A Framework to Simulate Bulk-Data Migration Using Vehicular Traffic

By

Murk

00000170349

Supervisor

Dr. Asad Waqar Malik

Department of Computing
School of Electrical Engineering and Computer Science (SEECS)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

November 2018

A Framework to Simulate Bulk-Data Migration Using Vehicular Traffic

By

Murk

00000170349

Supervisor

Dr. Asad Waqar Malik

A thesis submitted in conformity with the requirements for the degree of

Master of Science in

Information Technology

Department of Computing
School of Electrical Engineering and Computer Science (SEECS)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

November 2018

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by ***Ms. Murk***, 00000170349, of School of Electrical Engineering and Computer Science (SEECS) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ________________________

Name of Supervisor: _**Dr. Asad Waqar Malik**_____

Date: ________________________

Signature (HOD): ________________________

Date: ________________________

Signature (Dean/Principal): ______________________

Date: ______________________

Approval

It is certified that the contents and form of the thesis entitled “A Framework to Simulate Bulk-data migration Using Vehicular Traffic" submitted by Murk have been found satisfactory for the requirement of the degree.

Advisor: **Dr. Asad Waqar Malik**

Signature: __________________

Date: __________________

Committee Member 1: **Dr. Imran Mahmood**

Signature: _________________

Date: _________________

Committee Member 2: **Dr. Anis ur Rahman**

Signature: _________________

Date: _________________

Committee Member 3: **Dr. Muhammad Muneebullah**

Signature: _________________

Date: _________________

Certificate of Originality

I hereby declare that the research work titled ***A Framework To Simulate Bulk-data migration using vehicular traffic*** is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at National University of Sciences & Technology (NUST) School of Electrical Engineering & Computer Science (SEECS) or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: Murk _

Signature: _____________

I dedicate this research work to my parents and my siblings whose cooperation and exceptional support led me to this amazing accomplishment.

Declaration

I, Murk declare that this thesis titled "A Framework to simulate Bulk-data migration Using Vehicular Traffic" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated

3. Where I have consulted the published work of others, this is always clearly attributed

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work

5. I have acknowledged all main sources of help

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Murk,

00000170349

Copyright Notice

Acknowledgments

I thank my Creator Allah S.W who provided support and guidance to me throughout this amazing journey of my research work. No doubt, that without Your exceptional support and pure guidance, nothing could have been done by me.

I am thankful to every individual who helped me throughout this work and I am grateful to my beloved parents who raised me and continued to assist me in each and every aspect.

Thanks to my advisor Dr. Asad Waqar Malik for guidance, tremendous support, and cooperation throughout the research work. His patience and proper guidance throughout my thesis is greatly appreciated. I am thankful to him for his support from the initial to the final level and for enabling me to build a proper understanding of my subject.

Also, special thanks to Dr. Imran Mahmood, Dr. Anis ur Rahman, and Dr. Muhammad Muneebullah for being on a supportive and helpful thesis committee and also for being kind enough to help me whenever I needed their guidance and support.

Abstract

Data explosion in Smart cities is facing an exponential increase. Not only the field of Computer Science and Information Technology, but all other fields such as Science, medicine, arts, engineering etc. perform data-acquisition for future business analysis and organization's growth. Although there are many fields but the progress of smart cities seems to be the leading cause of data explosion. Nowadays, sensors and Internet of Things generate an amount of data which is far beyond our imaginations. With the emerging data there occurs a need to assess data dissemination techniques, such that the delivery of big-data shall be within optimal time and least cost. Data transfer techniques include PLC, Wi-Fi, DSRC, Optical-fiber etc. which have their own merits but demerits are more. Some are expensive to purchase, while the others cost more in terms of delay and energy. Hence, a technique is required for big-data transfer which is affordable by all means. In our proposed framework, we have put forward a centralized system which provides data transfer using daily Road traffic. Vehicles act as data-carriers picking up data from the source, where data hops over intermediate data-spots by vehicles and finally is delivered to the destination. Our aim in this research is to prove that the data transfer by using Road traffic is efficient as compared to Internet in terms of delay and as well as energy. We propose a technique in which data transfer is availed by using already available resource i.e. vehicles, thus saving time, resources and money.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

1.1.1. Big Data

Big data means a big volume of data. It is the data that is generated from huge data-centers, large business organizations and big enterprises or corporations. Big-data includes both structured and unstructured data [1]. The structured is the type of data that is stored in a structural and organized manner. It is stored in relational databases and can be used in a very effective way as compared to the semi-structured and unstructured data. A Structured type of data can be found in SQL data-bases. Unstructured data is the type of data that is completely un-organized. Examples of unstructured data can be text-messages, emails, written digital documents in Ms. Word files, etc. Semi-structured data is half-structured and half- unstructured. This type of file can be found in No-SQL data-bases.

Only 10 percent of the data is structured, 5 to 10 percent is semi-structured and 80 percent of the whole data being stored today lies in unstructured category [2].

Big-data can be very much helpful in making better and optimal decisions, also helps in making improvements in the conditions of a current system. There are different properties of big-data that are famous [1,3]. Some of them are volume, variety, velocity, variability, veracity, vulnerability, validity, volatility, value, visualization and complexity. Based on the nature of big-data, one can derive many other properties apart from the ones described above.

1.1.2. Digital Storage Device

Data which is fetched from various environments by different technologies need to be stored somewhere. The devices required for storage of big-data can be divided into three types. First is HDD, which is a hard disk drive is a traditional storage drive or a type of mechanism or a device that is used for reading and writing or even to control the positioning of the hard disk. These are termed as hard-drives because these include physical tangible moving parts which are responsible of performing the operations on a system [4]. Its merits are that it is cheap, has large storage capacity and is easily available. The drawbacks are that it has smaller lifespan, is less reliable, has more chances or wear and tear due to mechanical nature, is prone to damage, has low speed of data transfer and suffers with more power consumption. Second type of digital storage device is SSD. These are the devices that similar to the HDD are used to store big-data, but does not include moving parts as HDD does. SSD uses flash memory [5]. It is smaller in size and light weight. Its merits are that it has a very high speed, has good performance, is shock resistant, has low power consumption and longer life span as compared to HDD. But, SSD is very costly, has less storage capacity and is rarely available in market. The third type of digital storage device is SSHD. It is a hybrid form of storage device. It uses the advantages of both SSD and HDD. A SSHD system usually uses both these drives. The frequently accessible data is stored in SSD and a less used data is stored in HDD [6].

1.1.3. Cloud Storage

A cloud storage is a virtual storage system where data is being stored in servers remotely located from the user. This storage is accessed by means of Internet. The name of this storage is Cloud. The storage, transfer and maintenance is provided by the service providers of cloud storage [7]. There are different types of services offered by Cloud i.e. Saas, PaaS and IaaS [8].

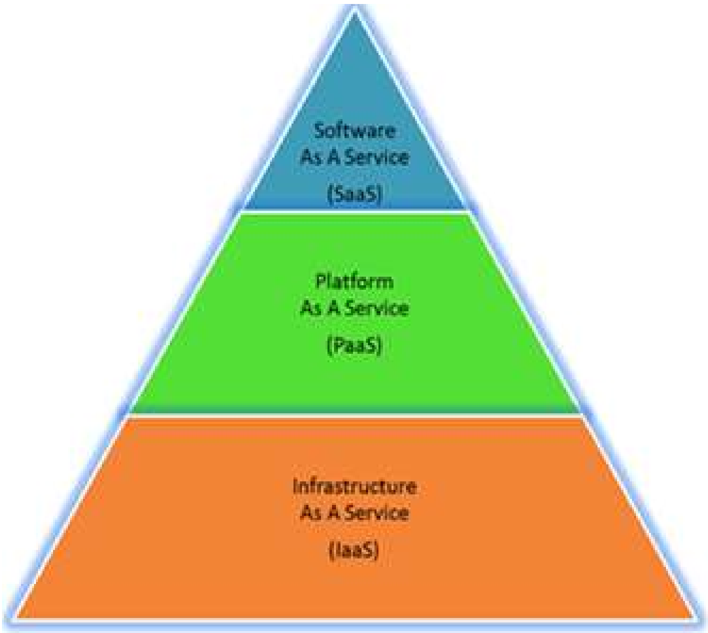


Figure 1.1: Types of Cloud Services [9]

The cloud service providers sometimes offer complete software to their customers to use. Such type of service is called Software As A Service or SaaS. Customers can access and manage the software on their own (just a part of it which is accessible to them), but the disadvantage is that they can use it only when they are online.

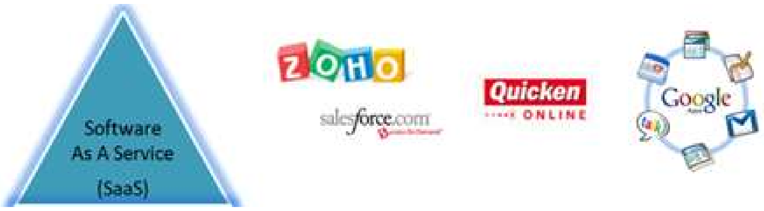


Figure 1.2: SaaS and its applications

A type of cloud service which offers its customers the complete development environment to use, or a complete development platform e.g. Azure. This is also known as PaaS. The disadvantage of this service is that although it provides ease but it can cause problems such as shifting the developed application to another environment.

Platform
As A Service
(PaaS)
Google
force.com
platform as a service
Windows Azure

Figure 1.3: PaaS and its applications

In this type of service, the cloud offers some computing power. Such as storage services or database services etc. This helps the customers to use a type of service which usually would need big data-servers running in the background. Examples include Amazon Web Services. This saves a lot of cost. This type of service also known as IaaS.

Infrastructure
As A Service
(IaaS)
amazon
webservices
rackspace cloud

Figure 1.4: IaaS and its applications

1.1.4. Data Explosion

The volume of big-data arriving from the systems is increasing with such a rapid pace which seems as if it is exploding. The amount with which big-data is constantly being produced every minute is increasing exponentially. Big data, now is not only an issue to be stored and managed. But, it is now of great concern to the consumers, customers and end-users to get this data correctly, efficiently, quickly and in the most appropriate manner which seems good while visualizing [10].

The big-data is not only used for technical purposes, but now it is mostly used for better customer or end-user service. One can consider the example of watching movies on a website. The data of every customer is stored and tracked in order to know the customer behavior and then provide recommendations for next movies he/she will prefer to watch.

Furthermore, one can take the example of any shopping website. A website that includes selling customers a variety of products, usually tracks the customer’s past experience to provide better and valuable recommendations in future.

These recommendations are a result of using the big-data of millions of customers in valuable and effective manner. Hence, one can think of many other websites such as Google Ads, YouTube videos, Facebook pages etc. These all websites use the data of end-users who are registered in the website and whose data is being stored in remote servers. Hence, every minute there is an explosion of hug data streams arriving on the way.

Followings are the data explosion values for different websites and applications. Fig. 1.5 shows the data explosion values in the year 2015.

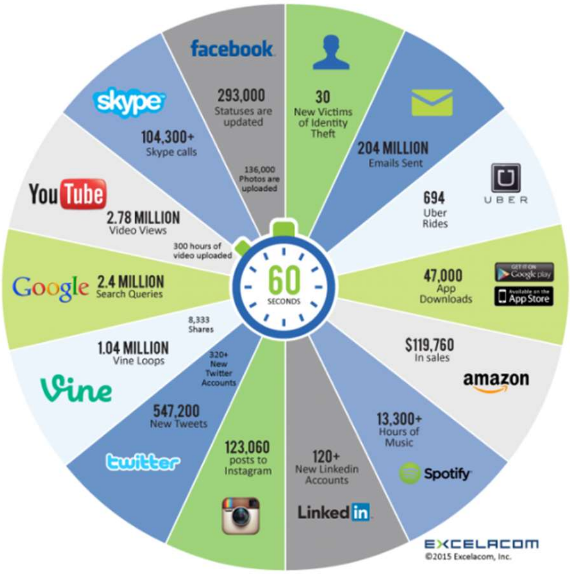


Figure 1.5: Data Explosion per minute in 2015 [10]

1.1.5. Data Transfer Techniques

Big data is generated from a variety of systems and devices. This data is of no use if it is not being transferred to an appropriate system which is capable enough to use it valuably. Moreover, there can be various other situations or needs to transfer the bulk of data to another system. Transferring data requires a valid medium between the source and the destination where the data is to be transferred. Data transfer techniques can be divided into three parts i.e. Wired, wireless and data transfer by physical shipment.

A wired data transfer is a type of data transfer technique which enables data to be transmitted by means of some wire. E.g. Optical-fiber communication, telephone and cable networks, internet by means of LAN etc. Nowadays USB versions are widely used to transfer data from one device to another. Such as USB 3.0, USB 3.1, USB 3.2.

Second type of data transfer media is wireless, which allows the data to be transmitted without a wire in the form of signal waves e.g. Wi-Fi, Bluetooth, DSRC etc.

Nowadays the data can also be transferred from one location to another by means of physical shipping devices e.g. vehicles.

1.1.6. Amazon Web Services (AWS)

It is a platform offerring cloud services. It is secure and offers services e.g. storage and database services, computation power etc. [12]

AWS offers big-data transfer by means of physical shipment. It has three famously known strategies to transfer customer's big-data from customer's location to AWS Data-center, where it is then moved to AWS Cloud. These services are Snowball, Snowball edge and Snowmobile [13,14,15]. Upon request, AWS offers its storage device and ships it to the customer's location. Customer then takes the device and transfers his/her big-data onto the device. The device is then sent back to the AWS datacenter. Different types of AWS big-data shipping services are offered i.e Snowball, which is used to transfer customer's 80 Tera-bytes of data, Snowball Edge which is a service used to transfer a 100 Tera-bytes of customer's data and Snowmobile which is a service used to transfer a 100 Peta-bytes of customer's data.

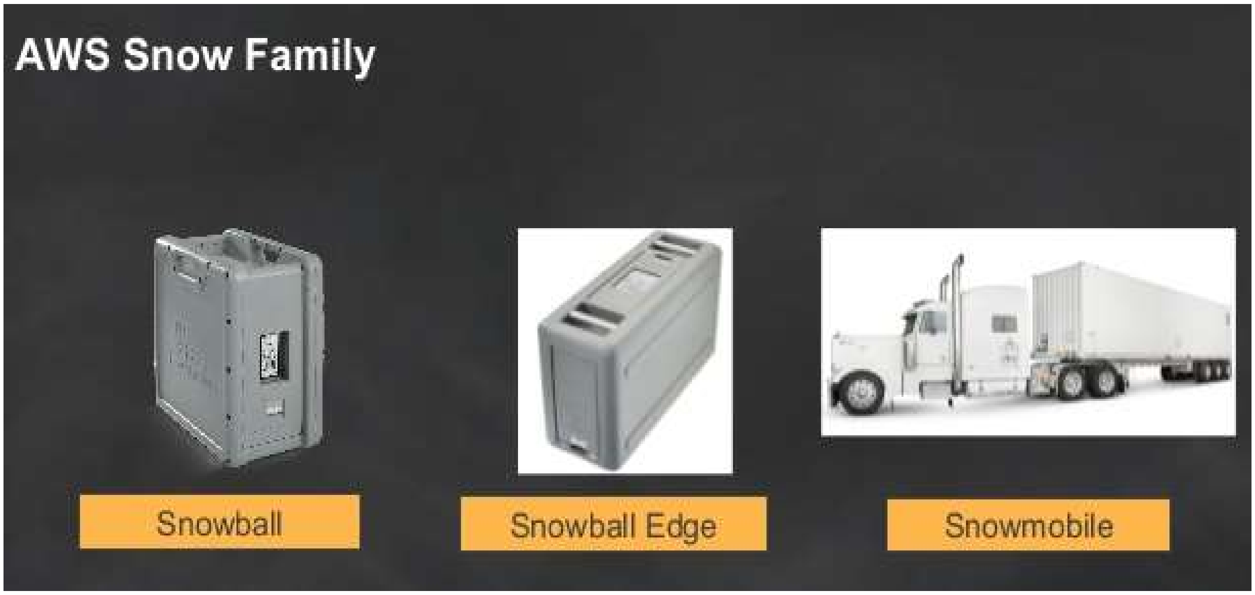


Figure 1.6: AWS SnowFamily [12]

1.1.7. Delay Tolerant Networking

It is also known as DTN is a network of computers used to send data and is facing with a continuous issue of loss in network connections [16]. Another term named as Disruption tolerant networking is also used in this regard. In DTN the routing protocols use Store and Forward strategy while building end to end paths. Since there are more chances of incomplete data transfer following end to end route, therefore the devices in between the two ends, store the data and then make a decision to choose the next hop. In this way the data moves from hop to hop and finally reaches its destination.

1.1.8. Vehicle Based Communication

When the data is being transferred either from the vehicle or by the vehicle it is termed as Vehicle based communication. There are two such types of communication i.e. V2V and V2I. V2V stands for Vehicle-to-Vehicle communication, in which the vehicles interact with each other by means of some short range waves. This can be helpful in transferring data of accidents ahead on the road and other letting the vehicles know so that they can choose another path. V2I stands for Vehicle-to-Infrastructure communication in which the vehicles communicate with the Road-side Units also known as RSUs which might offer some type of service for the smart vehicles such as Internet.

1.1.9. Points of Presence

Points of Presence or POP is a system which permits the users located remotely to connect over the Internet [17]. Any Internet Service Provider or a Telecommunication service provider can consist of a POP. A POP contains different types of devices such as switch, router, communication device, server etc.

1.1.10. Smart City

Smart city is a term used for those cities or areas which aim to make technological developments in the environment to optimize the surrounding functions and to provide a quality life to its citizens [18].

1.1.11. IoT

System of devices inter-related to one another, forming intelligent network, used to transfer data which is independent of any interaction with humans [19].

1.2.Motivation

The word data to explain it is smaller for itself. Data is exploding. It is not just the fields of computer science, Information Technology, Software Engineering and online web based systems, but the big data is associated with all the fields imagined so far. It might be medicine, sociology, history, education, finance etc., big-data is everywhere. All the data that a company or an organization gets from its clients or as a result of some test or by any other means, it needs to be stored somewhere. The data that is so huge is growing so fast that the pace is difficult to match up. Also, the data storage is not only a just alone a problem. There are many other factors to be dealt too. These factors include data security, privacy, maintenance, processing, transfer or dissemination, organization and much more [4]. These challenges are also growing along the growth of big-data. The data explosion is shown in the figure 1. This figure represents data growth over Internet. One can see that how large the values are. Facebook, twitter, Uber, Google etc. all have many customers and the data that is stored in their data centers is not just about the customer himself/herself. But this data includes all the activities a customer performs while surfing the website, all the links he/she clicks, all the comments, likes, reviews etc. are being tracked, and then later on used to provide best of the recommendations to the end-user [6]. This helps the company grow. But all the storage and processing need some effective approach to handle big-data.

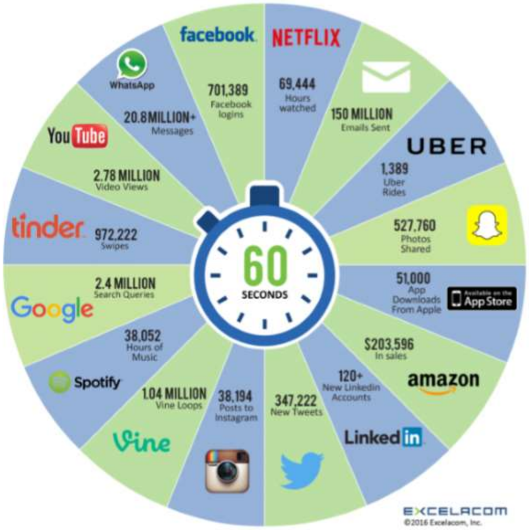


Figure 1.7: Data Explosion per minute in 2016 [6]

The data increase estimated for upcoming ten years reaches trillions of gigabytes [5]. According to Cisco the data increase is in the factor of Zeta-bytes per month [7]. Figure 1.8 shows a clear picture of data increase as a prediction of Cisco.

Data transfer techniques that are most commonly preferred are usually the core networks. These networks are connected by means of routers, switches and many other intermediate devices. People believe that the data transfer by this network is less costly and is an environment-friendly approach. This is not completely true. The device that are connected together to form a core network also emit carbon-dioxide gas. These devices also consume a big amount of energy which is not known by many. The Internet is easy to use and quick to access. But in order to keep it easy and quick there are a number of data center equipment behind which work and manage the big-data of millions of end-users surfing over the web. The network is overloaded and is costly in terms of energy and delay.

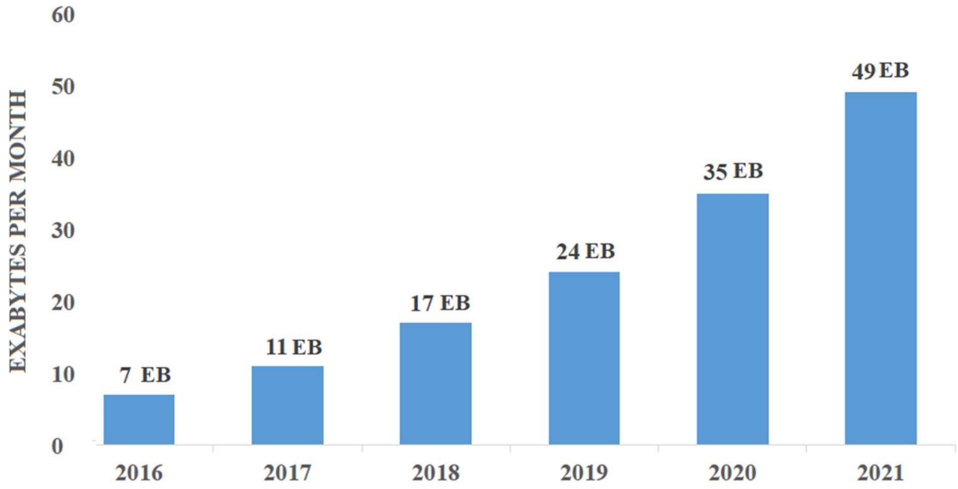


Figure1.8: Increase of data in EB per month

Other approach used by researchers for data dissemination is by means of physical shipment of storage device from one place to another. Physical shipment actually costs less than the overall calculated data transfer by means of Internet [8], as shown in figure 1.9. Since physical transport also has its costs of transportation and fuel consumption, hence the better approach shall be, using combination of both.

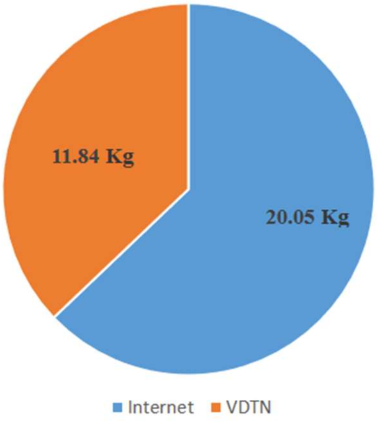


Figure 1.9: Carbon Emission Internet Vs. VDTN for 20TB

Amazon Web Services follow data transfer by means of shipment. Whoever wants to move their big-data of Petabytes to AWS cloud, then he should request AWS. AWS sends a storage device to the customer by means of a truck or any other vehicle. The customer gets the device and moves the entire data into the AWS storage device. The device will be then moved back to the AWS data-center. Then in the AWS data center, the customer's big-data is moved to the AWS cloud.

This system is quite easy but there are few drawbacks we have observed as follows:

- Client has to pay for to move data to AWS.
- Client will pay additional money which increases per day if the data transfer is late.
- If any problem occurs or the device is lost, then the client shall pay penalty.
- If the vehicle goes through some accident or the device is stolen, then client's entire data can be lost.
- Client in few cases has to pay the driver as well.

These drawbacks are enough to know that the system is good but there is additional improvement needed.

Our framework provides a solution to all the above problems. It provides a combination of both wired and physical shipment approach. Our framework aims to minimize the burden of the core network and helps to utilize the already being consumed road resources.

1.3.Research Objectives

With the growth of vehicles in today's world, we see that more researches are being performed over vehicles in order to make them an effective part of Smart cities. Much of the work is being done on keeping the vehicles more secure for humans and for minimizing the accident rates. Different kinds of intelligent devices and sensors are placed inside vehicles to make them more safe and intelligent while driving. Automated cars and electric trains have been emerged which need no driver.

Apart from making the vehicles more intelligent for human safety and easy transportation, an area of focus nowadays is more towards making vehicles a smart member of smart cities. Different types of services are taken from them such as running applications, providing internet accessibility to nearby devices. This carries vehicles more towards V2V and V2I approach. These smart vehicles not only act intelligently but now they act for human service. The reason for advancing the vehicles is because of the increased road traffic. Since fuel and energy is being consumed over vehicles, so people are tending to use these available resources for a more valuable purpose apart from just transportation.

Since the core network is already over-burdened by so many tasks. The big exploding data does not find its way within it. To reduce the burden over other data dissemination systems we move a step further in using vehicles for big-data delivery tasks. There are other systems too, which provide data delivery services by means of vehicles.

One of such famous system is Amazon Web Services. AWS has a huge cloud storage where the different cloud based services are being offered to customers. To utilize these services, many of the big data centers move their entire data center to AWS cloud. In order to follow this approach, AWS offer their vehicle for delivery of a big storage device. The vehicle delivers the storage device to the customers and the customer move their data to the device, which this returned to AWS data center and moved to AWS Cloud. This method of moving data from one place to another is costly and time consuming.

We have developed a framework which resolves this issue. In our system, we have split the job of big-data transfer to multiple vehicles which reduces burden from one vehicle. Secondly we have used vehicles as a road resource and we do not aim to hire them. This saves a lot of cost. The vehicles can be provided with credit points for moving the data from one place to another. This shall encourage the people to participate more.

By the help of vehicle data transfer, we aim to minimize the costs in terms of delay as well as energy. Furthermore, we aim to perform big data transfer job by using a combination of two approaches i.e. wired data network and the physical data transfer.

Following are the objectives of our research:

✓ **Data security**

The data is moved from the data center to the vehicle’s device and vice-versa by means of wired network. A wired network neither gets interrupted due to outside environment nor it gets attacked by other networks outside. This helps to keep the data secure. Furthermore, encryption shall be applied to the data so that the driver does not misuse it. Backup of the data is kept in the source data center and the data spots until the receiver receives the whole data.

✓ **Data integrity**

There is a continuous check inside our framework, where entities continuously make sure whether the data which has moved from the current spot has reached the next spot or has lost in transit. In case the data is lost due to vehicle breakdown, then the data is requested for retransmission. Data in our framework shall be divided in chunks and each chunk is assigned a unique chunk id. This helps in retransmitting only that chunk which has been lost.

✓ **Energy efficient**

By the help of equations and experimental results we shall prove that the data transfer by our framework costs much less in terms of energy as compared to the traditional data transfer approach i.e. Internet.

✓ **Less delay**

Also, we shall prove our framework in the form of results and graphs that it takes less time to transfer Terabytes of data by our system rather than Internet.

✓ **Saving Resources**

We in our research aimed to utilize the already being consumed resources i.e. vehicles. Since the roads are full of vehicles, we aim to save resources being used in traditional data dissemination techniques and we use the common vehicular resource over some more valuable purpose apart from just road transportation.

Following are the Sustainable Development Goals of our research:

✓ **Affordable and clean energy**

Bulk of the data is transferred by utilizing the vehicular resources that are already in use, saving energy from other methods specially designed for the big data transfer.

✓ **Sustainable cities and communities**

By providing smart vehicular data transfer, the framework shall approach to manage cities more effectively and holistically, such that intelligent transport systems allow efficiency with managed energy consumption.

✓ **Responsible consumption**

Energy that is consumed in by our framework shows that how one can efficiently consume the resources for multiple ways at same time.

✓ **No poverty**

Vehicle drivers can make their daily earnings just by helping the data-centers to carry data from one location to another. This system might be extended to a bigger-level helping to reduce un-employment.

Chapter 2

Literature Review

Big-data challenges are uncountable. It is being observed that every minute or even less, the data streams are flowing with an unstoppable rate [11]. Although dealing with the challenges of big-data is very hard but the benefits that can be achieved from it are immense.

Our framework deals with big-data transfer while saving a lot of costs. This is done by using vehicles as the transfer media. In this section, the work related to big-data and its transfer are highlighted briefly, which greatly relate to our research domain.

Sheikh Mohammad Idrees et al. [21] have highlighted very notable set of challenges that the world is facing while dealing with big-data. One can observe, that this increase in data is growing with the increase in vast number of organizations, businesses and companies that run their business some way or the other online. Also, the data that is generated online through these businesses is being saved for later valuable analysis. Storing big-data not only helps in analyzing the business current situations but also helps to increase the organization's progress. The main issues of big-data highlighted in [21] include Storage, data security and the data itself. Data is not only produced by the professionals or the large companies, but it is also being continuously generated at the pace of Tera-bytes by a normal end-user working from home. Storing such large amount of data is a big issue because it shall need large storage devices extending up to Peta-bytes. Not only handling and storing the big-data is an issue but the data itself is an important issue to be highlighted. There are different types of data being generated e.g. audios, text messages, emails, videos, photographs, written files and documents etc. Data security is also a great concern. Once the data is handled it needs to be so properly encrypted that it should not be accessed by the third party. Maintaining security is itself another domain. Apart from these, there are many others such as issues of data processing, management, forecasting etc.

Looking through the above big-data challenges there seems a need for an efficient big-data handling system.

Nowadays the data is not only being traveled through wires or waves but is being carried by vehicles. This is an Amazing approach which saves a lot of cost and time, along with minimum resources. Example of Amazon Web Services shall be at the top while explaining data transfer by means of physical shipment. Upon request, AWS offers its valuable customers with storage device. This device helps in the transfer of large data from the customer's data centers who want to shift their huge data on to AWS Cloud. There are three such types of devices used for data transfer in AWS [13,14,15]. These are Snowball, Snowball Edge and Snowmobile.

A very large company named Digital Globe shifted its entire data on to the AWS Cloud [24]. Enormous data which arrives from satellite, be it a weather data or maps or earth imagery [25]. This large amount of data was not only the one arriving currently but the company had to save the past data too.

AWS offers best cloud services i.e. IaaS. Based on this reason the Digital Globe decided to shift their entire data to the AWS Cloud, which was achieved by AWS Snowmobile. Snowmobile is AWS truck which is shipped to the customer such as Digital Globe, and has the capacity of 100 Peta-bytes of storage [15]. Once the truck arrives the customer's data center, all the data is moved to the truck which might take days. After that the truck is taken back from the customer to the AWS data center and then customer's data is moved to AWS Cloud.

A similar project experimented was in France, which is the work of Baron et al. [26]. This work focuses on the Delay Tolerant Network (DTN). The research aimed to transfer big amount of data by means of vehicles. The purpose to use this approach for data transfer was minimizing the network load. Furthermore, max-min algorithm and other strategies were also adopted to obtain overall effective results. Also, an analysis was done, in which the two approaches were compared i.e. centralized and decentralized. The results of analysis found centralized approach as the optimal one. The problem with this system was that the data loaded on to the vehicles was by the help of Dedicated Short Range Radio waves which, due to its small range and less data transfer capability, does not load large amount of data chunks to the vehicle, making the whole process slower.

Our framework uses the shipping approach but to load the data from the system to the vehicle, the medium used in wire. Wired devices offer better reliability, security and are not susceptible to the environment whereas the wireless waves easily get interrupted by other waves or by obstacles [27].

Various types of researches have been performed on providing vehicles the capability to offer Cloud based services. These smart vehicles make the world an easy place. Following are a few of many works done in this research area.

Onur Altintas et al. [28] proposed a framework named Car4ICT. In this framework the vehicles act as members. The job of a member is to provide any service the client asks for. Clients can be cars or any device used by the end-user. The member cars offer different services. For example, a temporary storage can be provided to a tourist capturing pictures and wants to upload the data over the cloud to access them later. Cars then later will share data over the cloud, so that when the current car has left, the tourist can his regain his previously stored data from any other car available. These types of services can be very helpful when other medium does not work e.g. in case of disasters.

Bo Li at el. [29] proposed a framework similar to the above. In this framework, different vehicles work together to accomplish a certain task. The task might be any computation or data processing which takes a lot of computation power and space. Therefore, vehicles share their resources and work in parallel sharing parts of an application.

Falko Dressler et al.[30] proposed a framework in which there are two type of vehicles. The ones that are parked and the other which are moving. The job of the parked vehicles is to provide services to the ones on the road. These services include storage of information over the cloud and to provide other cloud services to the vehicles.

Ivana Marincic et al. [23] proposed a research work, in which two approaches for transferring big-data were compared. One of the approaches was transferring data by means of physical shipping. This means that the data will be moved from the source to destination by copying it on a disk and then this disk will be moved to the destination by some vehicle. Different types of vehicles were used in this regard. A bicycle, an SUV, airplane etc. were considered and the energy was computed for each vehicle. Second approach was to use the Internet as a data transfer medium. The energy in this case was also computed. Comparison of the two ways i.e. vehicle and the Internet, it was concluded based on the results and analysis that the Physical shipping of the data consumed less energy and is more efficient as compared to the Internet. One of its results is shown in Fig. 2.1.

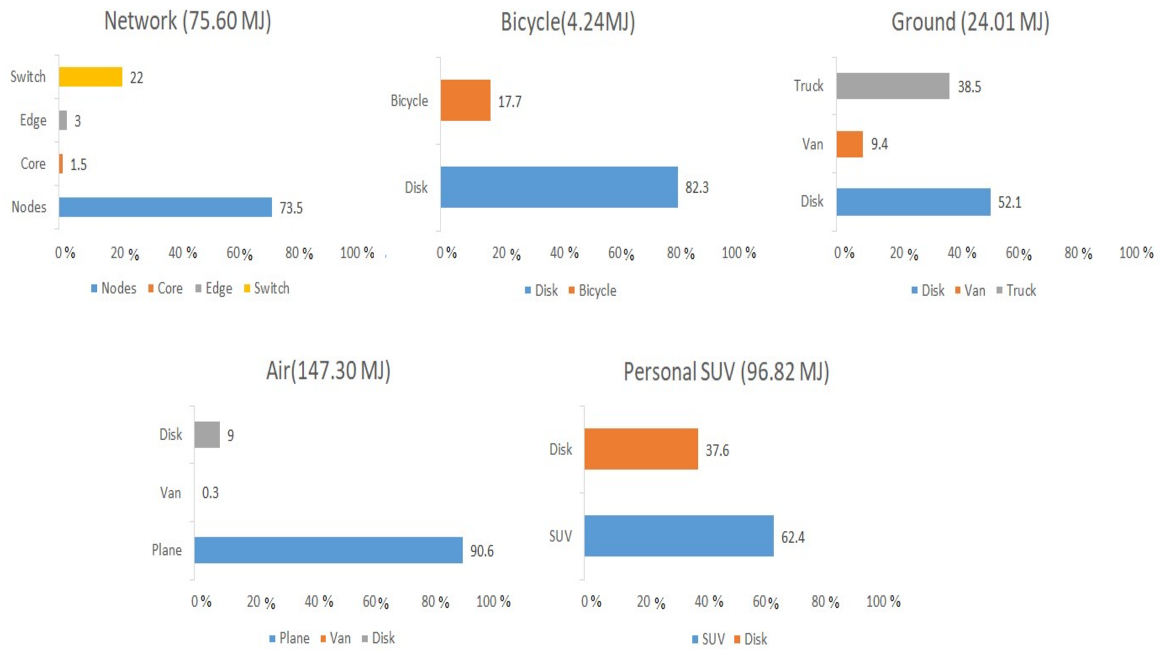


Figure 2.1: Average energy consumption in data transfer[23]

Another work done by Salman Naseer et al. [22] focused on the similar idea. The idea was to take two approaches. One approach was to transfer a big-data by the help of Internet data transfer, as done in a normal case. Another case was to transfer data by the help of vehicles. The difference between this research and the research of Ivana Marincic et al. [23] is that the shipping approach is not solely a responsibility of single vehicle in Salmaan Naseer et al. [22], there can be many vehicles distributing the task, whereas in [23] the shipping was solely a concern of single vehicle. Comparative analysis for the two cases i.e. Internet and vehicle case showed many results. Firstly, the time it takes to transfer the data increases exponentially with the increase in amount of data. Similarly, the energy consumption for the Internet case was found to be more as compared to the vehicle based approach. The result of this comparison showed that with the increase in amount of data to be transferred, there occurs more increase in energy consumption for Internet as compared to the Vehicles.

EV stands for Electric vehicles and ICE stands for vehicles with Internal Combustion Engine. A research done by David A Howey et al. [31], compared energy consumption of two types of vehicles. One category was EV and the other was ICE-vehicles. The results showed that the EV released about 93CO_2 per km and the ICE-vehicles released about 118CO_2 per km. This shows that EV are more energy efficient. Hence, if the physical shipping is chosen for data transfer then one should approach EV.

David Costenaro et al. [32] presented a scenario in the proposed research work. In their research work, data was to be transferred by means of Internet. Upon transferring big-data over the Internet, energy consumption was calculated for every device which was used in this cause. The purpose of this research was to bridge the gaps between Information Technology department and the Energy Management Industry. The reason is that the amount of energy consumed by this department is very high and yet ignored. But if certain actions are applied to manage energy consumption by the help of Energy Industry then this problem can greatly be reduced. Research divides the whole process of Internet data transfer to three tiers. Tier1 POPs consists of Huge data centers, Network Access Points etc. These systems are at higher-level. These are usually seen at National or International level. Tier2 consists of Medium level data centers which can be at a regional scale. Smaller than Tier1. Tier3 consisted of small data centers or ISPs which are involved at local scale to carry the data over the above tiers. In the last, there are end-users which use and avail services from the above tiers as shown in figure 2.2.

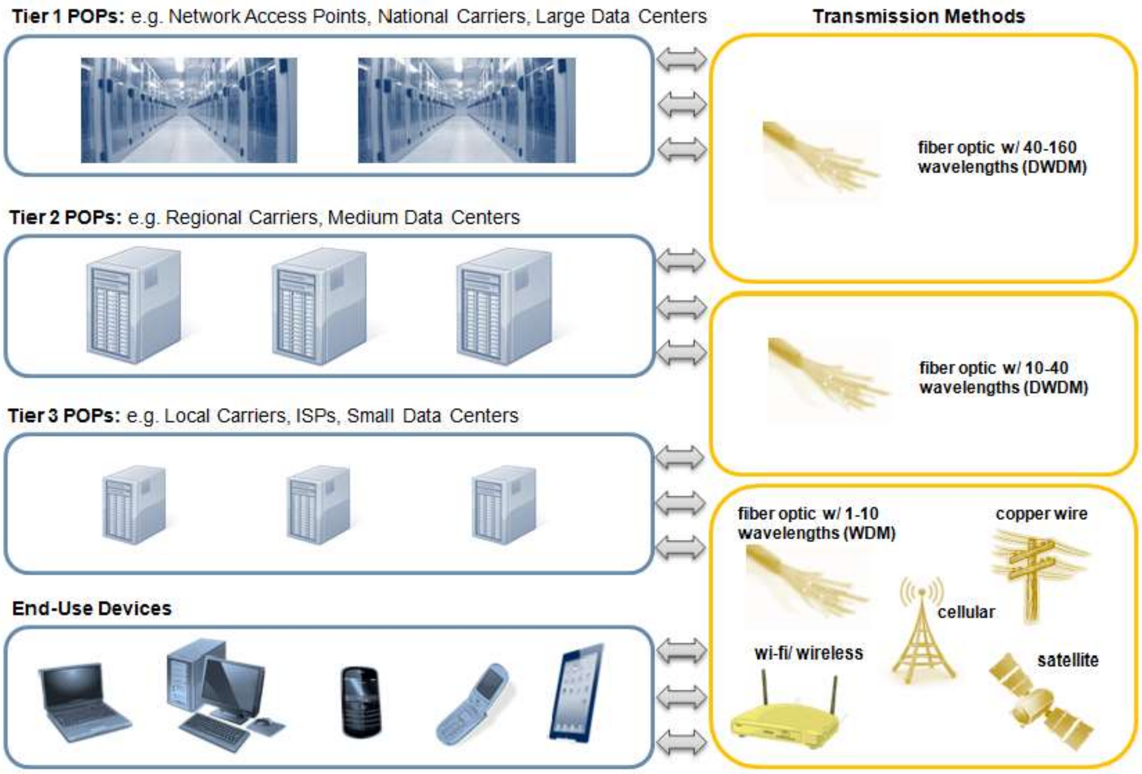

Tier 1 POPs: e.g. Network Access Points, National Carriers, Large Data Centers
Transmission Methods
fiber optic w/ 40-160 wavelengths (DWDM)
Tier 2 POPs: e.g. Regional Carriers, Medium Data Centers
fiber optic w/ 10-40 wavelengths (DWDM)
Tier 3 POPs: e.g. Local Carriers, ISPs, Small Data Centers
fiber optic w/ 1-10 wavelengths (WDM)
copper wire
cellular
End-Use Devices
wi-fi/ wireless
satellite

Figure 2.2: Three tiers of Internet Infrastructure [32]

Based on the equipment these tiers use and the energy they consume; the overall energy consumption was calculated if data was to transferred in GB. Upon analyzing the outputs of calculations, incredible results were observed. Results concluded that out of the total energy consumed over the Internet while data transfer in GB, the costs associated with the end-users was only 38%. The costs associated with the transportation of data was about 14%, whereas the data-centers costed 48%.

The main results of this research highlight the average amount of power consumed while data transfer via Internet, which was found to be 5.12 kWh for every GB. In the end of the research, awareness was provided that how one can keep from consuming more energy over the Internet.

Chapter 3

Problem Formulation and Proposed System Design

3.1.Overview:

The overview of our framework can be easily understood by looking at Fig.4. Our framework represents an example of a smart city. In this system, the smart vehicles and the other system entities are inter-connected with each other by means of some cellular network. There is a centralized entity named as Central controller, which is heart of our system. There are data-centers allocated in smart cities and any of the data-center can send big-data to another by means of our framework. There are set of data-spots which actively take part in order to send data hop to hop from source to destination. Vehicles in our framework can have fixed capacities i.e. they all might have same storage space or the vehicles may have different storage capacities.

Furthermore, there can be more than one data request or job. The sender data-center or the source has a choice to transfer the jobs either in a parallel manner or sequentially.

Each of the jobs which comprises of big-data, is divided into set of smaller chunks which can be easily loaded on to the vehicles and moved forward.

The process of data transfer is initiated by the sender. Whenever the sender data-center wants to transfer set of jobs to another location, it starts to signal this request to the nearby vehicles. Any vehicle participating in this environment shall approach the sender. On reaching, the sender data-center shall upload a chunk from the set of jobs to the vehicle.

Vehicle on carrying the chunk, is assigned a data-spot nearby which is one step ahead from the sender, towards the destination. Each data-spot has a set of stations, where the data is offloaded by the vehicle.

Similar to vehicles, each data-spot has a defined storage capacity. In case the data-spot is full with its capacity, the vehicle is informed prior to moving from the sender's location. In such case vehicle shall leave the task and is free.

Chunks carried by vehicles are uniquely identified and their information is stored in the global data-base which can be accessed by the sender as well as the receiver. By the help of this database the receiver can later check for data integrity.

A vehicle which has carried the chunk has a choice. It depends upon the driver, whether he/she wants to continue the task and offload the chunk to the next spot or leave the task. In this case the chunk that had to reach the next data spot remained undelivered. Cases similar to this such as, a road accident or vehicle breakdown etc. can also result in failure in data delivery.

Receiver data center always checks if all the chunks have been delivered successfully. In case any of the data chunk is lost in transit, then the receiver requests for retransmission of those specific lost chunks.

The routing process is done by the controller. Upon receiving the data transfer request from the sender, the controller creates a path from sender to receiver. This path comprises of set of intermediate data spots that arrive in between the source and the destination.

Fig 3.1 shows the two data centers, an examplary data spot, a central controller communicating with all the entities and the vehicles which carry data and turn the color into red to differentiate in the diagram below.

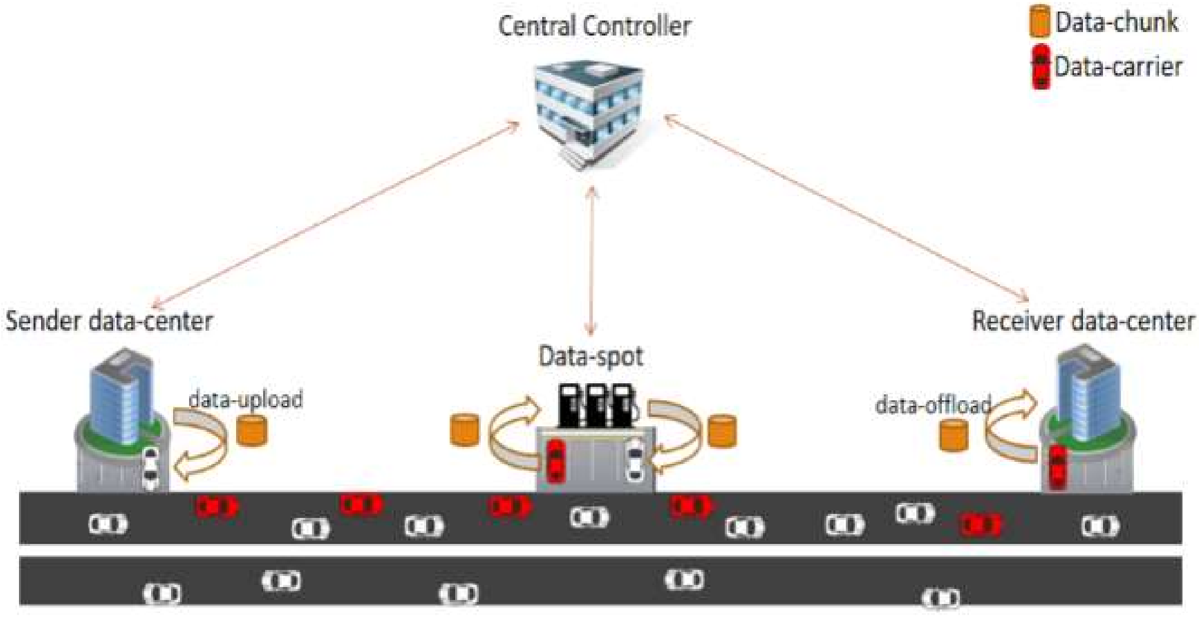


Figure 3.1: Overview of our framework

3.2.System Entities:

We have designed a framework in which there are two data centers and they are located about 340 km apart. The source data-center is located in the territory of Islamabad, Pakistan whereas the destination data-center is located in the city of Faisalabad, Pakistan. There are set of offloading spots in between the two data-centers. There is a central controller. The location of a central controller does not matter, but we have specified a location of it also. Following is the detailed explanation of each entity:

3.2.1. Controller:

The first one with which the source data center communicates while beginning the process is the central controller. This is the main entity of our framework. The sender initializes the process by sending the route request to the Controller. The Controller, upon receiving the request, calulates the route. The route is set of data spots in between the source and destination which forms an optimal path for the data to be moved. The data then shall be moved from the source to the data spots included in the route calculated by the controller.

Controller after calculating the optimal route, makes a decision. The decision is that which of the approach shall be preferred for data transfer i.e. either the data shall be transferred chunk by chunk by means of vehicles or it shall be transferred by means of Internet with a given datarate. Based on the above equations in the previous sections, the Controller if comes with the Internet as an efficient approach, then the data will be transmitted by means of Internet and the Source data center will be informed. Otherwise the calculated route will be sent to the Source data center, as well as to all the Data-spots included in the route and then the data will be transferred via vehicles.

To calculated Internet delay we have taken the average upload data rate in mbps for Islamabad region[33] and based on that the Controller calculates the route.

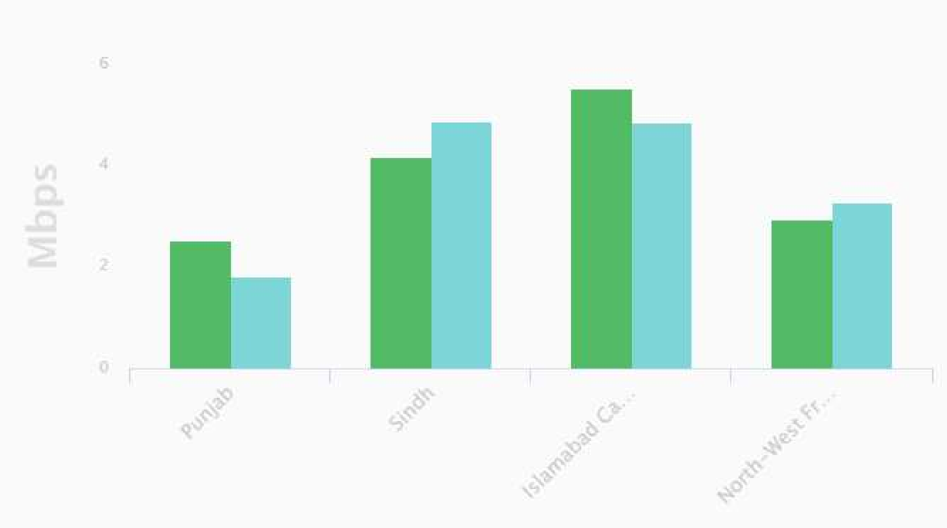


Figure 3.2: Average data-rate Results for Pakistan cities[33]

3.2.2. Dataspots:

Data spots shall be located at Petrol or Gas stations where the vehicles stop by. In smart cities these data spots can be located at Vehicle charging stations. Number of Data spots in our framework is than the number of data centers. Data spots are selected by the controller as intermediate stop points, where the vehicle can drop the data picked from the source. Similarly, the data spot, if has any data chunk on it, will signal the nearby vehicles. The vehicles shall pick the chunk from it and drop off to the next data spot ahead. In this way spots in our framework, give a helping hand to move the data hop to hop from source to spots, spots to spots and then from spots to destination.

Each data spot has a specific number of smaller stations where the vehicle will park and offload or load the data from the data spot.

Data spots, since are storage entities, hence they also have some storage capacities too. Due to limited storage capacities on data-spots, there are chances that the vehicle arrives to offload data to the data-spot but is full. In such a case, the vehicles are pre-informed.

3.2.3. Vehicles:

Vehicles are the data carriers in our framework. They can receive requests for carryung data either from the data centers or from the data spots. The vehicle upon receiving a data request has a choice whether to carry the data or to drop the request. If it aims to carry the data, then it shall move towards the entity which requested it. Vehicle shall then load the data from the datacenter/spot to its own storage device. This is done by means of USB 3.2 wire. Its data transfer rate is 20 Gbps [34], which helps in quick data transfer. Once the data is loaded on to the vehicle, the vehicle is then assigned dataspot where driver can offload the data. The data spot is assigned to the vehicle based on the destination located nearby. Vehicle again has a choice to offload the data or leave the task. Upon leaving the task, the data will be retransmitted. If the vehicle decides to offload then it moves towards the assigned data spot and offloads the chunk again by means of USB 3.2 wire. The process is same in case the data spot requests the vehicle to pickup data from it. In a deployed system in future, the vehicles can be paid to encourage larger number of participation.

3.2.4. Datacenters:

One of the data-centers in our system is source and the other is the destination. Any data-center can be a source or a destination. Whenever the source data-center finds a data to be sent in the queue of data, it requests the controller for the route. Upon receiving the route, source signals the nearby vehicles to pick up the data from it. Once the vehicle arrives, the sourcev checks the vehicle's storage capacity. Based on the vehicle's storage, the source picks up a chunk of data and loads to the vehicle. It creates a table consisting of all the metadata about each chunk and its status.

Every chunk is identified by a unique chunk id. This table of chunks is shared with the destination data center as well. If any of the chunk's status is found to be *"lost In transit"*

, then the receiver shall ask for retransmission of that specific chunk. Receiver periodically checks the status of all the chunks to maintain data integrity. The status *"lost In transit"* can be a result of either the vehicle loaded the chunk and did not offload it to a data spot, by choice. Or the vehicle intended to offload it but the dataspot had no more capacity.
A chunk that is to be retransmitted can be sub divided into sub-chunks, if the next vehicle arriving to pick it up has even smaller storage capacity.

3.3.Execution Of Framework:

Fig. 3.3 shows the flow of how each entity works in our framwork. As it is shown in Fig.3.3 that the execution is started by the Sender data center. Whenever the Sender data-center has a big-data that cannot be transmitted through other networks, then for confirmty of the transfer media, it asks the controller to calculate the route. If the data tto be transferred is efficient via Internet then the controller will inform the Sender. Otherwise, the controller will provide an optimal route from source to destination, consisting of intelligently chosen data spots in between where data can be offloaded. This route is also sent to the spots as well as the receiver to keep all aware of which entities are taking part. This approach helps in maintaining data integrity. As soon as the source gets the route it starts to signal for loading. The vehicle near by gets the request, decides to pick up data, sender loads the data chunk equal to the size of vehicle's data storage capacity, vehicle offloads the data to the next nearby assigned data spot and becomes free. Fig. 3.3 can be extended with the possibilities when the data is lost in transit. This loss might be a result of vehicle's choice of the spot gets full with its capacity.

Entities make a continuous check over different things:

- Sender checks whether there is some more data to be moved to the spots
- Dataspot checks whether there is any chunk left on it to be moved ahead
- Receiver checks whether all the chunks are in transit and none is lost. In case any of the chunk is lost, then the receiver request for retransmission of that chunk

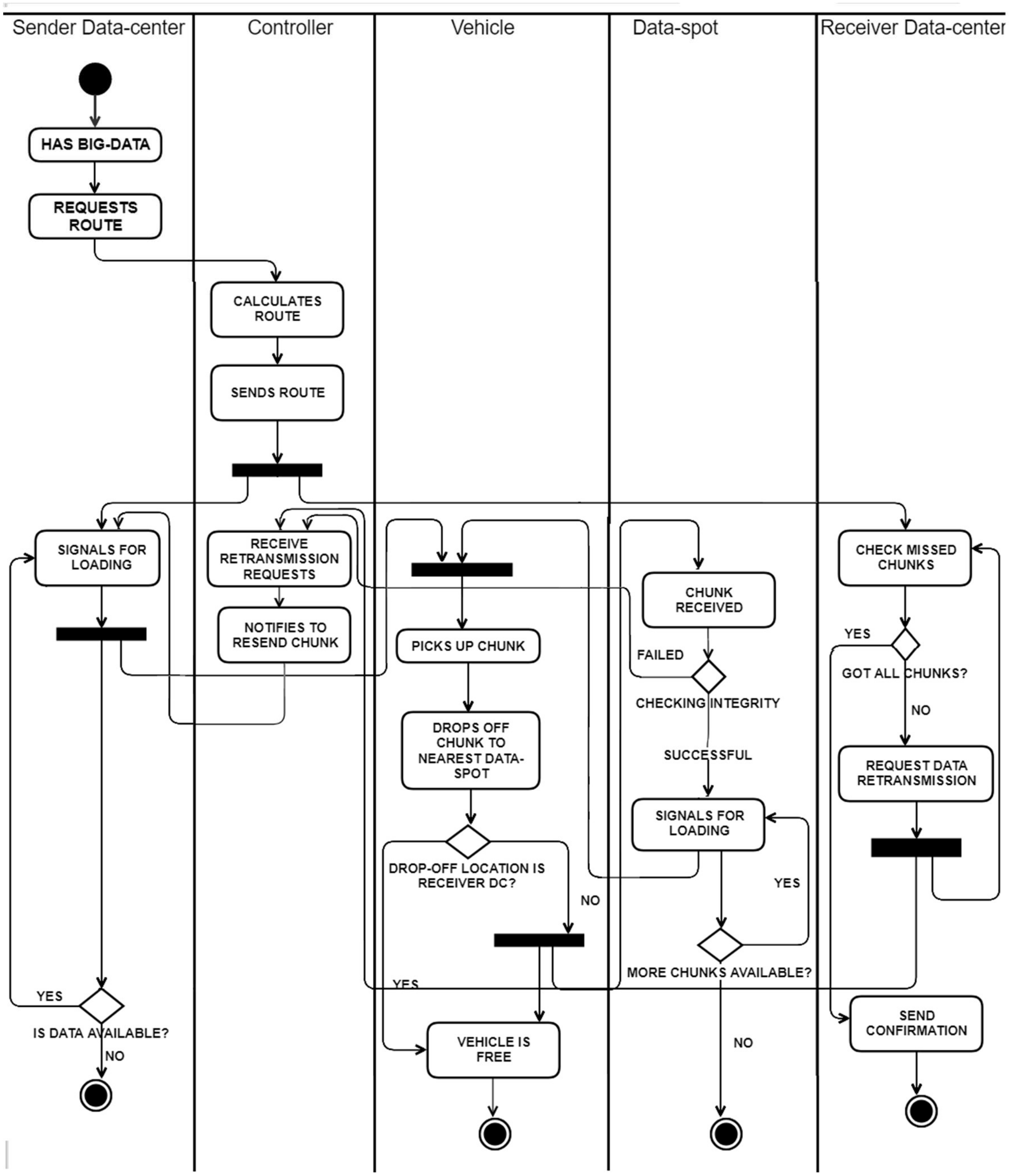


Figure 3.3: Activity Diagram of Our framework

3.4. Problem Formulation:

One of the important challenges to be noted in a smart city is Data-aggregation. Data-aggregation is a term which means that how chunks of data arriving from different locations can be combined together or brought under a single platform such that they can be accessed efficiently. This is done when there arrives a need of getting different variables of a specific type of data. It depends upon the need of the intended company or organization.

Table 3.1: Nomenclature

Symbols	Meaning
c	Controller
d_c	Data-center
d_s	Data-spot
v	Vehicle
J	Job
c_k	Chunks
$Dist_{SD}$	Distance between source S and destination D
Avg_{speed}	Average speed of vehicles
C_v	Capacity of each vehicle
N	Vehicle volume
ρ	Probability of participating vehicles
F_e	Fuel economy
W_{veh}	Average weight of vehicle
F_c	Fuel constant
E_{veh}	Total energy consumed by vehicles in data transfer
e_u	Energy consumed while unloading data
e_o	Energy consumed while offloading data
e_t	Energy consumed by vehicle
E_{trans}	Energy consumed by Internet data transfer
P_{st}	Power consumption during data storage

Different types of techniques and devices are used for this cause. Such as Sensors collecting geographical or weather data. These sensors are located at different places. The data these sensors collect is being aggregated at one point. This central location can be the organizations big data-center. These data centers use this aggregated data to provide an efficient output by the help of different analysis techniques.

Similarly, there are numerous data-centers with an explosion of data streams arriving per minute. So, the network due to big streams of data becomes measurably over-burden. The current network is slowly and gradually becoming unable to handle all of the data. Hence, there seems a need to approach some other data transfer mechanisms that shall help to minimize the burden from the data network.

Vehicle based data transfer is one of the possible data dissemination techniques. Efficiency of this technique depends upon a variety of factors.

3.4.1. Delay:

The delay while using this approach depends on factors which include the distance between source and destination geographically, delay in loading the data from the system to the vehicle, the delay in offloading the data from the vehicles to the system and the limit of the vehicle's speed in km/hr.

To calculate the delay, we have:

$$Delay = \sum_{i=1}^{n}(T_{vj} \times T_{dd} \times T_{dest} \times T_{du}) \quad (1)$$

Where T_{vj} is the travelling time a vehicle takes to reach the spot or the data-center i.e. $T_{vj} = distance/Avg_{speed}$, T_{dd} and T_{du} are the amount of time it takes to download and upload the data from or to the vehicle and T_{dest} is the amount of time it takes for a vehicle to reach the destination data-center i.e. $T_{dest} = distance/Avg_{speed}$.

Also $n = N \times \rho$, where N is the total vehicle volume and n is the number of vehicles participating with probability ρ.

Furthermore, if we consider C_v as the individual capacity of each vehicle, then the overall system's capacity can be calculated as:

$$S_c = \sum_{i=0}^{N} C_v \quad (2)$$

Where N is the total vehicle volume on road.

For all vehicles on road the total bandwidth can be computed as follows:

$$B_v = \sum_{v=0}^{N} C_v \times \rho \tag{3}$$

Where B_v is the amount of bandwidth per road for N vehicles, C_v is vehicle's storage capacity which can be fixed or may vary and ρ is the probability of vehicles that will participate effectively to fulfill the desired data transfer.

If all the available data is transferred to a single vehicle, then the for this worst-case scenario the gross data transfer can be computed as:

$$GD_{tran} = \frac{D_{vol}}{T_r} \tag{4}$$

Here D_{vol} is the total amount of big-data to be transferred and T_r is the data transfer rate. In our framework, we consider the usage of USB 3.2 which has a data transfer rate of 20Gbps so we consider T_r equal to it.

Also, for individual chunk we can compute chunk transfer time, as follows:

$$D_{tr} = \frac{D_{chk}}{T_r} \tag{5}$$

Where D_{chk} is the amount of chunk carried by individual vehicle.

$$Delay_{int} = \frac{D_{vol}}{B_{int}} \tag{6}$$

3.4.2. Energy:

The total amount of energy while vehicular data transfer also depends on certain factors. These factors include the amount of energy consumed while offloading the data from the vehicle to the spot or while loading from the spot to the vehicle. Also, energy consumed in transit affects the total energy.

There can be two cases in this regard.

Case-I:

Let the transportation cost vehicles is the framework's responsibility i.e. the vehicles are hired by the sender data center which runs our framework. Hence, it will be responsible for all the transportation costs of vehicles.

In this case the total energy can be expressed as follows:

$$E_{veh} = e_u + e_o + e_t \tag{7}$$

Where,

e_u is the energy consumption while uploading the data from the spot to the vehicle,

e_o is the energy consumption while uploading the data from the vehicle to the spot

and

e_t is the energy consumption in transit.

Case-II:

Let the transportation cost vehicles is the framework's responsibility i.e. the vehicles will not be hired by the sender data-center which runs our framework. Hence, it will not be responsible for any transportation costs.

In this case the total energy can be expressed as follows:

$$E_{veh} = e_u + e_o \tag{8}$$

Each of the e_u e_o e_t can be further formulated as below:

$$e_u = e_o = \sum_{i=0}^{p} e_{P2P_I} \tag{9}$$

Where $e_{P2P_I}$is the amount of energy consumption while upload or offload and p is the number of spots.

$$e_t = F_c \times \sum_{i=0}^{n} \frac{Dist_{SD}}{F_e} \tag{10}$$

Where F_c is the constant used to adjust the units into joule, $Dist_{SD}$ is the distance from S to D, where S means source and D means destination, F_e is the economy of fuel and lastly n is the number of vehicles.

3.4.3. Potential Objectives:

Followings are the potential objectives of our framework:

1) $min\{E_{ve}\ \}$

2) $min\{Delay\}$

Our aim is to minimize the total energy being consumed in data transfer and to utilize the capacity available within vehicles efficiently.

Chapter 4

Simulation Tool

Anylogic PLE:

We have developed our framework by using Anylogic Simulation Tool. There are multiple libaries in the pallete, out of which we have chosen GIS from Space-Markup library as shown in figure 4.1.

Our framework consists of six agents named Vehicles, Sender, Receiver, Dataspots, Controller and the big-data which is to be transmitted.

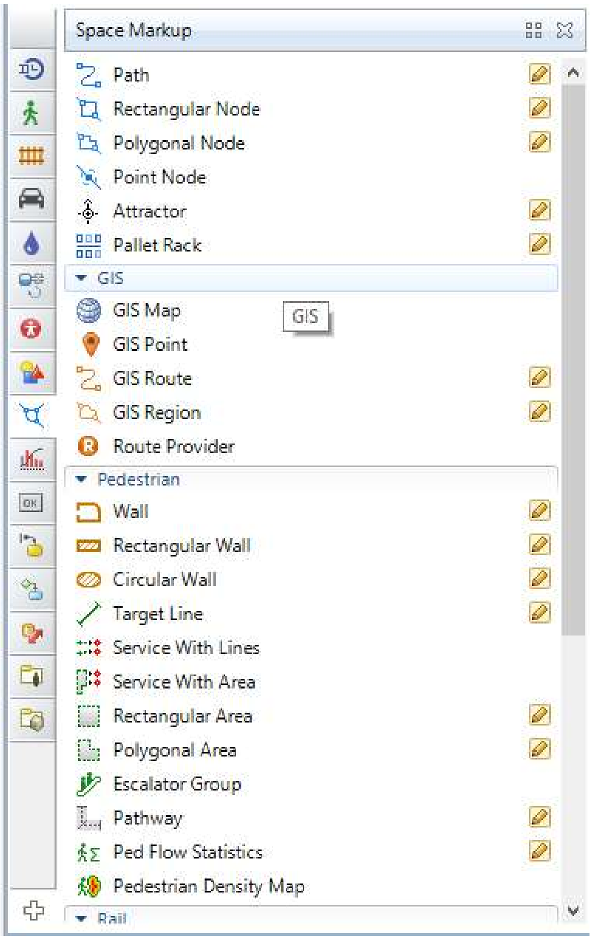
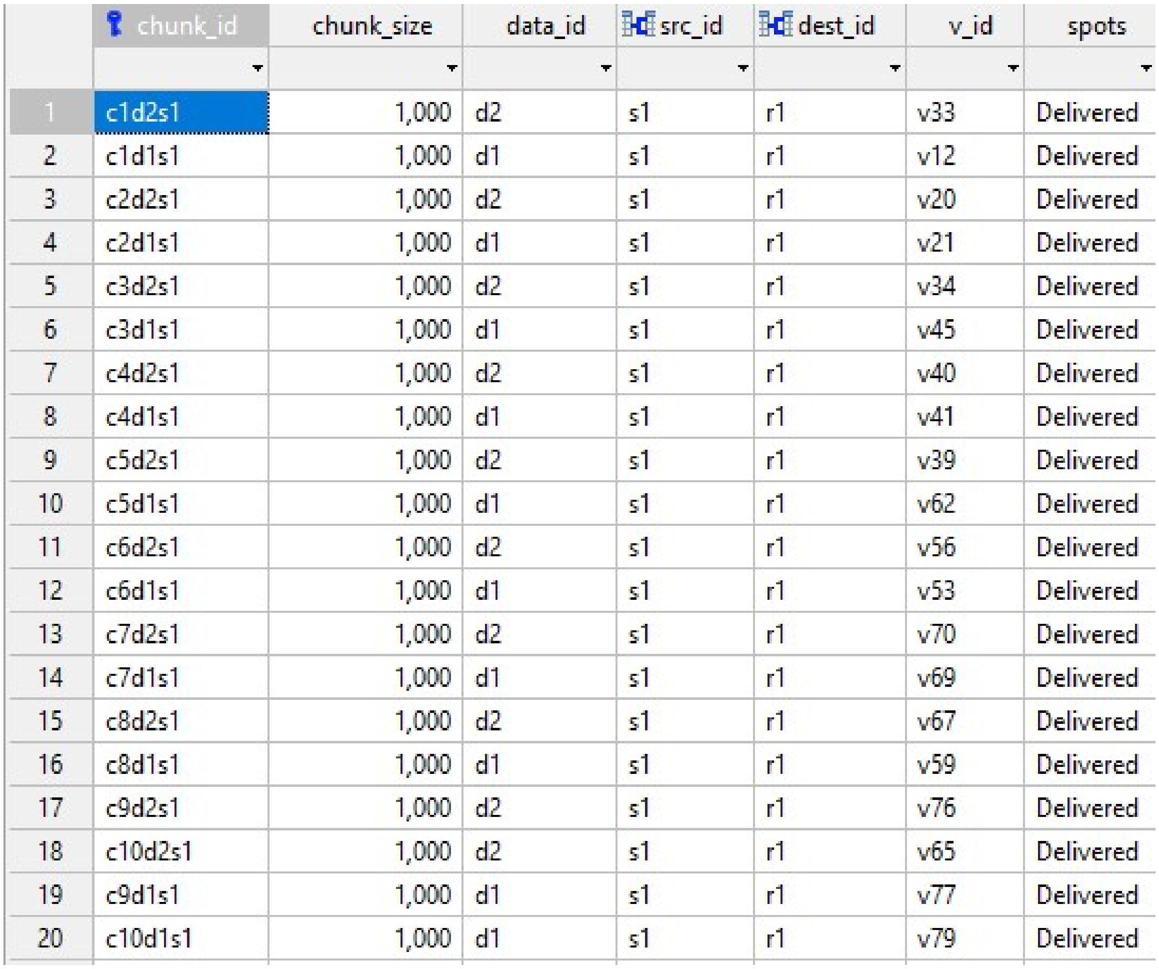


Figure 4.1: Space Markup Library used for our framework

The locations of source and destination datacenters, the intermediate data spots and the separate destinations i.e. home locations all have been fetched by Database tables. On of the most important table in a database in our framwork consists of is the table of Data-Chunks which is created and updated at runtime. Figure 4.2 shows the contents of database table after successful execution of our framework. Each chunk when created is assigned a unique chunk Id. This identification helps the chunks to be identified when they are lost so that only lost chunks are retransmitted, not the whole data. The column named ‘spots’ consists of the status of a chunk. If the chunk is at the spot currently then the row will contain spot id, if the chunk is lost in transit then it is mentioned as ‘Undelivered’ and if the chunk reaches its destination successfully then the status shows ‘Delivered’ as shown in the figure 4.2.



	chunk_id	chunk_size	data_id	src_id	dest_id	v_id	spots
1	c1d2s1	1,000	d2	s1	r1	v33	Delivered
2	c1d1s1	1,000	d1	s1	r1	v12	Delivered
3	c2d2s1	1,000	d2	s1	r1	v20	Delivered
4	c2d1s1	1,000	d1	s1	r1	v21	Delivered
5	c3d2s1	1,000	d2	s1	r1	v34	Delivered
6	c3d1s1	1,000	d1	s1	r1	v45	Delivered
7	c4d2s1	1,000	d2	s1	r1	v40	Delivered
8	c4d1s1	1,000	d1	s1	r1	v41	Delivered
9	c5d2s1	1,000	d2	s1	r1	v39	Delivered
10	c5d1s1	1,000	d1	s1	r1	v62	Delivered
11	c6d2s1	1,000	d2	s1	r1	v56	Delivered
12	c6d1s1	1,000	d1	s1	r1	v53	Delivered
13	c7d2s1	1,000	d2	s1	r1	v70	Delivered
14	c7d1s1	1,000	d1	s1	r1	v69	Delivered
15	c8d2s1	1,000	d2	s1	r1	v67	Delivered
16	c8d1s1	1,000	d1	s1	r1	v59	Delivered
17	c9d2s1	1,000	d2	s1	r1	v76	Delivered
18	c10d2s1	1,000	d2	s1	r1	v65	Delivered
19	c9d1s1	1,000	d1	s1	r1	v77	Delivered
20	c10d1s1	1,000	d1	s1	r1	v79	Delivered

Figure 4.2: Chunks Database table updated at runtime

Vehicle as an agent is the main entity of our framework. Figure 4.3 shows the statechart diagram of our agent vehicle. There are two statecharts of such kind. The first one is for those vehicles which take data from the sender or take the retransmitted data from the sender. The second statechart which is shown in figure 4.3 is the statechart diagram for those vehicles which take data from the data-spots. The vehicle firstly is signaled to pick up data if it is in the spots' range. After accepting request of the data spot the vehicle moves towards it. On reaching the respective data-spot, the vehicle's capacity is checked. If the vehicle has storage capacity less than the chunk present at the Dataspot, then the chunk will be further split into sub-chunks and loaded on to vehicle. Otherwise if the vehicle has storage capacity equal to or more than the chunk then the vehicle is simple loaded the whole data chunk. After data loading, there are few cases. Firstly, the vehicle moves towards the next data spot and successfully offloads the data and vehicle is free. Second case can be that the vehicle does not deliver the data by choice or by some personal problems. Third case can be that the vehicle aims to offload but the Dataspot cannot load more data onto it. The first case is a success, while the second and the third cases will lead to retransmission.

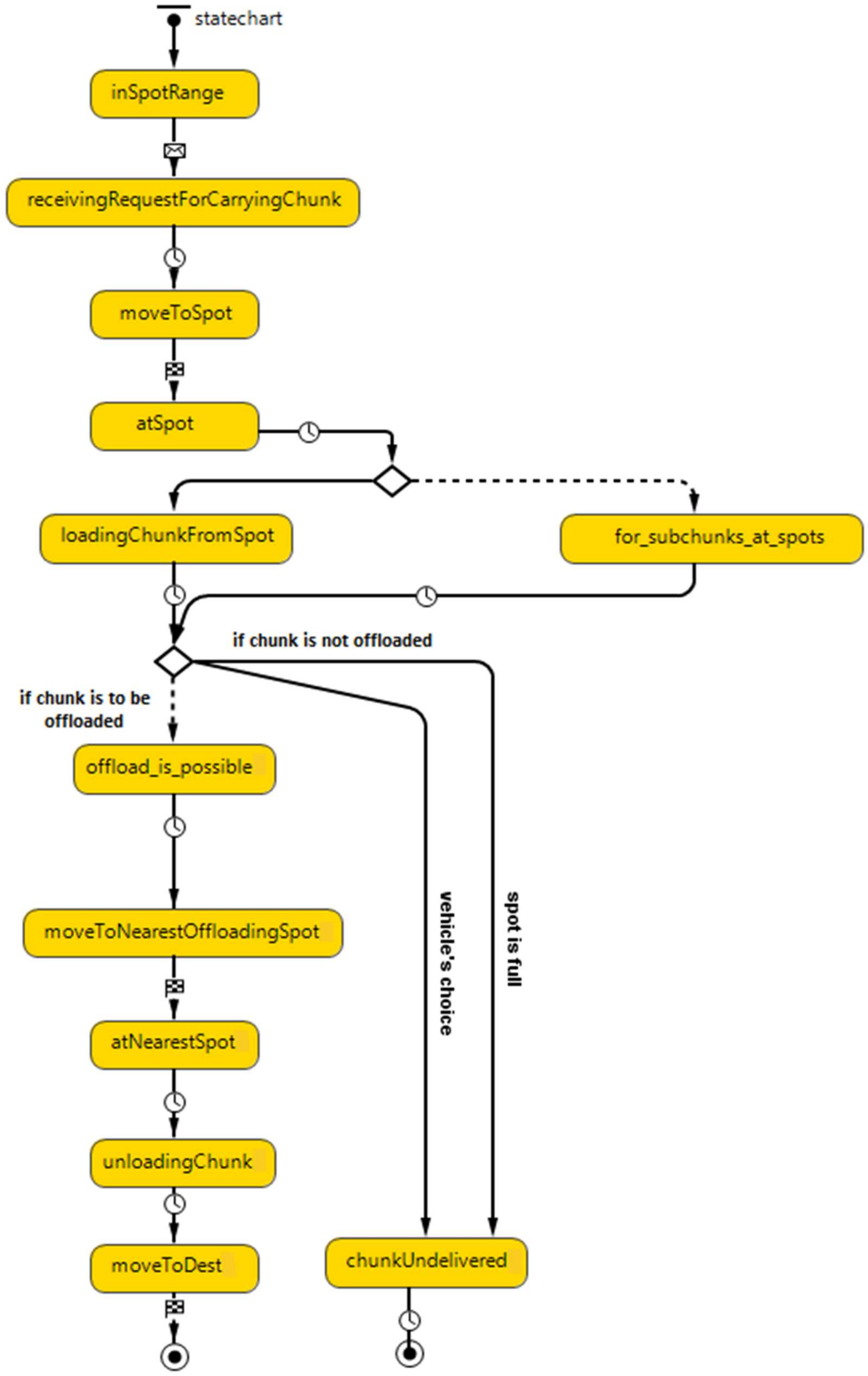


Figure 4.3: Statechart for vehicles as agents

Figure 4.4, shows a single simulation run of our framework. The green building at the north is Source data center or the Sender agent. The green building at the south is the destination data center or the receiver. The black spots represent Dataspots. Blue homes are the random destinations where the vehicle has to go. The yellow building represents central controller.

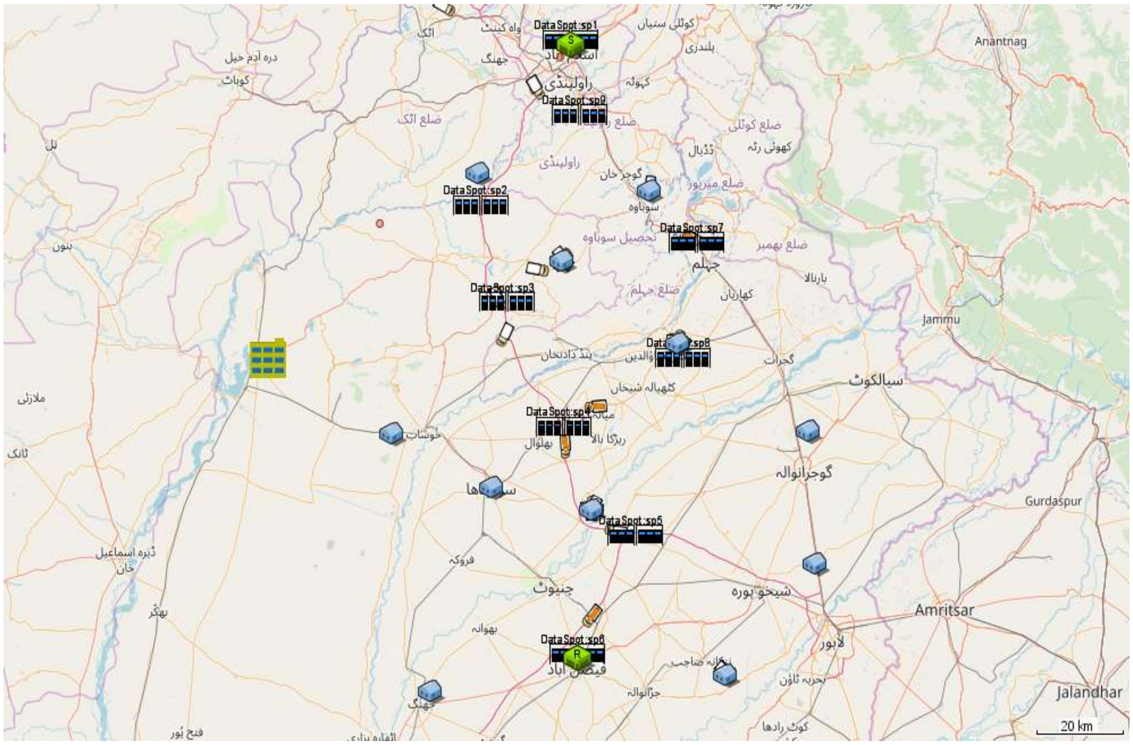


Figure 4.4: A single Simulation Run of our framework

Chapter 5

Evaluation and Results

We have considered parameters from Table 5.1 for computation and evaluation of results. Let there be one source data center, one destination data senter, nine data-spots in total, and two jobs $Job1$ and $Job2$. Each of the two jobs consists of 10 Terabytes of data to be moved from source to destination. This means that 20 Terabytes of data is expected to reach the same destination. Using data transfer by means of vehicles we have extracted the following results from our framework:

Table 5.1: List of Simulation parameters

Parameters	Values
Storage capacity per vehicle(GB)	Fixed /Variable
Total Vehicle density(N)	1000
ρ	0.65
Storage capacity per Data-spot	20,000 GB
Number of data-spots (p)	09
Data-transfer rate per spot	20 Gbps
Data block size (GB)	Fixed/Variable
D_{vol}	20,000 GB
Average Vehicle Speed	80 KM/h
$Dist_{SD}$	340 KM
Vehicle freuqency per KM	8
Storage Device used	Samsung SSD 1 terabyte
W_{veh}	990
F_e	17 KM/L
F_c	37624722.29 [35]

5.1.Fixed Vs. Dynamic Storage Capacity Of Vehicles:

As it is described in the previous section, that each vehicle has a specific storage capacity. This capacity is used to store the loaded data and offload it to another spot. In our first experiment we have evaluated results from two cases.

Case1: Fixed Capacity vehicles

The vehicles in this case have fixed storage values. Each vehicle had s storage capacity of 1000 GB or 1 TB.

Case2: Dynamic Capacity Vehicles

The vehicles having dynamic capacities ranging from 100 Gigabyte to one Terabyte.

5.1.1. Amount Of Utilization Of Data-Spots:

For two jobs $Job1$ and $Job2$ we have first calculated that how much each spot has been utilized for both the cases i.e. fixed and dynamic. Each spot has a specific capacity. This capacity is affected whenever any vehicle arrives for loading or offloading. When the route is selected by controller and set of spots are informed, then only those selected spots will be utilized for that respective process. In our case, for twenty terabytes of data the controller chose 5 spots. The utilization of each spot is shown below in Fig. 5.1 for fixed capacity case and Fig. 5.2 for variable capacity case:

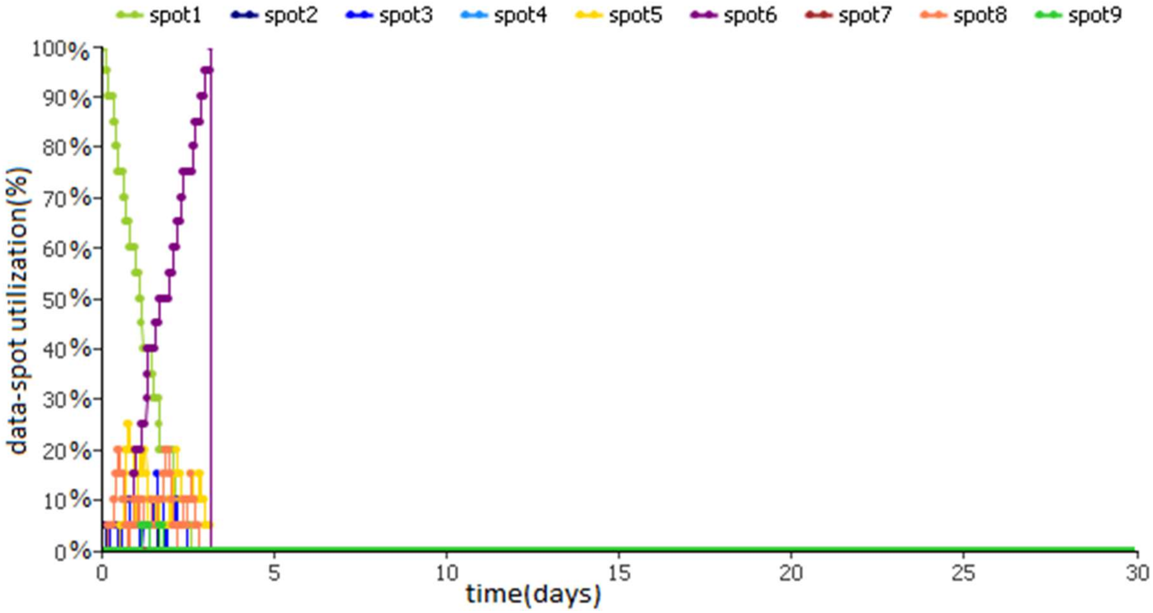


Figure 5.1: Amount of utilization of spots for Case1

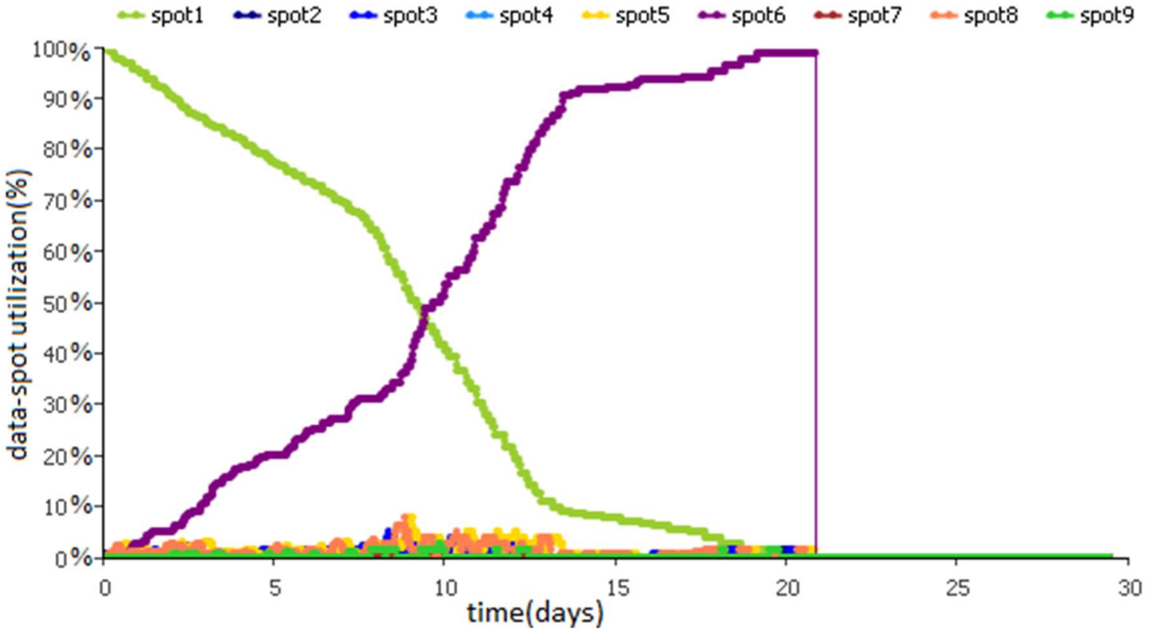


Figure 5.2: Amount of utilization of spots for Case2

It is quite clear in the above two figures, that the highest utilization of data-spots for Case1 rises up for 3 days, whereas for Case2 it rises up to 16 days. The reason is that for Case 1 in which vehicles have fixed storage capacities, there are less number of vehicles being used since they have fixed 1000GB capacity, hence they carry bigger chunks and finish the task earlier.

Whereas for the Case2 the vehicles have variable capacities, some vehicles will carry smaller chunks while others might carry bigger chunks. On the whole, there are more chances that more number of vehicles will have to participate. This ultimately imcreases the probability of vehicles leaving the task (as the probability of vehicles leaving the task is 0.65 which ultimately increases with the increase in number of vehicles).

Hence Case1 is fast and optimal for this evaluation.

5.1.2. Amount Of Utilization Of Vehicles:

Another evaluation for the two cases is performed in order to calculate the number of vehicles been used for this process. Some of the vehicles might leave the task in middle. But in this evaluation we have computed the overall vehicles which reponded to the source or spot's signals and participated to perform the task.

In the figure 5.3 it is shown that for case 1 of fixed capacity vehicles, we have less number of vehicles being used. The reason for this is the same as described in the previous evaluation i.e. the vehicles carry bigger chunks and hence less number of vehicles are enough to carry that much data. Whereas for case2 in fig 5.4, more number of vehicles were required for the whole 20 TB process.

Hence, it can be concluded that for the current evaluation more number of resources were being consumed for the dynamic case as compared to the fixed. So, fixed capacity vehicles seem optimal as a result of this evaluation.

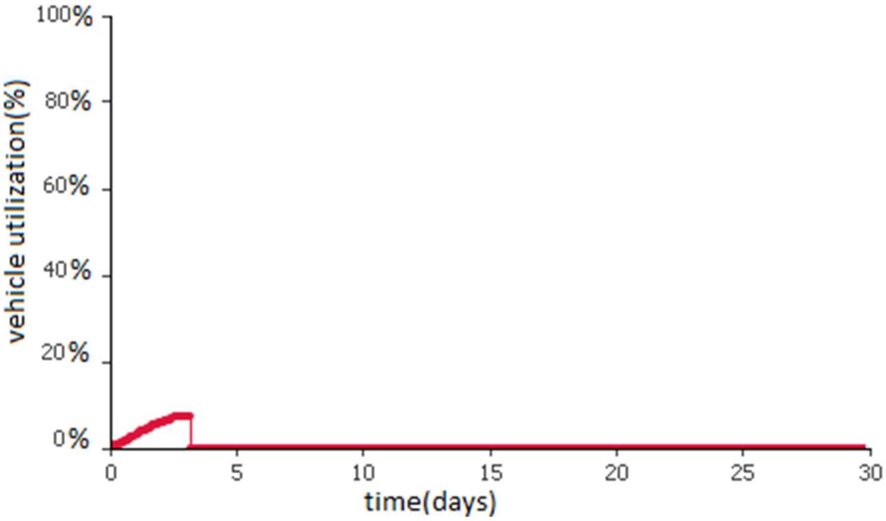


Figure 5.3: Amount of utilization of vehicles for Case1

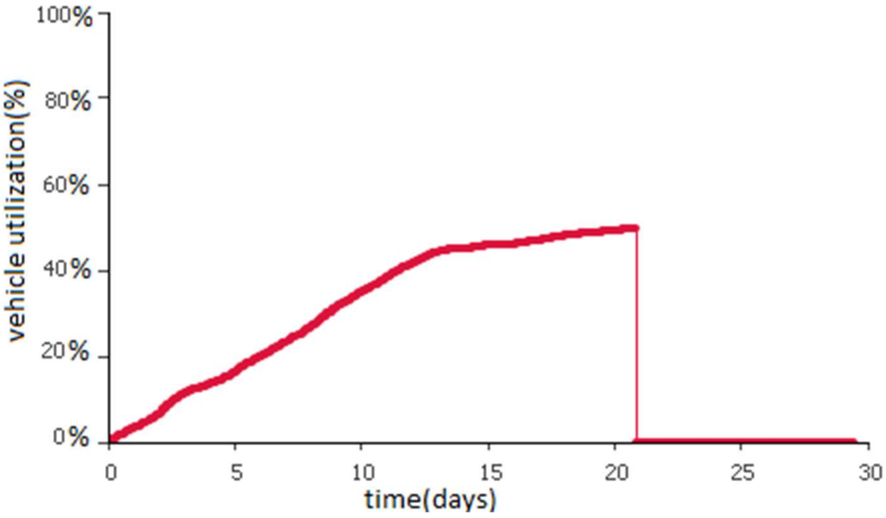


Figure 5.4: Amount of utilization of vehicles for Case2

5.1.3. Chunks Created Per Hour:

Furthermore, we have evaluated the number of chunks into which the data has been splitted. This split is caused based on the capacity of vehicles which carry chunks. If the vehicle has bigger storage capacity then bigger chunks will be picked and if the vehicle arriving for loading has smaller storage capacity then the chunks picked up will be smaller in size.

Also, it can be easily understood that with bigger chunks the data will be carried quickly with overall less number of chunks. On the other hand, if there are smaller chunks then the data will take time to deliver and large number of chunks will be created.

In figure 5.5, which is for fixed capacity vehicles, all the vehicles had fixed size hence no subchunks were to be formed. Secondly, the capacity was not too small i.e. 1000 GB. Hence less number of chunks were created, saving time and energy.

For the second case of dynamic vehicles, in figure 5.6, it is clearly shown that there are more number of chunks created due to dynamic capacity which might create subchunks. Secondly, the size is dynamic so there is a larger possibility that more number of smaller capacity vehicles might have arrived. This results in larger number of chunks.

Hence more cost and resource consumption. So, for this evaluation it can be concluded that case1 is more cost-efficient and energy-efficient as compared to case2.

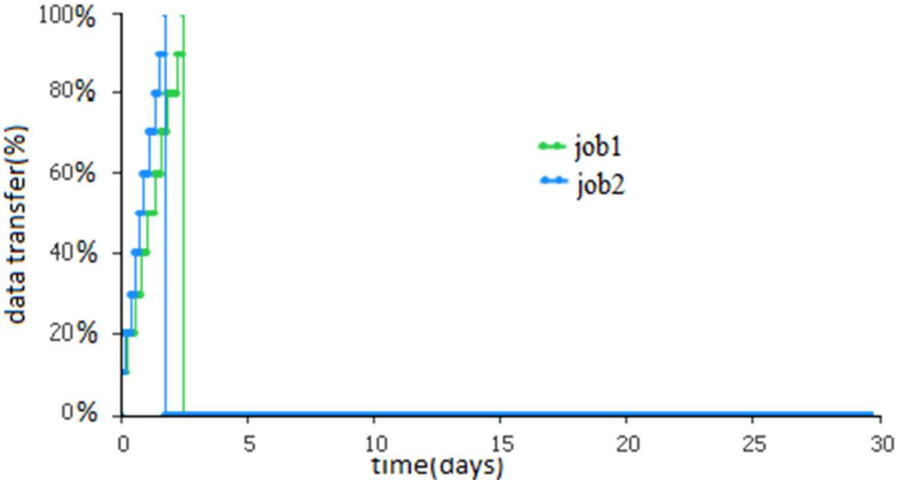


Figure 5.5:chunks per hour for Case1

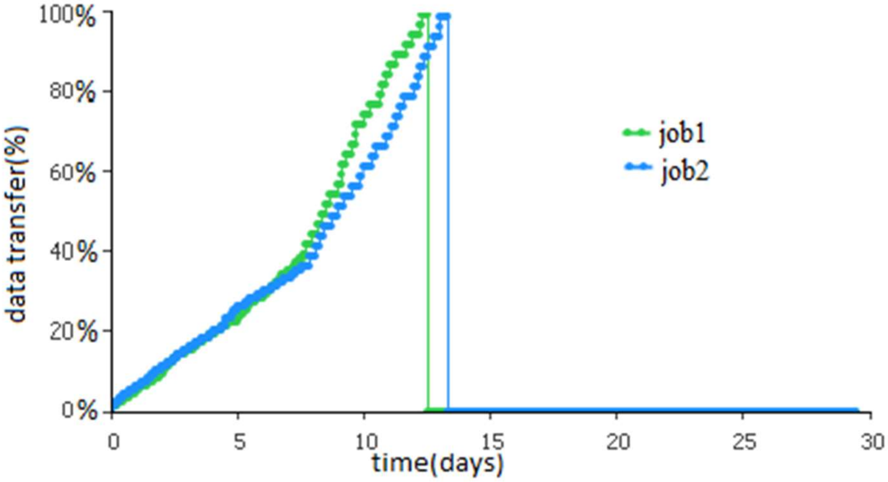


Figure 5.6: Chunks for hour for Case2

5.1.4. Number Of Quitted Jobs:

The probability of chunks undelivered while in transit can be due to one of the two reasons. The first is due to the possibility that the data spot where chunk had to be offloaded was full. The second reason can be that the vehicle did not choose to offload the data chunk by choice or due to any vehicle breakdown. In such cases, the probability of vehicles quitting the task seems to be directly proprtional to the number of vehicle participating in the job. For case1 of fixed capacity vehicles, as in the fig. 5.3 there were less number of vehicles participating, ultimately less number of chunks as shown in fig.5.5. Hence in fig. 5.7 there are less number of chunks undelivered for case1 which makes sense. Whereas for case2 of dynamic capacity vehicles, more number of vehicles participated and careted more number of chunks. Hence, with the increase in number of vehicles performing the task, the probability of vehicles quitting the assigned task also increased. Due to this fact more number of chunks went undelivered while in transit for case2 as compared to case1. Therefore, more retransmissions are expected for the dynamic vehicles case, which is again time and resource consuming, hence costly.

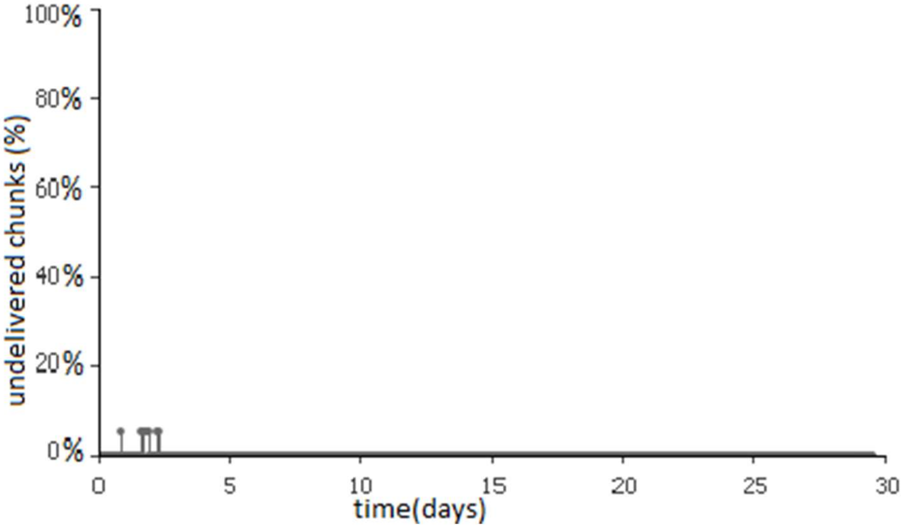


Figure 5.7: Percent chunks lost in transit for Case1

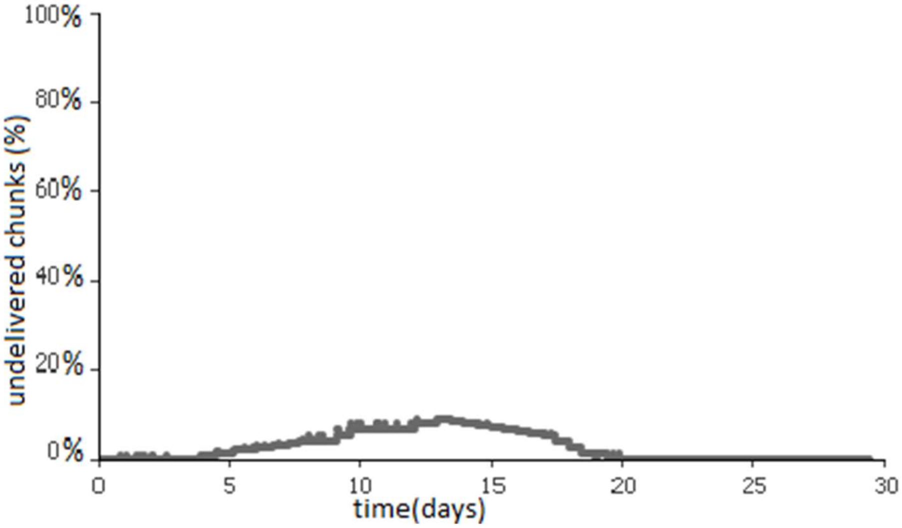


Figure 5.8: Percent chunks lost in transit for Case2

5.1.5. Task Distribution Chart:

The vehicles can be divided into three types:

- Vehicles which have been assigned a task and they have successfully completed it i.e. the vehicle was assigned a chunk. Vehicle loaded it and offloaded it to the the next assigned spot.

- Vehicles which were assigned a task and they did not complete it successfully i.e. a vehicle was assigned data chunk. Vehicle loaded it, but did not offload it to the assigned data spot, letting the chunk undelivered.
- Vehicles which were never assigned any task and did not participate

In figure 5.9, which is for case1, it shows the efficient utilization of vehicles as a resource. The overall cost shall be reduced due to less number of vehicles being utilized and even lesser number of unsuccessful vehicles.

Whereas, figure 5.10 shows that the utilization of vehicles was not efficient. More number of vehicles were called and more left in between the task, as compared to case1.

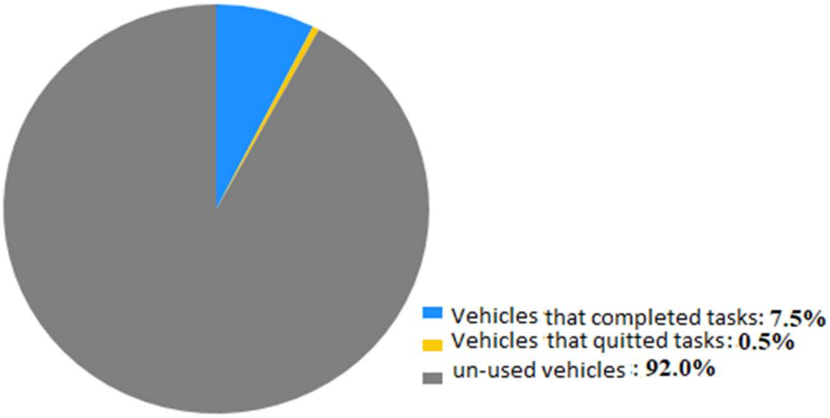


Figure 5.9: Task distribution chart for Case1

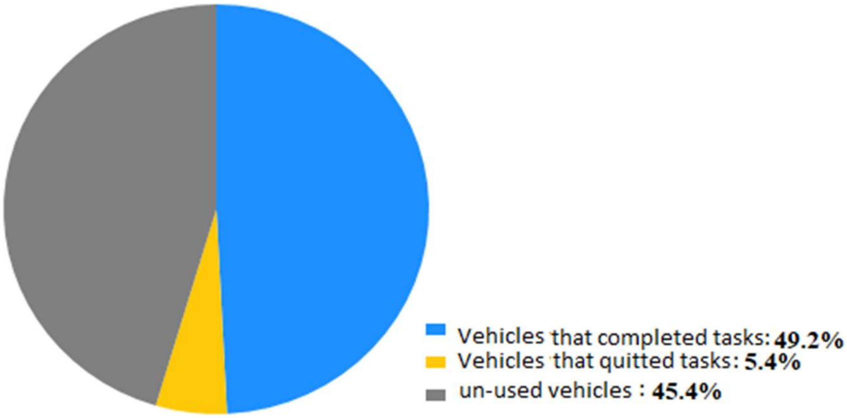


Figure 5.10: Task distribution chart for Case2

Out of 1000 vehicles 7.5% went successful, and 0.5% left the task in fig. 5.9. Whereas, in fig. 5.10, ot of 1000 total vehicles 49.2% were used and 5.4% left the task in between. Although the amount of data to be carried is same i.e. 20TB, but the difference in two approaches is huge.

5.1.6. Time Taken To Deliver Complete Data To Receiver:

Since for case1 of fixed capacity vehicles, there are less number of vehicles used and less number of chunks created. It can be said that due to faster chunk creation the chunks reached the destaination quickly. Also, subchunks and retransmission did not have a greater affect over the job completion hence the chunks reached the reciver within 3 days as shown in figure 5.11. On the other hand in figure 5.12, more number of chunks, subchunks and retransmissions caused more delay. Hence, for 20TB for case2 of dynamic vehicles, it took more than 16 days to deliver the two jobs to the receiver. Hence case1 is found more efficient tha case2 in this evaluation too.

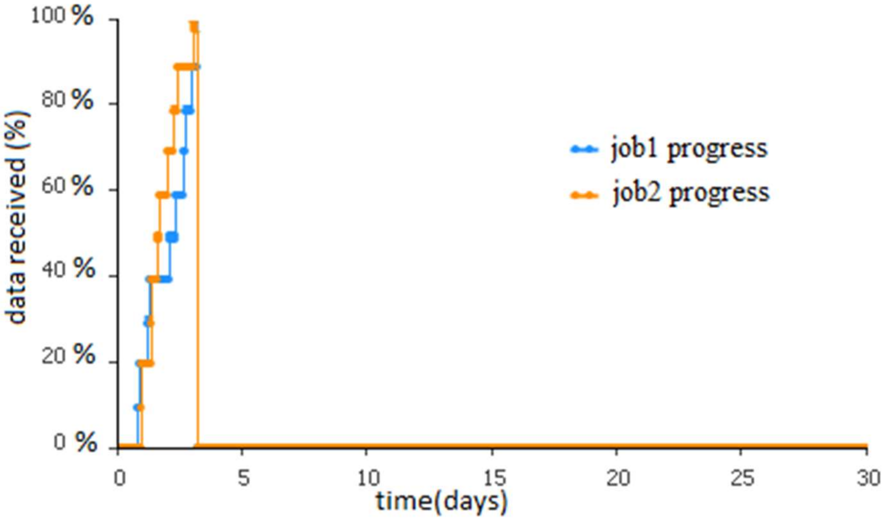


Figure 5.11: Amount of time for two jobs to complete for Case1

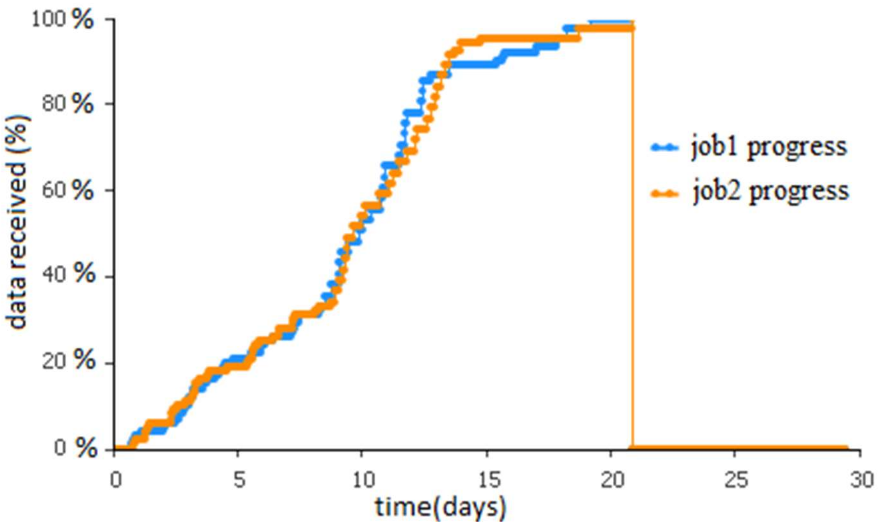


Figure 5.12: Amount of time for two jobs to complete for Case2

Looking through the above evaluations and results, one can observe that data transfer by means of vehicles is very efficient, but the amount of storage capacity of vehicles significantly affects the overall data transfer. Hence, before following this approach of data transfer, one must use better storage capacities to embed in vehicles, so that the output should not be better than Internet but be most appropriate version of itself.

5.2.Ideal Case:

In different researches such as Salman Naseer et al.[22] and other systems as described in Section-II (Literature review), we have observed that the scenario considered for data transfer by means of vehicles was ideal. No data losses were kept as constraints. It can be seen that in the real world there is definitely some probability that while the vehicle has carried the chunk, the vehicle faces some vehicle breakdown issue, the spot capacity might get full, any natural disaster might hit, etc. In such cases, the chunks which are moved might be lost while in transit. In previous works, the data transfer via shipment does not consider data loss aspect. Hence no retransmissions. In our framework, we have considered the retransmissions for worst case scenarios. Removing all the constraints we get ideal result in figure 5.13, as shown. But it can be seen, that our system works so well that even if we add constraints, the data for case1 reaches in the same time as for ideal case i.e. approx. 3 days.

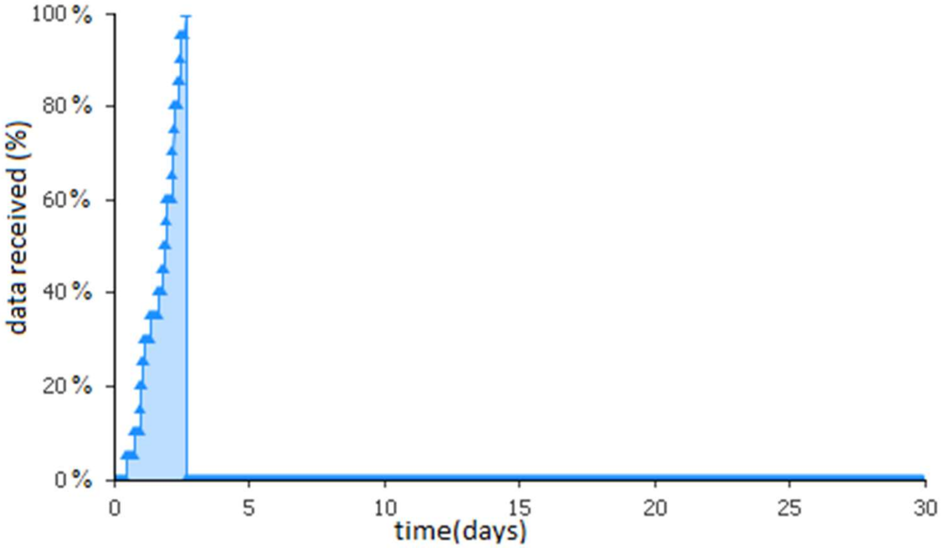


Figure 5.13: Ideal Scenario

5.3.Delay Cost Analysis

Internet Delay:

SpeedTest Results [36] provided measures for uploading data rates for main countries of every continent. We used its data to create a representation of average uploading delay for each continent. The measures are shown in Fig. 5.14. In this figure we have calculated the amount of delay for each continent when data of 20TB is to be uploaded. It is a very clear representation that it takes months to upload data over the cloud via Internet. The continent which faces maximum delay is Africa, which is about 38 months. The reason for such large delay in uploading big-data is that we are provided with a limited amount of bandwidth for uploading files. The data-rate for downloading an item is greater than the uploading rate. Due to such small bandwidth, it becomes very hard to upload big-data for an organization with little budget.

Hence, considering time as a unit of cost, Internet proves to be very costly.

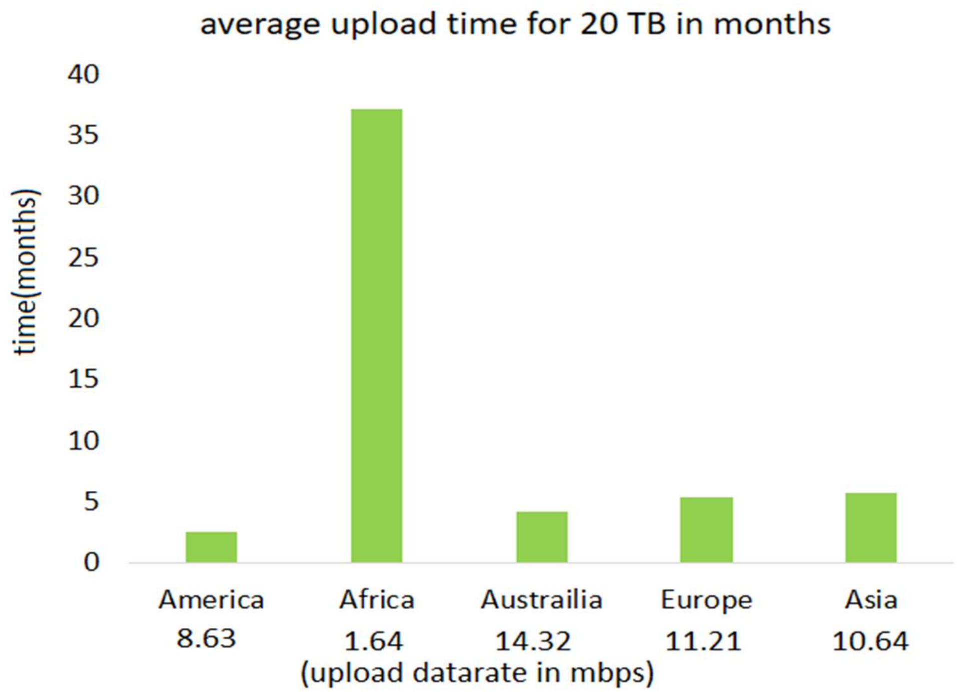


Figure 5.14: Average delay in data upload per continent

Vehicular Delay:

We have computed the amount of time it takes for the 20TB of data to transfer via vehicles. Further we analyzed the impact factor of delay while big-data transfer. Fig. 5.15 shows the number of data spots which arrive in between and the amount of time it takes to transfer 20TB for each case. It has been observed that avoiding retransmission and other delay factors such as travel distance, on can observe the relationship between the number of data spots and time taken to deliver the data. Figure 5.15 shows that with the increase in number of intermediate data-spots, there occurs increase in data transfer delay.

In our framework, the work of routing is done by the controller. It makes a decision when sender asks for route. The decision is either in the favor of data transfer by vehicles or by Internet. If the decision is in the favor of vehicle, then calculated route is sent to the source data center. The point of focus here is the Route. If the calculation of route aims for less number of data spots in between sender and receiver, then there are chances that less vehicles follow that route. Also, the spots will be distant from each other. Hence there come more chances that a vehicle who loads the data, gets to know the next spot to drop, but the next stop is so away that the vehicle quits the task. This means that by keeping less number of data spots there occurs more retransmissions and more delay.

Thus, one can observe that the number of data spots have a huge impact over delay cost of vehicle based data transfer.

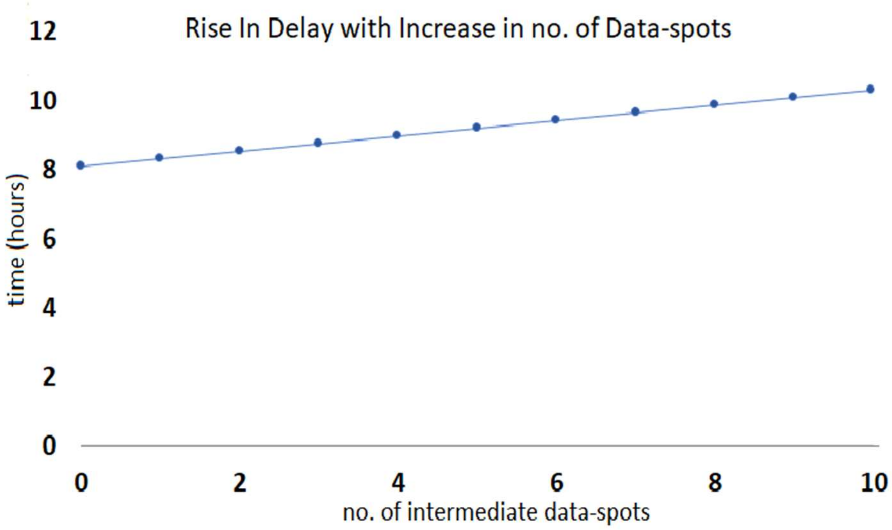


Figure 5.15: Relationship of delay and no. of data-spots

5.4.Energy Cost Analysis

Internet Energy Consumption:

Energy consumption over Internet is due to routers, switches, intermediate devices, the end-users, transportation and above all the data-centers running at the back-end [32].

Figure 5.16 represents the amount of energy consumed by three different energy consumption resources. The highest of all is the cost associated with the data-centers.

Cisco reported that by 2020, the data-usage shall reach about 278 Exabytes per month [37]. Based on the parameter values of core network devices used in Internet data transfer, David Costenaro et al. [32] estimated the energy consumed by the Internet in kWh per GB is 5.12

kWh/GB. Considering the above estimation to calculate data transfer of 20TB, 102.4MWh is the energy consumed by the Internet data transfer of 20TB.

Vehicular Energy Consumption:

On the other hand, energy consumed by the vehicular based data transfer approach, we consider average power consumed by a server equal to 145watt per hour. Using USB3.2 for vehicle to infrastructure communication, it takes 7minutes to upload or offload a 1TB chunk, hence, the average power consumed by a server for 7 minutes comes out to be 17watt-min approximately. For each upload or offload of 17 watt of power, using eq.9 we get cost upload or offload equal to 571.2kWh. Furthermore, using eq. 10, the transportation cost of a vehicle comes out to be 16723kWh approximately. Therefore, the total vehicular energy is equal to 17.864MWh.

To compute the cost in dollars, consider the average electricity cost standardized by U.S. Energy Information Administration [39] and that is $0.112 US per Kilowatt-hour (kWh); therefore, based on the energy consumption the cost of system energy for 20 TB, is estimated as $11469 approximately, whereas for vehicular approach the cost in dollars comes out to be $2000 approximately.

Comparing both the costs of energy as well as of delay, we conclude that for both the delay and the energy, Internet costs more as compared to the vehicle based data dissemination. Hence our vehicle–based data dissemination approach excels in terms of saving both delay as well as energy costs.

Chapter 6

Conclusion, Limitations and Future Work

Big-data is a digital material which is not only difficult to handle in terms of storage but also in terms of maintenance, processing and above all transfer. The big-data transfer techniques are many such as wired data transfer, wireless data transfer, optical-fiber communication, Bluetooth, Wi-Fi etc. The problem with all these and similar approaches is either the delay or the energy cost. Sometimes even the delay and energy do no cost more but the prices are high. With the exponential increase in data and increasing rate of data-explosion, we are in a need to create new data transfer techniques.

One of the new techniques is used by our framework in which the data is transfer by means of physical shipping. There are set of spots between the source and the destination. The vehicle picks up the chunk of data from the source and drops it to the nearest spot. Then, there are some vehicles which shall pick the data chunk from the spot and deliver it to the destination. In this way hop by hop the data is transferred successfully.

We have also considered the issues in case of data losses. The data loss might result in a situation when the spot is not able to offload more data onto it or the vehicle goes away without offloading the data to the next spot by choice. For such circumstances we have considered the retransmission technique in our framework. We have analyzed that how the whole two jobs, each of 10 Terabyte reach the destination successfully along with the retransmissions.

First we compared two approaches for the vehicles case. We had first assumed that all the participating vehicles will have dynamic storage capacities. Then we assumed that all the vehicles have static 1000 GB storage capacity, based on the number of chunks created, data transfer delay, delay due to retransmissions and losses and the spots' usage capacities we came out with the conclusion that our framework results best if all the vehicles have static 1000GB capacity.

Then we compared two data transfer techniques i.e. the Internet and the vehicle approach. Firstly, we used delay parameter for comparison. Results showed that the time it takes to transfer a 20 terabyte complete job by Internet takes months to upload, while it takes three days to complete the whole 20 terabyte job to reach its destination successfully.

Secondly, we analyzed energy as a parameter of comparing two approaches. Our experimental results prove that the energy consumption for the vehicle case is about 17.864MWh where as for the Internet case it comes out to be 102.4MWh. The clear difference between the energy and the delay shows that data transfer via vehicles is quite efficient as compared to the Internet.

Our work can be extended and can be further optimized by comparing different routing algorithms based on the density of the road traffic. Sensors can be deployed to recognize densely traffic areas and based on that the controller shall include nearby Dataspots so that more vehicles carry the data and quickly the data is transferred.

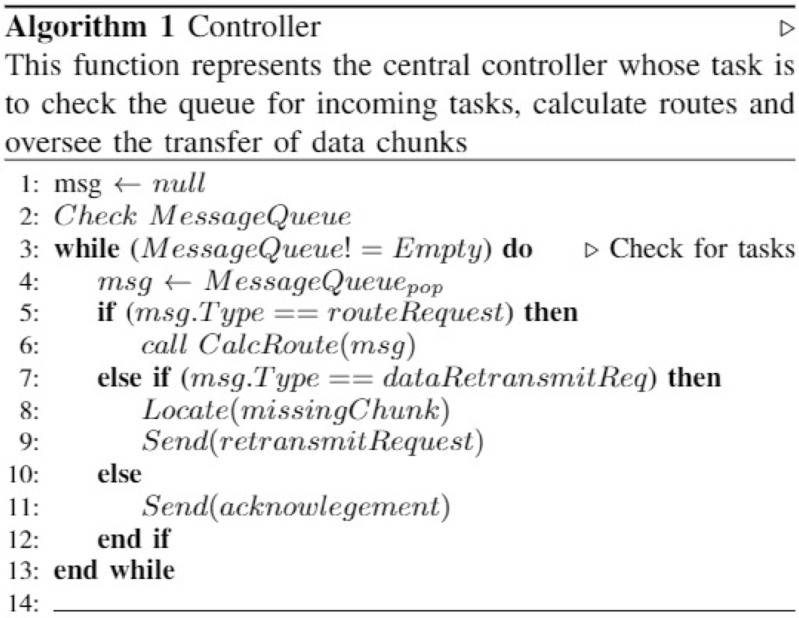
The system can be decentralized to minimize the burden of overall tasks performed by the controller. The data spots can communicate with each other and form a route based on traffic and distance. Retransmissions can be reduced by either increasing data-spots capacity or using some other strategy such as re-assignment of another spot if the assigned spot is full.

Analysis of payment costs can be performed based on the credits used by the participating vehicle which performs the task successfully and any case of misused credits by the participants which quit the task in the middle.

Appendix

Pseudocodes

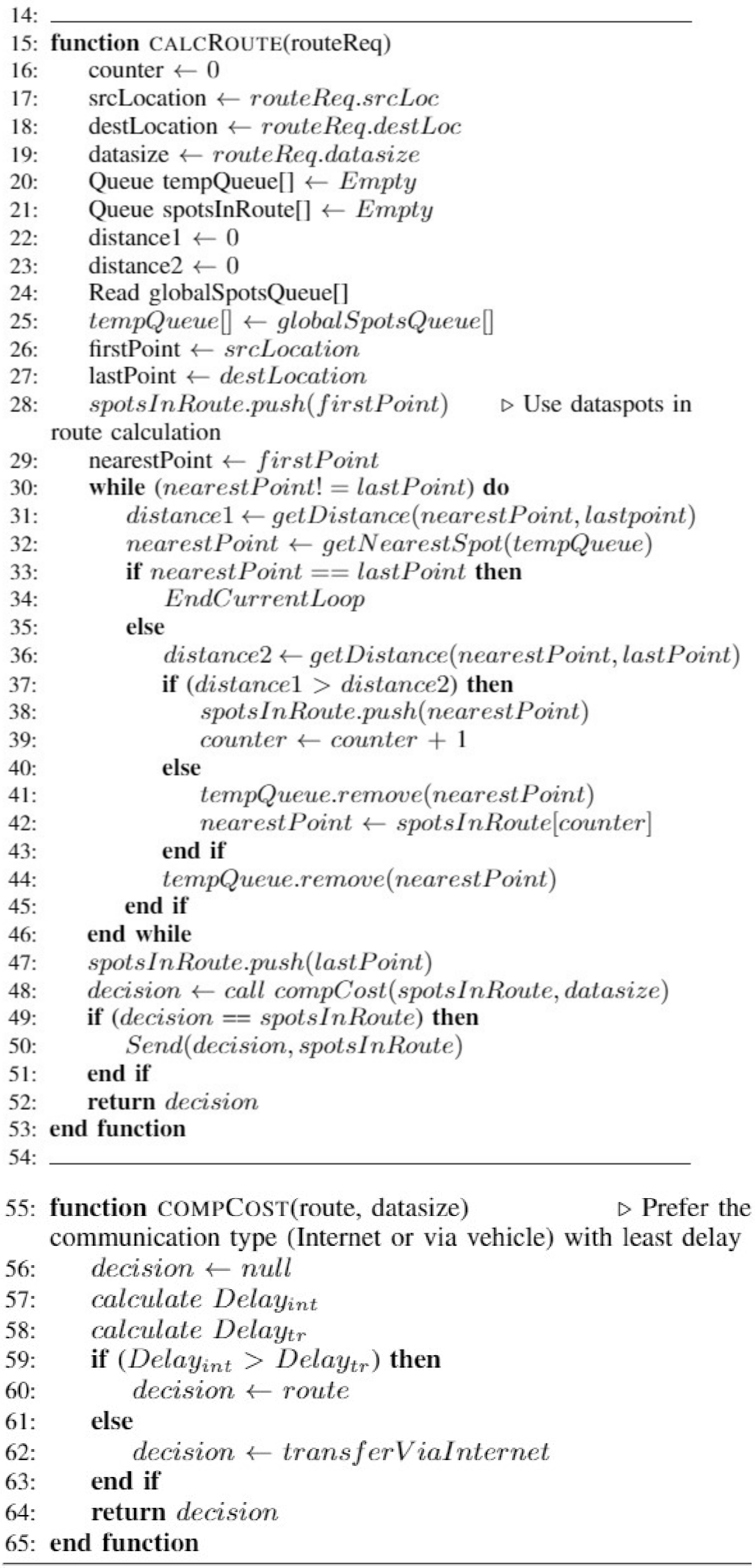
Controller:



Algorithm 1 Controller ▷

This function represents the central controller whose task is to check the queue for incoming tasks, calculate routes and oversee the transfer of data chunks

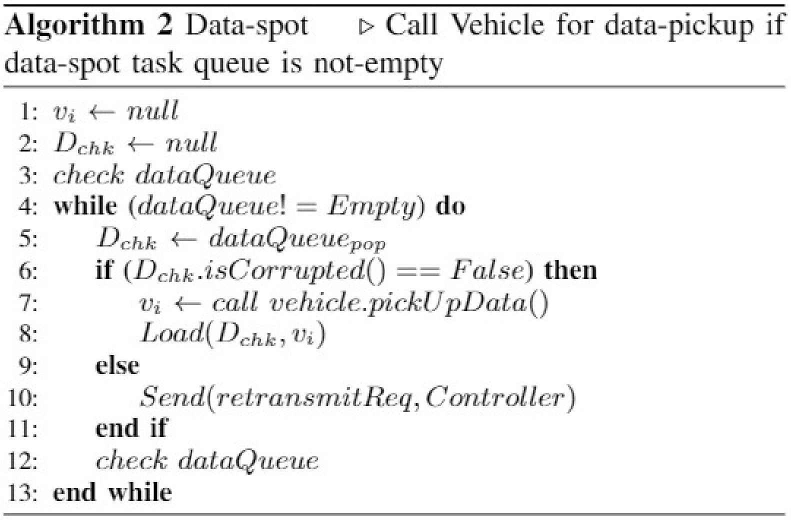
```
msg ← $null$
$Check\ MessageQueue$
while ($MessageQueue! = Empty$) do        ▷ Check for tasks
    $msg \leftarrow MessageQueue_{pop}$
    if ($msg.Type == routeRequest$) then
        $call\ CalcRoute(msg)$
    else if ($msg.Type == dataRetransmitReq$) then
        $Locate(missingChunk)$
        $Send(retransmitRequest)$
    else
        $Send(acknowlegement)$
    end if
end while
```



```
function CALCROUTE(routeReq)
    counter ← 0
    srcLocation ← $routeReq.srcLoc$
    destLocation ← $routeReq.destLoc$
    datasize ← $routeReq.datasize$
    Queue tempQueue[] ← $Empty$
    Queue spotsInRoute[] ← $Empty$
    distance1 ← 0
    distance2 ← 0
    Read globalSpotsQueue[]
    $tempQueue[] \leftarrow globalSpotsQueue[]$
    firstPoint ← $srcLocation$
    lastPoint ← $destLocation$
    $spotsInRoute.push(firstPoint)$        ▷ Use dataspots in route calculation
    nearestPoint ← $firstPoint$
    while $(nearestPoint! = lastPoint)$ do
        $distance1 \leftarrow getDistance(nearestPoint, lastpoint)$
        $nearestPoint \leftarrow getNearestSpot(tempQueue)$
        if $nearestPoint == lastPoint$ then
            $EndCurrentLoop$
        else
            $distance2 \leftarrow getDistance(nearestPoint, lastPoint)$
            if $(distance1 > distance2)$ then
                $spotsInRoute.push(nearestPoint)$
                $counter \leftarrow counter + 1$
            else
                $tempQueue.remove(nearestPoint)$
                $nearestPoint \leftarrow spotsInRoute[counter]$
            end if
            $tempQueue.remove(nearestPoint)$
        end if
    end while
    $spotsInRoute.push(lastPoint)$
    $decision \leftarrow call\ compCost(spotsInRoute, datasize)$
    if $(decision == spotsInRoute)$ then
        $Send(decision, spotsInRoute)$
    end if
    return $decision$
end function

function COMPCOST(route, datasize)        ▷ Prefer the communication type (Internet or via vehicle) with least delay
    $decision \leftarrow null$
    $calculate\ Delay_{int}$
    $calculate\ Delay_{tr}$
    if $(Delay_{int} > Delay_{tr})$ then
        $decision \leftarrow route$
    else
        $decision \leftarrow transferViaInternet$
    end if
    return $decision$
end function
```

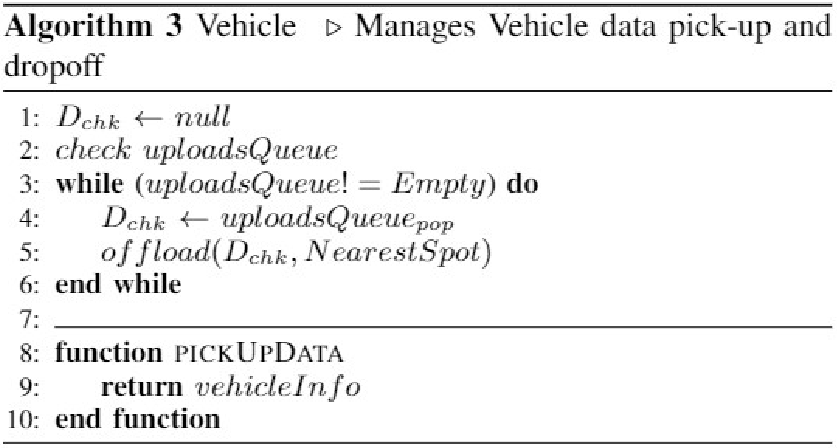
Data-spot:



Algorithm 2 Data-spot ▷ Call Vehicle for data-pickup if data-spot task queue is not-empty

```
$v_i \leftarrow null$
$D_{chk} \leftarrow null$
$check\ dataQueue$
while ($dataQueue! = Empty$) do
    $D_{chk} \leftarrow dataQueue_{pop}$
    if ($D_{chk}.isCorrupted() == False$) then
        $v_i \leftarrow call\ vehicle.pickUpData()$
        $Load(D_{chk}, v_i)$
    else
        $Send(retransmitReq, Controller)$
    end if
    $check\ dataQueue$
end while
```

Vehicle:

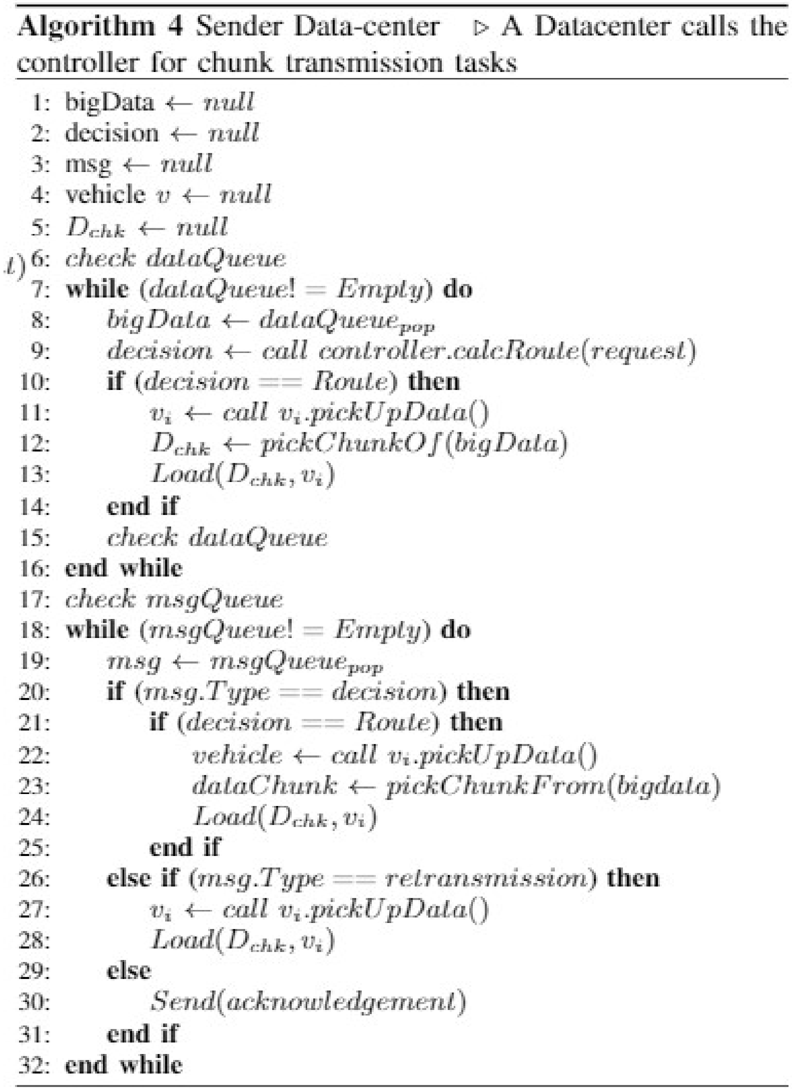


Algorithm 3 Vehicle ▷ Manages Vehicle data pick-up and dropoff

```
$D_{chk} \leftarrow null$
$check\ uploadsQueue$
while ($uploadsQueue! = Empty$) do
    $D_{chk} \leftarrow uploadsQueue_{pop}$
    $offload(D_{chk}, NearestSpot)$
end while

function PICKUPDATA
    return $vehicleInfo$
end function
```

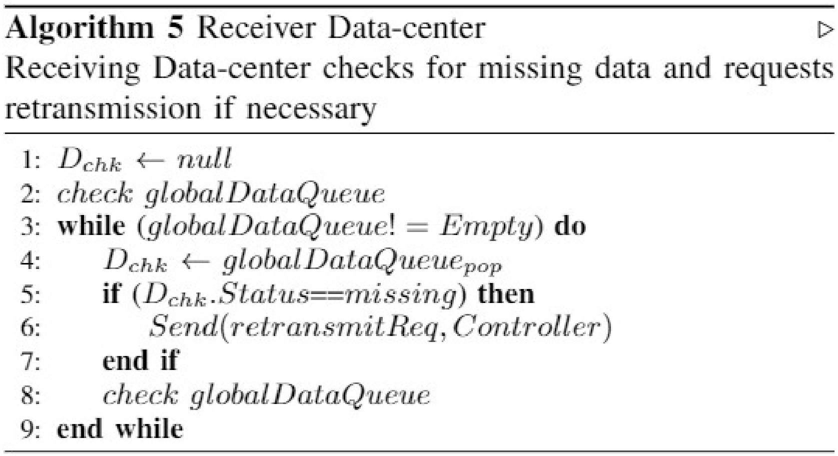
Sender data-center:



Algorithm 4 Sender Data-center ▷ A Datacenter calls the controller for chunk transmission tasks

```
bigData ← $null$
decision ← $null$
msg ← $null$
vehicle $v$ ← $null$
$D_{chk}$ ← $null$
$check\ dataQueue$
while ($dataQueue! = Empty$) do
    $bigData \leftarrow dataQueue_{pop}$
    $decision \leftarrow call\ controller.calcRoute(request)$
    if ($decision == Route$) then
        $v_i \leftarrow call\ v_i.pickUpData()$
        $D_{chk} \leftarrow pickChunkOf(bigData)$
        $Load(D_{chk}, v_i)$
    end if
    $check\ dataQueue$
end while
$check\ msgQueue$
while ($msgQueue! = Empty$) do
    $msg \leftarrow msgQueue_{pop}$
    if ($msg.Type == decision$) then
        if ($decision == Route$) then
            $vehicle \leftarrow call\ v_i.pickUpData()$
            $dataChunk \leftarrow pickChunkFrom(bigdata)$
            $Load(D_{chk}, v_i)$
        end if
    else if ($msg.Type == retransmission$) then
        $v_i \leftarrow call\ v_i.pickUpData()$
        $Load(D_{chk}, v_i)$
    else
        $Send(acknowledgement)$
    end if
end while
```

Receiver data-center:



Algorithm 5 Receiver Data-center ▷ Receiving Data-center checks for missing data and requests retransmission if necessary

```
$D_{chk} \leftarrow null$
$check\ globalDataQueue$
while ($globalDataQueue! = Empty$) do
    $D_{chk} \leftarrow globalDataQueue_{pop}$
    if ($D_{chk}.Status{==}missing$) then
        $Send(retransmitReq, Controller)$
    end if
    $check\ globalDataQueue$
end while
```

References

[1] Big-data Insights, what is Big-data? https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

[2] Structured Semi-structured and Unstructured data. https://jeremyronk.wordpress.com/2014/09/01/structured-semi-structured-and-unstructured-data/

[3] Ten V's of Big-data. https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

[4] Hard Disk Drive. https://searchstorage.techtarget.com/definition/hard-disk-drive

[5] Solid State Drive. https://techterms.com/definition/ssd

[6] What is the difference between SSD and SSHD? https://www.quora.com/What-is-the-difference-between-SSD-and-SSHD

[7] What is Cloud Storage? https://www.techopedia.com/definition/26535/cloud-storage

[8] Types of Cloud Computing Explained GlobalDots. https://www.globaldots.com/cloud-computing-types-of-cloud/

[9] KoiBriefing Cloud Computing and Africa. http://blog.koistrategy.com/2011/01/koibriefing-cloud-computing.html

[10] Abraham Y. What happens in an internet minute? how to capitalize on the big data explosion? https://www.excelacom.com/resources/blog/what-happens-in-aninternet-minute-how-to-capitalize-on-the-big-data-explosion , 2015.

[11] Kelly L. What happens in an internet minute? how to capitalize on the big data explosion? https://www.excelacom.com/resources/blog/whathappens-in-an-internet-minute-how-to-capitalize-on-the-big-dataexplosion , 2016.

[12] What is AWS? https://aws.amazon.com/what-is-aws/

[13] Aws snowball https://aws.amazon.com/snowball/ .

[14] Aws snowball edge https://aws.amazon.com/snowball-edge/ .

[15] Aws snowmobile https://aws.amazon.com/snowmobile/.

[16] Delay Tolerant Networking https://en.wikipedia.org/wiki/Delay-tolerant_networking

[17] What is Point Of Presence (POP)? https://www.techopedia.com/definition/1704/point-of-presence-pop

[18] What is Smart City? https://internetofthingsagenda.techtarget.com/definition/smart-city

[19] Rouse Margaret, What is IoT? https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT

[20] What is green cloud? https://searchstorage.techtarget.com/definition/green-cloud

[21] Sheikh Mohammad Idrees, M Afshar Alam, and Parul Agarwal. A study of big data and its challenges. International Journal of Information Technology, pages 1–6, 2018.

[22] Salman Naseer, William Liu, Nurul I Sarkar, Peter Han Joo Chong, Edmund Lai, and Rangarao Venkatesha Prasad. A sustainable vehicular based energy efficient data dissemination approach. In Telecommunication Networks and Applications Conference (ITNAC), 2017 27th International, pages 1–8. IEEE, 2017.

[23] Ivana Marincic and Ian Foster. Energy-efficient data transfer: Bits vs. atoms. In Software, Telecommunications and Computer Networks (SoftCOM), 2016 24th International Conference on, pages 1–6. IEEE, 2016.

[24] Digital globe company supported by amazon web services https://aws.amazon.com/solutions/case-studies/digitalglobe/.

[25] The digital globe company https://www.digitalglobe.com/.

[26] B. Baron, P. Spathis, H. Rivano, M. D. de Amorim, Y. Viniotis, and M. H. Ammar. Centrally controlled mass data offloading using vehicular traffic. IEEE Transactions on Network and Service Management, 14(2):401–415, June 2017.

[27] Musefiu Aderinola Jafaru Ibrahim, Tonga Agadi Danladi. Comparative analysis between wired and wireless technologies in communications: A review. In 99th The IIER International Conference, Mecca, Saudi Arabia, pages 45–48, March 2017.

[28] Onur Altintas, Falko Dressler, Florian Hagenauer, Makiko Matsumoto, Miguel Sepulcre, and Christoph Sommery. Making cars a main ict resource in smart cities. In Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on, pages 582–587. IEEE, 2015.

[29] Bo Li, Yijian Pei, Hao Wu, Zhi Liu, and Haixia Liu. Computation offloading management for vehicular ad hoc cloud. In International Conference on Algorithms and Architectures for Parallel Processing, pages 728–739. Springer, 2014.

[30] Falko Dressler, Philipp Handle, and Christoph Sommer. Towards a vehicular cloud-using parked vehicles as a temporary network and storage infrastructure. In Proceedings of the 2014 ACM international workshop on Wireless and mobile technologies for smart cities, pages 11–18. ACM, 2014.

[31] David A Howey, RF Martinez-Botas, B Cussons, and L Lytton. Comparative measurements of the energy consumption of 51 electric, hybrid and internal combustion engine vehicles. Transportation Research Part D: Transport and Environment, 16(6):459–464, 2011.

[32] David Costenaro and Anthony Duer. The megawatts behind your megabytes: Going from data-center to desktop. Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, ACEEE, Washington, pages 13–65, 2012.

[33] Average SpeedTest Results for Top cities of Pakistan. https://www.bandwidthplace.com/location/pakistan/

[34] Understanding USB 3.1 and USB 3.2 https://www.ptgrey.com/understanding-usb-31

[35] Samsungs ssd with 1tb storage inside and 26 gm weight https://www.samsung.com/uk/memory-storage/portable-ssd-t1/mups1t0beu/.

[36] Speedtest market reports of calculating upload and download data rates in 2017-18 http://www.speedtest.net/reports/.

[37] Cisco visual networking index: Forecast and methodology, 20162021 https://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/complete-white-paper-c11- 481360.html.

[38] Jayant Baliga, Robert WA Ayre, Kerry Hinton, and Rodney S Tucker. Green cloud computing: Balancing energy in processing, storage, and transport. Proceedings of the IEEE, 99(1):149–167, 2011.

[39] Annual energy outlook 2018, u.s. energy information administration https://www.eia.gov/outlooks/aeo/pdf/aeo2018.pdf , February 2018.