vehicular networks. In *Communications and Networking in China, 2008. ChinaCom 2008. Third International Conference on*, pages 29–33. IEEE, 2008.

[148] Michael Behrisch, Laura Bieker, Jakob Erdmann, and Daniel Krajzewicz. Sumo–simulation of urban mobility. In *The Third International Conference on Advances in System Simulation (SIMUL 2011), Barcelona, Spain*, 2011.

[149] Francisco J Martinez, Juan-Carlos Cano, Carlos T Calafate, and Pietro Manzoni. Citymob: a mobility model pattern generator for vanets. In *Communications Workshops, 2008. ICC Workshops' 08. IEEE International Conference on*, pages 370–374. IEEE, 2008.

[150] Chia-Ho Ou. A roadside unit-based localization scheme for vehicular ad hoc networks. *International Journal of Communication Systems*, 27(1):135–150, 2014.

[151] Narendra Kumar and Jong-Hyouk Lee. Peer-to-peer cooperative caching for data dissemination in urban vehicular communications. *Systems Journal, IEEE*, 8(4): 1136–1144, 2014.

[152] Neeraj Kumar, Sudip Misra, and Mohammad S Obaidat. Collaborative learning automata-based routing for rescue operations in dense urban regions using vehicular sensor networks. *Systems Journal, IEEE*, 9(3):1081–1090, 2015.

[153] Neeraj Kumar, Jong-Hyouk Lee, and Joel JPC Rodrigues. Intelligent mobile video surveillance system as a bayesian coalition game in vehicular sensor networks: learning automata approach. *Intelligent Transportation Systems, IEEE Transactions on*, 16(3):1148–1161, 2015.

[154] Chung-Ming Huang and Shih-Yang Lin. An early collision warning algorithm for vehicles based on v2v communication. *International Journal of Communication Systems*, 25(6):779–795, 2012.

[155] Saleh Yousefi, Mahmoud Siadat Mousavi, and Mahmood Fathy. Vehicular ad hoc networks (vanets): challenges and perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761–766. IEEE, 2006.

[156] Rasmeet S Bali, Neeraj Kumar, and JJPC Rodrigues. An efficient energy-aware predictive clustering approach for vehicular ad hoc networks. *Int J Commun Syst. doi*, 10:1002, 2015.

[157] John B Kenney. Dedicated short-range communications (dsrc) standards in the united states. *Proceedings of the IEEE*, 99(7):1162–1182, 2011.

[158] Sherali Zeadally, Ray Hunt, Yuh-Shyan Chen, Angela Irwin, and Aamir Hassan. Vehicular ad hoc networks (vanets): status, results, and challenges. *Telecommunication Systems*, 50(4):217–241, 2012.

[159] Mohamed Eltoweissy, Stephan Olariu, and Mohamed Younis. Towards autonomous vehicular clouds. In *Ad hoc networks*, pages 1–16. Springer, 2010.

[160] Kaushik Sekaran and P Venkata Krishna. Big cloud: a hybrid cloud model for secure data storage through cloud space. *International Journal of Advanced Intelligence Paradigms*, 8(2):229–241, 2016.

[161] Muhammad Kazim and Shao Ying Zhu. A survey on top security threats in cloud computing. *International Jouornal of Advanced Computer Science and Applications*, 6(3):113, 2015.

[162] Xiaodong Lin, Rongxing Lu, Chenxi Zhang, Haojin Zhu, Pin-Han Ho, and Xuemin Shen. Security in vehicular ad hoc networks. *Communications Magazine, IEEE*, 46 (4):88–95, 2008.

[163] Irshad Ahmed Sumra, Iftikhar Ahmad, Halabi Hasbullah, and Jamalul-lail Bin Ab Manan. Classes of attacks in vanet. In *Electronics, Communications and Photonics Conference (SIECPC), 2011 Saudi International*, pages 1–5. IEEE, 2011.

[164] Mounib Khanafer, Mouhcine Guennoun, and Hussein T Mouftah. Wsn architectures for intelligent transportation systems. In *New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on*, pages 1–8. IEEE, 2009.

[165] Norbert Bißmeyer, Christian Stresing, and Kpatcha M Bayarou. Intrusion detection in vanets through verification of vehicle movement data. In *Vehicular Networking Conference (VNC), 2010 IEEE*, pages 166–173. IEEE, 2010.

[166] Kleber Vieira, Alexandre Schulter, Carlos Westphall, and Carla Westphall. Intrusion detection for grid and cloud computing. *IT Professional Magazine*, 12(4):38, 2010.

[167] Claudio Mazzariello, Roberto Bifulco, and Roberto Canonico. Integrating a network ids into an open source cloud computing environment. In *Information Assurance and Security (IAS), 2010 Sixth International Conference on*, pages 265–270. IEEE, 2010.

[168] Nazaraf Shah, Kuo-Ming Chao, Nick Godwin, Muhammad Younas, and Christopher Laing. Exception diagnosis in agent-based grid computing. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 4, pages 3213–3219. IEEE, 2004.

[169] Neeraj Kumar, Sudhanshu Tyagi, and Der-Jiunn Deng. La-eehsc: Learning automata-based energy efficient heterogeneous selective clustering for wireless sensor networks. *Journal of Network and Computer Applications*, 46:264–279, 2014.

[170] Narendra Kumar, Naveen Chilamkurti, and Sudip Misra. Bayesian coalition game for the internet of things: an ambient intelligence-based evaluation. *Communications Magazine, IEEE*, 53(1):48–55, 2015.

[171] Paulo Manoel Mafra, JS Fraga, and Altair Olivo Santin. Algorithms for a distributed ids in manets. *Journal of Computer and System Sciences*, 80(3):554–570, 2014.

[172] Neeraj Kumar and Chun-Cheng Lin. Reliable multicast as a bayesian coalition game for a non-stationary environment in vehicular ad hoc networks: a learning automata-based approach. *International Journal of Ad Hoc and Ubiquitous Computing*, 19 (3-4):168–182, 2015.

[173] Neeraj Kumar and Jongsung Kim. Probabilistic trust aware data replica placement strategy for online video streaming applications in vehicular delay tolerant networks. *Mathematical and Computer Modelling*, 58(1):3–14, 2013.

[174] Bharati Mishra, Priyadarshini Nayak, Subhashree Behera, and Debasish Jena. Security in vehicular adhoc networks: a survey. In *Proceedings of the 2011 International Conference on Communication, Computing & Security*, pages 590–595. ACM, 2011.

[175] Narendra Kumar, Sudip Misra, Mohammad Obaidat, Jose Rodrigues, and Bibudhendu Pati. Networks of learning automata for the vehicular environment: a performance analysis study. *Wireless Communications, IEEE*, 21(6):41–47, 2014.

[176] Narendra Kumar, Joel JPC Rodrigues, and Naveen Chilamkurti. Bayesian coalition game as-a-service for content distribution in internet of vehicles. *Internet of Things Journal, IEEE*, 1(6):544–555, 2014.

[177] Neeraj Kumar, Sudip Misra, Joel JPC Rodrigues, and Mohammad S Obaidat. Coalition games for spatio-temporal big data in internet of vehicles environment: a comparative analysis. *Internet of Things Journal, IEEE*, 2(4):310–320, 2015.

[178] Teerawat Issariyakul and Ekram Hossain. *Introduction to network simulator NS2*. Springer Science & Business Media, 2011.

LIST OF PUBLICATIONS

SCI/SCIE Journals:

(J1) Rasmeet S. Bali, Neeraj Kumar, Joel JPC Rodrigues, "An Efficient Energy-Aware Predictive Clustering Approach for Vehicular Ad hoc Networks," *International Journal of Communication Systems*, Vol. 10, pp. 1099-1131, 2015. *(John Wiley & Sons, IF 1.10) (SCI/SCIE).*

(J2) Rasmeet S. Bali, Neeraj Kumar, "Learning Automata-assisted Predictive Clustering approach for Vehicular Cyber-Physical System," *Computers & Electrical Engineering*, vol. 10, pp. 1099-1131, 2015. *(Elsevier, IF 0.82) (SCI/SCIE).*

(J3) Neeraj Kumar, Jaskaran Preet Singh, Rasmeet S. Bali, Sudip Misra, Sana Ullah, "An Intelligent Clustering Scheme for Distributed Intrusion Detection in Vehicular Cloud Computing," *Cluster Computing*, vol. 18, no. 3, pp. 1263-1283, 2015. *(Springer, IF 1.51) (SCI/SCIE).*

(J4) Rasmeet S. Bali, Neeraj Kumar, "Secure Clustering for Efficient Data Dissemination in Vehicular Cyber-Physical Systems," *Future Generation Computer Systems*, vol. 56, pp. 476–492, 2016. *(Elsevier, IF 2.46) (SCI/SCIE).*

Non-SCI Publications:

(J5) Rasmeet S. Bali, Neeraj Kumar, Joel JPC Rodrigues, "Clustering in Vehicular Ad hoc Networks: Taxonomy, Challenges and Solutions," *Vehicular Communications*, vol. 1, no. 3 , pp. 134-152, 2014. *(Elsevier) (ESCI Indexed).*

International Conferences:

(C1) Rasmeet S. Bali, Neeraj Kumar, Joel JPC Rodrigues, "An Intelligent Clustering Algorithm for VANETs," In *IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, Vienna, Austria, 3-7 November 2014, pp. 974-979.