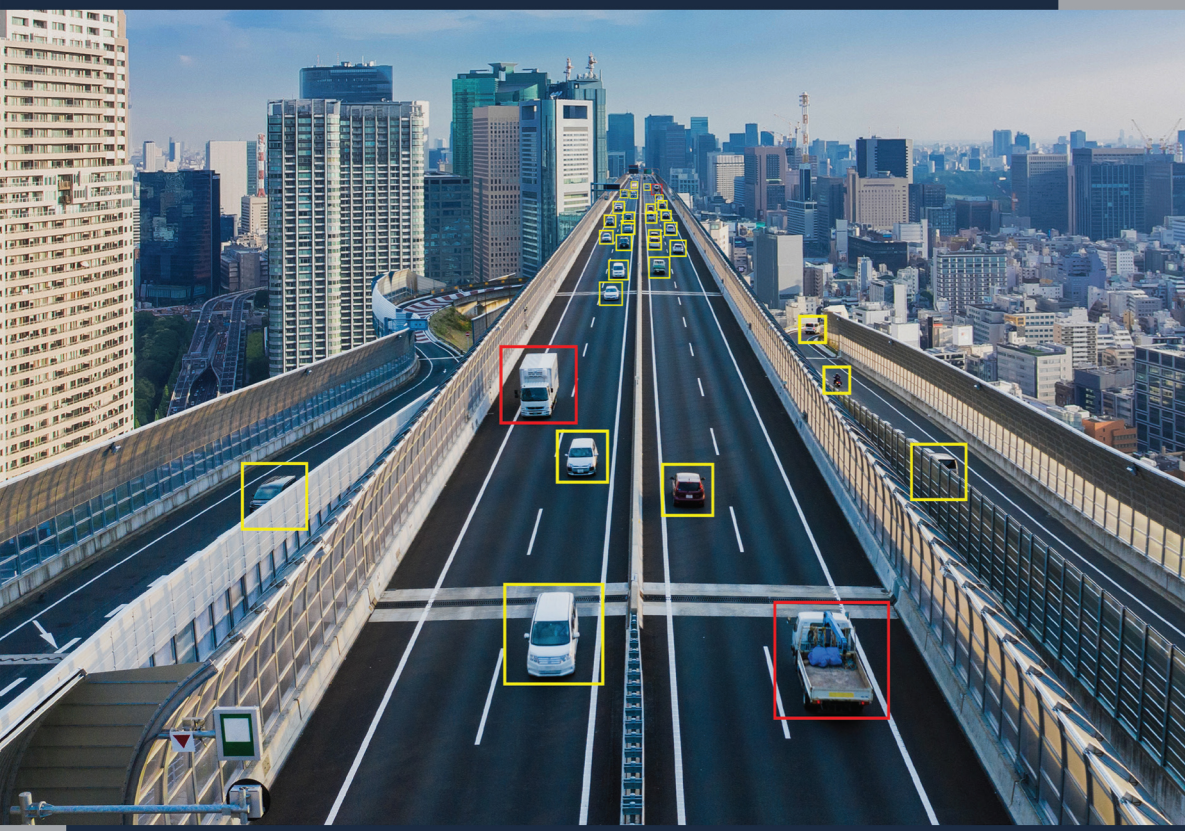


Vehicular ad Hoc Network (VANET)

Edited by: **Don Cooray**

BIBLIOTEX
Digital Library

Vehicular Ad Hoc network (VANET)

Vehicular Ad Hoc network (VANET)

Editor:

Don Cooray

BIBLIOTEX
Digital Library
www.bibliotex.com

Vehicular Ad Hoc network (VANET)

Editor: Don Cooray

www.bibliotex.com

email: info@bibliotex.com

e-book Edition 2022

ISBN: 978-1-98467-791-4 (e-book)

In Collaboration with Intelliz Press. Originally Published in printed book format by Intelliz Press with ISBN 978-1-68251-851-9

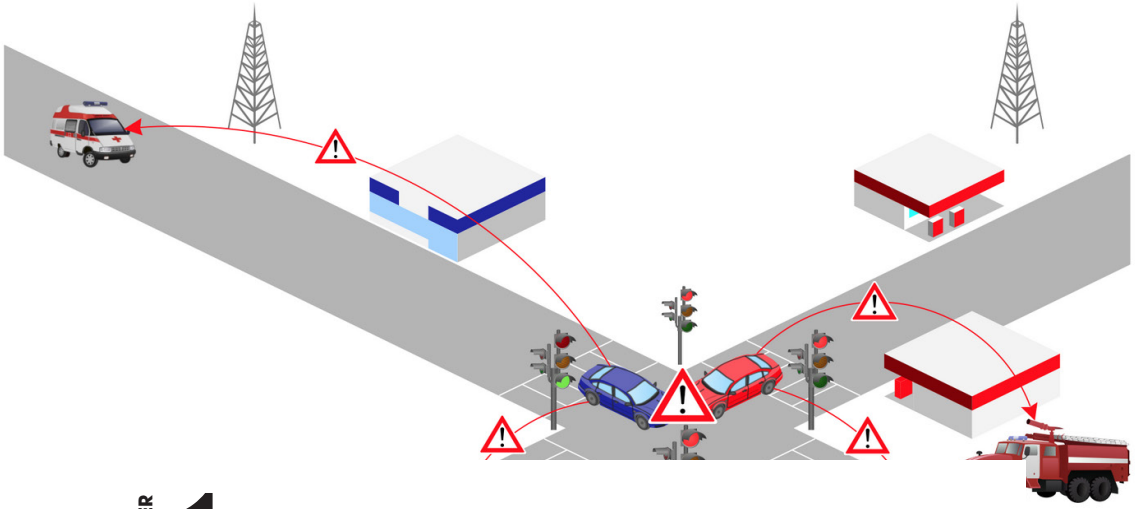
TABLE OF CONTENTS

PREFACE

Recently, with the development of vehicle industry and wireless communication technology, vehicular ad hoc networks are becoming one of the most promising research fields. Vehicular ad hoc networks (VANETs) have recently been proposed as one of the promising ad hoc networking techniques that can provide both drivers and passengers with a safe and enjoyable driving experience. VANETs can be used for many applications with vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. In the United States, motor vehicle traffic crashes are the leading cause of death for all motorists between two and thirty-four years of age. VANETs which use vehicles as mobile nodes are a subclass of mobile ad hoc networks (MANETs) to provide communications among nearby vehicles and between vehicles and nearby roadside equipment but apparently differ from other networks by their own characteristics. Vehicular ad hoc networks (VANETs) have been quite a hot research area in the last few years. Due to their unique characteristics such as high dynamic topology and predictable mobility, VANETs attract so much attention of both academia and industry.

This book deals with the basic architecture of networks, and discusses three popular research issues and general research methods, and ends up with the analysis on challenges and future trends of VANETs. A viable choice for spectrum sensing due to its simplicity, low computational cost, and ability to be applied on any kind of deterministic signal is energy

detection (ED). However, hidden terminal and low SNR problems due to shadow-fading put fundamental limits to the sensing performance and practical entailments in designing of cognitive vehicular networks. Extensive modeling efforts are then being carried out to cope with varying channel characteristics, particularly multipath fading and shadowing. The message routing in vehicular ad hoc networks (VANETs) is an attractive and promising area for research. These networks do not have a central coordination, the nodes are mobile, and the topology is highly dynamic, making the routing process a big challenge, since it is responsible for ensuring message delivery with small overhead and delay. This book presents vehicular ad-hoc networks (VANETs) from their onset, gradually going into technical details, providing a clear understanding of both theoretical foundations and more practical investigation.



CHAPTER 1

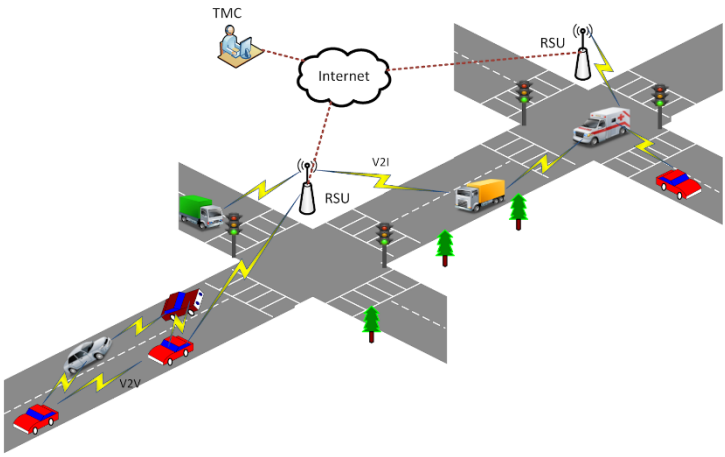
INTRODUCTION TO VEHICULAR AD HOC NETWORKS (VANETS)

INTRODUCTION

Vehicular Ad hoc Networks (VANETs) belong to a subcategory of traditional Mobile Ad hoc Networks (MANETs). The main feature of VANETs is that mobile nodes are vehicles endowed with sophisticated "on-board" equipments, traveling on constrained paths (i.e., roads and lanes), and communicating each other for message exchange via Vehicle-to-Vehicle (V2V) communication protocols, as well as between vehicles and fixed road-side Access Points (i.e., wireless and cellular network infrastructure), in case of Vehicle-to-Infrastructure (V2I) communications.

Future networked vehicles represent the future convergence of computers, communications infrastructure, and automobiles. Vehicular communication is considered as an enabler for

driverless cars of the future. Presently, there is a strong need to enable vehicular communication for applications such as safety messaging, traffic and congestion monitoring and general purpose Internet access.



VANET is a term used to describe the spontaneous ad hoc network formed over vehicles moving on the roadway. Vehicular networks are fast emerging for developing and deploying new and traditional applications. ANETs are characterized by high mobility, rapidly changing topology, and ephemeral, one-time interactions. Basically, both VANETs and MANETs are characterized by the movement and self-organization of the nodes (i.e., vehicles in the case of VANETs). However, due to driver behavior, and high speeds, VANETs characteristics are fundamentally different from typical MANETs. VANETs are characterized by rapid but somewhat predictable topology changes, with frequent fragmentation, a small effective network diameter, and redundancy that is limited temporally and functionally.

VANETS are considered as one of the most prominent technologies for improving the efficiency and safety of modern transportation systems. For example, vehicles can communicate detour, traffic accident, and congestion information with nearby vehicles early to reduce traffic jam near the affected areas. VANETs applications enable vehicles to connect to the Internet to obtain real time news, traffic, and weather reports. VANETs also fuel the vast

opportunities in online vehicle entertainments such as gaming and file sharing via the Internet or the local ad hoc networks.

Applications such as safety messaging are near-space applications, where vehicles in close proximity, typically of the order of few meters, exchange status information to increase safety awareness. The aim is to enhance safety by alerting of emergency conditions. Applications for VANETs are mainly oriented to safety issues (e.g., traffic services, alarm and warning messaging, audio / video streaming and generalized infotainment, in order to improve the quality of transportation through time-critical safety and traffic management applications,). At the same time, also entertainment applications are increasing (e.g., video streaming and video-on-demand, web browsing and Internet access to passengers to enjoy the trip).

Applications of alarm messaging have strict latency constraints of the order of few milliseconds, and very high reliability requirements. In contrast, applications such as traffic and congestion monitoring require collecting information from vehicles that span multiple kilometers. The latency requirements for data delivery are relatively relaxed i.e., they are "delay-tolerant", however, the physical scope of data exchange is much larger. In contrast, general purpose Internet access requires connectivity to the backbone network via infrastructure, such as Road-Side Units (RSUs).

Non-safety applications are expected to create new commercial opportunities by increasing market penetration of the technology and making it more cost effective. Moreover, comfort and infotainment applications aim to provide road travelers with needed information support and entertainment to make the journey more pleasant. They are so varied and ranges from traditional IP-based applications (e.g., media streaming, voice over IP, web browsing, etc.) to applications unique to the vehicular environment (e.g., point of interest advertisements, maps download, parking payments, automatic tolling services, etc.).

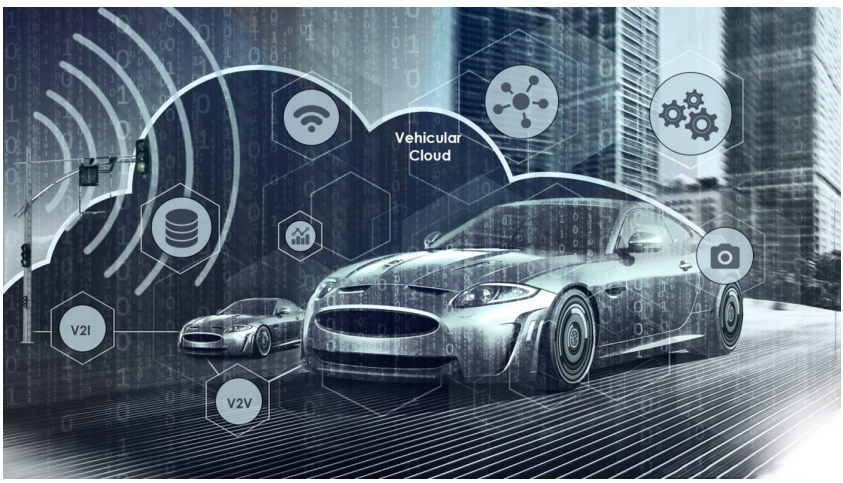
We can distinguish between intra and inter-vehicle communications. The first term is used to describe communications within a vehicle, while the second one represents communications between vehicles, or vehicles and sensors, placed in or on various locations, such as roadways, signs, parking areas, and so on. Inter-vehicle communications can be considered to be more technically challenging because vehicle communications need to be supported both when vehicles are stationary and when they are moving. As an instance, the use of a prepaid or automatic billing system when a vehicle slows down instead of stopping at a toll-booth is provided by using a small electronic transmitter.

Quality of service provided in a VANET is strongly affected by mobility of vehicles, and then dynamic changes of network topology. Different classes of vehicles can move in VANETs, depending on traffic conditions (i.e., dense and sparse traffic), speed limits in particular roads (i.e., highways, rural roads, urban neighborhoods), and also typology of vehicles (i.e., trucks, cars, motorcycles, and bicycles). In general, compared to traditional mobile nodes in MANETs, vehicles in VANETs move at higher speeds (i.e., from 0 to 40 m/s). All these unique features let VANETs well fit into the class of opportunistic networks that means the network behavior is changing and connectivity availability is not always satisfied. As a typical example, in order to maintain network connectivity in VANETs, it is a common technique to connect vehicles traveling on the roadway in opposite directions by means of opportunistic connectivity links. This situation is described as bridging technique. However, link breakages strongly hinder stable and durable V2V communications, and as a result communications are dropped. On the other hand, the limited infrastructure coverage, because of sparse fixed access points settling, may cause short-lived and intermittent V2I connectivity. It follows that interconnectivity and seamless connectivity issues in vehicular ad hoc networks represent a challenge for many researchers. Solutions based on both horizontal and vertical handover procedures have been largely investigated in recent works.

1.1 VEHICULAR AD HOC NETWORKS (VANETS)

Vehicular ad hoc networks (VANETs) are created by applying the principles of mobile ad hoc networks (MANETs) – the spontaneous creation of a wireless network of mobile devices – to the domain of vehicles. VANETs were first mentioned and introduced in 2001 under "car-to-car ad-hoc mobile communication and networking" applications, where networks can be formed and information can be relayed among cars. It was shown that vehicle-to-vehicle and vehicle-to-roadside communications architectures will co-exist in VANETs to provide road safety, navigation, and other roadside services. VANETs are a key part of the intelligent transportation systems (ITS) framework. Sometimes, VANETs are referred as Intelligent Transportation Networks. They are understood as having evolved into a broader "Internet of vehicles". Which itself is expected to ultimately evolve into an "Internet of autonomous vehicles".

VANET is a variation of MANET (Mobile Ad-hoc Network). MANET comprises of nodes which communicate without central network and where nodes are equipped with networking capabilities. VANET on the other side has emerged as a challenging and more liable class or variation of MANET. The freedom of nodes to enter or leave the network in VANET calls for different routing protocols than MANET.



This inter vehicle communication leads to passing and receiving of information so as to increase traffic efficiency, detect road conditions, decrease collisions, detect emergency situations and overall increase the efficiency of the network. VANET transfers the information to distant devices as well with the help of multi hops.

VANET can be characterized by following factors:

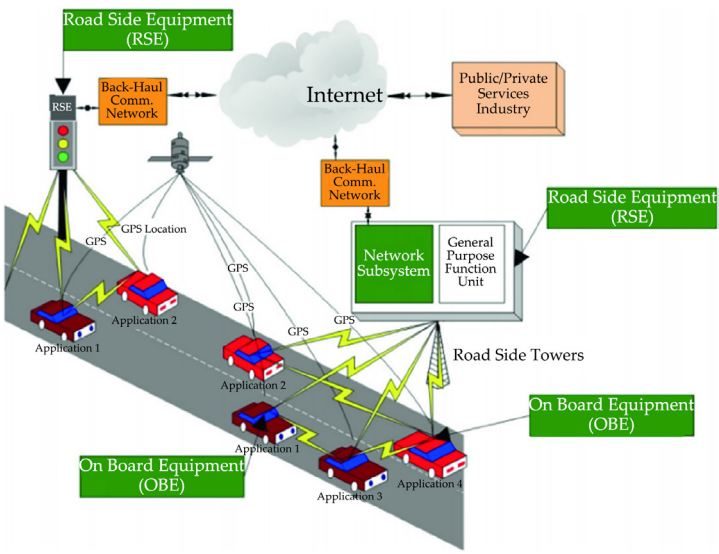
- *Dynamic topology*: The speed and direction of vehicles changes constantly thereby resulting in high dynamic topology
- *Intermittent connectivity*: Connectivity between devices changes very frequently like connection between two devices exchanging information can disconnect anytime. The reason behind frequent disconnection is high dynamic topology.
- *Mobility Patters*: A large section of vehicles follow a certain patterns to move which is generally a function of traffic signals, speed limits, highways, streets, road conditions etc. These patterns when observed help in the creation of routing protocols for VANET.
- *Unlimited power and storage*: It is assumed that the nodes in VANET are capable of possessing an unlimited amount of power as well as storage capacity. Therefore the nodes are free to exchange the data without the foundations of power consumption or storage wastage.
- *On board sensors*: VANET assumes that the nodes are seldom equipped with on board sensors which are capable of transmission of information to other devices or nodes.

VANET also forms a very important part in Intelligent Transport Systems as insights are produced from the information being exchanged by the vehicles and other devices in the VANET.

1.1.1 Architecture for VANET

VANET aims to provide communication between different neighboring vehicles. The entities in a VANET can be divided into three domains

- *Mobile domain*: Mobile domain comprises of two parts. First is vehicle domain which encompasses all the vehicles which are moving constantly such as buses, cars, trucks etc. Second part is mobile device domain which comprises of all the portable handy devices such as PDAs, laptop, GPS, smartphones etc.
- *Infrastructure domain*: It also comprises of two parts. Roadside infrastructure domain comprises of stationary roadside entities such as traffic lights, poles etc. Whereas, central infrastructure domain encompasses the central managing center such as vehicle management center, traffic management center etc.
- *Generic domain*: It comprises of Internet infrastructure and Private infrastructure. For instance, different nodes and servers and other computing resources working directly or indirectly for a VANET come under generic domain.



The mobile domain exchanges information and communicates to Infrastructure domain which processes data and does its own modulation. Then in the second step, infrastructure domain in turn communicates to generic domain and exchanges information with it. This data flow among the stationary and mobile resources result in efficient and effective utilization of road by the users.

Another form of VANET architecture is communication architecture where communication types are:

- *In vehicle communication*: It detects the inner system data or performance of the vehicle and determines factors such as driver exhaustion or drowsiness etc. Determination of such factors and their extent is crucial for public safety as well as driver safety.
- *Vehicle to Vehicle communication (V2V)*: The data exchange between different vehicles so as to assist the driver by informing them about warnings and other critical information to one another. V2V communication does not rely on fixed infrastructure for data exchange to happen and it helps in dissemination, safety and security applications.
- *Vehicle-to-road infrastructure (V2I) communication*: This communication taking place between mobile vehicles and roadside fixed infrastructure in order to gather data. It provides updates related to environmental sensing and monitoring such as real time traffic update or weather update.
- *Vehicle-to-broadband cloud (V2B) communication*: This allows communication of vehicles over broadband connections such as 3G/4G. This enhances the driver assistance and vehicle tracking as the broadband cloud may contain more of traffic information and other data.

Communication types take place in a single or multiple VANETs. The type of communication doesn't matter until and unless performance of VANET doesn't suffers. When vehicles move and an ad-hoc network, then information exchange begins. This transmission of information to other vehicles and nodes. The

vehicle works and leverages the VANET as long as it stays in that particular network.

VANET primarily supports two types of applications one is driver assistance and other is information dissemination. Driver assistance requires exchange of such information which assists the driver to maintain a more secure and efficient environment. Information dissemination focuses on delivering information to everyone such as drivers, nodes, passengers etc. Information dissemination applications range from critical safety applications to entertainment applications.

1.1.2 Protocols for Transmission

The life of VANET lies in the communication that takes place between different vehicles. The data being gathered and exchanged by the vehicles requires some protocols or rules through which transmission can take place in a systematic and organized way. The data exchange between nodes in a VANET happens via routing protocols. These protocols define how a packet of data will be distributed among different nodes. On the basis of senders and receivers involved, three types of protocols are defined for VANET communication which are briefed as:

Unicast: Such protocols aim to deliver or transmit data from one source to one destination over a wireless medium. There are two ways to transmit packets; one is via multi-hop transmission where an information of packet in transmitted further and further via hopping of packet to neighboring vehicle. Second one is carry and forward technique where a packet is carried by the vehicle as long as possible and then transmitted to reduce congestion or rebroadcast of packet. Third is trajectory based where nodes calculate various paths of data transmission and then transmit data by keeping in notice that minimum rebroadcast of packet happen.

Broadcast: Broadcasting protocols aim to deliver and communicate to as many nodes as possible. In situations like, road blocks, traffic

jams, places with high traffic density or emergency situations, broadcasting protocols are a must. They transmit data packet to more than one node at a time. On the counter side, broadcasting protocols also increase the chances of packet rebroadcast or storm problem. Figure 1 shows a list of broadcasting protocols.

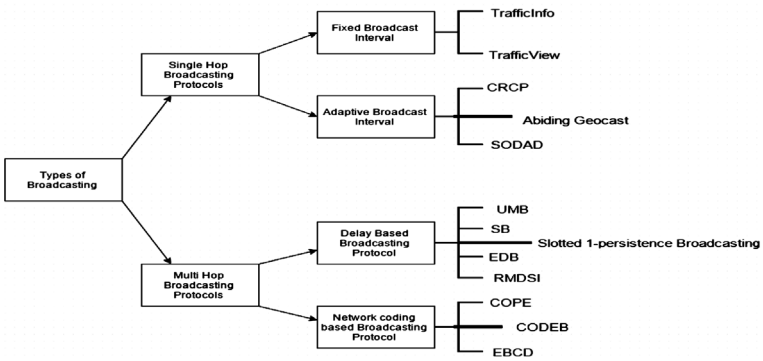


Figure 1: Classification of broadcasting protocols.

Challenges and Future Research Directions

Apart from the advantages one can take from adoption of VANET, there are numerous challenges as well which VANET has to face. These challenges can be viewed as a future research direction or open research issues where advancement and solutions are still required. Some of these challenges which user can take as research issues are:

Mobility: Ad-hoc networks comprises of mobile devices, PDA, laptops and other devices as nodes which have limited mobility or less mobile nature whereas in VANET the mobility factor of nodes is very high. Vehicles come and lose contact in a matter of seconds as speed is measured in miles per hour. Exchanging information in such a small amount of time and with such highly heterogeneous nodes is an open research issue which calls for development of more advanced and rich network topology model which has to differ from traditional models which require a larger interaction level between sender and receiver.

Data administration and storage: Any number of vehicles and other mobile as well as stationary devices can participate in a VANET. For large scale VANET's, number of nodes participating in a VANET can increase up to millions which will in turn generate a large amount of data. Monitoring, managing and storing such a large amount of data is still a challenge which researchers face.

Security and Privacy: VANET is an open network where any node is allowed to join the network. There is no certain mechanism which can ensure the trustworthy nature of the nodes. Therefore, security becomes a major concern for researchers as communication between nodes happen over a wireless medium where any node can transfer malicious data and may cause significant harm to other nodes. Moreover the identification of such a vehicle is also difficult and calls for better and robust security models to ensure security in VANET. Moreover, untrustworthy nodes can detect the activities, habits and patterns of other users by peeking the VANET and may cause serious threat to privacy of the individual.

Quality Service Delivery: The nodes participating in a VANET are very mobile and have very dynamic nature. Factors such as node position, topology, distance between nodes, connectivity etc. vary significantly and thereby the routing strategies and protocols applied become incapable of delivering a good quality of service. Designing, modelling and developing mechanism which ensure a good quality of service throughout the VANET is also an open research issue as well as a challenge which seeks solution.

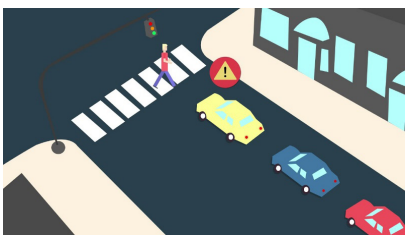
Heterogeneity and Standardization: The nodes participating in a VANET are highly heterogeneous in nature such as cars, buses, trucks, roadside units, and traffic lights and other nearby computing resources. Such diverse nature of nodes or vehicles have different modes of communications and every node follows its own set of communication mechanisms. Dealing with such high density networks and their varied modes of communication requires standardization of protocols via which nodes can communicate with any node on the go.

Routing Protocols: Traditional routing protocols are not appropriate for a VANET as the nodes participating in a VANET have very high mobility and thus they change the network topology in a matter of seconds. Moreover establishing a connecting between nodes and exchanging information between source and destination nodes and further propagation of information to other nodes requires the development of robust algorithms and routing protocols so as to deliver higher throughput, better service and enhanced packet delivery ratio.

1.2 TRAFFIC MONITORING SYSTEM FOR VANET

The area of automated surveillance systems is currently of immense interest due to its implications in the field of security. Surveillance of vehicular traffic and human activities offers a context for the extraction of significant information, such as scene motion, traffic statistics, object classification, human identification, anomaly detection, and the analysis of interactions between vehicles, between humans or between vehicles and humans.

Detecting moving objects in the video stream is essential for understanding behavior and tracking objects. To detect moving objects accurately, segmentation from sequences of frames is performed using adaptive Gaussian mixture models (MGM), because it became a standard in the recent years due to their analytical representation and theoretical basics. In Gaussian mixture models, the distribution of the background pixels is represented with a mixture of Gaussian functions, where some modes correspond to the background and the other associate with active regions.



As an applicable idea, image processing-based monitor systems are implemented in communication systems, such as Vehicular Ad-hoc Network (VANETs) in such a way the monitored events are reported to a remote traffic control unit through the communication lines. As depicted in Fig.2, VANET is a kind of Ad-hoc networks where every vehicle is attached with an On-Board Unit (OBU) to communicate with other vehicles through a Dedicated Short Range Communication (DSRC) protocol. Those kinds of communications are used for vehicular safety and infotainment applications. In addition, the OBU installed inside a vehicle also uses DSRC connections to communicate with Road-Side Units (RSUs) that are disseminated along the road and can easily enable the vehicle to communicate with the internet. Those traffic condition reports are then transmitted to the traffic Central Control Unit (CCU) to take some actions; for instance, in case of "emergency vehicle warning" report, the CCU may change the traffic light in order for the ambulance to pass. However, from security point of view, a malicious driver may transmit some fake traffic information, which in turn results in traffic congestion. Therefore, we propose a security scheme to detect and revoke fake messages and malicious vehicles from the VANET system. The security requirements for such monitoring systems are: Authentication, Data Confidentiality, Driver anonymity, Vehicle's location privacy, and Non-repudiation

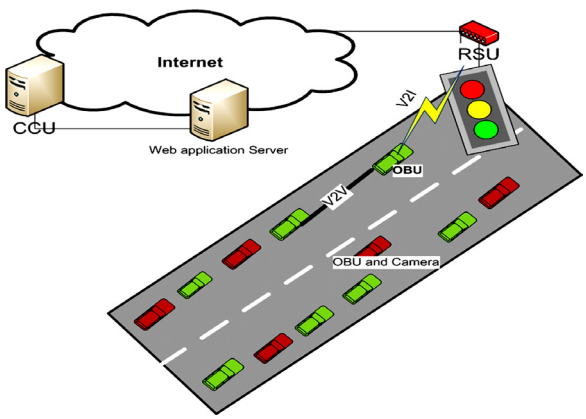


Figure 2: VANET System Model.

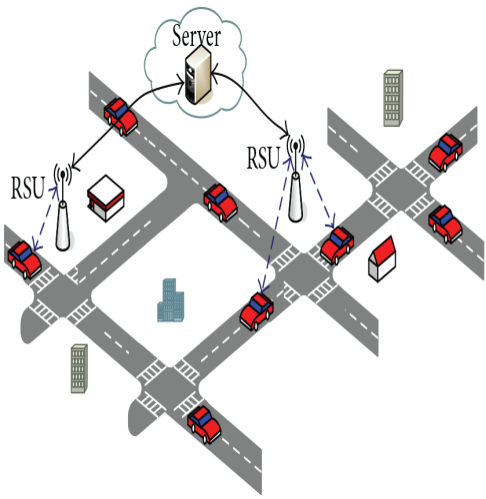
1.2.1 Work

In a video surveillance and monitoring systems, there are lots of techniques to detect moving objects. The most common object detection techniques are: point detectors, background subtraction and segmentation. In a video sequences, all moving objects are required to be segmented by automatic image analysis techniques. The most common segmentation techniques depend on statistical models of the background with estimated probability distribution, such as a mixture of Gaussians, to represent each pixel in the video stream. Applying security and privacy schemes for VANETs, many schemes have been proposed. An authentication scheme is proposed to anonymously authenticate a vehicle to an RSU, which validates all subsequent messages received from the authenticated vehicle. The vehicle afterwards employs the generated key to encrypt its messages to the RSU. However, this scheme is computationally expensive since the transmitted messages are delayed until the RSU verifies them. In addition, to verify message authentication, the RSU in categorizes the vehicles into three groups: high priority, registered, and unregistered vehicles. The registered vehicle receives a public-private key pair from a trusted authority (TA) and signs the messages using those keys. Both the RSU and neighbor vehicles can verify the received messages. But, unlike our proposed security scheme, this scheme cannot detect insider attackers, in which a register vehicle sends fake messages.

1.2.2 System Models

The VANET system contains an OnBoard Unit (OBU) installed in every vehicle, and a group of Road-Side Units (RSUs) installed in every traffic-light all over the road. The communication among cars (or OBUs) are called Vehicle- to-Vehicle (V2V) communication while the communication between a car (OBU) and an RSU is called Vehicle-to-Infrastructure (V2I). The network protocol used for these types of communication is the dedicated Short-Range communication (DSRC) at 5.9GHZ frequency band wireless radio signals. Every OBU contains a monitoring camera in order to

collect the images. In addition, the RSU connects to the Internet through wired communication and uses the Internet connection to communicate to the CCU, which works as a trusted authority in the system. The OBU transmits the generated monitoring reports for specific event to their neighboring OBUs as well as to the CCU. Considering that every OBU has a built-in GPS, the monitoring reports transmitted from the cars should contain the position of the monitored event, P_{me}. Without loss of generality, the generated traffic report contains the following fields: Vehicle's Pseudo Identity, V P ID, current time, T_i , Position, P_{me}, Direction, Z_i , Speed of the vehicle, and the traffic event, which contains two main fields: the traffic type and traffic information, the former includes the type of the reported event, while the latter contains more information related to the event. The traffic type field may be one of 3 types: accident location, Emergency vehicle warning, and group of people warning, such as protests. Every OBU transmits a traffic report every 200 ms to its nearest RSU, which in turn transmits the report to the CCU. The CCU also contains a web server in order for home users to check the traffic condition anytime from the Internet. In our model, we consider all RSUs and the CCU as trusted parties to all vehicles in the system.



1.2.3 Proposed Model

The proposed model is comprised of three phases: improved sensory, monitoring analyzer that detects the traffic monitoring event, and applying security and privacy preserving scheme to detect fake transmitted information as well as revocate malicious vehicle from the system. In the context of active vision, the driver externally directs the camera towards the spatial region of interest in a scene. An improved monitoring model for detecting moving objects is proposed. The inputs to the model are video sequences to detect and recognize moving objects. A preprocessing stage aims to avoid the problems that arises during video acquisition process is implemented. In the preprocessing stage, filters select visual areas from an internal representation of the scene to improve the monitoring process. Hence, image processing techniques are used to automatically generate the monitoring report that is distributed along with the application subscribers. Finally, the detected monitoring events are securely transmitted to the CCU to take an action. The proposed security and privacy scheme can detect fake reports as well as malicious vehicles.

The input data to the system is the color video that is acquired through the fixed camera in the vehicle. The movements and vibrations of the vehicle usually raise instability in the input video data. The proposed model is tested with input AVI (audio video interleave) format video file of 640 x 480 frame size and frame rate 25fps is captured in real time with digital camera.

Improved Sensory

To reduce the effect of non smooth motions of vehicles in video sequences, the image stabilization techniques are commonly used [8]. The sensory process forms with series of selective filters located in the sensory analyzers to improve the images of the scene in the monitoring process. The selective filters work consequently as follows: first we estimate the noise of the image using the standard deviation, σ_n of the noise from the noisy image that is computed from (2). Repress the image structures by the following Laplacian

operator of the ”Fast Estimation” method

$$\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \quad (1)$$

Then compute the standard deviation of the noise as follows:

$$\sigma = \sqrt{\frac{\pi}{2}} \frac{1}{6(W-2)(H-2)} \sum_{imageI} |I(x,y) * N| \quad (2)$$

Where W, and H are the width and height of the image, respectively. Hence, according to the value of σ_n, we make a decision to use median, wiener, or both median and wiener filters, for noise removal. The intervals for σ_n values were partitioned to three ranges that perform well for the proposed model as follows: Median filter for noise removal will be used Wiener and Median filter for noise removal will be used, and Wiener filter for noise removal will be used.

Monitoring Analyzer

In the monitoring analyzer phase, the proposed model is considered to automatically generate the report of monitoring that consists of 124 bits message formatted as follows: (current time, current location, traffic event type). To generate the monitoring report automatically, the proposed system passes through three steps. The current time is determined automatically via the connection to the device clock, and the current location is determined automatically from the GPS via the designed android application. Hence, the proposed system detects the traffic event type, crowd, detect the existence of demonstration, ambulance , and accident detection.

In the first step, to detect the crowd in the monitoring scene, we need to extract the number of moving objects in the monitoring scene of dynamic input video. The object detection and tracking composed of two components: moving object segmentation, and object tracking. Moving object segmentation is simply based on a comparison between the input frame and a certain background

mode, and different regions between the input and the model are labeled as foreground based on this comparison. The background and static regions contains most pixels in the frame, hence to detect individual targets in the scene suitable algorithms are needed. The key pointer of target objects in surveillance videos is the motion, so motion-based segmentation schemes will be used to detect individual moving objects in video frames.

Common algorithms for moving object segmentation include average background modeling, Gaussian background modeling, and MGM (Multiple Gaussian Model). In the proposed model, we use MGM algorithm that represents the distribution of the background pixels with a mixture of Gaussians to associate the pixels with the background or with active regions. Moreover, we use concept modification that aims to avoid the problems that arises during the video Acquisition process as discussed in the improved sensory phase with threshold value 50 to facilitate accounts for the data that should belong to the background.

Morphological opening filter was used to remove the undesirable noise foreground that includes after segmentation process. Blob detection, based on the Laplacian of the Gaussian with an area of blob greater than 2000, was used to obtain regions of interest for further processing, determining the number of moving objects, and detecting the connected components of the specified regions. The proposed system calculates the average number of detected moving objects in the video frames every minute to send for the generated report. If the number of detected moving objects is greater than 20, the system detects that the perspective scene is crowd.

In the second step, to detect the existence of demonstration or ambulance in the monitored scene, we employ feature detection and image matching techniques that have an important impact for the recognition system of objects. To discriminate one object from the scene, a superior set of perceptive information was extracted, the categorical feature of the object appearing in the image area corresponding to the focus of attention is detected. We used the Speed-Up Robust Feature detector (SURF) algorithm of

feature detection for matching step because it is more suitable for textured objects and are more robust with respect to illumination and variations they pose. The object recognition system works in the following consequence; extract the salient feature points from the image, construct regions around the salient points and finally match the images based on extracted features.

1) *Accident detection system in the monitoring scene*: In accident detection event, earlier researches related to traffic accident detection system involved detecting abnormal incidents. In this paper, a new technique is proposed that depends on the result of three-feature extracted factors: area (size), orientation, and centroid of the tracked car. Hence, to determine accidents we propose a general equation that consists of the combination of the three extracted involved factors from the analysis of traffic images.

2) *The steps of vehicles accident detection system*: The moving vehicles on each frame are detected as illustrated above using MGM algorithm. Several morphological operations (opening, closing, fill region and holes) are applied to avoid the problems of background clutter, different view aspects and the variation of high-low contrast object. When tracking the detecting vehicle or multiple vehicles over time in each video frame using kalman filter, the output is the moving regions that should be monitored to detect the significant changes. Then, extract the features of the tracking objects, such as the area (size), orientation, and the centroid of the tracked vehicle.

To determine accidents, we propose a general equation that consists of the combination of the percentage of the three extracted features that strongly participate for vehicles accident detection from the analysis of traffic images. The overall accident detection equation is defined as follows:

$$\text{Accidentpercentage} = \text{Area}+\text{orientation}+\text{centroid} \quad (3)$$

And give each concerned factor possibility percentage of 33.3%

according to the following extracted features:

If Accident percentage > 60% Then T he system detects that there is an accident (4)

Area variation: The accidents cause rapid change to the size via the contact of two vehicles; therefore, we used the variation rate of area as a factor for traffic accident detection as follows: Let A_i , i = 1, 2, 3...denotes the individual vehicle area detected in the input image if Ai of the tracked vehicle changed rapidly from frame to the next frame as the change in area of the vehicles exceeds the area threshold then area variation have the rate of 33.3% for possibility of vehicle accident detection.

Centroid: The Centroid is defined as the center of mass of the vehicle region with the region with the expression:

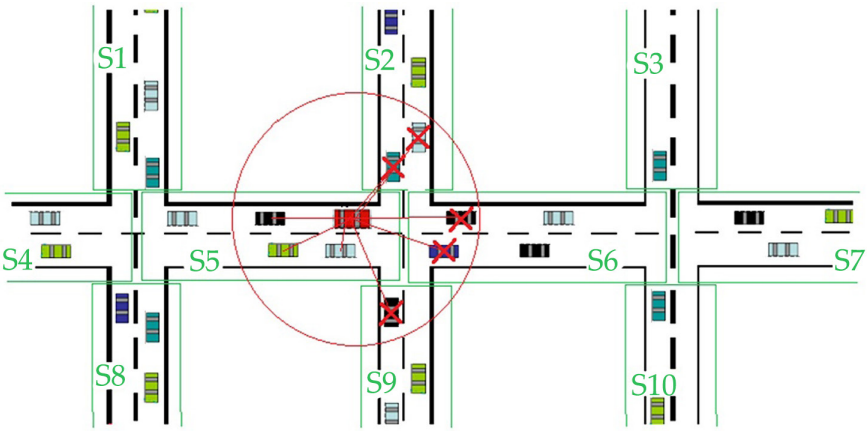
$$\left(\frac{\sum_i x_i}{total\ no.\ of\ pixels}, \frac{\sum_i y_i}{total\ no.\ of\ pixels}\right) = (\overline{x}, \overline{y})$$

Where x_1, x_2, .xnandy$_1$, y_2, .y_n denote the points along the horizontal and vertical plane of the image, respectively. If the centroid position of tracked vehicle changes from frame to the next frame, this means that there is an intersection between two objects which cause a change in overall position of the vehicles. Therefore, centroid has the rate of 33.3% for vehicle accident.*Orientation*: Orientation is defined as the angle in degrees between the x axis (the horizontal dotted line) and major axis of the bounding box ellipse of each vehicle where the orientation ranges from −90 degrees to 90 degrees. If the orientation changes (Δθ) of the tracked vehicle changes significantly than a given threshold (−45, 45) from frame to the next frame, then a possibility of accident occurrence is determined and the orientation factor have the rate of 33.3% for vehicle accident.

1.3 ROAD TRAFFIC CONGESTION

Traffic congestion is now considered one of the biggest problems around the world. Traffic problems will be also much more

widely increasing as an expected result of the growing number of transportation means and current low-quality infrastructure of the roads. There are many factors causing traffic congestion such as rush hour, road construction, accident and even bad weather. All this factors and many other can cause traffic congestion, Drivers who are unaware of congestion eventually join it and increase the severity of it. The more severe the congestion is, the more time it will take to clear once the cause of it is eliminated. The ability for a driver to know the traffic conditions on the road ahead will enable him/her to seek alternate routes saving time and fuel. When many drivers have this ability, traffic congestions will be less severe and only the vehicle in the center of congestion area will be affected. This would lead to a much more efficient use of road infrastructure. In order to face Traffic congestion there is a need for a system that provide drivers with useful information about traffic conditions, Information such as congestion type, location and boundaries. Traffic Congestion system must relay this information to drivers within the congestion and those heading towards it.

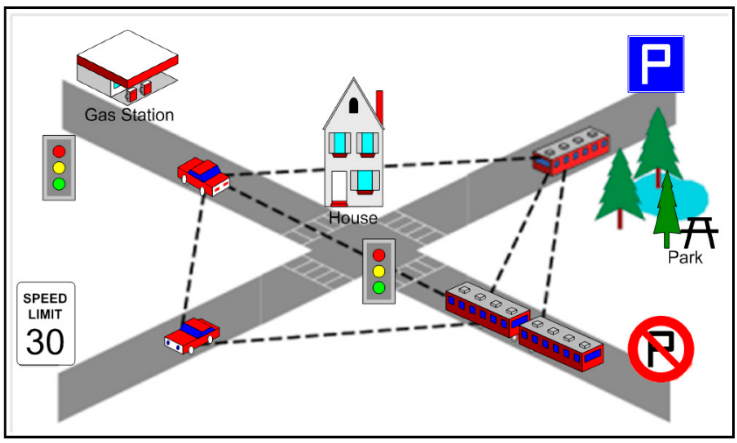


For vehicles within the congestion to form their own picture of congestion they need to collaborate using Vehicle-to-Vehicle (V2V) or vehicle-to-infrastructure (V2I) communication. Once a clear picture of the congestion has formed, this information needs to be relayed to vehicles away from the congestion so that vehicles heading towards it can take evasive actions avoiding further escalation its severity. The improvement of traffic flow

and congestion reduction can be achieved by means of traffic information systems (TIS).

1.3.1 Different Approaches of Traffic Information Systems

Most current navigation systems are static and do not provide traffic information. Route selection is based solely on static map data which leads to the system that fails to give the driver the most efficient route to his/her destination. In the last year or so, some of these devices have incorporated "real-time" traffic information to aid in route selection. Such "real-time" traffic systems such as the services provided from NAVTEQ and other commercial services today rely on humans and/or road infrastructure like traffic cameras and radars to maintain a central database of current traffic condition based systems.



Traffic Information Systems (TIS) are one of the key non-safety application areas of VANETs. As such, TIS are much less delay sensitive compared to safety applications, which have recently attracted a lot of attention in VANET. In general, TIS can be classified as either Infrastructure less or Infrastructure based as shown in figure (3).

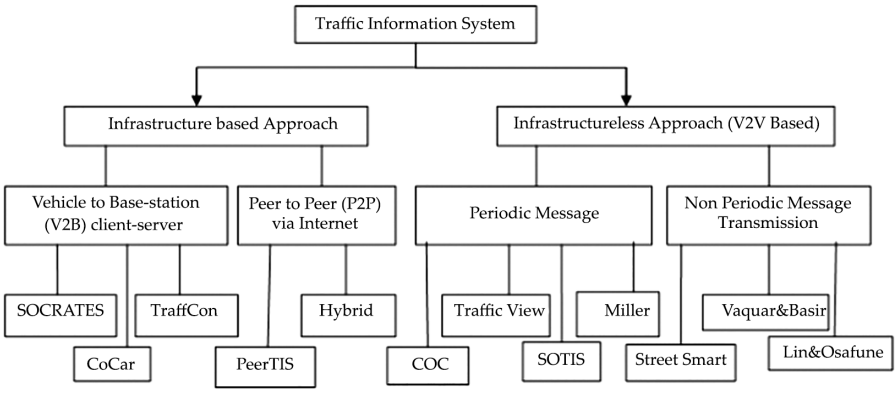


Figure 3: Traffic Information Systems Classification.

Infrastructure-based: TIS can rely on client-server or peer-to peer (P2P) models of data storage and communication that based on centralized architecture. Most traffic information systems are based on a centralized architecture focused around a traffic management center that collects data from the street network, via sensing devices, and processes them.

The resulting traffic information is made available to the drivers via broadcast service or alternatively on demand via cellular phones. The centralized approaches are dependent on fixed infrastructure that demands public investments from government agencies or other relevant operators to build maintain and manage such infrastructure: a large number of sensors are needed to be deployed in order to monitor the traffic situation.

The traffic information service is then limited to streets where sensors are integrated. Besides centralized designs, having the disadvantage of being rigid, difficult to maintain and upgrade, require substantial computing/communications capabilities, and are susceptible to catastrophic events (sabotage or system failures). Moreover, such systems require much lower market penetration compared to infrastructure-less approaches. There are many systems have been established depending on this approach such as SOCRATES system, TraffCon system, CoCar system, PeerTIS system and Hybrid system

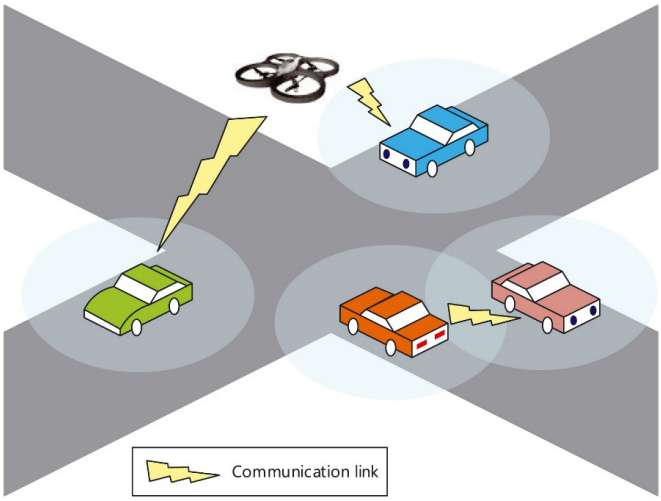
Infrastructure less approaches: Typically apply data aggregation techniques to limit bandwidth use and maintain scalability. Usually, with increasing distance, observations regarding a given area become less precise. Thanks to store-and forward techniques, traffic information can be disseminated in multiple partitions of VANETs. As the information is a subject of interest to many vehicles in a given geographical area, the broadcast nature of V2V communication fits very well the objectives of Infrastructure less TIS. However, such systems have two main drawbacks in long distance information dissemination. Firstly, they have a relatively high delay and secondly the information is limited in its details (due to the distance-based data aggregation). Another problem is that several overlapping aggregates for the same area may exist, making it difficult to compare them. Therefore, the quality of V2V communication based approaches greatly depends on the quality of the aggregation techniques. There are many systems that use this approach such as Contents Oriented Communication (COC), Traffic View, SOTIS, Miller, Street Smart, Vaqar and Basir and Lin and Osafune.

1.3.2 Infrastructure less Traffic Information System

The existing traditional ITS traffic information systems are based on a centralized structure in which sensors and cameras along the roadside monitor traffic density and transmit the result to a central unit for further processing. The results will then be communicated to road users. These systems require substantial public investment in sensing, processing and communication equipment's. Moreover, such systems are characterized by long reaction times and thus are not useable by all the applications requiring reliable decision making based on accurate and prompt road traffic awareness.

Infrastructure less approach of TIS provides a completely decentralized mechanism for the estimation of traffic density in city roads. This decentralized approach is based on traffic information exchanged, updated and maintained between vehicles in the roads. The estimated road traffic density information is useful for several ITS-related applications. Particularly, this schema is suitable for

integration to real-time traffic congestion warning systems. It may also be used as a critical metric for determining optimal vehicular data routing paths in Vehicular Ad Hoc Networks (VANET).



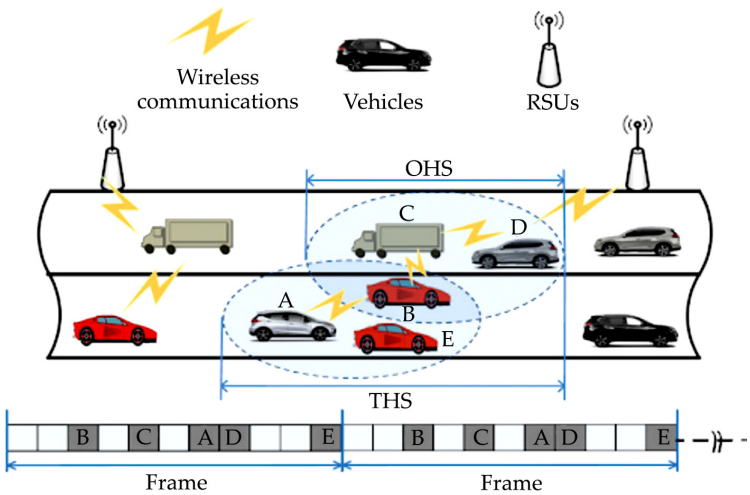
Infrastructure traffic information systems depends mainly on V2V (Vehicle-to-Vehicle) communication. This kind of architecture allows vehicles to send information between each other via multi-hop communication.V2V is better suited for safety applications because the vehicles can almost immediately detect collision or congestion warning that is transmitted within the affected area. the main advantage of this type of communication is low cost of deployment where there is no infrastructure or road side units also less delay than infrastructure to vehicle communication so it is suitable for sensitive application such as traffic congestion detection and collision avoidance and detection. Those systems used for traffic congestion detection based on V2V communication.

Periodic Message Techniques

1) Contents Oriented Communication (COC)

COC is a technique where vehicles estimate road traffic density from received beacon messages, and periodically transmit this

information to other vehicles. Vehicles can then detect traffic congestion conditions by comparing the exchanged traffic density estimates with average density values for the road segments under evaluation. With COC, each vehicle collects original information that each vehicle has by communicating each other, and creates contents which may be useful for drivers, by analyzing original information. COC deliver the analyzed contents to other vehicles. The simulation results show that COC provide timely information of vehicular accidents and congestion to drivers. This technique is seen to be one of very good ways to estimate traffic congestion and accidents. COC gets the content that is the local information immediately after dangerous events like terrorism, the situation information immediately after generation of catastrophe, or the local information of vehicular accidents and congestions. COC exchanges the information that consists of own status each other and acquired by surroundings. COC analyzes the situation in the surrounding in real-time. Moreover, COC shares the analyzed information among vehicles. In a word, people can recognize the vehicular accidents and congestions in real-time by using COC. This capability is obtained at the expense of overloading the communications channel through the continuous exchange of traffic density estimates so appear the need for better techniques such as Traffic View and SOTIS.



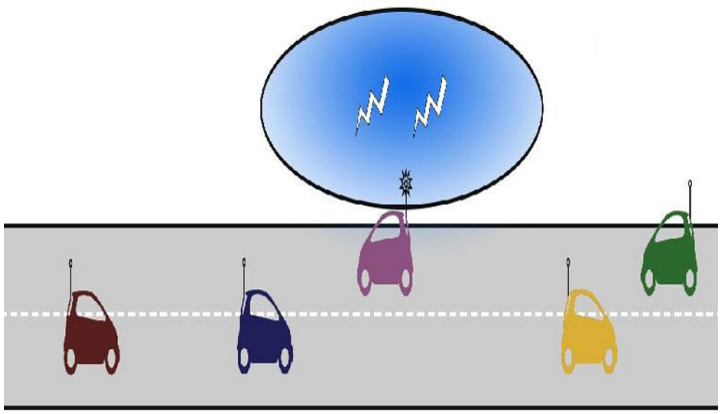
2) Traffic View Technique

Its main objective is to gather and disseminate information about the position and speed of vehicles. The information is restricted only to the vehicles positioned ahead of the current vehicle. The approach for message exchange is very similar to that used in SOTIS: vehicles periodically broadcast reports (contained in a single packet) about themselves and other vehicles they know about. Whenever a vehicle receives a report, it updates its stored information, and sends the updated report in the next broadcast period. Although the average size of the stored records is very small (on the order of 50 bytes) data aggregation is performed in order to fit all information in a single packet. Performance evaluation of Traffic View was carried out using the ns-2 network simulator and the CORSIM vehicular traffic simulator. The main drawbacks of this techniques it very costly and provide a great overhead to the network also this technique don't have the capacity to select the vehicle that will be in charge of disseminating the detected traffic congestion conditions to approaching vehicles or road authorities (e.g. through nearby road side units or cellular links).

3) SOTIS Technique

A Self-Organizing Traffic Information System (SOTIS) works within an approximate radius of 50 to 100 km of an individual user, even if as few as only 2% of all vehicles are using it. The precision of the information it provides decreases as the distance to the area of interest increases. A distributed receive-analyze-send algorithm. The information received from other vehicles is first analyzed, and only results of the analysis are transmitted. Roads are divided into variable size segments, and the information is exchanged on per segment basis. Each vehicle stores information for every segment of the road in its local database called Knowledge Base (KB). A GPS-based time-stamp is used to determine the accuracy of the information: more recent reports are assumed to be more accurate, and thus replace older ones. The reports are exchanged between vehicles travelling in both directions. The system was evaluated

using the ns-2 network simulator and a simple vehicular traffic simulator based on cellular automata. SOTIS is a good technique for aggregating traffic information these systems cannot be deployed in the near future, as one has to wait until the necessary market penetration of V2V communications technologies has been reached. The drawback of this approach is that the selection of the cluster head usually generates additional signaling overhead. In addition, it is important to emphasize that the definition of road segments is usually challenging. In fact, many techniques define road segments based on the vehicle's transmission range, but this might significantly vary, in particular when applying transmits power and congestion control protocols.



Cars with SOTIS system

4) Miller Technique

In the Miller technique or V2V2I architecture, the transportation network is broken into zones in which a single vehicle is known as the Super Vehicle. A zone can be as granular as desired, though it assume that the zones consist of sections of a freeway or individual lanes of sections of a freeway. Only Super Vehicles are able to communicate with the central infrastructure or with other Super Vehicles, and all other vehicles can only communicate with the Super Vehicle responsible for the zone in which they are currently traversing. That technique represent the freeway system as a

graph with edges consisting of sections of a freeway system and the weights of the edges being determined by the amount of time to traverse that section of the freeway with current speeds. They perform an analysis using FreeSim to determine how accurate the represented freeway system. This technique proposes that only one vehicle in each road segment is in charge of collecting and aggregating road traffic data. This information is then transmitted to adjacent road segments. However, the selection of the vehicle responsible for the data aggregation usually generates additional signaling overhead. The techniques described require the periodic exchange of packets different from the beacon messages already included in the IEEE802.11p/WAVE or ITSG5A standards.

Non Periodic Message Techniques

1) Street Smart

The Street Smart Traffic technique proposes to aggregate traffic information using distributed clustering algorithms with an epidemic diffusion model. It is designed to perform well even if only a small fraction of vehicles participates in the system. Moreover, the system does not require constant connectivity. Each vehicle records its speed and on this basis builds a local traffic map. Vehicles that are close to each other exchange their speed maps. Data aggregation is performed using clustering techniques, which combine related recordings of an unusual speed. The system was evaluated using the authors own simulator of Manhattan's grid of highways and a random way point mobility model. Street Smart obviously reduce the overhead generated by traffic messages than other techniques discussed where Street Smart limits the exchange of traffic information to only situations of unexpected or abnormal traffic conditions, e.g. traffic jams. VANET congestion systems had limited scope. The Traffic View project focused the congestion of the road directly ahead. The Traffic View project was able to demonstrate that it is possible to monitor vehicle congestion using a real VANET. The idea was extended to both

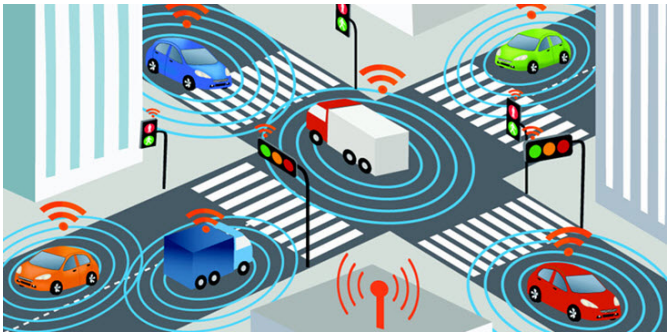
sides of the road by SOTIS. This technique the first to address the problem of discovering traffic on a road network.

2) Vaqar and Basir Technique

The traffic information gathered by a node in an ad hoc network is viewed as a snapshot in time of the current traffic conditions on the road segment. This snapshot is considered as a pattern in time of the current traffic conditions. The pattern is analyzed using pattern recognition techniques. A weight-of-evidence-based classification algorithm is presented to identify different road traffic conditions. The algorithm is tested using data generated by microscopic modeling of traffic flow for simulation of vehicle or node mobility in ad hoc networks. Test results are presented depicting different percentage levels of vehicles equipped with communication capability. The mechanism reported by Vaqar and Basir reduces the risk of communications overload by only estimating traffic congestion locally at each vehicle using pattern recognition techniques that exploit the beacon messages received from nearby vehicles. However, the lack of mechanisms to validate or correlate the traffic congestion estimates among various vehicles may lead to unreliable detections.

3) Lin and Osafune Technique

The technique relates to traffic condition detection by vehicle-to-vehicle communication systems. More particularly, the present technique relates to a method and apparatus for detecting and diffusing traffic condition information by distributed vehicle-to-vehicle communication systems. This system achieved by the method and apparatus according to the present invention as defined by the independent claims. The dependent claims relate to preferred embodiments of this technique.



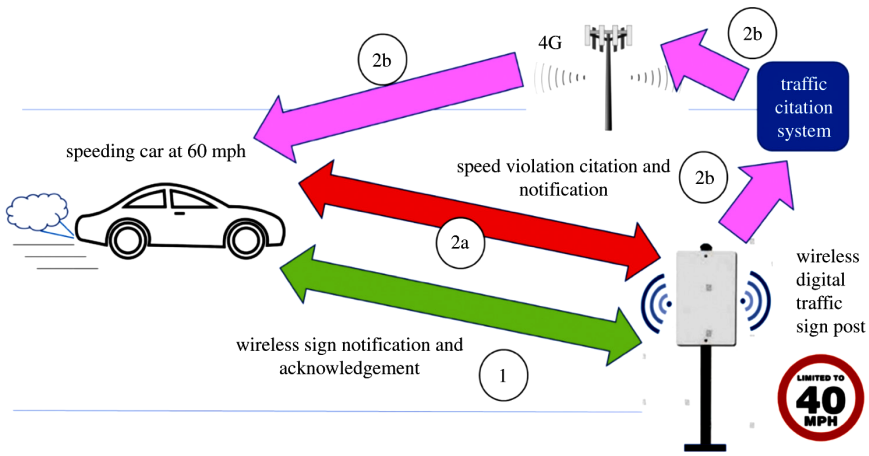
The present technique proposes a Wireless communication system, such as a Wireless vehicle-to vehicle communication system, by which a traffic condition can be determined. Exemplary traffic conditions that can be determined are free flow of traffic, traffic jam, and complete halt and/ or restricted flow of traffic. This technique makes a voting procedure so that neighboring vehicles exchange their traffic estimates and try to reach a consensus decision. A cooperative detection process that calculates the number of vehicles in a traffic jam using a tree based counting algorithm. However, the formation and management of the tree requires the exchange of a large number of packets, with the consequent risk of overloading the communications channel.

Table 1: Traffic Monitoring Techniques

Technique	Congestion Detection	Detection correlation	Congestion Level	Traffic Jam Length	Limited Overhead	Dissemination
COS	Yes	Yes	Yes	Yes	No	No
Traffic View	Yes	Yes	Yes	Yes	No	No
SOTIS	Yes	Yes	Yes	Yes	No	No
Miller	Yes	Yes	Yes	Yes	No	No
Street Smart	Yes	Yes	Yes	Yes	Yes	No
Vaquar &Basir	Yes	No	Yes	No	Yes	No
Lin & Osafune	Yes	Yes	No	No	Yes	Yes

1.4 VANET-ENABLED IN-VEHICLE TRAFFIC SIGNS

Understanding traffic sign information correctly is crucial. It helps the driver to anticipate future situations, make decisions and respond in an appropriate way. There are many different kinds of road signs and they are mostly placed above or beside highways and streets. However, these traditional static traffic signs have known limitations. The period of time the drivers have to analyze the information is limited, and even if the road signaling is predominantly standardized, most of the signs use text to convoy meaning restricting the understanding to readers of the language. They are likely to be overlooked during complex driving tasks, and sometimes, the adverse weather conditions or vehicles blocking the line-of-sight between the driver and the sign, make its recognition very difficult. In contrast to traditional traffic signs, traffic signs displayed within the vehicle will solve a big amount of these limitations, as well as they will provide additional help to the driver on his driving tasks. In fact, in-vehicle traffic signs are one of the main ITS technologies.



Regulatory signs play an important role in road safety and traffic efficiency. Traffic light systems are used to regulate and control conflicts between opposing vehicular, or pedestrian traffic movements. While these systems can improve junction capacity and road safety, they also still have some limitations. Most of the

signal control systems rely on timing plans generated offline by traffic engineers using optimization models. Traffic, however, seems to be an adaptation problem rather than an optimization problem. Improperly operated traffic lights cause excessive delays that sacrifice productivity, waste fuel, pollute the air, and increase the levels of stress.

Traditional traffic signs have static information and cannot be easily updated to cope with the frequent changes in the environment. Once the information is placed, it remains there until the sign is removed or replaced. Modern traffic signs should allow faster adaptation to the current road or traffic situations like icy roads or accident warnings.

Temporary road works, road accidents, or natural disasters are responsible for changes in the road network topology. Nevertheless, traffic should be able to recover rapid and efficiently from these changes.

Drivers should be informed about the best actions to follow in order to achieve a efficient traffic flow. In a simple situation of road works, static traffic signs are placed in a near position to inform the driver about the new action to follow. But, in road accidents or disaster situations, placing static traffic signs in order to provide rapid and efficient traffic adaptation is impracticable.

Mobile Ad-hoc Networks (MANETs) are often associated to crisis management applications, such as in a disaster recovery, where the entire communication infrastructure is destroyed and establishing communications quickly is crucial. As VANETs are considered one of the MANETs real-life applications, we consider these networks as a key technology to provide, not only rapid creation and adaptation of wireless networks (that already inherits from MANET features), but also as a mean to rapidly and spontaneously create new traffic rules that provides fast adaption in traffic change situations. For instance (see Figure 4), imagine a lane obstruction caused by an accident in a road which has two lanes, one in each direction. As one of the lanes is blocked, vehicles approaching the zone of the obstruction must

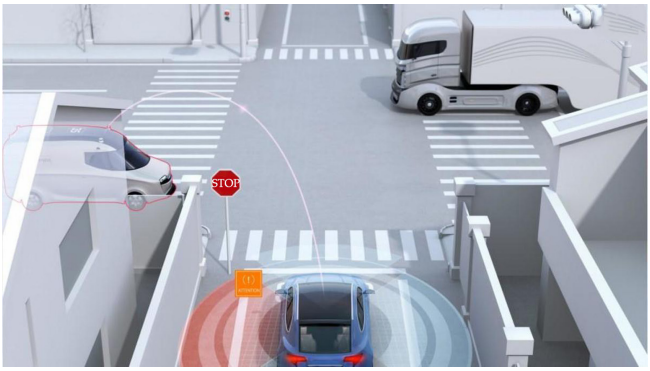
act cooperatively to create a just-in-time and just-in-place virtual traffic light to regulate and control conflicts between vehicles in different lanes, improving the traffic efficiency and safety in the only lane available. With the introduction of intelligent vehicles, we believe that road transportation and traffic management also needs intelligent traffic signs. Hence, we propose a novel approach to spontaneously create traffic signs through the collaboration of the local vehicles, in a synchronized manner, but without a centralized control infrastructure.

Figure 4: Lane obstruction example

Intersection Management with Virtual Traffic Lights

Virtual traffic lights can be used not only at intersections, but in a variety of environments where the crossing conflicts needs to be solved. Regarding the lane obstruction example illustrated in Fig. 5, virtual traffic lights can be used to coordinate the traffic flow in the only two available lanes.

However, we can see this example as an intersection scenario, where the exact location of the accident is the center of the junction, and the two available lanes represent two entry roads. In fact, every crossing conflict between vehicles can be seen as a conventional intersection of roads. When an accident occurs, the involved vehicles must inform their vicinity about the exact location of the accident. Once the neighbors are informed, they must temporary update its map topology assuming that a new intersection is now present in the exact location of the accident. Then, the virtual traffic light is created as if an intersection really exists in that location.



There are several goals that can be taken into consideration when designing an intersection control system based on traffic lights. As we envision a system where the traffic light infrastructure does not physical exists, a distributed and cooperative protocol must solve the crossing conflicts, minimize the average delay of vehicles approaching an intersection, reduce the queue length of all approaches, and even reduce overall fuel consumption and pollutant missions. Nevertheless, traffic does not follow a well defined pattern, and thus, traffic lights should be adapted to different traffic demands. In the example illustrated in Figure 5(a), when the vehicle A approaches the intersection and notices (looking to its LT) that there are no other vehicles moving to cross the same intersection, there is no need for traffic lights. However (Figure 5(b)), when vehicle A notices the presence of vehicle B, which is driving to the same intersection, the two vehicles must solve this conflict with a traffic light.

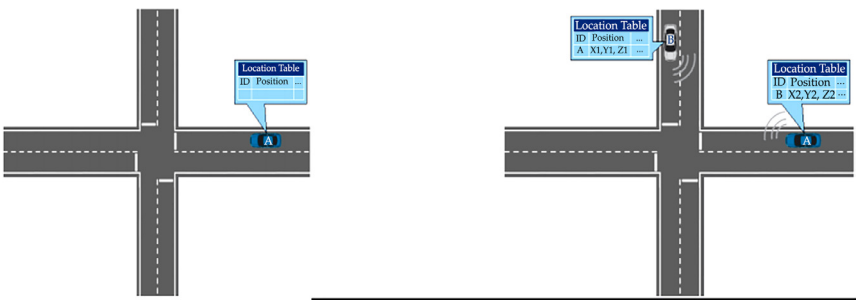


Figure 5: Vehicles approaching an intersection.

The decision process that might create a Virtual Traffic Light (VTL) at an intersection is illustrated in Figure 6. When a vehicle approaches an intersection and notices that are neighbors attempting to cross the same intersection, it must communicate cooperatively with them in order to create a VTL just-in-time and justin-place. However, as such VTL does not physically exist; a distributed algorithm is required to ensure that all vehicles can be aware of the traffic light. Thus, as result of the cooperation between all the vehicles that are approaching the intersection, one of them must be chosen to create the VTL and to propagate it - called Leader. This temporary elected Leader will work as an infra-structure and has the responsibility of controlling the VTL. Once the Leader is elected and starts broadcasting (periodically) the VTL information to its vicinity, the other vehicles act as passive nodes in the protocol (just listen the channel for traffic lights).

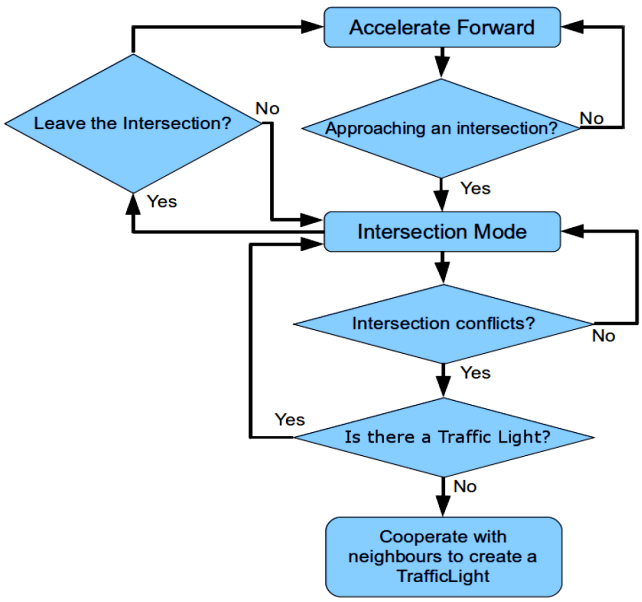


Figure 6: Flowchart representing the process for creating a virtual traffic light in an intersection

REFERENCES

1. Altayeb, M., & Mahgoub, I. (2013). A survey of vehicular ad hoc networks routing protocols. International Journal of Innovation and Applied Studies,3(3), 829-846.
2. Bako, B., & Weber, M. (2011). Efficient information dissemination in VANETs. INTECH Open Access Publisher.
3. Da Cunha, F. D., Boukerche, A., Villas, L., Viana, A. C., & Loureiro, A. A. (2014). Data communication in VANETs: a survey, challenges and applications (Doctoral dissertation, INRIA Saclay).
4. Emery, D., & Hilliard, R. (2009, September). Every architecture description needs a framework: Expressing architecture frameworks using ISO/IEC 42010. In Software Architecture, 2009 & European Conference on Software Architecture. WICSA/ECSA 2009. Joint Working IEEE/IFIP Conference on(pp. 31-40). IEEE.
5. Hartenstein, H., & Laberteaux, K. P. (2008). A tutorial survey on vehicular ad hoc networks. Communications Magazine, IEEE, 46(6), 164-171.
6. Jakubiak, J., & Koucheryavy, Y. (2008, January). State of the art and research challenges for VANETs. In Consumer communications and networking conference, 2008. CCNC 2008. 5th IEEE (pp. 912-916). IEEE.
7. Kumar, R., & Dave, M. (2012). A review of various vanet data dissemination protocols. International Journal of u-and eService, Science and Technology,5(3), 27-44.
8. Liang, W., Li, Z., Zhang, H., Wang, S., & Bie, R. (2015). Vehicular ad hoc networks: architectures, research issues, methodologies, challenges, and trends. International Journal of Distributed Sensor Networks, 2015, 17.
9. Maier, M. W., Emery, D., & Hilliard, R. (2001). Software architecture: introducing IEEE Standard 1471. Computer, 34(4), 107-109.

10. Maier, M. W., Emery, D., & Hilliard, R. (2004). ANSI/IEEE 1471 and systems engineering. Systems Engineering, 7(3), 257-270.
11. Nekovee, M., & Bogason, B. B. (2007, April). Reliable and efficient information dissemination in intermittently connected vehicular adhoc networks. In Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th (pp. 2486-2490). IEEE.
12. Panichpapiboon, S., & Pattara-Atikom, W. (2012). A review of information dissemination protocols for vehicular ad hoc networks. Communications Surveys & Tutorials, IEEE, 14(3), 784-798.
13. Ranjan, P., & Ahirwar, K. K. (2011, January). Comparative study of vanet and manet routing protocols. In Proceedings of the International Conference on Advanced computing and communication Technologies (ACCT 2011).
14. Willke, T. L., Tientrakool, P., & Maxemchuk, N. F. (2009). A survey of inter-vehicle communication protocols and their applications. Communications Surveys & Tutorials, IEEE, 11(2), 3-20.
15. Yousefi, S., Mousavi, M. S., & Fathy, M. (2006, June). Vehicular ad hoc networks (VANETs): challenges and perspectives. In ITS Telecommunications Proceedings, 2006 6th International Conference on (pp. 761-766). IEEE.

CHAPTER 2

MESSAGE SETS FOR VEHICULAR COMMUNICATIONS

INTRODUCTION

VANET technology includes Vehicle-to-Vehicle and Vehicle-to-Infrastructure communications. It supports the realization of a large variety of Cooperative Intelligent Transport System (C-ITS) applications and services by enabling real-time data exchanges between vehicles, between vehicles and infrastructure systems. C-ITS technologies extend the driver perception to the traffic ahead, avoid potential road hazard situations, collision risks and improve traffic efficiency. Automobile and road operator stakeholders in Europe and in North America have been jointly driving research and development of the C-ITS for more than a decade. Message sets specifications, standardization, and validation consist of one key activity for VANET technology development in world wide. Even though some differences are observed in technical features of message sets in order to satisfy regional specific requirements, commonality is often found in terms of basic features and application usages.

The Cooperative ITS (C-ITS) technologies include Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. They enable the real-time information exchanges between vehicles, between vehicles and infrastructure systems, in order to extend the driver perception to the traffic ahead and avoid potential road hazard situations. A variety of services may be provided to road users with C-ITS, categorized in road safety, traffic efficiency, environment friendly and infotainment services. VANET messages and the corresponding message exchange protocols enable exchange of application data in vehicular communication networks. A large set of real-time raw data or processed data are transported in message sets.

The C-ITS deployment does not only rely on the technical aspects such as message sets specifications, communication protocol design and access technology specifications, etc., but also on many non-technical aspects such as organizational, legal, or operational aspects. C-ITS deployment requires cooperation among stakeholders from different domains including automobile industries, road operators, public authorities, technology providers, telecommunication operators, etc., to reach agreement on a common deployment roadmap at vehicle and infrastructure side. The roles and responsibilities of different stakeholders need to be defined. These nontechnical aspects have strong impacts on final choice of the appropriate technical solutions. For example, car makers and road operators work together to define data sharing needs and to set requirements on data quality and information dissemination. These requirements are guiding the message content definition, message exchange protocol design as well as the applications that may be realized.

2.1 APPLICATION REQUIREMENTS

A wide variety of applications may be developed using C-ITS technologies. There exist in the research community many methods to classify the application categories. One popular method is the classification based on Time-To-Collision (TTC) parameter. TTC

is a parameter that measures the traffic conflict probability as the time required for two vehicles to collide if they continue at their present speed and on the same path. It has been proposed by Hayward.

At its establishment in 2008, ETSI TC ITS has defined a Basic Set of Application (BSA), that may be enabled by C-ITS technologies and deployable within three years' time frame after the standards have been completed.

A high level classification of the BSA applications according to TTC is presented in Figure 1.

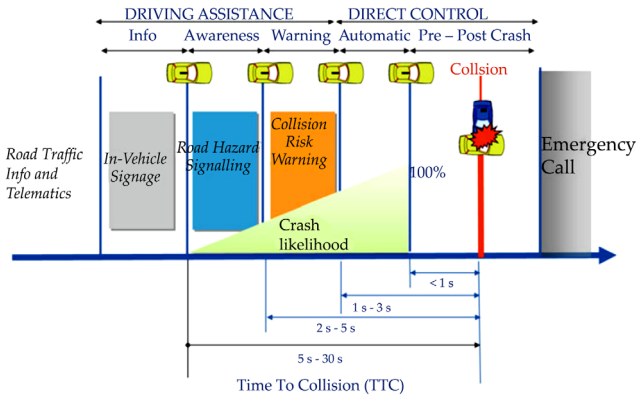


Figure 1 High level classification of applications based on TTC.

The TTC values are illustrated as examples.

- Road traffic info and telematics applications are telematics services provided by public operators or by private service providers. Upon reception of road traffic information, vehicles check the relevance of the received information and show to vehicle users as information, when no imminent danger is detected. This information may help users to adjust the navigation plan and be informed of road traffic status, to improve the driving comfort. At large scale, telematics services may also play a role in improving traffic efficiency and reducing road traffic pollution. An example of road traffic information

application is Transport Protocol Experts Group (TPEG) application.

- Driving assistance applications are proposed by automobile Original Equipment Manufacturer (OEM)s to assist driver in dense traffic or in dangerous situations where attention should be paid by driver to overcome potential safety risks. The information provided to driver may be a remind information for traffic rules (e.g., speed limit) or road conditions (traffic jam), an awareness information of detected road hazards (e.g., stationary vehicle at road side, hard break vehicle ahead), or even a warning message that requires immediate actions of driver to avoid potential collision. At this stage, there is no automatic control to the in-vehicle systems. Driver keeps the control on the vehicle maneuvering. An example of awareness application is Road Hazard Signaling (RHS) application. Examples for warning application are Longitudinal Collision Risk Warning (LCRW) application and Electronic Emergency Brake Light (EEBL) application.
- Direct control application takes over the control of vehicle by automatic control systems to avoid potential collision in case the TTC is further reduced. If the collision cannot be avoided, the pre-crash application may be launched to mitigate the collision impact and reduce the damage to vehicle driver or passengers. Automated driving applications are within this category.
- Post crash application consists of cooperating with rescue organizations to reduce the accident rescue latency time and improve the rescue efficiency. In Europe, eCall is a pan-European in-vehicle emergency call system for this purpose. It aims to deploy an in-vehicle device that will automatically dial 112 in case of road accident and establish voice connection with back-end office. In addition, a set of in-vehicle sensor data such as airbag deployment, impact sensor information, and Global Positioning System (GPS) coordinates will be

transmitted to local emergency agencies via the wireless communication infrastructure such as cellular networks. The emergency agencies can then organize the rescue plan properly.

The C-ITS standardization and validation activities in the EU are mainly focused on the driving assistance applications. During the standard development process, technology agnostic approach is used for the C-ITS applications and message sets specifications. It is assumed that any communication technology may be used to realize a C-ITS application, as long as it is suitable to satisfy the application functional and operational requirements. For example, ITS G5 technology at 5.9 GHz is considered by automobile stakeholders in the EU as main candidate technology to initiate the deployment of V2V-based road safety applications and a set of V2I applications. A MOU on deployment strategy for C-ITS in Europe has been signed by OEMs in Car-to-Car Communication Consortium (C2C-CC) to promote the ITS G5 enabled C-ITS and a set of day one applications. This is in order to ensure the interoperability between different implemented systems. This initiative has been followed up by main road operator association Amsterdam Group.

It has been demonstrated in various FOT projects that V2V and V2I technologies can address a large majority of road safety issues and as well improve the traffic efficiency. Table 1 summarizes a non-exhaustive list of C-ITS applications that are selected by different European initiatives, e.g., ETSI, DRIVE C2X, C2C-CC, etc.

In North America, US DOT National Highway Traffic Safety Administration (NHTSA) has announced in early 2014 that it will begin taking steps to enable DSRC V2V communication technology for light vehicles. Subsequently, in September 2014, NHTSA announced an Advanced Notice of Public Rulemaking (ANPRM) and published an accompanying research report on readiness of V2V technology for application. The US objective is to progress toward a Notice of proposed rulemaking (NPRM) in 2016, which formally solicits public commentary that will shape a mandate for V2V equipment, the suite of DSRC transceivers

and positioning equipment, to be on new production vehicles by a few years. Data analysis has been conducted based on data collected from 3,000 DSRC-equipped vehicles participating to one-year duration Safety Pilot Model Deployment in Ann Arbor, Michigan, USA. The results show significant reduction in crashes, as a consequence lives saved and economic benefit increased. This motivation, when coupled with emerging minimum performance standards and a proposed standardized solution for system security form the basis of a safety-of-life argument for a mandate to initiate C-ITS deployment in North America. Using the TTC classification in Figure 1 as a reference, the North American model for initial deployment thusly focuses almost exclusively on in the collision risk warning (2–5 s) window. To that end, the Safety Pilot Model Deployment was implemented and mined as significant data source to critically examine the technical performance of standardized DSRC messages and to project widespread crash avoidance or societal benefit for following V2V applications:

EEBL: Onset of braking from the forward vehicle is passed to the next vehicle, and a Human Machine Interface (HMI) is actuated to provide EEBL warning to the driver.

Forward Collision Warning (FCW): An algorithm to warn the driver and as the TTC continues to diminish, graduated intensity of HMI warning, followed by brake control is provided.

Blind Spot Warning/Lane Change Warning (BSW/LCW): Vehicles in the rear-view mirror blind spot or fast-approaching vehicles elicit a warning through the HMI.

Do Not Pass Warning (DNPW): A warning drivers in the presence of hazardous oncoming vehicles in a passing situation is provided through the HMI.

Intersection Movement Assist (IMA): As drivers move directly through or conduct a turning movement within an intersection, HMI alert with warnings of crossing-path hazards posed by vehicles approaching on other intersection legs is provided.

Left Turn Assist (LTA): At unprotected left turns (in countries where vehicles are driven on the right), warnings of vehicles approaching from the opposite direction are provided when TTC thresholds are exceeded.

The aforementioned ANPRM focused on just IMA and LTA as justification for a potential NHTSA mandate, most likely because the other crash types listed here are also being addressed through other NHTSA crash avoidance rulemakings and because these intersections crash categories represent significant risk.

Table 1. Examples of C-ITS application

Application	**Short description**
Longitudinal collision risk warning	Warns the driver when a longitudinal collision risk with neighbor vehicles is detected by processing the received messages from these vehicles.
Intersection collision risk warning	Warns the driver of a potential collision risk with other vehicles at an intersection area by processing the received messages from these vehicles.
Traffic light violation warning	Warns the driver of a potential traffic light violation if speed is not reduced, by processing the traffic light status information received from road side.
Lane change warning	Warns the driver who plans to change lanes if there is a vehicle in the blind spot or an overtaking vehicle.
Cooperative awareness	Vehicles are made aware of other vehicles' position, speed, and basic sensor status in its vicinity by processing the received messages from them.
Emergency vehicle approaching	Emergency vehicles transmit a message to other vehicles in its path for them to take appropriate actions to give priority.
Slow vehicle approaching	Vehicle driving at low speed on highway, e.g., a roadworks vehicle or a road surface cleaning vehicle transmits a message to announce its presence.
Emergency electronic brake light	A hard brake vehicle transmits a warning message to vehicles behind in order to avoid rear end collision.

Stationary vehicle	Stationary vehicle on the road surface due to accident, breakdown or other reasons transmits a warning message to oncoming vehicles from upstream traffic.
Hazardous location	Vehicle detecting hazardous road conditions e.g., obstacles on the road using its on-board sensors transmits a message to oncoming vehicles.
Roadworks warning	Roadwork vehicles or road side units transmit roadwork information as well as relevant speed limit information to oncoming vehicles.
Green light optimal speed advisory	Upon reception of traffic light phase and timing information from road side, vehicles may calculate optimal speed to cross the road intersection and avoid traffic light violation or reduce fuel consumption.
In vehicle signage	Road side infrastructure transmits static or dynamic road sign information to vehicles for in-vehicle presentation.
Speed limit information	Road side infrastructure transmits static or dynamic speed limit information to vehicles for in-vehicle presentation.

The first release of standards are close to finalization in the EU and in North America, in preparation of a day one deployment. At this first phase deployment, the system penetration rate in the overall car park and at road side may be limited and requires some time to progress. In addition, the communication reliability and quality of some key data e.g., position accuracy, may not be sufficient to develop direct control applications. On the other hand, new research initiatives have already been launched to study new features to support more advanced applications such as automated driving based on vehicular communication technologies. These research inputs will be used for future standard development.

2.2 MESSAGE SETS OVERVIEW

A VANET message refers to an Application or Facilities layer entity in the standardized ITS communication reference architecture, as

illustrated in Figure 2. A system that implements the ITS protocol stacks and ITS applications is denoted as an ITS station (ITS-S). An ITS-S may be integrated in a vehicle (vehicle ITS-S) or at road side (road side ITS-S). A vehicle ITS-S is also named as On Board Unit (OBU), a road side ITS-S is also named as Road Side Unit (RSU).

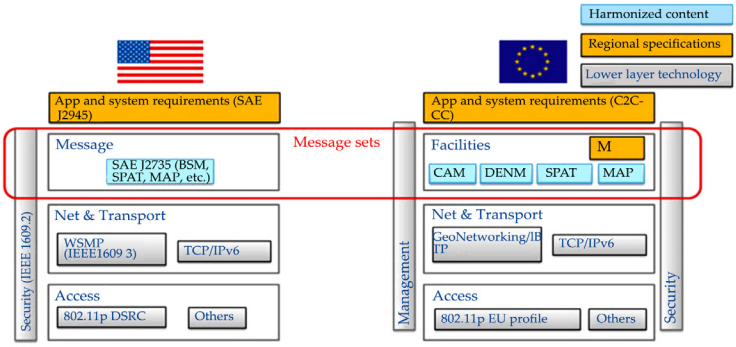


Figure 2. Message sets in reference ITS communication architecture.

In general, specifications of a message include a data dictionary definition (or message format definition) that defines the syntax and semantics of the message, and a related communication protocol for the message exchange. A message entity relies on the lower layer protocol stacks for dissemination. The Service Access Point (SAP)s enable data exchange between layers. The exchanged data via SAP for message transmission includes the message payload as well as communication requirement parameters. Figure 2 illustrates the set of messages developed in the EU and North America as well as lower layer standards applied for the message dissemination.

Stakeholders both at vehicle side and infrastructure side are cooperating together for message set specifications. The targeted applications include road safety, traffic efficiency, and added value applications. Among these applications, road safety applications represent the most stringent requirements in terms of time latency, data quality, communication reliability, and system reliability. A set of V2V and V2I messages are specified, in order to support the selected set of applications for deployment. Standardization of message format and communication protocols plays an essential

role to ensure the communication interoperability between vehicles as well as between vehicles and infrastructure systems. This standardization work needs to take into account:]

- Application requirements that define data exchange needs and communication requirements between ITS-Ss;
- Message format and data dictionary that defines unambiguous syntax and semantics of data being included in a message, in order to avoid any potential misinterpretation at receiving ITS-S;
- Communication capacities in order not to bring unnecessary overload traffic to communication channel; and
- Harmonization of message sets. The harmonization mainly includes two aspects: On the one hand, harmonization of message sets standards at International level enables cost reduction for industrial implementation at different markets. On the other hand, harmonization with existing relevant standards in the same region, in particular the existing road infrastructure message sets standards, would facilitate the interoperability between vehicles and road side infrastructures.

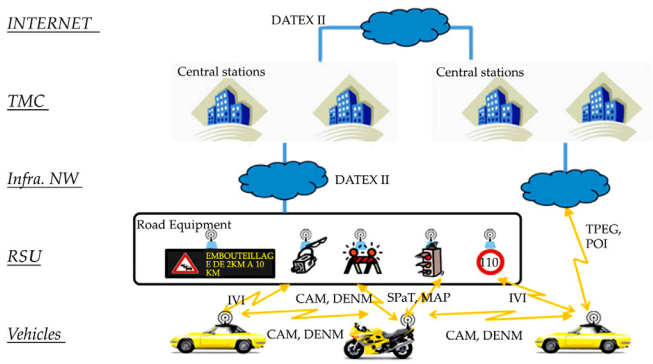


Figure 3. High level overview of ITS protocols used in the EU.

For illustration purpose, Figure 3 presents a high level overview of ITS message sets and application layer protocols that are used in ITS domain in the EU. In North America, similar architecture

may be observed, with corresponding message sets and protocols applied locally.

We will provide detailed description for message sets that are used for VANET networks, i.e., Cooperative Awareness Message (CAM), Decentralized Environmental Notification Message (DENM), Basic Safety Message (BSM), Signal Phase and Timing (SPAT), Map data (MAP), and In Vehicle Information (IVI) message. For other message sets and application protocols included in Figure 3, they are not used in VANET, but the harmonization with these message sets are taken care during the specifications phase.

The message sets is summarized as follows:

- CAM is specified by ETSI TC ITS for European deployment. It is a heartbeat message that is transmitted periodically from OBU and RSU to announce the position, movement, and basic attributes of the transmitting ITS-S.
- DENM is specified by ETSI TC ITS for European deployment. It is an event driven message that is transmitted from OBU or RSU at the detection of a traffic event or road hazard.
- BSM is specified by SAE TC DSRC for North America deployment. It provides functions equivalent to CAM and DENM adapted to targeted applications selected in North America.
- SPAT is specified by SAE TC DSRC, it has been extended for deployment in North America, EU, and Japan. It is transmitted from an RSU to provide phase and timing information of one or a set of traffic lights.
- MAP is specified by SAE TC DSRC, it has been extended for deployment in North America, EU, and Japan. It is transmitted from an RSU to provide road topology and geometry information of a road segment or an intersection area. It is used by a receiving ITS-S to map the SPAT data to the intersection topology.

- IVI is initiated by CEN and specified by CEN/ISO as world side standard. It is transmitted from an RSU to provide static or dynamic road signage information for in-vehicle presentation
- DATEX 2 is a web service based application protocol already deployed in the EU for the communication between Traffic Management Center (TMC)s to exchange road traffic or traffic management information. DATEX 2 specification includes a set of data dictionary for the description and definition of traffic situation, traffic rules, and road topology information. This data dictionary is considered during the I2V message specifications, e.g. the IVI message.
- TPEG specifications provide traffic and travel information services to road users. TPEG messages and protocols are developed by stakeholders in Traveler Information Services Association (TISA) and adopted as International Organization for Standardization (ISO) standards for Road Traffic and Traveler Information (RTTI) Service. TPEG also includes a data dictionary for road traffic and traveler information. In particular, the traffic event data dictionary is considered during the specifications of event-driven message DENM.
- POI refers to a set of messages and application protocols that provide Point of Interest Information to road users. A POI message set has been specified in ETSI TC ITS. For example, the POI message for electric vehicle charging spot information provides up-to-date availability and characteristics of charging stations to electric vehicle users; The POI message for tire pressure pump and gauge station provides availability of the tire pressure gauge and pump station information to road users.

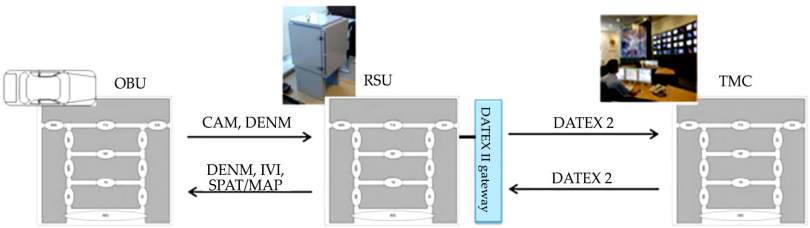


Figure 4. RSU gateway function for the inter-networking between VANET and road infrastructure network.

In the EU, road operators are testing functionalities to interconnect VANET and road infrastructure networks. In one representative application, VANET messages transmitted from vehicles are collected by RSU for aggregation then provided to TMCs as probe data. For example, in French FoT project SCORE@F each RSU collects CAM, DENM messages transmitted from vehicles in its vicinity, then realizes a simple processing of the collected data to derive local traffic status such as average driving speed, average travel time, etc. The aggregated traffic status data is then transmitted up-link to TMC using DATEX 2 protocol. In down-link, a TMC may send a DATEX 2 message to an RSU, providing traffic management, speed limit, and traffic status information that is relevant to the local traffic close to the RSU communication coverage range. After processing the received DATEX 2 message, the RSU generates a VANET message accordingly and transmits it to vehicles located nearby. A gateway function is implemented in RSU to interface the VANET and road infrastructure network, as illustrated in Figure 4. Such gateway function further facilitates the vehicle and infrastructure integration.

Table 2. Matching of ITS applications and VANET messages

Application	CAM	DENM	BSM	SPAT	MAP	IVI
Longitudinal collision risk warning	X	X	X	_	_	_
Intersection collision risk warning	X	X	X	X	X	X

Traffic light violation warning	X	X	X	X	X	_
Lane change warning	X	_	X	_	_	_
Cooperative awareness	X	_	X	_	_	_
Emergency vehicle approaching	X	X	X	_	_	_
Slow vehicle approaching	X	X	X	_	_	_
Emergency electronic brake light	X	X	X	_	_	_
Stationary vehicle	X	X	X	_	_	_
Hazardous location	_	X	X	_	_	_
Roadworks warning	X	X	_	_	_	X
Green light optimal speed advisory	_	_	_	X	X	_
In-vehicle signage	_	_	_	_	_	X
Speed limit information	_	_	_	_	_	X
FCW	X	X	X	_	_	_
BSW/LCW	X	_	X	_	_	_
DNPW	X	X	X	_	_	_
IMA	X	X	X	X	X	X

The transmission and dissemination of a VANET message rely on the lower layer protocol stacks. In the EU, the GeoNetworking and Basic Transport Protocol (BTP) stacks are specified, enabling a multi-hop, geographical position-based addressing scheme for message dissemination. This allows the dissemination of data packets to a geographical area according to the ITS application needs.13 In North America, IEEE 1609.3 standard specifies a network layer protocol for DSRC message dissemination in one hop dissemination mode. For access technology, IEEE 802.11p protocol at 5.9 GHz spectrum provides an ad-hoc and low latency access to the radio channel, enabling broadcast and unicast of data packets without the needs of communication infrastructure.

In addition, other protocol stacks such as legacy IPv6 protocol and cellular technologies may also be used. This is made possible thanks to the technology agnostic approach adopted for message sets standard development.

2.3 COOPERATIVE AWARENESS MESSAGE

In Vehicular Ad-hoc Networks (VANETs), vehicles gather Cooperative Awareness Messages (CAMs) sent from other vehicles via wireless broadcast. Each received message has to be processed by an in-vehicle system. In series implementations, such an in-vehicle system needs to cope with limited resources whose capacity is not yet defined. Therefore, information about the received CAM rate is a crucial input for the development of series VANET products. CAMs from distant vehicles are less likely to be received than those of nearby vehicles. Designers of applications leveraging CAM information are interested in the frequency of received CAMs originating from vehicles depending on their distance. We study future CAM rates depending on various parameters. We set up a road traffic simulation for selected highway scenarios. We estimate the rates of generated CAMs and introduce the notion of relative channel load. We present a new approximative channel model to determine a vehicle's message reception probability. That model is used to simulate the rates of received CAMs for each vehicle. Moreover, we investigate the origin of received CAMs and times between consecutive CAMs received from the same sender (inter-reception times) depending on distance. Most results depend on the penetration rate of VANET technology that will increase in the near future. We derive approximative formulae and use them to validate our simulation results. They are quite accurate, and so they may also serve for simple forecasts. The results from our analysis show that the rates of generated and received CAMs lead to several challenges for the design of an efficient and robust VANET implementation.

The generation, transmission, and management of CAM is realized by the Cooperative Awareness (CA) basic service component. The

CA basic service specifications have been initially introduced by European C2C-CC as a European Automobile stakeholder's joint effort. The CA basic service has been prototyped, validated, and tested in multiple EU FOT projects. Technical findings and lessons learned of these initiatives are fed back to ETSI, which has developed a European Norm (EN). CAM is a core message that is required to support the day one deployment of VANET system in Europe.

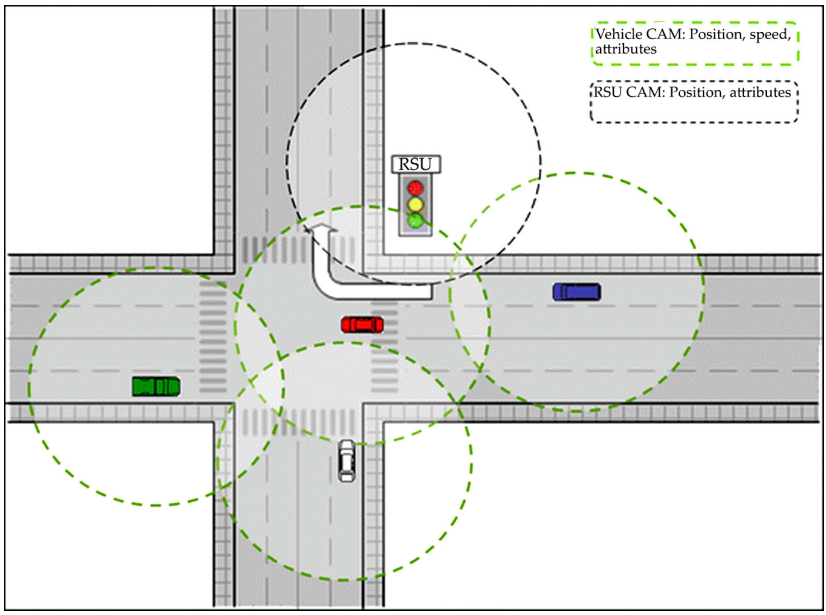


Figure 5. Example scenario of CAM transmission.

2.3.1 The CA Basic Service Overview

CAM contains real-time vehicle data or RSU data. It is transmitted with high frequency from an OBU or RSU to other OBUs or RSUs located in the close vicinity, in order to create and maintain awareness of each other and to support cooperative performance of vehicles using in the road network.

CAM is mainly designed for road safety applications. In one example use case where ITS G5 technology is used, an ego vehicle

receiving CAMs from the surrounding vehicles has awareness of the movement and basic sensor status of these vehicles. The ego vehicle can therefore detect in a short time latency an abnormal maneuvering of the surrounding vehicles such as a hard brake situation. The transmitting vehicle also provides its historical path and path prediction information in CAM, allowing the receiving ego vehicle to estimate the probability of path crossing with the transmitting vehicle in order to estimate the collision risk. Once a safety risk is detected, a warning is delivered to the driver of the ego vehicle, who may take appropriate actions to reduce the risk and improve the driving safety.

CAM may also be used in non-safety applications. In particular, CAMs transmitted by vehicles may be collected by RSUs and forwarded to TMC. TMC may further process the received CAM data for traffic monitoring and traffic management applications.

2.3.2 CAM Dissemination and Transmission Protocol

The ITS G5 technology is considered as main technology for CAM dissemination in support of road safety applications. However, CAM dissemination is not limited only to the ITS G5 technology. For example, for CAM collection by TMC application, cellular network may be used to transmit CAM and DENM directly to the TMC as road traffic floating car data.

CAM Dissemination

In case ITS G5 is used, CAM is broadcasted over Control Channel (CCH). The Single-Hop Broadcast (SHB) protocol of the GeoNetworking/BTP is used for CAM dissemination. As consequence, CAM is broadcasted to vehicles located in the direct communication range of the transmitting node. Given the high update and transmission frequency, the CAM packet life time set for the SHB protocol is set to a small value (1s) to avoid unnecessary queuing at network layer. The priority for CAM is set to a high value. This is to guarantee a prioritized access to the radio resource

for CAM transmission. The CAM transmission is independent to any specific applications. Its transmission is activated as long as the vehicle is located in the public road and the OBU is activated.

CAM Transmission Protocol

The CAM transmission frequency varies between 1 and 10 Hz. This requires the OBU to collect up-to-date information at least 10 Hz rate for CAM construction. Typically, the construction of a vehicle CAM requires the CA basic service to access to geographic positioning systems such as GPS and to the in vehicle network, e.g. Controller Area Network (CAN).

A protocol dynamically adjust the CAM transmission interval between the upper limit (1,000 ms) and lower limit (100 ms), according to the vehicle dynamics and the ITS G5 channel congestion status. This protocol defines a set of conditions under which a new CAM shall be generated, as follows:

1. If the absolute difference between the current heading of vehicle and the heading included in the previous generated CAM exceeds 4 deg, or
2. If the distance between the current position of the vehicle and the position included in the previous generated CAM exceeds 4 m, or
3. If the absolute difference between the current speed of the vehicle and the speed included in the previously generated CAM exceeds 0.5 m/s, and
4. The transmission interval is equal or greater than the allowed transmission interval set by the Decentralized Congestion Control (DCC) functionality. DCC includes a set of functionalities including transmission interval control in order to limit the channel congestion level for ITS G5 radio channels under a target threshold. Its standardization work is ongoing in ETSI TC ITS.
5. In addition, if the CAM transmission interval is reduced, i.e. the vehicle dynamics is reduced, the same transmission interval value should be maintained for at least a

predefined number of consecutive CAM generations. This is to improve the CAM reception probability even at vehicle low dynamics situation, in particular when the packet lost may be increased in some radio propagation conditions, e.g. in non-Line-of-Sight (LoS) conditions at road intersection area.

As for RSU CAM, the default transmission interval is set to 1 Hz. However, this may be adjusted depending on the application needs. For RSU CAM that provides CEN DSRC position, 1 Hz transmission rate is set to allow vehicles entering the direct communication range of the RSU have opportunity to receive at least one CAM during the passing-by time period.

CAM Format and Data Requirements

Depending on the type of transmitting node (e.g., light vehicle, public transport vehicle, or RSU, etc.), a CAM provides:

- ITS-S attributes such as vehicle type, vehicle role (e.g. emergency vehicle), vehicle size, road side equipment type information.
- ITS-S movement status such as position and time, moving speed, heading and the path history of the vehicle.
- Vehicle basic sensor data including acceleration status, exterior light status, steering wheel angle, yaw rate information, vehicle moving path curvature, etc.
- Special vehicle information including additional data for special vehicles such as light bar and siren status for emergency vehicle, roadwork type information for roadwork trailer cars, etc.

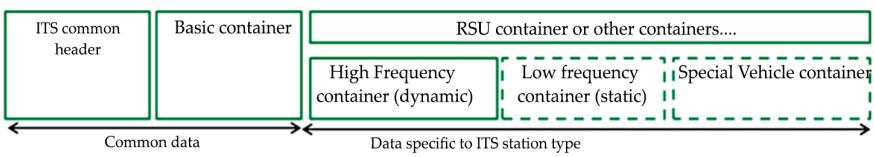


Figure 6. Overview of CAM structure.

A CAM is constructed with a set of data containers, either mandatory or optional. Mandatory container mainly contains highly dynamic data, i.e. vehicle movement status and vehicle basic sensor data. Mandatory container is present in all CAMs transmitted by an ITS-S. Optional containers contain low dynamic data of the vehicle or data that is available under specific conditions when vehicle is operating in a specific role like roadwork vehicle, emergency vehicle, public transport vehicle, etc. Figure 6 provides an overview of the CAM format as defined in. The main purpose of container-based CAM format design is to enable the flexibility of the message structure for future extensions of the message content, when new application needs are identified. Optional containers reduce unnecessary transmission of the low dynamic data at high frequency, in order to reduce the message size for ITS G5 communications.

The Abstract Syntax Notation One (ASN.1) unaligned Packed Encoding Rules (PER) are used for CAM encoding and decoding, in order to optimize the message size for ITS G5 technology.

One main performance indicator for CAM data quality is sensor data freshness. Data freshness indicates up-to-date level of the provided data, essential for the estimation of collision risk. Each CAM is time stamped according to the position data. Other sensor data included in CAM should ensure that the data age with regard to the timestamp is less than a certain predefined threshold. In V2V road safety application requirement standards defined in ETSI TC ITS, the data age of high dynamic sensor data is used to classify the OBU performance into two categories (class A and class B system). A threshold data age of 150 ms is used for this classification, denoted as the time interval between time at which a sensor data is available and the time at which a CAM is time stamped. This threshold is derived from the assumption that an overall end-to-end time latency of 300 ms is required for CAM dissemination, so that a receiving vehicle is able to perceive the vehicle dynamics or any potential sudden maneuvering in time for the realization of a collision risk warning application. Otherwise, an awareness information should be provided to driver, when applicable.

In addition to the data age, another important performance indicator is the position and time accuracy. Position and time accuracy determines to which level the ego vehicle may correctly estimate its relative position with regard to the transmitting vehicle (lane alignment or road alignment). Depending on the application requirements, the position accuracy requirement may vary. Typically, for an application that relies on CAM to estimate the longitudinal collision risk, the position accuracy requirement is set to 1 m, so that the ego vehicle can judge if it is located in the same lane as the transmitting vehicle. Otherwise, for applications that provide awareness information to driver, the position accuracy requirement can be relaxed to 10–15 m.

CAM Security

The security mechanism for CAM dissemination considers the authentication of messages transferred between ITS-Ss with certificates. The CAM signing and verification are processed at lower layer GeoNeworking/BTP protocol stack. The objective of this mechanism is to provide authentication service not only to the CAM message payload, but also to the GeoNetworking packet headers, which contain also security sensitive data such as node position information.

Table 3 presents an overall message frame for a secure CAM, when the GeoNeworking/BTP protocol stack is used for CAM dissemination. A CAM is delivered to or received at the GeoNeworking/BTP stack, which will send a request to the security functionalities for signing and verification, together with the GeoNetworking headers and CAM data.

A CAM certificate indicates the permission of the transmitting ITS-S using the parameter Service Specific Permissions (SSP). CAM SSP is defined based on CAM content, in particular the vehicle role and vehicle permission relevant data. More specifically, a CAM certificate indicates if an ITS-S is entitled to transmit a CAM with a specific role setting such as emergency vehicle, or with a specific permission to override some traffic regulations such as

requiring the traffic light preemption in an emergency situation. An incoming signed CAM is accepted by the receiving ITS-S if the CAM content is consistent with the SSP in its certificate.

Table 3. Message frame for a secure CAM

MAC	LLC header	GeoNetworking Bask header	Secure packet (Geonet-working common header, extended header and message)

The SSP indicates if the transmitting ITS-S is compliant to certain rules defined by a deployment stakeholder such as industrial consortium, to a compliance assessment standard, or to the local law or regulations. The SSP assignment procedure will be ensured by Public Key Infrastructure (PKI) in a real deployment.

2.4 DECENTRALIZED ENVIRONMENTAL NOTIFICATION MESSAGE

The generation, transmission and management of DENM is realized by the Decentralized Environmental Notification (DEN) basic service component. The DEN basic service has also been initially introduced by European C2C-CC. In complementary to CAM, the DEN basic service is considered as another core message required for the vehicular communication system deployment since day one in the EU.

2.4.1 The DEN Basic Service Overview

The DEN basic service is an entity that supports the exchange of event-driven DENM in vehicular communication networks. A DENM contains information related to a road traffic event, e.g., traffic jam, break down vehicle, roadworks, etc. The DENM transmission is triggered by an application, upon the detection of an event. DENM may be transmitted with high frequency

(1–10 Hz) to other vehicles or road users within a predefined geographical area.

The DENM is defined as follows:

A DENM contains information related to an event that has potential impact on road safety or traffic condition. An event is characterized by an event type, an event position, a detection time and a time duration. These attributes may change over space and over time. . . The DENM protocol is designed to manage the event detection, event evolution and event termination

Figure 7 provides an example scenario of the DENM transmission to announce a roadwork event. DENMs may be transmitted by a road work vehicle equipped with a vehicle ITS-S or by an RSU upon the request of a road operator. Currently, a variety of event types are specified, ranging from events detected by vehicles with in-vehicle sensors such as electronic brake lights warning, collision risk warning, to events related to the traffic and driving environment such as roadworks, extreme weather conditions, etc.

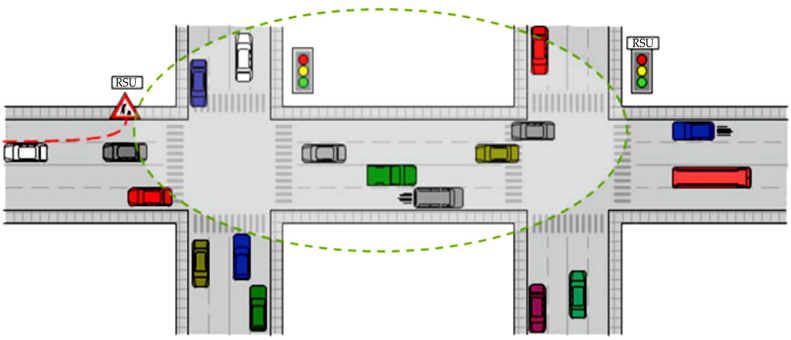


Figure 7. Example of DENM transmission.

Compared to CAM which is an application independent protocol, the DENM transmission is controlled by applications. The standardization of the DEN basic service is therefore scoped by applications selected for initial deployment, while keeping in mind the potential extension needs in the future.

The DEN basic service is mainly used for road safety applications. Its dissemination area may range from several hundreds of meters to several kilometers, depending on the application requirements. A node receiving a DENM may forward it to further distance, in order that vehicles located within the dissemination area may also receive the event information. If the received event information is considered relevant, a warning or information may be provided to driver, who takes appropriate actions to bypass the event position safely.

2.4.2 DENM Dissemination and DENM Protocol

The ITS G5 technology is considered as one of the candidate technologies for DENM dissemination in support of road safety applications. In example illustrated in Figure 7, RSU roadworks warning application requires DENMs to be broadcasted to all vehicles located within an area of relevance (or area of destination) located in the upstream traffic. In complement, other communication technologies such as cellular network may also be used.

DENM Dissemination

The GeoBroadcast protocol of the GeoNetworking functionality may be used for DENM dissemination. It supports multi-hop packet forwarding functionalities, in order to route a DENM packet from the source to the defined geographical destination area. In case ITS G5 is used, DENM is broadcasted over CCH for the 1st hop. Service Channel (SCH) may be used for broadcast from the following hops until reaching the destination area. In case the detected event requires immediate actions to avoid potential collisions, the Traffic Class for DENM may be set to the highest value, in order to guarantee a prioritized access to the radio resource.

Alternatively, the DEN basic service also includes a forwarding mechanism at the facilities layer, in complementary to the network

layer routing protocols. The main objective is to enable forwarding of the most updated DENM among multiple received ones of the same event. This forwarding mechanism may be useful in situation where the event is highly dynamic and requires continuous update of the DENM content (e.g., a moving emergency vehicle event). Another useful situation is when an event covers an area and/or persists during some time, more than one passing vehicle ITS-Ss may detect this event at different positions and times (e.g. an extreme weather condition event).

It should be noted that the DEN basic service standard does not specify requirements on the conditions under which the DENM transmission is triggered or terminated. These conditions are defined per application as application requirements. For example, automobile stakeholders at C2C-CC and road operators in Amsterdam Group define a set of triggering condition documents that specify DENM transmission triggering and termination conditions for different events.

DENM Protocol

The main technical features of the DENM protocol are related to the management of DENM transmission during different phases of the event evolution. For this purpose, multiple types of DENMs are defined:

- new DENM refers to a DENM generated by an ITS-S that detects an event for the first time. A new DENM contains event information, such as event position, event type, and optionally an estimated (or pre-set) validity duration of the event.
- update DENM refers to a DENM generated when an ITS-S detects an update of the event, such as the event position change.
- cancellation DENM refers to a DENM generated by the same ITS-S that has generated the corresponding new DENM, when this ITS-S detects that the event has ended before the originally set validity duration.

- negation DENM is generated by an ITS-S to announce the event termination. This ITS-S did not generate new DENM for this event but has received one from another ITS-S some time ago before arriving to the event position. When this receiving ITS-S arrives the event position and detects that the event has terminated before the expiration of the received event validity duration, it may generate a negation DENM according to a set of triggering conditions. A negation DENM differs from a cancellation DENM, in the sense that the negation DENM provides a possibility for the event termination from an ITS-S other than the one that has originally detected the event for the first time, referred as third-part termination. In situation where the ITS-S that has generated the new DENM has
- lost the capability to transmit a cancellation DENM (e.g., OBU fail in accident) or has moved away from the event position therefore cannot detect the event termination by itself, a negation DENM may be useful. Even though the feature is enabled by the standard, there remains research topics for third-part event termination, in particular how to ensure the liability of the information and how to avoid the misuse of the negation DENM.

The type of DENM to be generated is determined according to the type of application request when the event is newly detected, updated, or terminated. The DENM protocol operation is realized using several parameters:

- actionID is composed of the station ID of the detecting ITS-S and a sequence number. The concept of the actionID is introduced as the event identifier. An actionID enables a receiving ITS-S to distinguish an event detected by different ITS-Ss, or different events detected by the same ITS-S.
- referenceTime is the parameter that enables the distinction of different DENM updates about one event.
- termination allows the receiving ITS-S to derive the DENM type. If present in DENM, it includes two values

i.e., cancellation DENM or negation DENM.

- validityDuration parameter indicates the end of a DENM validity. It may be used to indicate an estimated or preset duration of the event persistence, in case such duration is known in advance. This parameter may not be present in a DENM, in case the detecting ITS-S is not able to provide the event duration information. In this case, a default value is set by ITS-S for internal protocol operation.
- repetitionDuration and repetitionInterval are parameters to control the DENM repetition. In case DENM includes event information that is static, a DENM repetition may be triggered, to transmit DENMs to oncoming vehicles entering the destination area. These parameters are used for protocol operation at transmitting ITS-S, therefore not included in a DENM.
- transmissionInterval is present in DENM when facilities layer forwarding is activated. It indicates the time interval of DENM transmission at the originating ITS-S.

The DEN basic service processes received DENMs using these parameters, then it provides up-to-date event information to applications or redelivers the DENM to the ITS networking and transport layer for forwarding. In one possible forwarding protocol, a receiving ITS-S may forward the most up-to-date DENM of a specific actionID, if it does not receive any repeated or updated DENM from other ITS-Ss within a time period (e.g., three times of transmissionInterval) and is still located inside the DENM destination area when this time period is expired.

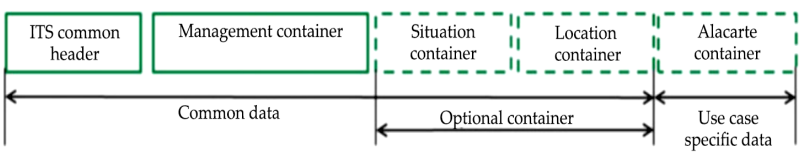


Figure 8. Overview of DENM structure.

2.4.3 Format and Data Requirements

Figure 8 provides an overview of the DENM format Similar to CAM, DENM is structured with data containers. Each container is extensible to support potential extension of the DENM content for future ITS application needs.

- ITS PDU header in a common header for all VANET message types. It includes the protocol version, message ID and station ID
- Management container contains information for the DENM protocol operation. This container is mandatory and shall be present in all transmitted DENM. The receiving ITS-S relies on the content of this container for protocol operation.
- Situation container contains event type information and an indicator of the event detection performance. Each event type is identified with an integer type event code. This list of event codes is extensible
- Location container contains information that describes the location referencing information at the event position. The location referencing information for DENM is a list of traces. Each trace is composed of a list of waypoints that construct a path approaching to the event position. This location referencing information enables receivers to estimate its relevance to the event, by comparing its own itinerary path to each trace contained in the received DENM. In addition, the location container may also include information that represents the detection history of a plain event (e.g., an extreme weather condition event), if the same event was detected by a moving vehicle along its travel path in the past.
- Alacarte container contains optional data specific to an ITS application. For example, in a pre-crash application, an alacarte container is defined to enable exchange of detailed vehicle size, shape, and passenger presence information for collision mitigation purpose.

The ASN.1 unaligned PER encoding rules are used for DENM encoding and decoding, in order to optimize the message size.

2.4.4 DENM Security

Similar to CAM, the security mechanism for DENM considers the authentication of messages transferred between ITS-Ss with certificates. The DENM signing and verification are realized by lower layer, i.e. GeoNeworking/BTP protocol stack. A DENM security profile is standardized.

DENM SSP is defined based on event type code included in the DENM situation container. It indicates if an ITS-S is entitled to transmit a DENM with a specific event code. This permission implies that the ITS-S is compliant to certain event detection and data quality requirements defined by Industrial Consortium such as C2C-CC, to a compliance assessment standard, or to the local laws or regulations.

2.5 BASIC SAFETY MESSAGE

An overview of the full set of standards underway to enable DSRC deployment in the USA can be found. A brief description of the protocols germane to generation message sets are those developed, and in some cases still under development, and described in that thorough overview. However, in keeping with the safety focus of the initial C-ITS deployment in the USA, the primary discussion of this section will detail the definition and use of the BSM. The BSM is a broadcast message with data elements that provide kinematic and other state information sufficient to develop V2V safety-of-life applications.

Despite this dominant set of applications, it is important to first conceptually understand the standardization work in defining the middle layer to enable the host of eventual C-ITS applications, as the initial implementation of V2V safety services portends those other mobility and environmental services that will be delivered

with the V2V and V2I market penetration. This middle layer is developed by standards produced by the IEEE P1609 Working Group. Specifically, IEEE 1609. We define the Provider Service Identification (PSID), IEEE 1609.3 for completeness, note also that IEEE 1609.0 defines the WAVE architecture enabling the use of IEEE 802.11p WAVE, and IEEE 1609.2.

This short operationally oriented description of the so-called middle layer begins by considering that IEEE 1609.3 defines the WAVE Short Message (WSM) to contain three extensions: channel number, data rate, and transmit power. These parameters enable higher layer to indicate the communication requirements for a message dissemination. It also contains a PSID, WSM length, followed by a higher layer message payload, e.g. BSM. The Wave Short Message Protocol (WSMP) is a networking protocol that delivers a WSM, when requested by higher layer, to required destination nodes. In case for BSM dissemination, the WSMP transmits the corresponding WSM to direct network neighbors (one hop broadcast).

At receiving side, the parameter EtherType is used to distinguish WSMP stack from IP stack, and PSID is used to deliver the received WSM to corresponding higher layer entities. A PSID value is assigned to each message, e.g. BSM, based on a well-defined registration procedure as specified in IEEE 1609. In addition, the WSMP protocol specified includes a management plane protocol WAVE Service Advertisement (WSA). A WSA is sent in CCH channel to indicate whether a delivered C-ITS service (i.e., denoted as PSID and other service context information) is through IPv6 or the WSMP stack. A WSA contains well-defined fields that indicates repeat rate, transmit power, location and confidence, among other values. These parameters enable an ITS-S receiving WSM to properly configure the system to access to the announced service, if interested. It is again worth noting that the associated standards are comprehensively described. At this writing there is considerable activity within IEEE P1609 to refine the WSMP and the security protocol, IEEE 1609.2, so the community should anticipate some changes in the detail of the standard. The current

work is very focused on maturing the IEEE 1609 set of standards to accommodate a potential V2V rulemaking or mandate.

Given the impending mandate and associated V2V deployment, the set of BSM content and performance standards drive the key C-ITS standardization activities in the USA. The content or data dictionary is given in SAE J2735 and described in Tables 5 and 6; the performance standards work is in progress at SAE J2945. In light of the impending V2V mandate in the United States, both are presently undergoing scrutiny and revision within the SAE DSRC Technical Committee. To set the context, consider that the SAE J2735 defines nearly 150 data elements, organized into data frames. These are fifteen explicit standard message sets in the standard, as illustrated in Table 4.

Table 4. VANET message sets specified in SAE J2735

No.	Standard message name
1	A La Carte (ACM)
2	Basic Safety Message (BSM)
3	Common Safety Request (CSR)
4	Emergency Vehicle Alert (EVA)
5	Intersection Collision Avoidance (ICA)
6	MAP
7	NMEA (GPS) Corrections (NMEA)
8	Probe Data Management (PDM)
9	Probe Vehicle Data (PVD)
10	Road Side Alert (RSA)
11	RTCM Corrections (RTCM)
12	SPAT
13	Signal Request Message (SRM)
14	Signal Status Message (SSM)
15	Traveler Information Message (TIM)

While the ACM (Standard Message 1) is purposely designed to be flexible, there is by design significant flexibility within SAE J2735. The re-use of modular data elements and data frames is anticipated in the standard. Message sets can be user defined to

address an assortment of C-ITS applications from a wide variety of data elements. The rather extensive SAE J2735 (2009) is therefore designed to be comprehensive. It is analogous to the combination of CAM, DENM, and IVI messages used in Europe. Since the original creation of SAE J2735, US DOT commissioned a contracting team to undertake a requirements-driven system-engineered process wherein a set of primarily public sector stakeholders was engaged to derive user needs, and where existing and prospective concepts of operations were extracted from the ongoing Dynamic Mobility Applications Program.18 This work was documented in what has become the SAE J3067 Information Report. From the list of 15 SAE J2735 standard messages, the SPAT (Standard Message 12) and MAP (Standard Message 6) are close matches to the EU SPAT and MAP messages. In order to achieve this harmonization, the SAE versions of the SPAT and MAP are under revisions, with a final SAE J2735 step in final ballot stages at this writing. This activity will enable SPAT and MAP from Europe (and Japan) to be additionally described within SAE J2735 with regional extensions. This harmonization is fostered under ISO/TC204 Working Group 18 (Cooperative ITS) development of Technical Standard (TS) 19091, which addresses the dynamic messages to enable harmonized mobility and safety applications, particularly for public sector transit and traffic operations. With the SPAT and MAP revisions within TS 19091 essentially complete, the SRM (Standard Message 13) and SSM (Standard Message 14) are the next targets for potential ISO harmonization. Similar to the EU messages, all the messages specified in J2735 except the BSM use ASN.1. Outside the separate DSRC Message ID, the BSM has only one encoding tag and the message, a BSM blob is fixed in length and order. This saves message size and for good reason: the most pertinent message for V2V or initial deployment in North America is the BSM, as the BSM Part 1 is a mandatory representation of the vehicle state, consisting of its kinematic state and other pertinent information. It is defined by the data elements enumerated in Table 5. In North America Safety Pilot Model Deployment tests, BSM transmission rate is fixed to 10 Hz. This transmission rate may be reduced in case channel load is too high.

The BSM Part II is optional, need and broadcast frequency in V2V safety will likely be lower. BSM Part II data elements and broadcast frequencies are listed in Table 6. If no frequency is specified for a certain data element, it will be transmitted when status changes are detected.

The BSM Part II data elements necessary to implement V2V safety are at this writing not normative. Event flags (consisting of application designer-selected values from Element Numbers 1 to 13 above), path history (Element Number 14), path prediction (Element 14), and RTCM correction (Element 16) have been used in an optional vehicle safety extension frame field tested within the Safety Pilot Model Deployment. Changes to the above may be considered in future revisions of the applicable SAE standards in support of the potential NHTSA V2V rulemaking. As an example, the safety extension frame might be transformed to a mandatory element. Moreover, the earlier-referenced SAE J2945.1 DSRC Vehicle BSM Communication Minimum Performance Requirements may rigorously specify the BSM Part I and Part II data elements based on a flowdown of the overall V2V safety performance and the data from the Safety Pilot Model Deployment. Sensor accuracy will be considered, along with the necessary BSM sending rate and transmit power, the latter two contingent on channel congestion issue currently under study by NHTSA and their CAMP partners.

Table 5. BSM Part I content

No.	BSM Part I data element
1	DSRC message ID
2	Message count
3	Temporary ID
4	Current time, 1 ms resolution
5	Lat/Long, resolution 0.1 μ deg
6	Elevation from sea level, 0.1 m
7	Position accuracy, 1 standard deviation per major and minor axes

8	Transmission (gear) and Vehicle speed, 1 cm s^{-1}
9	Heading, 1/80 deg
10	Steering wheel angle, 1.5 deg
11	Acceleration (three axes and yaw rate)
12	Braking state (control, boost, auxiliary) for each of four wheels
13	Vehicle size (length and width), 1 cm
14	Path history (sequence of position vectors for recent past)
15	Path prediction (radius of curvature)
16	Differential GPS corrections
17	Lights status (headlights, running lights, hazard lights, turn signals)
18	Light bar status (for emergency responders, school buses, special vehicles)
19	Front wiper status (on, off, intermittent)
20	Front wiper rate (sweeps per minute)
21	Rear wiper status (on, off, intermittent)
22	Rear wiper rate (sweeps per minute)
23	Braking status (brake applied, ABS, stability, traction control auxiliary, boost systems active)

Table 6. BSM Part II content.

No.	BSM Part II data element	Frequency
1	Hazard lights active	N/A
2	Vehicle expected to violate stop bar	N/A
3	Antilock brake system active over 100 ms	Sec.
4	Traction control system active over 100 ms	Sec.
5	Stability control system active over 100 ms	Sec.
6	Vehicle placard as HazMat carrier	N/A
7	Public safety vehicle responding to an emergency	N/A
8	Recent or current hard braking (> 0.4g)	0.1 s
9	Light status changed	N/A
10	Wiper status changed	All
11	Flat tire	N/A
12	Vehicle is disabled	N/A

13	Airbag has deployed	N/A
14	Path history (sequence of position vectors for recent past)	N/A
15	Path prediction (radius of curvature)	0.1 s
16	Differential GPS corrections (RTCM)	N/A
17	Lights status (headlights, running lights, hazard lights, turn signals)	Sec.
18	Light bar status (for emergency responders, school buses, special vehicles)	N/A
19	Front wiper status (on, off, intermittent)	Min.
20	Front wiper rate (sweeps per minute)	Min.
21	Rear wiper status (on, off, intermittent)	Min.
22	Rear wiper rate (sweeps per minute)	Min.
23	Braking status (brake applied, ABS, stability, traction control auxiliary, boost systems active)	Sec.
24	Level of brake application	Sec.
25	Road coefficient of friction	Sec.
26	Sunlight level	Min.
27	Rain type	Min.
28	Ambient air temperature	Min.
29	Ambient air barometric pressure	Min.
30	Confidence- steering wheel angle	N/A
31	Confidence- steering wheel rate of change	N/A
32	Front wheel angle	0.1 s
33	Vertical acceleration over threshold	N/A
34	Confidence- yaw rate	N/A
35	Confidence- acceleration	N/A
36	Confidence- set of values	N/A
37	Distance to obstacle on the road	N/A
38	Azimuth to obstacle on the road	N/A
39	Date/Time of obstacle detection	N/A
40	Confidence- time	N/A

2.6 IN-VEHICLE INFORMATION

IVI is a message that enables the transmission of road side sign information to road users for in-vehicle presentation. It provides

static sign or dynamic sign (e.g., Variable Message Sign (VMS)) data from infrastructure to vehicles or to mobile devices. The receiving ITS-S processes the received IVI data and estimates the relevance of the information to the driver. When appropriate, the information is delivered to driver as warning or as information. If required by the local regulation, presentation format of a signage, e.g. layout, font size, color, etc. may also be included in IVI for the in-vehicle presentation.

IVI is one of the infrastructure messages that is selected by stakeholders in the EU for day one deployment. The standardization of the IVI is undertaken by ISO TC 204, taking into account the feedbacks from Industrial Consortium such as C2CCC, Amsterdam Group and FOT projects. Multiple technologies can be appropriate candidates for IVI dissemination to satisfy requirements like covered area size, availability of the communication infrastructure, communication cost, etc.

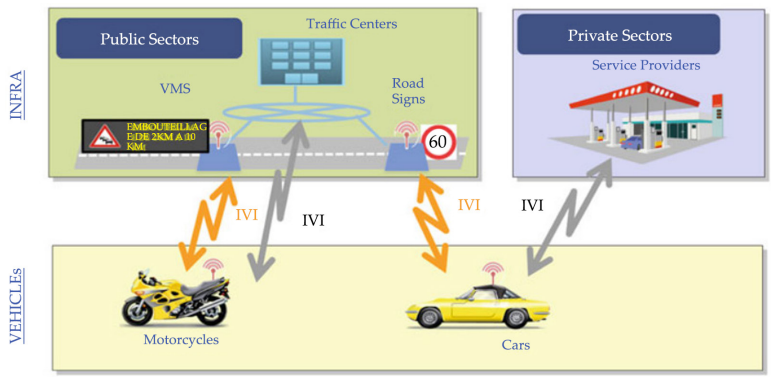


Figure 9. Example scenario of IVI application.

2.6.1 IVI Application

The IVI application consists of providing authenticated road sign information from road side to vehicles and to road users. However, the information chain may go beyond the transmission of messages itself. A series of back-end actions in road ITS infrastructure are required before hand, from the data collection, data processing,

to data generation and authorization/authentication procedures. The IVI message communicates the results of previous back-end actions in a specific message format with corresponding syntax and semantics definitions, enabling the receiving ITS-S to present the information in a proper manner and timing. The IVI application is one representative C-ITS application to further facilitate the vehicle and infrastructure integration.

Figure 9 provides an example scenario of the IVI application. The road signage information, e.g. VMS, speed limit sign, fuel station sign, etc., may be transmitted by RSU, by TMCs, or by private service providers. Standardization of IVI application is ongoing in ISO TC 204 ISO TC 204 Technical Specification 17425.

2.6.2 IVI Message Overview

The structure of an IVI message is derived from DENM. This is motivated by the similarity between the two, in terms of data exchange needs and content requirements. For example, a dynamic road side sign information is also characterized by a sign type, duration, position, and a relevance area, comparable to an event as indicated by a DENM. Generally speaking, an IVI contains information authorized and authenticated by public authorities or by road operators. Compared to a distributed ITS applications where the relevance check is completely at the shoulder of the receiving side, road operators or public authorities may have the control on the sign systems on the road and set requirements on the relevant road sections that a sign information should be informed to road users. This requirement is reflected in the IVI message structure, where the location container is extended with enriched content. Nevertheless, the receiving ITS-S remains the decision maker on the final presentation of a sign to driver or to road user, based on the overall in vehicle information processing load and the HMI system design.

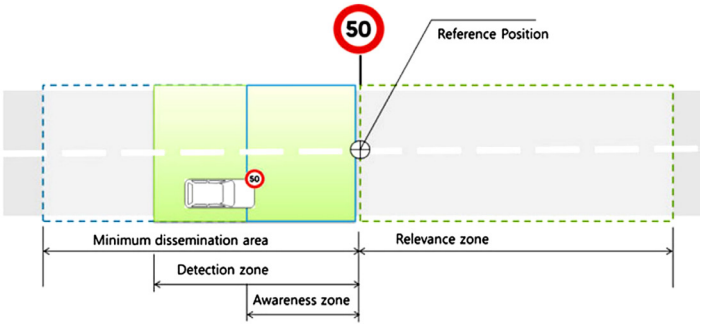


Figure 10. Overview of IVI geographical validity.

Typically, an IVI is transmitted from an RSU which is in connection with a data center (e.g., a TMC) providing the IVI content. The IVI message is under standardization in ISO TC 204 Technical Specification 19321.

Figure 10 provides an overview of the area concepts being relevant to the IVI application, described as follows:

- Minimum Dissemination Area defines the area of the IVI message dissemination. It is provided to lower layer stacks, for example, the BTP/GeoNetworking stack for routing.
- Detection Zone gives indication of an area in which the received IVI message should be processed by an ITS-S. Vehicles entering this zone would probably passing by the road sign in the short future and the information of IVI may be relevant for in-vehicle presentation.
- Driver Awareness Zone is typically determined by receiving ITS-S, taking into account a minimum time at which an IVI data is shown to driver before entering the IVI relevance zone, based on its motion status, e.g. driving speed, driving itinerary. Alternatively, it may also be included in an IVI message and transmitted from road side. In this case, it can be seen as an indication from IVI information providers (e.g., a TMC operator) when the IVI should be informed to drivers. However, the final decision on whether and when to show in vehicle

resides on vehicle side. For example, in case when vehicle encounters dangerous situations near the sign, the OBU application may ignore or delay the IVI presentation in order not to avoid further increasing driver's work load. Typically, a Driver Awareness Zone is a sub-set of the detection zone.

- Relevance Zone defines the zone of relevance for a sign, for example as shown in Figure 10 for Speed Limit Sign, the relevance zone is located in the downstream traffic from the sign position during a certain distance, e.g. until the next speed limit sign.
- Reference Position is the starting point for the definition of all different zones. By definition, the reference position point is covered by all zones of the IVI. In the example shown in Figure 10, the reference position is in the middle of the road surface of speed limit sign position. However, the reference position does not necessarily always correspond to the position where a sign is physically installed. Actually, road signs are often installed a certain distance prior to the position where the sign information becomes effective. For these signs, the reference position should therefore refer to the sign effective position.

Each IVI message includes at least one reference position and one area. One IVI message may be linked to other IVI messages, in case relevance (or overriding) of multiple signs occurs. For example, for a heavy vehicle, a speed limit sign for heavy vehicles should override the speed limit sign in case of bad weather, the lowest speed limit should apply for in-vehicle presentation.

Similar to CAM and DENM, IVI standard does not specify the receiving side protocol of the message processing. This is left at the discretion of implementors.

2.6.3 IVI Format

The container-based structure of a IVI message is illustrated in Figure 11. Each IVI contains at least one management container and optionally one or more location containers or one or more application containers.

The management container contains management information for an IVI message. The IVI protocol also supports similar life cycle management functions like the DENM protocol, including IVI trigger, IVI update, IVI cancellation, and IVI negation. This life cycle management is adopted to mainly support the dynamic sign information management. In addition, the management container may include an identifier of the provider authorities, or identifier of other relevant IVI messages.

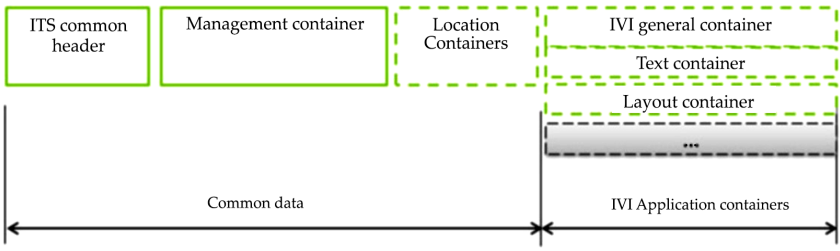


Figure 11. Overview of IVI structure.

The location containers contain one or more reference position and description information for one or more zones. A zone may be described using the combination of geographical points and distance attributes starting from the reference position, forming a circular (combination of reference position and a radius distance) or a polygon shape. Alternatively, the location referencing may also include a road segment ID of a map database or other relevant road topology information.

The application containers contain IVI application data. Currently three types of application containers are defined, namely the IVI general container, IVI text container, and IVI layout container:

- IVI general container provides information on the content and type of a sign, denoted as road sign code.

Each sign is linked with one or more zones described in the location containers, either as detection zone, driver awareness zone or relevance zone. The road sign codes are inherited from a road sign catalog such as defined in Vienna Convention for EU road signs. Other catalogs may apply such as ISO14823 codes, for usage in other regions or according to local regulations. In Vienna Convention catalog, a road side code includes typically a class, an integer code, a value with unit, some text and additional information. For example, the speed limit sign is assigned as code C.2, in addition to a value that indicates the speed limit value with a unit, e.g. 70 km/h. A primary sign may be associated with secondary sign, which limits the application conditions of the primary sign. Optionally, an IVI general container may also be linked to a specific layout container, in case a certain layout is required to be respected for in-vehicle presentation.

- Text container provides possibility to include free text information to road users as it may be the case for VMS text in some European Regions. Like IVI general container, it should be linked to one or more zones, be the relevance zone, driver awareness zone or a detection zone.
- Layout container defines a set of layout information for road sign. A layout container may be linked by an IVI general container or a text container. At receiving the IVI, the in-vehicle system may present the information to driver as defined in layout container. This is to enable an identical road sign layout in vehicle as at road side.

2.6.4 IVI Security

The current draft IVI standard defines the IVI SSP based on the type of applications that it may support, including road signage information, contextual speed information, roadwork warning, restriction information, rerouting, and traveler information, etc.

The SSP indicates if the transmitting ITS-S is entitled to provide the corresponding IVI content in support of the ITS application in question.

2.7 SIGNAL PHASE AND TIMING MESSAGE

SPAT is a message that provides the traffic light phase and timing information from road side ITS-S to vehicles or to mobile devices. One SPAT message may include traffic light status information of one or multiple intersections. The receiving ITSS should process the received SPAT data together with the intersection topology data, in order to match each individual traffic light status data to the corresponding road segment in the intersection to which the traffic light status information is relevant. The intersection topology information may be made available by an RSU, by transmitting a road topology MAP message.

The SPAT and MAP message are standardized by SAE DSRC Technical Committee in the standard SAE J2735, currently under revision to take into account regional requirements of USA, Europe, and Japan.

SPAT and MAP may be used by different ITS applications at intersection area. By knowing the light status and status switch timing before approaching to an intersection, OBU may provide speed advice and warning to driver to avoid traffic light violation or to smooth the intersection crossing.

2.7.1 SPAT Overview

Typically, SPAT is broadcasted from an RSU to vehicles located near the intersection area in question. In order to obtain the real-time traffic light phase and timing information, the RSU should interface with traffic light controller systems of the traffic light system. Depending on the deployment of the traffic light control system, a traffic light controller may be equipped locally in the intersection or in a specific traffic light control network. In urban

environment, one traffic light controller may coordinately control a series of traffic lights within a road segment or within an area. Therefore, the availability and accuracy of the content in a SPAT depends directly on the data made available by the traffic light controller.

Figure 12 provides an example scenario of the Green Light Optimal Speed Advisory (GLOSA) application realized by SPAT and MAP messages. This application is developed and validated in European FOT project DRIVE C2X. This application provides speed advice to pass the next traffic lights during a green phase. In case it is not possible to provide a speed advice, the remaining time to green is displayed. This helps to reduce stop time and unnecessary acceleration/brake in urban traffic situations to reduce fuel consumption and pollution emissions.

The SPAT message sends the current movement state of each active phase of the traffic light, including values of what lights are active and values of phase duration if available. Movements are mapped to specific lanes and approaches by use of the lane numbers included in the message. These lane numbers correspond to the specific lanes described in the MAP message for that intersection.

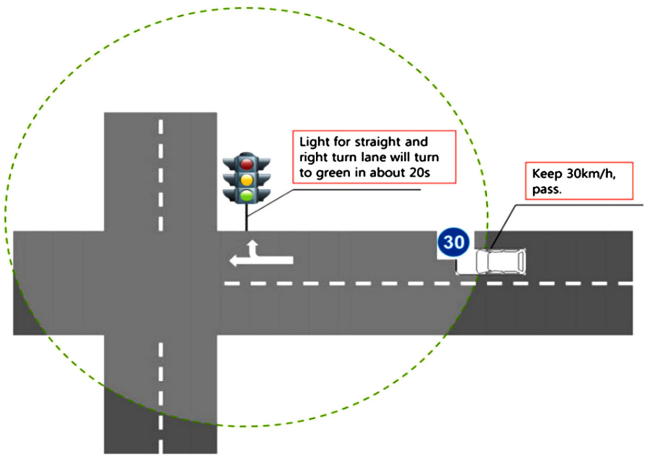


Figure 12. Example scenario for GLOSA application.

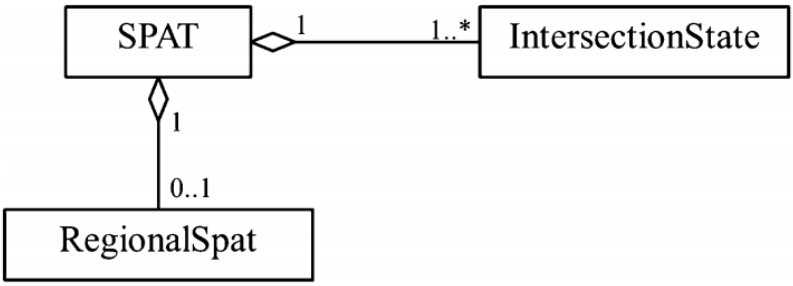


Figure 13. SPAT message format.

2.7.2 SPAT Format

The SPAT message format can be illustrated in UML class diagram in Figure 13. A SPAT message contains traffic light status of one or more intersections, each denoted as IntersectionState data frame. Optionally, SPAT message may contain regional SPAT content extensions, to provide regional additional content to describe the traffic light status, e.g. traffic light priority or preemption information.

Each of the InstersectionState data frame may be composed of the following categories of information to describe the traffic light status of one intersection:

- Message management data includes general information of the message, including a message count and a message generation time.
- Intersection data includes general information of the intersection, including an intersection ID, an enabled lane list, and general status information. Optionally, the enabled lane list may be used to indicate the list of lanes that are active (or open to traffic) within the intersection. This list may be changed over time, e.g. a right turn lane may be enabled even when the straight traffic light is in red, or a lane is open to traffic only during some period of the day. The intersection status data indicates the traffic light controller state of the intersection in question.

Depending on the local configuration of the traffic light control system, a traffic light controller may be operating in active, stand by, off or failure mode, the operation time interval maybe fixed or dynamic, the priority mode may be activated or deactivated. In addition, the intersection status may also indicate the status of the SPAT and MAP transmitting system itself e.g. if an MAP update is expected to process the SPAT, if the active lanes are updated etc.

- Traffic light status data provides in turn all traffic light status of the intersection, denoted as movement state, each applying to a set of lanes inside the intersection. An ID is assigned to each movement state, which will be used to match to lane descriptions of the corresponding MAP message. This ID should be made unique, at least within the intersection. In one movement event, the movement state may provide the phase type, preset or estimated phase change time, and optionally the advisory speed information.
- Maneuvering assistance data provides additional information to assist receiving vehicles to exit the intersection. For example, the SPAT message may include estimated queue length or presence of pedestrian information, if such information is made available by, e.g., equipped sensors.
- Other data for regional extensions.

REFERENCES

1. Amsterdam Group (2013) Roadmap between automotive industry and infrastructure organisations on initial deployment of cooperative ITS in Europe.
2. ETSI (2009) Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions.
3. ETSI (2012) ETSI TS 101 556 -1 (V1.1.1) - Intelligent Transport Systems (ITS); Infrastructure to Vehicle Communication; Electric Vehicle Charging Spot Notification Specification
4. ETSI (2013) ETSI EN 302 636-4-1 (V1.2.0): "Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 1: Media-Independent Functionality"
5. ETSI (2013) ETSI EN 302 637-2 (V1.3.0) - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service
6. ETSI (2013) ETSI EN 302 637-3 V1.2.0 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specification of Decentralized Environmental Notification Basic Service
7. European Commission (2009) Standardisation mandate addressed to CEN, CENELEC and ETSI in the field of information and communication technologies to support the interoperability of co-operative systems for intelligent transport in the European Community, p 5
8. Kenney JB (2011) Dedicated short-range communications (DSRC) standards in the United States. Proc IEEE 99(7):1162–1182
9. Malone K, Rech J (2013) User related results from DRIVE C2X test sites. http://www.drivec2x.eu/tl_files/publications/3rd%20Test%20Site%20Event%20TSS/6%20DRIVE%20C2X %203rd%20Test%20site%20event_Kerry%20Malone_Users_20130715.pdf

10. SAE (2009) SAE J2735: Dedicated short range communications (DSRC) message set dictionary
11. SAE DSRC Technical Committee (2011) SAE Draft Std. J2945.1, Revision 2.2.: Draft DSRC message communication minimum performance requirements: basic safety message for vehicle safety applications.
12. SAE DSRC Technical Committee (2014) SAE J3067: Information report on candidate improvements to dedicated short range communications (DSRC) message set dictionary [SAE J2735] using systems engineering methods
13. Song X (2014) Cooperative vehicle-infrastructure system - activities in China. In: 6th ETSI TC ITS workshop, Berlin, February2014.http://docbox.etsi.org/Workshop/2014/201402_ITSWORKSHOP/S02_ITS_SomeBitsFromtheWorld/CHINA_NATIONALCENTREofITS_ SONG.pdf
14. Stevens S (2013) HS63A3 Project memorandum: safety pilot—preliminary analysis of the driver subjective data for integrated light vehicles
15. Union E (2013) Commission delegated regulation (EU) No 305/2013 of 26 November 2012 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the harmonised provision for an interoperable EU-wide eCall

CHAPTER 3

NETWORKING ISSUES

INTRODUCTION

Networks are networks. Despite best efforts to keep things smooth all the time every day, things happen. Here's a look at some common network issues, some tips for quickly resolving them, and even better, how to prevent them from occurring again.

1. **Duplicate IP Addresses:** When two devices attempt to share a single IP, you see the dreaded "Address Already in Use" Kill — with no ability to access the network.

The Quick Fix: The blame for this often rests with your router's default DHCP configuration. DHCP is probably trying to assign your new device an address at the beginning of your subnet, and another device may already occupy these low-numbered addresses with static IPs. If you've just introduced a new device or server to your network, it may have its own DHCP server. Simply disable the DHCP server on that device to restore sanity to your network.

The Preventive Measure: You can take one simple step to avoid IP conflicts by modifying your router's configuration to begin assigning DHCP addresses near the top end of your subnet, leaving the lower addresses available for devices that require static IPs.

2. **IP Address Exhaustion:** To troubleshoot this issue, use the ipconfig command. If the workstation has assigned itself an IP address that begins with 169.x.x.x, it means that no IP address was available from the DHCP server.

The Quick Fix: Some users on cable internet might not have a local router, in which case IP addresses are assigned on a limited basis directly from your ISP. You have probably run out of allowed IP addresses from your ISP. The solution to this is to purchase either a standalone router or WiFi access point with an integrated router. This creates your own local pool of internal addresses, ensuring you won't run out.

If you already have a local router with DHCP, the default address pool might be too small for your network. By accessing the DHCP settings on the router, you can adjust the size of the address pool to meet your network's needs.

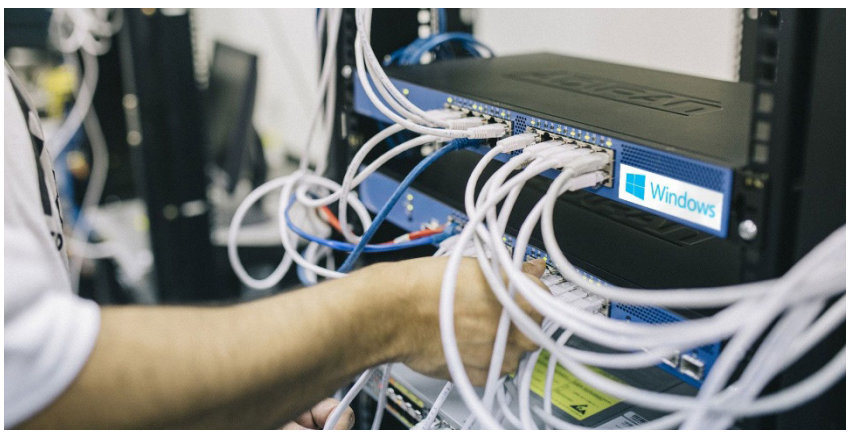
The Preventive Measure: It's important that any internet-connected network have a local router in operation with NAT and DHCP, both for security reasons and to prevent IP address exhaustion.

The router needs to be the only device connected to the modem, with all other devices connecting through the router.

3. **DNS Problems:** Errors such as The Network Path Cannot Be Found, IP Address Could Not Be Found, or DNS Name Does Not Exist, can usually be traced to a DNS configuration issue. The command line utility nslookup can be used to quickly show a workstation's DNS settings.

The Quick Fix: Workstations and other network devices can be configured to use their own DNS servers, ignoring the server assigned by DHCP. Checking the 'Internet Protocol Version 4 (TCP/IP)' settings for your adapter will show if an incorrect DNS server is specified, so just select "Obtain DNS server address automatically" instead.

*The Prevention Measure***:** Your local router might be configured to operate as a DNS Server, creating a DNS pass-through to your ISPs servers. On busy networks, this may overload the capabilities of the router. Change your network's DHCP settings to directly access your DNS servers.



4. **Single Workstation Unable to Connect to the Network:** If only a single workstation is displaying the "No internet" message when opening a web browser, we can usually assume that the rest of the network is healthy and turn our attention to any hardware and software

that is particular to this system.

The Quick Fix: To resolve this network issue, start by eliminating the obvious communication barriers such as a bad cable, poor WiFi signal, failing network card or incorrect drivers. Ensure that the workstation's network adapter is configured with the correct IP, subnet, and DNS servers.

If that doesn't solve the problem, check any firewall software on the device to ensure that necessary ports are open to the external network. Common ports include 80 and 443 for web traffic, plus 25, 587, 465, 110, and 995 for email.

The Preventive Measure: It's usually best to leave all workstation TCP/IP settings to "Automatically assigned." Use a DHCP server to hand out a uniform configuration to all devices on the network. If a static IP is needed on a particular workstation or server, most DHCP servers allow the ability to create static IP mappings.

5. **Unable to Connect to Local File or Printer Shares:** Sharing problems are among the most difficult network problems to solve, due to the number of components that need to be configured properly.

Most commonly, sharing problems arise due to conflicts between mixed security environments. Even different versions of the same operating system sometimes use slightly different security models, which can make interconnection of workstations difficult.

The Quick Fix: We can cure sharing problems most efficiently by drilling down through the possibilities in this order:

1. Ensure that the required services are running. On Windows systems, the server, TCP/IP NetBIOS Helper, workstation, and computer browser services all need to be running. On Linux machines, Samba is the primary component required to share with Windows systems.
2. Check your firewall(s). It's very common for a workstation's firewall to be configured to block file and printer sharing traffic, especially if a new antivirus package is installed that introduces its own firewall.

Firewall issues can also exist at the hardware level, so ensure that routers or managed switches are passing share traffic within the subnet. Speaking of subnet….

3. Ensure all workstations are on the same subnet. This problem typically only appears on complex networks, however, even simple networks sometimes have static-IP equipment with an improperly configured subnet. The result is that external traffic will move about just fine, while internal traffic will hit unexpected roadblocks.
4. All Windows network adapters will need File and Printer Sharing for Microsoft Networks, Client for Microsoft Networks, and NetBIOS over TCP/IP enabled.
5. Once the above checks have passed, it's finally time to check the most likely culprit, permissions. There are multiple layers of access required, each with their own interface within the OS. Check for:

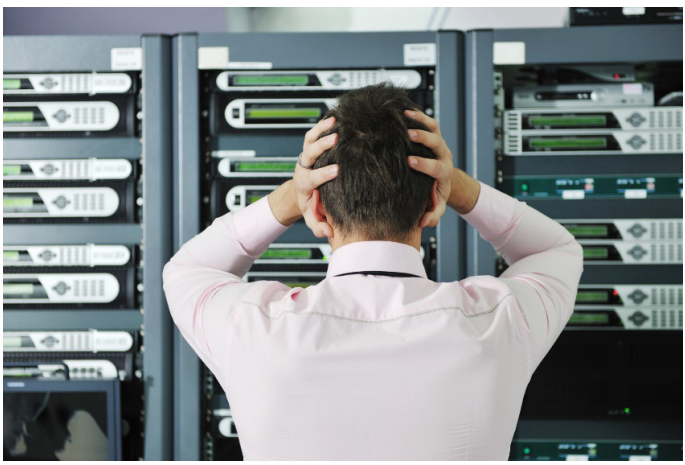
- Systems configured with the wrong workgroup or domain.
- Incorrectly configured HomeGroup.
- Network type set to Public.
- Incorrect NTFS permissions.

6. **Local Network is Unable to Connect to the internet:** This situation can either be intermittent or persistent. Often times, the most difficult aspect of dealing with any external network problem is finding the company responsible. And then tasking them to solve the issue, particularly with intermittent failures that are difficult to trace. It can sometimes be such a problem that organizations will have to switch internet providers in order to solve the issue.

The Quick Fix: A router and modem reboot is the first order of business. The tracert then utility can be used to identify communication breaks. It will clearly hiccup on the particular router hop that is causing the problem. Contact your ISP with your findings, providing screenshots as necessary.

The Preventive Measure: To avoid the finger-pointing that can prevent rapid resolution of external issues, do some research to ensure that you procure connectivity only from local Tier 1 providers. Other ISPs are more than happy to sell you service, however, they are simply piggybacking the Tier 1 connection, since they don't actually own the infrastructure in your area.

The goal is to remove as many middle-men as possible, so that when (not if) you experience a problem, one phone call is all that is required to identify the issue and get technicians to work on it.



7. **Slow Internet Performance:** Slow performance is typically due to congestion, or sometimes poor quality connections that have corroded or otherwise deteriorated. Congestion may not be directly related to bandwidth exhaustion, as a single overloaded port on a switch or router can diminish network performance.

This can be especially true on leased lines where dedicated bandwidth is to be expected, but speed tests indicate the network is not reaching it's rated potential.

The Quick Fix: Use speed test websites, conducting tests from geographically remote servers. This can pinpoint areas of congestion on the ISP's network. In the case of cable internet, the local network is shared amongst your neighbors, committing your ISP to a costly bandwidth upgrade when saturation occurs. Report your findings to your ISP so that they can take steps to resolve the issue.

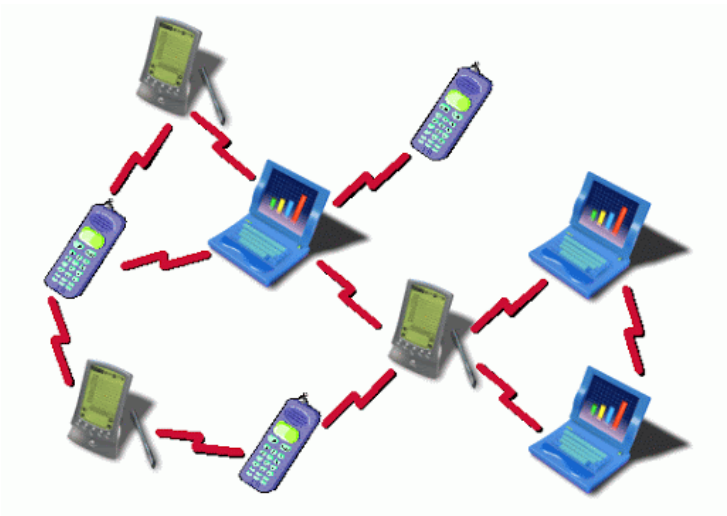
DNS servers are an often overlooked aspect of internet performance. Using incorrect DNS servers can result in routing congestion or load balancing problems. While you should typically use your ISP's DNS settings whenever possible, they may actually be routing traffic through overloaded web caches. You can temporarily adjust your DNS settings to use OpenDNS instead.

The Preventive Measure: if internet performance is critical, you'll need to procure adequate connectivity. While cable internet may be inexpensive, you could be setting yourself up for frequent jeers from employees. A local DSL operator may offer improved reliability for a slightly higher cost, but for the most consistent performance, you may find that an expensive leased line is a requirement for your organization.

3.1 ROUTING IN MANET

In Mobile Ad hoc Network (MANET), nodes do not know the topology of their network, instead they have to discover it by their

own as the topology in the ad-hoc network is dynamic topology. The basic rules is that a new node whenever enters into an ad-hoc network, must announce its arrival and presence and should also listen to similar announcement broadcasts made by other mobile nodes.

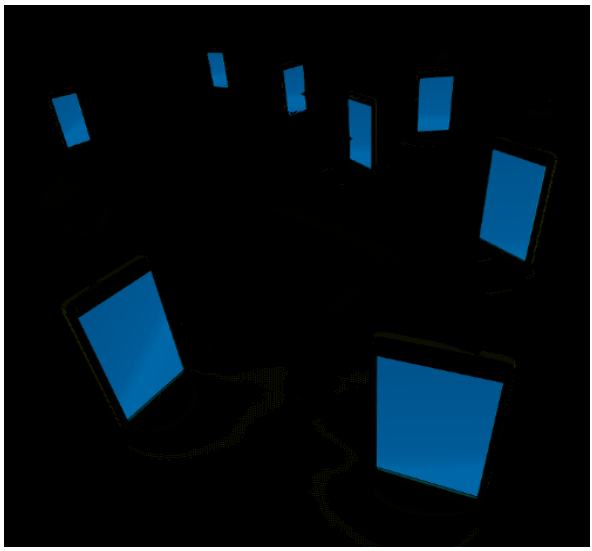


3.1.1 Characteristics of MANETs

A MANET consists of mobile platforms (e.g., a router with multiple hosts and wireless communications devices)--herein simply referred to as "nodes"--which are free to move about arbitrarily. The nodes may be located in or on airplanes, ships, trucks, cars, perhaps even on people or very small devices, and there may be multiple hosts per router. A MANET is an autonomous system of mobile nodes. The system may operate in isolation, or may have gateways to and interface with a fixed network. In the latter operational mode, it is typically envisioned to operate as a "stub" network connecting to a fixed internetwork. Stub networks carry traffic originating at and/or destined for internal nodes, but do not permit exogenous traffic to "transit" through the stub network.

MANET nodes are equipped with wireless transmitters and receivers using antennas which may be omnidirectional

(broadcast), highly- directional (point-to-point), possibly steerable, or some combination thereof. At a given point in time, depending on the nodes' positions and their transmitter and receiver coverage patterns, transmission power levels and co-channel interference levels, a wireless connectivity in the form of a random, multihop graph or "ad hoc" network exists between the nodes. This ad hoc topology may change with time as the nodes move or adjust their transmission and reception parameters.



MANETs have several salient characteristics:

1) Dynamic topologies: Nodes are free to move arbitrarily; thus, the network topology--which is typically multihop--may change randomly and rapidly at unpredictable times, and may consist of both bidirectional and unidirectional links.
2) Bandwidth-constrained, variable capacity links: Wireless links will continue to have significantly lower capacity than their hardwired counterparts. In addition, the realized throughput of wireless communications--after accounting for the effects of multiple access, fading, noise, and interference conditions, etc.--is often much less than a radio's maximum transmission rate.

One effect of the relatively low to moderate link capacities is that congestion is typically the norm rather than the exception, i.e. aggregate application demand will likely approach or exceed network capacity frequently. As the mobile network is often simply an extension of the fixed network infrastructure, mobile ad hoc users will demand similar services. These demands will continue to increase as multimedia computing and collaborative networking applications rise.

3) Energy-constrained operation: Some or all of the nodes in a MANET may rely on batteries or other exhaustible means for their energy. For these nodes, the most important system design criteria for optimization may be energy conservation.

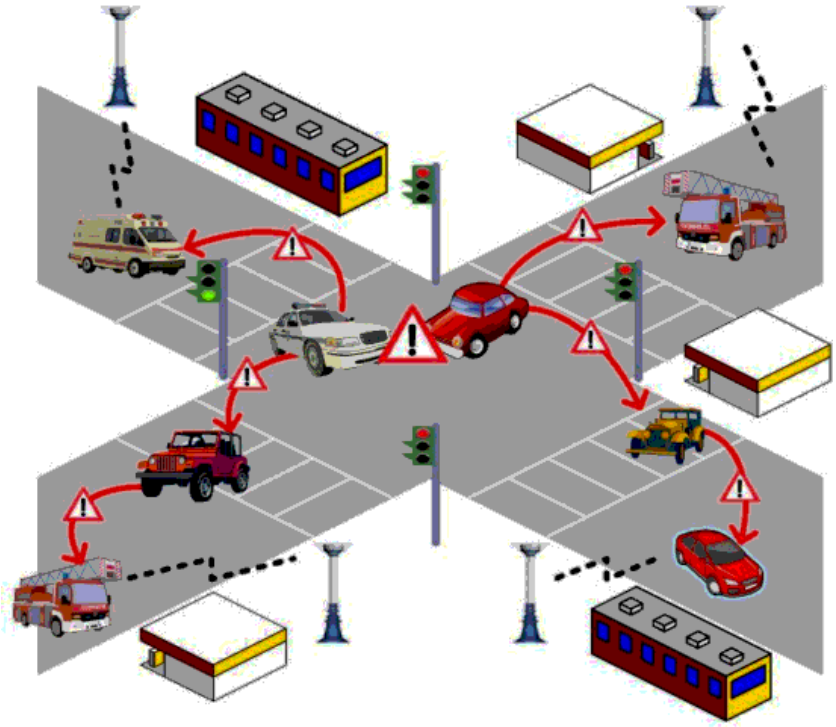
4) Limited physical security: Mobile wireless networks are generally more prone to physical security threats than are fixed- cable nets. The increased possibility of eavesdropping, spoofing, and denial-of-service attacks should be carefully considered. Existing link security techniques are often applied within wireless networks to reduce security threats. As a benefit, the decentralized nature of network control in MANETs provides additional robustness against the single points of failure of more centralized approaches.

In addition, some envisioned networks (e.g. mobile military networks or highway networks) may be relatively large (e.g. tens or hundreds of nodes per routing area). The need for scalability is not unique to MANETS. However, in light of the preceding characteristics, the mechanisms required to achieve scalability likely are.

These characteristics create a set of underlying assumptions and performance concerns for protocol design which extend beyond those guiding the design of routing within the higher-speed, semi-static topology of the fixed Internet.

3.1.2 Goals of IETF Mobile Ad Hoc Network (manet) Working Group

The intent of the newly formed IETF manet working group is to develop a peer-to-peer mobile routing capability in a purely mobile, wireless domain. This capability will exist beyond the fixed network (as supported by traditional IP networking) and beyond the one-hop fringe of the fixed network.



The near-term goal of the manet working group is to standardize one (or more) intra-domain unicast routing protocol(s), and related network-layer support technology which:

- provides for effective operation over a wide range of mobile networking "contexts" (a context is a set of characteristics describing a mobile network and its environment);
- supports traditional, connectionless IP service;

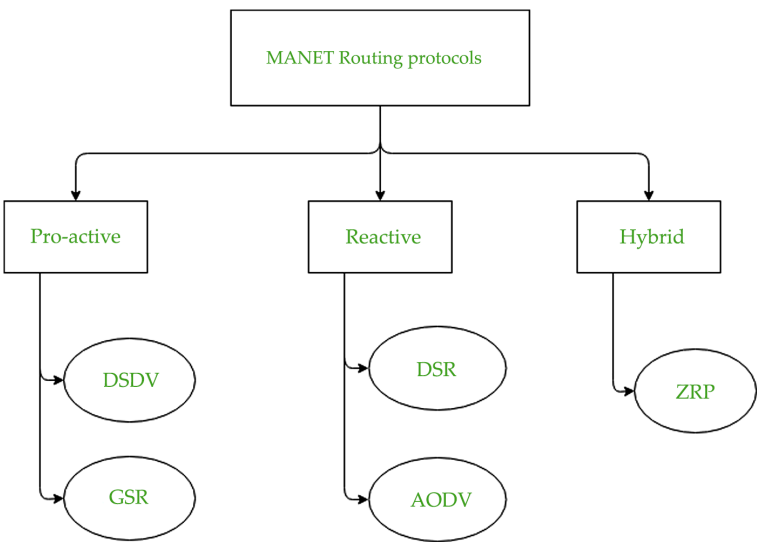
- reacts efficiently to topological changes and traffic demands while maintaining effective routing in a mobile networking context.

The working group will also consider issues pertaining to addressing, security, and interaction/interfacing with lower and upper layer protocols. In the longer term, the group may look at the issues of layering more advanced mobility services on top of the initial unicast routing developed. These longer term issues will likely include investigating multicast and QoS extensions for a dynamic, mobile area.

3.1.3 MANET Routing Protocols

Most routing protocols for mobile ad hoc networks (MANETs) can be categorized as being either reactive or proactive. INET contains various routing protocols for MANETs from both categories, and other categories as well.

This showcase demonstrates the configuration and operation of three MANET routing protocols with three example simulations, using a reactive (AODV), a proactive (DSDV), and a location-based (GPSR) routing protocol.



Pro-active routing protocols

These are also known as table-driven routing protocols. Each mobile node maintains a separate routing table which contains the information of the routes to all the possible destination mobile nodes.

Since the topology in the mobile ad-hoc network is dynamic, these routing tables are updated periodically as and when the network topology changes. It has a limitation that is doesn't work well for the large networks as the entries in the routing table becomes too large since they need to maintain the route information to all possible nodes.

Destination Sequenced Distance Vector Routing Protocol (DSDV)

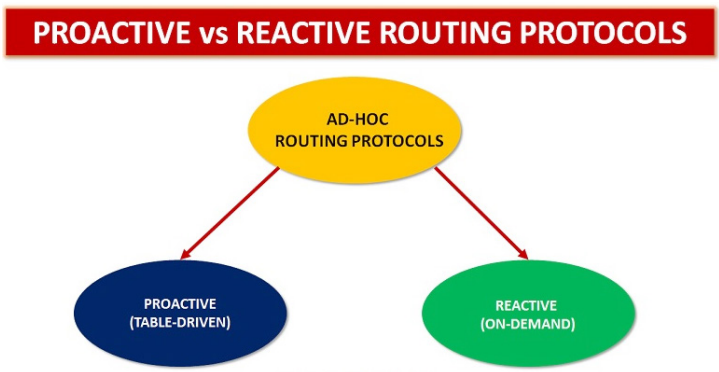
It is a pro-active/table driven routing protocol. It actually extends the distance vector routing protocol of the wired networks as the name suggests. It is based on the Bellman-ford routing algorithm. Distance vector routing protocol was not suited for mobile ad-hoc networks due to count-to-infinity problem. Hence, as a solution Destination Sequenced Distance Vector Routing Protocol (DSDV) came into picture.

Destination sequence number is added with every routing entry in the routing table maintained by each node. A node will include the new update in the table only if the entry consists of the new updated route to the destination with higher sequence number.

Global State Routing (GSR)

It is a pro-active/table driven routing protocol. It actually extends the link state routing of the wired networks. It is based on the Dijkstra's routing algorithm. Link state routing protocol was not suited for mobile ad-hoc networks because in it, each node floods the link state routing information directly into the whole network i.e. Global flooding which may lead to the congestion of control packets in the network.

Hence, as a solution Global State Routing Routing Protocol (GSR) came into the picture. Global state routing doesn't flood the link state routing packets globally into the network. In GSR, each of the mobile node maintains one list and three tables namely, adjacency list, topology table, next hop table and distance table.



Reactive routing protocols

These are also known as on-demand routing protocol. In this type of routing, the route is discovered only when it is required/needed. The process of route discovery occurs by flooding the route request packets throughout the mobile network. It consists of two major phases namely, route discovery and route maintenance.

Dynamic Source Routing protocol (DSR)

It is a reactive/on-demand routing protocol. In this type of routing, the route is discovered only when it is required/needed. The process of route discovery occurs by flooding the route request packets throughout the mobile network.

It consists of two phases:

- Route Discovery

This phase determines the most optimal path for the transmission of data packets between the source and the destination mobile nodes.

- Route Maintenance

This phase performs the maintenance work of the route as the topology in the mobile ad-hoc network is dynamic in nature and hence, there are many cases of link breakage resulting in the network failure between the mobile nodes.

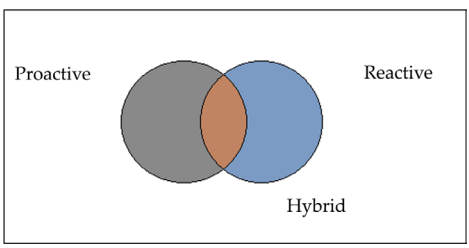
Ad-Hoc On Demand Vector Routing protocol (AODV)

It is a reactive/on-demand routing protocol. It is an extension of dynamic source routing protocol (DSR) and it helps to remove the disadvantage of dynamic source routing protocol. In DSR, after route discovery, when the source mobile node sends the data packet to the destination mobile node, it also contains the complete path in its header. Hence, as the network size increases, the length of the complete path also increases and the data packet's header size also increases which makes the whole network slow.

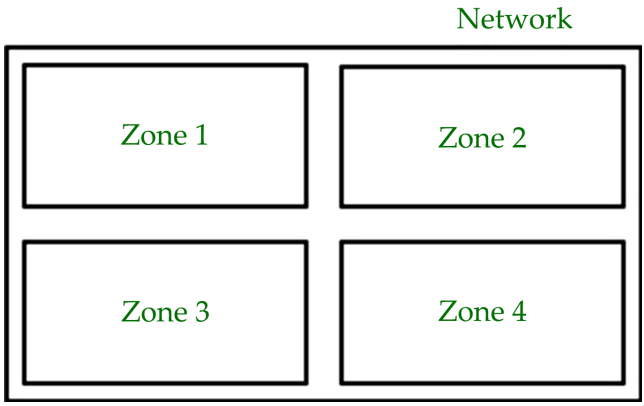
Hence, Ad-Hoc On Demand Vector Routing protocol came as solution to it. The main difference lies in the way of storing the path, AODV stores the path in the routing table whereas DSR stores it in the data packet's header itself. It also operates in two phases in the similar fashion: Route discovery and Route maintenance.

Hybrid Routing protocol

It basically combines the advantages of both, reactive and pro-active routing protocols. These protocols are adaptive in nature and adapts according to the zone and position of the source and destination mobile nodes. One of the most popular hybrid routing protocol is Zone Routing Protocol (ZRP).



The whole network is divided into different zones and then the position of source and destination mobile node is observed. If the source and destination mobile nodes are present in the same zone, then proactive routing is used for the transmission of the data packets between them. And if the source and destination mobile nodes are present in different zones, then reactive routing is used for the transmission of the data packets between them.



3.1.4 Configuration and Results

This section contains the configuration and results for the three simulations, which demonstrate the MANET routing protocols AODV, DSDV and GPSR. The AODV and DSDV simulations use the ManetRoutingProtocolsShowcaseA network, which features moving hosts. The GPSR simulation uses the ManetRoutingProtocolsShowcaseB network, featuring stationary hosts. Both networks contain hosts of the type ManetRouter (an extension of WirelessHost), whose routing module type is configurable. Just as WirelessHost, it uses Ieee80211ScalarRadio by default. It also has IP forwarding enabled, and its management module is set to Ieee80211MgmtAdhoc. In the network, there is a source host named source, a destination host named destination, and a number of other hosts, which are named node1 up to node10 (their numbers vary in the different networks). In addition to mobile nodes, both networks contain an Ieee80211ScalarRadioMedium, an Ipv4NetworkConfigurator, and an IntegratedMultiVisualizer

module. The nodes' default PHY model (IEEE 802.11) will suffice because we're focusing on the routing protocols.

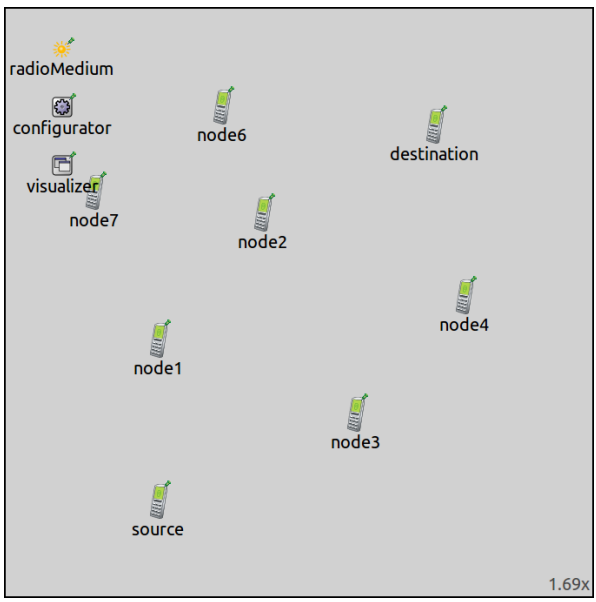
In all three simulations, the source node pings the destination node. The two nodes are out of communication range of each other, and the other nodes are responsible for forwarding packets between the two. Since routes are managed dynamically by the MANET routing algorithms, the Ipv4NetworkConfigurator module is instructed not to add any routes (it will only assign IP addresses). The netmask routes added by network interfaces are disabled as well. The following keys in the General configuration in **omnetpp.ini** achieve this:

*.configurator.addStaticRoutes = false

**.netmaskRoutes = ""

AODV

The example simulation featuring AODV is defined in the Aodv configuration in **omnetpp.ini**. This configuration uses the ManetProtocolShowcaseA network. The network looks like the following:



The nodes are scattered on the scene. The source and destination nodes are stationary, and the other nodes are configured to move in random directions. The communication ranges are set up so that source cannot reach destination directly but through the intermediate nodes. The routing protocols will adapt the routes to the changing network topology.

The mobility settings are defined in the MobileNodesBase configuration in **omnetpp.ini**. The simulations for AODV and DSDV, which feature moving nodes, are based on this configuration. The nodes will be moving on linear paths in random directions with a speed of 25 meters per second, bouncing back from the edge of the scene. The mobility settings are the following:

.node.mobility.typename = "LinearMobility"

.node.mobility.initialMovementHeading = uniform(0deg,360deg)

.node.mobility.speed = 25mps

**.constraintAreaMaxX = 400m

**.constraintAreaMaxY = 400m

**.constraintAreaMinX = 0m

**.constraintAreaMinY = 0m

The ping app in source will send one ping request every second to destination.

In INET, AODV is implemented by the Aodv module. This module is configured in **omnetpp.ini** as the routing protocol type in ManetRouter:

..routingApp.typename = "Aodv"

The Aodv module has many parameters for controlling the operation of the protocol. The parameters should be set according to the number of nodes in a network, the nodes' mobility levels, traffic, and radio transmission power levels/communication ranges. All of the parameters have default values, and Aodv

should work out of the box, without setting any of the parameters. We will fine-tune the protocol's behavior to our scenario by setting two of the parameters:

..routingApp.activeRouteTimeout = 1s

..routingApp.deletePeriod = 0.5s

The activeRouteTimeout parameter sets the timeout for the active routes. If the routes are not used for this period, they become inactive. The deletePeriod parameter sets the period after which the inactive routes are deleted. The activeRouteTimeout parameter is lowered from the default 3s to 1s, and the deletePeriod parameter is lowered from the default 15s to 0.5s to make the protocol react faster to the rapidly changing network topology. Higher mobility results in routes becoming invalid faster. Thus the routing protocol can work better - react to topology changes faster - with lower timeout values. However, setting the timeout values too low results in increased routing protocol overhead.

Successful data link layer transmissions are visualized by colored arrows. Note that only the routing protocol and ping packets are visualized, not the ACKs. Here is what happens in the video:

At the beginning of the simulation, source queues a ping request packet for transmission. There are no routes for destination, so it broadcasts an AodvRreq message. The RREQ is re-broadcast by the adjacent nodes until it gets to destination. The destination node sends a unicast AodvRrep. It is forwarded on the reverse path the RREQ message arrived on (destination->``node6``->``node1``->``source``). As the intermediate nodes receive the RREP message, the routes to destination are created. The routes are visualized with black arrows, and the RoutingTableVisualizer is configured to visualize only the routes leading to destination. When the route is established in source, it sends the ping request packet, which gets to the destination. The ping reply packet gets back to source on the reverse path.

When source sends the next ping request packet, host6 has already moved out of range of destination. The ping packet gets

to host6, but can't get to destination (host6 tries to transmit the packet a few times, but it doesn't get an ACK). So host6 broadcasts an AodvRerr message, indicating that the link no longer works. When the RERR gets back to host1, it initiates route discovery by broadcasting an RREQ message. When a new route is discovered (source->``node1``->``destination``), the ping traffic can continue.

The following log excerpt shows node6 handling the first RREQ and RREP messages:

```
ManetprotocolsShowcaseB.node6.routing (Aodv, id=470)  on aodv::Rreq (inet::Packet, id=568)
B.node6.routing: AODV Route Request arrived with source addr: 10.0.0.2 originator addr: 10.0.0.1 destination addr: 10.0.0.8
seB.node6.routing: Updating existing route: destination = 10.0.0.2, prefixLength = 32, nextHop = 10.0.0.2, metric = 1, interface = wl
seB.node6.routing: Route updated: destination = 10.0.0.2, prefixLength = 32, nextHop = 10.0.0.2, metric = 1, interface = wlan0
seB.node6.routing: Updating existing route: destination = 10.0.0.1, prefixLength = 32, nextHop = 10.0.0.2, metric = 2, interface = wl
seB.node6.routing: Route updated: destination = 10.0.0.1, prefixLength = 32, nextHop = 10.0.0.2, metric = 2, interface = wlan0
B.node6.routing: Forwarding the Route Request message with TTL= 2
ManetprotocolsShowcaseB.node6.routing (Aodv, id=470)  on aodv::Rrep (inet::Packet, id=892)
B.node6.routing: AODV Route Reply arrived with source addr: 10.0.0.8 originator addr: 10.0.0.1 destination addr: 10.0.0.8
seB.node6.routing: Adding new route destination = 10.0.0.8, prefixLength = 32, nextHop = 10.0.0.8, metric = 1, interface = wlan0
colsShowcaseB.node6.ipv4.routingTable: ManetprotocolsShowcaseB.node6.ipv4.routingTable: add route dest:10.0.0.8  gw:10.0.0.8  mask:25
seB.node6.routing: Updating existing route: destination = 10.0.0.8, prefixLength = 32, nextHop = 10.0.0.8, metric = 1, interface = wl
seB.node6.routing: Route updated: destination = 10.0.0.8, prefixLength = 32, nextHop = 10.0.0.8, metric = 1, interface = wlan0
B.node6.routing: Forwarding the Route Reply to the node 10.0.0.1 which originated the Route Request
```

DSDV

The example simulation featuring DSDV is defined in the Dsdv configuration in **omnetpp.ini**. Just like the AODV configuration, this one uses the ManetRoutingProtocolsShowcaseB network. The mobility settings are also the same as in the AODV simulation. The ping app in source will send a ping request every second.

The DSDV protocol is implemented in the Dsdv module. The routing protocol type in all hosts is set to Dsdv:

..routing.typename = "Dsdv"

Currently, complete routing table broadcasts are not implemented, only the broadcasting of changes in the routing table using periodic hello messages.

Like Aodv (and most routing protocol modules), Dsdv has many parameters with default values that yield a working simulation without any configuration. In this simulation, similarly to the previous one, we set two parameters of the protocol:

..routing.helloInterval = 1s

..routing.routeLifetime = 2s

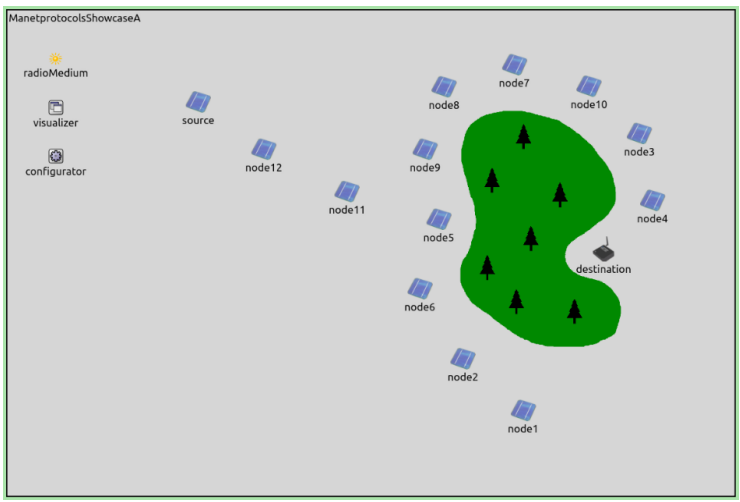
The helloInterval parameter controls the frequency of the periodic updates, or hello messages. Setting this parameter to a higher value decreases the protocol overhead, but the network will react more slowly to changes in topology. We lower it from the default 5s to 1s to make the network more adaptive to rapid changes. When a route is not used or updated after a time, it gets deleted. The routeLifetime parameter sets after how long the routes are deleted after not being used or updated anymore. We lower this from the default 5s to 2s.

The following video shows the nodes sending hello messages and routes being created at the beginning of the simulation. Note that the black arrows represent routes, and routes from all nodes to all destinations are visualized here.

The following video shows source pinging destination:

GPSR

The example simulation featuring GPSR is defined in the Gpsr configuration in **omnetpp.ini**. It uses the ManetRoutingProtocolsShowcaseB network. The network looks like the following:



Just as with the previous two configurations, the nodes are ManetRouters. The nodes are laid out along a chain. The transmitter power of the radios is configured so that nodes can only reach their neighbors in the chain (except for node9, which can reach nodes 11, 5, and 8). There is a forest, which represents a void that GPSR can route around. In this example simulation, the nodes will be static (though GPSR is suitable for scenarios with moving nodes). The source node will ping the destination node, which is on the other side of the void. (The ping app in source will send one ping request every second.)

The hosts' routing protocol type is set to Gpsr:

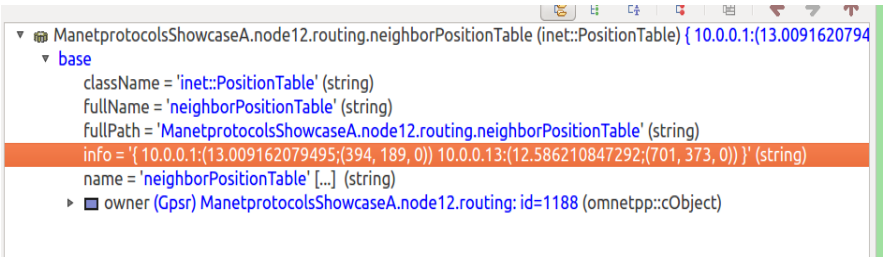
..routing.typename = "Gpsr"

The nodes start sending out GPSR beacons (and learning about the positions of their neighbors). Then, source sends a ping request packet. It gets forwarded along the chain to node9, which sends it to node5, as it is the closest to destination among node9's neighbors. However, node5 doesn't have any neighbors closer to the destination (and it is out of range of destination), thus it switches the packet to perimeter mode. It forwards the ping packet according to the right-hand rule. The packet gets to node1 and then back up along the chain through node9 again. Then node10 switches it back to greedy routing mode because node10 is closer to the destination than node5, where it was switched to perimeter mode. Then the packet arrives at destination.

The reply packet starts off in perimeter mode, as the destination is closer to source than destination's only neighbor, node4. The packet is switched back to greedy mode at node10 because it's closer to source than destination. From there, it gets to source through node9 and node11.

Note that the reply packet didn't get back on the same route as the request packet. Also, a packet is might not be routed to a closer neighbor because the sender doesn't yet know about it (and its position).

Also, note that there are no IP routes; the ipv4 module routing tables are empty. Instead, Gpsr maintains the positions of the nodes in communication range (those that a beacon was received from) and uses that for routing decisions. Here is node12's neighbor position table:



The table links node positions with IP addresses (it also contains the beacon arrival time).

3.2 ROUTING PROTOCOLS FOR VANET

As we discussed in earlier sections VANET inherits same characteristics as MANET. Due to high mobility, frequent changes in topology and limited life time are such characteristics of this network that make routing decisions more challenging. Several other factors such as road layout and different environments such as city and highway makes routing more challenging in VANET. As opposed to topology based routing of MANET, VANET uses position information of the participating nodes within the network to take routing decisions. Further we will discuss how position based routing used for VANET.

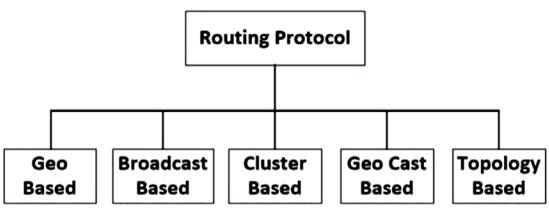
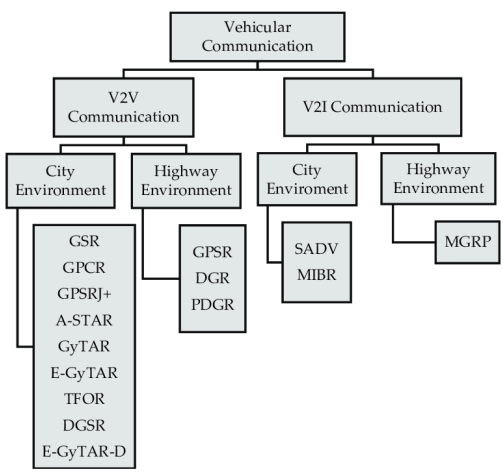


Figure 1: Classification of Routing Protocols.

3.2.1 Position Based Routing

The dynamic and highly mobile nature of VANET, where nodes behave very rapid and changes its location frequently demands such routing method that can deal with the environment of such network. These demands tend the researchers to use positions of nodes in order to provide successful communication from source to destination. Such method in which geographical positions of nodes are used to perform data routing from source to destination is called position based routing. Position based routing assumes that each node have knowledge about its physical/ geographic position by GPS or by some other position determining services. In it each node also has the knowledge of source, destination and other neighboring nodes. As compared to topology based routing, position based routing uses the additional information of each participating node to applicable in VANET, that additional information is gathered through GPS. Position based routing provides hop-by-hop communication to vehicular networks. A position based routing protocol consists of many major components such as "beaconing", "location service and servers" and "recovery and forwarding strategies". *Beaconing*: In it a node forwards packet with the current physical position and the unique id (IP ADDRESS). If node receives beacon from its neighbours then it updates its information in location table. Thus beaconing is used to gather information of node's one- hop neighbor or node's next hop neighbor.



Location service and servers: When a node does not contain current physical position of a specific node in its location table or want to know current physical position of any specific node then location service assisted to find current position of a specific node. To trace the current physical position of desired node, the requesting node sends location query with the unique ID of the desired node, sequence number and total number of hops. The neighbours reply this message until desired node found and if desired node lies among near neighbour's of the requested node then it replied with its current physical position message. In this way originating node updates desired node physical position information in the location table.

Forwarding and Recovery strategy: Forwarding and recovery strategy are used to forward data from source to destination node. Position based routing protocols used three types of forwarding methods for VANET in order to forward data packets from source to destination: 1) "restricted directional flooding" 2) "hierarchal forwarding" 3) "greedy forwarding" .

Restricted directional flooding sent data packets into the geographical area of specific node and the part of geographical area known as "forwarding zone". This method does not require information of neighboring nodes. The forwarding zone is created between source and destination nodes and the source node flood packet into the forwarding zone in order to send the packets towards destination. Overhead may be occur if large number of packets sent to the forwarding zone by source node that may results in expanding the area of forwarding zone. These issues can be overcome by adopting efficient flooding method such as "Distance-aware-timer-based Suppression method". Restricted directed flooding uses broad based protocols such as "Mobility-centric data dissemination algorithm for vehicular networks" (MDDV).

Another forwarding strategy for position based routing protocols is hierarchal forwarding in which protocols hierarchy is used as different steps to forward packets. The hierarchal forwarding performs routing for neighboring nodes and also for nodes at greater distance. Forwarding strategy for hierarchy routing used

"*geodesic packet forwarding*" (GPF) and anchored GPF that is defined in the terminodes project.

Another efficient forwarding strategy for position based routing is greedy forwarding in which a node sends packet to nodes closest to destination. The sending node calculated minimum hops for sending packet to destination. In fail situation where there is no node closest to the destination recovery strategy is used to overcome this kind of situation. Greedy perimeter stateless routing is an example of greedy forwarding strategy.

Unlike topology based routing, position based routing does not require any route maintenance. The route determined only when there is a need for forwarding packet. Another advantage of position based routing is that it contains information of source, destination and their neighboring nodes.

The aforementioned characteristics makes position based routing suitable for VANET. Several routing protocols have been proposed by many researchers that uses nodes position information for routing decisions. Although these routing protocols are most suitable for the vehicular communication but these protocols still have some challenges. We will discuss some of recently suggested protocols and the issues in these routing protocols. Furthermore, we will also investigate which recent advancement has been carried out to overcome these issues.

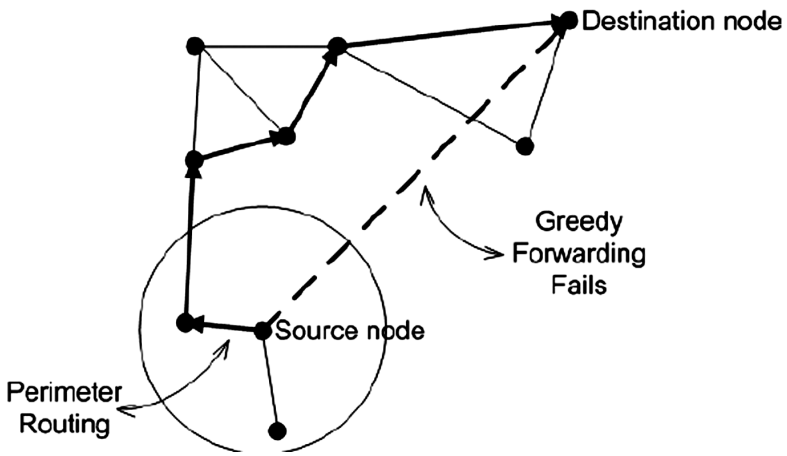
3.2.2 Greedy Perimeter Stateless Routing-GPSR

Greedy Perimeter Stateless Routing (GPSR) is one of the best examples of position based routing. GPSR uses closest neighbor◎s information of destination in order to forward packet. This method is also known as greedy forwarding. In GPSR each node has knowledge of its current physical position and also the neighboring nodes. The knowledge about node positions provides better routing and also provides knowledge about the destination. On the other hand neighboring nodes also assists to make forwarding decisions more correctly without the interference

of topology information. All information about nodes position gathered through GPS devices. GPSR protocol normally devised in to two groups:

- Greedy forwarding: This is used to send data to the closest nodes to destination.
- Perimeter forwarding: This is used to such regions where there is no closer node to destination. In other words we can say it is used where greedy forwarding fails.

Further we will see in detail how these forwarding strategy works and what are issues in them.



Greedy Forwarding

In this forwarding strategy data packets know the physical position of their destination. As the originator knows the position of its destination node so the greedy regions/hops are selected to forward the packets to the nodes that are closer to their destination. This process repeats until the packet successfully delivered to desired destination. Nearest neighbor's physical position is gathered by utilizing beaconing algorithms or simple beacons. When a neighboring node forwards packet to closer region to destination, the forwarding node receive a beacon message that contain IP address and position information. Then it updates its information in the location table. If forwarding node does not

receive beacon from its neighboring node within a specific time period, it assumes that either neighbor fails to forward packet to region closer to destination or neighbor's is not in its radio range. So it removes its entry from location table . The major advantage of greedy forwarding is that it holds current physical position of forwarding node. Thus by using this strategy total distance to destination becomes less and packets can be transmitted in short time period. Besides its advantages there are few drawbacks of this strategy i.e. there are some topologies used in it that limits the packet to move to a specific range or distance from the destination. Furthermore, this strategy fails when there are no closer neighbors available to destination.

Perimeter Forwarding

Perimeter forwarding is used where greedy forwarding fails. It means when there is no next hop closest neighbor to the destination is available then perimeter forwarding is used. Perimeter forwarding uses nodes in the void regions to forward packets towards destination. The perimeter forwarding used the right hand rule. In "right hand rule", the voids regions are exploited by traversing the path in counterclockwise direction in order to reach at specific destination. When a packet forward by source node, it forwarded in counterclockwise direction including destination node until it again reached at the source node. According to this rule each node involved to forward packet around the void region and each edge that is traversed are called perimeter. Edges may cross when right hand rule finds perimeter that are enclosed in the void by utilizing "heuristic approach". Heuristic has some drawbacks besides it provides maximum reach ability to destination. The drawback is that it removes without consideration of those edges which are repeated and this may cause the network partitions. To avoid this drawback another strategy is adopted that is described below.

Planarized Graph

When two or more edges cross each other in a single graph is called planar graph. "Relative Neighborhood Graph (RNG)" and "Gabriel Graph (GG)" are two types of planar graphs used to remove the crossing edges. Relative neighborhood graph (RNG) is defined as, when two edges intersect with radio range of each other and share the same area. For example, x and y are the two edges that share the area of two vertices x and y. The edge x, y are removed by using RNG because another edge from x towards v is already available.

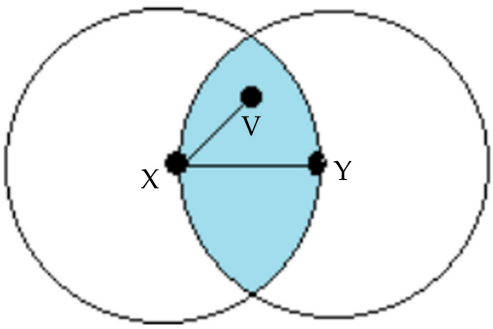


Figure 2: Example of RNG

Gabriel Graph (GG) is used to remove only those crossing edges which are in between the shared area of two nodes having the same diameter as the other nodes have. Figure 3 depicts GG:

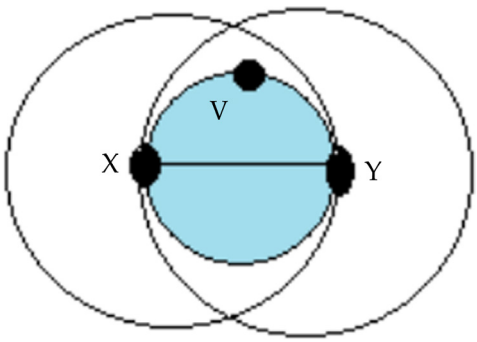


Figure 3: Example of GG

Figure 3 shows that the midpoint diameter is less than the diameter of node x or node y. Thus the edge from the x, y cannot be removed. So there is less network disconnection in the GG as compared to RNG.

Features of GPSR

GPSR combines the greedy forwarding with the perimeter forwarding to provide better routing decision on both full and planarized network graph by maintaining neighbor's information in the location table. For the forwarding decisions in perimeter mode GPSR packet header include the following distinct characteristics.

- GPSR packet header has the flag identity that is used to identify whether packet is in greedy forwarding or in perimeter forwarding.
- It contains destination node physical address.
- GPSR packet header also contains location of packet in the perimeter mode and the location of the new face to take a decision whether to hold the packet in the perimeter mode or to return it to the greedy mode.
- GPSR also have the record of sender and receivers address of the packet when the edge's crosses in the new face.

GPSR also have several distinct characteristics that are if the packet is in perimeter mode then its location address is compared to forwarded node address and if distance to location and destination node is less then packet it switched to greedy mode to forward packet towards destination. GPSR discard those packets that are repeatedly forwarded as destination for such packets are not in range. The packets in perimeter mode never send twice through the same link if destination is in range. Overall GPSR is an efficient example of the position based routing that uses the geographic location of nodes and reduced usage of routing state on each node. Furthermore, it provides maximum robustness in highly dynamic wireless ad hoc networks.

Issues in GPSR

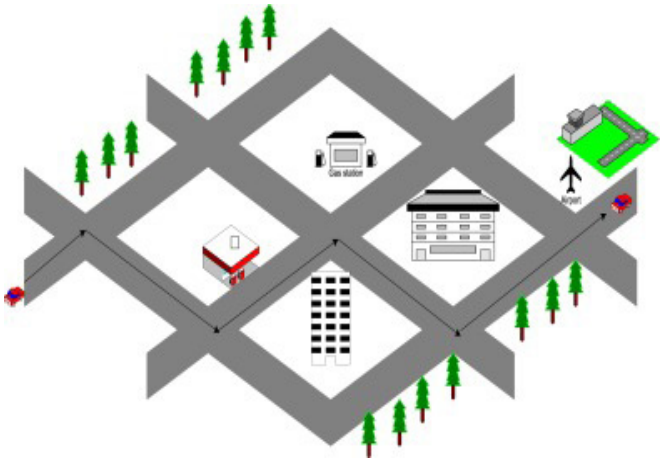
Besides GPSR certain characteristics, it suffers from several drawbacks. Greedy forwarding measured as unsuitable for the vehicular networks where the nodes are highly mobile and the node may not be able to maintain its next hop neighbors information as the other node may gone out of range due to high mobility. This can lead to data packets loss.

The second problem may occur during beaconing mechanism that beacons may lost due to channel destruction or bad signal. This problem can lead to removal of neighbor information from location table. GPSR uses planarized graphs as its repair strategy where greedy forwarding fails.

But these graphs perform well in the highway scenario due to their distributed algorithms. These graphs does not perform well in such environment of vehicular communication where a lot of radio obstacles involves, in addition to this their distributed nature may lead to certain partition of network and may lead to packet delivery impossible. Hence there is need of such position based routing protocols, which merge position information with the road topological structure in order to make possible vehicular communication in presence of radio obstacles.

3.2.3 Geographic Source Routing- GSR

Due to deficiencies of GPSR in presence of radio obstacles, network demanded new routing strategies that can compete with challenges occured due to radio obstacles. Therefore, Geographic Source Routing (GSR) is proposed . It deals with high mobility of nodes on one hand, on the other hand it uses roads layout to discover routes. GSR finds the destination node using “Reactive Location Service (RLS)”. GSR combines both geographic routing and road topology knowledge to ensure promising routing in the presence of radio obstacles .



Motivation

In city area there are buildings and tress etc that may create problems in direct communication among nodes. Hence, previously proposed protocol GPSR for highways may not perform well in city environment. The motivation for new routing protocol for city is stated below in details.

Frequently Network disconnection

Due to building and trees in city area, pure greedy position-based routing and its recovery mechanisms do not fully applicable . Nodes that can directly connect in free space cannot communicate in city area due to radio obstacles. As greedy position-based routing uses position of the nodes to find destination and planarization methods uses distance between nodes as connecting factor, that may not applicable in city due to unavailability of direct communication .

Multiple hops

In planarized connectivity, node send a packet to neighboring nodes until it reach at destination. In city area, planarized connectivity graph can increase delay due to the large number of nodes .

Routing Loops

Routing loops can be occurred in packets while using perimeter method due to mobility . Participation of a node in the network when the mobility is high can create routing loops. In city area, when there are many nodes participating in the communication at the same time, there are more chances of routing loops.

Incorrect route selection

In high mobility and too many hops, perimeter routing method can select a long route using "right hand rule" . The possibility of selecting and longer than necessary route is increased when there is more than one route available. High mobility and too many hops in city area may lead to incorrect route selection.

Working of GSR

GSR routing was proposed to deal with challenges faced by GPSR in city environment. There are two main issues in the city environment, one is dealing with high mobility issue in the city and other is topology structure of a city . In GSR position based routing is used that support the city map also. Vehicles have navigation system installed so getting map of city is normal. GSR use reactive location service to find the physical location for node.

RLS is used for position discovery in reactive position-based routing. In RLS a source node broadcast "position request" with some identification for the required node. When the node with that identification receives the position request, it responds with "position reply" containing its current physical position .

The sender node reaches the destination by using the road topology map and the above information. In other words in GSR the source node finds the shortest path to destination on the graph using simple graph algorithms and mark the packet with destination´s location. In this the packet travels through junctions to reach the destination.

Local recovery

GSR use "switch back to greedy" method for local recovery. After a packet reach to its local maximum, it switch back to greedy forwarding .

3.2.4 Anchor-based Street and Traffic Aware Routing- A-STAR

Anchor-based Street and Traffic Aware Routing (A-STAR) is position based routing protocol. The development of A-STAR was inconsideration with city environment. In city area, almost all roads and streets are covered by big buildings and there are close ends in the streets and so frequent stop signal, turns and speed breakers make routing more challenging. Problems faced by the position based routing protocols in city environment defined before in GSR. The capability of A-STAR protocol to overcome these problems will be defined here. A-STAR is anchor based routing protocol. In anchor based routing before transmitting the packet, source node address add in the header of packet and information of all intermediate node junction that packet must travel to reach the destination . To use city maps and road information of town to make routing decisions called "Spatial Aware Routing". Spatial awareness is used to get topology information and different nodes position in the network. Mostly anchor based routing and spatial aware routing used together .

Issues in city environment

In position based routing, every node sends it current position by a beacon message and every node knows its neighbor nodes. When a source send message to the destination it uses the geographic location of the destination. There are some limitations in position based routing protocols that are discussed and defined in previous topic as well. The challenges in city environment can be better understand by following example.

A source node wants to send packet to the destination. There are buildings between source and destination and there is no node closer to the destination. Two separate paths are available to the destination; one is shorter than other. But when in GPSR this situation occurs, GPSR will select the route according to its right hand rule. So GPSR will not look for shortest path, it will look for right hand rule. And packet will traverse hop-by-hop until it finds a node nearer to the destination. This takes much longer time and processing.

Working of A-STAR

Same like GSR, A-STAR was proposed for city environment. Both GSR and A-STAR compute the number of junctions to reach the destination but A-STAR also use traffic information and street awareness in path finding . In street awareness, A-STAR gets the anchor information according to the street map. A-STAR has two new features that make it differ from GSR in working. A-STAR uses statically and dynamically rated maps to find the number of junctions. In statistically rated maps, ASTAR uses schedule of buses to ensure the high connectivity e.g. some streets are served by regular city buses their connectivity can be high due to presence of city buses. In dynamically rated maps, A-STAR collect the latest information of traffic to find the anchors/junctions to compute the path e.g. some roads are wider than other so there are more traffic. It means that connectivity is high on wider roads with high traffic (more vehicles). Using this traffic information A-Star assign the weight to the street e.g. more vehicles less weight and less vehicles more weight. This dynamic process helps this protocol to calculate anchors more accurately .

Local recovery

Both the recovery strategies of GPSR i.e. perimeter mode and GSR i.e. switch back to greedy are insufficient in city environments of VANET.

A-STAR uses a new recovery method. When a packet face problem

to pass from a junction, that junction is marked as "out of service" so other packets are restricted to traverse that junction until that junction changed to "Operational" state . When any junction is out of order each node in the network is informed about that junction and updates their routing information and city maps by marking that place out of order. Therefore, no node will use that junction as anchor to be traverse to reach destination. When the out of order junction becomes operational each node aware about the usage of that junction and may adopt that junction for forwarding the packet towards destination. So as compared to other position based routing protocols, A-STAR adopt higher connectivity anchor based paths to find the route towards destination in large city environments.

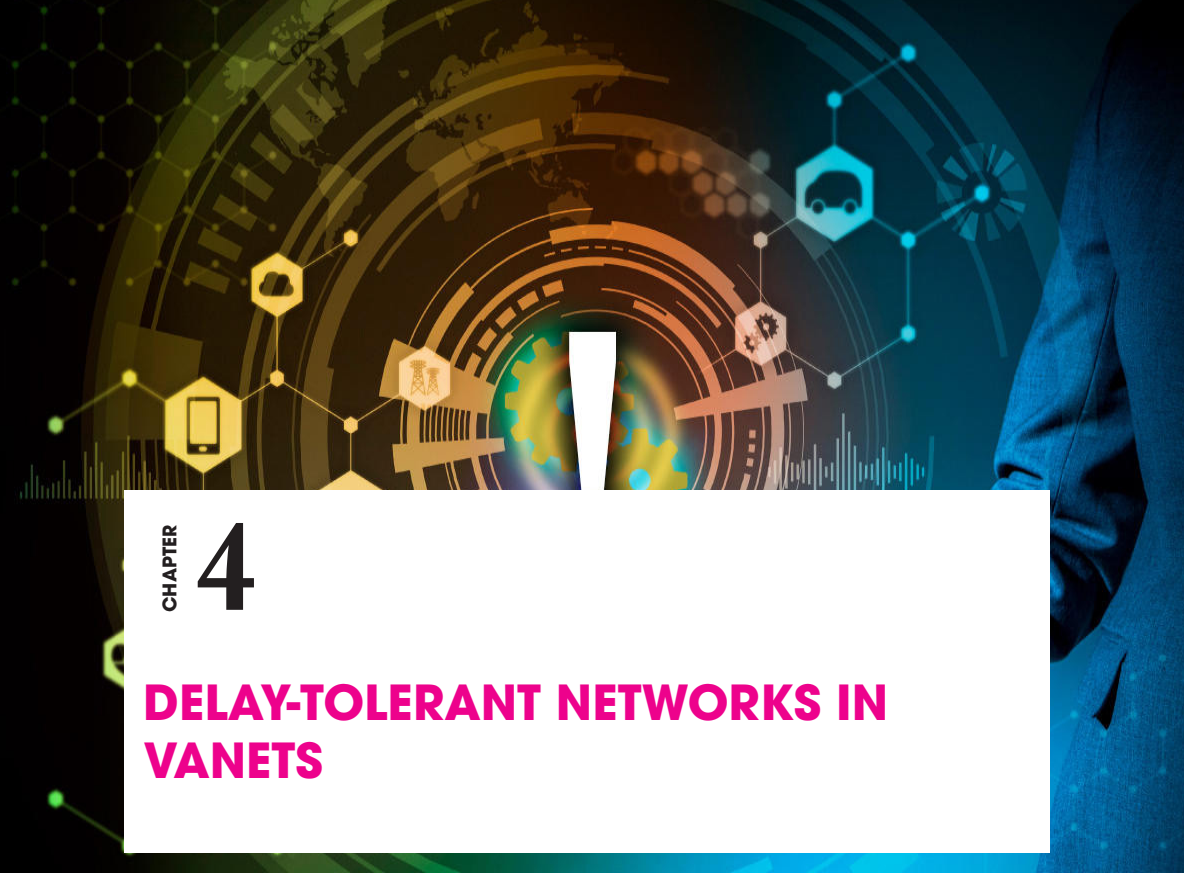
Table 1: Comparison of Various Protocols

Protocols	Proactive Protocols	Reactive Protocols	Position based Greedy Protocols	Delay Bounded Protocols	Cluster Based Protocols	Broadcast Protocols	Geo cast Protocols
Prior Forwarding Method	Wire less multi hop Forwarding	Wire less multi hop Forwarding	Heuristic method	Carry & Forward	Wireless Multi hop Forwarding	Wire less multi hop Forwarding	Wire less multi hop Forwarding
Digital Map Requirement	No	No	No	No	Yes	No	No
Virtual Infrastructure Requirement	No	No	No	No	Yes	No	No
Realistic Traffic Flow	Yes	Yes	Yes	No	No	Yes	Yes
Recovery Strategy	Multi Hop Forwarding	Carry & Forward	Carry & Forward	Multi hop Forwarding	Carry & Forward	Carry & Forward	Flooding
Scenario	Urban	Urban	Urban	Sparse	Urban	Highway	Highway

REFERENCES

1. Ad hoc On-Demand Distance Vector (AODV) Routing
2. Cano, Jose; Cano, Juan-Carlos; Toh, Chai-Keong; Calafate, Carlos T.; Manzoni, Pietro (2010). "EasyMANET: an extensible and configurabi'le platform for service provisioning in MANET environments". IEEE Communications Magazine. 48 (12): 159–167. doi:10.1109/mcom.2010.5673087. S2CID 20381835.
3. David Adrian, Karthikeyan Bhargavan, Zakir Durumeric, Pierrick Gaudry, Matthew Green, J. Alex Halderman, Nadia Heninger, Drew Springall, Emmanuel Thomé, Luke Valenta, Benjamin VanderSloot, Eric Wustrow, Santiago Zanella-Béguelin, Paul Zimmermann, "Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice", preprint at weakdh.org, May 2015.
4. Destination-Sequenced Distance Vector (DSDV) Protocol
5. Djenouri, D.; Kheladi, L.; Badache, N. (October 2005). "A Survey of Security Issues in Mobile Ad hoc and Sensor Networks". IEEE Communications Surveys and Tutorials. 7 (4): 2–28. doi:10.1109/COMST.2005.1593277. S2CID 11135536.
6. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks
7. Jhaveri, Rutvij H.; Patel, Narendra M. (2015). "A Sequence Number Based Bait Detection Scheme to Thwart Grayhole Attack in Mobile Ad-hoc Networks". Wireless Networks-The Journal of Mobile Communication, Computation and Information. 21 (8): 2781–2798. doi:10.1007/s11276-015-0945-9. S2CID 19934099.
8. Jhaveri, Rutvij H.; Patel, Narendra M. (2017). "Attack-pattern discovery based enhanced trust model for secure routing in mobile ad-hoc networks". International Journal of Communication Systems. 30 (7): e3148. doi:10.1002/dac.3148.
9. Maihöfer, C. (April 2004). "A Survey on Geocast Routing Protocols". IEEE Communications Surveys and Tutorials. 6 (2): 32–42. doi:10.1109/COMST.2004.5342238.

10. Mauve, M.; Widmer, J.; Hartenstein, H. (December 2001). "A Survey on Position-Based Routing in Mobile Ad Hoc Networks". IEEE Network. 1 (6): 30–39. CiteSeerX 10.1.1.25.2774. doi:10.1109/65.967595.
11. Satyajeet, D.; Deshmukh, A. R.; Dorle, S. S. (January 2016). "Article: Heterogeneous Approaches for Cluster based Routing Protocol in Vehicular Ad Hoc Network (VANET)". International Journal of Computer Applications. 134 (12): 1–8. Bibcode:2016IJCA..134l...1S. doi:10.5120/ijca2016908080.



CHAPTER 4

DELAY-TOLERANT NETWORKS IN VANETS

INTRODUCTION

Delay Tolerant Networks (DTN) have been utilized in various operational communication paradigms. This includes the communication scenarios that are subject to disruption and disconnection as well as the scenarios with high delay and frequent partitioning, i.e., Vehicular Ad hoc Networks (VANETs). Due to several characteristics match, a new research paradigm named as Vehicular Delay Tolerant Network (VDTN) is introduced. Through relays and store-carry-forward mechanisms, messages in VDTNs can be delivered to the destination without an end-to-end connection for delay-tolerant applications.

In many commercial applications and in road safety systems, vehicular delay-tolerant networks have been envisioned to be useful. For example, a vehicular ad hoc network (VANET) can be used to alert drivers of traffic jams ahead, help balance traffic loads, and reduce traveling time. It can also be used to propagate

emergency warnings to drivers behind the vehicles in an accident in order to prevent compounding on accident that has already taken place.

4.1. DELAY-TOLERANT NETWORKS

Delay Tolerant Networking (DTNs) is a new way of communication that facilitates the data transfer between source and destination even if a fully connected path may not exist between two end nodes. The Delay Tolerant Network (DTN) is an emerging area that has attracted keen research efforts from both academia and industry. DTNs consider an extreme network condition that is different from the traditional communication networks. There may not exist a complete end-to-end path between the data source and destination, and thus network is subject to dynamic node connections and unstable topologies. The communication in DTN is done by exploiting the characteristic of nodes i.e. mobility, available connections, and provided buffer space etc. DTNs find broad applications in the situations where legacy networks cannot work effectively, such as data communications in rural areas where stable communications infrastructure is not available or is costly. DTN is useful for extreme environments like battlefields, volcanic regions, deep oceans, deep space, developing regions etc., where they suffer challenging conditions as military wars and conflicts, terrorist attacks, earthquakes, volcanic eruptions, floods, storms, hurricanes, severe electromagnetic interferences, congested usage, etc. These challenging conditions result in excessive delays, severe bandwidth restrictions, remarkable node mobility, frequent power outages and recurring communication obstructions. Vehicular networking is a wide and growing field of DTNs, where many applications are being explored. One of these applications is to provide Internet access to vehicles by connecting to roadside wireless base stations. Non-commercial applications include monitoring and tracking wildlife animals, and environmental monitoring, such as lake water quality monitoring and roadside noise monitoring. DTNs can be applied in a variety of other fields ranging from healthcare to education to economic efficiency.

The idea of Delay Tolerant Network (DTN) was taken from Inter Planetary Networks (IPN), this was started in 1970s. The IPN was invented to communicate between earth and mars. The DTN is a type of wireless ad-hoc network which tolerates the intermittent connectivity. The intermittent connectivity can be defined as the sudden change of state (up/down) of any communication link between the nodes. The DTN can also be defined as intermittently connected wireless ad-hoc network ("Mobile Ad-Hoc and", n. d.) that can tolerate longer delays, intermittent connectivity and prevent data from being lost by using store-carry-forward approach. The Store-carry forward approach enables the nodes to take the message, store it in the buffer provided at each node and forward the same whenever new node comes in its communication range. DTN technology has become a new research focus in many fields including deep space communications, military tactical communications, and disaster rescue and internet access in remote areas. Internet Research Task Force (IRTF) has organized Delay-Tolerant Research Group (DTNRG) to research OTN technology, and as an important research theme.

With the advent of the Internet of Things (IoT) a number of new devices will become part of our day today life. Constrained Application Protocol (CoAP), and its extensions, are specially designed to address the integration of these constrained devices in IoT environment. However, due to their limited resources, they are often unable to be fully connected and instead form intermittently connected and sparse networks in which Delay Tolerant Networking (DTN) is more appropriate, in particular through the Bundle Protocol (BP).

4.1.1. Characteristics of Delay Tolerant Networks

As the node's mobility and energy are limited, DTN frequently disconnects, thus resulting in continue change in DTN topology. That is to say, the network keeps the status of intermittent connection and partial connection so that there is no guarantee to achieve end-to-end route.

High Delay, Low Efficiency, and High Queue Delay

End-to-end delay specifies the sum of the total delay of each hop on the specified route. The end-to-end delay involves queuing time, waiting time and transmission time. Each hop delay might be very high due to the fact that DTN intermittent connection keeps unreachable in a very long time and thus further leading to a lower data rate and showing the asymmetric features in up-down link data rate. In addition, queuing delay plays a main role in end-to-end delay and frequent fragmentations in DTN make queuing delay increasing.

Limited Resource

Node's computing and processing ability, communication ability and storage space is weaker than the function of an ordinary computer due to the constraints of price, volume and power. In addition, the limited storage space resulted in higher packet loss rate.

Limited Life Time of Node

In some special circumstances of the restricted network, the node is common to use the battery power on the state of hostile environment or in harsh conditions, which will cut the life time of node. When the power is off, then the node cannot guarantee normal work. That is to say, it is very possible the power is off when the message is being transmitted.

Dynamic Topology

Note that the DTN topology is dynamic changing for some reasons such as environmental changes, energy depletion or other failures, which results in dropping out of network. The requirements of entering DTN also make topology change.

Poor Security

Due to the lack of specialized services and maintenance in real world DTN is vulnerable to threats like eavesdropping, message modification, routing spoofing and Denial of Service (DoS) etc.

Heterogeneous Interconnection

The architecture of DTN is based on asynchronous message forward and operates as an overlay above the transport layer. DTN can run on different heterogeneous network protocol stacks and DTN gateway ensures the reliable transmission of interconnection message.

The above mentioned characteristics make DTNs different from traditional wired networks and mobile ad-hoc networks.

4.1.2. Types of DTNs

DTN for Satellite Communications

Space communication can be generally characterized by long link delay and frequent link disruptions. Despite of these characteristics of space communication DTN has been developed to enable automated network communications. DTN was originated from a generalization of requirements identified for interplanetary networking (IPN). Ordinary TCP/IP architectures fail to provide satisfactory performance because of the presence of one or more of the following impairments: long delays, disruptions, intermittent links, network partitioning etc. Satellite network is one among the challenged network. It is the network that includes one or more satellite links. LEO (low earth orbit) satellite networks were immediately recognized as a perfect candidate for DTN applications, because of the satellite link intermittency. In deep space and LEO satellite networks the communication opportunities or contacts are known in advance i.e. are fully deterministic as

they are related to the orbital characteristics of planets and space assets. This kind of connectivity is addressed by specific DTN solutions such as "scheduled contacts" where transport protocol connections start and stop at the beginning and at the end of contacts. In such networks routing must be designed to cope with scheduled contacts and not with opportunistic connectivity as in other challenged networks, therefore specific routing algorithms as contact graph routing designed by NASA. On the other hand GEO satellite networks are not pure challenged networks because they can offer a continuous connectivity at least for fixed terminals. However they are classified as challenged networks because of its long propagation delay of order 600 Ms.

DTN for Deep Space Communications

Delay/disruption tolerant networking (DTN) technology offers a novel way to significantly stressed communications in space environments, especially those with long link delay and frequent link disruptions in deep space missions. DTN was considered as the most suitable technology to be employed in space internetworking by NASA and hopes to fly with it on space missions soon. There are numerous research work has been done related with DTN for space communications in the past several years. The Space Internetworking Strategy Group (SISG), which is composed of technical experts appointed by the Interagency Operations Advisory Group (IOAG) agencies, considers DTN to be the only mature candidate protocol available to handle long propagation delays, frequent and lengthy network disruption inherent in space missions involving multiple spacecraft.

Vehicular DTN (VDTN)

Vehicular Delay-Tolerant Networks (VDTNs) are DTNs where vehicles communicate with each other and with fixed nodes placed along the roads in order to disseminate messages. Some of the potential applications for these networks are the following: notification of traffic conditions (unexpected

jams),road accident warnings, weather reports (ice, snow, fog, and wind),advertisements (free parking spots, nearby fuel prices, etc.),cooperative vehicle collision avoidance, web or email access, or even the gathering of information collected by vehicles such as road pavement defects. Vehicular networks have also been proposed to implement transient networks to benefit developing communities and disaster recovery networks.

DTN for Underwater Communications

Underwater networks (UWNs) have the potential to find applications in a wide range of aquatic activities, such as oceanographic data collection, pollution monitoring, offshore exploration, seismic monitoring, assisted navigation and tactical surveillance. In most cases, these networks will operate in harsh and constrained environments where communication disruption (and, hence, delay) is frequent. In this respect, an underwater network can be viewed as a delay/disruption-tolerant network (DTN) requiring specialized communication protocols.

DTN for Emergency Communications

Delay-tolerant networks can be used to improve situational awareness during the response to a largescale disaster. Delay/Disruption Tolerant Networks (DTNs) can be used in man-made or natural disaster stricken areas with communication infrastructure breakdown or power outages. DTN has been developed as a solution to wireless networks experiencing frequent disruptions. DTNs can provide communication support in disaster relief and rescue operations. An evaluation carried out by using DTN MapEx a disaster map generator that operates over a DTN with responders and volunteers, carrying mobile devices shows that DTN can improve information availability in disaster stricken areas.

4.1.3. Applications of DTNs

Deep Space Exploration

NASA and other agencies will plan a series of projects of lunar exploration, Mars exploration and others. In September, 2003, Cisco router (CL EO) was launched by satellite to monitor disaster in UK. Till to December 2008, CL EO has done a lot of routing tests in space environment including using Saratoga protocol of bundle layer instead of pervious protocol making full use of the link source to overcome serious asymmetry link conditions. The experiment shows it is feasible to use Bundle Protocol in space.

Studies of Wild Zebra

The Zebranet project has installed a global positioning system (GPS) in a zebra collar to study the habits of zebra activities, which is one of the early DTN projects and was started in 2004. Collars start every few minutes to record GPS location information, and every 2 h open radio function, when two collars' distance is in communication range they would exchange information (adopted Epidemic routing algorithms). After a period of time, every horse collar stores the position information of others activities. In this experiment, the researcher can know the exact location of zebra only with little information. The further experiment of this project is to resolve the issues of equipment energy, adaptability and data compression.

Rural Communication

There are many rural communication projects in remote villages to provide the access to Internet. Some of which is try to reduce the cost of communications using the way of asynchronous information transmission. For example, Wizzy digital courier service provides Internet access for some village schools in South Africa. This project adopted a simple one-hop delay network,

letting couriers drive a motorcycle with USB storage device to come and go between rural schools and cities with permanent Internet connection (such a round-trip may take several hours of time), so as to realize the connection between the school and the Internet.

Lake Quality Monitoring

European Union advises state and local government to launch protect water quality activities, in this project, the researchers didn't choose end-to-end communication mode, but using special node (data mule) in the lake to cruise, realizing DTN storage and forwarding mechanism. When the ship (data mule) back to dock, mule can exchange information with the gathering nodes accessing Internet. In this project, using data mules--besides low overhead--still can be independent with infrastructures and set flexibly in various carries. Other one such as Ad hoc being used for collecting battlefield information or collecting data in depopulated area is actually one application of DTN. That's to say, DTN has come into people's lives. Notice that with further development of DTN research, its range of applications will be larger, and more fields will be benefited.

Military Applications

Military communication network is a multi-hop wireless network, and is an ad hoc network. As a result of the impact of the battlefield special circumstances, such as, mobile nodes, enemy interference, geographical environment, etc, the connection between network nodes is intermittent, uncertainties and non-periodic. Therefore military communication network is a typical DTN network. DTN technology can be fully applied in military communication networks.

Public Transportation System

There are several promising applications of DTN in public transport.

Data Dissemination Application

For high volume and non-urgent data, it is not wise to use expensive network transmission techniques. Instead, DTN technique could be used as a low cost data dissemination method in Fog computing, particular for data dissemination among Fog servers and mobile devices. DakNet is a DTN technique based application and developed by researchers from the MIT Media Lab. It has been deployed in remote parts of Cambodia and India at a cost two orders of magnitude less compared to traditional landlines networks.

4.1.4. DTN Architecture

The existing TCP/IP-based internet, while fabulously successful in many environments, does not suit all environments. The ability of the "TCP/IP suite" to provide service depends on a number of important assumptions:

(i) Existence of end-to-end path between source and destination during communication session;

(ii) (For reliable communication) that the maximum round-trip time over that path is not excessive and not highly variable from packet to packet; and

(iii) That the end-to-end loss is relatively small.

Delay Tolerant Networks may not satisfy some of the assumptions due to their different characteristics such as long or variable delays, frequent partitioning, data rate asymmetry and interoperating among differently-challenged networks. The DTN architecture should provide the means for dissimilar networks to interoperate. The network architecture used for the conventional networks may not be used as it is for DTNs.

The DTN architecture provides a common solution for interconnecting heterogeneous gateways or proxies that employ store-and-forward message routing to overcome communication disruptions. At its inception, the concepts behind the DTN

architecture were primarily targeted at tolerating long delays and predictably-interrupted communications over long distances (i.e., in deep space). At this point in time, the work was architecture for the Interplanetary Internet (IPN). When the first draft of the eventual RFC 4838 was published, one of the authors had coined the term Delay Tolerant Networking suggesting the intention to extend the IPN concept to other types of networks, specifically including terrestrial wireless networks. Terrestrial wireless networks also suffer disruptions and delay, and the DTN architectural emphasis grew from scheduled connectivity in the IPN case to include other types of networks and patterns of connectivity (e.g., opportunistic mobile ad-hoc networks with nodes that remain off for significant periods of time).

The DTN architecture creates a "network of Internets" by providing an end-to-end layer above the transport layer. We call this the "bundle layer". The name "bundle" derives from considering protocols that attempt to minimize the number of round-trip exchanges required to complete a protocol transaction, and dates back to the original IPN work. By "bundling" together all information required completing a transaction (e.g., protocol options and authentication data), the number of exchanges can be reduced, which is of considerable interest if the round trip time is hours, days or weeks. Bundles comprise a collection of typed blocks. Each block contains meta-data; some also contain application data. The first or primary block of each bundle, illustrated in Figure 1, contains the DTN equivalents of the data typically found in an IP header on the Internet: version, source and destination EIDs, length, processing flags, and (optional) fragmentation information. It also contains some additional fields, more specific to the bundle protocol: report-to EID, current custodian EID, creation timestamp and sequence number, lifetime and a dictionary. Most fields are variable in length, and use a relatively compact notation called self-delimiting numerical values (SDNVs). Early designs for the primary bundle block used more fixed-length fields, but the relative merit of choosing a fixed-length field for simplicity was ultimately found to be less compelling than the flexibility offered by SDNVs. By setting various bits in the

bundle processing control flags, the sender can request a report for any of the following events: receipt at destination node, custody acceptance at a node, bundle forwarded/deleted/delivered route, and receipt by destination application.

Version(1 byte)	Bundle Processing Control Flags (SDNV)
Block Length (SDNV)	
Destination Scheme Offset (SDNV)	Destination SSP Offset (SDNV)
Source Scheme Offset (SDNV)	Source SSP Offset (SDNV)
Report-to Scheme Offset (SDNV)	Report-to SSP Offset (SDNV)
Custodian Scheme Offset (SDNV)	Custodian SSP Offset (SDNV)
Creation Timestamp (SDNV)	
Creation Timestamp Sequence Number (SDNV)	
Lifetime (SDNV)	
Dictionary Length (SDNV)	
Dictionary (byte array)	
Fragment Offset (SDNV, optional)	
Application data unit length (SDNV, optional)	

Figure 1. The structure of the primary block of a bundle.

The DTN architecture uses store-and-forward message switching technique by overlaying a new transmission protocol, called the bundle protocol on top of the lower-layer protocols such as Internet protocols. The bundle protocol ties together the lower-layer protocols so that application programs can communicate across the same or different sets of lower-layer protocols under conditions that involve long network delays or disruptions. The bundle-protocol agent stores and forwards entire bundles (or bundle fragments) between nodes. A single bundle protocol is used throughout a DTN. On the other hand, the lower-lower protocols below the bundle protocol are chosen depending on the characteristics of each communication environment. The figure below (top) illustrates the bundle-protocol overlay and (bottom) compares the Internet protocol stack (left) with a DTN protocol stack (right).

Store-and-Forward Message Switching

DTNs use store-and forward message switching technique to resolve the problems associated with intermittent connectivity, long or variable delay, asymmetric data rates, and high error rates. A DTN enabled application sends messages of arbitrary length, also called Application Data Units or ADUs. Whole messages (ADUs) or pieces (fragments) of such messages are forwarded from a storage place on one node (switch intersection) to a storage place on another node, along a path that eventually reaches the destination.

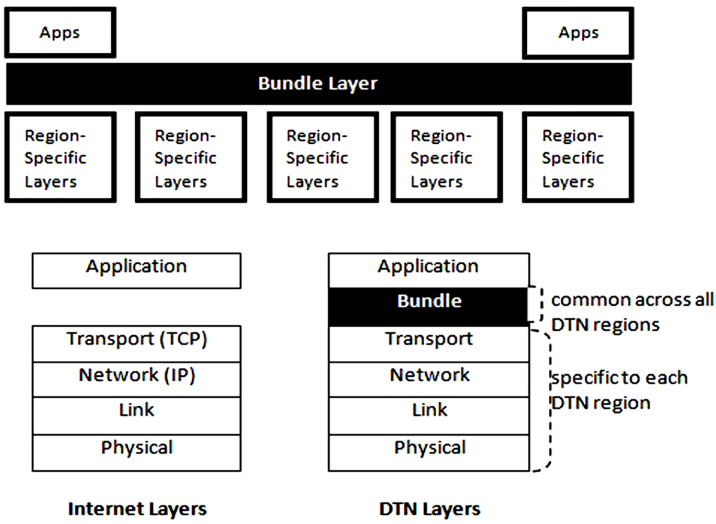


Figure 2. Bundle-protocol overlay with DTN protocol stack.

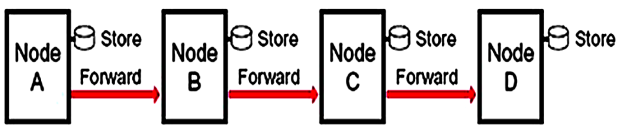


Figure 3. Store-And-Forward Message Switching.

Store-and-forwarding methods are also used here are not node-to-node relays (as shown above) but rather star relays where both the source and destination independently contact a central storage device at the center of the links.

DTN routers need persistent storage for their queues for one or more of the following reasons:

- A communication link to the next hop may not be available for a long time.
- One node in a communicating pair may send or receive data much faster or more reliably than the other node.
- A message, once transmitted, may need to be retransmitted if an error occurs at an upstream (toward the destination) node, or if an upstream node declines acceptance of a forwarded message.

By moving whole messages (or fragments thereof) in a single transfer, the message-switching technique provides network nodes with immediate knowledge of the size of messages, and therefore the requirements for intermediate storage space and retransmission bandwidth.

Nodes and Endpoints

A node is an entity with a bundle-protocol agent overlaid on lower-layer communication protocols in DTN. At any moment, a given node may act as a source, destination, or forwarder of bundles:

Source or Destination Function

As a source or destination, a node sends or receives bundles to or from another node, but it does not forward bundles received from other nodes. If the node operates over long-delay links, its bundle protocol requires persistent storage in which to queue bundles until outbound links are available. The node may optionally support custody transfers.

Forwarding Function

A DTN node can forward bundles between two or more other nodes in one of two situations:

Routing-Equivalent Forwarding. The node forwards bundles between two or more other nodes, each of which implement the same lowerlayer protocols as the forwarding node. If a forwarding node operates over long-delay links, its bundle protocol requires persistent storage in which to queue bundles until outbound links are available. The node may optionally support custody transfers.

Gateway-Equivalent Forwarding. The node forwards bundles between two or more other nodes, each of which implement different lowerlayer protocols while the forwarding node implements all such protocols. The node must have persistent storage; support for custody transfers is optional but typically advisable.

A bundle endpoint is a set of zero or more nodes that all identify themselves by the same endpoint ID. The common case in which only one node has a given endpoint ID is called a singleton endpoint. Every node is uniquely identified by at least one singleton endpoint. Source nodes are always singleton endpoints or null (anonymous source) endpoints, and destination nodes may or may not be singleton endpoints. Endpoints may also be multicast (multiple destination nodes with the same endpoint ID) or null (no nodes). Endpoints may contain multiple nodes, and nodes may be members of multiple endpoints.

Priority Classes

The DTN architecture offers relative measures of priority (low, medium, high) for delivering ADUs. These priorities differentiate traffic based upon an application's desire to affect the delivery urgency for ADUs, and are carried in bundle blocks generated by the bundle layer based on information specified by the application.

Three relative priority classes are defined to date. These priority classes typically imply some relative scheduling prioritization among bundles in queue at a sender:

Bulk

Bulk bundles are shipped on a "least effort" basis. No bundles of this class will be shipped until all bundles of other classes bound for the same destination and originating from the same source have been shipped.

Normal

Normal-class bundles are shipped prior to any bulk-class bundles and are otherwise the same as bulk bundles.

Expedited

Expedited bundles, in general, are shipped prior to bundles of other classes and are otherwise the same. Applications specify their requested priority class and data life time for each ADU they send. This information, coupled with policy applied at DTN nodes that select how messages are forwarded and which routing algorithms are in use, affects the overall likelihood and timeliness of ADU delivery. The priority class of a bundle is only required to relate to other bundles from the same source. This means that a high priority bundle from one source may not be delivered faster (or with some other superior quality of service) than a medium priority bundle from a different source. It does mean that a high priority bundle from one source will be handled preferentially to a lower priority bundle sent from the same source.

Congestion Control

The Delay Tolerant Networking architecture (DTN) supports a custody transfer concept implemented by an acknowledged transfer of data to persistent, reliable storage. A node "taking custody" of a message makes a commitment to deliver the message to its destination or another custodian node, effectively migrating one or both of the ends described in the end-to-end argument to new locations. The goal of custody transfer is to use hop-by-

hop (custodian-to-custodian) reliability to improve end-to-end reliability and to free retransmission buffers at a sender as soon as possible. To implement this facility, the node taking custody ("custodian") must generally reserve storage for messages it takes custody of, resulting in a reduced amount of storage remaining for either taking custody of subsequent messages or for merely doing its ordinary task of switching messages. When faced with persistent demand, a custodian unable to release or otherwise transfer custody of its messages will ultimately exhaust its storage resources– a form of DTN congestion. This type of congestion can easily result in head-of-line blocking, preventing further traffic from flowing even when some outgoing connections are available. Easing congestion at a custodian is a nontrivial task. The options include discarding messages, moving them toward their ultimate destination (typically the most desirable case), or moving them to some other place. The potential of long delays and interruptions of custody transfer operations between custodians makes the management of message migration to combat congestion especially difficult.

4.1.5. Routing and Buffer Management in DTN

DTN uses store-carry-and-forward protocols: there, a node may store a message in its buffer and carry it along for long periods of time, until an appropriate forwarding opportunity arises. Additionally, multiple message replicas are often propagated to increase delivery probability. This combination of long-term storage and replication imposes a high storage overhead on unbounded nodes (e.g. handhelds). Thus, efficient buffer management policies are necessary to decide which messages should be discarded, when node buffers are operated close to their capacity. This section highlights the issues and challenges in buffer management in DTN. This section will also cover some efficient approaches for buffer management in DTN.

Delay Tolerant Networks are wireless networks where disconnections may occur frequently due to propagation phenomena, node mobility, and power outages. Propagation

delays may also be long due to the operational environment (e.g. deep space, underwater). In order to achieve data delivery in such challenging networking environments, researchers have proposed the use of store-carry-and-forward protocols: there, a node may store a message in its buffer and carry it along for long periods of time, until an appropriate forwarding opportunity arises. Additionally, multiple message replicas are often propagated to increase delivery probability. This combination of long-term storage and replication imposes a high storage overhead on untethered nodes (e.g. handhelds). Thus, efficient buffer management policies are necessary to decide which messages should be discarded, when node buffers are operated close to their capacity.

In DTN, the "store-carry-forward" mechanism is used for message transmission. These messages are delivered to their final destinations in a hop-by-hop manner. As a result, many problems arise such as how to drop and how to schedule the messages, in the buffer due to the impulsive nature of the nodes. Many changeable situations may occur like limited storage node capacity, short contact duration between the two nodes, and so on. Buffer Management technology is a fundamental approach that manages the various resources among different situations as per the technique used. An efficient buffer management technique decides at each step which of the messages is to be dropped first, when the buffer is full as well as which messages are to be transmitted, when bandwidth is limited.

The nodes in the DTN require proper buffer management approach to get low delay and high data delivery. The buffer management, in this case, refers to the proper use of scheduling and dropping policies used by the nodes at the time of the buffer overflow and congestion.

Buffer Management Policies

The popular dropping policies techniques for buffer management used in DTNs are described.

Drop Least Recently Received (DLR)

In the DLR buffer management technique, as the name implies, the packet which is stays for a long time in the buffer will be dropped first. This is due to the fact that it has less probability of being conceded to the other nodes.

Drop Oldest (DOA)

In the DOA technique, the message with the shortest remaining life time (TTL) is dropped first. The idea behind dropping such messages is that of the messages whose TTL is small, then these are in the network from a long period of time and, thus, have the high probability of having already been delivered.

Drop Front (DF)

This technique drops the messages on the basis of the order in which they enter into the buffer. For example, the first message that enters the queue will be the first to be dropped.

Drop Largest (DLA)

In the Drop Largest (DLA) buffer management technique, the message with a large size will be selected in order to be dropped.

Evict Most Forwarded First (MOFO)

MOFO attempts to maximize the propagation of the messages through the network by dropping those messages that have been forwarded the maximum number of times. As such, the messages with a lower hop count are able to travel further within the network.

Drop Last (DL)

Drop the newly received message, irrespective of whether it is new or old, that is why responsible for maximize drop ratio.

Evict Most Favourably Forwarded First (MOPR)

MOPR maintains the value of each message in its queue. Thus, each time when a message is replicated the value in the message is increased based on the predictability of the message being delivered. Therefore, the message with the highest value is dropped first.

Evict Shortest Life Time First (SHLI)

This technique uses the timeout value of the message, which indicates when it is no longer useful. This means that a message with the shortest remaining life time is dropped first.

Evict Least Probable First (LEPR)

This technique works by a node ranking the messages within its buffer based on the predicted probability of delivery. The message with the lowest probability is dropped first .Basically buffer management policies can be divided into three types:

- Global buffer management policy which utilize network-wide information regarding all messages.
- Local buffer management policy which use partial network knowledge like number of copies of message in the network, instead of all network-wide information correlated with messages and additional message properties like remaining TTL, size etc.
- Traditional buffer management policies like drop head, drop tail, drop random.

Routing in DTN

The traditional routing protocols which consider an essential platform for most traditional mobile networks do not work well in DTN since these protocols assume an existing of the continuous route between the source node and destination. Since the DTN are intermittently connected networks where a continuous end-to-end path may not exist, the main objective of routing in DTN is to maximize message delivery to the destination while minimizing end-to-end delay. The routing protocols in DTN can be differentiated based on queue management, the amount of information available when making the forwarding decisions and the number of destinations a message can have.

Routing in DTN is the main issues and challenging because of frequent and long duration periods of disconnectivity. The properties of DTN certainly raise a number of interesting issues in routing which are summarized as follow:

Routing Objectives

The main and most important routing objectives in DTN are to minimize resource consumption such as network bandwidth, battery energy, and network bandwidth as well as maximize message delivery probability.

Buffer Space Since

DTN are intermittently connected networks, messages in these networks must be buffered for long periods of time. This means that the intermediate nodes require enough buffer space to store all messages until that intermediate nodes meet the specific destination nodes. The process of storing messages requires sufficient buffer space to store all pending messages as required.

Energy

Nodes in these networks normally have a low level of energy because of the mobility of nodes and the difficulties of connection to the power station. Much of energy is consumed during messages routing, as well as energy consumed for sending, receiving, storing, and computation of messages.

Reliability

Routing protocols in DTN should have some acknowledge for reliable delivery of data, which guarantee successful and stable delivery of information. Where some acknowledgment messages should be sent back when messages correctly reach to the final destination.

Security

Security has always been a significant problem for both traditional and DTN networks. The messages may go along arbitrary path through intermediate nodes before reaching their final destination. Therefore, based on the requirements of security of applications, users may require securing guarantees about the authenticity of a message. The cryptographic mechanisms may be useful to secure intermediate routing. To overcome the problem of intermittent connectivity and partitions in the networks, routing in DTN utilizes nodes mobility and messages buffering which makes it possible for a node to carry a message and bridge partitions in the networks.

Routing in DTN can be classified into different categories based on their characteristics as deterministic and stochastic. In deterministic scheme the network topology and/or its characteristics are assumed to be known. Contrarily, for stochastic case no exact knowledge of topology is assumed.

Processing Power

Processing Power is one of the goals of delay-tolerant networking is to connect devices that are not performed by traditional networks. These devices may be very small having small processing capability, in terms of CPU and memory. These nodes will not be capable of running complex routing protocols. The routing strategies presented here could still be used on more powerful gateway nodes, in order to connect the sensor network to a general purpose delay-tolerant network.

Classification of Routing Protocols

The existing routing protocols in DTNs are classified with respect to their strategies for controlling message copies and making the forwarding decision shown in Figure 4.

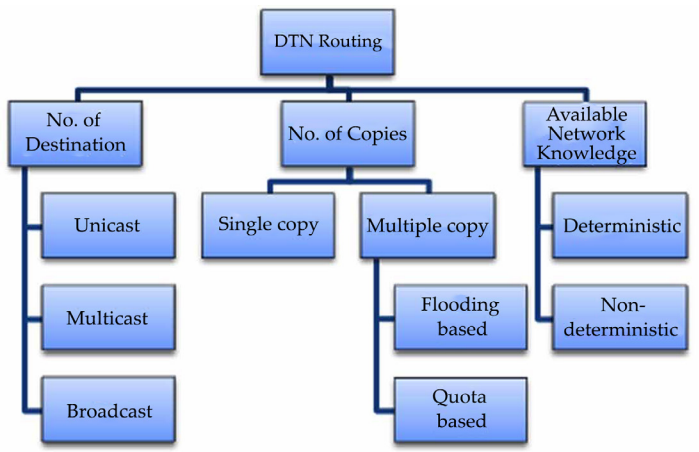


Figure 4. Classification of DTN Routing Protocols.

Number of Destination

According to the number of destination nodes of a message, routing protocols can be classified into three categories: unicast routing, multicast routing, and broadcast routing.

- Unicast Routing: Single destination for each message.
- Multicast Routing: Group of destination nodes for each message.
- Broadcast Routing: All the nodes in the network are destination nodes for each message.

Number of Copy

Depending on the number of message copies utilized in the routing process, protocols can be classified into two categories: single-copy and multiple-copy.

- Single-Copy Routing Protocols: Only a single copy for each message exists in the network at any time.
- Multiple-Copy Routing Protocols: Multiple copies of same message can be generated and distributed into the network. Moreover, multiple copy routing protocols can be further divided into flooding-based and quota based.
- Flooding-Based Routing Protocol: Dissemination copies of each message to as many nodes as possible.
- Quota-Based Routing Protocol: Intentionally limit the number of message copies.

Available Network Knowledge

In addition, according to whether the forwarding decision is based on the knowledge derived from the nodes' encounters or not, protocols can as well be classified into two categories: Deterministic and Nondeterministic (Opportunistic).

- Deterministic Routing Protocol: Complete knowledge of node trajectories, encounter probability of nodes and node meeting times and period to make the forwarding decision.
- Non-Deterministic Routing Protocols: Zero knowledge of predetermined path between source and destination. These algorithms either forward the messages randomly or prediction based (Probabilistic based).

DTN Routing Protocols

The routing protocols used in DTN are listed below:

Epidemic Protocol (EP): In this protocol all nodes can become the carrier, and it is ensured that messages can be delivered with a high probability. However, the network re-sources are consumed heavily. In other words, to deliver messages to the final destination, EP provides a redundant number of random messages exchange. This leads to guaranteeing the destination node receiving the messages in anyway.

Spray and Wait (SnW): The SnW algorithm is the advanced version of the epidemic routing protocol. In this algorithm the nodes are not distributing the message to each and every node but an optimal number of nodes (L) are selected to which the source node will relay the message. This algorithm consists of two phases spray phase and wait phase. In the spray phase, the source node replicates the message to the L-nodes and these L-nodes will further relay the message to L relay nodes. The relay nodes will store the message and perform direct transmission if the destination is not found in spray phase.

PROPHET: PROPHET is proposed in. The protocol estimates a node metric called delivery predictability, P (a, b), at each node a for each destination b. When two nodes meet, they update their delivery predictability toward each other. Then, the two nodes exchange their delivery predictability list toward other nodes update their delivery predictability.

MaxProp: MaxProp is a flooding-based routing protocol designed for vehicle-based delay tolerant networks. The buffer of this protocol is divided into two phases. First, messages are stored from low to high based on hop count information. Secondly, messages are arranged by cost from high to low. The first phase uses the front end of the buffer, while the second phase uses the back end of the buffer.

Routing in DTN is a big challenge because of frequency and length of the disconnection time between nodes in the network. However,

the main role of routing in DTN is to find an opportunity to connect nodes and to transmit data between them when the nodes meet each other if possible. In general, DTN routing protocols are designed to be as efficient as possible in cases of highly sparse networks and intermittent connectivity. Furthermore, an efficient routing protocol should be simple, scalable and capable of working at both low and high message load. Moreover, it should have optimal delivery probability, low delay and low overhead ratio.

4.2. DETERMINISTIC DELAY-TOLERANT ROUTING

In general, deterministic techniques are based on formulating models for time- dependent graphs and fi nding a space–time shortest path in DTNs by converting the routing problem to classic graph theory or by using optimization techniques for end-to-end delivery metrics. Deterministic routing techniques for networks with intermittent connectivity assume that local or global information on how the network topology evolves in time are available to a certain degree.

Good performance with less resource usage than stochastic routing techniques is provided by deterministic routing protocols using single-copy unicast for messages in transit. Deterministic routing mechanisms are appropriate only for scenarios where networks exhibit predictable topologies. Th is is true in applications where node trajectory is coordinated or can be predicted with accuracy, as in interplanetary networking.

4.2.1. Deterministic Delay-Tolerant Routing with Oracles

The distribution of network state and mobility information under sporadic connectivity, long delays, and sparse resources is a major problem facing deterministic routing protocols. Present a deterministic routing framework that takes advantage of increasing levels of information on topology and traffic demand (oracles) when such information is predictable. A DTN multigraph is defined where vertices represent the DTN nodes and edges

describe the time-varying link capacity between nodes. It is called a multigraph because multiple directed links between two nodes may exist.

One of the routing objectives is to minimize the end-to-end delay. Reducing the message transit times in the network also reduces contention for limited resources, such as buffer space and transmission time. Four knowledge oracles are defined: contacts summary oracle (for aggregate or summary contact statistics), contact oracle (for the time-varying contact multigraph), queuing oracle (for instantaneous queue state), and the traffic demand oracle (for present and future messages injected in the network). The authors adapt Dijkstra's shortest-path algorithm to support time-varying edge weights defined by the oracles available, and propose six algorithms for finding the optimal contact path.

Time-invariant edge weights is assumed in the first two algorithms. The First Contact (FC) algorithm is a zero-knowledge approach that chooses a random edge to forward a message among the currently available contacts. If no contact is available, the message will be forwarded on the first edge that comes up. The Minimum Expected Delay (MED) algorithm applies the Dijkstra algorithm where the edge weight is timeinvariant and is determined by the sum of the average waiting time (from the Contacts Summary oracle), propagation delay, and transmission delay. MED ignores congestion and does not recompute routes for messages in transit.

A time-varying edge cost, defined as the sum of the waiting, transmission, and propagation delays, is used in the following four proposed partial-knowledge algorithms. The waiting delay includes the time waiting for a contact and the queuing delay. The Earliest Delivery with Local Queuing algorithm (EDLQ) is equal to the local queue size at a particular node, and "0" for all other edges. EDLQ routes around congestion for the first hop and ignores queue occupancy at subsequent hops. Therefore, this algorithm must recompute the route at every hop. Cycles are avoided by using path vectors. Still, EDLQ is prone to message loss due to lack of available buffer space at reception.

The contacts oracle and the queuing oracle are used in the Earliest Delivery with All Queues (EDAQ) algorithm. EDAQ predicts the correct queue space for all edges at all times. In EDAQ, routes are not recomputed for messages in transit because the initial route accurately predicts all delays. EDAQ works only if capacity is reserved for each message along all contact edges. In practice, EDAQ is very difficult to implement in most DTNs with low connectivity, as it requires an accurate global distribution of queuing state. Limited connectivity also severely limits practical implementations of edge capacity reservations.

Simulation results indicate that algorithms that use the knowledge oracles (ED, EDLQ, and EDAQ) outperform the simpler MED and FC algorithms in terms of latency and delivery ratio. The more constrained the network resources are, the better the performance is for the algorithms that are more informed (i.e., use more oracles). A promising result is that routing with EDLQ (using only local queuing information) has a very similar performance to the EDAQ algorithm. This means that similar network performance can be achieved without expensive queue state dissemination and capacity reservations.

4.2.2. Deterministic Delay-Tolerant Routing with Space–Time Graphs

The trajectories and mission objectives of nodes may change. Therefore, in practice, contacts are deterministically predictable for only a finite time horizon. Merugu propose a deterministic routing framework where a space–time graph is built from predicted contact information. It starts with a time-varying link function defined as “1” when the link between two nodes is available and “0” otherwise. This function is defined as a function of time, where the time is discretized.

The space–time graph is built in multiple layers where the network nodes are replicated at each layer for each time unit t. Each layer has a copy of each network node. A column of these vertices maps to a single network node. A temporal link in the

space–time graph connects graph vertices from the same column at successive time intervals. When it is traversed, it indicates that the message is buffered. A spatial link connects two vertices from different columns, representing message forwarding. Forwarding delay is modeled by the number of layers traversed by a spatial link.

The objective of the least-cost routing in this DTN is to find the lowest cost (shortest) path from the source space–time node (column: layer) associated with the message arrival time to a vertex from the column corresponding to the destination DTN node. The end to-end latency for a message becomes equal to the length of the path traversed in the space–time graph. The routing problem is solved using the Floyd–Wars Hall all-pairs shortest paths algorithm, modified to account for the particular characteristics of the space–time graph. Multiple message sizes are supported by a path-coloring scheme.

4.2.3. Delay-Tolerant Routing with Link State

ASCoT, a dynamic routing mechanism for space networks and the Positional Link State routing protocol (PLS) to implement positionbased routing that enables the prediction of the trajectories of satellites and other space assets. Link state updates with predicted contacts and their link performances are disseminated in advance in the network through reliable flooding. Nodes execute a modified Dijkstra algorithm to recompute routing tables when link state updates are received.

The authors propose a data-centric approach similar to directed diffusion to support proximity routing for space assets in close formation. Note that in deterministic routing techniques using shortest-path algorithms, routing tables and forwarding schedules are recomputed whenever the contact graph state has changed, and selection of the next contact is done for a message at each hop along the path, as opposed to source routing. Thus, loops become possible because nodes may use outdated topology information. Cycles are avoided with path vectors.

For a limited range of applications, deterministic DTN routing protocols are effective where the contact schedule can be accurately modeled and predicted. Otherwise, it is necessary to frequently disseminate nodes' states throughout the network. In networks with constrained capacity or limited connectivity, this becomes very expensive and difficult to implement without an out-of-band broadcast channel. When contacts cannot be accurately predicted, routing must consider stochastic mechanisms that can only hint to predilection for future contacts based on historic information.

4.3. VEHICLE TRAFFIC MODEL

Vehicle traffic models are important for DTN routing in vehicle networks because the performance of DTN routing protocols are closely related to the mobility model of the network. The car-following model is used in civil engineering to describe traffic behavior on a single lane under both free-flow and congested traffic conditions. The is model assumes that each driver in the following vehicle maintains a safe distance from the leading vehicle and the deceleration factor is also taken into account for braking performance and drivers' behavior. The complete mathematical model is given by

$$S' = L + \beta' V + \gamma V^2$$

where S' is the headway spacing from rear bumper to rear bumper, L is the effective vehicle length in meters, and V is the vehicle speed in meters/second. β' is driver reaction time in seconds, and the γ coefficient is the reciprocal of twice the maximum average deceleration of a following vehicle. Both the β' parameter and the γ coefficient are introduced to ensure that the following vehicle can come to a complete stop if the leading vehicle suddenly brakes. As in many other civil engineering studies, we use a so-called "good driving" rule, which assumes that each vehicle has similar braking performance. In this case, the car following model can be simplified as

$$S' = L + \beta' V.$$

The car-following model has some limitations in modeling freeway traffic behavior for the purpose of wireless networking research, but is one of the most popular models in civil engineering. These limitations can be summarized as follows:

1. The car-following model is limited to the situation where driver reaction time is believed to be a dominant factor. Therefore, it is only an appropriate model under free-flow traffic or heavy traffic scenarios. Empirical studies confirm that during rush hour β' is typically a small number that represents the reaction time of a driver, following a log-normal distribution. However, in light to moderate traffic, β' can be as large as 50 to 100 sec and cannot be interpreted as driver reaction time. Instead, interarrival time between vehicles should be used to describe this spacing.
2. This is the focus of vehicular safety research in civil engineering. Therefore, the car-following model describes headway spacing between two adjacent vehicles of the same lane (i.e., lane-level spacing). From the network connectivity standpoint, however, we observe that the most relevant metric is spacing from the leading vehicle to the nearest following vehicle on a multilane road (i.e., road-level spacing), regardless of whether the following vehicle is on the same lane or on a different lane from the leading vehicle.

To address both of the aforementioned limitations, the car-following model is extended to the road level by replacing the lane-level reaction time β' with a road-level interarrival time β_b (the interarrival time of vehicles on any lane on the same road as observed from a fixed observation point). The lane-level car-following model can be generalized as

$$S = L_{\min} + \beta V$$

where L_{min} is the minimum spacing between any two adjacent vehicles, which is assumed to be zero in this study. By focusing on road-level intervehicle spacing S, the proposed model not only models rush-hour heavy traffic but also captures the sparse or intermediate traffic during nonrush hour times.

4.4. VEHICLE–ROADSIDE DATA ACCESS

Although a lot of research has been carried out on intervehicle communication, vehicle– roadside data access is also an important issue in vehicle DTN network. Medium access control (MAC) issues have been addressed, where slot-reservation MAC protocols and congestion control policies for emergency warning2 are studied.

Vehicle–roadside data access the roadside unit (RSU) can act as a router in a delay-tolerant network or as an access point for vehicles to access the Internet. Although this can bring many benefits to drivers, the deployment cost and maintenance cost are very high. As another option, RSU can also be used as a buffer point (or data island) between vehicles. This section focuses on the latter paradigm due to its low cost and easy deployment.

All data on the RSUs are uploaded or downloaded by vehicles in this paradigm. For example, some data, especially those with spacial/temporal constraints, only need to be stored and used locally. Applications that also belong to this case where the data is buffered at the RSUs and will not be sent to the Internet include the following:

1. Real-time traffic. Vehicles can observe real-time traffic observations and report them to nearby RSUs. The traffic data are stored at RSUs, providing real-time query and notication services to other vehicles. The data can be used to provide traffic conditions and alerts such as road congestion and accidents.
2. Value-added advertisement. To provide efficient advertisements, stores may want to advertise their sale or activity information in nearby area. Without Internet

connection, they can ask the running vehicles to carry and upload the advertisement information to nearby RSUs. At the same time, other vehicles driving around can download these advertisements and visit the stores.

3. Digital map downloading. It is impossible for vehicles to install all the most up-to date digital maps before traveling. This would help to solve the storage limitations of memory cards and changes resulting from frequent road construction. Hence, vehicles driving to a new area may update map data locally for travel guidance.

Vehicles are moving and they only stay in the RSU area for a short period of time. This makes vehicle networks different from traditional data access systems in which users can always wait for the service from the data server. As a result, there is always a time constraint associated with each request. Meanwhile, to make the best use of the RSU and to share the information with as many vehicles as possible, RSUs are oft en set at roadway intersections or areas with high traffic. In these areas, download (query) requests retrieve data from the RSU, and upload (update) requests upload data to the RSU. Both download and upload requests compete for the same limited bandwidth. As the number of users increases, deciding which request to serve at which time will be critical to system performance. Hence, it is important to design an efficient scheduling algorithm for vehicle–roadside data access.

4.4.1. A Model for Vehicle–Roadside Data Access

An architecture of vehicle–roadside service scheduling is shown in Figure 5, where a large number of vehicles retrieve (or upload) their data from (or to) the RSU when they are in communication range. The RSU (server) maintains a service cycle, which is non preemptive; that is, a service cannot be interrupted until it finishes. When one vehicle enters the RSU area, it listens to the wireless channel.

All vehicles can send requests to the RSU if they want to access the data. Each request is characterized by a 4-tuple: <v-id, d-id,

op, deadline> where v-id is the identifier of the vehicle, d-id is the identifier of the requested data item, op is the operation that the vehicle wants to do (upload or download), and deadline is the critical time constraint of the request, beyond which the service becomes useless.

All requests are queued at the RSU server upon arrival. Based on the scheduling algorithm, the server serves one request and removes it from the request queue. Unlike traditional scheduling services, data access in vehicular networks has two unique features:

1. The arrival request is only active for a short period of time due to vehicle movement and coverage limitations of RSUs. When vehicles move out of the RSU area, the unserved requests have to be dropped.
2. Data items can be downloaded and uploaded from the RSU server. The download and update requests compete for the service bandwidth.

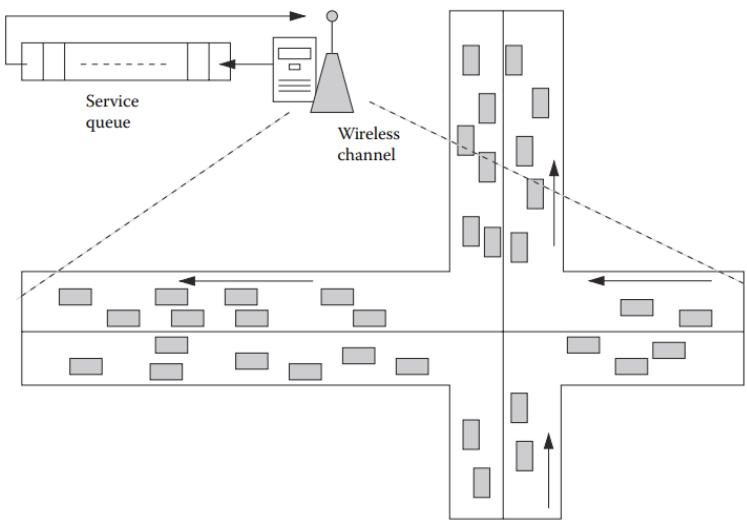


Figure 5. The architecture of vehicle–roadside service scheduling.

It is assumed that each vehicle knows the service deadline of its request. This is reasonable because when a vehicle with a GPS device enters the coverage area of a RSU, it can estimate its departure time based on the knowledge of its driving velocity and

its geographic position. A vehicle establishes connectivity with one RSU, it can get the geographic information and radio range of the RSU through beacon messages. With its own driving velocity and position information, the vehicle can estimate its departure time, which is its service deadline.

4.4.2. Performance Metrics

The metrics for scheduling algorithms are responsiveness (e.g., average/worst-case waiting time) or fairness (e.g., stretch) and are commonly used in works. In most of these works, requests do not have time constraints, and the data on the server is either not updated, or updated only by the server. However, in the vehicle–roadside data access scenario, requests that are not served within a set time limit will be dropped as the vehicles move out of the RSU area. As update requests compete for bandwidth with other download requests, some data may become stale after an update is missed, degrading service quality. Therefore, we use the following metrics for scheduling vehicle–roadside data access compared with responsiveness and fairness, providing fresh data to more vehicles.

1. Data quality. Good data quality means data is not stale. Data become stale if a vehicle has the new version of the data but fails to upload it before the vehicle moves out of the RSU range. The staleness of the data will degrade the data quality for the download service. The percentage of fresh data access to represent the data quality of the system. Therefore, a good scheduling scheme should update data in time and try to avoid data staleness.
2. Service ratio. A good scheduling scheme should serve as many requests as possible. The ratio of the number of requests served before the service deadline to the total number of arriving requests is the service ratio.

4.4.3. Roadside Unit Scheduling Schemes

Giving more bandwidth to download requests can provide a higher download service ratio, but a higher update drop ratio and hence low data quality. Therefore, achieving both high service ratio and good data quality is very difficult. If update requests get more bandwidth, the service ratio decreases.

There is always a trade-off between high service ratio and good data quality. Our focus now switches to improving the service ratio. The primary goal of a scheduling scheme is to serve as many requests as possible. We identify two parameters that can be used for scheduling vehicle–roadside data access:

1. Deadline. The request is not useful and should be dropped if a request cannot be served before its deadline.
2. Data Size. Usually, vehicles can communicate with the RSU at the same data transmission rate. The data size decides how long the service will last.

Three naive schemes for roadside unit scheduling are as follows:

1. First Deadline First (FDF). In this scheme, the request with the most urgency will be served first.
2. Smallest Data Size First (SDF). In this scheme, the data with a small size will be served first.
3. First Come First Serve (FCFS). In this scheme, the request with the earliest arrival time will be served first.

The service ratios under these three naive scheduling schemes are compared in Figure 6. The interarrival time of the requests is determined by the percentage of vehicles that will issue service requests, which is varied along the x-axis. As shown in the figure, when the request arrival rate is low, FDF outperforms FCFS and SDF. This is because, when the workload is low, the deadline factor has more impact on the performance.

After the urgent requests are served, other pending requests can still have the opportunity to get services. However, when the request arrival rate increases, the service ratio of FDF drops quickly while SDF performs relatively better. Because the system can always find short requests for service, SDF can still keep a

higher service ratio. FCFS does not take any deadline or data size factors into account when making scheduling decisions. It has the worst performance.

Data size and request deadlines are not considered in FCFS. FDF gives the highest priority to the most urgent requests while neglecting the service time spent on those data items. SDF takes the data size into account but ignores the request urgency. It is clearly shown in the figure that FDF and SDF can only achieve good performance for certain workloads.

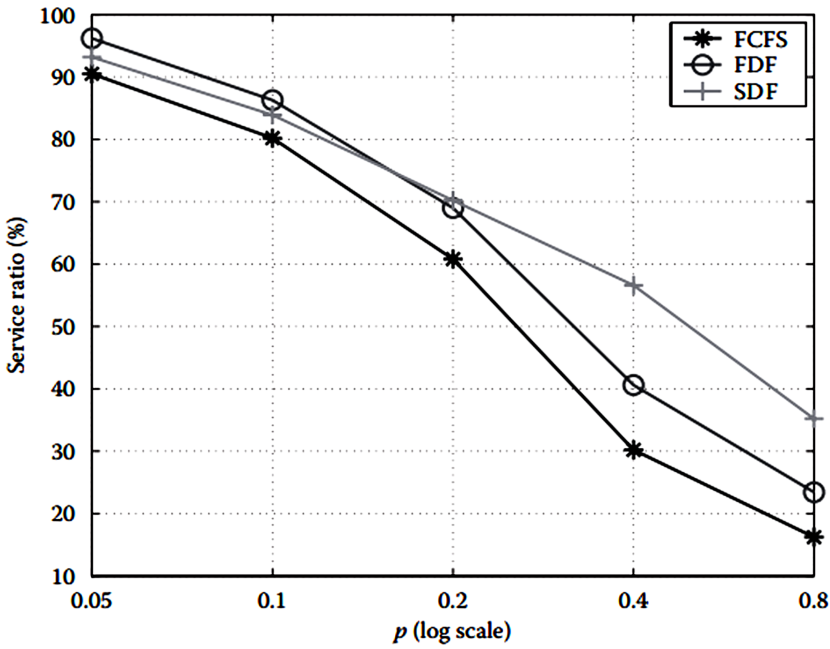


Figure 6. Service ratio for FCFS, FDF, and SDF schemes.

This motivates the integration of the deadline and data size to improve the performance of scheduling. None of them can provide a good scheduling as a result. D* S^{30} considers both data size and deadlines when scheduling vehicle–roadside data access. From the above observations, there are two principles are:

1. Given two requests with the same deadline, the one asking for a small size of data should be served first.

2. Given two requests asking for data with same size, the one with the earlier deadline should be served first.

Each request is given a service value based on its deadline and data size, called DS_value, as its service priority weight:

DS_value = (Deadline - CurrentClock) *DataSize

In this equation, the deadline and data size factors are multiplied because these two factors have different measurement scales and/or units. With product, different metrologies will not impose any negative effect on the comparison of two DS_values. At each scheduling time, the D* S scheme always serves the requests with the minimum DS_value.

4.5. DELAY-TOLERANT ROUTING IN VANETS

Although most of the existing work on vehicle networks is limited to 1-hop or short-range multi hop communication, vehicular delay-tolerant networks are useful to other scenarios. For example, without Internet connection, a moving vehicle may want to query a data center ten miles away through a VANET. The widely deployed wireless LANs or infostations can also be considered.

Vehicle delay-tolerant networks have many applications, such as delivering advertisements and announcements regarding sale information or remaining stocks at a department store. Information such as the available parking spaces in a parking lot, the meeting schedule at a conference room, and the estimated bus arrival time at a bus stop can also be delivered by vehicle delay-tolerant networks.

For the limited transmission range, only clients around the access point can directly receive the data. However, this data may be beneficial to people in moving vehicles far away, as people driving may want to query several department stores to decide where to go. A driver may query the traffic cameras or parking lot information to make a better travel plan. A passenger on a bus may

query several bus stops to choose the best stop for bus transfer. All these queries may be issued miles away from the broadcast site. With a vehicular delay-tolerant network, the requester can send the query to the broadcast site and get a reply from it. In these applications, the users can tolerate up to a minute of delay as long as the reply eventually returns.

The problem of efficient data delivery in vehicular delay-tolerant networks is studied in this section. Specifically, when a vehicle issues a delay-tolerant data query to some fixed site, we must know how to efficiently route the packet to that site and receive the reply with a reasonable delay. We will present a vehicle-assisted data delivery (VADD) based on the idea of carry and forward.

Some of the carry-and-forwarding approaches either pose too much control or no control at all on mobility or hence are not suitable for vehicular networks. They include the ones proposed for delay-tolerant network. In contrast, VADD makes use of predictable vehicle mobility, which is limited by the traffic pattern and road layout. For example, the driving speed is regulated by the speed limit and the traffic density of the road, the driving direction is predictable based on the road pattern, and the acceleration is bounded by the engine speed. VADD exploits the vehicle mobility pattern to better assist data delivery.

4.5.1. The VADD Protocol

In the model assumed by the VADD protocol, vehicles communicate with each other through a short-range wireless channel, and vehicles can find their neighbors through beacon messages. The packet delivery information such as source ID, source location, packet generation time, destination location, expiration time, and so on, are specified by the data source and placed in the packet header. A vehicle knows its location by triangulation or through a GPS device, which is already popular in new cars and will be common in the future.

Geographical information is also assumed to be available in the vehicles. Vehicles are equipped with preloaded digital maps, which provide street-level maps and traffic statistics such as traffic density and vehicle speed on roads at different times of the day. Such digital maps have already been commercialized. The latest one is developed by Map Mechanics, and includes road speed data and an indication of the relative density of vehicles on each road. Yahoo! is also working on integrating traffic statistics in its new product called Smart View, where real traffic reports of major U.S. cities are available.

It is expected that more detailed traffic statistics will be integrated into digital maps in the near future. The cost of setting up such a vehicular network can be justified by its application to many road safety and commercial applications, which are not limited to the proposed delay-tolerant data-delivery applications.

The most important issue is to select a forwarding path with the smallest packet delivery delay. VADD is based on the idea of carry and forward. Although geographical forwarding approaches such as GPSR, which always chooses the next hop closer to the destination, are very efficient for data delivery in ad hoc networks, they may not be suitable for sparsely connected vehicular networks.

Suppose a driver approaches intersection I_a and he wants to send a request to the coffee shop (to reserve a sandwich) at the corner of intersection I_b, as shown in Figure 7. To forward the request through $I_a \rightarrow I_c, I_c \rightarrow I_d, I_d \rightarrow I_b$ would be faster than forwarding through $I_a \rightarrow I_b$, even though the latter provides a geographically shortest-possible path. The reason is that, in the case of disconnection, the packet has to be carried by the vehicle, whose moving speed is significantly slower than the wireless communication. In sparsely connected networks, vehicles should try to make use of the wireless communication channel, and resort to vehicles with faster speed. Thus, VADD follows the following basic principles:

1. If the packet has to be carried through certain roads, the road with higher speed should be chosen.

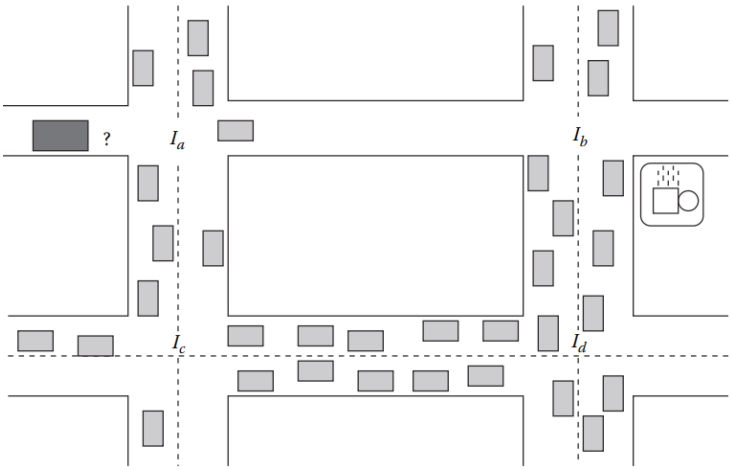


Figure 7. Find a path to the coffee shop.

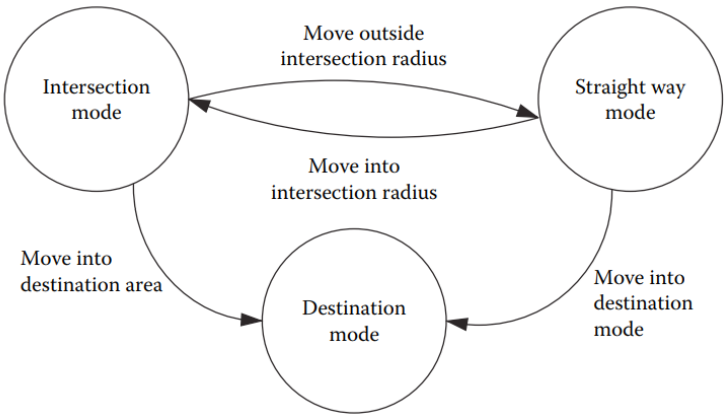


Figure 8. The transition mode in VADD.

2. Transmit through wireless channels as much as possible.
3. Owing to the unpredictable nature of VANETs, the packet cannot be expected to be successfully routed along the precomputed optimal path, so dynamic path selection should continuously be executed throughout the packet-forwarding process.

VADD has three packet modes (Figure 8): Intersection, Straight Way, and Destination, based on the location of the packet carrier

(i.e., the vehicle that carries the packet.) By switching between these packet modes, the packet carrier takes the best packet-forwarding path. Among the three modes, the Intersection mode is the most critical and complicated one, because vehicles have more choices at the intersection.

4.6. DATA DISSEMINATION IN VANETS

Data dissemination protocols have been proposed to disseminate information about traffic, obstacles, and hazards on the roads. Similar applications such as real-time video streaming between vehicles have been studied. A conventional way to report accidents or traffic conditions is to use certain infrastructures such as roadside traffic sensors reporting data to a central database, or cellular wireless communication between vehicles and a monitoring center. The problem with this design is the expensive deployment. In addition, these infrastructure-based networks are not scalable due to their centralized nature. VANETs, as an alternative to infrastructure-based vehicle networks, are constructed on-the-fly and do not require any investment besides the wireless network interfaces that will be a standard feature in the next generation of vehicles.

How to exchange traffic information among vehicles in a scalable fashion in VANETs is an interesting but challenging problem that has to be solved. Solutions to this problem can be categorized into two main mechanisms: a flooding-based approach and a dissemination-based approach. In the flooding mechanism, each individual vehicle periodically broadcasts information about itself. Every time a vehicle receives a broadcast message, it stores it and immediately forwards it by rebroadcasting the message. This mechanism is clearly not scalable due to the large volume of messages flooded over the network, especially in high-traffic-density scenarios. The flooding-based mechanism can be further divided into three categories: priority-based approaches, distance-based approaches, and geocast approaches. On the other hand, in the dissemination mechanism, each vehicle broadcasts

information about itself and the other vehicles it knows about. Each time a vehicle receives information broadcasted by another vehicle, it updates its stored information to the next broadcast period, at which time it broadcasts its updated information. The dissemination mechanism is scalable, because the number of broadcast messages is limited, and they do not flood the network. The dissemination-based mechanism can be further divided into two categories: approaches utilizing the bidirectional mobility of vehicles and forwarding-based approaches. We can see the classification of mechanisms of multicast/broadcast in VANETs in Figure 9.

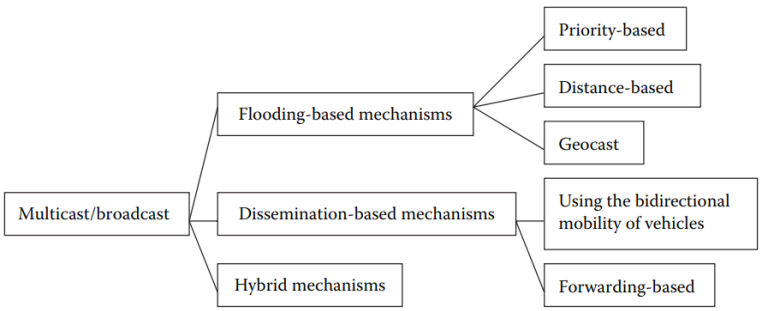


Figure 9. The classification of multicast/broadcast mechanisms.

4.6.1. Flooding-Based Mechanisms

A number of safety applications require communications to a group of vehicles, not just pairwise communications supported by unicast protocols. Safety applications require propagation of information to a large number of nodes quickly and reliably. Flooding is the most common approach for broadcasting without explicit neighbor information. However, flooding is known to be inefficient due to the so-called broadcast storm problem, especially in networks with high node density. Most existing flooding-based information dissemination approaches in VANETs aim to achieve a high message delivery ratio by avoiding contention and collision caused by the broadcast storm phenomena.

4.6.2. Dissemination-Based Mechanisms

Compared to the flooding-based approaches, dissemination-based mechanisms are more scalable because the number of broadcast messages is limited, and they do not flood the network. The dissemination mechanism can either broadcast information to vehicles in all directions, or perform a directed broadcast restricting information about a vehicle to vehicles behind it.

4.6.3. Hybrid Mechanisms

Flooding-based data dissemination mechanisms are unscalable due to the large amount of contention and collision, especially in dense networks. On the other hand, dissemination-based mechanisms are not suited for delay-sensitive safety message dissemination, albeit scalability is achieved. Hence, hybrid mechanisms that combine the strengths of each are proposed. Reference 55 proposes an approach (called Directional Propagation Protocol, or simply, DPP) using clusters of connected vehicles where flooding-based data dissemination mechanisms are used in a cluster and dissemination-based mechanisms are used among clusters.

DPP uses the directionality of data and vehicles for information propagation. DPP comprises three components: a Custody Transfer Protocol (CTP), an Inter-Cluster Routing Protocol, and an Intra-Cluster Routing Protocol.

In order to overcome the lack of an end-to-end path between source and destination, the Custody Transfer Protocol is introduced which is derived from delay-tolerant networking concepts. On the one hand, the Inter-Cluster Routing Protocol controls the message exchange between nodes within a cluster. On the other hand, the communication between clusters is governed by the Intra-Cluster Routing Protocol. As illustrated in Figure 10, interconnected blocks of vehicles can be formed by vehicles traveling towards the same direction. Gaps are allowed between consecutive blocks. The traffic density has a significant impact on the cardinality of each block. For example, a long continuous block can be formed under

dense conditions, while under sparse conditions, the cardinality of each block could be one.

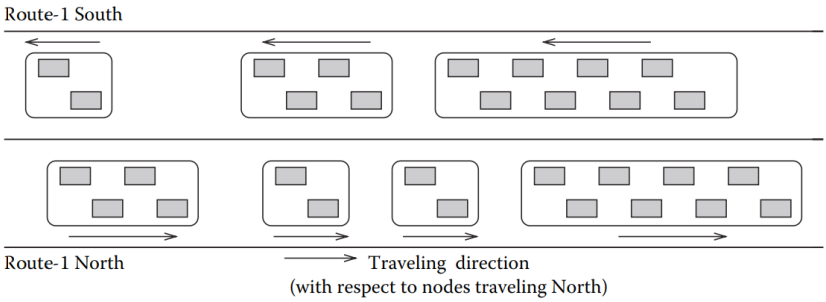


Figure 10. An example of blocks of vehicles.

Additionally, vehicles that are within range R and maintain connectivity for a minimum time t are said to be part of a cluster. Thus, a block may comprise several clusters.

Under sparse traffic conditions, gaps between blocks are frequent and network partitions are common, which prevents an end-to-end path between source and destination. Accordingly, the speed of the vehicle that carries the message may influence the data dissemination performance. Under dense traffic conditions, an end-to-end path between source and destination exists with high probability where the data dissemination performance is mainly determined by contentions and collisions.

The effects of speed differentials within the cluster are not considered as the faster vehicles will leave one cluster and join another as they progress on the road. Also, there are intersections on a highway where vehicles may join or leave the clusters. Once a cluster becomes very large, the cluster is split to better manage intracluster traffic.

Each cluster has a header and a trailer, located at the front and rear of each cluster, entrusted with the task of communicating with other clusters. A node at the head or tail of the cluster will elect itself as the header or trailer for our protocol. (Node election is not covered here.) This limits congestion caused by the large number

of participating nodes. The remaining nodes in the cluster, nodes that are not header or trailer, are described as intermediate nodes. Within a cluster, communicated messages are shared with all nodes to both facilitate header/trailer replacement and general awareness of disseminated messages.

The intermediate nodes retain a passive role of receiving messages and acknowledgments from opposing blocks and forwarding them to the header or trailer sharing the information within the cluster. Similarly, messages originating from intermediate nodes are immediately routed to header or trailer depending upon the direction in which information needs to propagate. Any duplicate messages received at any of the nodes are dropped. End-to-end path formation can be assumed to be taking place within a cluster.

In most message-passing schemes, a message is buffered until an acknowledgment from the destination is received. However, due to network fragmentation in a VANET and the resultant lack of continuous end-to-end connectivity at any given instant, the message can require buffering for an indeterminate amount of time. The result translates to the requirement for large buffer sizes or dropped messages and difficulty in exchanging acknowledgments. For applications that do not require continuous end-to-end connectivity, a store-and-forward approach can be used.

With the custody transfer mechanism, a message is buffered for retransmission from the originating cluster until it receives an acknowledgment from the next-hop cluster. The custody is implicitly transferred to another cluster that is in front along the direction of propagation and is logically the next hop in terms of the message path. The traffic in the opposing direction acts as a bridge but is never given custody of the message. The custody is not released until an acknowledgment is received from the cluster in front. Once the message reaches the next-hop cluster, it has custody of the message and the responsibility for further relaying the message is vested with this cluster. The custody of the message may be accepted or denied by a cluster by virtue of it being unable to satisfy the requirements of the message.

The propagation is called reverse propagation if the data are headed in a direction opposite to the direction of motion of the vehicles and forward propagation if data are headed along the direction of motion of the vehicles. In forward propagation, as illustrated in Figure 11, the vehicle is assumed to be traveling along the N direction and the data are also to be propagated in the N direction. The data can travel at a minimum rate of the speed of the vehicle because the data are traveling along with the vehicle. The data are propagated to the header of the cluster. The header now tries to propagate the data further along the N direction, trying to communicate with other clusters located ahead of this cluster. If the clusters are partitioned, the header attempts to use the clusters along the S direction, which may overlap with other clusters along the N direction to bridge this partition. Thus, the data are propagated to nodes traveling along the N direction that are otherwise partitioned from each other, by using clusters along the S direction. This temporary path occurs due to opportunistic contact with nodes in the overlapping clusters. Once the data are forwarded to the next hop and an acknowledgment is received, the custody is transferred to that cluster. The entire process is repeated until the data reaches its required destination. The reverse propagation scheme can be modeled as an extension of the forward propagation scheme.

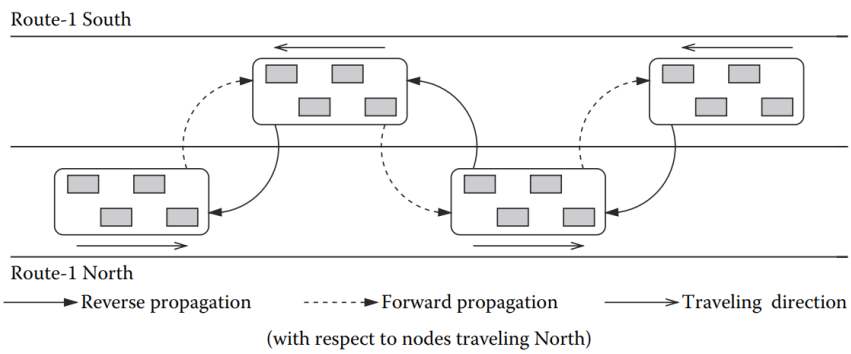
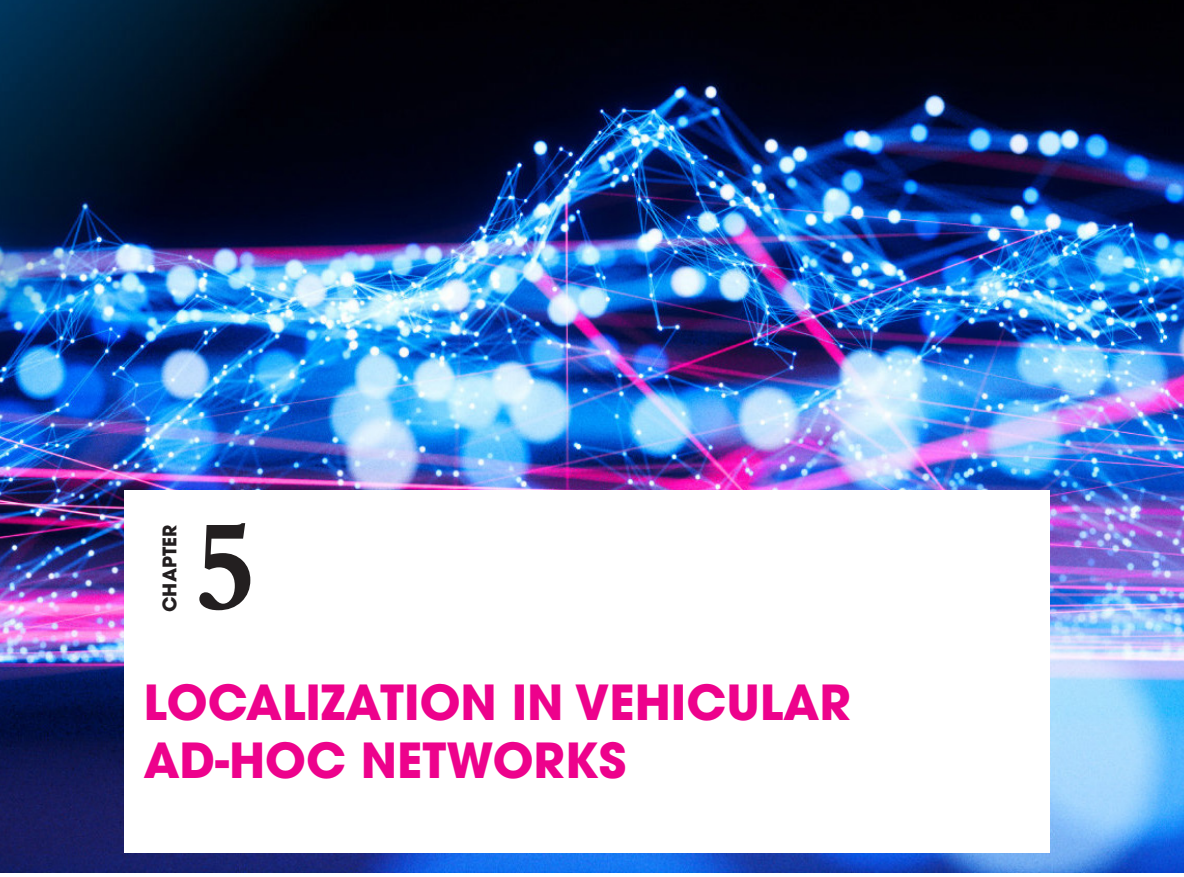


Figure 11. An example of forward propagation and reverse propagation.

REFERENCES

1. Burns, B., Brock, O., and Levine, B.N., Mv Routing and Capacity Building in Disruption Tolerant Networks, in Proc. IEEE INFOCOM, 2005.
2. Cardei, I., Liu, C., and Wu, J., Routing in Wireless Networks with Intermittent Connectivity, in Encyclopedia of Wireless and Mobile Communications, CRC Press, Taylor & Francis Group, 2007.
3. Gnawali, O., Polyakov, M., Bose, P., and Govindan, R., Data Centric, in Proc. Of IEEE Aerospace Conference, 2005.
4. Intanagonwiwat, C., Govindan, R., and Estrin, D., Directed Diff usion: A Scalable and Robust Communication Paradigm for Sensor Networks, in Proc. of ACM MOBICOM, 2000.
5. Leguay, J., Friedman, T., and Conan, V., Evaluating Mobility Pattern Space Routing for DTNs, in Proc. of IEEE INFOCOM, 2006.
6. Lindgren, A., Doria, A., and SchelSn, O., Probabilistic Routing in Intermittently Connected Networks, in Poster of ACM MOBIHOC, 2003.
7. Lott, M., Halmann, R., Schulz, E., and Radimirsch, M., Medium Access and Radio Resource Management for Ad Hoc Networks Based on UTRA TDD, in Poster of ACM MOBIHOC, 2001.
8. Merugu, S., Ammar, M., and Zegura, E., Routing in Space and Time in Networks with Predictable Mobility, in GIT-CC-04-7, 2004.
9. Namboodiri, V., Agarwal, M., and Gao, L., A Study on the Feasibility of Mobile Gateways for Vehicular Ad-Hoc Networks, in Proc. of ACM VANET, 2004.
10. Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S., A Scalable Content Addressable Network, in Proc. of ACM SIGCOMM, 2001.
11. Vahdat, A. and Becker, D., Epidemic Routing for Partially Connected Ad Hoc Networks, April 2000.

12. Wisitpongphan, N., Bai, F., Mudalige, P., Sadekar, V., and Tonguz, O.K., On the Routing Problem in Disconnected Vehicular Networks, in Proc. of IEEE INFOCOM Minisymposia, 2007.
13. Xu, Q., Mark, T., Ko, J., and Sengupta, R., Vehicle-to-Vehicle Safety Messaging in DSRC, in Proc. of ACM VANET, 2004.
14. Yang, X., Liu, J., Zhao, F., and Vaidya, N., A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning, in Proc. of ACM MOBIQUITOUS, August 2004.
15. Yin, J. et al., Performance Evaluation of Safety Applications Over DSRC Vehicular Ad Hoc Networks, in Proc. of ACM VANET, 2004.
16. Zhao, W., Ammar, M., and Zegura, E., Controlling the Mobility of Multiple Data Transport Ferries in a Delay-Tolerant Network, in Proc. of IEEE INFOCOM, 2005.

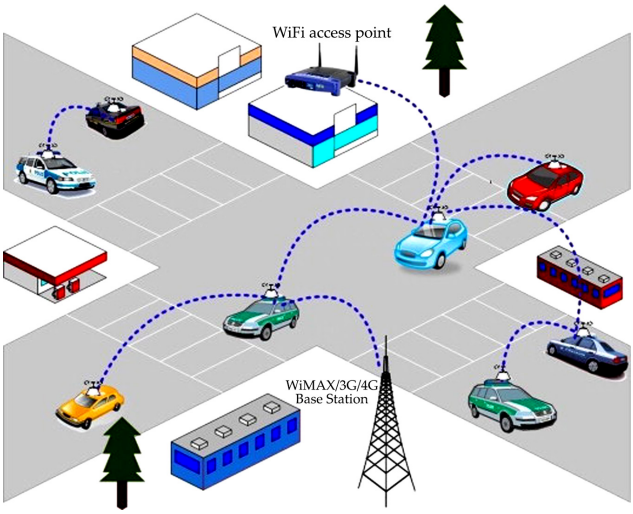


CHAPTER 5

LOCALIZATION IN VEHICULAR AD-HOC NETWORKS

INTRODUCTION

Localization systems play a major role in many applications for vehicular ad hoc networks (VANETs). One of the most interesting problems to be solved in vehicular networks is how to provide anywhere and anytime highly accurate and reliable localization information. Unique characteristics of VANETs such as mobility constraints, driver's behavior, and the high speed displacement nature of vehicles cause rapid and constant changes in network topology, leading to dissemination of outdated localization information. To circumvent this problem, an alternative is the use of predicted future locations of vehicles.



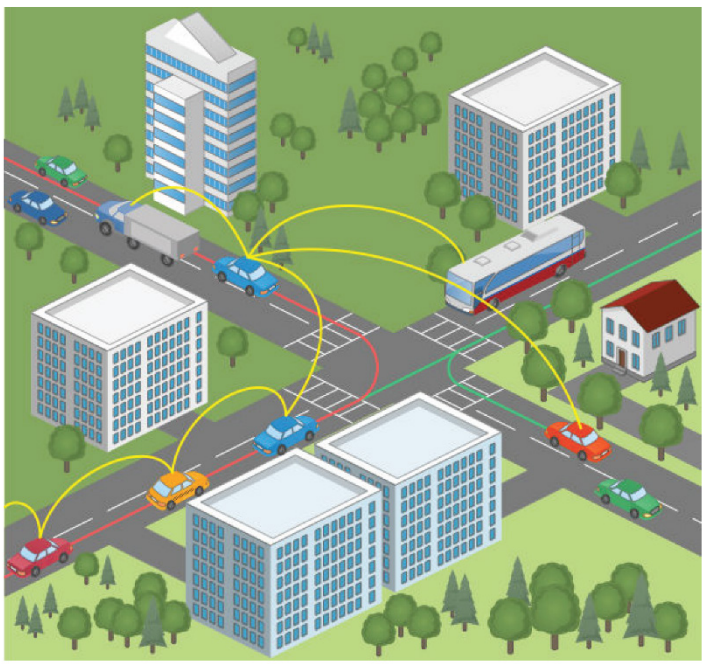
The main idea of this approach is to use the localization prediction as an extension of a data fusion localization system. In such an approach, a future position of a vehicle is predicted for a given future time and used to take advantage of a future time-space window of a vectorial trajectory rather than a static localization point. We discuss this subject by studying and analyzing the use of localization prediction as natural way to improve VANET applications. We survey proposed approaches for localization, target tracking, and time series prediction techniques that can be used to estimate the future position of a vehicle. We also highlight their advantages and disadvantages through an analytical discussion visualizing its potential application scenarios in VANETs.

5.1 LOCALIZATION EFFECT IN VEHICULAR AD-HOC NETWORKS

Global Positioning System (GPS) has been widely used for positioning service. Although, it is possible to find the position of each vehicle in a VANET network with the aid of Global Positioning System (GPS) installed in all vehicles. Equipment such as GPS receivers and digital compasses provide good positioning

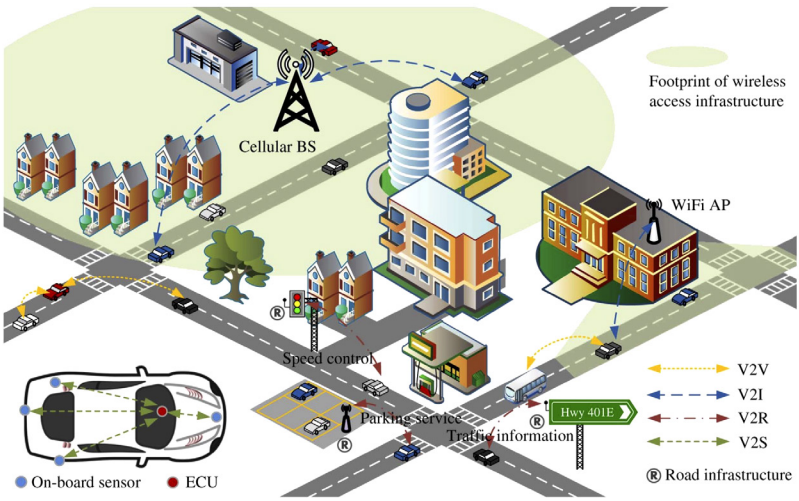
and orientation outdoors, there are many applications requiring the same services indoors, where line of sight access to satellites is unavailable, or earth magnetic readings are unavailable.

To be located compared to an accident when a vehicle is informed of the existence of an accident or an imminent danger is a task of great importance. That can avoid pile-up of vehicles and avoid the loss of human life. All remember fire started in the tunnel of the Mont-Blanc between France and Italy. If vehicles were equipped with communication and localization means of new technology with automatic release the dramas could have been avoided. Thus, the localization is very simple when the vehicle is equipped GPS. Finding position without the aid of GPS in each vehicle is important in cases where GPS is either not accessible, or not practical to use due to form factor or line of sight conditions such as when vehicles are in tunnels.



Recently, the democratisation of GPS technology and the progress in mobile ad hoc networking have led to the appearance of new inter vehicle communication protocols. Based on the use of GPS devices, these protocols have been mainly designed for safety

driving by the dissemination of urgent information, called alarm messages, in the case of accidents, fogs, etc, among the vehicles. The proposed solution called RBM Role Based Multicast was designed to overcome fragmentation in the ad hoc network composed by the vehicles and to reduce the number of redundant broadcasts of alarm messages. However, since RBM is based on neighbor detection it consumes more bandwidth, presents longer delays, and more packet loss. Moreover, RBM does not always overcome fragmentation, and sometimes the alarm message arrives out of date to the destination vehicles. In TRADE, each vehicle wanting to disseminate an alarm message has to determine positions and driving directions of its neighbors. Thus, each vehicle has to designate the vehicle that has to rebroadcast after it in order to avoid the broadcasts redundancy. However, since TRADE is based on neighbor maintenance, it consumes more bandwidth and presents longer delays. This limits the scalability of TRADE in dense networks.



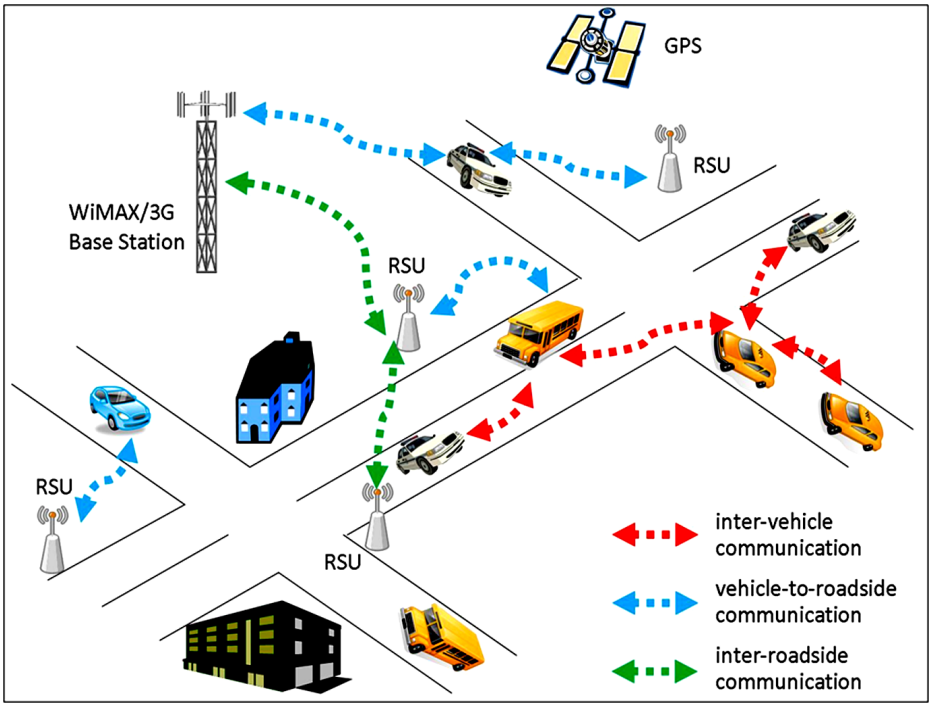
In order to cope to problems resulting from neighbors maintenance, DDT was proposed. DDT does not rely on neighbors maintenance, but inserts distance-based defer time slots for each rebroadcast alarm message. When a vehicle executing DDT receives an alarm message, it sets-up a timer in order to determine if it is useful to rebroadcast that message. However, DDT does not take into

account multiple alarm messages broadcasts by the same vehicle. Consequently, it can't overcome fragmentation in the ad hoc network composed temporally by vehicles in highways. The lack of supporting fragmented ad hoc networks makes DDT unsuitable for light-crowded highways. So it can't be used in emergency situations. Simulation results show that even for a transmission range up to 2000 m the reliability of DDT is not perfect in rural areas where the vehicle penetration rate is low. Since all proposed algorithms: ODAM, DDT, RBM and TRADE are based on geographical positioning system (i.e. GPS). The solution is based on cooperation between GPS-E vehicles in order to help GPS-U vehicles to get their positions. Although the knowledge of the exact position is not always possible, the GPS-U vehicle can obtain some useful information such as driving direction and distance from the accident.

5.1.1 Overview of Radiolocation Methods

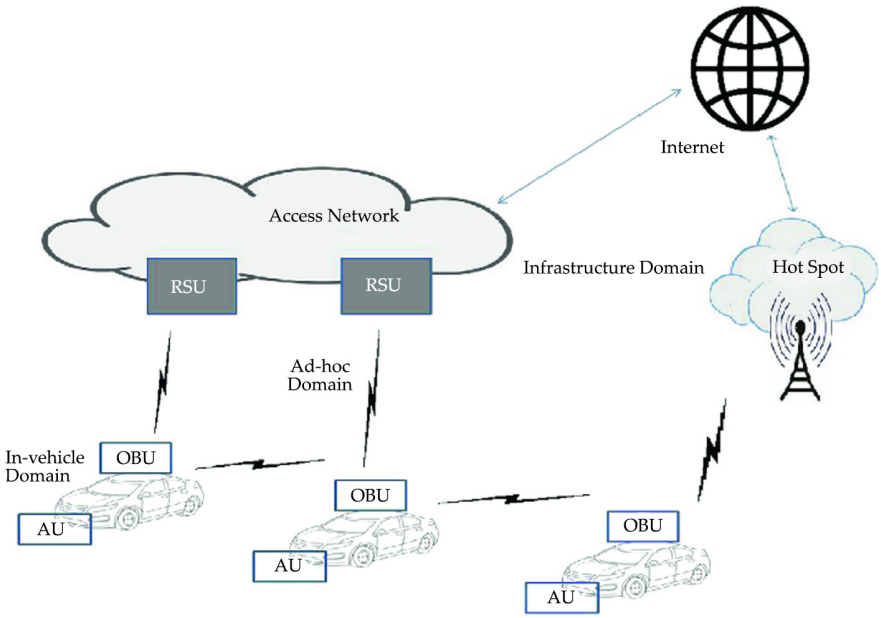
There exist several ranging techniques. Most generally used ones are based on signal propagation duration. Several Radio Location systems have been proposed for locating the Mobiles Stations (MS) in cellular systems. The main three different measurement methods are used today: Received-Signal-Strength Indicator (RSSI), Time based measurements which include Time-of-arrival (TOA) and Time-Differenceof-arrival (TDOA) and Angle-of-arrival (AOA).

RSSI measures the power of the signal at the receiver. Based on the known transmit power, the effective propagation loss can be computed. Since a measurement of signal strength provides a distance estimate between the MS and the BS (Base Stations), the MS must lie on a circle centered at the BS. $P_r = P_t (c_1/d)^n . c_2$, where P_t is the power level on which the message is sent, P_r is the power level on which the message is received and n, c_1, c_2 are constants related to physical environment, the carrier's wavelength and antenna gains, respectively. Since P_r and P_t can be measured, the distance d can be estimated from this formula.



Time based methods record the time of arrival (TOA) or time-difference of arrival (TDOA). The propagation time can be directly translated into distance, based on known signal propagation speed. The distance from the mobile target to the receiver is directly proportional to the propagation time. If the signal propagates in time t from the target transmitter to the receiver, then the receiver is at the range R given by R=c.t where c is the speed of light. In general, direct TOA requires synchronization and timestamp transport. So, TDOA allows to determine relative position, rather than the absolute arrival, of the transmitter by examining the difference in time at which the signal arrives at multiple receivers.

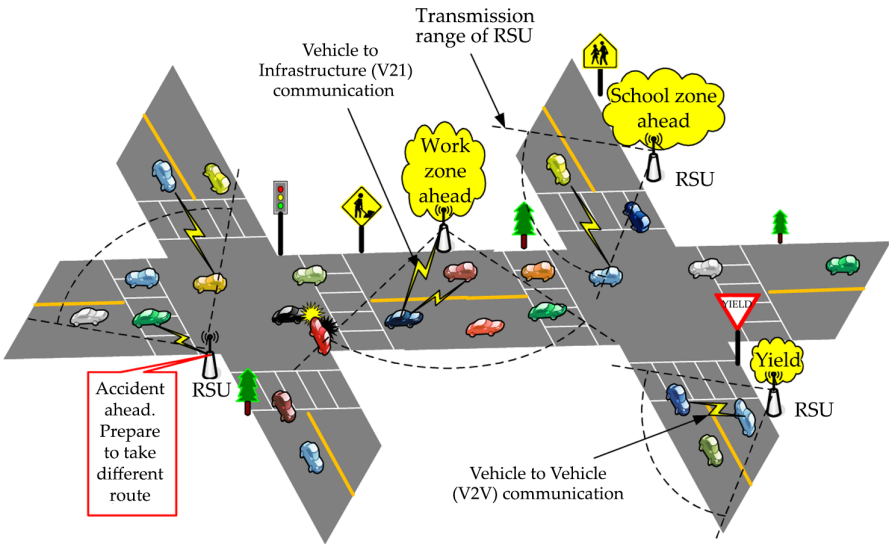
AOA techniques estimate the desired target by measuring the angle at which signals from several BSs through the use of directive antennas or antenna arrays. These radio location techniques are corrupted by many sources of errors such as multipath fading and shadowing, the non-line of sight (NLOS) propagation, and the multiple access interference. Thus, several researches have been leading to detect and correct the NLOS propagation.



Recently, a new algorithm Self-Positioning Algorithm (SPA) has been proposed for positioning mobile nodes in wireless ad hoc networks without relying on GPS. The algorithm uses the distance between the nodes to build a relative coordinate system in which the node positions are computed in two dimensions. The time of arrival TOA method is used to obtain the range between two mobile devices. The algorithm provides enough stability and location accuracy despite the range measurement errors and the motion of the nodes. However, a node executing SPA algorithm fails in some situations to computes its coordinates since it is based on some conditions on its neighbors (Local View Set). Moreover, while SPA is based on neighbor's maintenance by the exchange of hello messages between neighbors, it is expected that it present more bandwidth consummation, longer delays and more packet loss leading to inaccurate information in the nodes. Consequently, SPA is not well suited for inter vehicle communication. For example, in the case of an accident, vehicles without GPS have to be informed in the right moment. The algorithm should be lightweight and give to the vehicle enough accurate information about the accident.

5.1.2 ODAM Overview

ODAM is mainly designed for effective alarm message dissemination in the ad hoc network of vehicles in a highway or a road. ODAM is based on geographical multicast, which consists in determining the multicast group according to the driving direction and the positioning of the vehicles. The multicast is restrained to the so-called risk areas. First, broken vehicle (or accident, vehicle x in figure 1) begins to broadcast an alarm message to inform the other vehicles of the situation. Since the accident vehicle can just inform its one-hop neighbors, some other vehicles have to rebroadcast the alarm message to inform the vehicles located at more than one hop from the accident. The vehicle that performs the rebroadcast is called relay. Relays in ODAM are designated in fully distributed manner. The way with which a node is designated as relay is based on distance defer time algorithm.



The node that receives an alarm message does not rebroadcast it immediately but has to wait some time to take a decision about rebroadcast. When the defer time expires, if it does not receive the same alarm message from another node behind it, it deduces that there is no relay node behind it. Thus it has to designate itself as a relay and starts to broadcast the alarm messages in order to inform

the vehicles which could be behind it. When vehicle (x) broadcasts an alarm message for the first time, vehicles (a) and (b) can receive it because they are in the transmission range of (x). So, the defer time of (a) must be greater than the one of (b) because the distance from (x) to (a) is less than the one to (b). The defer time must be inversely proportional to the distance separating the receiver to the sender in order to favorite the farthest node to wait less time and to rebroadcast faster. To see the importance of distance defer time, remark that if node (a) were taken as relay (Figure 1.), node (c) could not be reached because it was out of the transmission range of node (a). Against, if node (b) were selected as relay, node (c) would have been reached and informed via (b) rebroadcast.

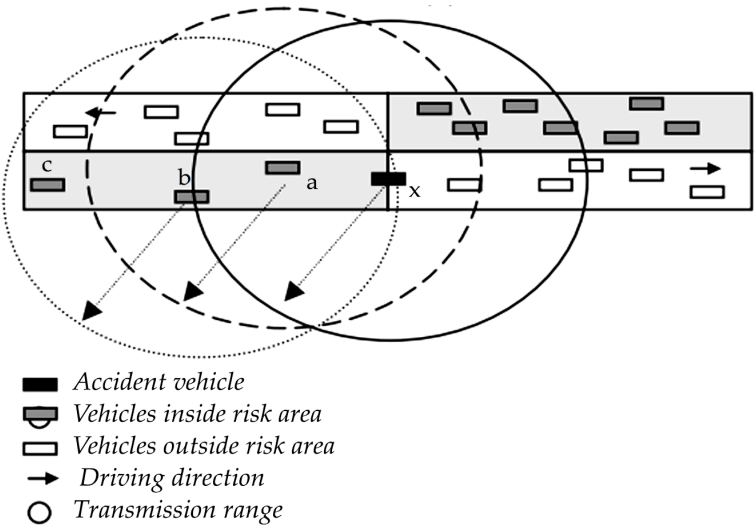


Figure 1. Road areas and transmission range

The alarm message must contain some information such as accident position, previous and current positions of the relay from which the message is received. This information is used by the vehicle that received the alarm message in order to determine its location according the accident vehicle. The message is relevant if the vehicle is located in a relevant area and it is received for the first time. When a vehicle receives the same alarm message before it's defer timer expires, it concludes that there is another vehicle behind it which is broadcasting the same alarm message. The

process of message dissemination with ODAM depends on the rate of vehicles equipped GPS in the road. We believe that the success of ODAM depends on its performances with GPS-unequipped vehicles. We propose a solution that allows the well-functioning of ODAM even with GPS-unequipped vehicles. The performances of that solution depend on the rate of GPS-unequipped vehicles and on the density of vehicle in the highway.

5.1.3 GPS-unequipped algorithm

Since each vehicle executing ODAM relies on the periodic computation of its driving direction, previous and current positions, some modifications have to be envisaged to make GPS-U vehicles know these positions when the communication with the GPS satellite is not possible. ODAM can be executed normally if these positions are accurately known, however, this is not always possible. In some situations, GPS-U vehicles can't obtain their exact previous and current positions. However, they can obtain some information about the driving direction and the distance from the accident. This can help the driver to take decisions. For example, if the accident happens in the opposite driving direction according to the accident in a divided highway there will be no need to brake.

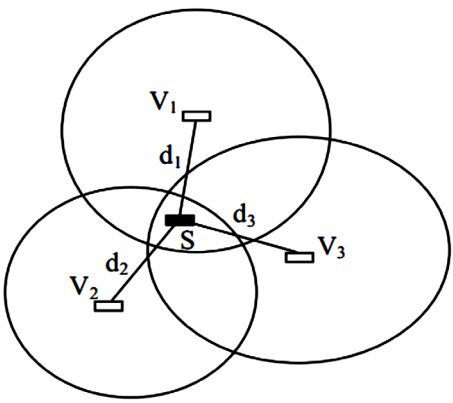


Figure 2. Location using three non-aligned GPS-E vehicles.

In order to obtain and refresh its position, a GPS-U vehicle, say S, periodically broadcasts a PREQ (Position Request) message to its one-hop neighbors. When a GPS-E vehicle receives a PREQ, it creates a PREP (Position Reply) message, includes its current position in that message, and sends it back to S. The knowledge of the exact position of S depends on the number and the positions (not all aligned) of neighbors sending PREP messages. S can compute its exact position if it receives at least three PREP from three different vehicles (Figure 2). When S(x,y) receives three PREP messages from three different vehicles, say $V_1(x_1,y_1)$, $V_2(x_2,y_2)$ and $V_3(x_3,y_3)$, it uses one of the several methods (ex. Received signal strength indicator) in order to determine the distances d_1,d_2 and d_3 from V_1,V_2 and V_3. In this case the exact position of S can be easily calculated by triangulation. The algorithm of ODAM can be executed normally if the GPS-U vehicles can compute their positions. In fact, GPS-U vehicle uses PREP messages in order to get its position instead of GPS satellite. However this is not always possible because in some cases, where the number of PREP messages is less than three, the exact position cannot be known. In what follows, we study these cases, when S receives two, one, or zero PREP. S receives answers when it moves from position called previous position to a new position called current position. Suppose that the previous position of vehicle S were $S_p(x_p,y_p)$ and its current position is $S_c(x_c,y_c)$. To allow computation of positions and driving direction of vehicles, we distinguish the following situations:

a) **Situation 2-3 and 3-2 :-** This situation is when S had two neighbors when it was in S_p and three neighbors in S_c (Fig 3.a), or three neighbors in S_p and two neighbors in S_c (Fig 3.b). In both the two situations, exact values of Sp and Sc can be obtained with the help of the distance D spent by vehicle S. For example, in Fig. 3.a (Resp. Fig. 3.b) we start by calculating the value of S_c (Resp. S_p) from the distances d1, d2 and d3, than the value of S_p (Resp S_c) from D, d_4 and d_5.

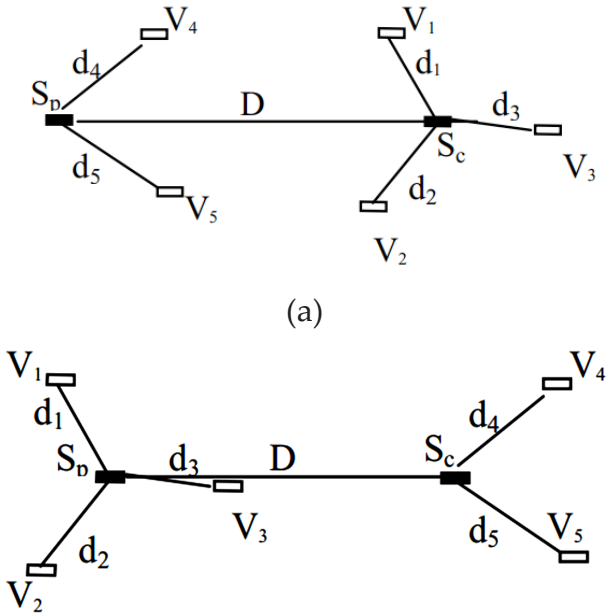
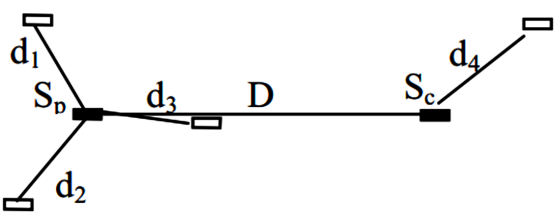
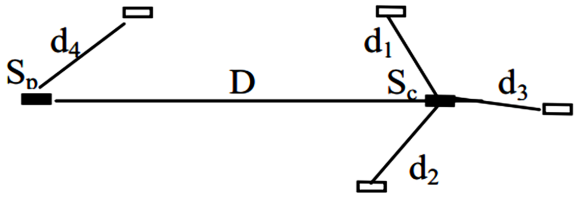


Figure 3. Location when a GPS-U vehicle moves: Fig. 3.a: from a position with two GPS-E neighbors to a position with three GPS-E neighbors; Fig.4.a: from a position with three GPS-E neighbors to a position with two GPS-E neighbors.

b) **Situation 3-1 and 1-3:-** This situation occurs when S had three neighbors when it was in S_p and one neighbor in S_c (Fig. 4.a), or one neighbors in S_p and three neighbors in S_c (Fig. 4.b).





(b)

Figure 4. Only one exact position can be known when (Fig. 4.a) a vehicle moves from a position with 3 GPSE neighbors to a position with 1 GPS-E neighbour and (Fig. 4.b) from a position with 1 GPS-E neighbour to a position with 3 GPS-E neighbors.

In both these two situations, one exact position S_p (Resp. S_c) in Fig 4.a (Resp. Fig 4.b) can be calculated, the second one called the lacking position is the intersection of two circles. Hence, if this intersection is in one point, the exact value of the lacking position S_c (Resp. S_p) in Fig 4.a (Resp. Fig 4.b) can be known. Otherwise, the lacking position can be one of the two points of the intersection of the two circles. For example, the value of S_c in Fig. 4.a is the intersection of two circles, the first has S_p as center and D as ray, and the second has V_4 as center and d4 as ray.

Table 1. Possibilities according to $N(S_c)$ and $N(S_p)$

$N(S_c)$ / $N(S_p)$	3	2	1
3	Exact position	Exact position	Probably exact position Probably driving direction
2	Exact position	Exact position	Probably driving direction
1	Probably exact position Probably driving direction	Probably driving direction	
0	Probably exact position Probably driving direction	Probably driving direction	

In some cases, even when the exact values of previous or current positions are not accurately known, the driving direction of vehicle S can be guessed. This is the case where the two possible solutions fall in the same driving direction. In Fig. 5, the two possible values of S_c are in the same driving direction with the vehicle V_4.

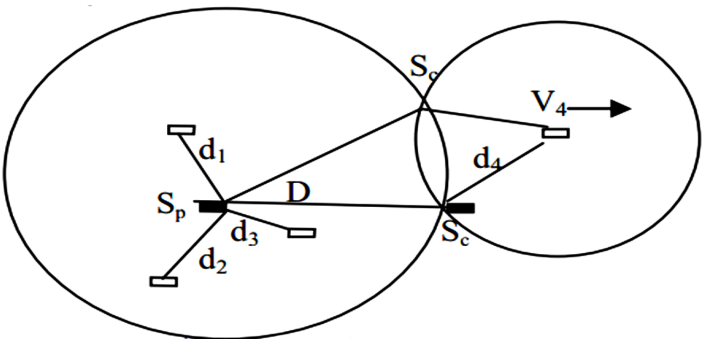


Figure 5. Driving direction can be known in situation 3-1

c) Situation 2-2 :-This situation occurs when S had two neighbors when it was in Sp and two neighbors in Sc (Figure 6).

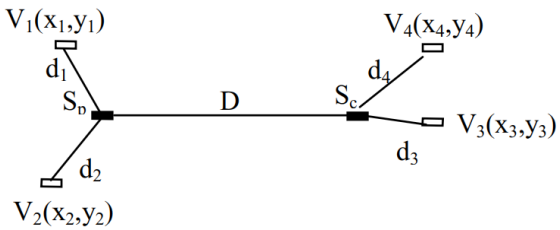


Figure 6. Exact position be known in situation 2-2.

The solution is given from the following system of equations.

$$(x_p - x_1)^2 + (y_p - y_1)^2 = d_1^2 \quad \text{(A)}$$

$$(x_p - x_2)^2 + (y_p - y_2)^2 = d_2^2 \quad \text{(B)}$$

$$(x_c - x_3)^2 + (y_c - y_3)^2 = d_3^2 \quad \text{(C)}$$

$$(x_c - x_4)^2 + (y_c - y_4)^2 = {d_4}^2 \quad \text{(D)}$$

$$(x_c - x_p)^2 + (y_c - y_p)^2 = D^2 \quad \text{(E)}$$

From the equations (A) and (B) we get two possible values for S_p, and from the equations (C) and (D) we get two possible values for S_c. Equation (E) restraints the number of solutions by eliminating bad solutions. Given the possible values of S_c and S_p, the position can be computed. If we call $N(S_c)$ (Resp $N(S_p)$) the number of GPS-E neighbors at the current (Resp. previous) position and using the same reasoning as above, we get the following table that lists all the cases that we can have.

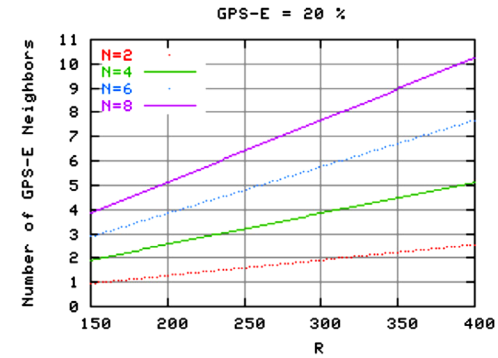
5.1.4 Simulation and analyses

In order to evaluate the performance of the ODAM with the help of GPS-U algorithm, we model a straight road 10 km long with C lanes in each direction (two directions). Each vehicle on the road moves at a constant, randomly chosen velocity. For sake of simplicity, we do not model complex maneuvers like lane changes and overtaking. Furthermore, we uniformly distribute the number of vehicles per kilometer per lane N to model the traffic density in the road. We take R as transmission range and τ as the rate of GPS-E vehicles. Since the knowledge of the position of a GPS-U vehicle depends on the number of its GPS-E neighbors, we derive a formula giving the mean number of GPS-E neighbors of a GPS-U vehicle. The reader can find mathematical calculations in the.

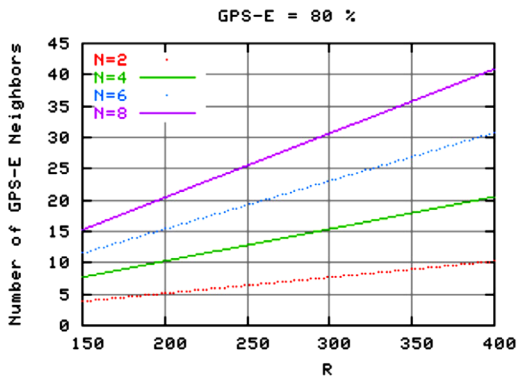
$$N(\text{GPS - E}) = \tau \cdot \left(\frac{N}{10^3 W}\right) \cdot \lceil \overline{H} - 1 \rceil \text{ where}$$

$$\overline{H} = \frac{2R^3}{3CW}\left(3Arc\cos T + (T-1)(2-T^2-T)\right) \text{ and}$$

$$T = \sqrt{1-\left(\frac{CW}{R}\right)^2}$$



(a)

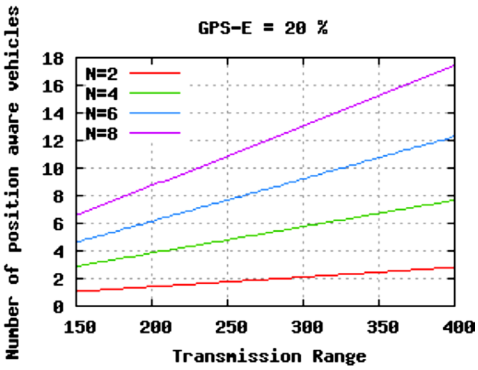


(b)

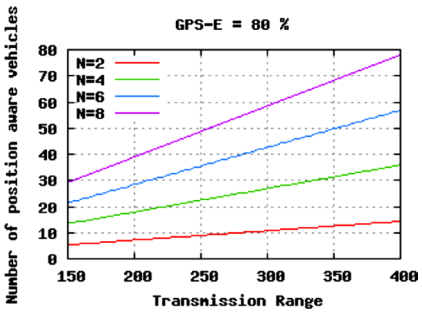
Figure 7. The average number of GPS-E neighbors with different τ rates.

Figures in 7 show the variation of the mean number of GPS-E neighbors of a vehicle according to the variations of the rate of GPS-E vehicles, transmission range and traffic density. In simulations, we have considered the density of traffic (N=2, 4, 6 and 8) and the rate of GPS-E (τ = 0.2, 0.4, 0.6 and 0.8). However, since the document must have a restricted size, we present results only for (τ = 0.4 (figure 7a) and 0.8 (figure 7b)). Simulations show that the mean number of GPS-E vehicles is proportional to the transmission range and the GPS-E vehicles rate. We remark that when τ is greater than 60% then the mean number of GPS-E neighbors is greater than three even with a low transmission range (R=150). This means that all GPS-U vehicles can obtain their positions and ODAM performs well.

We have observed also that when τ is around 40% and traffic density is low (N=2) that the mean number of GPS-E neighbors can be less than three when the transmission range is less than 250m. In this situation, the performances are not optimal since not all the GPSU vehicles can obtain their positions. However, we can envisage that the GPS-U vehicles increase their transmission power to reach ranges more than 250m in order to get more than two GPS-E neighbors, therefore they can compute theirs exact positions. When τ is around 20%, simulations show that the number of GPS-E neighbors is always less than three even the transmission range is 400m when the traffic density is low (N=2). In this situation, not all GPS-U vehicles can compute their exact positions. Hence, these vehicles can't be relays in ODAM, they are just passive elements.



(a)



(b)

Figure 8. The average number of position aware vehicles with different τ rates.

To improve the performance of ODAM, we have extended the ns-2 code of ODAM in order to support the presence of some GPS-U vehicles. Indeed, the performances of the proposed method are better than those presented in the mathematical analyzes because some GPS-U vehicles can get their positions and help other GPS-U vehicles. This means that average number of GPS-E vehicles can be higher than the one presented previously. Compared to the pictures in figure 7, the average number of aware position vehicles of an initial node un-ware of its position is increased according to the transmission range, the density of vehicles and the rate of GPS equipped vehicles. Thus the performance of ODAM can be optimal even with less than 40% initially GPS-E vehicles. Further simulations must be done to draw other characteristics allowing to highlight our purpose.

5.2 SELF-CORRECTING LOCALIZATION SCHEME FOR VEHICLE TO VEHICLE COMMUNICATION

Vehicular ad-hoc networks (VANETs) are greatly dynamic adhoc network topology. VANETs have two types of infrastructure: centralized and distributed. The centralized architecture merges cellular and ad-hoc technologies (V2I) while the distributed architecture is based on ad-hoc technology and known as Vehicle to Vehicle communication (V2V).

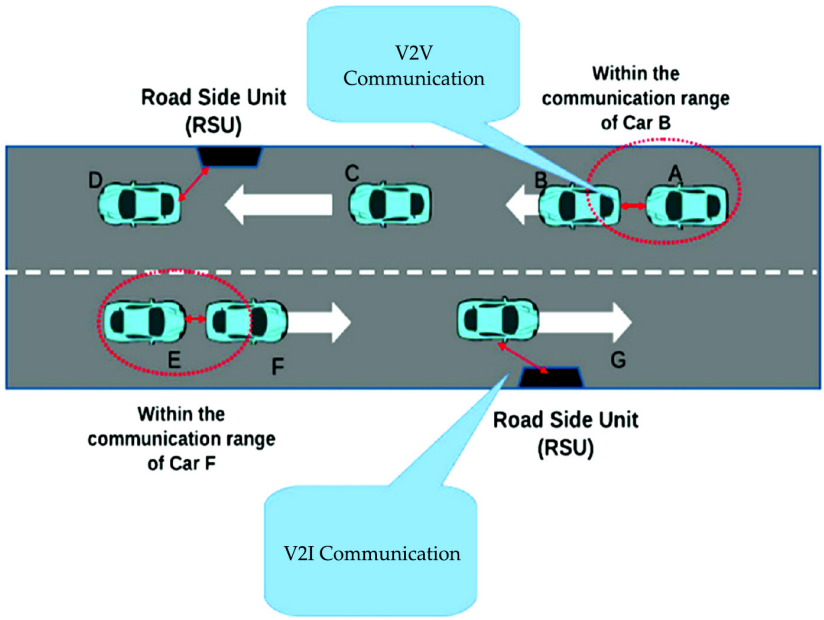
Vehicular ad-hoc technology has derived many applications such as road safety, traffic management and entertainment. Such applications depend on an accurate estimation for vehicular node position. Recently, most vehicles come with positioning technology, i.e.

Global Positioning System (GPS) devices, to precisely estimate their positions. In practice, vehicular nodes may suffer from GPS unavailability for productive or environmental reasons. The productive reasons include vehicular nodes are not originally equipped with GPS device, or have limitations in the GPS device. On the other hand, the environmental reasons attenuate GPS signals

such as dense foliage in rural regions, or compact high buildings in urban regions. Therefore, vehicular nodes particularly in V2V communication can be classified into two types. The first type is that vehicular nodes are denied from GPS service to estimate their locations while the other type is that vehicular nodes predict their locations due to GPS signals (i.e., called vehicular beacon nodes). A vehicular node can internally communicate with surrounding beacon nodes to estimate its location using a radio ranging technique (i.e., RSSI).

5.2.1 Future Tendency

In the future, merging roadside to vehicle communication with 5G networks will solve the problem of end-to-end latency in node localization because higher data rates are expected in 5G networks in which the latency will be less than one millisecond. In addition, 5G networks will provide us with higher capacity, reduced cost, consistent Quality of Experience provisioning and massive device connectivity. The future networks (i.e., 5G networks) will help also in developing autonomous vehicles (i.e., self-driving cars). Another localization technique in vehicular ad hoc networks uses local relative positioning to estimate the distance between vehicular nodes. Smart phones with GPS are exploited in location-based services, for example finding the nearest gas station. However, such solutions still suffer from the problem of GPS positioning and unavailability. The map matching is considered a way to correct GPS signal errors by alignment method. Since GPS data is inaccurate, map matching aligns vehicular node locations with the road on a digital map. The work appropriate to estimate positions of nodes moving on freeways. Roadside units are used to collect information and dead reckoning method is exploited to compute the current position of nodes based on the node's initial position.



Since vehicle to vehicle communication achieves minimum localization cost. The basic RSSI model can only operate well with high density of vehicular nodes with GPS; for example, a vehicular node requires at least three nearby nodes with GPS to estimate its location and that is unguaranteed all the time. In addition, when a vehicular node estimates its position based on basic RSSI model in the presence of physical changes such as shadowing and multipath (noise levels), the localization accuracy rapidly decreases. Afterwards, a grid method is used to decrease the localization error and increase the number of nodes that can estimate their locations. However, GOT scheme suffers from the effect of environmental parameters that lead to poor localization accuracy when their values increase.

In what follows, radio ranging limitations in vehicular node localization is introduced to show how the loss of received signal strength decreases the localization accuracy. Afterwards, the proposed scheme is introduced illustrating how it overcomes the basic RSSI problems and how it solves those problems.

5.2.2 Radio Ranging Limitations in Vehicular Node Localization

RSSI based localization is radio ranging technique to estimate vehicular node position. This work is based on such localization type; accordingly, a free propagation model is suggested which only one clear line-of-sight path between the transmitter and receiver is assumed. In free space propagation model, we can compute the received signal power at distance d as follows:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (1)$$

Where L is system loss including path loss, G_t is transmitter antenna gain, G_r is receiver antenna gain, λ is the wavelength and P_t is transmitted signal power. For hypothetical case, the parameters G_t, G_r, and L equal one. The communication range of vehicular nodes is assumed as a circular disk area around the sending node (i.e., beacon node) with radius R in which a vehicular node entering such area can receive from the sending node.

The free space propagation model can be simplified by assuming a reference point d_0 and a constant (K) as follows.

$$P_r = K\left(\frac{d_0}{d}\right)^2 \quad (2)$$

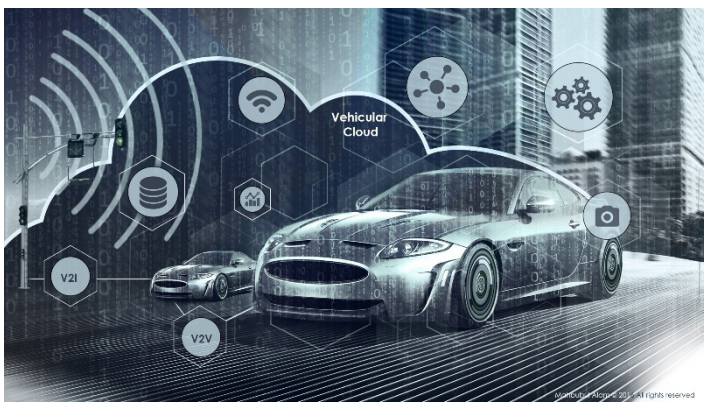
Particularly, signals in physical environments (i.e. urban and suburban) suffer from more noise due to shadowing and multipath effect resulting in complex path loss. Accordingly, the received power can be described as follows.

$$P_r = P_0\left(\frac{d_0}{d}\right)^L \quad (3)$$

Where L is the path loss exponent which equals two in free space as shown in Eq. (2). Generally, for physical environments as shown in Eq. (3), path loss exponent increases to reach a value greater than two and less than 6.5 due to many effects such as diffraction,

refraction, reflection, propagation medium (dry or humid air), absorption, height of antenna and distance between transmitter and receiver.

In what follows, RSSI limitations for vehicular node localization are explained. Radio ranging localization is mostly related to trilateration analytical model. In such model, when a node begins to estimate its location, it first measures signal strength (P_r) from three nearby beacons. By choosing a proper value for path loss exponent (L) and substituting P_r and L in Eq. (3), the measured distances to beacons can be determined which are analytically used to estimate the vehicular node position. For example, path loss exponent, in urban regions, changes from 2.7 to 3.5; accordingly, the average value (3.1) is chosen as a particular value for L. The localization accuracy decreases due to the expected changes in environmental parameters (i.e., path loss, shadowing and multipath) and network topology. The change in path loss occurs due to the actual shadowing/multipath level (noise level).



Furthermore, the rapid changes in network topology may lead to low localization accuracy. For example, when all beacon nodes have few meters spacing to the current node, trilateration analytical model achieves high localization accuracy where is poorly influenced by environmental parameters. On the other hand, when the distance between a beacon node and the current node is approximately big enough, localization accuracy largely may obviously decrease due to a small change in the noise level

because this change leads to an error in measured distances as shown below.

Accordingly, RSSI based localization is reasonable to estimate a vehicular node position when vehicular beacon nodes are near from the current vehicular node by few meters. In practice, vehicular beacon nodes randomly move in lanes; accordingly, RSSI limitations appear when vehicular beacon nodes communicate with the current node and the measured distance nears from the maximum communication range. As shown in Figure 9, there are four vehicular beacon nodes B_1, B_2, B_3 and B_4. Assuming vehicular node V_7 is located at (x, y) and discovers three 1-hop vehicular beacons B_2, B_3 and B_4 at (x_1, y_1), (x_2, y_2) and (x_3, y_3), respectively. Assume the measured distances at V_7 for signals received from those beacons are m_1, m_2 and m_3, respectively.

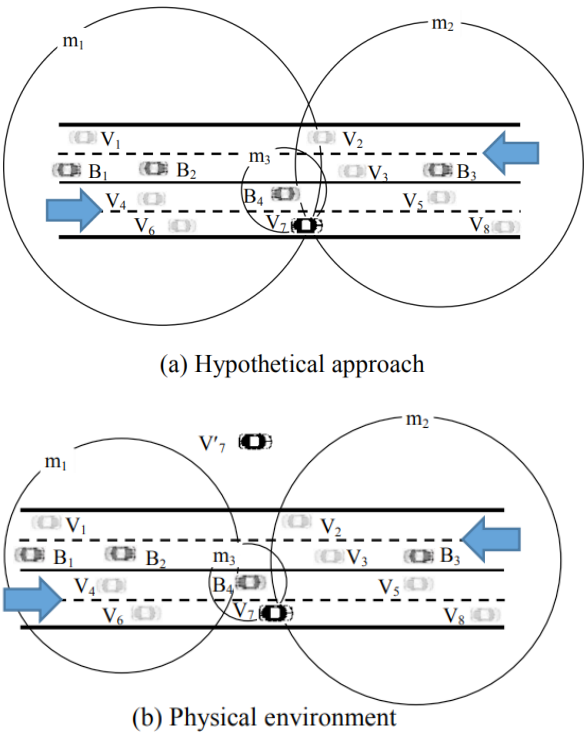


Figure 9. RSSI trilateration method (two far 1-hop beacons).

According to trilateration analytical model, the estimated location

(x_e, y_e) for node V_7 can be determined as follows.

$$x_e = \frac{A(y_3 - y_2) + B(y_1 - y_3) + C(y_2 - y_1)}{2[x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1)]} \quad (4)$$

$$y_e = \frac{A(x_3 - x_2) + B(x_1 - x_3) + C(x_2 - x_1)}{2[y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1)]} \quad (5)$$

Where

$$A = x_1^2 + y_1^2 - m_1^2, B = x_2^2 + y_2^2 - m_2^2, \text{ and } C = x_3^2 + y_3^2 - m_3^2$$

This work extends the trilateration analytical model as follows. The estimated distance between V_7 and beacons B_2, B_3 and B_4 are d_1, d_2 and d_3, respectively.

$$d_i = \sqrt{(x_i - x_e)^2 + (y_i - y_e)^2} \quad (6)$$

Where $i \in \{1, 2, 3\}$; the estimated distance difference (ΔSi) is an absolute value of the difference between V_7 estimated location and measured distance to beacon (i).

$$\Delta S_i = |d_i - m_i| = \left|\sqrt{(x_i - x_e)^2 + (y_i - y_e)^2} - m_i\right| \quad (7)$$

The total value for distance differences (ΔD) is determined as follows.

$$\Delta D = \Delta S_1 + \Delta S_2 + \Delta S_3 \quad (8)$$

Figure 9(a) shows the hypothetical approach (i.e., no noise occurs) which no measurement errors are observed in m_1, m_2 and m_3. As a result, actual and estimated positions for V_7 are matching ($\Delta D = 0$ due to $\Delta Si = 0$) which means no localization error happens. In practice, for physical environments, received signal strength is influenced by noise added to the path loss. Localization accuracy decreases due to different measurement errors in m1, m2 and m3, respectively.

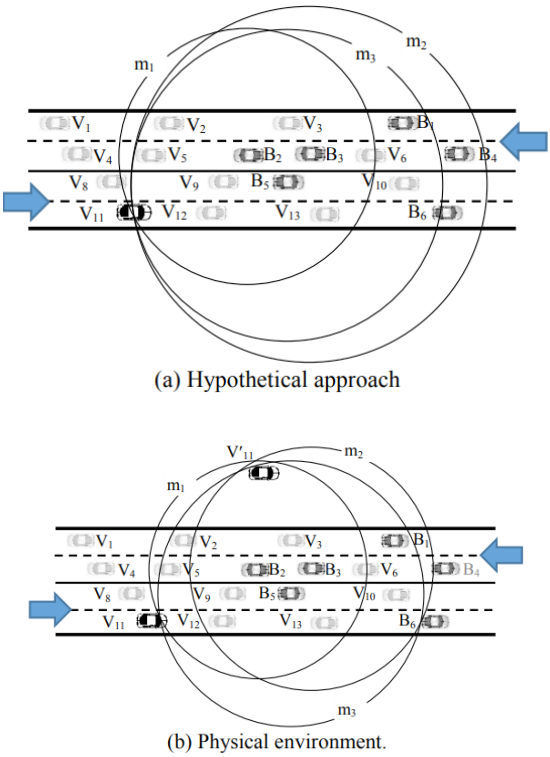


Figure 10. RSSI trilateration method (three far 1-hop beacons)

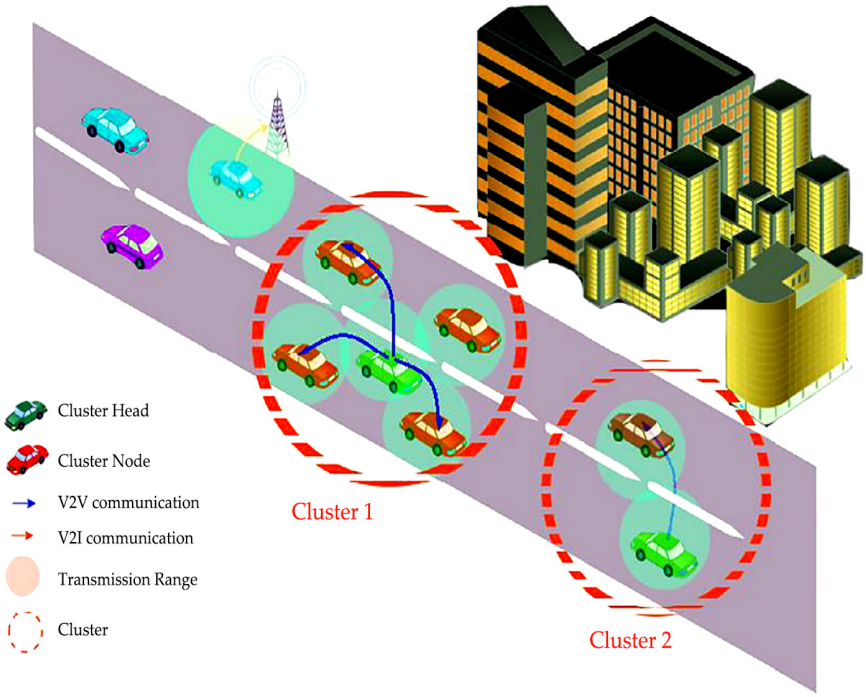
As shown in Figure 9(b), assume urban area where L is chosen (3.1) while the actual path loss due to different noise levels at B_2, B_3 and B_4 equals 2.76, 2.98 and 3.02, respectively. B_2 is more influenced by change in path loss than B_3 and B_4; accordingly, m1 circle radius decrease more than m_2 and m_3. The estimated location of V_7, computed by Eq. (4) and Eq. (5), is indicated by V′7 in Figure 9(b). The localization error can be easily detected from ΔD in Eq. (9) where ΔD is greater than zero. Generally, ΔD can be used in this work to indicate to the localization accuracy which the localization accuracy increases as long as ΔD decreases and approaches zero. In this scenario, the localization error exceeds half of the maximum communication range when comparing the actual position V_7 to estimated location V_7' in Figure 9(b).

The following scenario shows another case, at the same environmental parameters, when beacon nodes are close to each other and have large spacing with the current node. As shown in Figure 10, there are six beacon nodes from B_1 to B_6. Assuming vehicular node V_{11} discovers three nearby vehicular beacons B_2, B_3 and B_5 where are close to each other. For hypothetical approach, estimated location and actual position for V_{11} are congruent as shown in Figure 10(a).

In contrast, when an error occurs in measured distance for one or two beacons as shown in Figure 10(b), the estimated position V'_{11} is imprecise where localization error increases to values greater than R where R represents the maximum communication range for vehicular node.

5.2.3 The Proposed Localization Scheme

When radio ranging localization is used in vehicle to vehicle communication, the basic RSSI analytical model cannot correctly estimate vehicular nodes' locations due to expected noise in physical environments. For high dynamic network topology, it is unguaranteed that three close beacons are discovered for most vehicular nodes all the time. Furthermore, in lightweight traffic (i.e., rural regions or nightly driving), a vehicular node may originally fail to communicate with three beacons. A vehicular node can exploit farthest beacons (2-hop beacons) besides nearest beacons (1-hop beacons) to estimate its location. In this work, many extensions and improvements are introduced to reduce the time complexity and effectively deal with all expected changes in environmental parameters. In what follows, SCL-VNET algorithm is introduced to show those extensions and improvements. A location packet is broadcasted from a beacon node in which each packet contains sender ID and its location at time (t_i). The sender waits for (Δt) and when the time interval (Δt) passed, it broadcasts the next location packet.



SCL-VNET algorithm, for V2V communication, contains three phases: communication, estimation, correction and alignment. Each vehicular node (VN) separately runs such algorithm to estimate its location. The first phase is illustrated in Lines (1- 6). When a VN receives a location packet from a beacon node, beacon ID and its current location are extracted from the received packet. The measured distance can be determined using the received signal power that can be obtained from PHY/MAC layer. A VN node establishes a new data record containing beacon ID, beacon location and measured distance to save it in a list called (nearest beacon list) as shown in Line (3). A VN changes its mode from receiving mode to sending mode to broadcast new location packet (it is called here forwarded packet) containing nearest beacons information.

SCL-VNET **algorithm**

BN (b) at time (t_i)

```
1. oneHopBoadcast(ID, LOC);
2. scheduleNextBroadcast (Δt);
```

VN (v) at time (t_i)

Communication phase:

```
1.  WHILE t < t_i + ΔT
2.    IF msg_recieved THEN
3.       nearestBeaconList.add (beacon(ID, LOC, mdist));
4.       oneHopBoroadcast(ID, LOC, mdist, ownerID);
5.    IF forwarded_msg_recieved THEN
6.          farthestBeaconList.add (beacon record);
```

Estimation phase:

```
7.    beaconList = nearestBeaconList;
8.    IF beaconListSize >= 3 THEN
9.       selectBestThreeNearestBeacons ( );
10.   estimateLocationByTrilateration ( );
11. ELSE   /*less than three nearest beacons are discovered*/
12.   IF previous_location_is_available THEN
13.      beaconList.add (previousLocation);
14.   IF beaconListSize < 3 THEN
15.      selectedBeacon = chooseBestFarthestBeacons( );
16.      beaconList.add(selectedBeacons);
17.      IF beaconListSize = 3 THEN
18.         sampleList =
    samplingBeaconMeausredCircle(selected_beacons);
19.         FOREACH sample(i) IN samplesList
20.            estimateLocationByTrilateration ( );
21.    ELSE
22.      estimateLocationByTrilateration ( );
```

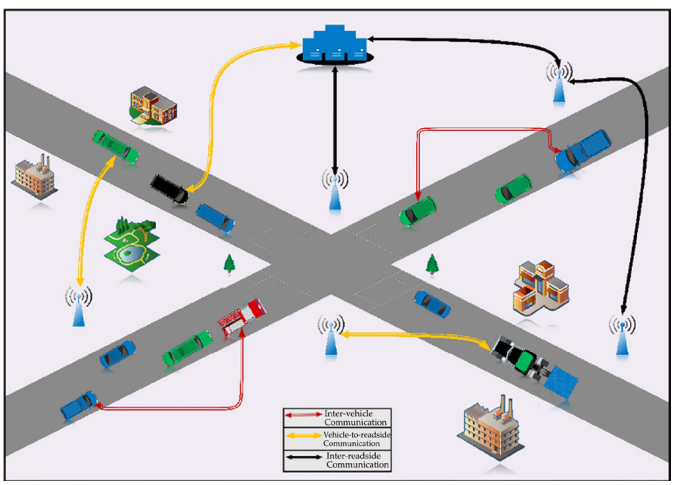
Correction and alignment phase:

```
23. IF beacon (i) is a nearest beacon and m_i > m_th THEN
24.     measuredDistanceList (i) =
    generateNewMeasuredDistance (m_i , Δm);
25. visibleSolutionsList = createVisibleSolution
    (measuredDistanceList (i));
26. FOREACH visibleSolution (j) IN visibleSolutionsList
27.     estimateLocationByTrilateration( );
28. chooseBestSolution( );
29. bestSolutionAlignment( );
```

Finally, a VN returns to receiving mode to receive forwarded packets from VN neighbors. Beacon ID, its current location and its measured distance to its owner (forwarding node) are extracted from forwarded packet. The measured distance to forwarding node can also be determined using received signal power from that can be obtained from PHY/MAC layer. A VN node establishes a new data record containing five attributes, 1) beacon ID, 2) beacon

location, 3) measured distance between beacon and forwarding VN, 4) forwarding VN id, and 5) measured distance between forwarding VN and receiving VN. This record is stored in new list called (farthest beacon list) as shown in Line (6).

The second phase is called an estimation phase, illustrated in Lines (7-22), which a new list called (beacon list) is established to contain all existing 1-hop beacons. Such phase is performed whether the number of 1-hop beacons is greater than, equal, or less than three. The first part described in Lines (8-10) shows how SCL-VNET works when the number of beacons is greater than or equal three. When the number of beacons is greater than three, a NV chooses a best three 1-hop beacons. The best three 1-hop beacons have the shortest measured distances to the current vehicular node. Afterwards, VN location is estimated using basic RSSI analytical model, represented by Eq. (4) and Eq. (5) as shown in Line (10), whether there are either exactly three 1-hop beacons, or more and best three 1-hop beacons are chosen.



When the current vehicular node, that already discovered three 1-hop beacons, estimates its location, it moves to the correction and alignment phase. Such phase is based on the extended trilateration analytical model described in equations (6), (7) and (8). When ΔD is examined, a VN decides correcting its estimated position when ΔD is greater than zero which means measured distances

to 1-hop beacons suffer from noise in path loss. The correction and alignment phase is described in Lines (23-29). A measured distance list is created for each beacon (i) where i ∈ {1, 2, 3}, as shown in Line (24). The measured distance can be corrected by applying a small change in the distance step by step to reach the actual distance (it is called a correction step). A new measured distance for beacon (i) equals ($m_i \pm k\Delta m_i$); k ∈ {0, 1, 2, …., n} where nΔmi represents the expected maximum correction step based on the maximum noise level. Each beacon (i) contains at least one element (at k = 0) equals m_i. Meanwhile the time complexity of correction phase is $O(n^3)$, the correction step (CS) parameter (in meters) controls the number of visible solutions in which each visible solution contains one measured distance ($m_i \pm k\Delta m_i$) for each beacon (i).

Figure 11 shows the correction process when SCL-VNET scheme is applied. As shown in Figure 11, new measured distances are generated from ($m_i \pm k\Delta_m$). The measured distance list size is 7, 5 and 3 for B_2, B_3 and B_4 constrained by the maximum expected noise level as shown in Figure 11. The visible solution space contains 105 solutions. After scanning such solutions, the optimal solution that achieves minimum ΔD is determined and the corrected location for VN is indicated by black dot as shown in Figure 11.

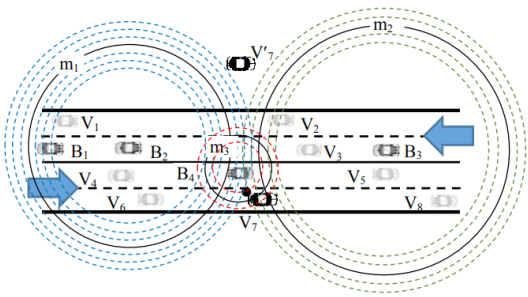
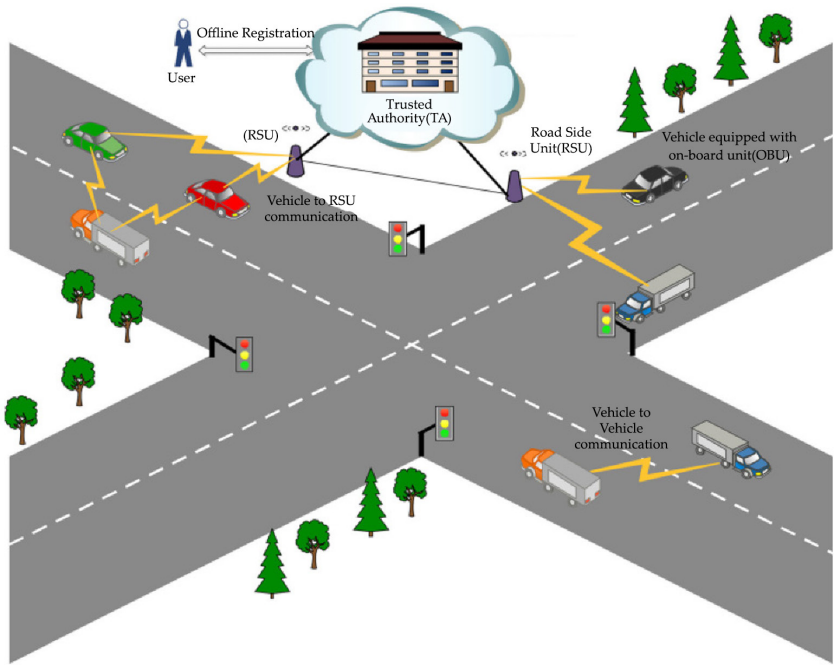


Figure 11. Correction phase

After the correction process finishes, SCL-VNET runs the alignment process, as shown in Line (29), to achieve more localization accuracy. Such process matches the estimated location and the

preloaded map. When a vehicular node locates between lanes or outside the road, SCL-VNET aligns a vehicular node and relocates it on the nearest lane.

The second contribution in this work is to estimate a vehicular node position in lightweight traffic density (sparse network). When the number of nearest beacons is less than three, the trilateration analytical model fails to estimate VN location; however, SCL-VNET scheme solves this problem as shown in Lines (11-22) of estimation phase. The estimated location that obtained at (t-Δt) is examined. Supposing VN moves with regular velocity s during Δt, when estimated location obtained at (t-Δt) (it called pervious location) is available, the distance from such location to current location can be considered with value less than or equal sΔt. SCL-VNET generates a virtual beacon node at previous location and add it with its measured distance to beacon list as 1-hop beacon.



When the number of 1-hop beacons is still less than three as shown in Line (14), 2-hop beacon nodes can be exploited. When there are many 2-hop beacon nodes, best three nodes are chosen

which they have shortest measured distances to forwarding VNs. Furthermore, the forwarding VNs have also shortest measured distances to current VN. However, there is a problem to use 2-hop beacon nodes in trilateration analytical model because the direct measured distance between VN and 2-hop beacons is unavailable. SCL-VNET solves this problem by firstly estimating a forwarding VN position; afterwards, it estimates the current VN location as follows.

Assuming a vehicular node (V1) receives from two 1-hop beacon nodes B_1 and B_2 by measured distances m1 and m2 in physical environment (i.e. measured distance may be less than actual distances). Assume V1 receives from a 2-hop beacon (B_3) via a forwarding VN (V_2) as shown in Figure 12. Meanwhile a circle circumference of beacon node (B_3), with a radius represented by m_{23} where m_{23} is a measured distance between a beacon (B_3) and VN (V_2), is the locus of V_2, then SCL-VNET scheme takes samples from the circle circumference. Sampling a circle circumference means considering the integer values that represent the intersection points between the circle circumference and the four lanes.

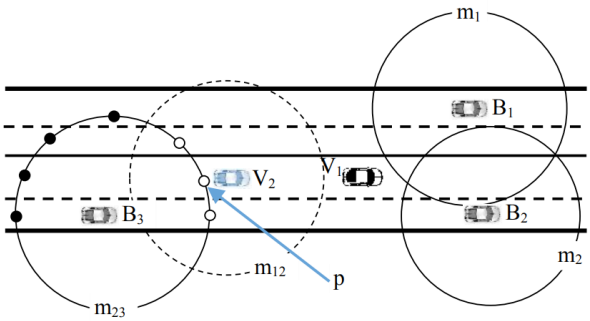


Figure 12. Sampling 2-hop beacon circle circumference

The number of considered samples represented by small circles, as shown in Figure 12, are seven only which reduce the proposed algorithm time complexity. SCL-VNET estimates V1 position by iterative trilateration process for such samples. Generally, the number of samples (N) taken over a 2-hop beacon circle circumference with measured distance (m) is described in Eq. (9)

where W is the lane width and R is the maximum communication range.

$$N = \begin{cases} 2 & m < W \\ 3 & m = W \\ 4 & W < m < 2W \\ 5 & m = 2W \\ 6 & 2W < m < 3W \\ 7 & m = 3W \\ 8 & 3W < m < R \end{cases} \quad (9)$$

When a vehicular node requires two 2-hop beacons to estimate its location, the number of pair samples may reach 8^2 pair samples at the worst case. When three 2-hop beacons are used, the number of triple samples may increase to reach 8^3 triple samples. Therefore, further constraint is used to decrease the number of iterations by picking candidate samples. Such constraint is based on the distance between two 1-hop beacons must be less than 2R where R represents the maximum communication range. As shown in Figure 13, a vehicular node (V_1) discovers two 1-hop beacon nodes B_1 and B_2. The maximum distance between two 1-hop beacons is 2R.

When applying such constraint on the first scenario (two 1-hop beacons and one 2-hop beacon), we assume 1-hop beacons are located at (x_1^b, y_1^b) and (x_2^b, y_2^b), respectively. Also, we assume samples (x_j, y_j) are taken from 2-hop beacon circle circumference where $j \in \{1, 2, \ldots, 8\}$. The proposed constraint can be analytically described as shown in Eq. (10).

$$\forall_i \forall_j \left(\sqrt{(x_i^b - x_j)^2 + (y_i^b - y_j)^2} \right) \leq 2R \quad (10)$$

Where $i \in \{1, 2\}$. The result of applying the constraint is shown in Figure 13; seven samples are filtered to three samples only indicated by whole circles. The candidate sample achieves minimum ΔD. Afterwards, the correction process is applied for 1-hop beacon nodes that satisfies the correction conditions to improve the localization accuracy as shown above.

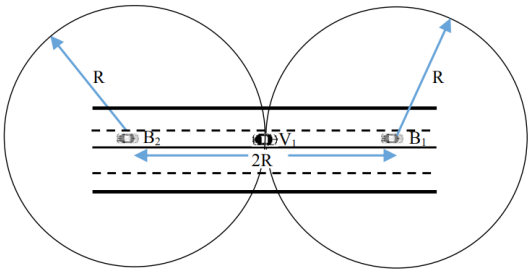


Figure 14. Maximum distance between two 1-hop beacons.

Similarly, the second scenario assumes one 1-hop beacon and two 2-hop beacons are available. The sampling process is performed for two 2-hop beacons to obtain (x_j^1, y_j^1) and (x_j^2, y_j^2). Assume 1-hop beacon is located at (x, y); accordingly, samples are filtered as shown in Eq. (11).

$$\forall_i \forall_j \left(\sqrt{\left(x - x_j^i\right)^2 + \left(y - y_j^i\right)^2} \right) \leq 2R \tag{11}$$

Where i ∈ {1, 2} and j ∈ {1, 2, ……, 8}. The candidate sample pair achieves minimum ΔD and the correction phase can be also applied for 1-hop beacon, when satisfying the correction conditions, to improve the localization accuracy.

The final scenario assumes three 2-hop beacons are available. The sampling process is performed for three 2-hop beacon nodes to obtain (x_j^1, y_j^1), (x_j^2, y_j^2) and (x_j^3, y_j^3). Samples are filtered as shown in Eq. (12).

$$\forall_i \forall_j \left(\sqrt{\left(x_j^i - x_j^k\right)^2 + \left(y_j^i - y_j^k\right)^2} \right) \leq 2R \quad , \mathrm{i} < \mathrm{k} \tag{12}$$

where i ∈ {1, 2, 3}, k ∈ {1, 2, 3} and j ∈ {1, 2, ……, 8}. The candidate sample triple achieves minimum ΔD.

In what follows, the simulation results are introduced to show how SCL-VNET scheme decreases the localization error for V2V communication in physical environments, and how it works in case of lightweight traffic density. In addition, a comparative study is illustrated to show how SCL-VNET scheme overcomes the existing localization schemes.

5.3 VEHICULAR AD HOC NETWORKS: A NEW CHALLENGE

A number of interesting and desired applications of Intelligent Transportation Systems (ITS) have been stimulating the development of a new kind of ad hoc network: Vehicular Ad Hoc Networks (VANets). In these networks, vehicles are equipped with communication equipment that allows them to exchange messages with each other in Vehicle-to-Vehicle communication (V2V) and also to exchange messages with a roadside network infrastructure (Vehicle-to-Roadside Communication – V2R). A number of applications are envisioned for these networks, some of which are already possible in some recently designed vehicles (Fig. 15):

- Vehicle collision warning
- Security distance warning
- Driver assistance
- Cooperative driving
- Cooperative cruise control
- Dissemination of road information
- Internet access
- Map location
- Automatic parking
- Driverless vehicles

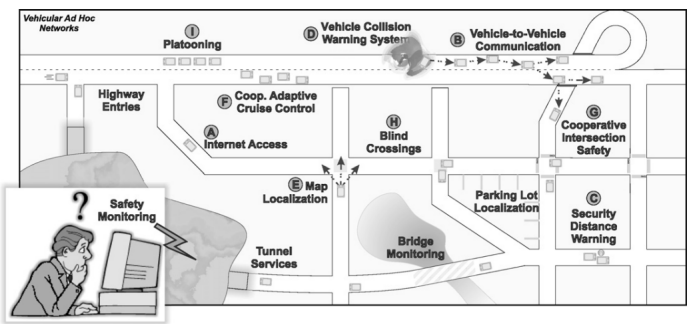


Figure 15. Several VANet applications.

All of these applications require, or can take advantage of, some sort of localization technique. In the localization problem, the definition of a reference system among nodes is performed by identifying their physical location (e.g., latitude, longitude, and altitude) or their relative spatial distribution in relation to each other. For instance, Map Location is usually done using Global Positioning System (GPS) receivers with a Geographic Information System, while Vehicle Collision Warning Systems can be implemented by comparing distances between nodes' locations combined with geographic information dissemination.

As ITS and VANets technology advances toward more critical applications such as Vehicle Collision Warning Systems (CWS) and Driverless Vehicles, it is likely that a robust and highly available localization system will be required.

Unfortunately, GPS receivers are not the best solution in these cases, since their accuracy range from up to 20 or 30 m and since they cannot work in indoor or dense urban areas where there is no direct visibility to satellites.

For these reasons and, of course, for security reasons, GPS information is likely to be combined with other localization techniques such as Dead Reckoning, Cellular Localization, and Image/Video Localization, to cite a few. This combination of localization information from different sources can be done using such Data Fusion techniques as Kalman Filter and Particle Filter.

The localization requirements of a number of VANet applications. We then survey several proposed localization techniques that can be used to estimate the position of a vehicle, and we highlight their advantages and disadvantages when applied to VANets.

By concluding that none of these techniques can achieve individually the desired localization requirements of critical VANet applications, we show how the localization information from multiple sources can be combined to produce a single position that is more accurate and robust by using Data Fusion techniques.

5.3.1 Location-aware VANet applications

Most VANet applications consider the availability of real-time updated position information. They differ, however, on the localization accuracy required in order to be able to function properly.

For instance, some applications can work with inaccurate localization information in which computed positions can have errors from 10 to 20 or 30 m, while other applications, especially critical safety applications, require more accurate and reliable localization systems with sub-meter precision.

We divide VANet applications into three main groups according to their localization requirements and show how position information is used by these protocols and algorithms. These localization requirements for VANet applications are then summarized in Table 2.

Applications able to work with inaccurate localization

Although some VANet applications do not require any localization to function, most of them can take advantage of localization and show better performance when the position information of vehicles is available.

Most of these applications are related to vehicle communication, which includes vehicle-to-vehicle (V2V) and vehicle-to-roadside (V2R) communication, and provide services such as information routing, and the data dissemination of accidents, road congestion, etc. The algorithms that deal with communication will accept localization errors mostly within 10–20 or even 30 m, since the long transmission range of the vehicles' transmitters can compensate for these localization inaccuracies. However, the greater the localization error, the worse the algorithms' performance. In the following paragraphs we will discuss some of these applications and algorithms.

Table 2. Required localization accuracy for some VANet applications.

Technique	Localization Accuracy		
	Low	Medium	High
Routing	X	–	–
Data Dissemination	X	–	–
Map Localization	X	–	–
Coop. Adapt. Cruise Control	–	X	–
Coop. Intersection Safety	–	X	–
Blind Crossing	–	X	–
Platooning	–	X	–
Vehicle Col. Warn. System	–	–	X
Vision Enhancement	–	–	X
Automatic Parking	–	–	X

Routing protocols for VANets usually use position information in order to improve their performance and be able to comply with such VANets requirements as dynamic topology changes and frequent network fragmentation. This routing technique has long been used in Ad Hoc networks and most of its protocols can also be applied to VANets. A classical example is Greedy Forwarding, in which, location information is used at each step to forward a packet to the neighbor nearest to the destination node. However, some geographic routing protocols have also been designed specifically for VANets, taking advantage of more geographical knowledge such as Maps and movement information. Routing techniques are also used to access local infrastructured networks via an Internet connection (Fig. 15A). In these cases, position information as well as future trajectory knowledge can be used to assist routing.

Several Data Dissemination protocols have been proposed for VANets that aim to inform both near and distant vehicles about transit conditions such as the road flow, congestion, and potentially dangerous situations. Most of these protocols also consider localization knowledge mostly to ensure that locally disseminated information reaches only the vehicles that should be interested in it. Driver direction can also be used, as proposed by the ODAM algorithm. In Fig. 15B, road information about a

dangerous situation is disseminated to interested vehicles. A widely known and already in-use driver assistance application is Map Localization, in which the current position of the vehicle is shown on a map. In these applications, a path direction between two points of the city, for instance, can be drawn on a map indicating the current location of the vehicle. This application can assist drivers in situations when they find themselves lost in a unknown part of the city, as depicted in Fig. 15E. Localization information with errors of about 10–20 m are proven to be useful for this kind of application, since map knowledge can be used to overcome this high localization inaccuracy.

Applications requiring accurate localization

This kind of application require a certain degree of confiability and accuracy in the computed positions and/or in the distance estimation between vehicles. Applications in this group are usually Cooperative Driving applications, where vehicles in a VANet exchange messages between them to drive and share the available space on the road cooperatively. In these applications, the vehicles can assume partial control over driving. In most cases, localization errors from 1 to 5 meters are acceptable. In the following paragraphs we will discuss some of these applications and algorithms.

In Cooperative Adaptive Cruise Control, the vehicle maintains the same speed whether traveling uphill or down without requiring driver intervention. Usually, the driver sets the speed and the system will take over, but in this case, vehicles can cooperate among themselves to set this speed adaptively (Fig. 15F). This application only takes care of speed, while the driver still has to control the direction of the vehicle. Another interesting application of VANets is Cooperative Intersection Safety, in which vehicles arriving at a road intersection exchange messages in order to make a safe crossing as depicted in Fig. 15G. Besides ensuring a safe crossing, it is also possible to make a Blind Crossing, where there is no light control and the vehicles cooperate with each other to make a cooperative crossing (Fig. 15H). In these applications, the

localization accuracy must allow the application to differentiate between the lanes as well as the sides of the street. Vehicle Following or Platooning is a technique used to make one or more vehicles follow a leader vehicle to form a train-like system, as shown in Fig. 15I. This application can be useful in situations where two or more vehicles are going to the same location. A minimum distance must be ensured between vehicles. Also, vehicles must track the position of the vehicle in front of them with a good precision, both of which can be accomplished by a localization system with accurate position information.

Applications requiring high-accurate localization

A third class of applications for VANets requires very precise and reliable localization systems. Most of these applications are critical safety applications such as Vehicle Collision Warning Systems (CWS) and other driver assistance applications. In driver assistance applications, VANet resources are used to enhance the driver's perception and knowledge of the road and environment. In these applications, the driver is informed about the surrounding environment in order to improve safety, and, in case of emergency, the vehicle can perform some automatic procedures. These are the most interesting applications for VANets, and since we are dealing with safety, position information reliability and accuracy are crucial. Accurate positioning ensures localization with a meter or sub-meter precision in order to estimate accurately the distances between vehicles, while a reliable localization will ensure that updated information will always be available. In the following paragraphs we will discuss some of these applications and algorithms.

Vehicle Collision Warning Systems are one of the most interesting applications of VANets for driver assistance. One part of these systems is Security Distance Warning, in which the driver is warned when a minimum distance to another vehicle is reached (Fig. 15C). It can also implement an emergency break when the distance between two vehicles or between a vehicle and an obstacle decreases too quickly, as shown in Fig. 15D. Another

part of these systems is when a collision has already occurred and nearby vehicles need to be warned (warn messages) so they can avoid pile-up collisions (Fig. 15D). In these cases, multihop communication can be used to disseminate collision information. Since they provide a critical application for safe driving, these applications require robust, accurate, and reliable localization systems

Another driver assistance application is Vision Enhancement, in which drivers are given a clear view of vehicles and obstacles in heavy fog conditions and can learn about the existence of vehicles hidden by obstacles, buildings, and by other vehicles. Automatic Parking is an application through which a vehicle can park itself without the need for driver intervention. In order to be able to perform an automatic parking, a vehicle needs accurate distance estimators and/or a localization system with sub-meter precision.

REFERENCES

1. M. Joe, and B. Ramakrishnan, Review of Vehicular Ad hoc Network Communication Models including WVANET (Web VANET) Model and WVANET Future Research Directions, Wireless Networks, Springer, vol. 22, no. 7, pp. 2369-2386, 2015.
2. M. Joe, and B. Ramakrishnan, WVANET: Modelling a Novel Web Based Communication Architecture for Vehicular Network, Wireless Personal Communications, Springer, vol. 85, no. 4, pp. 1987-2001, 2015.
3. S. Biswas, R. Tatchikou, and F. Dion, Vehicle-to-Vehicle Wireless Communication Protocols for Enhancing Highway Traffic Safety, IEEE Communications Magazine, vol. 44, pp. 74-82, 2006.
4. N. Patwari, J. N. Ash, A. O. Hero, R. L. Moses, and N. S. Correal, Locating the Nodes: Cooperative Localization in Wireless Sensor Networks, IEEE Signal Process. Mag., vol. 22, no. 4, pp. 5469, Jul. 2005.
5. Boukerche, H. Oliveira, E. Nakamura, and A. Loureiro, Localization Systems for Wireless Sensor Networks, IEEE Wireless Communications, vol. 14, pp. 6-12, 2007.
6. R. Parker, and S. Valaee, Vehicle Localization in Vehicular Networks, IEEE VTC, 2006.
7. Benslimane, Localization in Vehicular Ad hoc Networks, Systems. Communications, pp. 1925, 2005.
8. R. Parker, and S. Valaee, Vehicular Node Localization using Received Signal-strength Indicator, IEEE Trans. Veh. Technol., vol. 56, no. 6, Nov. 2007.
9. T. Yan, W. Zhang, and G. Wang, A Grid-Based On-Road Localization System in VANET with Linear Error Propagation, IEEE Trans. on Wireless Communications, vol. 13, no. 2,pp. 861-870, 2014.
10. Ou, A Roadside Unit-based Localization Scheme for Vehicular Ad Hoc Networks, International Journal of Communication Systems, vol. 27, pp. 135-150, 2014.

11. H. Li, X. Chen, L. Huang, and D. Yao, A GPS/Wi-Fi Integrated System for Positioning in Cooperative Vehicle and Infrastructure System, in Proc. of IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 285-289, 2012.
12. Gupta, and R. Jha, A Survey of 5G Network: Architecture and Emerging Technologies, IEEE Access Journal, Special Section on Recent Advances in Software Defined Networking for 5G Networks, vol. 3, pp. 1206-1232, 2015.
13. J. Rodriguez, Fundamentals of 5G Mobile Networks. John Wiley & Sons, Ltd, 1st Edition. 2015.
14. P. Goyal, and A. Buttar, A Study on 5G Evolution and Revolution, International Journal of Computer Networks and Applications (IJCNA), vol. 2, no. 2, 2015.
15. R. Parker, and S. Valaee, Vehicle Localization in Vehicular Networks, in Proc. of IEEE 64th Vehicular Technology Conference, pp. 1-5, 2006.
16. V. Kukshya, H. Krishnan, and C. Kellum, Design of a System Solution for Relative Positioning of Vehicles using Vehicle-to-Vehicle Radio Communications during GPS Outages, in Proc. of IEEE 64th Vehicular Technology Conference, pp.1313-1317, 2005.
17. M. Ikram, and J. Cazalas, Efficient Collaborative Technique using Intrusion Detection System for Preserving Privacy in Location-based Services, International Journal of Computer Networks and Applications (IJCNA), vol. 2, no. 5, 2015.
18. O. Pink, and B. Hummel, A Statistical Approach to Map Matching using Road Network Geometry, Topology and Vehicular Motion Constraints, in Proc. of the 11th International IEEE Conference on Intelligent Transportation Systems, 2008.
19. S. Lin, W. Ying, X. Jingdong, and X. Yuwei, An RSU-assisted Localization Method in non-GPS Highway Traffic with Dead Reckoning and V2R Communications, in Proc. of the 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 149-152, 2012.

20. Niculescu and B. Nath, Ad hoc Positioning System (APS) using AOA, in 22nd IEEE Annual Joint Conference on the Computer and Communications Societies(INFOCOM 03), pp. 17341743, April 2003.
21. Ramakrishnan, R. Nishanth, M. Joe and R. Shaji, Comprehensive Analysis of Highway, Manhattan and Freeway Mobility Models for Vehicular Ad hoc Network, International Journal of Wireless and Mobile Computing, vol. 9, no. 1, pp. 78-89, 2015.

CHAPTER 6

VEHICULAR APPLICATIONS

INTRODUCTION

Vehicular services and applications can be classified into three main families: safety, traffic management and monitoring, and comfort. Road safety certainly is the main motivation and represents the major issue in Intelligent Transportation Systems (ITS). About 40.000 people died on roads every year only in the European Union, with around 1.7 million people incurring several injuries. These accidents are often caused by a faulty driver behavior, bad weather conditions or mechanical problems. Hence, one of the most important solutions relies on using vehicular communications to anticipate road accidents, extend road visibility and disseminate safety information. In addition to road safety applications, traffic information and monitoring systems are another important application field of vehicular networks. They aim at improving traffic flow and road usage, providing timely information about

traffic state along many kilometers. Finally, the goal of comfort applications is offering novel on board services to improve the travel experience, improving common multimedia capabilities of current commercial vehicles. In order to analyze this wide world of vehicular applications, some of the most representative ones have been chosen as cases of study. Thus, three reference applications which best represent these three families are described, such as Cooperative collision warning, Platooning or Parking place management

Vehicular Ad hoc NETworks (VANETs) belong to a subcategory of traditional Mobile Ad hoc NETworks (MANETs). The main feature of VANETs is that mobile nodes are vehicles endowed with sophisticated "on-board" equipment's, traveling on constrained paths (*i.e.*, roads and lanes), and communicating each other for message exchange via Vehicle-to-Vehicle (V2V) communication protocols, as well as between vehicles and fixed road-side Access Points (*i.e.*, wireless and cellular network infrastructure), in case of Vehicle-to-Infrastructure (V2I) communications. Future networked vehicles represent the future convergence of *computers*, *communications infrastructure*, and *automobiles*. Vehicular communication is considered as an enabler for driverless cars of the future. Presently, there is a strong need to enable vehicular communication for applications such as safety messaging, traffic and congestion monitoring and general purpose Internet access.

Applications such as safety messaging are near-space applications, where vehicles in close proximity, typically of the order of few meters, exchange status information to increase safety awareness. The aim is to enhance safety by alerting of emergency conditions. Applications for VANETs are mainly oriented to safety issues (*e.g.*, traffic services, alarm and warning messaging, audio / video streaming and generalized infotainment, in order to improve the quality of transportation through time-critical safety and traffic management applications,). At the same time, also entertainment applications are increasing (*e.g.*, video streaming and video-on-demand, web browsing and Internet access to passengers to enjoy the trip).

Applications of alarm messaging have strict latency constraints of the order of few milliseconds, and very high reliability requirements. In contrast, applications such as traffic and congestion monitoring require collecting information from vehicles that span multiple kilometers. The latency requirements for data delivery are relatively relaxed *i.e.*, they are "delay-tolerant", however, the physical scope of data exchange is much larger. In contrast, general purpose Internet access requires connectivity to the backbone network via infrastructure, such as Road-Side Units (RSUs). Non-safety applications are expected to create new commercial opportunities by increasing market penetration of the technology and making it more cost effective. Moreover, comfort and infotainment applications aim to provide road travelers with needed information support and entertainment to make the journey more pleasant. They are so varied and ranges from traditional IP-based applications (*e.g.*, media streaming, voice over IP, web browsing, etc.) to applications unique to the vehicular environment (*e.g.*, point of interest advertisements, maps download, parking payments, automatic tolling services, etc.).

6.1 SAFETY RELATED VEHICULAR APPLICATIONS

VEC and the concept of utilizing smart vehicles as infrastructures have opened an arena of various associated vehicular applications, for example, driving safety, AR, infotainment services, and video streaming. For the applications where high computational processing is the demand, the VEC networks play an important role in rushing up computing, thereby minimizing the delay like if an accident takes place, we need to formulate a solution to reschedule traffic lights and to dissipate large traffic backlog in a suitable way. This has an exceptional demand in the computation resources. Applications have been divided into two groups, that is, safety applications and non-safety applications.

6.1.1 Classification of Vehicular Applications

Vehicular applications are typically classified in (*i*) active road *safety* applications, (*ii*) traffic efficiency and *management* applications, and (*iii*) *comfort* and *infotainment* applications. The first category aims to avoid the risk of car accidents and make safer driving by distributing information about hazards and obstacles. The basic idea is to broaden the driver's range of perception, allowing him/her to react much quicker, thanks to alerts reception through wireless communications. The second category focus on optimizing flows of vehicles by reducing travel time and avoiding traffic jam situations. Applications like enhanced route guidance/navigation, traffic light optimal scheduling, and lane merging assistance, are intended to optimize routes, while also providing a reduction of gas emissions and fuel consumption.

Finally, although the primary purpose of VANETs is to enable safety applications, non-safety applications are expected to create commercial opportunities by increasing the number of vehicles equipped with *on-board* wireless devices. Comfort and infotainment applications aim to provide the road traveler with information support and entertainment to make the journey more pleasant. In the next subsections we will describe the main aspects of *safety* and *entertainment* applications for VANETs.

The applications regarding *safety* are strictly tied to the main purpose of vehicles: moving from a point till to destination. Car collisions are currently one of the most frequent dead causes and it is expected that till 2020 they will become the third cause. This leads to a great business opportunity for infotainment, traffic advisory service, and car assistance.

Safety applications are always paramount to significantly reduce the number of accidents, the main focus of which is to avoid accidents from happening in the first place. For example, TrafficView and StreetSmart inform drivers through vehicular communications of the traffic conditions in their close proximity and farther down the road. Vehicle platooning is another way to improve road safety. By eliminating the hassle of changing

lane and/or adjusting speed, platooning allows vehicles to travel closely yet safely together. Fuel economy can also benefit from reduced aerodynamic as a vehicle headway is tightened (*e.g.*, the spacing can be less than 2 m). Together with adaptive cruise control assisted by V2V communications, the problem of vehicle crashes due to human error can be alleviated.

Some of the most requested applications by polls, currently under investigation by several car manufacturers are Post Crash Notification (PCN), Congestion Road Notification (CRN), Lane Change Assistance (LCA) and Cooperative Collision Warning (CCW). In the following, a brief overview of the above-cited applications is provided.

In PCN, a vehicle involved in an accident would broadcast warning messages about its position to trailing vehicles so that it can take decision with time in hand as well as to the highway patrol for asking away support. The PCN application may be implemented both on V2V and V2I network configurations. In fact the V2V presents the advantage of giving quickly the information through a *discover-and-share* policy. Through the use of specific sensors, it consists in measuring possible changes in the rational behavior of the driver (*e.g.*, quick brake use, rapid direction changes, and so on), which are then communicated back via directional antennas to the other vehicles along the same direction. Once received, the closest vehicle can share this information with the other nodes with a flooding routing. In the particular case of false alarm by the first vehicle experiencing the irrational behavior of the driver, this information floods on the VANET. It is then important to fix the issue of false alarms.

Let us suppose a driver has been distracted by something on the panorama and moves the steering wheel, so that the vehicle direction changes accidentally. Once recognized the error, the driver will react by quickly changing direction or with a quick and strong use of breaks. This behavior is not rational since there is no danger for the VANET community, but only the behavior of a single is irrational. This represents a false indication of alarm. If the first following driver does not experience some accidents, then

the vehicle does not forward this information, and false alarm probability is reduced, otherwise if it discovers the same problem, it shares such information with the other vehicles.

Dealing with the use of V2I architecture, the access points should gather information (*e.g.*, alarms for quick speed changes), coming from different vehicles, and merging the data so reducing the signaling from the vehicles. The V2V has the drawback of not allowing a quick communication if the vehicles are far away from each other (*e.g.*, in low traffic density scenarios), while the V2I is more energy consuming since it should be on all the time.

The LCA application constantly monitors the area behind the car when passing or changing lanes, and warns the driver about vehicles approaching from the rear or in the next lane over. This application has two different modalities, the first one is the so called passive mode, while the other one is the active mode. In the passive mode the vehicle simply measures distances, by means of detection and ranging procedures, while in the active mode it communicates to the other vehicles that they are too close, so they should change their direction / behavior.

Traffic monitoring and management are essential to maximize road capacity and avoid traffic congestion. Crossing intersections in city streets can be tricky and dangerous at times. Traffic light scheduling can facilitate drivers to cross intersections. Allowing a smooth flow of traffic can greatly increase vehicle throughput and reduce travel time. A token-based intersection traffic management scheme is presented, in which each vehicle waits for a token before entering an intersection. On the other hand, with knowledge of traffic conditions, drivers can optimize their driving routes, whereby the problem of (highway) traffic congestion can be lessened.

CRN detects and notifies about road congestions, which can be used for route and trip planning. This kind of application is partially implemented in current GPS-based applications where a new route is evaluated when heavy congestion has been detected on a route or in a portion of it. Up till now several commercial tools

are available for smart-phones and special purpose devices. These are currently based on GPS coordinates and local resident software able to indicate the shortest or fastest routes from a starting point till to a destination by considering one ways streets and so forth. A new generation of this kind of software integrates some control messages coming from the so-called Radio Data System-Traffic Message Channel (RDS-TMC) that gathers information about unavailable routes or congested streets. TMC messages contain a considerable amount of information:

- *Identification*: what is causing the traffic problem and its seriousness;
- *Location*: the area, road or specific location affected;
- *Direction*: the traffic directions affected;
- *Extent*: how far the problem stretches back in each direction;
- *Duration*: how long the problem is expected to affect traffic flow;
- *Diversion advice*: alternative routes to avoid the congestion.

The service provider encodes the message and sends it to FM radio broadcasters, who transmit it as an RDS (Radio Data System) signal within normal FM radio transmissions. There's usually only about 30 seconds between the first report of an incident to the traffic information Centre and the RDS-TMC receiver getting the message.

Also this application may be implemented according to a V2V configuration or a V2I one. In fact, it is possible to encapsulate information about the position, the direction, and the average speed, which are then communicated back to the vehicle following on the street the information. As it appears clear, this solution suffers for a large amount of data to be processed by the vehicles themselves. What is worth in this environment is the use of V2I since the access points can process information coming and communicate to the incoming vehicles the new route after request information about their destination. So, with a software that implements what is current available on the market (with a

special purpose processor in this case and without strict bounds on energy consumption for processing) it is possible to develop an instance of ITS.

Finally, the CCW system works with a cutout revealing a stopped car, or a stopped or slow-moving car before arrival at the curve or downhill. All these applications require radio transceivers for message exchange, GPS and sensor on board car and road infrastructure units. The dualism between V2V and V2I is renovated. Not so different from PCN, the behavior of the driver must be *understood* by the vehicles and then forwarded to the following cars and the vehicles coming with an opposite direction. This can be set up in a V2V modality since once recognized (*i.e.*, just happened) it is important to spread and flood this information. In the second phase, about 30 seconds or one minute, depending on traffic level, the information should be managed by the access point so to advice upcoming vehicles in time.

For what concerns *non-safety* applications, they have very different communication requirements, from no special real-time requirements of traveller information support applications, to guaranteed Quality-of-Service needs for multimedia and interactive entertainment applications. In general, this class of applications is motivated by the desire of the passengers to communicate either with other vehicles or with ground-based destinations (*e.g.*, Internet hosts or the Public Service Telephone Network (PSTN)). Also, various traveler information applications belong to this category. As an instance, the driver could receive local information regarding restaurants, hotels and, in general, Point of Interest, whenever the vehicle is approaching there (*i.e.*, *context aware* applications).

The aim of *infotainment* applications is to offer convenience and comfort to drivers and/or passengers. For example, Fleetnet provides a platform for peer-to-peer file transfer and gaming on the road.

A real-time parking navigation system is proposed in to inform drivers of any available parking space. Digital billboards for

vehicular networks are proposed in for advertisement. Internet access can be provided through V2I communications; therefore, business activities can be performed as usual in a vehicular environment, realizing the notion of mobile office. On-the-road media streaming between vehicles also can be available, making long travel more pleasant. An envisioned goal is to embed *human-vehicle-interfaces*, such as color reconfigurable head-up and head-down displays, and large touch screen active matrix Liquid Crystal Displays (LCDs), for high-quality video-streaming services. Passengers can enjoy their traveling time by means of real-time applications *e.g.*, video streaming and online gaming, using individual terminals next to their seats. Figure 1 (a) and (b) depict the use of LCD devices for entertainment applications.

(a) (b)

Figure 1. Video-streaming applications for passengers in a smart vehicle,

6.1.2 Non-Safety Applications

VEC applications stress not only upon the safety services but also on developing non-safety applications, for example, multimedia applications like video streaming, AR, and infotainment services. The number of streaming applications has increased significantly and these contribute a large proportion of the network traffic.

- *Video Streaming*. Video streaming is an essential form of IoT communication and it involves smartphones, for example, for video crowdsourcing. The Internet

of Vehicles (IoV) is used by various applications like intelligent transportation system and mobile multimedia. In IoV, the users connect their mobile phones through Internet connectivity and access the multimedia content from the remote servers. However, it is challenging to maintain the QoS while considering the parameters of the video streaming applications like jitter, buffering, throughput, and transmission delays. This challenge is faced due to the high mobility of vehicles in IoV. In a distributed reliable real-time streaming in vehicular cloud-fog networks was proposed. A utility function is considered to improve the QoS of real-time streaming as well as the fairness of resource reservation between mobile devices. The utility function takes into account the provision of content for streaming and the number of tokens for the reservation of content from the service providers, edge, and cloud. Every mobile device, in a network, inquires its probable location, the amount of data for streaming, and thus the needed tokens for content provisioning expected to effectively reserve streaming content from computing service providers. In this way, the mobile device will have a high probability to receive a sufficient amount of content by paying a certain number of tokens. Therefore, the streaming utility can be augmented, hence making it more reliable.

Now, it is quite possible to visualize innovative solutions to the parking lot monitoring problem. As, in, a scheme based on an edge computing is projected in which each vehicle uploads the contents of the street collected by the camera for video analytics. ParkMaster makes the system capable of estimating the precise location for each parked vehicle. It can also track and count parked vehicles through the information provided by the vehicle's camera, GPS, and inertial sensors.

- *Augmented Reality*. AR is an evolving multimedia application that incorporates real scenes into virtual scenes without a glitch and it can superpose the virtual scenes onto the real scenes and increase the conventional

real image information. This technology can augment the traffic awareness of the vehicles or pedestrians within the vicinity of the drivers. In addition, the head-up display curtails the distractions for the drivers and thus increases driving safety. The head-up-display (HUD) based navigation system with AR-based contents is examined. Likewise, the HUD-based navigation system has also been studied for safety and convenience services. Recently, an application is developed as walk navigation, which uses camera and GPS to execute a car navigation system with AR technology. It is very much convenient as it directs the driver through a virtual path. The driver is able to get the real-time navigation to get driving condition without hindering his safety.

The output of the device's camera is analyzed by the edge computing application, thereby covering the objects viewed with AR content. AR involves complex storing operations and tedious data processing tasks; therefore, it needs an increased level of data storage, computation, and communication. The VEC is regarded as the best alternative to fulfill the demand of AR applications in a vehicular network with specific requirements of mobility, location awareness, and low latency.

- *Infotainment Services*. These types of services help to entertain the passengers and keep them informed while navigating towards their destination. The infotainment services are made part of many research projects in order to enhance the comfort of the vehicle occupants. For instance, a European research project is developed while emphasizing on the establishment of Cooperative Vehicle-Infrastructure Systems (CVIS). To ensure uninterrupted V2V and V2I communication, the CVIS is expected to deliver a number of convenience and business applications and safety applications. Fleet-net is a German project and its prime goal is to provide a platform for V2V communications to provide a non-safety application (e.g., web access) and safety services (e.g., cooperative driving). TracNet is an automotive

vehicle Internet access system brought out by Microsoft. This project can bring internet services to the vehicular video screen. This can also turn the whole vehicle into an IEEE 802.11-based Wi-Fi hotspot; therefore, the handheld devices or laptops can also be connected to go online. Additionally, the aim of the infotainment services is to provide the travelers with the required information services and to entertain them to make the journey enjoyable, for example, maps downloads, automatic tolling systems, and parking payment systems.

6.1.3 Work Related Vehicle Safety (WRVS)

Work Related Vehicle Safety (WRVS) is the management of the hazards and risks associated with work activities involving vehicles and mobile equipment. This includes the risks to employers, self-employed people, employees and members of the public. WRVS encompasses both workplace transport safety and work related road safety.

- *Workplace Transport Safety (WTS)* is the management of hazards and risks associated with any vehicle or piece of mobile equipment that is used by an employer, employee, self-employed person or a visitor in a fixed or temporary workplace but excludes work related road safety.
- *Work Related Road Safety (WRRS)* is the management of the hazards and risks to persons engaged in or affected by work related driving or work activities on or near a road.
- *Driving for Work (DFW)* is the activity of driving on the road for work purposes. This includes the risk posed to workers themselves and those not at work who may be affected by the work activity, such as pedestrians and other road users. Driving for Work excludes commuting to work, except where the person's journey either starts from their home and they are travelling to a work location that is not their normal place of work or their journey involves travel in a company provided vehicle.

- *Working on or near a road (WNR)* is work activity carried out on or near a road and includes, for example, the safety of those working to maintain roads and street furniture, traffic wardens, engineers, emergency service personnel, road users and members of the public affected by the work activity.

6.1.4 Use of Infrastructure in VANETs

Vehicular ad hoc networks (VANETs) are created by applying the principles of mobile ad hoc networks (MANETs) – the spontaneous creation of a wireless network of mobile devices – to the domain of vehicles. VANETs were first mentioned and introduced in 2001 under "car-to-car ad-hoc mobile communication and networking" applications, where networks can be formed and information can be relayed among cars. It was shown that vehicle-to-vehicle and vehicle-to-roadside communications architectures will co-exist in VANETs to provide road safety, navigation, and other roadside services. VANETs are a key part of the intelligent transportation systems (ITS) framework. Sometimes, VANETs are referred as Intelligent Transportation Networks. They are understood as having evolved into a broader "Internet of vehicles". Which itself is expected to ultimately evolve into an "Internet of autonomous vehicles".

While, in the early 2000s, VANETs were seen as a mere one-to-one application of MANET principles, they have since then developed into a field of research in their own right. By 2015, the term VANET became mostly synonymous with the more generic term inter-vehicle communication (IVC), although the focus remains on the aspect of spontaneous networking, much less on the use of infrastructure like Road Side Units (RSUs) or cellular networks.

VANETs support a wide range of applications – from simple one hop information dissemination of, e.g., cooperative awareness messages (CAMs) to multi-hop dissemination of messages over vast distances. Most of the concerns of interest to mobile ad hoc networks (MANETs) are of interest in VANETs. Rather

than moving at random, vehicles tend to move in an organized fashion. The interactions with roadside equipment can likewise be characterized fairly accurately. And finally, most vehicles are restricted in their range of motion, for example by being constrained to follow a paved highway.

Example applications of VANETs are:

- *Electronic brake lights,* which allow a driver (or an autonomous car or truck) to react to vehicles breaking even though they might be obscured (e.g., by other vehicles).
- *Platooning,* which allows vehicles too closely (down to a few inches) follow a leading vehicle by wirelessly receiving acceleration and steering information, thus forming electronically coupled "road trains".
- *Traffic information systems,* which use VANET communication to provide up-to-the minute obstacle reports to a vehicle's satellite navigation system
- *Road Transportation Emergency Services* – where VANET communications, VANET networks, and road safety warning and status information dissemination are used to reduce delays and speed up emergency rescue operations to save the lives of those injured.
- *On-The-Road Services* – it is also envisioned that the future transportation highway would be "information-driven" or "wirelessly-enabled". VANETs can help advertise services (shops, gas stations, restaurants, etc.) to the driver, and even send notifications of any sale going on at that moment.

CVIS is a developed for the purpose to increase road safety and effectiveness and reduce the environmental impact of road safety. CVIS tests technologies to permit vehicles to communicate with each other and nearby road side.

- Development of standards for the vehicle-to-vehicle and vehicleto-vehicle infrastructure communication.

- Bringing more precision in the vehicle location and generation of more dynamic and accurate local maps using satellite navigation and other modern methods of location referencing.
- New systems for cooperative traffic and network monitoring in both vehicle and roadside infrastructure and to detect incidents immediately.
- Range of cooperative applications for traffic management, mobility services and driver assistance.
- Development of toolkits to address key deployment.

6.1.5 Vehicular Network Simulators

Vehicular networks have emerged as a result of advancementsin wireless technologies, ad-hoc networking, and the automobile industry. These networks are formed among moving vehicles, road side units (RSUs), and pedestrians that carry communication devices.

VANET Simulation Problem

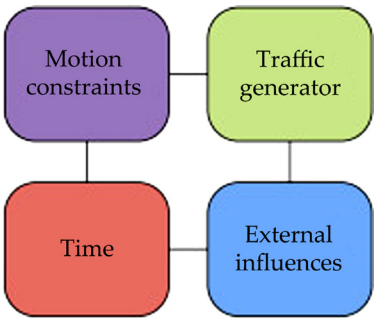
To evaluate VANET protocols and services, the first step is to perform an outdoor experiment. Many wireless technologies such as GPRS, IEEE 802.11p and IEEE 802.16 have been proposed for reliable traffic information. For test purpose software simulations can play a major role in imitating real world scenarios.

VANET relies on and is related to two other simulations for its smooth functioning, namely traffic simulation and network simulation. Network simulators are used to evaluate network protocols and application in a variety of conditions. The traffic simulators are used for transportation and traffic engineering. These simulations work independently but to satisfy the need of VANET, a solution is required to use these simulators together. There are a large number of traffic and network simulator and they need to be used together into what can be called VANET simulator. There are few tools for VANET simulation but most

of them have the problem of proper „interaction®. Therefore, comparison of present tools will suggest the choice with proper „interaction®. This paper consists of a survey of various traffic simulators, network simulators and VANET simulators resulting in the selection of a preferred recommended choice.

6.1.6 Mobility Models and Simulation Tools

It has been identified that the following attributes are important while considering any simulator for vehicular mobility modeling.



Mobility Model

- Accurate and realistic topological maps (intersections, lane and categories of streets with speed limits, etc.)
- Attraction/repulsion points (specifying source and destination points)
- Vehicle characteristics (heavy duty, light duty, emergency vehicles, etc.)
- Smooth deceleration and acceleration
- Human driving patterns (overtaking, traffic jam, preferred paths, etc.)

Intersection Management (Traffic lights, stop signs, obstacles, etc.): Based on control over the road traffic behavior and on the simulation performance, two types of modeling techniques are used: the Agent Centric and Flow Centric approaches. An Agent Centric approach is able to fully control individual vehicles with increased computational cost. However, the bidirectional interaction between the road traffic simulator and the network

simulator with a very low latency and with high accuracy is essential for developing applications such as traffic management or the Intelligent Road Traffic Signaling System (IRTSS). Modeling sophisticated applications and safety message dissemination protocols require full control on the road traffic mobility patterns.

6.1.7 Smart Vehicles

Vehicles will be equipped with multi interface cards, as well as sensors, both on board and externally. With an increasing number of vehicles equipped with on-board wireless devices (*e.g.*, UMTS, IEEE 802.11*p*, Bluetooth, etc.) and sensors (*e.g.*, radar, ladar, etc.), efficient transport and management applications are focusing on optimizing flows of vehicles by reducing the travel time ad avoiding any traffic congestions. As an instance, the on-board vehicle radar could be used to sense traffic congestions and automatically slow the vehicle. In other accident warning systems, sensors are used to determine that a crash occurred if air bags were deployed; this information is then relayed via V2V or V2I within the vehicular network.

Forgetting traditional vehicles, in the next few years we will drive *smart —intelligent— vehicles*, with a set of novel functionalities (*e.g.*, data communications and sharing, positioning information, sensor equipment, etc.). It is then necessary that for specific applications (*i.e.*, safety messages and alerts, gossip-based applications, etc.) the majority of mobile vehicles within a vehicular network be equipped with *on-board* wireless device, namely On-Board eqUipment (OBU).

A number of systems and sensors are used to provide different levels of functionality. Among the major systems and sensors exploited for intra-vehicle communications we cite: the braking system, crash sensors, the data recorder, the engine control unit, the electronic stability control, the electronic steering, the infotainment system, the integrated starter generator, the lighting system, the power distribution and connectivity, seat belt sensors, the tire pressure monitoring system, etc. Particularly, for the

brake systems, there are also the parking brake and the antilock brake system. The parking brake, which is also referred to as an emergency brake, controls the rear brakes through a series of steel cables. This allows the vehicle to be stopped in the event of a total brake failure. Moreover, also vehicle-mounted cameras are largely used to display images on the vehicle console.

Commonly, a smart vehicle is equipped with the following devices and technologies: (*i*) a Central Processing Unit (CPU) that implements the applications and communication protocols; (*ii*) a wireless transceiver for data transmissions among vehicles (V2V) and from vehicles to RSUs (V2I); (*iii*) a Global Positioning Service (GPS) receiver for positioning and navigation services; (*iv*) different sensors laying inside and outside the vehicle to measure various parameters (*i.e.*, speed, acceleration, distance from neighboring vehicles, etc.); (*v*) an input/output interface for human interaction with the system.

The basic idea of *smart vehicles* is addressed to safety issues, and then by a proper combination of functionalities like *control*, *communications*, and *computing* technologies, it will be possible to assist driver decisions, and also prevent wrong driver's behaviors. The *control* functionality is added directly into smart vehicles to connect the vehicle's electronic equipment. The technology used for control should take into account the need of limit vehicle weight; as a matter, the added wiring increases vehicle weight, and weakens performance. It has been proven that for an average well-tuned vehicle, every extra 50 kilograms of wiring —or extra 100 W of power— increases fuel consumption by 0.2 liters for each 100 kilometers traveled.

Based on such considerations, today *control* and *communications* in a vehicular ad hoc network counter the problems of large amounts of discrete wiring. In the following Figure 2 we show the sheer number of systems and applications contained in a modern vehicle's network architecture.

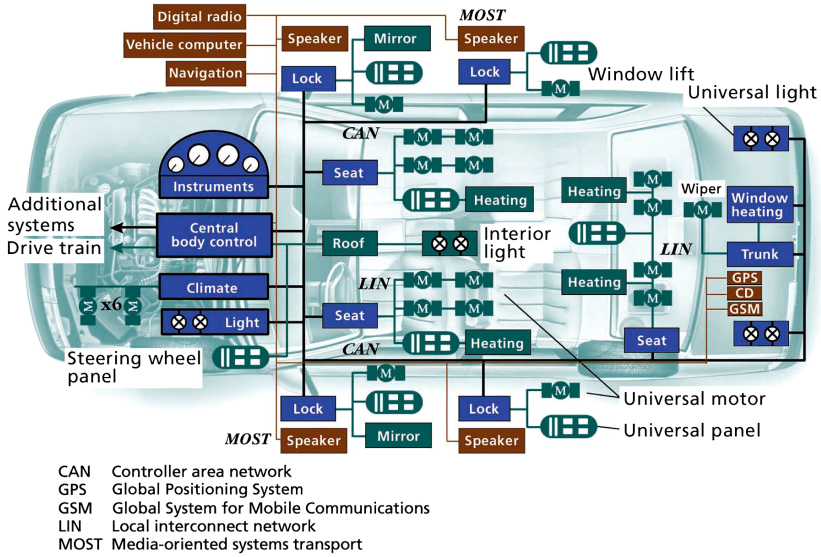


Figure 2. Design of a modern vehicle's network architecture.

In the mid-1980s, Bosch developed the Controller Area Network (CAN), one of the first and most enduring automotive control networks, and now being used in many other industrial automation and control applications. CAN (ISO 11898) is currently the most widely used vehicular network, with more than 100 million CAN nodes sold in 2000.

CAN is a high-integrity serial data communications bus for real-time applications, operating at data rates of up to 1 Mbit/s and having excellent error detection and confinement capabilities. A typical vehicle can contain two or three separate CANs operating at different transmission rates. A low-speed CAN running at less than 125 Kbps usually manages a car's "comfort electronics," like seat and window movement controls and other user interfaces. Generally, control applications that are not real-time critical use this low-speed network segment. Low-speed CANs have an energy-saving sleep mode in which nodes stop their oscillators until a CAN message awakens them. Sleep mode prevents the battery from running down when the ignition is turned off. A higher-speed CAN runs more real-time critical functions such as engine management, antilock brakes, and cruise control. Although

capable of a maximum baud rate of 1 Mbps, the electromagnetic radiation on twisted-pair cables that results from a CAN's high-speed operation makes providing electromagnetic shielding in excess of 500 Kb/s too expensive.

CAN is a robust, cost-effective general control network, but certain niche applications demand more specialized control networks. For example, X-by-wire systems use electronics, rather than mechanical or hydraulic means, to control a system. These systems require highly reliable networks.

In 2011 a novel enhanced version of CAN, called *CAN with Flexible Data-Rate* (CAN FD), supports payloads higher than 8 byte per frame. CAN FD protocol controllers are also able to perform standard CAN communication: this allows the use of CAN FD in specific operation modes, like software download at end-of-line programming, while other controllers that do not support CAN FD are kept in standby. In automotive electronics, engine control units, sensors, anti-skid-systems, etc. are connected using CAN. At the same time, CAN is cost effective to build into vehicle body electronics, *e.g.* lamp clusters, electric windows etc. to replace the wiring harness otherwise required.

A well-known communication system designed for automotive applications is the *FlexRay Communications System i.e.*, a robust, scalable, deterministic, and fault-tolerant digital serial bus system. The core concept of the FlexRay protocol is a time-triggered approach to network communications. This is a different approach to some earlier successful networking schemes. Indeed, FlexRay is an option for upgrading existing network systems using CAN in the automotive industry.

It could be useful for applications, where safety and reliability in a work environment is of most importance due to its deterministic approach and two channel topologies. Due to its high data rate of 10Mb/s over each of its two channels, this protocol suits as the basis of a network backbone. The FlexRay protocol developed by the FlexRay consortium has already found applications in the automotive industry and looks set to become the network scheme

favored especially in x-by-wire applications and other safety critical systems. There is on-going research into the migration from CAN based systems to FlexRay based systems and as such the protocol could find itself being used in many areas outside the automotive industry. With its deterministic time-triggered approach and the high data rates achievable it is also suitable for safety and control applications.

In recent past, more number of multimedia and telematics applications has been integrated into premium class vehicles. These include Sound system, CD player, navigation systems, video players, voice input. High bandwidth requirement of all these applications has led to the development of infotainment communication system, *i.e. MOST* (Media Oriented Systems Transport). MOST is the de-facto standard for *multimedia* and *infotainment* networking in the automotive industry? The technology was designed from the ground up to provide an efficient and cost-effective fabric to transmit audio, video, data and control information between any devices attached even to the harsh environment of an automobile.

MOST technology is the result of the collaboration between car makers and suppliers, working to establish and refine a common standard within the MOST cooperation. MOST Cooperation was founded in partnership by BMW, Daimler Benz, Becker and OASIS silicon system. Now SMSC is a leading provider of MOST, the de-facto standard for high bandwidth automotive multimedia networking.

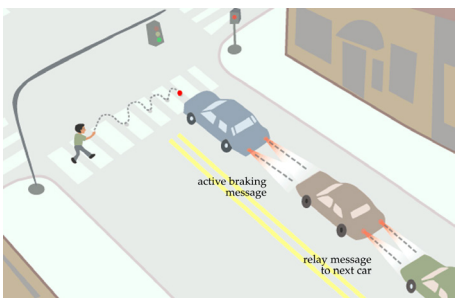


Figure 3. Safety application (i.e., brake messaging) by using VLC devices.

Finally, our attention will be focused on the use of Visible Lighting Communications (VLC) can provide a valid technology for communication purposes in VANETs. The use of the visible spectrum provides service in densities exceeding femtocells for wireless access. It represents a viable alternative that can achieve high data rates, while also providing illumination. This configuration minimizes packet collisions due to Line Of Sight (LOS) property of light and promises to alleviate the wireless bottleneck that exists when there is a high density of rich-media devices seeking to receive data from the wired network.

Possible applications of VLC impact the quality of life, including control of auto / traffic signaling for safety and enabling communications where high noise interferes with WiFi. However, the main issues are related to medium-range LED-based communication derived from outdoor lighting (*i.e.*, sunlight and interference illuminations) that is common in public and private infrastructure (*e.g.*, street, building, and signage illumination).

It is also common on vehicles and in traffic infrastructure including car, rail and air transportation. Specific scenarios include V2V, V2I, as well as V2X communications, all of them supporting goals of improved safety, reduced carbon emissions, energy conservation, enhanced connectivity and network performance.

Although LEDs are commonly used in automotive and infrastructure lighting (*i.e.*, brake and traffic lights, as depicted in Figure 3), there remain key challenges to achieving effective modulation and communication between devices, especially while they are moving or in the presence of sunlight. Enabling communications in mobile outdoor systems, particularly in dense, fast moving safety-critical automotive environments is one of the main benefits of VLC for VANETs. In vehicular applications, mobile communications are particularly suitable for adoption of directional communications using LOS links. Applications such as safety and emergency messaging require very high reliability, and this can be provided through short-range inter-vehicular communications. As an instance, vehicles can be equipped with optical transceivers, such that they can communicate with other similarly equipped

vehicles. As a practical example, it is easy to understand how through the use of electronic or light pulse or radar, it is possible to measure the round-trip delay when pulses bounce off an object. For decreasing delays, the distance between the vehicle and any obstruction in the roadway decreases as well. This information can be used to automatically adjust the speed of the vehicle and inform the driver of the potential occurrence of a frontal collision condition (see Figure 3).

6.1.8 Technologies in Vehicular Ad hoc Networks

Several technologies are involved in Vehicular Ad hoc Networks, especially as enablers of Intelligent Transportation Systems (ITS). These are GSM, UMTS, Wi-MAX limited Wi-Fi and a new and specific technology thought for this kind of applications, namely Wireless Access in Vehicular Environments (WAVE), also known as IEEE 802.11/*p*. This implicitly suggests that a car should have on board different radio interfaces (and/or network card). About WAVE, it is member of the IEEE 802.11 family, this implicitly suggests that this solution (currently at the stage of draft) is borrowed from IEEE 802.11 and adapted for the vehicular context.

Recent advances in the area of ITS have developed the novel Dedicated Short Range Communication (DSRC) protocol, which is designed to support high speed, low latency V2V, and V2I communications, using the IEEE 802.11*p* and WAVE standards. In 1999, the Federal Communication Commission (FCC) allocated a frequency spectrum for V2BV and V2I wireless communication. DSRC is a communication service that uses the 5.850-5.925 GHz band for the use of public safety and private applications.

The allocated frequency and newly developed services enable vehicles and roadside beacons to form VANETs in which the nodes can communicate wirelessly with each other without central access point. Specifically, the communication is in the bandwidth 5.850 GHz—5.925 GHz, so allowing a band of 70 MHz with some guard bands. This band is partitioned in 7 different sub-bands presenting a bandwidth of 10 MHz. The first channel is for V2V

communication with public safety purposes while the second and third channels are private channels and used for public safety too in a medium range environment. The fourth is a control channel while the fifth and sixth channels are for public safety services with short range. The seventh channel is dedicated to manage public safety intersections.

Data-rates offered by VANET strictly depend on the kind of service and its own specifications. As an example the smaller data rate is for toll and payment services (highways) where the transmission ate is of the order of few Mb/s at tens of meters. The same distance range is for the Internet access even if the required rate can rise till to 54 Mb/s. Safety message service should allow proactive actions so the range is much higher, of the order of hundreds of meters and the required rate is below 20 Mb/s, down to 6 Mb/s. Finally, services regarding emergency vehicles require rate of the order of 5 Mb/s at a very high distance with respect to previous ones (*e.g.*, 3000 m).

The worldwide ISO TC204 / WG16 has produced a series of draft standards, known as CALM (Continuous Air-Interface, Long and Medium Range). The main goal of CALM is to develop a standardized networking terminal, able to seamless connect vehicles and roadside systems, avoiding disconnections. This can be well accomplished through the use of a wide range of communication devices and networks, such as mobile terminals, wireless local area networks, and the short-range microwave (DSRC) or infrared (IR).

The CALM architecture separates service provision from medium provision via an IPv6 networking layer, with media handover, and will support services using 2G, 3G, 5 GHz, 60 GHz, IEEE 802.16*e*, IEEE 802.20, etc. A standardized set of air interface protocol is provided for the best use of resources available for short, medium and long-range, safety critical communications, using one or more of several media, with multipoint (mesh) transfer.

The CALM concept is now at the core of several major EU sixth framework research and development projects. In the United

States, the Vehicle Infrastructure Integration (VII) initiative will be operating using IEEE 802.11*p* / 1609 standards at 5.9 GHz, which are expected to be aligned with CALM 5.9-GHz standards, although the IEEE standards do not have media handover.

Due to the recent strides made in VANETs, a new class of in-car entertainment systems and enabling emergency services using opportunistic spectrum has increased by means of Cognitive Radio (CR) technology. These CR-enabled Vehicles (CRVs) have the ability to use additional spectrum opportunities outside the IEEE 802.11*p* specified standard band.

The growing spectrum-scarcity problem, due to the request of high-bandwidth multimedia applications (*e.g.*, video streaming) for in-car entertainment, and for driver-support services, such as multimedia-enabled assistance, for opportunistic spectrum use, which directly benefits various forms of vehicular communication. In such a network, each CRV implements spectrum management functionalities to (*i*) detect spectrum opportunities over digital television frequency bands in the Ultra-High Frequency (UHF) range, (*ii*) decide the channel to use based on the QoS requests of the applications, and (*iii*) transmit on it, but without causing any harmful interference to the licensed owners of the spectrum.

CRVs have many unique characteristics that involve additional considerations than merely placing a CR within a vehicle. As an example, unlike static CR systems, the spectrum availability perceived by each moving vehicle changes dynamically over time, as a function not only of the activities of the licensed or Primary Users (PUs) but also based on the relative motion between them. Spectrum measurements need to be undertaken over the general movement path of the vehicles, leading to a path-specific distribution, instead of focusing on the temporal axis alone.

The CRV network can also leverage the constrained nature of motion, *i.e.* along linear and predecided paths corresponding to streets and freeways. At busy hours or in urban areas, spectrum information can be exchanged over multiple cooperating vehicles, leading to know more about the spectrum availability.

This also allows the vehicles that follow to adapt their operations and undertake a proactive response, which is infeasible in both static and non-stationary scenarios with random motion. The CRV networks fall under three broad classes, such as (*i*) V2V only, (*ii*) V2I only, and (*iii*) centralized V2I. The first class, a network can be formed between vehicles only that rely on cooperation for increasing accuracy. The second class deals with periodic interactions between vehicles and roadside BSs, where the latter acts as a repository of data that is subsequently used by passing vehicles. Finally, a completely centralized network is possible, in which the BS autonomously decides the channels to be used by the CRVs, without relying on information from the vehicles.

The access problem can be solved on two different layers. The first access problem is the selection of the network providing the service. The second one is the access within the selected network. This is mainly true for the V2I environment. Once specified the network quality metric, the vehicle should select the best network (this is the principle of *vertical handover*). Regarding the V2V communications, requiring network synchronization appears complicated so static access procedure x-DMA usage is by fact discouraged. Dynamic access best suit the typical channel features of the multi-hop network so Carrier Sensing Multiple Access – Collision Avoidance may be used. In this context the lack of synchronization at network level is not dramatic and requires only a node-by-node synchronization.

About routing procedures, these are a key point since, especially in the V2V scenario, each kind of message generated after a sensing action should be forwarded, in principle, to all the interested vehicles. In the V2I environment, the routing is not so critical even if vertical handover procedures should be considered.

Regarding V2V connections some routing *philosophies* can be considered. These are Geo-Broadcast, when a node send to all its neighbors an update about a region, Geo-any cast when a vehicle interrogates other nodes about road status and Fleetnet Routing, when a Greedy approach is used, that is, each node tries to forward the information according to a metric (variation of flooding) and

it can be implemented via a beacon-based scheme that requires to each node to periodically transmit its position. It this last the positioning is really important and it can be derived or via an absolute service —like GPS— or by triangulation, so requiring more signaling.

The use of GPS (and, more in general, the GNSS) unit within the vehicles allows knowing the vehicles' positions. The awareness of precise locations is very important to every vehicle in VANET so that it can provide accurate data to its neighbors. Currently, typical localization techniques integrate GPS (GNSS) receiver data and measurements of the vehicle's motion.

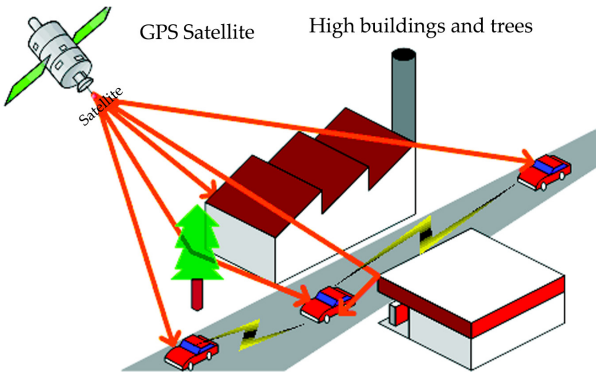


Figure 4. The use of satellite (GPS system) for outdoor localization. However, multipath effect affects the accuracy.

GPS is a positioning system developed and operated by the U.S. Department of Defence. A GPS system is formed from a network of satellites that transmit continuous coded information, which makes it possible to identify locations on Earth by measuring distances from the satellites. At the same time, the receiver has the ability to obtain information about its velocity and direction.

With respect to VANET, many techniques have been proposed to the use of GPS as a localization technique, as shown in Figure 4. However, in many applications a simple GPS receiver is not a satisfactory tool for location estimation (*e.g.*, to discriminate vehicles between those vehicles on a particular highway and others outside the highway), also due to the multipath effect,

specially affecting urban areas. Such a system requires highly accurate location estimation. Solutions integrating a GPS with an Inertial Navigation System (INS) can increase the accuracy of the localization application. Also augmented GPS solutions like Differential GPS are largely used for increasing accuracy.

Finally, hybrid approaches comprising of both V2V and V2I mode are largely used in order to improve network performance, while limiting the packet forwarding delay. For example, the ubiquitous integration of existing high-speed WLANs with wide-range 3GPP Long Term Evolution (LTE) results in the service extension of the backbone cellular network. LTE is the upcoming 4G cellular network with high data rate support for multimedia services, and robustness to high speed. The use of small cells will be massively deployed for increasing coverage areas; as a result they can be good candidates for V2I communications due to their reduced cost.

Novel solutions exploit the use of LTE technology in VANETs. The authors propose LTE4V2X approach, for the framework of a centralized vehicular network, whose effectiveness has been proven with respect to decentralized protocols. LTE4V2X uses both the IEEE 802.11*p* and 3GPP, LTE to provide an efficient way to periodically collect messages from vehicles and send them to a central server. As a result, the use of heterogeneous wireless network architectures achieves seamless data connectivity among separated vehicular clusters.

6.1 TRAFFIC CONTROL

Traffic control, supervision of the movement of people, goods, or vehicles to ensure efficiency and safety. Traffic is the movement of people and goods from one location to another. The movement typically occurs along a specific facility or pathway that can be called a guideway. It may be a physical guideway, as in the case of a railroad, or it may be an agreed-upon or designated route, marked either electronically (as in aviation) or geographically

(as in the maritime industry). Movement—excepting pedestrian movement, which only requires human power—involves a vehicle of some type that can serve for people, goods, or both. Vehicle types, often referred to as modes of transportation, can be broadly characterized as road, rail, air, and maritime (*i.e.*, water-based).

Traffic evolves because of a need to move people and goods from one location to another. As such, the movement is initiated because of decisions made by people to transport themselves or others from one location to another to participate in activities at that second location or to move goods to a location where they have higher value. Traffic flows thus differ fundamentally from other areas of engineering and the physical sciences (such as the movement of electrons in a wire), because they are primarily governed and determined by laws of human behavior. While physical attributes are critical in the operation of all modes (*e.g.*, to keep airplanes in the air), the demand or need to travel that gives rise to traffic is derived from the desire to change locations.

One of the principal challenges in traffic control is to accommodate the traffic in a safe and efficient way. Efficiency can be thought of as a measure of movement levels relative to the objectives for a particular transportation system and the finances required for its operation. For example, a railroad can be thought of as efficient if it can accommodate the travel requirements of its customers at the least cost. It will be thought of as inefficient if an alternative (*e.g.*, a trucking service) can also meet customer needs but at a lower cost.

Safety, the management of traffic to reduce or eliminate accidents, is the other critical reason for traffic control. An airline pilot needs

to be warned of high winds at the destination airport just as an automobile driver needs to be warned of a dangerous curve or intersection ahead. Traffic control has as its principal objective to manage the movement of people and goods as efficiently and safely as possible. The dual objectives, however, frequently conflict or, at least, compete. For example, there are frequent cases in which commercial airlines are held on the ground at their originating airport until they receive a clearance to land at a destination. The clearance is given only when the destination airport determines that the number of airplanes expected to arrive at a particular time is small enough that local air traffic controllers can assist the plane in landing without overtaxing their human limitations and compromising safety.

In road traffic, intersections with traffic lights (*i.e.*, green, amber, and red indications) will often add a separate lane with a lighted green arrow to allow left turns with no opposing traffic. This frequently results in longer non-Greek periods at the intersection, causing an increased delay and a reduction in efficiency and mobility. Traffic control will always be burdened with seeking to satisfy the frequently conflicting goals of safety and mobility.

Safety is not the exclusive concern of the traffic control community. Nearly every transportation mode has organizations that regulate operators through a series of licensing procedures, sanctions for inappropriate operating practices, and requirements for continuing training to retain certification to operate. Examples include federal aviation authorities that oversee pilot training (*e.g.*, the U.S. Federal Aviation Administration); road agencies that administer driver's licenses may exist at the provincial level (as in Canada) or at the national level (as is more common in Europe). Transportation safety management is thus accomplished through a complex set of interactions between different agencies at different levels (*e.g.*, national, regional or state, and local) using both formal legal requirements and administrative actions. The following discussion will necessarily focus on safety concerns that evolve from and are a component of the traffic control function.

6.2.1 Overview of Traffic Control

Traffic control is a critical element in the safe and efficient operation of any transportation system. Elaborate operational procedures, rules and laws, and physical devices (*e.g.,* signs, markings, and lights) are but a few of the components of any traffic control system. At the centre of any system is the operator: a driver or pedestrian in a roadway system, a pilot in aviation or maritime systems, and a locomotive engineer in railway systems. While traffic control can be considered initially as a need to control or influence large numbers of vehicles, it is important to realize that traffic is made up of a large number of individual operators who collectively must make consistent decisions in order for the systems to work safely and efficiently.

The operator is the principal decision-making unit in any traffic control system. As such, the entire system is organized to assure the safe and efficient movement of vehicles along a guideway or separational infrastructure by providing adequate, accurate, timely information to the operator. The operator accepts inputs from a variety of sources, enters into a decision-making process, and determines the appropriate control actions to maintain vehicle operation.

The operator receives most immediate and direct information from the vehicle. In addition to visual inputs regarding vehicle status that are provided by instrumentation (*e.g.,* speed, direction), the operator receives information through physical sensation of movement (*i.e.,* through forces acting on the muscles and sensory organs). The slowing and turning of a vehicle, for example, are sensed not only visually but also physically by the operator's body as the vehicle decelerates and changes course. Different vehicles have vastly different performance characteristics that directly affect the physical forces acting on an operator. An automobile is highly responsive and gives virtually immediate response (certainly less than a second) to braking or steering inputs. A large vessel or airplane, because of its design and the "guideway" in which it operates, is slow (on the order of minutes) to respond to

steering or speed change inputs. Small aircraft and boats, however, have response attributes much more similar to an automobile than to their larger counterparts.

In addition to vehicle inputs, the operator's decision making is influenced by the information provided by the guideway and its associated infrastructure. Because infrastructure is man-made, it is one of the places where proper design and procedures provide an important foundation for operating safety. For example, roadway systems set precise standards for the size, shape, color, and use of road signs and markings. These standards have the goal of improving road safety and efficiency by providing the driver with consistent information regarding hazards, control of right-of-way (*e.g.*, stop signs or signals), and direction guidance (*e.g.*, "Highway 66 next left"). Aviation, maritime, and rail systems also have elaborate standards, all with one goal in mind: to reduce accidents and increase efficiency through the consistent and effective use of standard traffic control devices. Clearly aviation, and to some degree maritime, systems cannot place physical signs in the sky or sea. Electronic signs or signals, particularly communication devices, are used instead to guide the vehicle and operator. The guideway includes the attributes of the physical infrastructure upon which the vehicle operates (*e.g.*, a roadway for automobiles, trucks, bicycles, and pedestrians or a set of rails for trains). There are similar corridors within which planes and ships operate, although they are not defined by physical elements so much as geographic location (*i.e.*, longitude and latitude, and altitude for aviation). The ambient environment poses both direct and indirect limitations on an operator's ability to control a vehicle. Snow, rain, sleet, fog, and darkness all serve to limit visibility. Electronic devices such as radar are particularly helpful in aviation and marine contexts in providing supplementary information that allows operators to make safe and efficient control decisions.

6.2.2 Road Traffic Control

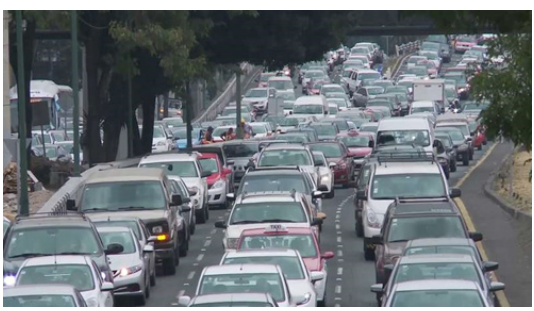
At the broadest level, road traffic control includes the layout of streets to serve a variety of travel needs in a region. Highways or

expressways carry through traffic at high speed; arterial streets carry traffic within and across urban areas; and local streets provide low-speed travel but access to many local destinations. The hierarchy of streets that perform at different levels of speed and provide different levels of access form the foundation upon which traffic control problems evolve. Long delays and frequent accidents are common outcomes of inadequate road planning, which results in an insufficient number of roads to meet travel needs. While traffic control may help, it is not a substitute for adequate provision of transportation supply.

History

Traffic congestion, often bad enough to require drastic control measures, was a feature of city life at least as early as Roman times. A basic cause, then as now, was poor city planning, with roads laid out in such a way as to bring traffic from all quarters to a central crossing point. In the 1st century BC Julius Caesar banned wheeled traffic from Rome during the daytime, a measure gradually extended to cities in the provinces. Late in the 1st century AD the emperor Hadrian was forced to limit the total number of carts entering Rome.

About 1500 Leonardo da Vinci, envisioning a revolutionary solution to urban traffic problems—then acute in the crowded and busy Italian cities—proposed separating wheeled and pedestrian traffic by creating routes at different levels. Except for the railway, however, few segregated route systems were established before the 20th century.



Congestion was severe enough in European cities of the 17th century to require ordinances prohibiting parking on certain streets and establishing one-way traffic. The advent of the railroad brought temporary relief to the growing problem of road traffic control, though it created congestion at terminals inside cities. The automobile, with its increase first in speed and then in numbers over horse-drawn transport, rapidly created a new situation that was to become one of the characteristic problems of urban industrialized society in the 20th century.

Traffic Elements

Road traffic control at its most elemental level is achieved through the use of a system of signs, signals, and markings. Elaborate engineering standards are used to assure that the traffic control devices convey a clear and simple meaning to the motorist. A comparable and matching education program is needed, through driver-licensing authorities, to assure that those who operate motor vehicles understand the rules of the road and the actions that they are required or advised to take when a particular control device is present.

Each traffic control device is governed by standards of design and usage; for example, stop signs always have a red background and are octagonal in shape. Design standards allow the motorist to quickly and consistently perceive the sign in the visual field along the road. Standard use of colors and shape aids in this identification and in deciding on the appropriate course of action.

Standards also exist on the use of the control device, such as guidelines as to when circumstances warrant the use of two-way stop signs or traffic signals. Standards also are used to locate control devices in a particular circumstance. For example, signs on high-speed expressways or motorways need to be placed well in advance of exits to allow sufficient time for drivers to choose a course of action. Standards for location allow drivers to expect and anticipate these devices at certain distances from decision points. Adhering to these standards promotes safety; failure to adhere

increases the risk of driver error and, ultimately, accidents. The design and use of traffic control devices must also recognize the tremendous mix of vehicles that use highway systems. The devices must be useful for pedestrians and bicyclists as well as drivers of 80,000- to 120,000-pound trucks that are up to 100 feet long. It is not the size and weight differences per se that are important but what they imply for vehicle performance. On a road that is heavily used by trucks, for example, the location of warning signs for a dangerous intersection must be placed sufficiently in advance to compensate for a truck's longer stopping distance (as compared to that of a car). Design of devices such as guardrails must take into account the larger mass and higher Centre of gravity of trucks as well. Because trucks serve so many purposes, highly specialized vehicles have evolved to meet different needs. While principles of standardization would indicate a desire to limit the type and configuration of trucks in use, characteristics of transportation markets often lead to specialized vehicle developments. The conflict between standardization and market need, and the sheer size and bulk of many trucks, has led to a series of controversies concerning their safety performance. As long as private, personally owned automobiles must share roadway space with very large commercially owned trucks, the conflicts and controversy are likely to continue.

Common Control Techniques

Traffic signal controllers are electronic devices located at intersections that control the sequence of the lights. Along with computers, communications equipment, and detectors to count and measure traffic, the controllers are frequently grouped together to control large numbers of traffic signals, either at intersections in a city or on ramps approaching expressways and motorways. While the detailed brand and type of equipment vary greatly, the functions performed by the systems are generally consistent.

There are four basic elements in a computerized traffic control system: computer(s), communications devices, traffic signals and associated equipment, and detectors for sensing vehicles.

Traffic flow information is picked up by the detectors from the roadway and transmitted to the computer system for processing. The detectors are normally embedded in or suspended above the roadway. Vehicle counts and speeds are typically measured; vehicle type (*e.g.*, auto or truck) also may be obtained. The computer processes the traffic flow data to determine the proper sequence for the lights at the intersections or ramps. The sequencing information is transmitted from the computer through communications equipment to the signals. In order to assure safe and proper operation, information is also transmitted from the traffic signals to the computer, confirming proper operation. Humans can interact with the system by accessing the computer system in some way.

While these are the general principles, important variations are possible. First, it is common to find some form of computer as part of the traffic signal at the intersection or ramp to be controlled. This allows the local computer to process traffic flow data directly, reducing communications needs and costs. Another variation is that selected vehicles themselves may transmit traffic data directly to the computer system. This is frequently combined with the ability to receive information in the vehicle regarding points of congestion, so the driver can choose to avoid them. If the two-way communication exists between the vehicles and computer system, it may not be necessary to have separate physical detectors.

Another area of application for traffic control devices is their use in traffic restraint (often called traffic "calming"). Rather than use traffic control to increase efficiency of movement, controls are used to create impediments that restrain traffic from sensitive areas. Most commonly applied in older cities whose road network does not match current needs, traffic restraint seeks to funnel traffic onto particular routes by creating impediments to movement on others. These other routes typically have some special value—a historic site or a residential character—that requires protection. Devices typically used include speed bumps, barricades to block streets, turn prohibitions, stop signs, and raised pavement markers.

Traffic restraint also includes programs to foster bicycle and pedestrian travel. Wider sidewalks, sometimes including tables and benches, and bicycle lanes frequently accompany restraint actions. These programs recognize that what is good for vehicular travel may not necessarily be positive for other road users, the environment, or the neighborhood. An unfortunate aspect of these programs is that their benefits and costs are highly localized. Those living on the "right" side of the restraint device generally experience slow speed and lower traffic volume. Those living along the routes onto which the traffic is funneled must endure increased vehicle volumes and speeds.

Traffic control also can be used to give priority to high-occupancy passenger modes. The objective of such actions is to emphasize people rather than vehicle movement. A variety of techniques are available and are employed in priority treatment approaches. The most common is the dedication of special lanes to the use of priority, or high-occupancy, vehicles. Buses and car pools can use the lanes to move at high speeds along congested expressways and motorways, bypass queues at expressway ramps, and move along congested arterial streets. Because these special lanes are designed to operate uncongested, they provide an incentive, through reduced travel times, for travelers to leave private single-passenger automobiles and travel by multipassenger modes. Buses also may be given priority by allowing only them to turn at intersections and to be provided with extra green time at a traffic signal. The undesirable feature of such systems is that they provide improved service to high-occupancy modes while sustaining or increasing congestion for others. The residual congestion for other road users may result in continued wasteful fuel consumption and high vehicle pollutant emission.

New concepts

Rapid and continuous advances in communications and computer technology are spurring a host of new concepts in road traffic control. Automobiles equipped with on-board computers, driver displays, and communications devices will receive instructions

about the optimal path to a destination from a traffic control Centre. The vehicle also will periodically report its travel time and speed to be used as part of the information for the computer to give advice. In more advanced systems, the timing of traffic signals at intersections and ramps will be coordinated with the routing advice. Rather than simply accommodating vehicles that travel through the network, the system will cause patterns of travel to be altered. Computers and sensors within the vehicle will monitor the operation of critical safety systems (*e.g.*, brakes, steering), warning the driver when conditions exceed nominal values.

Communications and computers also will aid the movement of trucks and other commercial vehicles in urban areas. A dispatcher will be able to alter the schedule while the driver is on the road. For these companies, this means reduced costs, and for their customers, improved service. Drivers on long-distance intercity trips can be warned of impending bad weather. They also can receive warnings if they are entering a curve too quickly or an intersection too fast. Road safety should be greatly enhanced by such systems.

Public transit users should be able to receive more accurate information concerning travel time and seat availability on buses and trains. If accurate information can be provided in the home or office, the systems can spread peak loads, making service less expensive to provide and the trip more comfortable for the traveler. Those who are informed of congestion or uncomfortable conditions can use another system to find a match for a car pool participant. Alternatively, individuals may "telecommute," staying at home and working with their office electronically.

Lastly, the ultimate system is viewed as an automatic vehicle-control system in which a driver's vehicle is checked at an authorized station, then proceeds on a highway, lane, or local street. The spacing to the vehicle ahead and lateral control within the lane are determined by on-board computers. Maximum flows are expected to increase from 2,000 vehicles per hour per lane to as many as 10,000 to 20,000. The increased flows will mean substantial

reductions in congestion and, because vehicles are automatically controlled, improvements in road safety through the elimination of accidents due to driver error.

6.2.3 Air Traffic Control

Air traffic control (ATC) is a service provided by ground-based air traffic controllers who direct aircraft on the ground and through controlled airspace, and can provide advisory services to aircraft in non-controlled airspace. The primary purpose of ATC worldwide is to prevent collisions, organize and expedite the flow of air traffic, and provide information and other support for pilots. In some countries, ATC plays a security or defensive role, or is operated by the military.

Air traffic controllers monitor the location of aircraft in their assigned airspace by radar and communicate with the pilots by radio. To prevent collisions, ATC enforces traffic separation rules, which ensure each aircraft maintains a minimum amount of empty space around it at all times. In many countries, ATC provides services to all private, military, and commercial aircraft operating within its airspace. Depending on the type of flight and the class of airspace, ATC may issue instructions that pilots are required to obey, or advisories (known as flight information in some countries) that pilots may, at their discretion, disregard. The pilot in command is the final authority for the safe operation of the aircraft and may, in an emergency, deviate from ATC instructions to the extent required to maintain safe operation of their aircraft.

History

The air age arrived on Dec. 17, 1903, when the Wright brothers succeeded in a 120-foot flight in a heavier-than-air craft at Kitty Hawk, N.C., U.S. It is difficult to imagine the rapid technological advances that now allow interplanetary travel by unmanned, but directly controlled, satellites and probes. The earliest common uses of aviation were by the military and the civilian postal service.

With infrequent flights and virtually no carriage of passengers, the primary concern was for the integrity of the aircraft and the management of safe takeoffs and landings. One of the principal distinguishing characteristics of aviation, compared to other transportation modes, is the high speed and "vertical" nature of operations. Because of these unique features, aviation has always posed the highest risk of severe injuries and fatalities, given an accident, of almost any transportation mode. When passengers began to be carried in significant volumes in the 1920s, it became clear that a systematic set of air traffic control principles were needed to handle the increasing volumes at several critical airports.

Airplanes travel along established routes called airways, which are analogous to guideways, even though they are not physical constructions. They are defined by a particular width (*e.g.*, 32 miles) and also have defined altitudes, which separate air traffic moving in opposite directions along the same airway. Because of the ability to vertically separate aircraft, it is possible for through traffic to fly over airports while operations continue underneath. The economics of air travel require relatively long-distance travel from origin to destination in order to retain economic viability. For the vehicle operator (*i.e.*, the pilot), this means short periods of high concentration and stress (takeoffs and landings) with relatively long periods of low activity and arousal. During this long-haul portion of a flight, a pilot is much more concerned with monitoring aircraft status than looking around for nearby planes. This is markedly different from highways, in which a collision threat is nearly always apparent. While midair collisions have occurred away from airports, the scenario most feared by safety analysts is a midair collision near or at an airport because of a traffic control misunderstanding. These concerns led to the evolution of the present air traffic control system.

The first attempt to develop air traffic control rules occurred in 1922 under the auspices of the International Commission on Air Navigation (ICAN) under the direction of the League of Nations. The first air traffic controller, Archie League of St. Louis, Mo., U.S., began working in 1929. The long distances traveled by aircraft

show why aviation quickly became an international concern. The capabilities of aircraft to fly hundreds or thousands of miles at several hundred miles per hour created a market for long-distance, high-speed transportation. Two immediate concerns were in the areas of language and equipment compatibility. Pilots from many countries and with many native languages needed to communicate with each other and with controllers on the ground. Electronic equipment including radios and, more recently, computers needed to exchange information. English was established as the international language of air traffic control, but even within this context, there was a need for precise use of phrases and strings of words. These common practices have their conceptual roots in the same issues of uniformity that are directly applied to highways. The operator needs to be given clear and simple information that meets a direct need. In road transportation, this is conveyed through verbal or symbolic visual images; in aviation, it is achieved through the spoken word, supplemented by aircraft instrumentation. The initial international activity in navigation also distinguishes air transport: finding a way to a destination was an area of principal concern in the early years of aviation. Because aircraft could not operate without fixed land references (particularly on long-distance trips), it became necessary to develop an elaborate system of navigation aids (first visual, using beacons, now electronic, using radar) to help indicate the current aircraft position. Availability of inertial navigation units for commercial aircraft has reduced the need for this communication in the passenger sector; en route information is still provided through a variety of communication media on long-distance trips to warn of impending delays or other conditions.

Traffic Elements

The elements that make up the air traffic control system must provide the capability to assist aircraft in traveling between airports as well as in landing and taking off. Air route traffic control centers are responsible for controlling and monitoring movement between origin and destination airports. Each Centre is

responsible for a defined geographic area; as an aircraft continues on a flight, crossing these areas, the responsibility for monitoring the plane is transferred ("handed off") to the next air route Centre. The flight continues to be transferred until it reaches the control area at its destination. At this point, typically within five miles of the destination airport, the air traffic control function is turned over to an airport controller, and the plane is guided through a sequence of locations in order to land.

The airport traffic control tower has direct responsibility for managing handling, takeoffs, and all movement within the airport terminal control area. Flight service stations are located at airports and air route centers, providing updated weather and other information of relevance to incoming and departing pilots.

Air traffic controllers and aircraft pilots occupy a unique position in the air traffic control system. There is no other mode of transportation that relies so heavily on the communication and coordination of these two sets of individuals. As part of an overall objective to maintain safe and efficient air traffic flow, the pilot is required to comply with requests and instructions directed to him by the controller, subject to the pilot's ultimate responsibility for the safety of the aircraft. Particularly in the vicinity of airports, and particularly when arranging for landing or takeoff, clear communication is essential. Conflicts can arise between the control responsibilities of the air traffic controller and the authority of the pilot in the aircraft. Traditional approach control using

stacks (see below) placed a heavy burden on the airport traffic controllers to monitor many planes in the air. After the 1981 air traffic controller strike in the United States and the subsequent dismissal of approximately 10,000 controllers, the Federal Aviation Administration instituted a policy of flow controls. These controls required an aircraft to remain at its origin airport unless a landing opportunity was estimated to be available at the destination airport at the estimated arrival time. This results in a significantly reduced workload for the terminal air traffic controllers at the destination airport. It is an understandable source of frustration for travelers because they are not informed of a flow control delay until after the plane is pushed away from the gate at its origin and the pilot requests a landing slot. While air traffic controller staffing levels have gradually increased, the flow control system is retained because it reduces air traffic controller stress and workload by delaying flights on the ground, not in the air.

Aids to navigation are a critical element in the air traffic control system. The navigation function needs to be satisfied by a variety of technologies to supplement destination finding when visual references are limited by weather or ambient light. The earliest navigation aids were lighted beacons along the ground; these suffered obvious problems during adverse weather and were replaced by radio direction-finding equipment. The radio technologies are able to transmit the heading and distance to an intended destination. These aircraft-mounted technologies are supplemented by air route surveillance radar, which monitors aircraft within each designated sector of the air route traffic control system. The radar-based systems form the backbone of the navigation aids for privately owned aircraft and small passenger-carrying planes. Major commercial jets are now supplied with inertial navigation units, which allow an aircraft to independently navigate to a destination. A computer and gyroscope are used to sense direction and, with speed sensors, track direction and distance to the destination. The navigation units can fly virtually automatically until in the vicinity of an airport, at which time the pilot and controller interact to safely control the landing.

The landing aids most often employed are illustrated in Figure 5. An aircraft leaves the holding stack (a series of elliptical patterns flown at assigned altitudes while awaiting clearance to land), if there is one, and approaches a runway through an outer and an inner marker. Airport surveillance radar and approach lights are used to assist the pilot. The landing occurs on a runway that is designed to carry the impact load of the aircraft on landing. An important role is played by exit taxiways in expeditiously clearing aircraft from the runway in order to allow another operation (either landing or takeoff). The electronic landing aids, approach lights, and exit taxiways should work as a system to safely land and clear the runway for another operation.

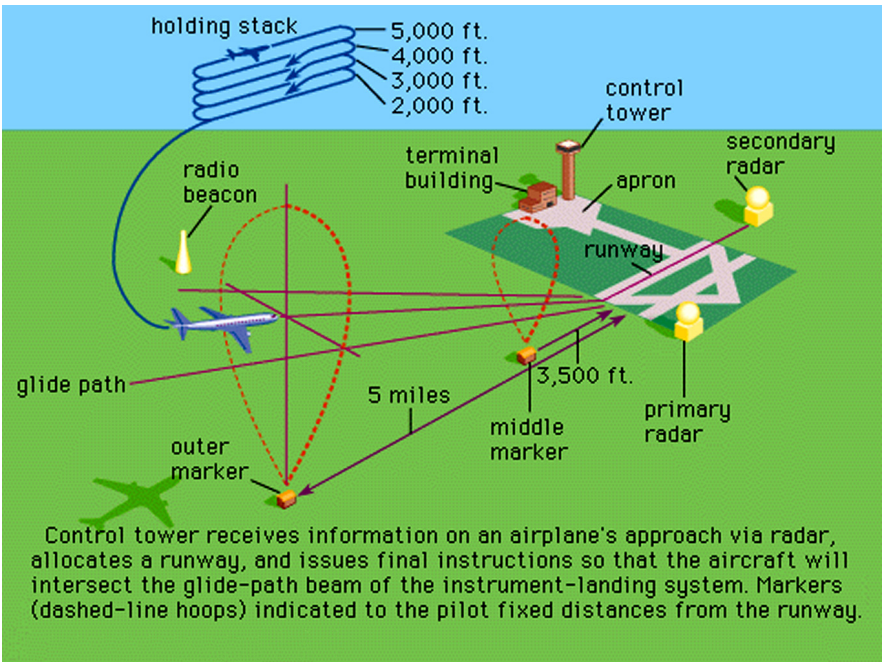


Figure 5: Aircraft landing sequence.

The final element in the air traffic control system is the ability to control and direct aircraft on the ground. Arriving flights must be safely guided to a terminal, departing flights to the proper runway. For smaller airports, under satisfactory weather conditions, this can be done visually. At larger airports, ground movement radar

is needed to track planes on the ground, just as in the air. Part of an air traffic controller's duties is to conduct this guidance of planes along taxiways and near terminals. Ground movement problems have been exacerbated in the United States by the hub-and-spoke network that has evolved for most carriers since deregulation in 1978. Carriers now operate in and out of hub airports, which are the focal points of large numbers of flights. Waves of aircraft arrive tightly spaced in a narrow time window and depart similarly bunched. Passengers frequently reach their destinations by changing planes at the hub. This allows airlines to minimize transfer times and schedule efficiently, but it can result in extreme ground delays when many aircraft exchange gate positions simultaneously. Airlines generally resist attempts to move flights significantly from on-the-hour or half-hour departures because of a perception of passenger inconvenience. Expansion of hub-and-spoke operations will continue the pressure on ground operations.

Conventional control techniques

Airspace is divided by flight levels into upper, middle, lower, and controlled airspace. Controlled airspace includes that surrounding airports and airways, which define the corridors of movement between them with minimum and maximum altitudes. The degree of control varies with the importance of the airway and may, for private light aircraft, be represented only by ground markings. Airways are usually divided by 1,000-foot levels, with aircraft assigned specific operating levels according to direction and performance. Normally all such movements are controlled by air traffic control centers. In upper airspace, above about 25,000 feet (7,500 meters), pilots may be allowed free route choices provided that flight tracks and profiles have been agreed on in advance. In middle airspace, all pilots entering or crossing controlled airspace are obliged to accept control, and notification must therefore be given to the control Centre in advance. There is a continuing trend toward expanding areas requiring positive control. Besides vertical spacing's in airways, horizontal separations are important, usually taking the form of a minimum time interval of 10 minutes

between aircraft on the same track and elevation with a lateral spacing, typically, of 10 miles.

The simplest form of flight control is called the visual flight rule, in which pilots fly with visual ground reference and a "see and be seen" flight rule. In congested airspace all pilots must obey the instrument flight rule; that is, they must depend principally on the information provided by the plane's instruments for their safety. In poor visibility and at night, instrument flight rules invariably apply. At airports, in control zones, all movements are subject to permission and instruction from air traffic control when visibility is typically less than five nautical miles or the cloud ceiling is below 1,500 feet.

Procedural control starts with the aircraft's captain receiving meteorologic forecasts, together with a briefing officer's listings of radio-frequency changes along the flight path and notice to airmen. Flight plans are checked and possible exit corridors from the flight path, in case of emergency, are determined. Flight plans are relayed to control towers and approach control centers. As the aircraft taxis out, under instructions from the ground controller, the pilot waits to be fitted into the overall pattern of incoming and outgoing movements. Controllers allocate an outgoing track, which enables aircraft separation to be maintained; this is determined from a check of the more recently used standard departure clearances. As the aircraft climbs to its initial altitude, on an instructed heading, the departure controller identifies the image produced by the aircraft on the radar screen before allowing any new takeoffs or landings. Further instructions clear the aircraft for its final climb to the en route portion of the flight and the pilots' first reporting point marked by radio devices. Progress reports on the en route portion of the flight are required and typically are tracked on radar.

At a reporting point en route, the receiving control Centre takes over the flight from the departure Centre, and all further reports and instructions are made to the new control Centre. Descent instructions are relayed to arrange the incoming aircraft at separations of perhaps five miles, in effect, on a slanting line.

As the aircraft closes in, speed adjustments or lengthening of flight paths may be necessary to maintain separations of three nautical miles over the airport boundary. Controllers determine the landing sequences and stacking instructions and may adjust takeoffs to handle surges in the incoming flights. The final stage is initiated by transfer of control to an approach controller. Under radar surveillance the final directions are given for landing. In the landing sequence, control passes to the control tower, where precision radar is used to monitor the landing, and ground-movement controllers issue taxiing instructions.

New concepts

Aviation interests also are taking full advantage of new computer and communications capabilities. In some cases, such as with on-board inertial navigation units, the computer systems will actually direct the aircraft. In most other circumstances, computer systems will provide a variety of decision-support and warning functions to pilots and air traffic controllers. Radar and plane-to-ground communications are used by air traffic control systems to predict midair conflicts and suggest actions to resolve them. Decision-support systems with voice recognition can be used to alert a controller as to when a risky or inappropriate command is given. Runway incursions (the simultaneous and conflicting use of a runway for arrival and departure) can be identified and prevented, for example. Minimum safe altitude warning also can be encoded within the air traffic control radar. Knowing the location, speed, and heading of all aircraft, the system can sound an audio and visual warning to the controller of an impending low altitude event. The low altitude systems are greatly facilitated by a capability to accurately digitally map the location of objects with particular attributes (*e.g.*, height above ground level) for use in low-altitude systems. Less fanciful but no less important is the continued expansion in use of microwave landing systems (MLS), which are replacing aging instrument landing system (ILS) equipment. The MLS is a more accurate and reliable contemporary technology.

6.2.4 Rail Traffic Control

Centralized traffic control (CTC) is a form of railway signaling that originated in North America. CTC consolidates train routing decisions that were previously carried out by local signal operators or the train crews themselves. The system consists of a centralized train dispatcher's office that controls railroad interlocking's and traffic flows in portions of the rail system designated as CTC territory. One hallmark of CTC is a control panel with a graphical depiction of the railroad. On this panel, the dispatcher can keep track of trains' locations across the territory that the dispatcher controls.

History

The first slow and cumbersome horse-drawn rail traffic posed few control problems not resolved by follow-the-leader principles. It was only after the development of swifter steam-driven trains, in the early years of the 19th century, that more frequent trains and their proximity to each other created dangers of collisions. The smooth contact between tracks and iron wheels allowed higher speeds and greater loads to be hauled at the same time that the low friction necessitated long stopping distances. Engines were fitted with brakes and, later, manned brake vans, whose guard could apply the brakes when the engine driver signaled with a whistle.

Trackside control also developed slowly with the first signalman, or "railway policeman," located at passenger and goods depots, or stations, sited along the line. These men indicated, by means of hand signals, the state of the track ahead. Red taillights were mounted at the rear of trains at night to improve safety. Later, signal flags were often replaced by swiveling colored boards, or disks, for daytime use and by colored lights at night. Later, signals were located well ahead of stopping points, giving rise to the term "distant signal." The first real method of control was the development of a time-interval system of train spacing. In the event of a breakdown or accident, however, there were no means

of delaying a following train from entering a section of track except by a physical check on entry and exit by sections—*e.g.*, a brakeman with a flag or lantern.

Because concise and standardized information was needed by the engineer, mechanical semaphore arm signals, operated remotely by wires from a lever in a signal box, were developed in 1841 as a principal means of communication. The angle of the arm indicated stop, proceed with caution, or clear ahead. For night use, colored lenses, mounted near the pivot of the arm, were passed across a light source, thus displaying, for the different arm angles, either the familiar red for stop, yellow for caution (approach, reduce speed), and green for clear (proceed as authorized). The time losses due to poor acceleration and deceleration characteristics of trains were obviated, to some extent, by the increasing use of personals, informing the driver that the signal ahead might be at stop and requiring him to reduce speed or to proceed slowly from a stop.

In the United States the railroads were provided land grants, which gave them ownership of lands adjacent to tracks as an incentive to expand service and access from the East Coast to the West. This led to a widely dispersed rail network, in private ownership with considerable duplication of service. Because the network was greatly dispersed, little congestion was experienced except in terminal areas. An unfortunate outcome of the land grant policy was oversupply of rail service and, in some cases, deliberate attempts to use rail expansion to acquire real estate. While these problems did not occur to the same degree in other smaller countries, they helped shape the scale of the U.S. system for years to come.

Traffic elements

Rail traffic control differs fundamentally from all other modes because the operator of the rail vehicle must exercise virtually all vehicle control through changes in speeds. Trains do not move vertically, and they are otherwise constrained to the guideway defined by the tracks. Rail's principal mechanical advantage is

the low friction between the wheels and the rails; this allows for efficient propulsion of the vehicle. Unfortunately it also causes rail's chief control problem: very long stopping distances. In virtually all situations, the rail vehicle operator must anticipate events very far in advance in order to take appropriate action. Unlike the highway system, in which signs and signals largely supplement what the operator sees, in many cases the rail control system must provide the operator with information beyond the immediate visual scene. This places even greater importance on the control system. Further, because the operator can adjust only speed, no other evasive action is possible to avoid an accident. These constraints in physical operation add a different imperative to rail traffic control than to any other mode.

While the technology of railroading might appear uniform, it is not, nor is the service that rail companies provide. Railroads were initially in the business of moving passengers and freight long distances (intercity service). In some countries, this dual function has remained with some or all aspects of the passenger and freight carriage being subsidized by national governments. In the United States, the long-distance passenger service, with isolated exceptions, is now conducted by airlines. Rail service is almost exclusively long-haul transportation of heavy, low-valued goods because of the comparatively long time to ship products. Because of the size of trains and their length, most control problems in the freight sector occur near cities and other termini.

Rail passenger transportation in the United States is principally conducted within urban areas and cities by urban mass transit systems. While these systems also have evolved from private to public ownership, they must contend with traffic congestion that is endemic to large urban areas. This problem is dealt with in many large cities by burying the track and stations, creating a subway or underground service. In some cases, the tracks were elevated and run one or two stories above ground. The nature of the service provided within urban areas is very different from intercity service, and so the methods of control differ. Urban service contains frequent stops. Further, some rail service (streetcars,

trams, or trolleys) runs on rails but in mixed street traffic with automobiles, buses, trucks, bicycles, and pedestrians. These rail vehicles use warning bells or buzzers to alert passengers regarding stops. They also contain all the lighting and signaling required of other road vehicles. Because of their importance in moving large numbers of passengers, urban rail transit vehicles are frequently given priority in their movement along the road network. The priority may take the following forms: separate right-of-way or lane in which other traffic may not operate; exclusively signaled turns at intersections, particularly those with heavy congestion; or portions of urban street space given to loading platforms to ease passenger boarding and alighting. Traffic signals at intersections may also be built to give priority to rail vehicles by interrupting or preempting the normal sequencing of the signals when a rail vehicle approaches. This allows the rail service to be more efficient while increasing the safety of the rail passenger. Frequent interruption of the normal signal sequence can, however, result in long delays for other road users.

Conventional control techniques

Modern railway traffic control techniques are principally automated developments of earlier systems based on timetabling, operating rules, and signals. The scheduling of trains in a working timetable predetermines the basic running patterns and the daily work pattern of personnel. Unscheduled operations require controllers to change the schedules. Minimum intervals between trains are determined on the basis of track conditions. Time-distance diagrams are often used to compare running conditions with those in the timetable and to indicate when and what type of regulatory intervention is needed.

Color light signals have now largely superseded semaphore types. Because they are operated electrically, color light signals can be sited at distances remote from the signal box. Combinations of colors are used to indicate different requirements to the driver. High-intensity lights, visible over great distances, are particularly advantageous in poor weather. Searchlights use a single lens and

bulb with different colors displayed by means of panels on color filters rotated in front of the lamp. Lights can be more appropriately sited in relation to the driver's cab position and permit a greater variety of information to be efficiently displayed.

The basic element in automatic control is an electric circuit built into the track, which operates track signals. When a train enters a section of track, or "block," it causes the current to detour through the locomotive's wheels and axles instead of completing its normal circuit, altering signals ahead. When a train has passed a section, the signal behind it is automatically switched by a track circuit immediately ahead to indicate danger. As the train advances to the next section, the first signal can automatically be changed to a lower state of warning and so on until a full clearance signal is set at a given number of sections behind the train. The number of intermediate sections left behind a train is determined by train speeds and section lengths and influences the capacity of a track.

The first recorded moving-train, two-way radio was used by the New York Central Railroad in 1928. Radio offers a number of advantages in improving communications between train crews and control dispatchers or maintenance gangs on the track. It also establishes a direct link between trains and obviates the need for crews to use wayside telephones. Equipment failures can be reported directly, and because of this and other advantages, particularly in automated marshaling yards, delay is reduced. Most railways throughout the world are equipped to some extent with two-way train radios.

Sorting freight cars is a complex operation. Various control systems have been installed in marshaling yards, enabling cars to be pushed over a raised track, known as a hump, so that the car travels freely down a grade and over switching points to its correct berth. Automatic humping includes sensors to detect car speed and weight, from which car rolling resistance is estimated. Once the uncoupled car has been allocated a train and siding, automatic switching sets the points along its predetermined path. Simultaneously the computer calculates the speed required for the car to reach the end of the train. Automatic braking devices

or boosters reduce or increase the car's speed off the hump to that needed to reach its train coupling point in the siding.

Other, more refined, methods remotely control the pushing locomotive. The spacing of cars rolling off the hump, the automatic control of the pushing speed, and the control of retarders or speed boosters are all directly controlled by computer. Identification of car destinations is an essential part of the process. Manual checking in the yard with radio links to the yardmaster have been displaced by closed-circuit television checking off the train against the makeup list that is forwarded by teleprinter.

The final scheduling and control of the freight train is integrated into the comprehensive rail control systems, and computers permit the computation of alternative strategies with an assessment of benefits. Finally, controllers impose their selection priorities.

Important traffic control and safety problems can exist where rail systems cross road networks at the same grade or level (*i.e.,* without a bridge or tunnel to separate them). These areas, called rail-highway grade crossings, pose particular control and safety problems. Because rail trains are of substantial mass and often travel at high speeds, any collision with a road vehicle is likely to severely damage the road vehicle and injure or kill its occupant(s). Because trains cannot readily slow and stop in response to an emergency, the driver of the road vehicle is most responsible for taking appropriate control actions at crossings. A well-known problem in vision perception frequently operates at railroad crossings: road drivers underestimate the closing speed and distance of the train to the crossing, because it is a relatively large object moving across the driver's field of view at a nearly 90° angle. The misperception makes it important that drivers be warned of the location of the crossing and whether trains are approaching.

Traffic devices at rail-highway grade crossings include signs, signals, and automatically controlled crossing gates. Simple warning signs advising the motorist of a crossing are the minimal type of control exercised. These may be supplemented with

flashing lights that are activated by the train when it reaches a particular distance from the crossing. The signals may be supplemented further by crossing gates that block the road based upon train detection as with the signal lights. The signal light and gate control are expensive because they require the installation and maintenance of the train detection and communication system. Simple warning signs, while useful, have the shortcoming that the degree of hazard posed by the crossing is not well known. Drivers who frequently pass the crossing with no trains nearby can become conditioned to be less alert, increasing their accident risk.

Traffic control also must be carefully managed in terminal areas where trucks are used to carry traffic to and from a train. Traditional road traffic control techniques are used in these circumstances, but particular attention must be paid to accommodating the size and performance characteristics of trucks. Intersections must have sufficient turn radii; freeway ramps must be of sufficient lengths to accommodate the limited acceleration capabilities of large trucks, and lane widths must be adequate. Special accommodation may be needed to handle longer trucks (*e.g.*, 60 feet or more) at rail-highway grade crossings.

New concepts

Just as advances in computers and communications technologies are facilitating advances in highway and air traffic control, so are they contributing to a new generation of automatic train control systems? These systems seek to make train schedules, and thus train service, more reliable. The technological infrastructure of automatic train control is particularly vexing for rail systems that are greatly dispersed geographically (*e.g.*, in the United States, Canada, Russia, and China). There, the blocks used to control train movement may be 30 or more miles long, meaning that no train may enter these long blocks while another is within them. The long blocks seriously constrain the movement of trains in a network. Components of an automatic train control system must include a capability to monitor every train in a rail network and,

associated with that train, the shipments contained in the railcars, including their expected arrival time at the destination. Vehicle location systems such as satellite-based global positioning systems are an important element in location tracking.

In order to be truly effective, automatic train control must reside within a broader companywide structure aimed at managing operations. The structure includes explicit long-term policy evaluation, which helps to plan resource allocations in support of operations.

The most basic decision that an organization must make is whether to schedule trains at all or whether it is adequate to dispatch a train when sufficient traffic is acquired (*i.e.*, a tonnage operation). This type of operation may be most wise for short-line railroads that feed specialized commodities (*e.g.*, ore or grain) to large railroads. The automatic train control system for the large railroad must be able to accommodate the movement of this train to a yard for subsequent dispatch. The tactical scheduling of trains occurs every two to four weeks, with real-time scheduling of tonnage loads in between. Computer-aided dispatching and automatic train control provide capabilities for real-time management of operations. They also provide evaluation data to use in modifying tactical or schedule policy decisions. The system, in addition to monitoring the location of all trains, must contain information on the status of every section of track and whether trains are complying with automated instructions. The system will thus use train control to improve efficiency but also improve safety by assuring compliance by train crews.

6.2.5 Marine Traffic Control

History

Navigation is still the principal means of controlling the paths of ships; direction measurements are made by a navigator using, as of old, a knowledge of the movements of the sun and stars

and, since the Middle Ages, the magnetic compass or the later development, the gyrocompass. From early times the need to exchange information between ships and with land stations led to the development of visual and audible signal systems. Markers were carried by ships and also laid in channels, and the transmission of messages was accomplished through flag, semaphore, horn, bell, whistle, and light signals leading to the establishment of first national and later international codes. The invention and use of radio, at the beginning of the 20th century, brought a marked improvement in ship communication.

Considerable advances in mapping were made over the centuries; modern navigation charts show all coasts, submerged obstacles, sea depths, and navigational aids such as lighthouses, lightships, buoys, and radio beacons.

New forms of steam propulsion and the design of iron ships in the 19th century led to increased ship size. The growth in world trade brought to the fore the problem of establishing consistent avoiding action when vessels approached each other. International rules of the road at sea were laid down in 1863 and have since been periodically updated.

Traffic Elements

Control of ships at sea and their ability to avoid potential collisions are a source of primary concern for marine safety. Because the "guideway" for a ship is water, there are limited frictional forces available to hold a ship on course. Laws of physics demonstrate that bodies in motion tend to stay in motion unless acted upon by outside forces. Because of the large mass of ships, large forces are needed to change their velocity and direction. The changes also occur very slowly and over distances of miles for large commercial ships, owing to the low friction of the guideway surface. In this respect, large ships are like trains in that they have very long stopping distances. While they can adjust their lateral position—unlike trains, which must remain on the track—they are unable to do so rapidly. Safety of large ships at sea is thus

dominated by concerns for the relative lack of longitudinal and lateral maneuverability of ships to avoid both fixed and moving hazards.

The maneuverability of any ship is heavily influenced by the environment at the time of the attempted maneuver. Wave actions, tides, and currents all result in water movement around the ship, which must be considered by the pilot in directing the vessel. Wind also can strongly influence ship movement, both for sailing vessels that use wind for power, and for motorized vessels. Limitations in visibility posed by nighttime conditions, fog, rain, or snow also strongly influence ship control and safety; indeed, environment plays the strongest role in ship and in airplane operations. Guideway-related information is important, but its effect is limited. Vessel characteristics, as described earlier, also are extremely important in marine traffic control.

Communications between ships and from ship to shore are important elements in marine traffic control. Radio frequencies are allocated for marine use on the FM band, but in busy port or shipping areas these can become quickly oversaturated. Vessel traffic systems (see below) have been proposed to ease communications and manage vessel traffic flow. In clear weather, communication is still conducted by flashing lights and flags. More than any other mode except aviation, communications play a crucial role in marine traffic.

Control devices for marine traffic include buoys, lights, sound-generating devices, and lighthouses. As with all other modes, rigid standards and regulations exist governing the use and performance of the devices. The International Maritime Organization (IMO) regulates operational procedures for avoiding collisions at sea as well as device design. Lights used to convey vessel status are regulated for specific levels of chromaticity and intensity (in order to be seen at a given distance). Sound-generating devices, including horns, bells and whistles, also are carefully allocated to particular frequencies. Lighthouses continue to be important; increasingly they are unmanned and are monitored by communications and computer equipment.

Conventional control techniques

Control of ships on the open sea still remains exclusively with the master of the vessel; when other ships are encountered, established rules of steering are practiced. This ancient arrangement—primitive by comparison with the sophisticated and centralized traffic control systems described for road, rail, and aviation—has survived, thanks to the expanse of sea and the relatively few ships sailing upon it. Communication between ships is, therefore, vital in their control, both at sea and within the confined channels of inland waterways. The principal methods of transmitting a signal are visual (that is, by flag, semaphore, or light) or audible (by means of horns or radio). The revised International Code of 1934 includes alphabetic, numeric, and answering flags. Urgent messages can be communicated by single flags, while three-letter groups are used for compass points, bearing, and times. Semaphore signaling employs hand flags, while Morse code can be transmitted visually by searchlights equipped with horizontal control slats or by radio. Ships also use sirens for "in sight" conditions to indicate impending course changes and, generally, for warning purposes in bad visibility.

The control of ships near coasts is facilitated, both for warning and navigational purposes, by the use of lightships and lighthouses. Channels on the approach to ports are clearly marked by floating buoys, usually fitted with lights and equipped with sound signals (horns, bells, and whistles) for use in bad weather or at night. The proper provision of buoys and beacons, anchored in their correct position and their subsequent maintenance, is essential for control and safety purposes.

Buoys are classified by their function into categories denoted by shape, markings, and color. The approach to an estuary, for example, is marked by a landfall buoy, and main channels by red can-shaped or black cone-shaped buoys. Where channels fork, at junctions, spherical buoys are used to indicate direction to either port or starboard. Other special buoys denote wreck positions, danger areas, and middle ground, the region near the Centre of the channel where ships can safely move.

New concepts

The management of traffic and safety on a given body of water has been previously described as an assemblage of related but distinct systems. These systems are integrated in a vessel traffic system (VTS), which can be defined as an assortment of personnel, procedures, equipment, and regulations assembled for the purpose of traffic management in a given body of water. A VTS includes some means of area surveillance, traffic separation, vessel movement reporting, a traffic Centre, and enforcement capability. These functions are not dissimilar to the advanced train control and management systems discussed in the rail section.

VTS seeks to meet the goals of the vessel traffic Centre (to manage traffic) and the ship (to move through the area) by integrating space management, position fixing, track monitoring, and collision avoidance. The vessel traffic Centre (VTC) coordinates ship passage in an area so as to be orderly and predictable. Position fixing may be done by both the VTC and ship and is critical to the next function, track monitoring, which is based upon cumulative position fixing. The last function, collision avoidance, is a new area of responsibility for VTCs. This function has traditionally been the responsibility of the respective ships' pilots and should remain so. VTC can, if so equipped, provide advance warning of impending collision and may allow the pilot extra time to maneuver.

VTSs are proposals to once again harness the power of advanced communications and computers to improve vessel safety and efficiency. The extremely large size of ocean vessels poses risk for the environment if hazards are not properly managed; the ecological disasters resulting from oil spills throughout the world are testimony to the importance of marine safety. While accidents involving loss of life are few, the prospect remains for high mortality given passenger loads (frequently in the thousands of passengers). VTS exists in limited application around the world and is likely to expand for several more decades.

REFERENCES

1. Agarwal, A, & Little, T. D. C. Opportunistic Networking in Delay Tolerant Vehicular Ad Hoc Networks, In Advances in Vehicular Ad-Hoc Networks: Developments and Challenges (Ed. M. Watfa), (2010). , 282-300.
2. De Felice M, Cuomo F, Baiocchi A, Turcanu I, Zennaro S. Traffic monitoring and incident detection using cellular and early stage VANET technology deployment. Proc. ACM 1st Int. Workshop Internet Vehicles Vehicles Internet; 2016.
3. Hartenstein H, Laberteaux KP. A tutorial survey on vehicular ad hoc networks. IEEE Communications Magazine 2008; 46(6):164–171,
4. Iglesias I, Isasi L, Larburu M, Martinez V, Molinete B. I2V Communication Driving Assistance System: On-Board Traffic Light Assistant. Vehicular Technology Conference, 2008; 1–5,
5. Pasin M, Scheuermann B, de Moura RF. VANET-based intersection control with a throughput/fairness tradeoff. Proc. IEEE 8th IFIP Wireless Mobile Netw. Conf. (WMNC); Oct. 2015;
6. Smitha A, Pai MMM, Ajam N, Mouzna J. An optimized adaptive algorithm for authentication of safety critical messages in VANET. Proc. 8th Int. Conf. Commun. Netw. China (CHINACOM); Aug. 2013; pp. 149–154.
7. VANET Vehicular Applications and Inter-Networking TechnologiesEd. H. Hartein- stein, K.P. Labertaux), John Wiley & Sons, Ltd., Mar. (2010).
8. Vegni, A. M, Inzerilli, T, & Cusani, R. Seamless Connectivity Techniques in Vehicular Ad-hoc Networks, in Advances in Vehicular Networking Technologies, Edited by Miguel Almeida, 978-9-53307-241-8INTECH Pub, April (2011).

CHAPTER 7

VANET SECURITY AND PRIVACY

INTRODUCTION

Nowadays, vehicles are equipped with high-technology devices, such as: GPS navigators, radars, and on-board units (OBUs). These wireless-enabled devices make vehicles intelligent and able to communicate with each other, and thereby form a self-organized vehicular ad-hoc network (VANET). Most proposed system architectures for VANET need to equip vehicles with a box that contains a radio interface to enable wireless communication between vehicles. The rapid mobility and dynamically changing topology of VANET cannot use the current IEEE wireless protocols 802.11 in its present state, so a modified version named 802.11p was developed by IEEE for vehicular networks. The modifications were mostly done in the MAC layer. Many wireless technologies like: a) IEEE 802.11p that is a standard for Dedicated Short Range Communication, DSRC, a Wifi type called Wireless Access in

Vehicular Environment, WAVE, b) General Packet Radio Service (GPRS), c) IEEE 802.16 that is a standard for WiMAX, and d) 4G-Long Term Evolution (LTE) have been proposed for reliable vehicular communications.

Besides that, VANETs are always looked upon as systems that would open innovative and path breaking applications. Also, before the real technology hits the road, a series of detailed research is carried out around the world to make the system reliable and robust. In addition, VANETs are a promising area for the creation of Intelligent Transportation Systems (ITS) that provide assistance to drivers to increase their safety and comfort by offering useful services to them. Moreover, VANET is a kind of network that has two main types of communication: V2V and V2I, which are vehicle–to–vehicle and vehicle–to–infrastructure respectively. The set of applications that offer comfort and convenience-based services are referred to as non-safety applications, while the safety applications are more concerned with life-saving services. With the assistance of V2V and V2I communications, potentially fatal road accidents can be avoided; dangerous driving behaviors can be alerted; city traffic flows can be optimized; and traffic jams can be alleviated. However, since vehicular communication systems (VCS) aim to serve people, any small error as unauthorized modification of data or system malfunction can be fatal.

VANETs command a unique grade of requirements to maintain liability and accountability of drivers involved in accidents, traffic violations, emission norms and irregularities in order to take punitive actions if a driver commits any crime. Besides that, location and context-aware services require pin-point user location and preferences to provide the most specific, exact and comprehensive list of personalized information. Despite that, communication of such information raises significant privacy issues that cannot be neglected. Also, privacy concerns in vehicular communications are necessary to provide protection for the user data from profiling and tracking. For example, location-based service applications have a high probability of privacy breaching and jeopardizing security-related issues, which decrease the

widespread of VANET. Moreover, quality and privacy are two divergent tendencies that exist with VANET applications and have undeniable importance to the user. Thus, both industry and academia have paid extensive attentions to address the various VANET security- and privacy related issues. On the account of improving and providing reliable services, many researchers have identified various privacy issues and came up with different techniques and approaches to maintain the user's privacy, like the use of pseudonyms, mix-zones, and group signatures. However, some applications, such as safety critical ones, are time sensitive and prevent the use of security protocols with high computational overhead and cost. Thus, security and privacy requirements in VANET should be taken into consideration when designing a robust system, otherwise, malicious attacks may ruin the original intention of VANET. In this context, prior to putting VANET into practice, it is important to have an efficient secure privacy preserving mechanism on board, which provides the needed security and privacy services while mitigates the well-known attacks in VANET.

7.1 SECURITY AND PRIVACY REQUIREMENTS

In VANETs, malicious vehicles may disrupt the network performance, for example, via modifying or inserting fake information in the network, which could incur life-endangering accidents. There are three basic requirements that should be met in VANETs to deal with any threat, which are: authentication, integrity, and conditional privacy. These requirements are fundamental so that every VANET system should follow. However, there are other security and privacy requirements discussed in the literature. Despite that, VANET brings in new challenges and conflicts between these security and privacy requirements in the system.

7.1.1 Security Requirements

There are several security requirements that should be taken into consideration when developing a secure architecture for VANET, which are explained in this as follows:

- ***Authentication:*** The basic and foremost requirement for vehicular network security is authentication. Authentication is essential for verifying a claim of authenticity. Particularly in VANET, authentication means verifying the identity of a vehicle and distinguishing legitimate vehicles from unauthorized vehicles. It is important to make sure that the transmitted messages originate from actual vehicles and not from non-existent nodes because transmission of malicious messages can lead to serious consequences like human injuries, traffic disruptions and in extreme cases may even lead to death. Also, an adversary may unnecessarily divert the traffic leading to chaos. Hence, message authentication is important in VANET. Besides that, authentication generally includes message integrity and sender verification. Moreover, for safety applications in V2V communication, the authentication requirement can deal with a masquerade attack. While, for commercial applications in V2I communication, authentication ensures that each user is authorized to access the needed service. Thereby, authentication is a fundamental access control mechanism in VANET.
- ***Integrity:*** A wireless channel is vulnerable to active attacks, e.g., data modification. Integrity is to assure that messages do not suffer from these attacks, and that all sent messages are not modified. Therefore, integrity protection is an essential requirement in vehicular communications.
- ***Accountability:*** In accountability, a node sending a message is obligated for its actions. A law enforcing agency should be able to identify malicious drivers and accounts them for their actions. Also, accountability is regarded as a crucial requirement due to the safety critical

runtime environment of vehicular networks. Moreover, accountability by its nature imposes another potential security requirement known as non-repudiation.

- ***Non-repudiation:*** It is avoiding denying that the contents of a certain message have been sent by a certain entity. Hence, non-repudiation is a critical requirement for the reliable use of VCS.
- ***Restricted Credential Usage:*** In order to achieve both authentication and accountability, a cryptographic token is used, which is called a credential. Restriction of parallel usage of authentication credentials at a particular time is a vital security requirement. It is quite necessary to protect the system from Sybil attacks, where in a fraudulent system an adversary may obtain an anonymous set of credentials to be used for impersonation of other vehicles in order to create network disturbances.
- ***Credential Revocation:*** Since VCS attaches an element of trust to a node's credential, there should be a methodology to invalidate a credential. In case of misbehaved or faulty nodes, isolation of these nodes from the network is a must that is performed through revoking their credentials.
- ***Data Consistency:*** It generally encapsulates accuracy, usability, authenticity and integrity of data in vehicular networks. It also warrants that all drivers in the system perceive a consistent view of the data. Besides that, it ensures that the data sent by a certain vehicle and its nearby vehicles are consistent.

7.1.2 Privacy Requirements

The need for privacy is addressed differently by various countries. Some countries enforce drivers' identification mechanism for crime prevention. While, some other countries may impose a mandatory privacy policy in the system. Moreover, the requirement for privacy is one of the vital reasons for public acceptance of VANET deployment. Communication in the network should be

anonymous where a message should not reveal any information about its sender. Also, the message sent should be protected in the presence of an unauthorised observer. Furthermore, the activities of a sender should be unlinkable to its source. In some schemes, a higher level of privacy is proposed where the identity of vehicles broadcasting announcements are protected even from the authorities. However, full anonymity may allow for misbehaviour occurrence as attackers would act maliciously without the fear of being caught. Whereas, some other schemes allow authorities to reveal the identity of vehicles in case of misbehaviour detection so as to achieve conditional anonymity, that is, the identity of users remain anonymous unless they misbehave.

There are several privacy requirements that should be considered when designing a privacy-preserving architecture for VANET, which are described as follows:

- ***Anonymity:*** A message's sender should be indistinguishable or anonymous among a group of senders. In order to preserve privacy of senders, VANET needs to provide anonymity to senders/drivers. Thus in theory, it should not be possible to link a message content to the person who sent the message. However, this imposes a conflict between accountability and anonymity. Therefore, the provision of conditional anonymity is needed in order to achieve both security and privacy requirements.
- ***Conditional Privacy:*** Undoubtedly, a driver benefits from the traffic-related messages that are automatically sent by other neighboring vehicles. However, these messages include a sender's private information, such as the vehicle's identity (plate license number), location, and direction. Clearly, people are not interested to expose these private information to third parties. Hence, a secure mechanism should prevent an unauthorized party from knowing the combination of the real identity and other private information. On the other hand, a trust authority (e.g., police officers) has the authority to reveal a vehicle's identity in case of criminal action occurrence.

Thereby, conditional privacy preservation is essential in VANET.

- ***Confidentiality:*** This security service prevents the disclosure of message contents to unauthorized entities in order to maintain the user's privacy.
- ***Unlinkability:*** An adversary cannot sufficiently distinguish whether the Items of Interest (IOI) (messages, actions, and / or subjects) used in vehicular networks are related or not. It is worth to note that unlinkability of sender to a certain message can be termed as anonymity, as this may breach the sender's anonymity.
- ***Minimum Disclosure:*** A user should reveal the minimum amount of information during communication. The user's data that is disclosed during a transaction should be minimum, in short no extra information than what is required for the job. The information collected should be adapted to the concerned specific requirement.
- ***Distributed Resolution:*** Distributing among authorities the process of identity resolution is an important privacy requirement, where authorities need to cooperate in order to link a credential to a specific entity. This property is crucial for maintaining conditional anonymity while still preserving the user's privacy.
- ***Perfect Forward Secrecy:*** Resolving a user's identity or credentials should not disclose any information that allows the linkability of future messages to that user.

7.1.3 Other System Requirements

There are other system requirements that should be thought of when developing a robust architecture for VANET, which are given as follows:

- ***Scalability:*** It may not be considered when designing a secure protocol for a traditional MANET because the number of users in MANET is not large and so failing to consider scalability would not lead to vital attacks.

However, in VANET, scalability is an extremely vital factor. The incoming messages should be authenticated by a vehicle in a timely manner even in a high density area. Otherwise, some messages will be dropped before being verified if the security scheme is not efficient in high density areas. Moreover, a scheme that is not scalable is vulnerable to denial-of-service (DoS) attacks.

- ***Storage Requirements:*** Cryptographic authentication techniques have been widely exploited to secure VCS. Cryptographic credentials should be securely stored and constantly updated due to various reasons. One of these reasons is to achieve privacy. Two techniques commonly used to satisfy the property of privacy, which are pseudonyms and group signatures. In pseudonymous authentication, vehicles store a large number of public/private key pairs, and their corresponding certificates. The changing of pseudonyms is required to make tracking of vehicles by an adversary difficult. Therefore, the size of an anonymous key should be kept as minimum as possible in order to minimize the storage space needed by a vehicle. While in group signature schemes, pre-storing a large amount of certificates is needed. However, the issue associated with group signatures is that the size of a signature is quite big.
- ***Availability:*** Some applications require high availability of a communication network, such as emergency services that are time sensitive. For example, in case of an emergency, the failure of instant reception of sent messages renders the application useless.
- ***Real-time Requirements:*** VANET applications, such as: safety-related applications, require updated or real-time information to be frequently broadcasted to vehicles via RSU or neighbouring vehicles.
- ***Robustness:*** System robustness implies that the communication channel is secure and privacy-preserving, e.g. authentic and integrity-protected even in presence of malicious or faulty nodes.

7.2 SECURITY AND PRIVACY REQUIREMENTS VERSUS SECURITY APPROACHES

We briefly discuss different VANET approaches presented in the literature, which are proposed in order to fulfil various security and privacy requirements in VANET. Table 1 provides a summary of these approaches that are discussed below in details.

7.2.1 Authentication and Privacy

In recent days, two methods are used for providing anonymous services, which are Group Signature and Pseudonymous Authentication schemes. Both of them address the problem of authentication and privacy. In group signature schemes, a vehicle is issued a group private key with which it signs a message; while in pseudonymous authentication schemes, each vehicle stores a set of identities. In addition, there are hybrid approaches that combine both group signature and pseudonymous authentication schemes, where a vehicle can maintain a pseudonyms set and a secret group signing key, and also can issue itself a certificate using the group key.

Group Signature Scheme

There are some approaches in VANET that use group signature scheme in order to maintain the signer's anonymity. This scheme allows every group entity to generate a signature without revealing its exact identity, while other group members could verify the message authenticity. This signature scheme has two components: a group manager and group members. A group manager is responsible for the key distribution, adding a group member, detecting and revoking a misbehaved group member. Each group member signs a message by a group user key issued by the manager, while other members would not be able to identify the exact identity of sender. At first, the manager of group issues different group user keys for every group member, then issues

group public key to all group members. A group member uses the group user key for signing messages, while uses the group public key for verifying the message authenticity. Only the group manager knows each member's real identity, thus it could detect and revoke the group members. Despite that this scheme provides some security services, yet it incurs a large revocation cost. Also, vehicles can join and leave groups very rapidly; therefore this scheme is not practical to be used in real-life scenarios.

Table 1: Summary of VANET Security and Privacy Approaches

Security and Privacy Requirements	Security and Privacy Methods	Security and Privacy Approaches
Authentication, Privacy	Credential Usage, Digital Signature, Encryption, Anonymizer Proxy	**Group Signature Approaches:** TACK [25], BGLS [26], Signcryption [27], Trusted Platform Module (TPM) [28], [29], Batch Verification [30], Re-encryption [31]
		Pseudonymous Authentication Approaches: PASS [32], DCS [33], Mix-zone [34], Fixed Mix-zone [10], RLC [8]
Authentication, Data Integrity	Credential Usage, Digital Signature, Encryption, Message Authentication Code (MAC)	**Multiple Approaches:** Decision Packet [37], Security Mechanisms [38], Multi Operating Channels Model [39], Public Key Infrastructure (PKI) [40]
		Identity-based Approaches: Identity-based Batch Verification (IBV) [36], Identity-based Aggregate Signature[41]
Anonymity, Unlinkability	Pseudonym Usage, Silent Period, Mix-zone	**Pseudonym Approaches:** Pseudonymous Technique [42], Variable Pseudonyms [35], Silent Period [34]
		Mix-zone Approaches: Independent Mix-zone [9], Multiple Mix-zones [43]
		Other Approaches: VANET-based Clouds [44]
Traceability, Accountability,	Credential Usage, Digital Signature,	**Traceability:** Challenge-response Protocol [11]
Misbehaviour Detection, Revocation	Intrusion Detection System (IDS), Certificate Revocation List (CRL), Reputation-based Methods	**Revocation Approaches:** CRL [32], Local and Global Revocation [48], Reputation-based Scheme [49], Certificate Revocation Scheme [50], Credential-based Protocol [51]
		Misbehaviour Detection Approaches: IDS [52], APDA [55], RRDA [56], Stable Community Detection [57], EAPDA [60], Verification Technique [61]

For example, a scheme based on group signature was presented, where a VANET key management approach using Temporary Anonymous Certificate Keys, TACKs, was examined. In, authors gave some valid assumptions and discussed the efficiency of their scheme. However, some issues in VANET still remain in their TACKs; such as: the detection and revocation of temporary keys is restricted by the expiration scheme. Also, the correlation attack could happen in the following situation: when only one OBU is changing its keys at a certain time, the new key could be associated with the old key of this OBU by an adversary, so the adversary can compute the exact identity of the OBU. Also, the use of Boneh, Gentry, Lynn and Shacham (BGLS) aggregate signature, which allows a receiver to verify a group of signatures in one operation. The authors also proposed a method to compress data and signatures to minimize the storage overhead of an OBU.

Used two mechanisms on top of a PKI system to achieve authentication, privacy, integrity, linkability, efficiency, and scalability. These mechanisms are known as Signcryption and group signature. The system depends heavily on distributed RSUs to run an onthefly group. Signcryption is a technique used to sign and encrypt a message at the same time. Also, it is used to enable a vehicle to ask for a secrete member key in order to join a group run by a RSU. Then, a RSU would check the validity of vehicle data and issues a key if the vehicle is legitimate. A RSU would use a group signature where every vehicle in the group is capable of communicating with other group members and RSU without revealing its identity. This is made possible using anonymous group certificate. The authors also proposed that a vehicle can check if the sender is revoked through the RSU instead of maintaining a CRL. Besides that, batch verification is used to reduce the computation time needed. An efficient group formation technique, where asymmetric cryptography is adopted for normal message communication and symmetric cryptography for event-driven messages. Vehicles form a trusted group using Trusted Platform Module (TPM) to improve security and privacy. While, in order to maintain conditional privacy for V2V communication in VANET, a batch verification method to prohibit unnecessary

group subscriptions via group signature method. This approach provides a variety of security requirements utilizing the group signature method. Whereas, A group signature method integrated with a re-encryption technique, where a third party can re-encrypt a message that can be decrypted by other authorized users. This approach allows the message broadcasted in a group to be read by any member of the group or other authorized groups using the re-encryption mechanism.

Pseudonymous Authentication Scheme

The idea behind Pseudonymous Authentication scheme is that each vehicle stores many pseudonymous certificates at first, and then randomly chooses one of these certificates to act as its identity at a certain time. Also, since a Trusted Authority (TA) has sufficient storage and could not be compromised; so, it is safe and feasible for a TA to store these pseudonymous certificates. When a vehicle first registers, the TA sends enough pseudonymous certificates to it, and a unique permanent identity. For privacy considerations, vehicles do not use the permanent identity to sign messages; they rather randomly choose one of the pseudonymous certificates that the TA has issued for digital signature. By this way, the temporary identity of each vehicle changes over time, and a malicious attacker can hardly trace a specific vehicle. This is because after altering the certificate, an attacker would not be able to link the new certificate with the old certificate, which means that the attacker has lost the target. However, this method still has some problems, such as high revocation cost. For example, when a vehicle is revoked, the number of pseudonymous certificates that needs to be added to the Certificate Revocation List (CRL) could be too large, where the size of CRL increases rapidly when the size of network increases.

In the context, many works on location privacy address the issue of pseudonyms' provisioning and update, i.e. how signatures using pseudonyms are assigned, and when pseudonyms are changed, as shown in Figure 1.

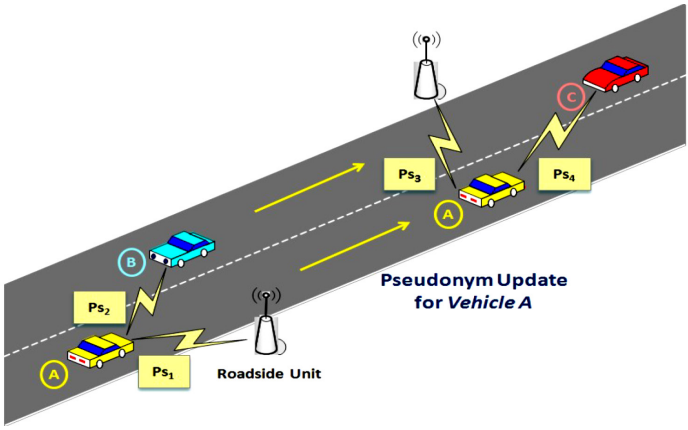


Figure 1: Vehicle A updates its pseudonym (Ps) for each new message sent.

For example, a Pseudonymous Authentication and Strong privacy-preserving Scheme (PASS), where a vehicle can update its set certificates via the neighbouring RSU. Compared with the previous pseudonymous schemes, PASS has a lower revocation overhead and lower certificate update overhead too. Also, a Distributed Certificate Service approach, DCS. In order to ensure that an OBU can update its certificate via a RSU; regardless whether this OBU currently exists in the same RSU domain where it has been registered or not. The DCS approach could rapidly certify many certificates and signatures. In addition, some schemes imply that the pseudonyms need to be changed every certain time in a mix-zone to preserve the user's location privacy, since a mixzone creates a k-anonymous region where a silent period should be preserved between each pseudonym update. For example, a method for pseudonym update that exhibits the same level of privacy for all traffic scenarios. This method depends on fixing the mix-zone that thwarts an adversary from linking a particular pseudonym with the real identity of a specific vehicle. Whereas, a pseudonym-enabled approach for maintaining privacy in VANET. In this approach, a RSU is used to assign a Reputation Label Certificate (RLC) for vehicles in its range in order to overcome the conflict between privacy preservation and reputation determination of a certain vehicle.

7.2.2 Authentication and Data Integrity

One of cryptographic mechanisms used to achieve message authenticity and data integrity is the usage of symmetric primitives. This includes Message Authentication Code (MAC) appended with a message that is computed using shared symmetric secret key. In order to validate a MAC, a third entity needs to see the secret key, but would not know which of the two parties computed the MAC. While symmetric-based techniques are computationally efficient, they do not provide the property of non-repudiation. So, a digital signature is used to solve this problem, since signing a message using valid credentials issued from a Trusted Party (TP) would satisfy both authentication and data integrity. Digital signature schemes used in the literature include: Group Signatures (GS), traditional Public Key Cryptography (PKC), identity-based signature & batch verification.

For example, an approach that utilizes a Decision Packet. In this approach, a node creates a route from departure node to destination node and performs the necessary verification using the Decision Packet's hash value, which thwart an attacker from changing the hop count. Besides that, different security mechanisms in order to provide security for the routing protocol and to protect the user's information from modification. While, Multi Operating Channels Model was proposed. To provide network protection against attacks that threaten the network functioning and data confidentiality in VANET, such as: DoS and modification attacks. In addition, a Public Key Infrastructure (PKI) in order to fulfil the major security requirements in VANET, such as: authentication, integrity, and non-repudiation. Also, an update certificate method is used in when a vehicle enters a new region. In this approach, the computation time needed for message authentication is decreased in order to minimize the message loss rate due to message check delay. Moreover, an Identity-based Batch Verification (IBV) scheme, where they proposed to randomly update the anonymous identity and location after a certain time. While, an approach that utilizes multiple TAs and a one-time identity-based aggregate signature. In this approach, vehicles are able to verify multiple

messages simultaneously, where signatures are compressed into a single signature in order to minimize the storage space needed by a vehicle.

7.2.3 Anonymity and Unlinkability

A common approach to achieve anonymity is by using pseudonyms. For example, a pseudonym as a public key in place of an identity in the ID-based announcement scheme. Each key could be either used once for each message or used to sign multiple messages over its short lifetime, where the key update frequency is varied depending on some factors, such as vehicle's speed. The pseudonyms are updated in order to prevent linking of different vehicle's activities. However, drawbacks of the pseudonymous technique include: secure key distribution and management, and storage complexity. Also, randomly chosen and changing pseudonyms are used. In order to prevent linking to the user's real identity. While, a silent period was proposed in order to achieve unlinkability. The level of unlinkability depends on the number of vehicles present during the time the change of pseudonyms takes place. Also, the high velocity of vehicles present at the time of pseudonym update decreases the ability of an adversary to probabilistically determine and link pseudonyms and vice versa. During silent period, transmission of messages is temporarily disabled for a period of time, and a vehicle does not receive any incoming messages. The drawback of this technique is that it restricts a vehicle from generating or receiving a message for a certain period of time, which defeats the purpose of VANET deployment.

An independent mix-zone mechanism to solve the problem of the pseudonym update in low density areas. This mechanism involves provisioning of certificates and pseudonyms, where a vehicle establishes a mix-zone through its periodically broadcasted messages. While other vehicles may produce random versions of a pseudonym respectively, which generate a k-anonymous mix-zone. A pseudonym update method with multiple mix-zones. Also, a cheating detection technique was presented in that enables

a vehicle to check if the pseudonym update process was successful or not. Whereas, an approach that offers the following services:1) Maintains the user's privacy using multiple anonymity. 2) Tracks the user's route by saving beacon messages in Cloud. 3) Provides conditional privacy via anonymity withdrawal application. This approach incurs less overhead as compared to other approaches proposed in the literature.

7.2.4 Traceability, Accountability and Non-Repudiation

In a pervasive VANET environment, misbehaviour may take place as a result of hardware malfunctioning or may be intentional. For instance, the safety-related messages may contain fake information, or may have been modified, discarded or delayed intentionally. In such situations, it is desirable to achieve accountability. Also, the source of misbehaviour should be traceable for liability purposes. Traceability is desirable when a dispute arises, however, it is difficult to achieve traceability due to the need of privacy requirement. So, different methods are studied to solve this problem. For instance, in schemes that use pseudonyms, the pseudonyms could be linked with a specific identity that possesses the unique Electronic License Plate (ELP), in order to trace the misbehaved user by the authorities. Meanwhile, in schemes such as: group signatures, a tracing manager is adopted to revoke the malicious vehicles by opening their signatures.

While, accountability is achieved if it satisfies traceability, non-repudiation and revocation requirements. The necessity for accountability in VANETs arises from the possibility of misbehaviour among users that may harm public road safety and jeopardize VANET future deployment. Misbehaviour in VANETs may occur due to malicious activities of users inside the system. Such activities may include: preventing message broadcasting to other vehicles; generating fake messages; injecting non-safety-related messages that may cause traffic jam in the network due to overload of the bandwidth; or escaping from an accident. While attacks performed by outsiders can be addressed by means of authentication, misbehaviour among legitimate senders is a more

challenging problem to address. This is because legitimate senders possess valid credentials issued by an authority and could deceive other vehicles to trust them to perform malicious actions.

Another aspect of accountability is non-repudiation, where an entity is not able to deny the act of sending a message signed using a key that belongs exclusively to it, assuming forgery is not possible. A challenge-response protocol is another approach to achieve non-repudiation that was proposed, where: given a signature on a message, the challenge-response protocol determines whether a vehicle is the signer of the message. While in some other schemes, non-repudiation is assumed in the presence of a fully TP. In addition, a mutual authentication approach with DoS resilience. This approach adopts an identity-based signature scheme, and offers security services such as: unlinkability, conditional privacy, and non-repudiation. Besides that, provided a mobile agent protocol for VCS that provides a variety of security services including accountability and non-repudiation, as well as mitigates well-known attacks. Also, this protocol allows fast response for a vehicle's request, where the collected data is not lost even if the mobile agent is lost.

7.2.5 Misbehaviour Detection and Revocation

Detection of misbehaving vehicles is an important issue in VANET that has recently received a lot of attention. The authentication mechanism itself only guarantees the message integrity but cannot ensure that the content of message is correct. Therefore, when a vehicle misbehaves, such as: modifies a message, gives bogus information to others, attacks the network by pretending to be another one; the network should have the ability to detect these false messages and the malicious vehicle in order to revoke that vehicle through some schemes. Some misbehaviour detection schemes (MDSs) are run by vehicles in order to detect any misbehaviour and then report the malicious vehicle to the Certificate Authority (CA). Meanwhile, the communication overhead is a big issue when distributing the CRL. Accordingly, some work have been done in order to decrease the size of CRL

to reduce the network traffic during the distribution phase. For example, an authentication mechanism, where the size of CRL depends on the number of revoked vehicles and independent on the number of pseudonyms that the misbehaved vehicle owns. Despite that, the CRL distribution process to all remaining vehicles in the whole network takes large time. During this interval, the attacks could still jeopardize other drivers' safety.

In this context, the existing revocation schemes are mainly of two types, which are local revocation and global revocation that are described as follows:

- *InLocal Revocation*, a local voting mechanism is used to identify and revoke a malicious vehicle. Two requirements should hold that are: the majority is honest, and other vehicles are able to detect any misbehaviour. Many legitimate nodes may be unable to vote as a result of the lack of detection ability, e.g. not within detection range. Also, there exists Sybil attacks that can affect the voting result.
- While, *inGlobal Revocation*, the CA identifies the accused vehicle, and determines whether to revoke it by the use of trust management. If one vehicle is judged as a misbehaving node, all its certificates are invalidated in the entire network. However, the main challenge in the global revocation scheme is that the CA is not always available and the latency may be unacceptable in real-life scenarios.

Moreover, in some revocation schemes, the CRL is no longer used. For example, in reputation based schemes, revocation is achieved by ceasing to provide misbehaved vehicles with their reputation credentials. A vehicle whose reputation score decreases to zero would not be able to continue its future participation in the network. On the other hand, an approach to maintain the security and privacy in VANET. This approach incurs a low computation time and enhances the certificate revocation process. Besides that, an anonymous credential-based protocol that enables the revocation of a misbehaved vehicle. Also, this protocol provides

the detection of fraudulent actions, where the revocation of subsequent credentials of a malicious entity is performed.

Furthermore, an Intrusion Detection System (IDS) in VANET through classifying the detection system into signature-based, anomaly-based, and specifications-based systems. Besides that, an Attacked Packet Detection Algorithm (APDA) to protect against some security attacks, such as: DoS attack. This algorithm reduces the time delay to enhance VANET security. In addition, a Request Response Detection Algorithm (RRDA) to determine DoS attack. This algorithm utilizes a hash table to minimize a DoS attack caused by a malicious vehicle, and sends messages to all vehicles from departure till destination as well as updates the hop count. Also, this algorithm minimizes the message delay as compared to other algorithms proposed in this context. While, a stable community detection algorithm after considering the vehicle's dynamic motion, where authors evolved the Label Propagation Algorithm(LPA) community detection algorithm with the Stability and Network Dynamics over a Sharper Heuristic for Assignment of Robust Communities (SandSHARC). Besides that, evaluated their work by testing the stability of the detected community. Not only that, but also an Enhanced Attacked Packet Detection Algorithm (EAPDA) to mitigate various attacks, such as: DoS attack, which results in performance deterioration of VANET. This algorithm incurs less time delay as compared to other proposed algorithms. Whereas, a verification technique to verify the vehicle's activities in a private manner. In this technique, a RSU decides if a message is trusted or not, and then notifies the neighboring vehicles with the decision.

7.3 TYPES OF VANETADVERSARIES, ATTACKS AND ATTACKERS

Vehicular networks are vulnerable to eavesdropping by adversaries in their wireless range as well as location samples can be collected for tracking purposes. Also, envisioned inter-vehicular communication protocols and applications provide information

about different identifiers ranging from the vehicle IP address and destination IP address to the protocol used. The interesting part is the association of these identifiers with location and time samples, i.e. identifier, location, and time. The identifier of a vehicle with its location and time-stamp are often referred to as the location sample. Many profiles of such location samples collected pose a serious threat to the privacy of the user. It is interesting to note that considering only location tracking of a car user doesn't violate the user's privacy until the user is mapped to the vehicle, where breaching of privacy takes place.

When considering VANET security, a large number of threat models may be assumed. A threat model includes an adversary that is a person or a group of people, which threatens the security and privacy of a given system. Moreover, in VANET, there are several possible attacks and attackers. In this, we explain the different types of adversaries and attackers, and also classify the attacks present in VCS based on the communication system layers. Table 2 shows the security and privacy requirements against the various attacks present in VANET.

7.3.1 Types of Adversaries

VANET safety and non-safety applications may perform as expected or deviate from their expected operations mainly due to adversarial activities. The adversarial incentives may be money, spying on the user, or some other personal benefits. Some of the previous works proposed in this context provide a survey of the adversaries relevant to the vehicular context. The different types of adversaries that are present in VANET are provided as follows:

- ***External and Internal Adversaries***: Some adversaries could be internal entities while others could be external entities in/to the system. External adversaries are entities that are not equipped with credentials and keys to access the data-handling systems/servers or applications, where the processing of the user's location data, personal data and preferences is performed. While entities

having access to the previously mentioned systems and legitimate participants in the system can be termed as internal adversaries.

- ***Passive and Active Adversaries***: A passive adversary can only learn and listen, for instance an eavesdropper that intercepts messages between a user machine and an infrastructure. A passive adversary gathers information from the collected messages and vehicle movements to draw inferences about a target user. Despite that a passive adversary learns about a specific user, it does not influence the user's behaviour. In contrast to a passive adversary is an active adversary that can affect the user's behaviour. For an adversary to be active, it vigorously participates in the network with intentions to cause disruption. This type of adversary may modify, replay or drop legitimate messages in order to present fake information to other vehicles. Other attacks that could be generated by this adversary type include generating and broadcasting bogus information to other vehicles.
- ***Local, Extended and Global Adversaries***: A local adversary has limited territorial effect and controls some entities in network, such as: vehicles or RSUs. On the other side, for an adversary to be extended, it controls several nodes in the network. While, the strongest adversary has a global coverage of the network that is known as a global adversary. The distinction between these types of adversaries is important to preserve the user's privacy.
- ***Independent and Colluding Adversaries***: Adversaries may perform independently, or may collude in order to exchange information and perform more effective attacks. For example, a group of vehicles may collude to perform a certain attack to achieve their mutual agenda or interest. Also, a group of colluder vehicles may clear the way for attackers by reporting false information of traffic jam, which would convince other innocent vehicles by that wrong information since the report had come from more than one vehicle.

Table 2: Security and Privacy Requirements versus VANET Attacks

	Security Requirements[15], [16]					Attack Scope - CommunicationSystem Layers
	Authentication	**Integrity**	**Accountability, Non-Repudiation, Credential Revocation**	**Restricted Credential Usage**	**Data Consistency**	
VANET Attacks [13]	Impersonation, Sybil Attack		Impersonation, Sybil Attack, Malicious Vehicle	Impersonation, Sybil Attack		Application and Transport Layers
	Bogus Information or Forgery, Jungle Communication, Tunnel Attack	Jungle Communication	Bogus Information or Forgery, Jungle Communication, Wormhole Attack		Bogus Information or Forgery, Wormhole Attack	Network Layer
	On-board Tampering orIllusion, Message Replay, Message Modification / Alteration, Denial-of-	On-board/Vehicle Information Tampering orIllusion, Message Modification	On-board/Vehicle Information Tampering orIllusion, Message Replay, Message Modification /		On-board Tampering orIllusion, Message Replay, Message Modification / Alteration	Physical Layer
	Service (DoS)	/ Alteration	Alteration			
	Privacy Requirements[17], [18]					**Attack Scope - Communication System Layers**
	Conditional Privacy	**Anonymity, Confidentiality, Unlinkability**	**Minimum Disclosure**	**Distributed Resolution**	**Perfect Forward Secrecy**	
	Movement Tracking	Movement Tracking	Movement Tracking	Movement Tracking – Internal Adversary [5]	Movement Tracking	Application and Transport Layers
VANET Attacks [13]	Location Disclosure, Trajectory Disclosure	Location Disclosure, Trajectory Disclosure	Location Disclosure, Trajectory Disclosure			Network Layer
	Eavesdropping, On-board Tampering orIllusion, Message Modification / Alteration	Eavesdropping				Physical Layer
	Other System Requirements[13], [14]					**Attack Scope - Communication System Layers**
	Scalability	**Storage Requirements**	**Availability**	**Real-time Requirements**	**Robustness**	
		Movement Tracking	Sybil Attack, Information Block	Sybil Attack, Information Block	All attacks in Application and Transport Layers	Application and Transport Layers

VANET Attacks [13]	Tunnel Attack, Wormhole Attack		Packet Dropping, Bogus Information or Forgery, Tunnel Attack, Wormhole Attack	Packet Dropping	All attacks in Network Layer	Network Layer
	DoS		On-boardTampering orIllusion, Message Replay, Message Modification / Alteration, Jamming, DoS, RSU Relocation	Message Replay, Jamming, DoS	All attacks in Physical Layer	Physical Layer

Moreover, there are other types of adversaries that are a mixture of the previous types, for example:

- ***Global Passive Adversary:*** A person or a group of people with enough privileges to eavesdrop on the whole network.
- ***Local Passive Adversary:*** An entity that has a restricted coverage range, e.g. through gaining access to a RSU, and eavesdrops on the wireless communication.
- ***Local Active Adversary:*** An entity that has a restricted coverage range and performs malicious actions, such as compromising a neighbouring vehicle (target vehicle) or a RSU.

7.3.2 Types of Attacks

In VANET, attacks may arise from faulty or hostile remote computing nodes. Also, the privacy attacks are of greater concern for a user than the security attacks. In this part, the possible attacks that may occur in VANET are classified based on communication system layers as follows:

Security Attacks on Application and Transport Layers

- ***Movement Tracking***: An attacker associates the identity of a vehicle to its messages then tracks the route of that vehicle.

- ***Impersonation Attack***: For malicious purposes, an attacker masquerades as another vehicle by using a false identity to attack and fool other vehicles. Furthermore, an attacker may pretend to be a RSU to send fake advertisements to vehicles in its coverage range.
- ***Sybil Attack***: An attacker uses a faulty entity to create multiple fake identities and then acts as a few vehicles to takeover part of the system. This allows an attacker to produce an illusion to other vehicles, for example, that there is a traffic congestion to force other vehicles to take an alternate route to free the route for itself.
- ***Information Block***: In this attack, an attacker makes use of the VANET protocol. If a vehicle sends a message to its neighboring vehicles, the information is stopped while being transmitted causing confusion to other vehicles.
- ***Malicious Vehicle***: When using pseudonyms, a malicious vehicle may change its identity and hence it may be hard to be tracked by authorities.

Security Attacks on Network Layer

- *Location Disclosure*: A location sample includes three components, which are: ID, location, and time. Any of these components could be modified and manipulated by attackers. Also, attackers may abstract the real identity of a target vehicle from its traffic-related messages, and further knows the vehicle's location sample.
- *Trajectory Disclosure*: An attacker may globally observe the broadcasts of a target vehicle and uses this information to reveal the vehicle's identity.
- *Packet Dropping*: In a multihop communication, an attacker may automatically or selectively drop some or all packets received.
- *Bogus Information or Forgery or Fabrication Attack*: In this attack, an adversary broadcasts fake messages into the network. For instance, a malicious vehicle may send a

fake congestion message or claim that it is an emergency car to make use of the lane alone. Moreover, this type of attack could lead to accidents. Therefore, verifying messages' freshness and validity in V2V communication is vital to make sure that the received messages are not forged.

- *Jungle Communication*: This attack is an evolution of the Bogus Information attack, where data is sent to other vehicles that continue to modify it in order to change original information.
- *Tunnel Attack*: In this type of attack, false information is sent to a vehicle moving in a place with no GPS coverage, e.g. a tunnel, where the vehicle may update false information.
- *Wormhole Attack*: In this attack, meaningless information is sent from authorized entities that results in network disturbance.

Security Attacks on Physical Layer

- *Eavesdropping*: In VCS, overhearing of messages could allow an attacker to easily collect vehicle-specific information and infer the personal data of driver thus violates ones privacy.
- *On-board / Vehicle Information Tampering or Illusion Attack*: It involves cheating with internal vehicle's information, such as: speed, position, via tampering hardware. The vehicle's information is provided incorrectly using sensors or internal devices, in order to trigger malfunction of a vehicle or to deceive other vehicles in the network. Also, sometimes an attacker may pretend to be another person by cloning the other's location.
- *In-transit Traffic Tampering Attack*: A malicious entity may intentionally cause delay, corruption, replay or modification of messages to damage the normal functioning of VANET communications. This attack includes:

  - *Message Replay*: An attacker records messages received from legitimate vehicles and then resends them back, for example, in order to disturb the traffic or cause some confusion. This attack can be done in two methods, which are: using one's OBU or using a special piece of hardware. The duplicated messages might make a vehicle fail to know its neighbour's correct driving status, e.g., direction, position, speed, etc.
  - *Message Modification / Alteration*: An attacker modifies information of a vehicle included in a message(e.g. position) for ones benefit, which could be a potential threat to the safety of other vehicles in network.
- *Jamming Attack*: An attacker deliberately generates too many messages to overcrowd the wireless channel in order to trigger malfunction of network.
- *Denial-of-Service (DoS) Attack*: An attacker attempts to disrupt the normal service of VANET by prohibiting services to be provided normally. This behaviour would cause serious consequences if the service is safety-related application. For example, an attacker may continuously broadcast a lot of dummy messages to flood the network aiming to bring down the transmission channel so that vehicles cannot exchange safety messages. For a sophisticated attacker, it may send a large number of messages with invalid signatures. In this case, a legitimate vehicle would spend a lot time verifying invalid signatures that causes a delay in verifying a legitimate message.
- *RSU Relocation*: An attacker may relocate a RSU in order to mislead vehicles.

Group-Related Security Attacks

Some attacks may be carried out if the system utilizes a group approach, such as:

- Disclosure of group secrets.

- Tracing based on group secrets, where a vehicle could be traced based on its group key.
- Collusion between new Group Leader (GL) and old GL, where random rotation of GLs would not prevent the attacks based on disclosure of group secrets.

7.3.3 Types of Attackers

An attacker is an entity that compromises the security and breaches the privacy of another entity. Different attackers have different impacts on VANET. Although some attackers look for amusement and fun, some others look for severe damage or privacy breach. The broad categories of attackers explained are as follows:

- *Malicious Attacker*: A dangerous person whose main goal is to cause a great damage in system. This kind of attacker brings jeopardy to legitimate drivers. For instance, a malicious attacker may deliberately tamper messages and give wrong information to mislead other network entities. Also, malicious attackers could slow down or stop vehicles on a highway to cause accidents. A malicious attacker may breakdown the network by compromising a RSU.
- *Selfish or Greedy Driver*: A driver who abuses the system to maximize ones benefit. Most drivers misbehave for selfish reasons, where they do not want to share lanes with other vehicles. For example, a vehicle may tell other vehicles behind it that "there is congestion ahead" by injecting false messages so that the road is cleared for it.
- *Snooper or Eavesdropper*: This person tries to collect information about a target vehicle via its broadcasts in order to identify that vehicle and hence could easily track it.
- *Prankster*: A hacker who performs some malicious actions or a person looking for fame. For example, a malicious entity that sends fake messages in order to slow down the network.

- *Industrial Attacker*: A person who belongs to the automotive manufacturer and could tamper the GPS system, vehicle's sensors, or other sensitive devices. So, a tamper-proof device (TPD) is recommended to be used to thwart such type of attacker.

7.4 SECURITY ISSUES IN VEHICULAR AD HOC NETWORKS

Communications are becoming more wireless and mobile than ever. Thus, in the near future, we can expect that vehicles will be equipped with wireless devices, which will enable the formation of Vehicular Ad Hoc NETworks (VANETs).

The main goal of these wireless networks will consist in providing safety and comfort to passengers, but their structure will be also taken advantage with many different aims, such as commercial, access to Internet, notification, etc.

From a general point of view, the basic idea of a VANET is straightforward as it can be seen as a particular form of Mobile Ad hoc NETwork (MANET). Consequently, in a first approach we could think on considering well-known and widely adopted solutions for MANETs and install them on VANETs.

A VANET is a wireless network that does not rely on any central administration for providing communication among the so-called On Board Units (OBUs) in nearby vehicles, and between OBUs and nearby fixed infrastructure usually named Road Side Unit (RSU). In this way, VANETs combine Vehicle TO Vehicle (V2V) also known as Inter-Vehicle Communication (IVC) with Vehicle TO Infrastructure (V2I) and Infrastructure TO Vehicle (I2V) communications (see Figure 2).

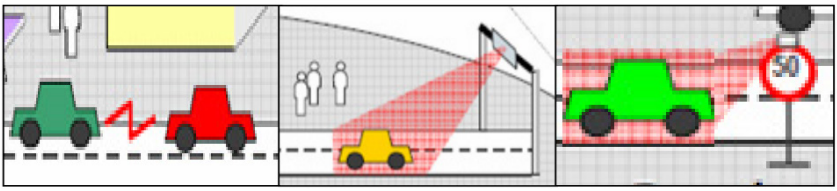


Figure 2: V2V, V2I & I2V Communications.

On the one hand, OBUs in vehicles will broadcast periodic messages with the information about their position, time, direction, speed, etc., and also warnings in case of emergency. On the other hand, RSUs on the roads will broadcast traffic related messages.

Additional communications can be also useful depending on the specific application. Among all these messages, routine traffic-related will be one hop broadcast, while emergency warnings will be transmitted through a multi hop path where the receiver of each warning will continue broadcasting it to other vehicles. In this way, drivers are expected to get a better awareness of their driving environment so that in case of an abnormal situation they will be able to take early action in order to avoid any possible damage or to follow a better route.

VANETs are expected to support a wide variety of applications, ranging from safety-related to notification and other value-added services. However, before putting such applications into practice, different security issues such as authenticity and integrity must be solved because any malicious behaviour of users, such as modification and replay attacks with respect to disseminated traffic-related messages, could be fatal to other users.

Moreover, privacy-regarding user information such as driver's name, license plate, model, and travelling route must also be protected. On the other hand, in the case of a dispute such as an accident scene investigation, the authorities should be able to trace the identities of the senders to discover the reason of the accident or look for witnesses. Therefore, specific security mechanisms for VANETs must be developed.

Great attention both from industry and academia has been received to this promising network scenario, and standards for wireless communications in VANETs are nowadays under preparation. In particular, IEEE 802.11p is a draft standard for Wireless Access in Vehicular Environment (WAVE), and IEEE 1609 is a higher layer standard on which IEEE 802.11p is based. At a superior level, Communications, Air-interface, Long and Medium (CALM) range is an initiative to define a set of wireless communication protocols and air interfaces for the so-called Intelligent Transportation System (ITS).

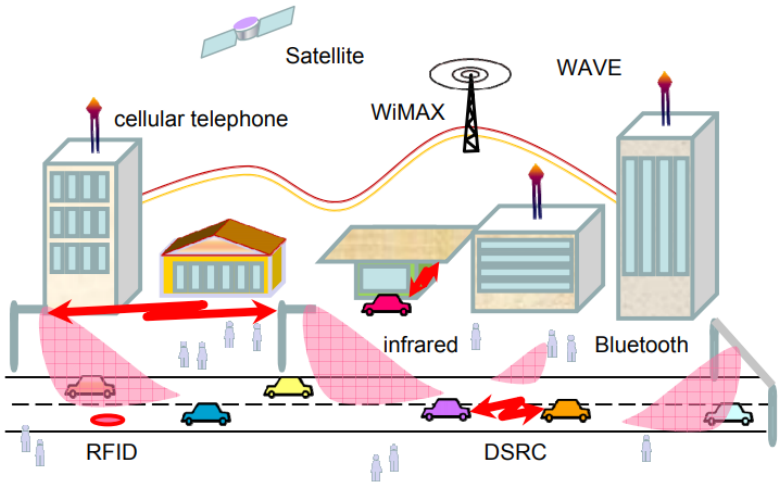


Figure 3: Convergence of technologies.

It is foreseeable that VANETs will combine a variety of wireless methods of transmission used by CALM and based on different types of communication media such as WAVE, infrared, cellular telephone, 5.9 GHz Dedicated Short-Range Communication (DSRC), WiMAX, Satellite, Bluetooth, RFID, etc. The current state of all these standards is trial use (see Figure 3).

In this way, the field of vehicular applications and technologies will be based on an interdisciplinary effort from the sectors of communication and networking, automotive electronics, road operation and management, and information and service provisioning. Without cooperation among the different

participants, practical and wide deployment of VANETs will be difficult, if not impossible.

In the future it could be expected that each vehicle will have as part of its equipment: a black box (EDR, Event Data Recorder), a registered identity (ELP, Electronic License Plate), a receiver of a Global Navigation Satellite System like GPS (Global Positioning System) or Galileo, sensors to detect obstacles at a distance lesser than 200 ms, and some special device that provides it with connectivity to an ad hoc network formed by the vehicles, allowing the node to receive and send messages through the network (see Figure 4). One of the most interesting components of this future vehicle is the ELP, which would securely broadcast the identity of the vehicle.

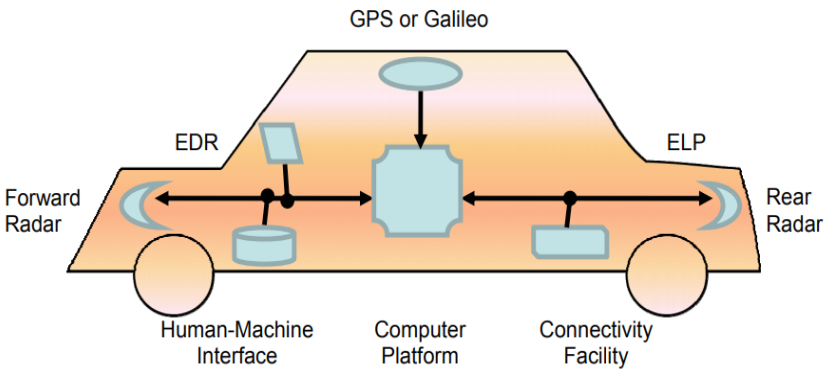


Figure 4: Components of a future vehicle.

This certificate is assumed to be emitted by a Certification Authority (CA) that is admitted as reliable by the whole network. The moments corresponding to the vehicle purchase and to the periodic technical inspections are proposed to be respectively associated to the emission and renovation of its public-key certificate. In general, symmetric authentication is acknowledged by most authors as not a valid option due to important factors in VANETs such as time and scalability.

Different security challenges of vehicular networks are here addressed, paying special attention to the application of several

known security primitives such as symmetric and asymmetric cryptography, strong authentication, data aggregation and cooperation enforcement.

7.4.1 Characteristics

There are several general security requirements, such as authenticity, scalability, privacy, anonymity, cooperation, stability and low delay of communications, which must be considered in any wireless network, and which in VANETs are even more challenging because of their specific characteristics such as high mobility, no fixed infrastructure and frequently changing topology that range from rural road scenarios with little traffic to cities or highways with a huge number of communications.

Consequently, VANET security may be considered one of the most difficult and technically challenging research topics that need to be taken into account before the design and wide deployment of VANETs.

Among the main key technical challenges the following issues can be remarked:

- The lack of a centralized infrastructure in charge of synchronization and coordination of transmissions makes that one of the hardest tasks in the resulting decentralized and self-organizing VANETs is the management of the wireless channel to reach an efficient use of its bandwidth.
- High node mobility, solution scalability requirements and wide variety of environmental conditions are three of the most important challenges of these decentralized self-organizing networks. A particular problem that has to be faced comes from the high speeds of vehicles in some scenarios such as highways. These characteristics collude with most iterative algorithms intended to optimize the use of the channel bandwidth or of predefined routes.

- Security and privacy requirements in VANETs have to be balanced. On the one hand, receivers want to make sure that they can trust the source of information but on the other hand, this might disagree with privacy requirements of the sender.
- The radio channel in VANET scenarios present critical features for developing wireless communications, which degrade strength and quality of signals.
- The need for standardization of VANET communications should allow flexibility as these networks have to operate with many different brands of equipment and vehicle manufacturers.
- Real-time communication is a necessary condition because no delay can exist in the transmission of safety-related information. This implies that VANET communication requires fast processing and exchange of information.
- The existence of a central registry of vehicles, possible periodic contact with it, and qualified mechanisms for the exigency of fulfilment of the law are three usual assumptions that are necessary for some proposed solutions.
- Communication for information exchange is based on node-to-node connections. This distributed nature of the network implies that nodes have to relay on other nodes to make decisions, for instance about route choice, and also that any node in a VANET can act either as a host requesting information or a router distributing data, depending on the circumstances.

Another interesting characteristic is the dependency of confidentiality requirements on specific applications. On the one hand, secret is not needed when the transmitted information is related to road safety, but on the other hand, it is an important requirement in some commercial applications.

As aforementioned, VANETs can be seen as a specific type of MANET. However, the usual assumption of these latter networks about that nodes have strict restrictions on their power, processing and storage capacities does not appear in VANETs. Another difference with respect to pure MANETs is that in vehicular networks, we can consider that access to a fixed infrastructure along the roadside is possible when RSU is available either directly or through routing.

When developing a simulation of a VANET (see Figure 5), some special features have to be considered:

- Each vehicle generally moves according to a road network pattern and not at random like in MANETs.
- The movement patterns of vehicles are normally occasional, that is to say, they stop, move, park, etc.
- Vehicles must respect speed limitations and traffic signals.
- The behaviour of each vehicle depends on the behaviour of its neighbour vehicles as well as on the road type.
- VANETs can provide communication over 5-10 Km.
- Two nodes cannot exist in the same location at the same time.
- Nodes usually travel at an average speed lower than 120 Km/h.

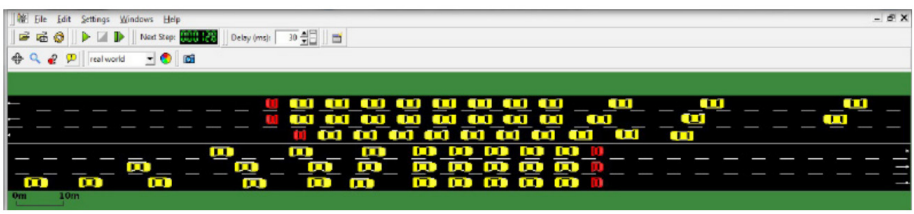


Figure 5: Example of simulation.

Despite the aforementioned differences between MANETs and VANETs, some security tools designed for their use in MANETs have been evaluated for their possible application in VANETs.

Such as it happens in MANETs, in VANETs the nodes are in charge of package routing. Up to now, several routing protocols originally defined for MANETs have been adapted to VANETs following different approaches.

Reactive protocols designed for MANETs such as Ad hoc On-demand Distance Vector (AODV) and Dynamic Source Routing (DSR) have been modified to be used in VANETs. Nevertheless, simulation results do not indicate a good performance due to the highly unstable routes. Consequently, we can conclude that those adaptations might be successfully used only in small VANETs.

In other routing protocols based on geographic location of nodes, the decisions related to package routing are taken based on street guides, traffic models and data collected with global positioning systems available in the vehicles.

According to simulations, this type of protocols based on geographic information seems to be the most promising for its use in different types of sceneries such as cities and highways. In particular, in VANETs it might be useful to send messages only to nodes in a precise geographic zone. Specific routing protocols with this characteristic have been designed, and mentioned in the bibliography as geocast routing. This way to proceed allows disseminating information only to interested nodes (for instance, in case of an accident, only to proximal vehicles, and in case of an advertisement, only to nodes that are in the zone of the advertised service). In a comparative study among different routing schemes is presented.

Also like in MANETs, routing in VANETs basically follows two ways of action:

- Proactive: All vehicles periodically broadcast messages on their present states (beacons) containing their ELP, position, timestamp, speed, etc., and resend such messages if it is necessary.
- Reactive: Each vehicle sends messages only after it detects an incident, generates a request, or must resend a received message.

We have an example of how to take advantage of the proactive mode when a parked vehicle is witness of an accident thanks to its sensors, and stores the corresponding data in its EDR, so that they could be later used to determine liabilities.

In the proactive mode, the frequent beacons are very costly. Furthermore, they imply the possibility of their use to track vehicles. This fact leads to the necessity of a solution that might consist in encrypted beacons. The high frequency of those beacons combined with the higher computational cost of asymmetric cryptography suggests the application of a hybrid solution combining it with symmetrical cryptography. This hybrid solution also seems the best option, independently of the routing protocol, for some specific applications.

7.5 VANETS SAFETY APPLICATIONS

After full deployment of VANETs, when vehicles can directly communicate with other vehicles and with the road side infrastructure, several safety and non-safety applications will be developed. Although less important, non-safety applications can greatly enhance road and vehicle efficiency and comfort.

7.5.1 Safety-Related

A possible application of VANETs for road safety, besides the warning dissemination of accidents or traffic jumps that constitute their main application, is the warning dissemination of danger before any accident or traffic jump has taken place. This would be the case for example of a high speed excess or a violation of a traffic signal (such as a traffic light or a stop sign). In these cases, when some vehicle detects a violation through its sensors, it must activate the automatic dissemination of warning messages communicating the fact to all neighbour vehicles in order to warn them about the danger.

An additional difficulty of this application is due to the fact that the dangerous vehicle is in motion. This implies that it is not clear what any vehicle that receives the message can do to avoid the danger without being able to identify the actual location of the guilty vehicle.

Another related application of VANETs in road safety is the warning dissemination of emergency vehicle approach.

The situations of vehicles that have suffered an accident or have met a traffic jump can be dealt in the same way as any other detection of anything that might be classified as an obstacle, such as extremely slow vehicles, results of possible natural phenomena on the road, stones, bad conditions of the pavement due to works on the road, or bad meteorological conditions like low visibility. In all these cases we have that the corresponding information is important for road safety, and that the incident can be characterized by a certain location and moment.

Consequently, in these cases of applications for driver assistance, the aforementioned hypothesis referring to the existence of a Global Navigation Satellite System in vehicles is fundamental because it allows locating both the own location and that of the detected incident (see Figure 6).

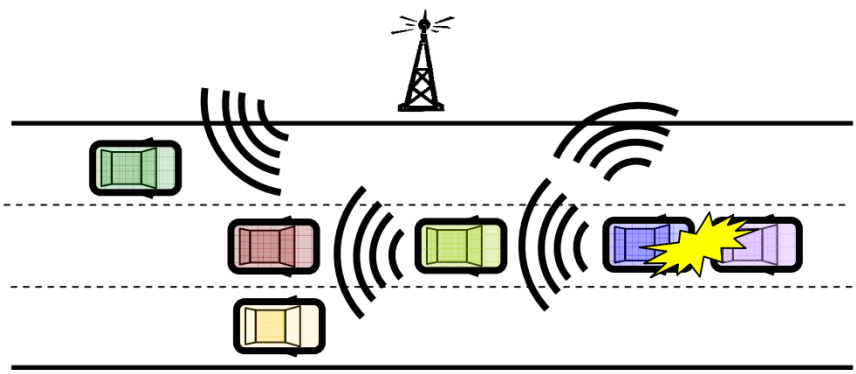


Figure 6: Accident warning.

Given the importance of the warnings of incidents for road safety, in these cases it would be advisable the use of an evaluation

system of messages previous to their massive dissemination. For example, we could stipulate that in the scenery of the incident at least a minimum number of vehicles higher than a pre-established threshold activates or signs the same warning. This can be implemented for example by means of a voting scheme among the vehicles in the area nearby the incident.

In addition, note that with this proposal, possible Denegation of Service (DoS) attacks and sending of false warnings are prevented. In this sense, note that, although privacy is an important aspect in VANETs, its protection cannot stop the use of information by the authorities in order to establish responsibility in case of accident. On the other hand, it is foreseeable that the reception of a warning of abnormal and/or potentially dangerous incident will have influence in the behaviour of the other drivers. For that reason, in these schemes it is necessary to consider possible attacks based on trying to inject or to modify messages in order to obtain an effect like for example a road free of vehicles.

In order to inform cars in their vicinity to warn their drivers earlier of potential hazards, so that they have more time to react and avoid accidents, vehicles exhibiting abnormal driving patterns, such as a dramatic change of direction, send messages including information derived from many sources like sensors, devices ABS, ESP, etc., use of airbags, speed, acceleration or deceleration of vehicle, as well as information originating from other sources like radars or video monitors, and SOS telephones or traffic lights used as repeaters to extend the dissemination rank of warnings.

From the combination of all these data, neighbouring vehicles can directly identify in many cases the type of incident by means of the interpretation of this information. A similar approach can be applied at intersections where cars communicate their current position and speed, making it possible to predict possible collisions between cars. There is another important case that does not correspond exactly to a warning of an incident with a determined location and moment, but has also important implications in road safety. That is the case of a warning of the presence of an emergency vehicle like police, ambulance, fire-fighters, etc. In this case, the

warning should include location, moment and foreseeable destiny or route of the emergency vehicle, and the objective is that the other vehicles can receive this information with enough time to clear the path of emergency vehicles in realtime, hence saving crucial time.

7.5.2 Non-safety-related

There is a whole variety of non-safety applications included in Value-Added Services (VASs), which can be provided through a VANET. Passengers in vehicles who spend a very long period in transit might be interested in certain application domain for vehicular networks consisting in the provision of many different types of information. Such information could be data about the surrounding area such as nearby businesses, services, facilities or road conditions, different entertainment-oriented services like Internet access (see Figure 7) or sharing multimedia contents with neighbours, and advertisement services. This diversity of possible applications comes from the fact that vehicular networks can be considered a form of pervasive network, that is to say, they operate anywhere and at any time.

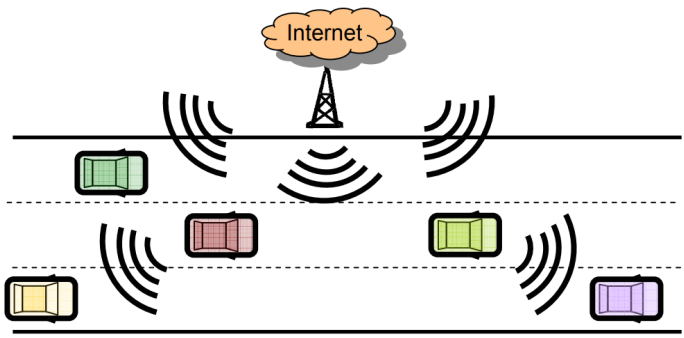


Figure 7: Internet access.

Vehicular networks could be also used for traffic monitoring. In particular, traffic authorities might be interested in obtaining information about road users so that for example they could get traffic flows to deduce current congestion levels and detect

potential traffic jams. In general, dissemination of that type of information among nodes can be used to manage traffic, not only in the aforementioned cases when an incident occurs, but also in normal conditions, when it can be used for the optimization of traffic flow.

Therefore, on the one hand, VANETs could be used for traffic management by extending drivers' horizons and supporting driving manoeuvres so that they provide drivers with information they might have missed or might not yet be able to see, in order to help them in decision making. A special traffic management application is a lane positioning system that uses inter-vehicle communication to improve GPS accuracy and provide lane-level positioning. Such detailed positioning allows the provision of services such as lane departure warning, as well as lane-level navigation systems.

On the other hand, if junctions are equipped with a controller that can either listen to communication between vehicles or receive messages from arriving vehicles, then the controller would be able to build an accurate view of the traffic at the junction through the aggregation of the received data corresponding to traffic conditions in the area, and could therefore adapt its behaviour to optimize the throughput. Traffic management applications could be also used to allow emergency vehicles to change traffic lights at signalized intersection in order to synchronize adequately to the objective of clearing the path.

An approach similar to the general case of traffic monitoring could be extended by the use of audio and video devices, which could be used for terrorist activities monitoring.

Closely related to traffic monitoring and a current particularly useful application of VANETs is traffic management. For instance, V2I solutions for road tolling are already deployed in certain places in the world to allow paying for road usage on congested roads, with prices depending on congestion levels. In the future, vehicular networks could enable that drivers are charged for their specific usage of the road network.

The idea of autonomous vehicles that are able to operate in urban areas while obeying traffic regulations is part of a collection of revolutionary applications called coordinated driving applications. This special type of safety-related applications improves performance and safety of participant vehicles through their collaboration with each other. Proposed coordinated driving applications focus mainly on three scenarios: adaptive cruise control, platooning and intersection management.

The simplest coordination application is adaptive cruise control, which performs control manoeuvres in order to maintain a safe distance for each vehicle to the vehicle in front by using forward sensors, wireless communication and cooperation among vehicles.

In a platoon, V2V communication is used to coordinate platoon members through a leader or a teamwork model in which autonomous vehicles follow a decentralized management scheme. The main benefits of platoon applications are: increase of road capacity and efficiency, reduction in congestion, energy consumption and pollution, and enhancement of safety and comfort. Demonstrations of cars travelling in platoons have already proven the feasibility of such a radical approach in certain protected settings. In particular, and have demonstrated the technique of coupling two or more vehicles together electronically to form a train.

Finally, the third mentioned coordinated driving application is intersection management for collaborative collision avoidance of autonomous vehicles while reducing delay in comparison to traffic lights or stop signs. This interesting application allows improving road safety through cooperative driving in dangerous road points where certain circumstances exist according to which several vehicles compete for a common critical point that all have to go over so that the VANET can offer support for certain driving manoeuvres. That is the case for example of the access to a highway or a road intersection without visibility or traffic lights, where it is convenient that vehicles act co-ordinately through group communications in order to avoid accidents.

Each application implies several important differences in the security schemes that are used. In order to use VANETs as practical support for advertisement dissemination, a system of incentives must be defined both for the advertiser and for the nodes of the VANET, so that both gain when disseminating the advertisements through the network. In this sense, the VANET can offer several advantages because the driver would be aimed to listen to advertisements, and even to help in their dissemination, if it obtains something in return, for example, some valuable good as gasoline. Obviously, in these cases it is necessary to define measures to prevent possible frauds of those who try to gain without receiving/redistributing the advertisements.

A similar incentive-based approach might be used for other Value-Added Services, like for example, the supply and demand of useful information like alternative routes, near parking zones, gas stations, hotels, restaurants, access points to Internet, etc. In all these cases it is fundamental that the information is encrypted in order to prevent access to non-authorized users who have not paid for the service. These other VAS applications have some similarities and differences with respect to the described advertising support service.

Both in the case when the information is a warning of incident or emergency vehicle, and in the case of dissemination of publicity or other VAS, it is remarkable that the messages have a definite origin (crashed/in traffic jump/emergency/VAS applicant vehicle, or advertiser business) but do not have a unique and definite destiny, what has clear implications in security issues. In fact, in all those cases the objective is to disseminate the message to the largest number of nodes but with different optimization criteria. In order to achieve such a goal the origin broadcasts the message to all the vehicles within its neighbourhood.

7.6 SECURITY PROPOSAL

This group formation is proposed as a valid strategy to strengthen privacy and provide authenticity, privacy and integrity protection,

while reducing communications in VANETs. To make it possible, group management within the network must be very fast to minimize time lost in that task.

In particular, we propose location-based group formation according to dynamic cells dependent on the characteristics of the road, and especially on the average speed. In this way, any vehicle that circulates at such a speed will belong to the same group within its trajectory. It is also proposed here that the leader of each group be the vehicle that has belonged to the same group for the longest time (see Figure 8).

According to our proposal, V2V between groups will imply package routing from the receiving vehicle towards the leader of the receiving group, who is in charge of broadcasting it to the whole group if necessary. If the cells have a radio that is greater than the wireless coverage of the OBU, the group communication may be carried out by proactive Optimized Link State Routing (OLSR).

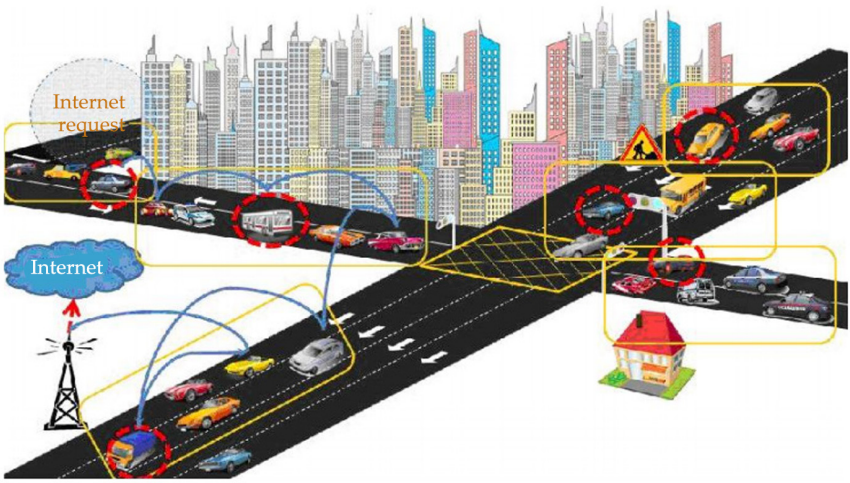


Figure 8: Proposal structure.

In the two phases corresponding to group formation and node joining, each new node has to authenticate itself to the leader through asymmetric authentication. Later, the leader sends a shared secret key to it, encrypted with the public key of the new

node. In particular, this secret key is shared among all the members of the group, and used both for V2V within the group and for V2V between groups. We propose the application of different cryptographic primitives for node authentication, while paying special attention to the efficiency of communications and to the need of privacy. In this way, we distinguish four different ways of authentication, which are analyzed in the following subsections.

7.6.1 I2V Authentication

Since privacy-preserving authentication is not necessary in I2V, we propose for such a case the use of Identity-Based Cryptography because it provides a way to avoid the difficult public-key certificate management problem.

Identity-Based Cryptography is a type of public-key cryptography in which the public key of a user is some unique information about the identity of the user (e.g. the ELP in VANETs).

The first implementation of an Identity-Based scheme was developed in, which allowed verifying digital signatures by using only public information such as the users' identifier. A possible choice for VANETs could be based on the modern schemes that include Boneh/Franklin's pairing-based encryption scheme, which is an application of Weil pairing over elliptic curves and finite fields.

7.6.2 V2I Authentication

Unlike I2V communication, in V2I communications privacy is an essential ingredient. Here we propose a challenge-response authentication protocol based on a secret-key approach where each valid user is assigned a random key-ring with k keys drawn without replacement from a central key pool of n keys.

According to the proposed scheme, during authentication each user chooses at random a subset with c keys from its key-ring, and uses them in a challenge-response scheme to authenticate itself to

the RSU in order to establish a session key, which is sent encrypted under the RSU's public key.

This scheme preserves user privacy due to the feature that each symmetric key is with a high probability (related to the birthday paradox and dependent on the specific choice of parameters) shared by several vehicles.

When a vehicle wants to communicate with the RSU, it sends an authentication request together with a set of c keys taken at random from its key-ring and a timestamp. All this information is then encrypted by the established session key. Note that a set of keys, instead of only one key, is proposed for authentication, because there is a high probability for the OBU to have one key shared by a large amount of vehicles. This makes it difficult to identify a possible malicious vehicle if just one key is used. However, there is a much lower probability that a set of keys be shared by a large number of vehicles, and so it is much easier to catch a malicious vehicle in the proposal.

After the RSU gets the authentication request from the vehicle, it creates a challenge message by encrypting a random secret with the set of keys indicated in the request, by using Cipher-Block Chaining (CBC) mode. Upon receiving the challenge, the vehicle decrypts the challenge with the chosen keys and creates a response by encrypting the random secret with the session key. Finally, the RSU verifies the response and accepts the session key for the next communications with the vehicle.

In the first step, in order to make easier the task of checking the key subset indicated in the request by the RSU, we propose a tree-based version where the central key pool of n keys may be represented by a tree with c levels. Each user is associated to k/c leaves, and each edge represents a secret key.

In this way, the key-ring of each user is formed by several paths from the root to the leaves linked to it. During each authentication process the user chooses at random one of its paths, which may be shared by several users. In this way, to check the keys, the RSU has to determine which first-level key was used, then, it continues

by determining which second level key was used but by searching only through those second-level keys below the identified first-level key.

This process continues until all c keys are identified, what at the end implies a positive and anonymous verification. The key point of this proposal is that it implies that the RSU reduces considerably the search space each time a vehicle is authenticated.

7.6.3 V2V Authentication inside Groups

At the stages of group formation and group joining, each new node has to authenticate itself to the group leader by using public-key signatures.

After group formation or group joining, the group leader sends a secret shared key to every new member of the group, encrypted with the public key of this new node (see Figure 9). Such a secret group key is afterwards used for any communication within the group both for node authentication and for secret-key encryption if necessary (e.g. for commercial applications).

In this way, the efficiency of communications inside the group is maximized because on the one hand certificate management is avoided, and on the other hand, secret-key cryptography is in general more efficient than public-key. Note that the use of a shared secret key also contributes to the protection of privacy.

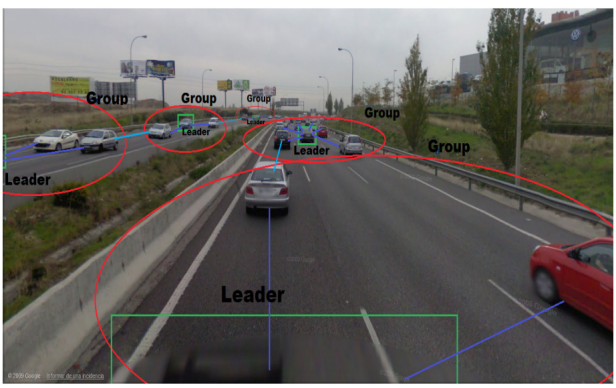


Figure 9: Group-based organization.

7.6.4 V2V Authentication between Groups

In order to protect privacy, group signatures might be proposed for node authentication between groups. A group signature scheme is a method for allowing a member of a group to anonymously sign a message on behalf of the group so that everybody can verify such a signature with the public key of the group. This group signature identifies the signer as a valid member of the group and does not allow distinguishing among different group members.

Essential for a group signature scheme is the group leader, who is in charge of adding group members and has the ability to reveal the original signer in the event of disputes. In this proposal, the group leader issues a private key to each vehicle within the group, which uniquely identifies each vehicle, and at the same time allows it to compute a group signature and prove its validity without revealing its identity.

In this way, any vehicle from any group will be able to communicate with any vehicle belonging to other group anonymously. In particular, a proposal for group signature might be based on the cryptographic primitive of bilinear pairings, which was also proposed for I2V authentication.

REFERENCES

1. Asuquo, P., Cruickshank, H., Morley, J., Anyigor Ogah, C.P., Lei, A., Hathal, W., and Bao, S., "Security and Privacy in Location-Based Services for Vehicular and Mobile Communications: An Overview, Challenges and Countermeasures," in IEEE Internet of Things, 2018.
2. Emara, K., "Safety-aware Location Privacy in VANET: Evaluation and Comparison," inIEEE Transactions on Vehicular Technology, vol. 66, issue 12, pp.10718-10731, 2017.
3. Feiri, M., Petit, J., and Kargl, F., "The Impact of Security on Cooperative Awareness in VANET," in IEEEVehicular Networking Conference, VNC, 2014.
4. Guo, N., Ma, L., and Gao, T., "Independent Mix Zone for Location Privacy in Vehicular Networks," in IEEE Access, 2018.
5. Laganà, M., et al., "Secure Communication in Vehicular Networks – PRESERVE Demo," in Proceedings of the 5th IEEE International Symposium on Wireless Vehicular Communications, 2013.
6. Mansour, M. B., Salama, C., Mohamed, H. K., and Hammad, S. A., "CARCLOUD: A Secure Architecture for Vehicular Cloud Computing," in 14th Embedded Security in Cars Europe Conference, ESCAR, Germany, Nov. 2016.
7. Pathan, A.S.K., "Security of self-organizing networks: MANET, WSN, WMN, VANET," CRC press2016.
8. Rasheed, A., Gillani, S., Ajmal, S., and Qayyum, A., "Vehicular Ad Hoc Network (VANET): A Survey, Challenges, and Applications," in Vehicular Ad-Hoc Networks for Smart Cities,pp. 39-51, Springer, Singapore, 2017.
9. Samara, G., and Al-Raba'nah, Y., "Security Issues in Vehicular Ad Hoc Networks (VANET): a survey," arXiv preprint arXiv:1712.04263, 2017.

10. Siddiqui, N., Husain, M.S., and Akbar, M., "Analysis of Security Challenges in Vehicular Adhoc Network,"in Proceedings of International Conference on Advancement in Computer Engineering &Information Technology, IJCSIT, pp. s87-s90, 2016.
11. Singh, A., and Kad, S., "A review on the various security techniques for VANETs," inProcedia Computer Science, vol. 78, pp.284-290, 2016.
12. Wang, S., and Yao, N., "A RSU-aided distributed trust framework for pseudonym-enabled privacy preservation in VANETs," in Wireless Networks, pp.1-17, 2018.
13. Zhu, L., Zhang, Z., and Xu, C., "Security and Privacy Preservation in VANET,"in Secure and Privacy-Preserving Data Communication in Internet of Things, pp. 53-76, Springer, Singapore, 2017.

INDEX

Vehicular ad Hoc Network (VANET)

Recently, with the development of vehicle industry and wireless communication technology, vehicular ad hoc networks are becoming one of the most promising research fields. Vehicular ad hoc networks (VANETs) have recently been proposed as one of the promising ad hoc networking techniques that can provide both drivers and passengers with a safe and enjoyable driving experience. VANETs can be used for many applications with vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. In the United States, motor vehicle traffic crashes are the leading cause of death for all motorists between two and thirty-four years of age. VANETs which use vehicles as mobile nodes are a subclass of mobile ad hoc networks (MANETs) to provide communications among nearby vehicles and between vehicles and nearby roadside equipment but apparently differ from other networks by their own characteristics. Vehicular ad hoc networks (VANETs) have been quite a hot research area in the last few years. Due to their unique characteristics such as high dynamic topology and predictable mobility, VANETs attract so much attention of both academia and industry.

This book deals with the basic architecture of networks, and discusses three popular research issues and general research methods, and ends up with the analysis on challenges and future trends of VANETs. A viable choice for spectrum sensing due to its simplicity, low computational cost, and ability to be applied on any kind of deterministic signal is energy detection (ED). However, hidden terminal and low SNR problems due to shadow-fading put fundamental limits to the sensing performance and practical entailments in designing of cognitive vehicular networks. Extensive modeling efforts are then being carried out to cope with varying channel characteristics, particularly multipath fading and shadowing. The message routing in vehicular ad hoc networks (VANETs) is an attractive and promising area for research. These networks do not have a central coordination, the nodes are mobile, and the topology is highly dynamic, making the routing process a big challenge, since it is responsible for ensuring message delivery with small overhead and delay. This book presents vehicular ad-hoc networks (VANETs) from their onset, gradually going into technical details, providing a clear understanding of both theoretical foundations and more practical investigation.

Don Cooray received the engineering degree in computer science and PhD in Information and Communication Engineering. He has written several articles, research reports, and research papers on data communication published in peer reviewed journals. His research interest topics include wired and wireless networks, intrusion detection systems, secure routing protocols in wireless ad hoc and sensor networks.

ISBN 978-1-98467-791-4

BIBLIOTEX
Digital Library