Deep Learning-based Data Dissemination Scheme in Content-Centric Internet of Vehicles

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Technology

in

Computer Applications

Submitted By

Amuleen Gulati

601634002

Under the Supervision of

Dr. Neeraj Kumar

Associate Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

PATIALA-147004

June 2018

Certificate

I hereby certify that the work, which is being presented in the thesis, entitled **Deep Learning-based Data Dissemination Scheme in Content-Centric Internet of Vehicles**, in partial fulfillment of the requirements for the award of the degree of **Master of Technology in Computer Applications** and submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of **Dr. Neeraj Kumar** and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for the award of any other degree of this or any other University.

Amleen

Amuleen Gulati
601634002
M.Tech(CSA)

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

Dr. Neeraj Kumar
Associate Professor
Computer Science and Engineering Department
TIET, Patiala

i

Acknowledgements

This research work would be incomplete without acknowledging the people who supported and guided me for the successful completion of this task.

First of all, I wish to acknowledge the benevolence of God who gave me courage and strength to face the challenges and to overcome the obstacles that occurred while working on this task.

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Neeraj Kumar**, Associate Professor, Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala for his valuable guidance and continual encouragement throughout this work. The appreciation and continuous support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

I am thankful to **Dr. Gagangeet Singh Aujla**, former PhD Research Scholar, Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala for being an indispensable part of my research work. Apart from directly contributing to some parts of my research, he has provided me with immense support and guided me in right direction. His guidance has helped me polish my skills and gain confidence in the research field.

I am grateful to **Dr. Maninder Singh**, Hon'ble Head of Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala for his kind support and providing basic infrastructure and healthy research environment.

I would also thank the Institution, all faculty members of Computer Engineering Department, Thapar Institute of Engineering and Technology, Patiala for their special attention and suggestions towards this work.

- Amuleen Gulati
601634002

ii

Abstract

The IP-based IoV networks aimed to provide vehicle-to-vehicle communication to enable information exchange between vehicular users. However, due to the address-content binding mechanisms used in IP-based protocols, the IP-based networking proves to be costly and inefficient in terms of overhead incurred due to DNS lookups, load balancing, etc. The Content Centric Networking (CCN) technology provides an alternative by enabling the users to share the information based upon the content, thereby providing a content-centric model as opposed to host-centric model of IP-based networks. Previous studies have shown that CCN uses transfer of named data and in-network caching mechanisms to enable communication in an infrastructure-less environment. But its broadcast-based access mechanism has given rise to new problems related to network congestion and increased delays. This problem is more severe in vehicular network scenario where the network nodes are mobile and require dynamic addressing. This thesis addresses this issue by putting forward the design of a data dissemination scheme in an IoV scenario using Content-Centric networking architecture is proposed.

The proposed scheme works in three phases: (1) Firstly, the vehicles will be screened based upon the remaining energy of the vehicles. This is done to ensure that the vehicles do not run out of energy during data transmission. (2) In the next step, the connection probability of each potential vehicle pair is computed with an aim to promote only stable and long lasting connections to the next level. (3) In the last step, a convolutional neural network (CNN)-model is used for community detection by using the social layer information of vehicular users.

CNN is used to identify the ideal vehicle pairs, which can share data to ensure minimum delay and high data availability. This application is based on the assumption that the users which belong to the same community are more likely to download the same type of information and hence, rather than using cellular links for data transfer, the vehicles can get the data from other nearby vehicles belonging to the same community and hence minimize the cost of cellular link usage and delay. To improve the performance of CCN, several hyper parameters are adjusted and tuned to achieve optimal performance.

Keywords: Content-centric Networking, Convolution Neural Network, Deep Learning, Internet of Vehicles.

Table of Contents

List of Figures

List of Tables

List of Notations

ξ_i	value of objective function
$U_i(N)$	utility gained by the vehicle
u_i	velocity of vehicle
V_i	vehicle
$H(t)$	headway distance at time t
e_{ij}	energy required for data exchange
ψ_{v_i}	remaining energy
E_o	threshold energy
$F_{ij}(t)$	connection probability
p_o	threshold connection probability
E_i	energy of vehicle
ϕ_{v_i}	energy indicator function
$\varphi_{v_{ij}}$	energy consumed in data exchange
τ	mean first passage time
μ	vehicle drift
$F_{ij}(t)$	CDF of connection time
$\theta_v ij$	connection function
L	communication range
η_{in}	input size
$\eta_{padding}$	padding size
ηout	output size
P_{ij}	interest matching probability
Z_{ij}	hidden layer output
S	social score

List of Abbreviations

IoT	Internet of Things
IoV	Internet of Vehicles
TCP	Transmission Control Protocol
IP	Internet Protocol
CCN	Content Centric Networking
FIB	Forwarding Information Base
PIT	Pending Interest Table
CS	Content Store
V2V	Vehicle to Vehicle
CNN	Convolutional Neural Network
TTL	Time To Live
VPN	Virtual Private Network
VoI	Voice over Internet
ICN	Information Centric Networking
ReLU	Rectified Linear Unit
GPU	Graphical Processing Unit
NLP	Natural Language Processing
D2D	Device to Device
DNS	Domain Name Servers
DT	Distance Table
MPR	Multi Point Relay
PMPR	Publisher Multi Point Relay
PDLP	Packet Diffusion Limited Protocol
CODIE	Controlled Data and Interest Evaluation in Vehicular Networks
CONET	COntrolled data packets propagation in vehicular NETworks
ISD	Interest Satisfaction Delay
ISR	Interest Satisfaction Rate
AP	Access Point
V2I	Vehicle to Infrastructure
NDN	Named Data Networking

MGRS	Military Grid Reference System
GPS	Global Positioning System
VANET	Vehicular Adhoc Network
DADT	Density-Aware Delay Tolerant
RSL	Recent Satisfied List
NSL	Neighbour Satisfied List
VIN	Vehicular Information Network
GOFP	Geographic Opportunistic Forwarding Protocol
VNDN	Vehicular Named Data Networks
POI	Position of Interest
TAG	Traffic Aware Geographic forwarding protocol
RNC	Region Needed-to-be Covered
DaS	Drive and Share
VSN	Vehicular Social Networks
BNL	Bayesian Non-parametric Learning

Chapter 1

Introduction

This chapter provides a brief introduction to the work carried out in this thesis work. It states the basic concepts of the work performed along with stating the motivation for carrying out the research and a brief outline of research gaps. Besides stating the definitions, it underlines the objectives to carry out this work and the research motivation. The chapter is closed with a brief summary of the work presented in this thesis.

The exponential increase in the number of connected devices in the recent years, together with the emergence of newer technologies has paved way for a world-wide network of interconnected objects, popularly known as Internet of Things (IoT) [1], which is aimed at embedding communication capabilities in the things that surround us. It is a network wherein billions of devices can be interconnected to share data and information among the users in a seamless manner. Figure 1 clearly shows the humongous increase in number of interconnected devices over the years. As per the report, the number of connected devices surpassed world population somewhere between the years 2008-2009, around the years in which the Internet of Things was born. It includes all 'smart' devices including computers, laptops, mobile phones, PDAs, home appliances, vehicles etc. Within the huge IoT domain, Internet of Vehicles(IoV) is an integral sub-domain, consisting of vehicles equipped with wireless communication interfaces to enable information exchange within a short range. With an increase in the number of on-road connected vehicles, the generation of massive amounts of heterogeneous data is inevitable from such a vast interconnection of vehicles as is evident from [1].

Vehicles in an IoV network can exchange safety related information including collision information, cooperative driving etc. or non-safety related data including toll service, entertainment(videos, music), weather information and so on. Traditionally, the vehicular users network were based on a fixed infrastructure consisting of an Access Point(AP) and Road Side Units(RSUs). The vehicles accessed the required data through cellular networks or WiFi based on 802.11 standard. But the limited coverage of RSUs, high cost and limited bandwidth of cellular networks and mobility of vehicular users posed serious challenges to the infrastructure-

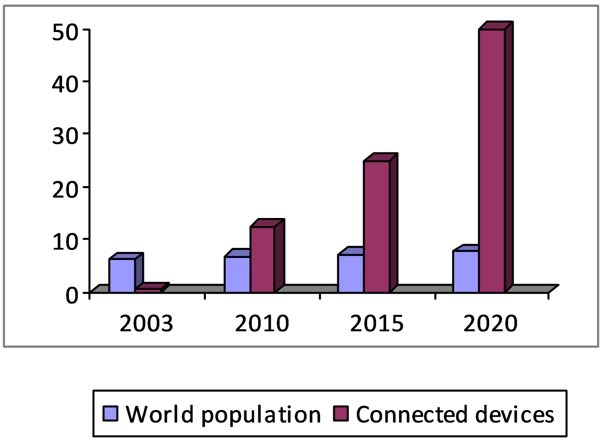

Figure 1.1: Number of connected devices vs world population (in billions)

based vehicular network. To overcome these issues, 802.11p standard was introduced to enable Intelligent Transport Systems(ITS). It allows vehicles to exchange information with each other(V2V) as well as with roadside infrastructure(V2X) in a fixed, licensed 5.9GHz frequency band. Data packets could be directly exchanged between vehicles without passing through the AP. 802.11p aimed to facilitate data dissemination in vehicular network with minimized latency. It uses the communication mechanisms provided by 802.11 standard and operates in Distributed Short Range Communication(DSRC), which is a communication technology operating in 5.9GHz frequency band. It offers a transmission rate of 3-27Mbps in a communication range of 1km, for a V2V and V2X communication.

In order to fully exploit the underlying potential of IoV technology, vehicles and services should be readily available, discoverable and useful by other vehicular users. This requires sharing large amounts of data across interconnected vehicles. The current IoV technology(802.11p) relies upon the underlying TCP/IP protocols [2] to discover and share data across multiple objects. Each object in an IoV network is uniquely addressable and uses standard communication protocols to provide the required content and services to the end users. This makes the network performance dependent on bandwidth resources which are limited compared to the massive amounts of data. Once the available bandwidth is exhausted, packets start getting dropped, network congestion occurs and the network performance goes down drastically.

The situation is worsened with the increase in number of vehicles and the amount of data in the network. In order to overcome this problem and to achieve a better network performance, TCP/IP implements several congestion control algorithms [3] including TCP Tahoe, TCP Reno, TCP Vegas, CUBIC and so on. However, these algorithms cannot be deployed in the current scenario to guarantee congestion-free networks. The reason for this is attributed to their implementation: Tahoe does not perform well for short-lived connections, Reno performs poorly in wireless scenarios subject to high packet losses, slow convergence in case of CUBIC and instability with small buffers in Vegas, being some of them [4] [5] [6]. Another common shortcoming in each of these algorithms is that these are based on host-centric TCP/IP model, which requires data sharing objects to be known in advance. This means an object needs to access the data by location of its host. However, the users of the content are only concerned about the data and are oblivious to its location. Therefore, address-content binding in this architecture is a big concern.

In order to decouple content from the location at which it is stored, the concept of content-centric networking (CCN) was proposed [7] [8]. In CCN, the user or object can access data by name rather than by address. The individual nodes in the network do not need to be IP-addressable. In order to achieve this, each node maintains a set of data structures: a Forwarding Information Base (FIB) which has the responsibility of maintaining forwarding information of data requests, a Pending Interest Table (PIT) which stores information regarding the data requests received by the respective node and a Content Store (CS) which caches the content received by it for future use. The node which wants to access data broadcasts the Interest packet specifying the data by its name and its own identifier. All intermediate nodes continue to rebroadcast this packet until it reaches a node holding the required content in its CS. This node, upon receiving the interest packet, sends back the requested content encapsulated in data packets through the reverse path of the interest packet. This architecture makes content addressable by name and hence eliminates the need for IP allocation to each node. This is particularly useful in IoT scenario where mobility of devices poses a problem in terms of IP-addresses' dynamic assignment and re-assignment. Also, it has potential to be the foundation of future Internet as it will eliminate the IP-address depletion problem. In spite of its many advantages, the broadcast-based access mechanism of this architecture poses serious congestion problems. The traditional TCP/IP congestion control protocols cannot be used in this scenario as the congestion window in TCP puts an upper limit on the number of unacknowledged packets in transit between a pair of hosts. But since, the hosts which would be communicating is not known in advance in CCN architecture due to address-content decoupling,hence, these protocols cannot be applied in CCN scenario.

1.1 Traditional Internet vs CCN

The hourglass design of traditional Internet made its outline exquisite and path-breaking. The core of this design is a straightforward and general network layer (IP) which implemented the protocols for executing all the functionality required to realise a world wide interconnection of smart devices within a single network. This thin waist of the Internet architecture was the prime influence on the Internets surging development rate; however, one of its design choices is the primary reason of recent Internet issues. The Internet was originally devised as a communication system so the communication end points are the only main units that can be named in IP packets. The increasing development in the fields of e-commerce, PDA applications, social networking, and digital media has led to the Internet essentially being utilized more as a distribution system rather than merely a communication system. Distribution networks are basically broader than communication networks and taking care of dissemination issues with a communications network is complicated as well as prone to errors.

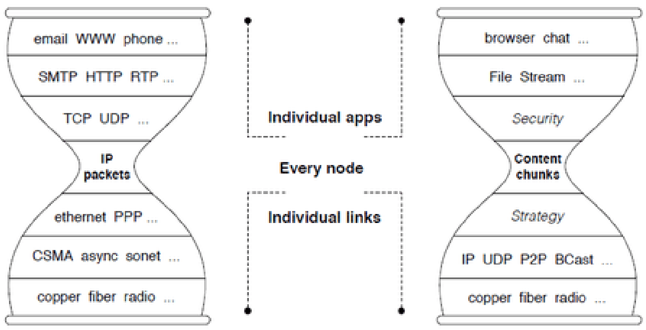

Figure 1.2: Comparison of traditional Internet and CCN architecture

Content Centric Networking architecture has broadened the scope of the 'thin waist' by allowing the formation of a totally general distribution system, while still retaining the Internet's hourglass design. The basic component of this advancement is the removal of hinderance which allowed only the endpoints to be named in the IP packets, whereas CCN architecture permits each and every network entity including end hosts, data chunks, commands etc. to be named in the packets. This seemingly straightforward change has made CCN architecture powerful by enabling the CCN network to utilize a majority of the Internet's existing and novel engineering principles to proficiently take care of communication issues as well as difficulties in network dissemination and control.

1.2 CCN Architecture

In general, CCN architecture is comprised of three data structures, each holding different kinds of data to facilitate sending and receiving of data across various nodes in a content-centric network. This is done by sending the query in the form of named data packets, also called Interest packets, in the network to fetch the desired content. When the data matching the queried content is found, it is returned to the requester of the data encapsulated in the form of Data packets, by following the reverse path of the query.

There are three major components that are essential to any node in a CCN model; these are Forwarding Information Base (FIB), Pending Interest Table (PIT) and Content Store (CS). All these components together serve the purpose of extracting named content while minimizing the amount of packets disseminated in the network.

1. FIB is responsible for storing the routing information at each node. Each entry in an FIB consists of the data identifier, along with the identifier of the face at which a request for that particular data is to be forwarded.

2. PIT at each node keeps record of similar requests that have been passed on by this node but whose reply has not yet been received. This structure helps in reducing the network congestion as the requests which want to fetch the same data objects are not forwarded again in to the network.

3. CS is a cache maintained at every node which serves to further decrease the network traffic by caching the content based upon some caching policy intrinsic to each network.

1.3 Data Dissemination in CCN

The data dissemination in a content-centric network is carried out in two-steps as follows:

1. The first step involves requesting the required data by a node in the form of 'Interest Packet'. This interest packet contains an identifier of the requesting node($node\_id$), the identifier of the data($data\_id$) which is requested and a Time To Live (TTL) field. This interest packet is broadcasted into the network.

2. Each node, upon receiving the interest packet, invokes a Content Store (CS) look up to

determine if it holds the requested content. The CS look up is carried out by matching the $data\_id$ from the interest packet with the entries of CS. In case a match is found, the node returns the requested content as 'Data Packets' through the reverse path of interest packet.

3. In the alternate case, when a node does not hold the requested data, it invokes a Pending Interest Table (PIT) lookup procedure, in which the $data\_id$ is matched with the entries of PIT to determine whether an awaiting request for the same data already exists on the node. If it exists, then the node simply adds the new request in its PIT and waits for the corresponding data packet.

4. In case a PIT entry does not exist, the node adds a fresh entry to its PIT corresponding to the received $data\_id$ and then invokes a Forwarding Information Base (FIB) lookup procedure, which returns the next face at which the interest packet must be forwarded.

5. The steps two to four are followed recursively, either until the requested data is returned by some intermediate node, or until the TTL of the interest packet expires. In the latter case, the requesting node must generate a fresh interest packet to request the data.

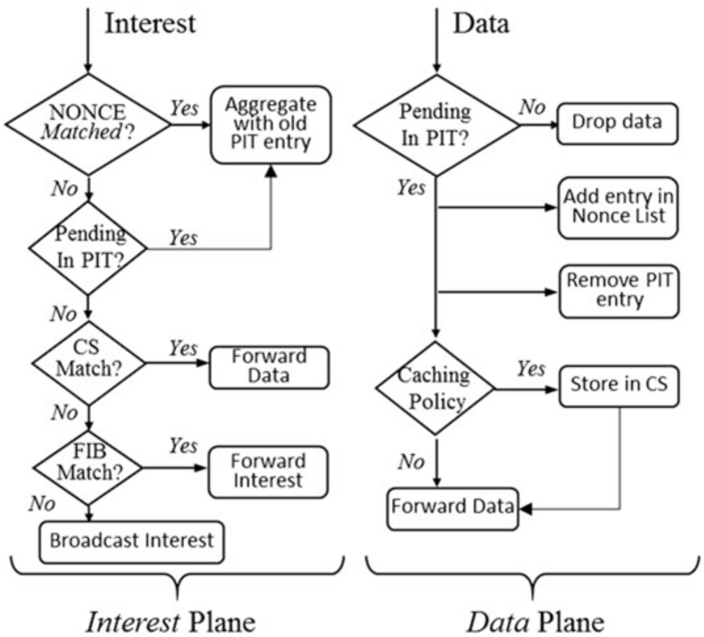

Figure 1.3: Flowchart of data dissemination process

1.4 Key Features of CCN

Following are some of the key features offered by CCN architecture, which makes it suitable for application in the current network scenario.

1. **Eliminates manual configuration:** It simplifies the usage of network by minimizing the set-up time and eliminating the need for manually configuring the network through firewalls, Virtual Private Networks or synchronization protocols.

2. **Supports mobility:** It provides a consistent and uninterrupted service to end users by enabling individuals to send and receive content digitally from diverse locations, multiple devices and varying networks.

3. **Reduces congestion:** It minimizes network congestion and delay due to its inherent design which does not allow it to send extraneous and unnecessary content through the network.

4. **Reduces cost:** It minimizes the operational cost of the network by eliminating the need for broadcasting routing information and control messages.

5. **Improves reliability:** It enhances the reliability of network by robustly delivering the content, irrespective of its location and network size.

6. **Provides security:** It reduces security issues by securing the content instead of the channel through which it is disseminated. CCN modifies the security paradigm by shifting from link protection to content protection.

7. **Supports data prioritization:** It places the power with the user by allowing end users to prioritize specific type of content over others and prioritizing their needs within the network.

Owing to its content-centric approach, ICN has lowered the complexity of user tasks in terms of generating queries to retrieve content. But its broadcast-based access mechanism has given rise to new problems related to network congestion and increased delays. This problem is more severe in vehicular network scenario where the network nodes are mobile and require dynamic addressing. To address this issue, we propose a design of a deep learning-based data dissemination scheme in vehicular scenario using Content-Centric networking architecture.

1.4.1 Convolutional Neural Networks (CNNs)

A convolutional neural network is essentially many layers of convolutions with nonlinear activation functions like ReLU or tanh applied to the results of the previous layers. In a customary neural network, each input neuron is connected to each of output neurons in the successive layers, and finally terminated at a fully connected affine layer which gives the output. But instead of fully connected layers, CNNs carry out convolutions over the input layers to calculate the final output. This generates local connectivity of neurons and divides the input layer into regions, whereby each region in the input layer is connected to a neuron in the output layer. Different filters are applied by each layer on the input neurons,and the results are combined in a fully connected output layer. In the training phase, based upon the application for which CNN is deployed, it automatically learns the values of several parameters.

One of the primary advantages of CNNs is their speed. They can perform the same task in lesser time as compared to conventional neural networks. Another advantage of CNNs is that these can be used to model an n-gram vocabulary by adjusting the number and size of their filters. The CNNs convolve the input layer to produce intermediate output layers in which the output is represented in a compact manner. At the final layer, a fully connected layer combines the results of all intermediate layers to compute the final output.

1.4.1.1 CNN Hyper-parameters

1. **Narrow vs. Wide convolution:** This parameter defines how the convolution is applied to different parts of the input sentence. The input text is given as a matrix to the Convolutional Neural Network. An n $\times$ n filter can be applied at the center of the input matrix but applying the same filter at the edges does not work well. This is due to the fact that the filter can not be applied to a matrix element that does not have any element in even one of its neighbourhood. This problem is eliminated by using zero-padding. By using the technique of zero-padding, each element that does not have a neighbouring element is padded with 0's, i.e., additional 0's are inserted at places where required, in order to apply the filter. As a result, the application of filter at each element becomes possible and a larger output matrix can be obtained. This ensures that none of the features of input matrix are ignored during convolution. This process of using zero-padding in CNN is called wide convolution and the absence of it is called narrow convolution.

Wide convolution is particularly necessary in cases where the filter size is considerable large as compared to the size of the input matrix to ensure a reliable output. The size of the output is computed according to the following equation:

$$\eta_{out} = \eta_{in} + 2 * \eta_p - \eta_f + 1 \tag{1.1}$$

where η_{out} is the size of output, η_{in} is the size of input, η_p is the size of padding, and ηf is the size of applied filter.

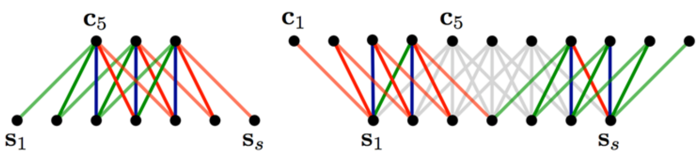

Figure 1.4: Narrow convolution (left) vs. wide convolution (right)

2. **Stride Size:** Another hyper parameter for convolution neural network is the stride size, which defines by how much the filter will be shifted at each step. In the instance where stride size is chosen as 1, the successive usages of the filters partly cover each other. On the other hand, if a large value is taken as the stride size, it results in a small-sized output due to fewer filter applications. The diagram below demonstrates the application of different stride sizes (1 and 2) to an input array:

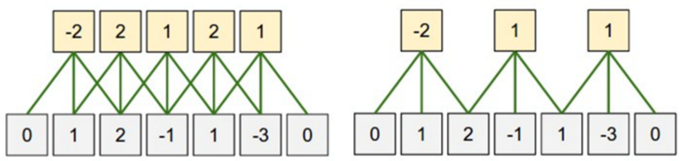

Figure 1.5: Effect of stride size in convolution

3. **Pooling Layers:** In general, a pooling layer is applied after each convolutional layer in a CNN. A pooling layer sub-samples the input given to it. There are three ways to apply pooling: max pooling, min pooling and average pooling. The main aim of applying the pooling layer is to combine the results of different filters at a single layer to detect the most common feature for that layer. The pooling layer can be applied to a small window or the complete input, as required. The following diagram depicts a max pooling layer applied to the 2×2 input matrix:

There are several advantages of using pooling layer listed as follows:

(a) In general, classification tasks require a fixed size output which is achieved by using pooling. It combines outputs of several layers to compute a single output. It allows us to use variable size input, and will give a fixed size output equal to the number of filters, regardless of the filter and input size.

(b) It helps in detecting the most important features of the input while suppressing the redundant information. It can successfully capture the local information like 'not exist' being different from 'exist not', however we lose the global information about content locality. It works by assigning a large value if a content type is present in the text.

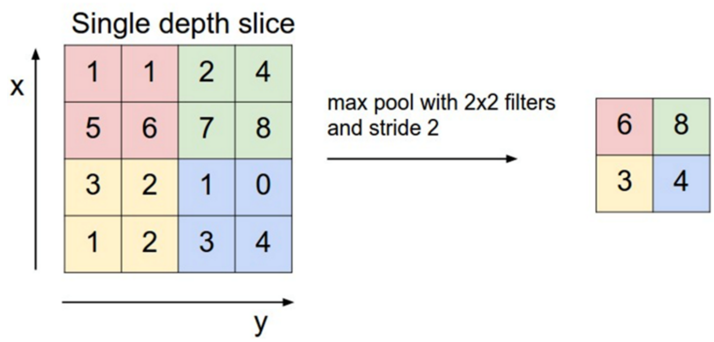

Figure 1.6: max pooling in CNN

4. **Channels:** These provide varied views of the input matrix. For instance, in image recognition there are essentially three channels, viz., RGB (red, green, blue). A convolution can be applied either with equal or varying weights across channels. In an NLP task, there are different channels for various word embeddings or there can be a single channel for a sentence represented in many languages, or expressed in multiple ways.

1.5 Research Orientation

This section describes the motivation to carry out this research and the key research areas selected for this research.

1.5.1 Research Motivation

The depletion of IP-addresses and an increasing demand for mobility support in IoT devices has forced the Internet architecture to move from traditional host-centric approach to a more flexible content-centric approach. CCN architecture allows end users and devices to remain oblivious to the location of content and access required data by name. However, this approach is based on broadcasting Interest and Data packets which leads to the problem of network congestion which, if not curbed, may have severe impact on network performance. The problem becomes more serious due to the fact that traditional TCP/IP congestion control protocols cannot be used in CCN scenario due to the sole crux of this architecture, which allows communication between unknown hosts. Therefore, there arises a need to develop an approach for data dissemination in CCN scenario to minimize delay and improve the overall network performance.

There has to be some mechanism to limit the number of potential objects with which an object may share data. Transient, short-lived connections must be avoided to reduce the overhead incurred by repeated link establishment. In addition, the energy dissipation of each vehicle participating in a data exchange also needs to be taken into consideration to ensure that vehicles do not run out of energy during a data exchange. The social information available from vehicular users can also be analyzed using deep learning approach based on CNN to screen vehicles with matching interests. This physical and social layer information can be then combined to select the optimal V2V pairs for vehicular communication.

1.5.2 Research Gaps

Owing to its content-centric approach, CCN has lowered the complexity of user tasks in terms of generating queries to retrieve content. But its broadcast-based access mechanism has given rise to new problems related to network congestion and increased delays. This problem is more severe in vehicular network scenario where the network nodes are mobile and require dynamic addressing.

Two major bottlenecks of implementing CCN architecture in V2V communications are limited frequency spectrum and mobility. Many authors over the years have proposed solutions for effective data exchange in Vehicular environment. For example, Ren *et al.* [9] proposed a device-to-device (D2D) channel selection framework on the basis of power control scheme in order to achieve optimal network performance. Sun *et al.* [10] formulated the radio resource allocation as an optimization problem for D2D-based vehicular communications. In another work, Yu *et al.* [11] proposed an enhanced C-RAN architecture with geographically distributed cloudlets, integrated with the original cloud, to provide local services.

All these works focus on the data associated with the physical layer and do not utilize the information related to the social layer available from the vehicular users. Social layer data available from the Vehicular environment contains a variety of information in terms of demographic, temporal and spatial attributes. An analysis of this information gives an insight of the interests of respective vehicles. As vehicles which have the same social interests have a higher probability of accessing the same type of data, a data exchange between such vehicles-to-vehicle (V2V) pairs leads to a higher hit ratio of finding requested content. Thus, in the Vehicular environment, social layer information also needs to be analyzed.

Some works based on the social interest information also exist in the literature [12]- [13]. Several vehicular applications take into account this information such as-Social Drive [12], Road Speak [15], Road Sense [16], and Verse [13]. These applications are based on the key-words in social information and invoke interest matching procedures to identify vehicles with similar interests so that they can share data.

But, one major drawback with all these works lies in their hard-coded parameters. Several parameters and their values must be defined prior to training the model. This reduces the flexibility of algorithm parameters as they must be calculated in advance and they cannot be auto-tuned according to the training set.

1.5.3 Key Research Contributions

In order to overcome the problems of IP address depletion along with IP address re-assignment, periodic updating of routing tables and increased network latency, content-centric networking (CCN) architecture is used, using which, instead of focusing on the physical location of the data which is to be retrieved, the main emphasis is given to the content which is required. Therefore, instead of accessing a data item by specifying the IP address of the host on which it is stored, the users can simply fetch the data by name. As a result, the problem of IP-address depletion is resolved, and at the same time, the Domain Name Servers (DNS) lookups are no longer required which reduces the latency in the network.

Further, to eliminate the need of using hard-coded parameters, convolutional neural network (CNN) is used, which is a flexible and completely trainable system where all the parameters are adjusted on the basis of the training data. Therefore, CNNs can be used to analyze the social data to perform interest matching of the vehicles. Social layer data available from the Vehicular environment contains a variety of information in terms of demographic, temporal and spatial attributes. An analysis of this information gives an insight of the interests of respective vehicles. As vehicles which have the same social interests have a higher probability of accessing the same type of data, a data exchange between such vehicles-to-vehicle (V2V) pairs leads to a higher hit ratio of finding requested content. Thus, in the Vehicular environment, social layer information can be analyzed using CNNs and combined with the physical layer information such as the energy dissipation and connection probability to build optimal V2V pairs, which can participate in data exchange.

Following are the key research contributions:

1. An energy estimation scheme for identifying the vehicles that can participate in data dissemination is designed. In this scheme, vehicles having energy value greater than threshold are selected for participation in data dissemination process.
2. A connection probability estimation scheme is presented to determine the probability of connection between V2V pairs using Weiner process model. Using this scheme, stable and long-lasting connections are identified.
3. In order to implement community detection, a social relationship estimation approach using CNN-based learning is designed. In this scheme, optimal V2V pairs are identified on the basis of trust, connection probability, energy available, and social score.

1.6 Thesis Organisation

This thesis is organized into different sections as follows:

1. **Chapter 1:** This chapter introduces the different concepts of Content Centric Networking paradigm used in the work along with its key features. Further, convolutional neural network concepts are introduced, which form the basis of this work. This is followed by

a brief description of the research motivation and the identified research gaps. Finally, the chapter is closed with a summary of contributions of this thesis.

2. **Chapter 2:** This chapter contains the literature relating to the field of convolutional neural network and content centric networking.

3. **Chapter 3:** This chapter formally states the problem that is addressed in this thesis and the objective function for the work.

4. **Chapter 4:** This chapter describes the complete details of the algorithms and procedures used to solve the problem stated in Chapter 3. It also contains all the formulas and equations as well as diagrams to explain the solution for solving the problem.

5. **Chapter 5:** This chapter consists of the simulation parameters and conditions used for implementing the algorithms. Futher, the performance parameters used and the results obtained from the proposed scheme are also included in this chapter.

6. **Chapter 6:** Finally, this chapter provides a brief conclusion to the work done in the thesis. Some limitations of the work done are also discussed and future research scope is provided in this chapter.

Chapter 2

Literature Review

The information centric networking concept was introduced in order to provide address-content decoupling to ensure a seamless content delivery to end users. The increasing amount of content being generated on the Internet made the content distribution a difficult task for traditional point-to-point networking architecture. As internet users are only concerned about content and not the physical location where it is stored, information centric networking was proposed. However, it suffers from several problems including interest packet broadcasting and network congestion. Several works exist in literature proposing several schemes to make information centric data dissemination more efficient.

Jian *et al.* [17] proposed a broadcast-based mechanism for data dissemination in a wireless network taking into account the mobility of the nodes as well as the available bandwidth for each node. Each node maintains a distance table (DT) containing information about the available content providers and the number of hops required for reaching it. A content request is broadcasted only if the corresponding entry is not found in DT. This scheme works by selected less mobile nodes as content deliverers and selects the nodes with higher available bandwidth as data forwarders, thus improving the data delivery rate while minimizing the network overhead. However, it is dependent on cellular network and may prove to be inefficient in vehicular environments due to highly mobile nodes resulting in frequent hand offs and link failures.

Kim *et al.* [18] proposed a topology aware data dissemination scheme. It performs packet flooding based on Multi Point Relay (MPR)protocol. Further, it uses the hop count metric in order to control the range of flooding and for disseminating the type of content stored, each node uses a Publisher MPR (PMPR) procedure. Overall, it achieves a minimum network overhead and mitigates the content broadcasting storming issue. However, the periodic updating of topology information tables incurs network overhead. Moreover, the velocity of nodes is considered to be low (1-4 m/s) which cannot adapt to vehicular environments.

Chan *et al.* [19] proposed a packet diffusion limited protocol (PDLP) reducing broadcast storming problem and network latency by exponentially reducing the hop count which eliminates unnecessary packet flooding by constraining transmission range. It also gives priority to nodes on the shortest path by allowing these nodes to choose a Deferred Time value from a smaller range of numbers ($0 \sim W_s$) as compared to other nodes' ($0 \sim W_n$).

Jamie *et al.* [20] proposed a novel energy efficient content distribution scheme in information centric networking scenario, dealing with the problem of multi casting and distributed caching. The problem is simulated as an information-centric linear programming model to minimize energy and storage costs at each node. It presents a general paradigm, independent of ICN implementation. However, there are some loopholes as the scheme is based on the assumptions that the network is private and that latency does not change with load, both of which are unreasonable in vehicular scenario. In addition, it does not take into account the mobility of devices.

Maryam *et al.* [21] introduced a novel routing scheme in order to mitigate the issues of packet flooding and link failures in vehicular content centric networks. It also takes into consideration the mobility of vehicles. In this scheme, an interest packet, including the id of best forwarder vehicle, is forwarded. The best forwarder vehicle is selected based upon vehicle trajectory and velocity. The best forwarder vehicle then checks its CS and in case the requested content is not found, the vehicle further re-broadcasts it within its transmission range. One major shortcoming in this scheme is that it requires a trade-off between energy utilization/load and number of collisions.

Ahmed *et al.* [22] put forward a routing scheme CODIE in which the routes of data as well as interest packets is computed in a VANET scenario. This scheme aims to reduce packet flooding by forwarding lesser number of content copies while retaining Interest Satisfaction Rate (ISR) and reducing Interest Satisfaction Delay (ISD) metric. CODIE successfully minimizes the number of data packets in the network in vehicular named data networking environment and achieves improved network performance by minimizing traffic and latency.

Kim *et al.* [23] introduced the CONET scheme in which the aim is to reduce the number of content/data packets in the network while preserving an acceptable value of Interest Satisfaction Rate (ISR), while at the same time reducing the Interest Satisfaction Delay (ISD) metric. In CONET, data requester sends an interest packet containing the TTL value and hop count of the respective interest packet such that the packet travels back along the reverse path to the consumer. The main aim of CONET is to reduce the number of network packets while maintaining a stable value of ISR.

Jeong *et al.* [24] put forth a trajectory-based forwarding scheme in a vehicular network scenario for data dissemination in a V2I scenario. In the proposed mechanism, multi-hop packet forwarding is achieved keeping in view the Access Point (AP) which has been selected. An algorithm is invoked for optimizing the AP selection in such that the delay in packet delivery is minimized. After access point is selected, consumer routing mechanism is chosen by the forwarding scheme. From the different trajectories available in the network, shortest path is selected to minimize delay.

Hommel *et al.* [25] introduced a forwarding scheme in which neighbouring node of requester sends the interest packet to the node which is at least distance from the data provider or to a node which is at greatest distance from requester. It is a position-based and greedy approach in which all routing decisions are based on distance from the data provider. After receiving an interest packet, each forwarder vehicle waits for a back-off time before re-broadcasting it which diverges proportionally as the distance between provider-consumer varies. This mechanism is successful in achieving low network latency.

Aamir *et al.* [26] introduced a novel scheme for sending interest packets into the network. In this scheme, the node requesting content sends an interest packet to the CCN router in the network. In these routers, a number of queues are generated simultaneously, in the same sequence according to which the router receives the interest packets. A pointer for checking the priority of the interest packets is deployed at the end of the queue. The priorities of received packets are checked against the priorities announced by the content publishers through control messages. On receiving an interest packet, the corresponding entry is recorded in pending interest table and is forwarded to the data producer based on its priority. As a result, higher priority content suffers reduced latency in comparison to the content with lower priority.

Vaidya *et al.* [27] put forward a forwarding scheme for named data networking environment. It is an adaptive mechanism and works by using the state of NDN datagram which is the primary unit of a packet switched network. The overall aim of the proposed mechanism is to successfully obtain data packets, while detecting content loss issues simultaneously as well as recovering network problems at once. At the end, some open forwarding research strategies have also been mentioned in some detail. However, transmission errors and congestion problems are not taken into consideration.

Chand *et al.* [28] proposed content-based routing scheme for data dissemination in ad hoc vehicular network environment. It works by building a network tree reflecting the topology of underlying network. Then the content can be routed based upon type of content rather than destination address. However, this technique is fundamentally based on Internet Protocol (IP) and the difficulty in maintaining the network tree in highly mobile vehicular networks limits its application to more static environments.

Grassi *et al.* [29] proposed a prototype for vehicular named data networking (VNDN) scenario in which, based on the priorities of receiver neighbouring nodes, interests are sent in the network. Upon the reception of the interest packet, a defer time is set by the receiver node which varies directly with the distance of the receiver from the sender. proportional to its distance from the sender. The main purpose of this is to achieve interest forwarding prioritization among all the neighbouring nodes.

Grassi *et al.* [30], in another work, put forward a novel routing protocol Navigo in which the interest packets are forwarded by utilizing information about the geographical location of vehicles. It makes use of the MGRS to identify every region of the world with the help of a label. Therefore, network data is named after being mapped to specific geographical areas where it is produced. This scheme has been developed, keeping in view the VANET applications that are dependent on location. These include information about the condition of traffic on roads and information regarding parking space availability. For the purpose of interest forwarding,the idea of geo-faces mapped to geographical locations has been put forward by Navigo. Overall, interest is forwarded on the shortest path, towards the content-producing area.

Hong *et al.* [31] introduced the DADT mechanism for forwarding interest packets in the network. In this scheme, the interest forwarding is carried out according to the node priority of the receiving node. DADT also makes use of defer time to determine the priority of interest packets before transmitting them into the network. The defer time is calculated at each node on the basis of its distance from the data producer and data requester. At each node, defer time is calculated based on its distance to the sender and its proximity to the data producer. There are two drawbacks of using this scheme. Firstly, it requires periodic sending of hello messages to neighbouring nodes for facilitate neighborhood discovery. Secondly, it assumes that each node in the network is aware of the content producer's location, which leads to producer-consumer coupling.

Ahmed *et al.* [32] introduced an interest forwarding scheme for vehicular content centric networks in order to reduce the interest broadcasting storming in CCN environments. This mechanism requires each node to share and maintain two kinds of lists: a Recent Satisfied List (RSL) which contains information about all the data requests satisfied by the particular node; and a Neighbour Satisfied List (NSL) which contains information regarding all the data requests which have been satisfied by the neighbouring nodes of a specific node. These lists serve as routing tables and serve to determine the next hop of interest packet for optimal performance.

Yan *et al.* [33] proposed the application of named data networking in the vehicular scenario to facilitate the deployment of Vehicular Information Networks (VINs). In the proposed scheme, each network content is named hierarchically and the interest packets are forwarded in the network by using the location information of the network vehicles. It also supports mobility management. In this work, several issues have been taken up including the high operational cost of the network and less efficiency in communication due to increased mobility. This scheme is successful in achieving optimal performance in case of large network areas and growing network size; however, the authors did not explicitly mention the strategies that must be used for forwarding interest and data packets.

Wang *et al.* [34] proposed the application of the concept of named data networking to facilitate vehicle to vehicle communication with minimum collisions. An application called Rapid Traffic Information Dissemination has been introduced by the authors in order to enable data propagation in V2V NDN networks successfully. The primary aim of the author in this work is to distribute content in the network in a wider coverage area with minimum network overhead. At the same time, the issue of interest packet broadcast storming problem has also been taken into account.

Li *et al.* [35] introduced a novel forwarding strategy, GOFP, which dependent on data names tagged with geographical location of the content publishers in a vehicular network. In this protocol, an opportunistic strategy has been used for interest forwarding. The next best forwarder is selected on the basis of two parameters, viz., vehicle trajectory and position of interest (POI). The data requester send out an interest packet in the network containing its current geographical location, the trajectory information including speed and mobility of relay nodes and a TTL value. This protocol successfully eliminates the transmission latency as well as delays in delivery of packets in the whole vehicular content centric network.

Zhao *et al.* [36] introduced an interest forwarding scheme based on the geographical location of nodes over the underlying named data network architecture. In this scheme, data is named using the conventional naming scheme used in the network appended with the information of its geographical location of the data requester. Multiple forwarding is used in this mechanism in order to make the network more efficient and reliable. This scheme also targets the neighbouring vehicles at the intersection with a view to overcome hurdles at road junctions.

Huang *et al.* [37] introduced a novel traffic aware routing protocol named TAG. The main aim of this protocol is to achieve effective resource utilization by minimizing the number of packets transmitted in the network. There are two basic elements in TAG framework namely, Region Needed-to-be Covered (RNC), which contains geographical information collected from the nearby road information as well as from real time road maps. Another element is the broadcasting mechanism which is used to distribute interest and data packets to network vehicles in an efficient manner.

Junaid *et al.* [38] introduced a socially aware scheme 'SAVING' to make optimal decisions regarding content placement in caches of vehicles to minimize resource utilization. A vehicle ranking system is used to identify CarRank and GRank for each node, which are then used to determine optimal locations for caching content based upon content popularity and other social layer information including content importance, vehicle spatio-temporal availability, and connectivity. A content distribution protocol is also introduced which uses FACILITATOR(), PROVIDER(), PUBLISH() and CONTENT() functions to route content optimally. However, in spite of all the advantages, the scheme does not address cache redundancy issue which may lead to resource wastage.

Lequerica *et al.* [39] introduced the DaS application, enabling the vehicular users to share information in real-time. The information can be gathered autonomously by the vehicles, and distributed into the vehicular network, which may include both personal, geographical and traffic information. This information is then utilized by vehicles to determine shortest routes, avoid congested routes and beware of slippery roads. However, the DaS application is dependent on the cellular network for content dissemination in the vehicular network. As cellular networks have limited bandwidth and coverage, frequent handovers and increased latency may affect the performance of the application.

Stephen *et al.* [40] introduced an application called 'RoadSpeak' for making the communications in vehicular networks more efficient by utilising social layer information. Using

RoadSpeak, different vehicular users can connect with each other using voice chat and may create groups according to their geographical locations and social interests. The users of a specific voice chat group may communicate with each other on the road through voice messages. It requires a centralized server deployed in the network to maintain the information of users and an Internet infrastructure to enable users to connect to the server either directly through cellular network or by using WiFi hotspots.

Azza *et al.* [41] proposed an Android application 'RoadSense' which is a vehicular social application that independently makes the information of the road and traffic conditions available to vehicular users in real time. It is based on tri-axial accelerometer and gyroscope, and utilizes Global Positioning System (GPS) to generate road trace on a geographical map. The social information on the road is used to train a Decision Tree, which is then used to predict the road conditions and shared among the vehicular users.

Victor *et al.* [42] proposed an application 'Social Drive', which is an Android application, mainly for vehicular users. It enables the users of vehicles to learn their respective driving patterns as well as share the information about their trip in real-time by making use of social networking applications including Facebook and Twitter. This applications aims to help the vehicular drivers in improving their driving habits by reducing the usage of fuel and hence move a step further towards green transportation.

Luan *et al.* [13] proposed an application 'Verse' takes into consideration completely distributed and short-lived vehicle-to-vehicle connections in a vehicular network scenario and aims to put forward a general architecture for facilitating instantaneous communications among vehicular users in a cheap and reliable manner by utilising social layer information. It uses the underlying vehicular networks existing on highways. However, this application works well only in highway scenarios and may not work effectively due to significant changes in vehicular mobilities.

Zhou *et al.* [14] put forward a content dissemination scheme for internet of vehicles based on social layer data, which utilizes social layer information of vehicular users to facilitate content distribution in a D2D-V2V based communication scenario. The physical layer information is used to calculate connection probability of vehicles and the social layer information is used to compute social relationship tightness. The social tightness represents the content selection similarity between vehicles and is calculated using Bayesian non-parametric learning (BNL) model. However, BNL does not impose an upper bound on the number of clusters, which may result in exponential increase in complexity.

Table 2.1: Summary of Literature Survey

S. No.	Ref. no.	P1	P2	P3	P4	P5	P6	P7
1	**[17]**	✓	10 - 15	802.11g	50	×	×	✓
2	**[18]**	✓	1 - 4	802.11g	32	×	×	✓
3	**[19]**	×	0	802.11a	20	×	×	✓
4	**[20]**	×	0	802.11a	79	✓	×	✓
5	**[21]**	✓	Random	802.11p	20 - 45	×	×	✓
6	**[22]**	✓	55 - 85	802.11p	50 - 120	×	×	✓
7	**[23]**	✓	55 - 85	802.11p	50 - 120	×	×	✓
8	**[24]**	✓	20 - 60	802.11p	250	×	×	✓
9	**[25]**	✓	36	802.11ac	200	×	×	✓
10	**[26]**	×	0	802.11	4	×	×	✓
11	**[27]**	✓	2.5	802.11p	30	×	×	✓
12	**[28]**	✓	6.3 - 21.2	802.11a	10	×	×	✓
13	**[29]**	✓	Random	802.11	812	×	×	✓
14	**[30]**	✓	Random	802.11p	30	×	×	✓
15	**[31]**	✓	55	802.11p	20 - 50	×	×	✓
16	**[32]**	✓	10 - 60	802.11p	Random	×	×	✓
17	**[33]**	✓	60	802.11a	200	×	×	✓
18	**[34]**	✓	2.2 - 13.9	802.11p	40 - 120	×	×	✓
19	**[35]**	✓	0 - 40	802.11p	200 - 600	×	×	✓
20	**[36]**	✓	Random	802.11	100 - 2000	×	×	✓
21	**[37]**	✓	2 - 6	802.11	100	×	✓	✓
22	**[38]**	✓	Random	802.11p	2986	×	✓	✓
23	**[39]**	×	0	802.11	3	×	✓	×
24	**[40]**	✓	Random	802.11p	2	×	✓	×
25	**[41]**	✓	Random	802.11n	2	✓	✓	×
26	**[13]**	✓	70 - 130	802.11b	1000	×	✓	×
27	**[14]**	✓	$\leq$50	802.11p	Random	×	✓	×

- P1: Mobility; P2: Speed (m/s); P3: Communication Paradigm; P4: Network Size; P5: Energy Estimation;P6: Social Information Analysis; P7: CCN architecture

Chapter 3

Problem Statement

3.1 Objectives

1. To select devices for data dissemination based on energy values.
2. To determine Connection Probabilities of device-pairs, simulated as a Weiner Process, to identify stable and long-lasting connections.
3. To implement community detection by finding social relationship using convolutional neural network (CNN) based learning approach. This is done to identify vehicle-pairs for optimal performance.

3.2 Problem Statement

Figure 3.1 shows the IoV scenario consisting of vehicles and the possible V2V links between vehicle-pairs. Each link has an associated connection probability indicating the duration t for which the end vehicles of the link will remain connected. Also, data dissemination in the network will incur a cost for each vehicle, with respect to energy dissipation. Each vehicle in the network has an associated remaining energy e, which indicates the time for which the vehicle will be active. Furthermore, each vehicular user has different social interests which are stored in terms of demographic, temporal and social attributes in each vehicle.

However, in CCN scenario, available network bandwidth is consumed by broadcasting and re-broadcasting of data and interest packets which may potentially lead to congestion in the network. This, in turn, incurs cost on the network in terms of delay and hence, degrades the overall network performance. Hence, our aim is to provide solution architecture for data dissemination which takes into account the various parameters of vehicles like relative velocities, headway distance, etc. to form disjoint V2V pairs which can then exchange data with

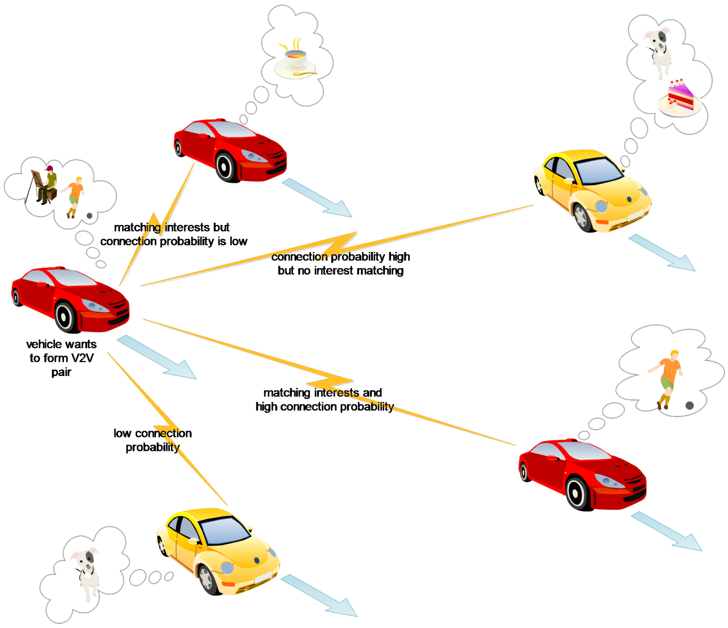

Figure 3.1: Problem Statement

minimal delay. In view of this, the objective function for our work is defined by the following equation,

$$\xi_i = max\{U_i(N) - |N|[\sum_{j=1}^{n} H(t)/|(u_i - u_j)| - e_{ij}]$$
$$|H(0) = h_o; N \geq 1\}$$

where $U_i(N)$ denotes the utility gained by the vehicle by sending *N* number of packets through the V2V link obtained as a solution from our architecture, *H(t)* refers to the headway distance between V_i and V_j at time *t*, (u_i-u_j)*H(t)* gives the latency for each packet in terms of distance between the vehicles, and e_{ij} denotes the energy dissipation of V_i to exchange *N* data packets with V_j. The aim of each vehicle is to maximize the value of ξ_i.

In order to achieve this, energy analysis of V_i is done to obtain an energy value ψ_{v_i} indicating the amount of energy that would remain with V_i, if it participates in the data dissemination process for which it requested. In case ψ_{v_i} is found to be greater than a threshold value E_o, connection probability $F_{ij}(t)$ of V_i is computed with each vehicle in the network which can potentially participate in data exchange, which gives the probability that V2V pair will remain connected for time *t*. This is done in order to make sure only stable and long-lasting connections are formed. Based upon these computations, those V2V pairs are screened for which the value of $F_{ij}(t)$ is greater than the threshold p_o. These vehicles then undergo social relationship estimation analysis wherein the interests of respective vehicular users are fed into a CNN to detect disjoint V2V pairs which must communicate for optimal performance.

3.3 Constraints and Assumptions

1. The learning rate of the vehicles is normalized and lies between 0 and 1 ($0 \leq \alpha \leq 1$).
2. The cumulative distribution function (CDF) of connection probability lies between 0 and 1 ($F_{ij} \in [0, 1]$).
3. The energy required for data dissemination from vehicle *i* to *j* is always greater than the minimum threshold energy value E_o ($e_{ij} > E_o$).
4. The probability that remaining energy is greater than threshold value lies between 0 and 1 ($\phi_{v_i} \in [0, 1]$).
5. The social score of V2V pairs is normalized and lies between 0 and 1 ($S \in [0, 1]$).
6. The initial headway distance between two vehicles is h_o ($H(0) = h_o$).
7. The minimum number of data packets exchanged between two vehicles should be at least 1 ($N \geq 1$).

Chapter 4

The Proposed Scheme

This chapter provides the details of approach used in solving the problem stated in previous chapter. The solution can be divided into three main sections, first is Energy Estimation, which gives details about the algorithm used to screen the vehicles based upon the energy values in order to ensure than only those vehicles participate in V2V communication whose energy value is greater than a minimum threshold value, the second section is Connection Probability Estimation, which gives details about the algorithm used to find vehicle pairs ideal for establishing a connection to ensure only stable and long-lasting connections are formed, and the third section is Social Relationship Estimation which gives details of algorithm used to determine the social score information of vehicles to facilitate interest matching and data exchange.

4.1 Energy Estimation

Each vehicle in the network has limited energy, which is dissipated as data packets are exchanged with other vehicles. The following function gives the total energy $\varphi_{V_{ij}}$ required to exchange a data packet between vehicles V_i and V_j:

$$\varphi_{V_{ij}} = max\{|N| \sum_{i=1}^{n} \sum_{j=1}^{n} e_{v_{ij}} \forall v \in \{v_1, v_2, ..., v_n\}; i \neq j\} \tag{4.1}$$

Here, N is the total number of data packets exchanged between V_i and V_j. Whether V_i can form a V2V link for data exchange with V_j is determined by the remaining energy (E_i) of V_i with a view that whole of V_i's energy should not be exhausted. For this purpose, a threshold value E_o is set, which indicates the minimum level of remaining energy that a vehicle must have in order to participate in data exchange. The remaining energy of V_i denoted as ψ_{v_i} is computed as:

$$\psi_{v_i} = E_{V_i} - \varphi_{V_{ij}} \tag{4.2}$$

Based upon the value of equation 4.2, an energy indicator function ϕ_{v_i} is defined, which indicates whether a vehicle V_i can participate in data exchange or not.

$$\phi_{v_i} = \begin{cases} 1 & \text{if } \psi_{v_i} \geq E_o \\ 0 & \text{if } \psi_{v_i} < E_o \end{cases}$$

Each vehicle in the network computes an Energy Indicator Function ϕ_{v_i} according to Algorithm 4.1. Firstly,for each vehicle V_j in the network, V_i calculates the energy required to exchange *N* data packets with V_j. Once done, it then computes the value of energy that would remain with it, if the data exchange occurs. If this remaining energy value is greater than a certain specified threshold, only then V_i would proceed further to estimate its connection probability with other vehicles. Otherwise, V_i is denied for data exchange.

Algorithm 4.1 Pseudo code for Energy Analysis of Vehicles

procedure $\phi_{v_i}(\psi_{v_i}, E_o)$ ▷ Energy Indicator Function ϕ_{v_i}
 for $i \in \{1, ..., n\}$ **do**
  for $j \in \{i+1, ..., n\}$ **do**
   $E_{V_{ij}}$=$|N|.e_{v_{ij}}$
  end for
  $\varphi_{V_{ij}}$ = max$\{E_{V_{ij}}$ $\forall$ V $\in \{V_1, V_2, ..., V_n\}\}$
 end for
 for $i \in \{1, ..., n\}$ **do**
  for $j \in \{1, ..., n\}$ **do**
   ψ_{V_i} = E_{V_i} - $\varphi_{v_{ij}}$
   if $\psi_{v_i} \geq E_o$ **then**
    ϕ_{v_i} = 1
    Call $F_{ij}(t, u_i, v_i, \sigma_i^2, \sigma_j^2)$
   else
    ϕ_{v_i} = 0
   end if
  end for
 end for
end procedure

4.2 Connection Probability Estimation

For two vehicles, V_i and V_j to form a V2V pair (provided $\phi_{v_i} = 1, \phi_{vj} = 1$), the probability of a stable and long lasting connection, for duration *t*, must be greater than a certain threshold value. To ensure this, the method proposed in [14] has been adopted for computing connection probabilities between V_i and V_j and then define a function θ_{ij} to determine if V2V pair V_iV_j should be formed or not for optimal performance.

Let the mean and variance of velocities of vehicles V_i and V_j is u_i, σ_i^2 and u_j, σ_j^2 respectively. Let *H(t)* denote the headway distance between V_i and V_j after time *t*. The initial headway distance *H* is set to h_o. Further, let the communication range between V_i and V_j is *L*.

$$H(t) = \begin{cases} V_i\, is\, ahead\, of\, V_j & \text{if } H(t) > 0 \\ V_j\, is\, ahead\, of\, V_i & \text{if } H(t) < 0 \end{cases}$$

To compute the connection time, the mean first passage time τ is evaluated as a random variable, depending upon the velocity differences and h_o.

$$\tau = \{min\, t | H(0) = h_o, -L < H(\tau) < L, 0 \leq \tau \leq t\} \tag{4.3}$$

To evaluate τ , *H(t)* is modeled as a Wiener process. The drift μ and variance σ^2 is denoted by the following equations:

$$\mu = u_i - u_j \tag{4.4}$$

$$\sigma^2 = \sigma_i^2 - \sigma_j^2 \tag{4.5}$$

Within an infinitesimally small time interval Δt , the increment of *H(t)* follows a normal distribution, as follows:

$$\Delta H(t) = H(t + \Delta t) - H(t) = \mu \Delta t + \sigma W \tag{4.6}$$

Here, *W* varies in accordance with a normal distribution with a mean value of zero and the variance of Δ t, i.e., *W* $\sim$ *N(0,Δ t)*. Further, the time variation of probability density function (PDF) of a vehicle's velocity in Weiner Process can be described by Kolmogorov equation,

$$\frac{\partial p(\tau|h_o,t)}{\partial t} = -\mu\frac{\partial p(\tau|h_o,t)}{\partial \tau} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial \tau^2}p(\tau|h_o,t) \tag{4.7}$$

Here, *-L* $\leq \tau \leq$ *L*, and *p(*$\tau|$ h_o*,t)* is the PDF of *H(t)* given *H(0)=*h_o. Define $\delta(.)$ as the Dirac delta function, the initial and boundary conditions are given by

$$p(\tau|h_0,0) = \delta(h_o) \tag{4.8}$$

$$p(-L|h_o,t) = p(L|h_o,t) = 0, t > 0 \tag{4.9}$$

Combining 4.7 $\sim$ 4.9, we can obtain p($\tau|h_o$,t) as

$$\begin{aligned} p(\tau|h_o,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \sum_{y=-\infty}^{\infty} [exp\frac{4y\mu L}{\sigma^2} - \frac{[(\tau-h_o)-4yL-\mu t]^2}{2\sigma^2 t}] \\ -exp\frac{2\mu L(1-2y)}{\sigma^2} - \frac{[(\tau-h_o)-2L(1-2y)-\mu t]^2}{2\sigma^2 t} \end{aligned}$$

The CDF (Cumulative Distribution Function) of the connection time can be derived based on equation 4.3 as,

$$F_{ij}(t) = Pr(T \leq t) = 1 - \int_{-L}^{L} p(\tau|h_o,t)d\tau \tag{4.10}$$

Equation 4.10 gives the probability that vehicles V_i and V_j are connected during time duration *t*. Based on this probability, a connection function $\theta_{v_{ij}}$ is defined as,

$$\theta_{v_{ij}}(t) = \begin{cases} 1 & \text{if } F_{ij}(t) \geq p_o \\ 0 & \text{if } F_{ij}(t) < p_o \end{cases}$$

where p_o is the lower bound on probability required to establish a stable connection.

Given the communication range *L*, V_i computes its connection probability with every other vehicle V_j based upon the information of velocities u_i and u_j, and velocity variances σ_i^2 and σ_j^2. The probability at time *t*, given the initial headway distance d_o is computed by line 5 of the Algorithm 4.2 by recursively calling the function *H(t)*. Finally, the CDF, giving the value of connection probability, is computed by equation 4.10 which is then compared with the threshold probability.

Algorithm 4.2 Pseudocode for Connection Probability Estimation

procedure F_{ij}(t,u_i,v_i,σ_i^2, σ_j^2,L) ▷ Connection Probability Estimation Function F_{ij}
    μ = u_i - u_j ▷ Drift of velocities u_i and u_j
    σ^2 = σ_i^2 + σ_j^2 ▷ Variance of velocities
    H(0) = d_o ▷ The initial headway distance
    p($x|d_o$,t) = P(x $\leq$ H(t) $\leq$ x+Δx)
    Function H(t):
    { W = Random.nextGaussian()
    H(t)= $\mu\Delta$t + Wσ }
    return 1- $\int_{-L}^{L}$ p($x|d_o, t$)dx
    if F_{ij}(t,u_i,v_i,σ_i^2, σ_j^2,L) $\geq p_o$ **then**
        $\theta_{v_{ij}}(t) = 1$
        Call Procedure: Social Relationship Estimation
    else
        $\theta_{v_{ij}}(t) = 0$
    end if
end procedure

4.3 Social Relationship Estimation

To ensure optimal network performance, it is required that the vehicle V_j holds the data requested by V_i so that the request need not be broadcasted to other vehicles. In order to ensure this, an interest-matching procedure is invoked, which establishes a social relationship between V_i and V_j in terms of probability. Based on this probability value, V2V pairs are established, which can share data, if required, for optimal network performance.

In this thesis, we have undertaken a deep-learning based approach for social relationship estimation. We have used Convolutional Neural Network for this purpose. The view behind choosing CNN over other approaches is two-fold:

1. CNNs automatically learn the values of its filters based upon the training set, with respect to the task at hand.

2. CNNs are very fast in terms of producing output (with reasonable accuracy) and it is possible to implement an *n*-gram model for community detection using CNNs by applying a

large filter, in a relatively inexpensive manner, as compared to other models.

To compute social scores of vehicles, we have utilized the Convolutional neural network. We have utilized the word2vec to pick up the vectors for the words as the input matrix. Word2vec is a neural network that processes message before that content is taken care of by deep-learning procedures. In the proposed scheme, sentences are needed to be classified using convolutional neural network, but CNNs find it difficult to identify the meaning of sentences without errors. Therefore, to solve this issue, word2vec is used to decode and interpret the context in which a word is used in the sentence. Also, the vector resulting from word2vec translation represents the distance of words, i.e., the words having similar meaning have a close value of the corresponding vector. Word2vec takes a sentence as input and transformed into a vector and then compared to related vectors in an n-dimensional space of vectors. The words which are related to each other appear close together and thus their similarities can be computed with accuracy.

Convolution is performed on the input sentence matrix by each filter and produces feature maps of variable length. At that point 1-max pooling is performed over each feature map, i.e., the biggest number from each of the feature maps is identified. Hence it creates a uni-variate feature vector by combining all feature maps, and these features are combined to generate a feature vector for the final output layer. The last softmax layer at that point gets this element vector as input and utilizes it to classify the information.

The field values of our training data set are given as input to the input layer of CNN on which filters are applied. The filter slides over the entire output to produce an intermediate hidden layer. At this layer, we perform pooling operation to produce a set of non-overlapping partitions over the input. Finally, we construct a fully connected layer to concatenate the results of each hidden layer neuron to get the output.

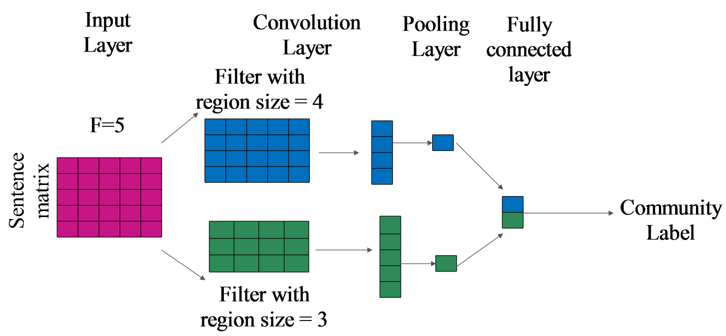

Figure 4.1: Convolutional Neural Network

4.3.1 CNN Hyperparameters

- Output Size: To obtain the results with high accuracy, the output size should be relative to the size of input layer. We have used wide convolution (zero-padding) which ensures that the filter is applied to every element of the input and thus produces a larger or equal-sized output relative to the input layer.
- Stride Size: This parameter determines the amount of shift at each step. Smaller stride size will result in overlapping of consecutive filters thus giving us high accuracy results whereas a larger stride size gives more coverage.
- Filter Size: The filter size is dependent on the value of *n* in the n-gram model that has been used for categorization of communities. Larger value of *n* will analyze the input text in a broader concept. CNNs have the potential to implement larger values of *n* in an inexpensive way.
- Pooling:In this work, max-pooling has been used as only the information regarding whether or not a word appeared in the text has to be maintained. A larger value indicates the appearance and a smaller value, otherwise. The size of padding is given as,

$$P = \frac{F - 1}{2} \tag{4.11}$$

  where *F* is the size of the filter.

The social relationship between V_i and V_j is computed based upon different temporal, demographic and social attributes of users according to Algorithm 4.3 **??**. The dataset is given as input to the convolution layer of CNN and several hyper-parameters including stride side, pool size, and number of filters are specified to form a network. The output layer of CNN gives the community label of the user based upon the field values.

For each unit Z_{ij} in the hidden layer, the output is obtained by summing up the contributions from previous layers as,

$$Z_{ij}^{l} = \sum_{u=0}^{m-1} \sum_{v=0}^{m-1} F_{uv} \cdot y_{(i+u)(j+b)}^{l-1} \tag{4.12}$$

where *m* is the size of the filter, *l* denotes the number of hidden layers and *F* is the filter applied.

After computing the non-liner value of Z_{ij} using equation 4.13, CNN applies *tanh* non-linearity over Z^l_{ij} as,

$$y^l_{ij} = tanh(Z^l_{ij} + B) \tag{4.13}$$

where, *tanh* is the non-linearity function and *B* is the bias term added to make the model robust.

At last, the probability of user belonging to that community is computed and combined with the probability of other vehicles in the network to generate the social score *S* of V2V pair as,

$$S = y^l_{ij} \times F_{ij}(t) \tag{4.14}$$

Two vehicles, whose combination gives the highest value of *S* can form a V2V pair for data exchange.

Algorithm 4.3 Pseudocode for Social Relationship Estimation

procedure SOCIAL RELATIONSHIP ESTIMATION
    data[] = url("dataset") ▷ Load dataset
    trainingSet = data[0:x]
    testingSet = data[x+1:n] ▷ n rows of dataset
    F = i ▷ Define filter size
    P = j ▷ Define pooling size
    S = k ▷ Define stride size
    Function tanh(x):
    { return (1 / (1 + exp(-x))) }
    Define ConvNNet(F,P,S,sigmoid(x))
    Label = ConvNNet(data[].load())
    P_i(Label|data) = P(Label).P(data|Label)
    if Pi(Label|data) $\geq P_o$ **then**
        for j = 1; j $\leq$ n ; j++ **do**
            $P_{ij} = P_i + P_j$
            return max(P_{ij})
        end for
        V_i and V_j can form V2V pair for data exchange
    end if
end procedure

4.3.2 Data Dissemination Scheme

The complete process of deep learning-based data dissemination in internet of vehicles takes place in the following four stages:

Stage 1: In the first stage, all electric vehicles are screened to eliminate those vehicles whose energy is less than the minimum threshold energy value E_o. This is done in order to ensure than the vehicle does not break down while the content distribution is in progress. In case of non-electric vehicles, this step is ignored.

Stage 2: In the next stage, V2V pairs are formed containing all vehicles selected in Stage 1 and the connection probability of all these pairs is computed. Those V2V pairs whose connection probability is above a minimum threshold value are selected for the next stage. This is done in order to ensure that only stable and long lasting connections are formed.

Stage 3: In the third stage, interest matching procedure is invoked for all V2V pairs selected in Stage 2. A convolutional neural network is used for this purpose. This step determines the social score S by combining interest matching score with connection probability, which indicates the probability of two vehicles storing the same content. Intuitively, a high value of S_{ij} maximizes the probability that the content requested by V_i will be found at V_j.

Stage 4 After finding ideal V2V pairs, vehicle V_i sends an interest packet containing the $content\_id$ of data to be retrieved to V_j. If V_j holds the requested content (the probability of this is maximized by a high value of S), it returns the corresponding data packet back to V_i. In this way, interest packet broadcasting problem is eliminated and the network delay and latency is reduced.

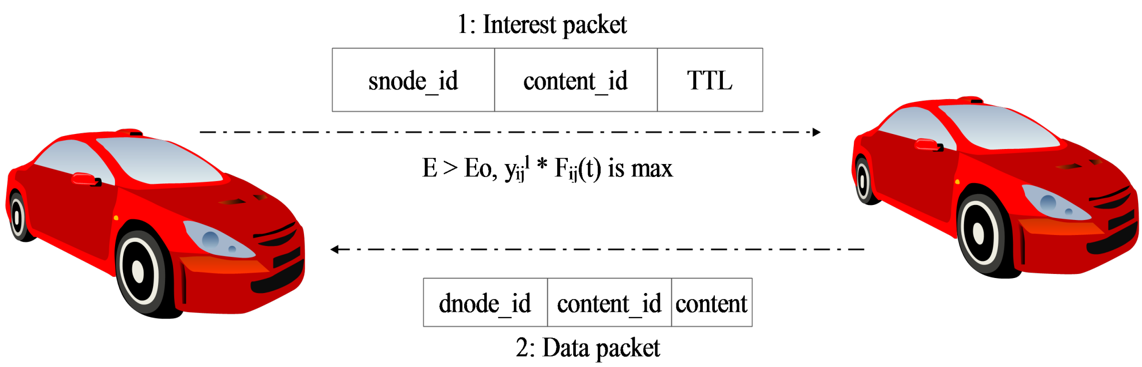

Figure 4.2: Complete content dissemination scheme

Chapter 5

Performance Evaluation

This chapter presents the performance evaluation of the proposed algorithm for data dissemination in content centric networking scenario. The proposed work is evaluated through simulation based implementation. This section presents details of simulation environment followed by results of implementation of the proposed algorithms.

5.1 Simulation Setup

The proposed scheme is evaluated using extensive simulations on a highway topology scenario with one thousand vehicles. The mean(μ) of the vehicle‘s velocity is assumed to be distributed uniformly in a range of (60-140) km/h. Similarly, for each vehicle, the standard deviation(σ^2) is also distributed uniformly in a range of 20-40 km/h. Initial placement of vehicles on the road is based on Poisson distribution with a headway distance between two consecutive vehicles following an exponential distribution (μ= 50 *m*). The physical-layer transmission rate follows a typical VANET communication standard, i.e., 6 Mb/s for each vehicle. The vehicles communication range (300 *m*) is scheduled for channel access on the basis of IEEE 802.11b [14].

5.2 Performance Metrics

The following metrics have been used in this thesis to evaluate the usefulness of this work and its impact on overall performance of the proposed algorithms.

1. **Latency:** It refers to the amount of round trip time(in ms) that a network packet takes to travel over the network. The network latency is calculated as the time lapsed between the sending of interest packet by the data requester and receiving of data packet sent by the receiver. Latency is inversely proportional to network performance and hence, it must be minimized.

2. **Delay:** It refers to the amount of time that a network packet takes for one-way transmission from one network end-point to another. The network delay is calculated as the amount of time lapsed between the sending of interest packet by data requester and corresponding receiving of the interest packet by the data provider. Delay is inversely proportional to network performance and hence, it must be minimized.

3. **Cache hit ratio:** It refers to the ratio of the number of times a received content request is satisfied to the number of content requests received by a particular node. It is expressed in the form of percentage. Cache hit ratio is directly proportional to network performance and hence, it must be maximized.

4. **Network traffic:** It refers to the amount of data packets in transit within the network at a given point of time. It includes all interest and data packets sent out in the network but not yet received. A high network traffic may lead to collisions and packet losses. It is inversely proportional to network performance and hence, it must be minimized.

5.3 Results

According to the proposed scheme, the vehicles that are having sufficient amount of energy available with them to participate in the data dissemination are selected. To validate this fact, the impact of energy available with a vehicle and the content disseminated is analyzed. Figure 5.1 shows the variation of content dissemination with respect to an increase in the amount of energy available with the vehicles. Now, once the vehicles are screened for the available energy, then their connection probability is estimated. So, in order to validate this scheme, two vehicles *i* and *j* are considered in the CCN-based network with a mean and variance of their velocities as μ_i = 110 km/h, σ^2= 30 and μ_j= 950 km/h, σ^2= 21.

Now, Figure 5.2 shows the CDF of link connection duration (considered as a function of the initial distance between two vehicles *h(i,j)*). In this analysis, as $\mu_j > \mu_i$, so the mean value of the link connection time decreases with an increase in the distance between *i* and *j*. So, the vehicle *i* is bound to move out of the vehicle *j*s communication range. Figure 5.2 elaborates on this fact where it is clearly shown that at a given time as the distance h_o increases from 0 to 200 m, the probability of disconnection increases as the curve shifts towards left side.

Now, to analyze the impact of change in the value of μ and σ^2 of velocities of vehicles,

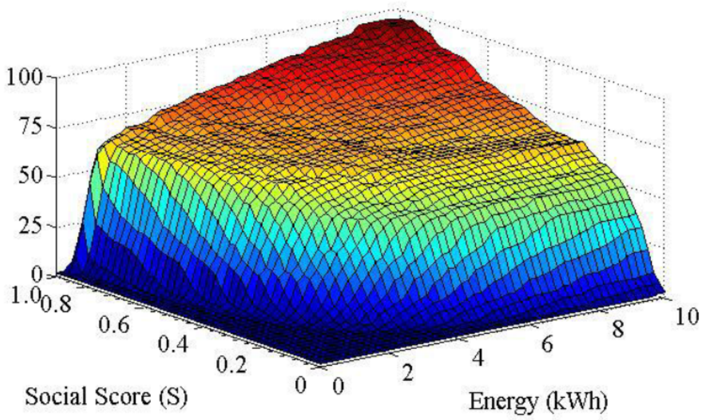

Figure 5.1: Content disseminated with respect to social score and energy

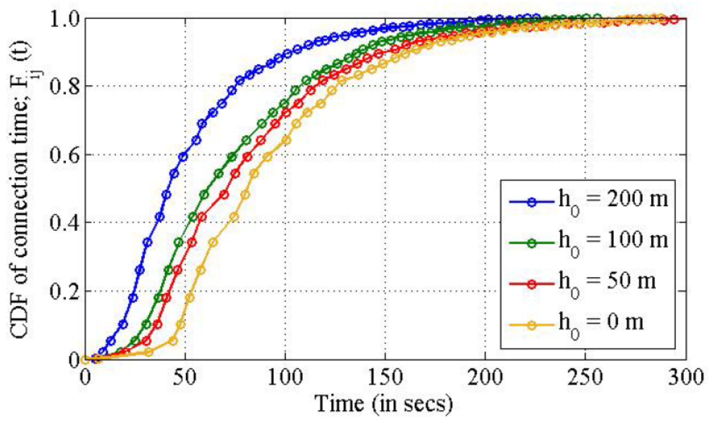

Figure 5.2: CDF of connection time vs distance

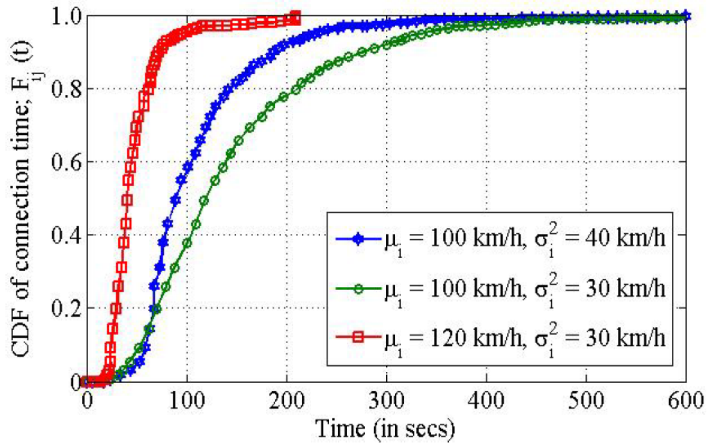

Figure 5.3: CDF of connection time vs μ_i and σ^2 of velocities

Figure 5.3 shows that the curve for $F_{ij}(t)$ shifts towards left side. This indicates that the probability of disconnection is enhanced with an increase in the gap between velocities of i and j. Moreover, with an increase in the value of σ^2, the curve for CDF expands in width.

In this direction, the impact of variation in the number of vehicles, and their corresponding velocities on the content dissemination is shown in Figure 5.4 . The results clearly show that the content disseminated exhibits an increase with respect to increase in the number of connected vehicles. However, with an increase in the velocity of vehicles, the possibility of vehicles moving out of the communication range increases, thereby increasing the changes of disconnection. Hence, the content disseminated decreases with an increase in the velocity of vehicles.

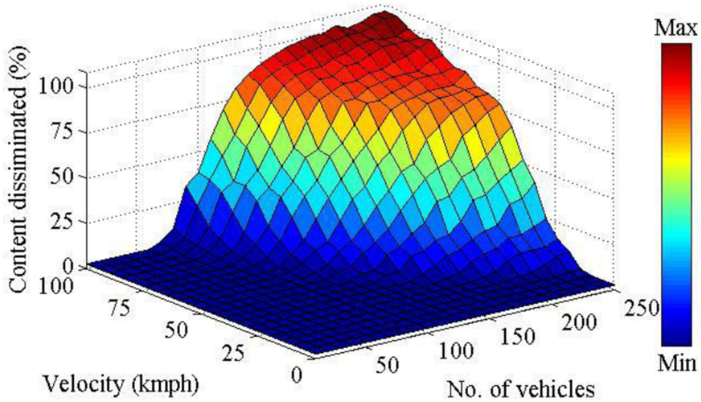

Figure 5.4: Content disseminated vs velocity and number of vehicles

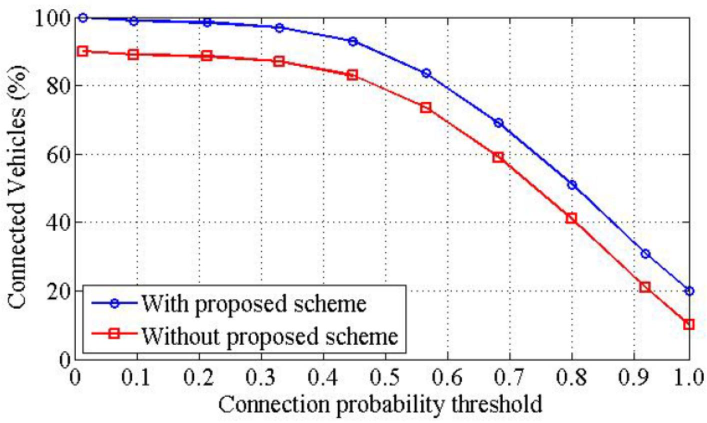

Figure 5.5: Number of connected vehicles vs connection probability threshold

The connection probability threshold also affects the number of connected vehicles. Figure 5.5 shows that the percentage of connected vehicles increases with a decrease in the connection probability threshold. It depicts that the lower the value of threshold, the more vehicles can participate in the CCN. Finally, the impact of social score on the content dissemination is evaluated. Figure 5.1 shows that with an increase in the social score, the percentage of content disseminated also increases. This is due to the reason that the more socially a vehicle is connected, the more actively it participates in the data dissemination scheme.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In this thesis, a content-centric data dissemination approach is presented for IoV. Initially, the vehicles are evaluated based on available energy with them to select the vehicles having sufficient amount of energy to provide a stable and reliable connection. After this, the connection probability of the selected vehicles is estimated using Weiner process model. Finally, the social relationship score of vehicles is estimated using CNN. After extensive simulation analysis, the results obtained show that the energy, connection probability, and social relationship score have a strong impact on the data dissemination. The impact of variation in mean and variance of velocities of vehicles, number of vehicles, social score, and energy were analyzed. The results depict that the data disseminated increases with an increase in social score, energy level, and number of connected vehicles. Thus, the proposed scheme successfully solves the interest broadcast storming problem by utilising the social layer information of vehicular users to make the routing decisions. It maximizes the probability of encountering a cache hit, while minimizing the latency and network delay. The operational cost of a content-centric network is also reduced to a large extent due to considerable reduction in the number of control messages in the proposed scheme.

6.2 Future Scope

The results depict that the disconnection probability increases with an increase in the velocities of the vehicles. This aspect is yet to be explored and improved upon. The performance of data dissemination in terms of delay and latency depends largely on the accuracy of the Convolutional neural network implemented for community detection. The accuracy of CNN can be improved further by training the model on larger data sets. The proposed system can further be tested in different environments like highways and urban environment, one-way/two-way traffic so that the system is adaptable to varying scenarios.

References

[1] Dave Evans. The internet of things: How the next evolution of the internet is changing everything. *Cisco Internet Business Solutions Group*, April, 2011.

[2] Van Jacobson. Congestion avoidance and control. *SIGCOMM Symposium on Communication Architecture and Protocols*, Nov, 2012.

[3] Van Jacobson. Modified tcp congestion control and avoidance algorithms. *Technical Report*, April, 1990.

[4] T. Henderson S. Floyd. The new-reno modification to tcp's fast recovery algorithm. *RFC 2582*, April, 1999.

[5] E.Altman O. Ait-Hellal. Analysis of tcp reno and tcp vegas.

[6] L.L. Peterson L.S. Brakmo. tcp vegas: End to end congestion avoidance on a global internet. *IEEE Journal on Selected Areas in Communication, vol. 13*, pp. 1465-1490.

[7] Md. daizul Bari et al. A survey of naming and routing in information-centric networks. *IEEE Communications Magazine*, December, 2012.

[8] B. Ohlman Bengt Ahlgen, Christian Dannewitz. A survey of information-centric networking. *IEEE Communications Magazine*, July, 2012.

[9] et al. Y. Ren, F. Liu. Power control in d2d based vehicular communications networks. *IEEE Transactions on Vehicular Technology*, 2015.

[10] et al. W. Sun, D. Yuan. Cluster-based radio resource management for d2d-supported safety critical v2x communications. *IEEE Transactions on Wireless Communications*, 2016.

[11] et al. R. Yu, J. Ding. Optimal resource sharing in 5g-enabled vehicular networks: A matrix game approach. *IEEE Transactions on Vehicular Technology*, 2016.

[12] et al. X. Hu, V. Leung. Social drive: a crowdsourcing-based vehicular social networking system for green transportation. *3rd ACM international symposium*, 2013.

[13] et al. Zhenyu Zhou, Caixia Gao. Social big data based content dissemination in internet of vehicles. *IEEE Transactions on Industrial Informatics*, 2017.

[14] et al. S. Smaldone, L. Han. Roadspeak: enabling voice chat on roadways using vehicular social networks. *Social Network Systems*, 2008.

[15] et al. P. Shankar, M. Muscari. A case for automatic sharing over social networks. *1st ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*,

2012.

[16] et al. T. H. Luan, X. S. Shen. Feel bored? join verse! engineering vehicular proximity social networks. *IEEE Transactions on Vehicular Technology*, 2015.

[17] Jian Kuang and et al. Shun-Zhang. Broadcast-based content delivery in information centric hybrid multihop wireless networks. *IEEE Communication Letters*, 2016.

[18] et. al Jaebeom Kim, Daewook Shim. Topology aware content-centric networking for mobile ad hoc networks. *IEEE International Conference on Networking*, 2013.

[19] et. al Chan Min Park, Asif Rehman. Packet flooding mitigation in ccn-based wireless multimedia sensor network for smart cities. *IEEE Access*, 2017.

[20] et. al Jaime Llorca, Antonia M. Tulino. Energy efficient dynamic content distribution. *IEEE Journal on Selected Areas in Communication*, 2015.

[21] et. al Hafra Maryam, Munam Ali Shah. Tfs: A sociable routing protocol for vehicular content-centric networks. *IEEE International Conference on Emerging Technologies*, 2017.

[22] et. al S.H. Ahmed, S.H. Bouk. Codie: Controlled data and interest evaluation in vehicular named data networks. *IEEE Transactions on Vehicular Technologies*, 2016.

[23] et. al S.H. Ahmed, D. Kim. Conet: Controlled data packets propagation in vehicular named data networks. *IEEE Transactions on Vehicular Technologies*, 2016.

[24] et. al J. Jeong, S. Guo. Trajectory based statistical forwarding for multi-hop infrastructure-to-vehicle data delivery. *IEEE Transactions on Mobile Computing*, 2012.

[25] et. al L. Briesemeister, G. Hommel. Localized group membership service for adhoc networks. *IEEE International Conference on Parellel Processing*, 2002.

[26] M. Aamir. Content priority-based interest forwarding in content centric networks. *arXiv:1410:4987*.

[27] et. al Yi.B.Ko, N.H. Vaidya. Geocasting in mobile ad hoc networks: Location-based multicast algorithms. *Mobile Computing Systems and Applications*.

[28] et. al R. Chand, P.A. Felber. A scalable protocol for content-based routing in overlay networks. *IEEE International Symposium on Network Computing and Applications*, 2003.

[29] et. al G. Grassi, D. Pesavento. Vanet via named data networking. *IEEE Conference on Computer Communications*, 2014.

[30] et. al G. Grassi, D. Pesavento. Navigo: Interest forwarding by geolocation in vehicular named data networking. *IEEE International Symposium on a world of Wireless, Mobile and Multimedia Networks*, 2015.

[31] et. al M. Kuai, X. Hong. Density-aware delay tolerant interest forwarding in vehicular named data networking. *IEEE Symposium for architectures for networking and communication systems*, 2012.

[32] et. al S.H. Ahmed, D. Kim. Rufs: Robust forwarder selection in vehicular content-centric networks. *IEEE Communication Letters*, 2015.

[33] et. al Z. Yan, S. Zeadally. A novel information architecture based on named data networking. *IEEE Internet of Things Journal*, 2014.

[34] et. al L. Wang, R. Kuntz. Rapid traffic information dissemination using named data. *Proceedings of 1st ACM Workshop on Emerging Name-Oriented Mobile Networking Design Architecture Algorithms*, 2012.

[35] et. al X. Li, S. Venticinque. Intelligent distributed computing. *Concurrent Computing*, 2016.

[36] et. al C. Bian, T. Zhao. Boosting named data networking for efficient packet forwarding in urban vanet scenarios. *IEEE Workshop on Local Metropolitan Area Networks*, 2015.

[37] et. al Y. Xian, C.T. Huang. Traffic aware geographic forwarding in vehicular adhoc networks. *Proceedings of 9th ACM International Workshop in Vehicular Inter-Networking*, 2012.

[38] et. al Junaid Ahmed Khan. Saving: Socially aware information centric networking. *IEEE Communications Magazine*.

[39] et. al I. Lequerica, M. Garcia Longaron. Drive and share: Efficient provisioning of social networks in vehicular scenarios. *IEEE Communications Magazine*, 2010.

[40] et. al Stephen Smaldone, Lu Han. Roadspeak: Enabling voice chat on roadways using vehicular social networks. *ACM workshop on Social Networks*, 2008.

[41] et. al Azza Allouch. Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. *IEEE Sensors Journal*, 2016.

[42] et. al Xiping Hu, Victor C. M. Leung. Social drive: A crowdsourcing-based vehicular social networking system for green transportation. *DIVANet*, 2013.

List of Publications

1. Amuleen Gulati, et al. *"Deep Learning-based Data Dissemination Scheme in Content-Centric Internet of Vehicles"*, IEEE International Conference on Communications, May 2018. [Accepted]