Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Ciudad de México

School of Engineering and Sciences

Getting Smarter Urban Mobility in Mexico City: Visualizing Streets Pavement Conditions and Anomalies Through Fog Computing V2I Networks and Machine Learning

A dissertation presented by

Alec García Barba

Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

In

Engineering

Mexico City, June 10th, 2021

Dedication

To my parents Adriana and Rogelio, and to Mariana. I hope I will always be able to share my joys and achievements with you.

Acknowledgements

Throughout the writing of this dissertation I have received an invaluable amount of support, assistance, and guidance.

I would first like to thank my advisor, Professor Rogelio Bustamante. From his constant guidance to his great support during the experimentation phase of the project, he was always “al pie del cañón”. Without the data obtained during these difficult times, this project would not have been what it is today. But mostly I want to thank him for believing in me and giving me the opportunity to do research with him for the last three years. It has been quite a ride.

I would like to thank my co-advisor, Dr. Javier Izquierdo Reyes for his great ideas, suggestions, and his patience over the years, pushing the results of this project way farther than I initially believed they could.

I would like to acknowledge MSc. Luis Curiel and Dr. Renato Galluzzi for taking the time to constantly review my work and provide great ideas for improvement, as well as the entire research group CIMB for their help.

In addition, I would like to thank my parents for their unconditional support and great example. You are always there for me.

Getting Smarter Urban Mobility in Mexico City: Visualizing Streets Pavement Conditions and Anomalies Through Fog Computing V2I Networks and Machine Learning

by

Alec García Barba

Abstract

One of the mayor problems of Mexico City today is its poor mobility. Despite the constant efforts and monetary investments to emerge as one of the largest smart cities in the world, City government has not found a solution to mobility problems. Without a complete coverage of public transportation, and around 35 million daily commuters, with a third of them lasting more than an hour, the country's capital has been considered several times one of the world's most congested cities.

One way to make better decisions about public spending on mobility is to analyze data related to the conditions of its cities' streets and avenues. Generally, the streets and avenues are fixed as soon as they have a citizen report or when a major incident happens. However, it is uncommon for cities to have real-time reactive systems that detect the different problems they have to fix on the pavement.

Today, the government provides internet connection in more than 13,690 Access Points distributed throughout the city. This document thoroughly reviews the context of Mexico City, its communication system range, and technological limitations. At the same time, it is proposed to implement a distributed computing network that couples with the existing infrastructure to capture, filter, and analyze the data that could potentially help decision-makers in terms of public spending on mobility through sensors within vehicles that travel those streets daily and connecting them to a fog-computing architecture on a V2I network.

This solution detects main road problems or abnormal conditions in streets and avenues by implementing Machine Learning (ML) algorithms to compare roughness against a flat reference. An equipped vehicle obtained the reference through accelerometry sensors and then sent the data through mid-range communication systems.

The present work compares the accuracy and F1 score metrics of a soft-Max Artificial Neural Network (supervised MLA) and a K-Nearest Neighbor (Unsupervised MLA) to select the best option to handle the acquired data, and compares both model's classification in two different avenues close to ITESM CCM premises.

List of Figures

List of Tables

Contents

Chapter 1: Introduction

Since the beginning of our history as humanity, we have focused on different communication techniques, in the coding of our language and the interpretation of the signs and symbols that we use to transmit the ideas that we generate inside our head. This is how various alphabets, codices, languages and languages have been generated with which we communicate as a species. Telecommunications are a symbiosis of the human desire to communicate with each other, with the desire to connect at a distance.

Thanks to last decade's technology development, humans are not the only ones connecting to one another. Inventions such as Internet of Things, Cloud computing, as well as telecommunication breakthrough technologies such as DSRC, CV2X or 5G, entire cities are slowly, but steadily connecting to the internet. It is projected that approximately 38.6 billion IoT devices will be connected to the internet worldwide by 2025 [1].

Concepts such as Smart Cities and Smart Mobility are transitioning to reality as governments start to provide free wireless connectivity within the city, as well as private competitors lower their costs and bring new alternatives to citizens. There seems to be no clear consensus on the specific definition of both terms, but generally speaking, Smart City refers to an ultra-modern urban area that addresses the needs of businesses, institutions, and especially citizens with focus on Environment, Economy, Mobility, Governance, way of living and people itself [2].

While some authors focus extensively on the sustainable aspect of smart urban cities [3], others focus on the integration and impact of Information and Communications Technology (ICT) throughout the whole city [4]. The term of smart city evolves constantly, and it is rediscovered through new lenses almost on a daily basis.

Smart mobility plays a decisive role in considering whether a city is considered smart or not. Even though a lot of definitions arise in the literature as well, the better suited to this document is the one presented by G. Lyons [5]:

“Generate and share data, information and knowledge that influences decisions; using technology to enhance vehicles, infrastructure, and services; and deriving improvements for transport system operators, users and for shareholders.”

The data sources for smart mobility are endless: vehicles, roads, buildings, pedestrians, even traffic lights, or sensors embedded in any of the above. However, the data created by any source that can help achieve a smarter mobility must be communicated for further filtering, processing, analysis, and finally be used to improve the city. This is where vehicular communications come into play.

Vehicle to Infrastructure (V2I) communications have been implemented based on numerous standards, such as IEEE 802.11n, DSRC, and Infrared techniques [6]. Generally consisting of single hop communication between the vehicle and the infrastructure. These vehicles can have a built-in On-Board Unit (OBU) or have them added afterwards. A Road-Side Unit (RSU) generally acts as a router and has a higher coverage than the OBU range. This is the link between the vehicle and any other possible computing device, and a tight integration system can have both the routing and the computing system.

Once the communication of such data is successful, it still needs processing, and analysis to be useful. In the traditional context of smart cities, such data is stored and handled in the cloud, whether it is public or private. Cloud technology provides ubiquity, almost unlimited storage and computing capacity, elasticity, and benefits from a cost efficiency model; turning the cloud into an advantageous technology. [7] However, the massive use of cloud computing unmasks a new set of challenges.

30.8 billion devices connecting to the internet, specially to the cloud, will lead to many issues, including network congestion, delay, and privacy concerns if the data is analyzed on cloud. More specifically, transmitting the data to the cloud consumes many network resources, congests networks, and leads to long latency. [8] Just on IP traffic alone, a staggering 19.5 zettabytes of information is expected to travel on data centers globally [9].

The concept of Edge computing was born in order to solve this problem. It leverages the storage and processing capacities of a large number of devices connected to the internet deployed for the purpose to provide an intermediate layer between the end devices and the cloud. With the presence of these “Edge devices”, the computation

load at the data centers is reduced by handling some of the requests directed to the cloud, locally, which do not require intervention from the cloud. Therefore, this approach reduces the latency in resolving the requests and allows real-time handling of a subset of requests. Edge devices also support mobility due to the abundant availability and geo-distributed nature. [10]

Another concept that proposed both by the industry and academia is fog computing. Fog Computing bridges the gap between the cloud and Internet of Things (IoT) devices by enabling computing, storage, networking, and data management on the network nodes within the close vicinity of IoT devices. Therefore, computation, storage, networking, decision making, and data management occur along the path between IoT devices and the cloud, as data moves to the cloud from the IoT, or in our case, Internet of Vehicles (IoV) devices.[11]

Edge computing is located at the edge of the network close to IoT devices; note that the edge is not located on the IoT devices, but as close as one hop to them. [11]

The Industrial Internet and Open Fog Consortium makes the distinction that fog computing is a system-level hierarchical architecture and it provides computing, networking, storage, control, and acceleration anywhere from cloud to things; while, edge computing tends to be limited to computing at the edge, being the boundary between the pertinent digital and physical entities, delineated by IoT devices. [12] Figure 1 and figure 2 present a general scheme of vehicular communications, and a comparison that show the difference between edge and fog computing.

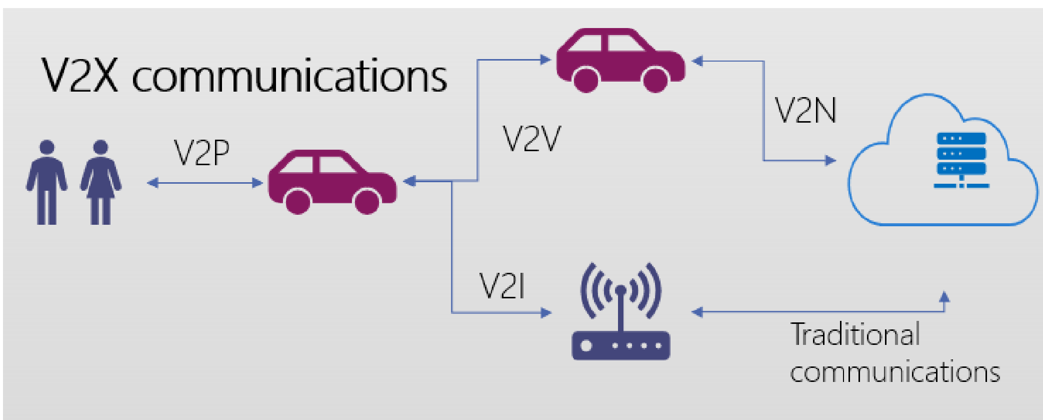


Figure 1. General scheme of Vehicular communications

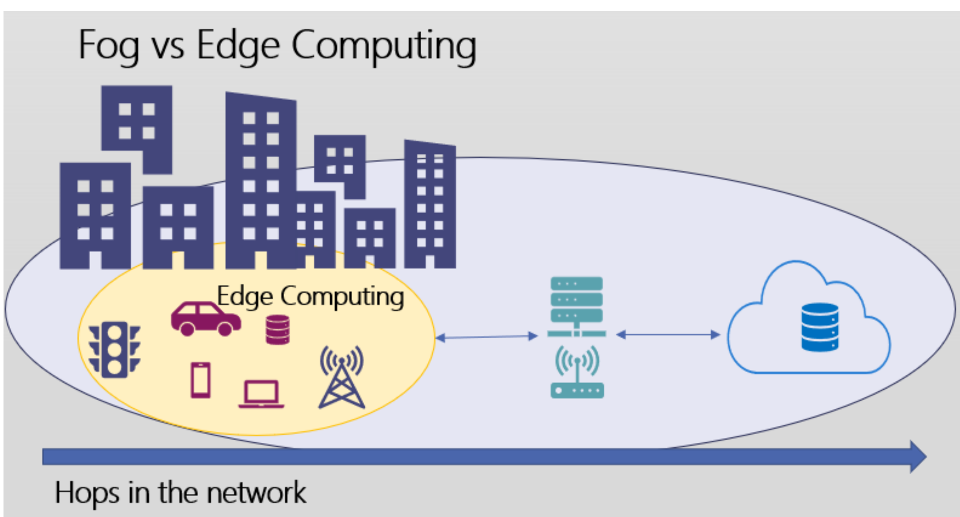


Figure 2. Fog vs Edge computing comparison

Chapter 1. Introduction

One of the questions this dissertation means to answer is how to approach the massive list of problems urban environments currently have, with the technology we currently possess, under the specific scope of smart urban mobility.

Each city is different from the others, and definitely not all cities have the same problems, nor the same technological capacity or economic resources. That is why there are no concrete methods or universal recipes that solve the mobility problems of every urban environment. It is imperative to solve the problems of the city in the same context in which the city finds itself, and adapt solutions to its resources and possibilities.

Mexico's capital, CDMX, as of 2020 is the home of 9,209,944 people [13]. With a city this enormous, and territorially extensive, problems regarding mobility arise on a daily basis. Public transport inefficiencies, heavy traffic, and a deterioration of the urban landscape are just the tip of the iceberg of the problems that millions of people must bear every time they want to move from one point to another.

It is imperative to know in depth the context, the available technology, and the possibilities of implementation, as well as the major areas for improvement within the city before proposing solutions to these problems.

In this dissertation, I present the context in which CDMX currently operates in general terms of urbanism and mobility related problems, technology, and communications. And afterwards, I propose a smart urban mobility approach to start solving mobility problems using a V2I-fog computing architecture, with data visualization tools on the cloud. At the same time, this solution will be aligned with the principles of affordability, effectiveness, attractiveness, and sustainability.

This document is divided as follows:

- Chapter 1: The motivation, problem statement and city and global context is provided. The research question, hypothesis and both general and particular objectives are stated. The limitations of this project are stated at the end of this section, as well as the main contributions.
- Chapter 2: A literature, theoretical framework and a state-of-the-art technologies review is detailed, in terms of smart mobility, vehicular communications, fog computing and anomaly detection.
- Chapter 3: a research-based methodology is proposed to implement a V2I-Fog network. Two main avenues near ITESM CCM campus are used for reference and experimentation.
- Chapter 4: the final results of the project are presented.

- Chapter 5: Both the results and the overall performance of the project are discussed.

- Chapter 6: A conclusion is derived and recommendations for further research are suggested.

1.1Motivation

With the invention of the automobile, new possibilities arose that unleashed great economic, social, and technological growth. These advances of course have been, and will, continue to be of great benefit to humanity. However, they also were accompanied by large costs.

The excessive use of cars has generated problems related to climate change, mobility, accidents, economic losses, and loss of lifetime of people who need to commute multiple times a day in these highly urbanized cities. These problems grow exponentially as the poor planning of cities and their accelerated growth rate limit emerging solutions. Specially in developing cities with high birth rate.

Scientists and engineers around the world are currently seeking solutions to this set of problems: including electric vehicles, better proposals for public transport, autonomous vehicles, and a myriad of small solutions that seek to improve the quality of the air we breathe, the lives and stress of people who spend entire hours daily in some form of transportation, the economy, and/or general well-being.

Despite the fact that many solutions have been studied in depth, and even tried some level of real-life implementation, carrying them out in cities in a practical way has been nearly impossible. Due to its practically null feasibility due to high costs, among many other factors, like solutions that require a much higher level of computation because the lack of information of the vehicle's surroundings, or simply because the vehicle requires data from the network.

This dissertation is focused on developing a platform in which all these individual solutions can be integrated into a large vehicular network, given that they are V2I solutions, where information can be collected, vehicles connected, data analyzed to give better and different proposals, in an effective way. This solution is based on 802.11 communication technology, fog computing architecture support subsequent projects related to the solution of vehicle problems in cities. However, as new technologies arise, migration to these technologies are heavily advised.

1.2 Problem Statement and Context

The exponential increase in automobile manufacturing, together with the population growth rate in urbanized cities, constantly generates new needs and problems. Traffic, a poor coordination within the public transport system, the safety of pedestrians and drivers, as well as the enormous financial losses due to the man-hours invested in transporting workers make it necessary to find a solution. Those solutions require concrete, data-based decisions.

A clear example of economic and time losses can be seen in Mexico City, where 35 million average trips are made on a weekday, of which 33.3% last more than an hour, and 7.5 million of those trips are work related activities. 36% of them last at least two hours [14].

It is easy to notice the great loss of both man hours and fuel use in a city with that much commuting. The total cost of vehicular congestion costs the Valley of Mexico $ 46,043,636,087 MXN annually [15].

In order to find a way out first, it is necessary to fully understand the dimension of the problem. Therefore, first we need to measure, and intelligently gather information regarding traffic and mobility to be able to find areas of opportunity, and thus, implement efficient and effective solutions that solve these problems related to modern urban mobility in Mexico City.

In the era of "Big Data", where efficient information processing adds high-reaching value in any company or government, solutions based on extremely large data sets processing set both a standard and an example to face the aforementioned problems. However, before going into this topic, it is necessary to ask what is needed before being able to process the data related to vehicular mobility in urban environments. The answer is to have an infrastructure to collect and ship them.

As our knowledge about mobility in the city increases, better proposals can be applied to solve the main issues. A V2I-Fog Computing architecture that, in addition to being an instrument for the connectivity of intelligent vehicles, can also enhance data analysis related to mobility, safety, and roads in the city, seems to be a great fit in this context.

Throughout this thesis project, it is sought that, at the same time that an infrastructure is generated to collect information and understand the problem, the same infrastructure generates data analysis regarding the mobility situation, using the same architecture to advance a large part of the process. In this way, a two-stage process where the appliance is first created and then the data will be processed after the collection and transportation will be time-reduced.

By having a large-enough network, that processes data in real time, and processes information faster, instead of using the most common storage services like cloud or private services, and that can provide information to the city's decision makers, or even the new generation of intelligent vehicles, we will be able to provide solutions worthy of our own technological advancement.

1.3 Research Question

How can a smart urban mobility visualization solution can be made for Mexico City that uses existing infrastructure and vehicles to propose solutions in terms of traffic and road safety?

1.4 Hypothesis

A V2I-Fog computing architecture that integrates different sensors in vehicles, preprocesses information at the edge of the network, and through fog computing detects problems in the state of the streets and brings data visualization through cloud will bring Smart urban mobility solutions in terms of affordability, effectiveness, attractiveness, and sustainability.

1.5 Aim and Objectives

This project involves several areas of knowledge, from hardware implementation and communications, all the way through application development and artificial intelligence. However, the purpose of this project is to unite hardware implementation, telecommunications and software development in an instrument that benefits the urban environment of Mexico's capital with use of their data, while understanding the context in which the city currently operates. The Aim and Objectives, therefore, are described below.

Aim:

> Design, develop and evaluate a V2I-fog computing network that receives, processes, and subsequently sends data related to traffic, accidents and/or vehicle safety automatically, in order to have a smarter urban mobility in Mexico City.

Objectives:

- Generate information through devices at the edge of the network and send it to micro servers for pre-processing.
- Build databases accessible through the internet with useful information for urban decision making, related to mobility.

- Design and evaluate an independent protocol tool that analyzes data provided by the same vehicles, detecting mobility, traffic, and safety problems in CDMX in order to make proposals for improvement.
- Produce an efficient, affordable, effective data visualization tool that allows better decision-making. Developing greater sustainability of the city in the long term.

1.6 Limitations and annotations of the dissertation

No autonomous driving algorithm, system, or process will be implemented. This project will only focus on establishing vehicular communications and defining the distance ranges per access point with a mean PER of less than 20%, an average latency of less than 300ms, and a network with up to 4 APs. The architecture will be tested with 1 OBU, and 1 RSU, and its scalability will be put to test via simulation.

The requirements in terms of PER is practically a quarter of the best result obtained from using radio frequency technologies operating in free band for close environments, and as good as a cellular-technology V2I communication in a LOS scenario, both measured in previous works: [16],[17]. The latency average limit was also taken from previous experimentation on V2I networks.

On the other hand, the maximum distance is directly dictated by the frequencies used by wireless technology, whether cellular, 802.11 or other alternatives. A decrease in the maximum distance is expected as well, due to the loss in LOS, partial obstructions with metal objects, weather conditions, or any other unexpected events that could alter the digital communications.

It should also be clarified that if there is a change in communication technology in the future, where they have higher frequencies, according to the increase in speed, their reach will be directly reduced.[18]

Various experiments will be carried out in two different avenues. A reference street, which has been recently paved, and with few anomalies to take as the “ideal street”, and another avenue with very little maintenance, where it will be sought to make a data processing system that detects different anomalies in the avenue.

Data from sensors connected to the vehicle will be used that subsequently will be processed in the available computing nodes. Afterwards, an analysis will be made on two different fronts. The first analysis will be regarding the condition and rugosity of the streets where routes will have been traveled by a standard vehicle with an OBU added on it. The purpose of this analysis is to prove if the architecture can bring benefits to the decision makers in terms of general street investment.

The second analysis will regarding how much information was avoided sending to the network, to determine the percentage of throughput release and storage use savings for then carry out a technical, electrical, and financial analysis comparison between a conventional V2N architecture, and a V2I-Fog computing architecture.

1.7 Contributions

Generate robust vehicular communications in local scenarios with the possibility of being replicated *en masse*, which can make the traffic of an entire city more efficient, generating savings for traffic and time that would avoid expenses of up to millions of pesos annually.

The technical goal of this dissertation is to make possible the implementation of various solutions to traffic problems. Haddad proposes that, under an efficient and robust vehicular communications scenario, eventually all traffic lights will be eliminated to allow a continuous flow of cars as a solution to traffic jams [19]. This can be achieved by doing an optimization model where each vehicle drives at its maximum possible speed, with its proper speed limit on each street or avenue, without encountering another car at any intersection. A possible first step to reaching a system of this nature is by designing a V2I communications network.

Leaving the scope of the project for a moment, and regarding vehicular collision avoidance, various systems have been made to avoid deaths, crashes, and increased pollution. Among them are proposals such as making predictive calculations for worst-case analysis for pedestrian movement [20], or the automatic collision avoidance system from Mercedes Benz [21]. On environmental issues, it is proposed to use the sensors of buildings, supermarkets and public spaces to contribute to the database necessary for autonomous driving [22].

All these scientific advances can be highly benefited on, or are completely dependent of, vehicular communications. It is very difficult to operate relatively inexpensive autonomous vehicles with proper environmental sensing if they do not obtain data from their surroundings.

Finally, a fog computing system directly contributes to an economic benefit, since vehicles avoid having built-in supercomputers, and can more easily use currently available resources due to the short physical and logical distance. Data centers will avoid handling heavy traffic and massive storage. Fog achieves this by using relatively inexpensive equipment right on the edge of the network, with otherwise unproductive devices which can operate in standby or asleep mode [7]. Considering the fog computing architecture benefits, this technology can lead to a substantial benefit to the scalability of the project.

Chapter 2 - Literature Review and State of the Art

Over the last decade, vehicular communications have gained popularity with both academia and industry, since they are getting closer to being a reality due to the emerging technologies surrounding it, such as Artificial Intelligence (AI), 5G, the Internet of Things (IoT), and autonomous driving. Efforts to develop and improve the former, have substantially benefited adjacent technologies.

The different technologies that surround this new generation have opened a huge spectrum of possibilities, including smart cities and therefore smart urban mobility. I will now focus on providing an overview of the most relevant technologies developed in the last decade in terms of smart urban mobility, V2I communications, fog-computing, and different mechanisms for street anomalies detection.

Two dictionary definitions actually help to understand the direction of this project:

- Smart – having or showing a quick-witted intelligence; far on or ahead in development and progress.
- Intelligent – having the ability to acquire and apply knowledge and skills.

Thus, intelligent transportation systems must have the ability to apply knowledge given by the data processing of different vehicular communications systems, and smart urban mobility must route the city's development to an accelerated progress.

This chapter is divided in three parts. Firstly, I provide an overview of previous work [16], [17] related to vehicular communications, as well as a theoretical framework in which these technologies operate. Then, a literature review is performed to help understand how smart urban mobility solutions have been carried out around the world. Lastly, state-of-the-art technologies are revised to compare, contrast, and guide the project with the best solutions within the context of Mexico City urban conditions.

2.1 Theoretical framework & Previous Work

2.1.1 Theoretical framework

Internet of Vehicles (IoV), as we know it today, focuses on intelligent integration of humans, vehicles, things, and environments. It is a larger network that provides services for large cities or even a whole country. IoV is an open integrated network system with high manageability, controllability, operationalization and credibility, composed of multiple users, vehicles, infrastructure and networks. [23]

Multiple applications arise from this technology. For instance, safety applications, such as notifying vehicles about dangerous situations within the roads ahead or identifying possible parking locations before the vehicle actually arrives to the destination.

Vehicles are expected to provide data about vehicle sensors, environment, the driver and passengers. A Road-Side Unit (RSU), or generally speaking, Road-Side Equipment (RSE) can exchange context-aware information with an On-Board Unit (OBU), or On-Board Equipment (OBE) installed inside any of the vehicles in the network. Context-aware information exchange between RSU/RSE and vehicles can help in generating real time information, e.g., traffic information [24].

The first popular communication architecture for interconnecting vehicles were called Vehicular Ad-Hoc Network, or VANET. The basic principle of a VANET is that a vehicle is a mobile node that enables the connection with other vehicles thereby creating a network [25]. Multiple technologies and protocols arose from this concept network, such as: IEEE 802.11p, Directional Medium Access Control (DMAC), Vehicular Cooperative Media Access Control (VC-MAC), Dynamic Source Routing (DSR), General Packet Radio Services (GPRS), and others [26].

This is considered a local network, connected in mesh, which in theory makes the information collected from one vehicle available to the others, through direct communication between vehicles, or via a Roadside Unit. One of the main problems of VANETs is its limited capacity for processing all the information that is collected by themselves and other actors (such as sensors and mobile devices) around the environment [25].

Current IoV systems must have at least a data source layer, a communication layer, a fog layer, and a cloud layer. This is a direct evolution from VANETs, in which the communication only stayed at the edge of the network.

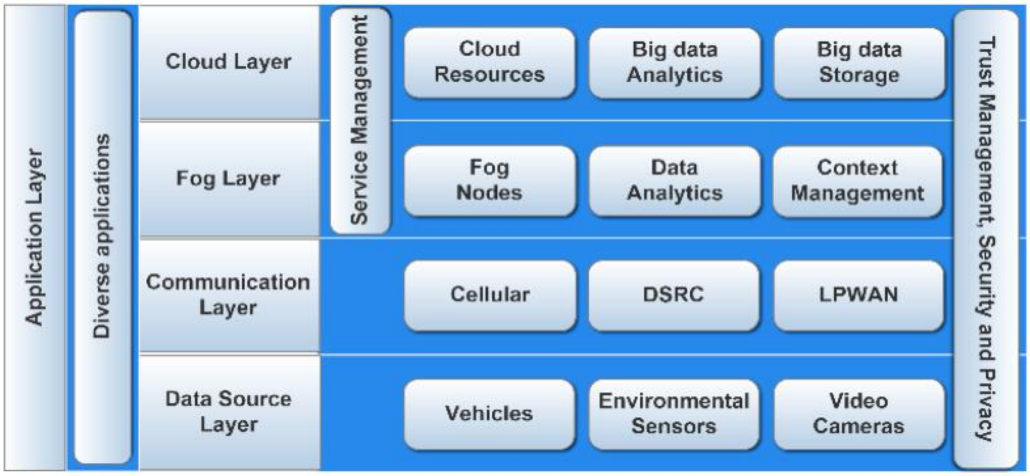


Figure 3. IoV layer architecture [24]

The communication layer represents a fundamental aspect of this project. Independently of the technology used in this layer, the following elements must be considered:

- Transmission frequency and Bandwidth (BW): Every wireless communication sends wave signals in a specific range of frequencies. Bandwidth is defined as the data transfer capacity of a computer, in terms of bits per second (Bps). There is a strong relation between the transmission frequency and its Bw.
- Data Rate. In wireless communications, a more complex modulation results in a higher data rate, but also requires a higher signal-to-noise ratio (SNR) in the receiver to decode correctly.
- Latency: the delay between the sending of a wireless signal, and the arriving of such a signal. Latency can be produced by multiple hops in the network, or mistakes in the communication and constant re-sending of the information.
- Throughput: Throughput is a measure of the rate of successful message delivery over a communication channel. Successful meaning data being transferred without errors.
- Packet error Ratio (PER): The packet error ratio is a value that indicates the number of data packets received incorrectly divided by the total number of received packets, with this we can know in a numerical way the fidelity of transmission in a medium, in this case, a wireless communication. A packet is declared incorrect if at least one bit of the packet or string is wrong.

- Bit Error Ratio (BER): The bit error ratio works in a very similar way as the PER. However, this type of ratio is not always measured in physical networks because an error in a bit implies an error in the packet.

- Signal to Noise Ratio (SNR): measure of signal strength, relative to the background noise in the medium. Usually measured in decibels (dB). Maximizing this ratio leads to minimizing errors in any communications system.

These elements provide enough tools to measure the performance of a wireless communications system, regardless of whether it is DSRC, cellular, satellite, or any other at a physical layer.

2.1.2 Previous work

In the publication “A Simulation Approach of the Internet of Intelligent Vehicles for Closed Routes in Urban Environments”, an implementation of the simulation of intelligent vehicular communications through the IEEE 802.11p protocol was implemented.

An algorithm for the generation of vehicular convoys to improve vehicle flow application through V2X communication was proposed and implemented in the simulation, and the communication quality was validated in a closed route in an urban environment during the simulations, omitting aspects of security in communications. A closed vehicle route was designed using the open traffic simulation software SUMO, and the communications during the route were simulated using the software framework OMNeT++.

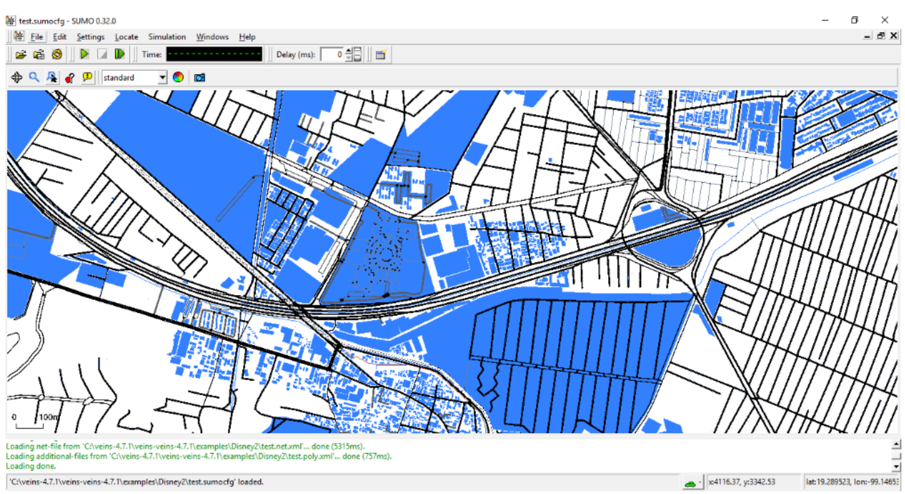


Figure 4. Simulated structural model for a 1km Route

Three antennas have been placed in the simulation model by car. Two vehicles were included that will communicate during the same trip in two configurations: with Line of Sight (LOS) and No Line of Sight (NLOS). 30 additional vehicles are added to observe the effects of adjacent communications.

It was chosen as measure of attenuation of the signal in the transmission between the two cars of 9 dB for each 0.4 of height of the buildings. A central reception frequency of 5.89 GHz places us in the band of the technology implemented in VANET networks. The RSU is located in a building at the entrance of the University, 3 meters above the ground. 80 bits were sent in a frame in 1 second intervals. The maximum interference distance was 2600 meters. The transmission power was 20 mW, and the sending rate is 6 Mbps, the reception sensitivity of the antennas, that is the vehicles, was -89dBm. The presence of thermal noise with a value of -110dBm was considered. The model implemented for the loss of signal was Simple Path Loss Model for the transmission of the RSU in the free space.

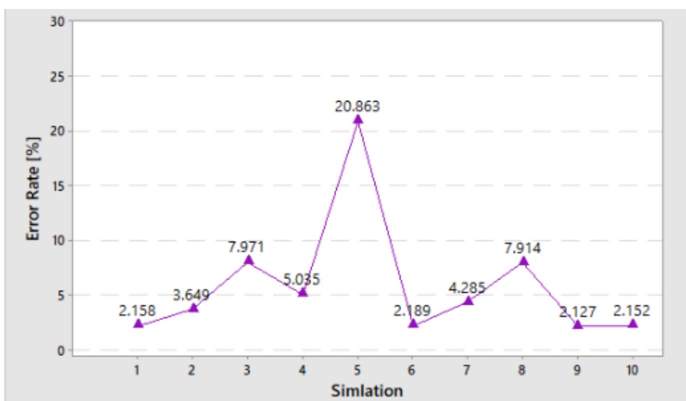


Figure 5. Results of PER in LOS simulations

Within a completely LOS environment, different simulations were run, and a maximum of 20% PER was obtained. However, it was an outlier in comparison to the average result.

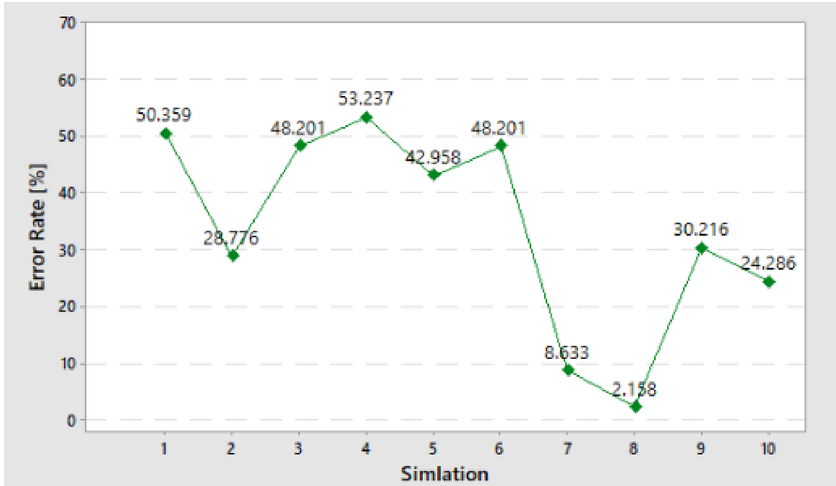


Figure 6. Results of PER in NLOS simulation

In contrast with the LOS scenario, the NLOS environment caused a considerable increase in the packet error rate. Figure 7 shows a comparison between 2 different OBUs in both a LOS and NLOS scenario, with a total of 10 route runs.

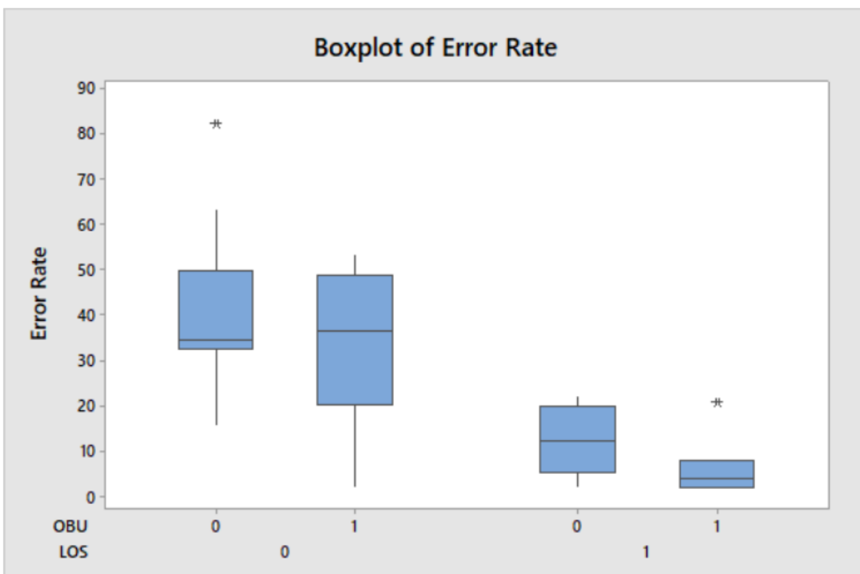


Figure 7. Packet Error Rate Box Plot comparison

In general, it was found that LOS is a very important factor to generate a viable communication on the IEEE 802.11p protocol because of its high transmission frequencies, so having NLOS generates a higher and more variable PER.

Within the bachelor thesis project called "*Emulation of V2X communications*", a V2I architecture was proposed, with only a central computer and a database to test the edge communications system using various technologies. A mixed communication system was proposed, with the 802.11ac/n protocol connecting different APs with a central RSU system, while the communication between OBUs in a VANET styled mesh network.

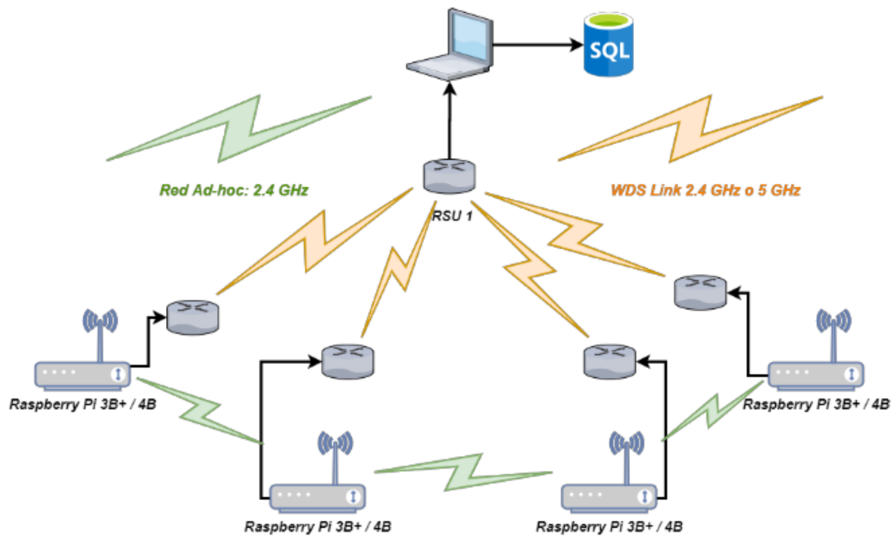


Figure 8. V2I Network Design

A comparison between different modulations used in 802.11 technology was carried out, to determine the BER in contrast with the SNR, to consider the amount of power an antenna with this protocol should have in order to have a low error rate. The other alternative, is to accept a lower modulation in the transmission, which will lead to a poor data transmission rate.

This gives us a theoretical upper-bound limit, depending on the power capabilities of the different appliances used both in the reception and transmission of data over this type of wireless communication protocol.

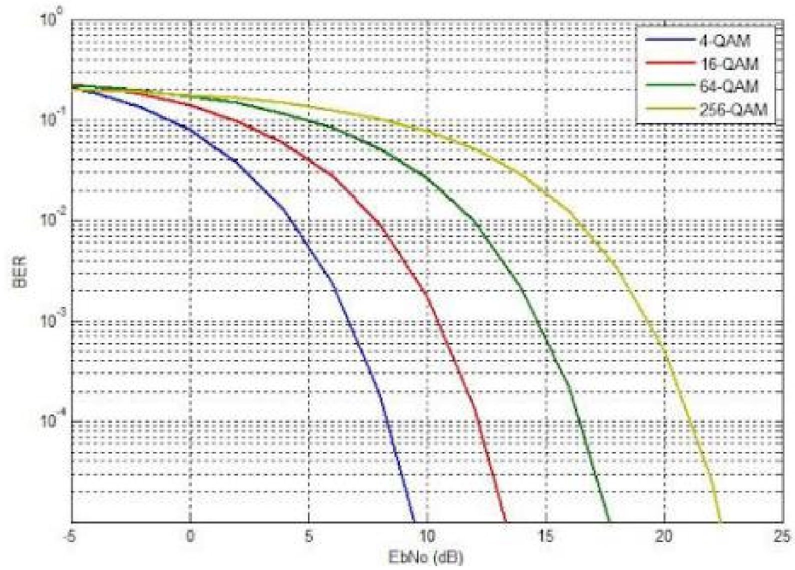


Figure 9. Analysis of BER in Wi-Fi communications, varying modulation and SNR [16]

Finally, an experiment was set up with the following variables: LoS, Cyclic Redundancy Check detection, packet sending retries in case of error, Power, Bandwidth channel, number of devices, distance, and acknowledging times. After 32 repetitions of the experiment, the most important factors within 802.11 protocol to affect wireless communications were found out to be the Line of Sight, Cyclic Redundancy Check detection, Power, and Bandwidth. Considering the maximum distance of the experiment was lower than the limit specified by the protocol.

Multiple tests were carried out where both the latency and PER were tested, considering them as the dependent variables. The main idea was to identify the most important factor during these specific wireless communications, and based on the correlation between controllable factors implement a wireless solution that minimizes both the latency and the packet error rate.

The results can be seen in the figures 10 and 12. Within range, the maximum latency obtained with this particular network configuration network was if 1.5 seconds. However, this latency included the generation of packets, sending through the network, data processing and comparison. This protocol can vary depending on the type of hardware, software, and processing of such communications. Figure 10 shows the total latency of the designed network, while Figure 11 shows different tests and characterization made by Nokia regarding 802.11n/ac communications, creating a Cumulative Distribution Function (CDF) of the latency of such protocol.

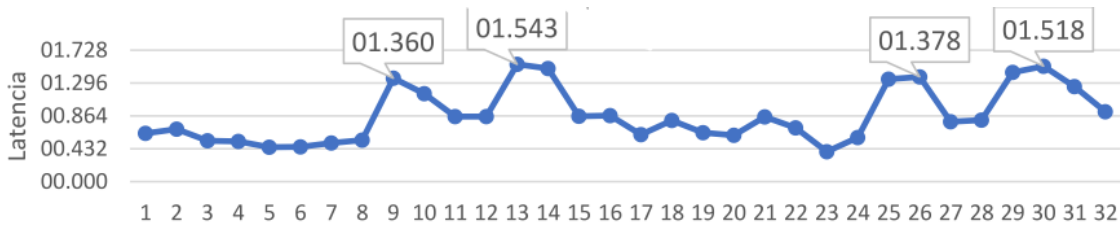


Figure 10. Latency of V2I network throughout several experiments

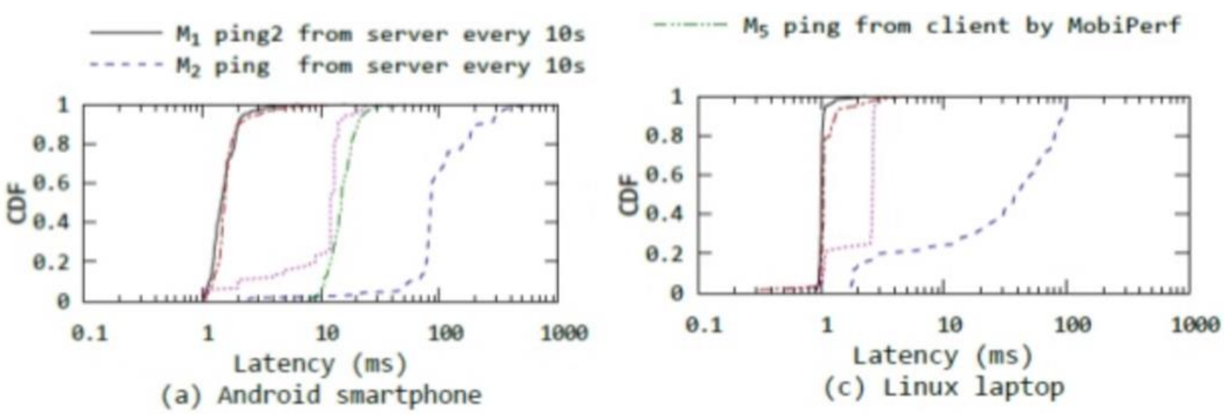


Figure 11. CDFs of Wi-Fi latency varying hardware [27]

With these graphics we can have a valid understanding of the delay in 802.11ac/n networks. While the upstream/downstream latency of a single packet communication, considering different hardware devices, normally takes up between 1-2 ms, a whole communication process normally takes between 400 and 500 ms.

The packet error rate of the proposed V2I network in the bachelor thesis project ended up with results of an average 2.47% PER, with a maximum of 8% given the worst configuration possible and unexpected anomalies in the communications, which can include interference, partial or complete obstruction in the line of sight, or unexpected weather conditions.

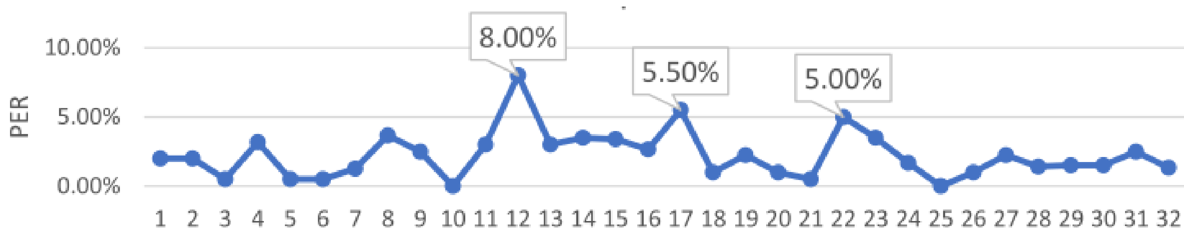


Figure 12. Packet Error rate throughout experiments

After all this previous work and previous work, some key points have to be considered:

- Wireless protocols for V2I communications cover a short distance in relation to the amount of distance a vehicle can travel in a short amount of time. Any real V2I implementation in CDMX must have an enormous amount of Access Points to the network.
- 802.11p protocol seemed promising, however it was substituted by further technologies such as C-V2X and in following years 5G. However, 5G is not currently available for commercial use in Mexico.
- Other alternatives similar to 802.11p, such as 802.11ac/n could represent a temporal alternative while other technologies are accessible to the masses.
- LoS is a main pillar in wireless communications in the range of 802.11 protocols. Whatever solution is proposed with this type of communications must take it into account.
- Given the packet error rate and general latency, other alternatives should be proposed in V2V communications, where ultra-low latencies and reliability are desired. This type of communications will not be possible in reality without 5G.

2.2 Literature Review

The concept of Vehicular Ad Hoc Networks (VANETs) was conceived over two decades ago, and it was a very active area of research, both in academia and industry [25], [26], [28]–[30] as part of the Intelligent Transportation systems(ITS).

ITS are a core component regarding smart urban mobility. Some of the main challenges that ITS try to solve are traffic assessment and management, In-vehicle and on-road safety management, Driver modeling techniques, and emergency management. [31]

The new solutions to the above problems started demanding more requirements to vehicular networks. Since multiple sensors, multiple vehicles, and non-vehicle devices and users were needed to be considered in the equation.

One of the main problems regarding VANETs is its reduced ability to both permit the integration of sensors and data-collecting technology, and integration, processing and forwarding of information to networks or devices outside that meshed network.

In this context, vehicles needed to evolve into smarter objects equipped with a multi-sensor platform, with a set of communication technologies, robust computational units, IP-based connectivity to the Internet, and a direct or indirect connection to other vehicles and with all devices around the environment. In this context, the concept of vehicular ad-hoc networks is evolving into the Internet of Vehicle (IoV) [25].

There are multiple technologies, and use cases of IoV that VANETs cannot simply handle. Some of the solutions that IoV enable in urban environments including traffic control and large-scale mobility coordination through V2I networks, pedestrian detection and avoidance systems in autonomous vehicles with V2V, V2P and V2I communications, collision avoidance within vehicles through V2V, etc. This type of communications are being constantly studied and evolving, both with the help of the academia, the industry and the government. [32]–[34]

Focusing on V2I communications, they enable us to innovate the current traffic control strategies, mainly in three verticals:

(i) Fixed time mobility management, in which pre-programmed signal plans are based on historical traffic data along the day.
(ii) Traffic-actuated, where real-time traffic conditions either change the length or order of signals phases, changing traffic strategies.
(iii) Adaptive, which also predicts near future traffic conditions in order to optimize signal timing using an objective function, minimizing mobility transfer times.[35]

IoV also provides other numerous benefits, including dynamic information services, intelligent vehicle control and applications to reduce insurance rates and increased productivity due to the reduction of traffic jams and car accidents [36], [37]

Since this type of communications needed, in most of the cases, some integration with the internet, as well as data processing, IoV started depending on the Cloud to operate. Due to the inefficiencies of the distance that traffic, latency, and lack of Local to Metro Area Network solutions, Vehicle Edge Computing was born.

While several authors have mixed opinions and thoughts on Edge computing's exact definition, most often refers to having intermediating infrastructure at the edge of the network that handles storage, computing, analysis, and networking issues. This infrastructure must exist, logically speaking, as near as possible to the end user.

In edge computing, the processing and analysis of data take place in the proximity to the end devices. Edge acts as an intermediary amongst the cloud and vehicles. The benefits of Vehicular Edge Computing (VEC) and Vehicular Cloud Computing can be seen at table 1.

Table 1: Comparison Between VEC and VCC, from [38]

Features	Vehicular Edge Computing	Vehicular Cloud Computing
Location	At user's proximity	Remote location
Latency	Low	High
Mobility support	High	Limited
Decision making	Local	Remote
Communication	Real Time	Constraints in Bandwidth
Storage Capacity	Limited	Highly Scalable
Context awareness	Yes	No
Device Heterogeneity	Highly Supported	Limited Supported
Computing Capability	Medium	High
Cost of Development	Low	High

One of the main reason Edge Computing has gained momentum in the industry is because of the reduced latency that it brings to applications that need real-time data processing or secure data treatment. A couple of examples unrelated to vehicular communications include Virtual and Augmented reality, smart manufacturing, or smart retail artificial intelligence applications.

Within Vehicular communications, considering them a subset of Internet of Things that require low latency for some of the real-time analysis applications, processing as close to the vehicle as possible can become critical to avoid any kind of accidents. This is one of the main reasons that Edge computing was considered a good suit to the IoV environment.

Multiple applications can depend on this type of technology, from secure safety data sharing of vehicles [39], to enhance passengers user experience using deep reinforcement learning [40]. IoV does not limits to specific vehicle technical requirements, it can also improve the user experience during the recurrent travels within the city.

However, With the growing quantity of data generated at the edge, speed of data transportation is becoming the bottleneck for the Edge-Cloud based computing paradigm [41]. Since only one Gateway and processing node or cluster generally exist in edge computing networks, most of the computing actually goes to the cloud for further processing. There is an evident tradeoff here, where you need to bring expensive computing power to the edge, or end up with a lot of bandwidth consumption and processing in the cloud. In both scenarios, economic costs can naturally rise up and scale out of proportion.

VEC is definitely an improvement from vehicular cloud computing networks, however, a better solution can be found within Fog Computing networks.

Fog Computing is a highly virtualized platform that provides compute, storage, and networking services between end devices and traditional Cloud Computing Data Centers, typically, but not exclusively located at the edge of network. The defining characteristics of the Fog are:

a) Low latency and location awareness, b) Wide-spread geographical distribution, c) Mobility, d) Very large number of nodes, e) Predominant role of wireless access, f) Strong presence of streaming and real time applications, and g) Heterogeneity [42]

Based on this type of computing network, the architecture layer in figure 3 describes its need to separate data, communication, fog cloud. A more device-oriented architecture can be seen on figure 13.

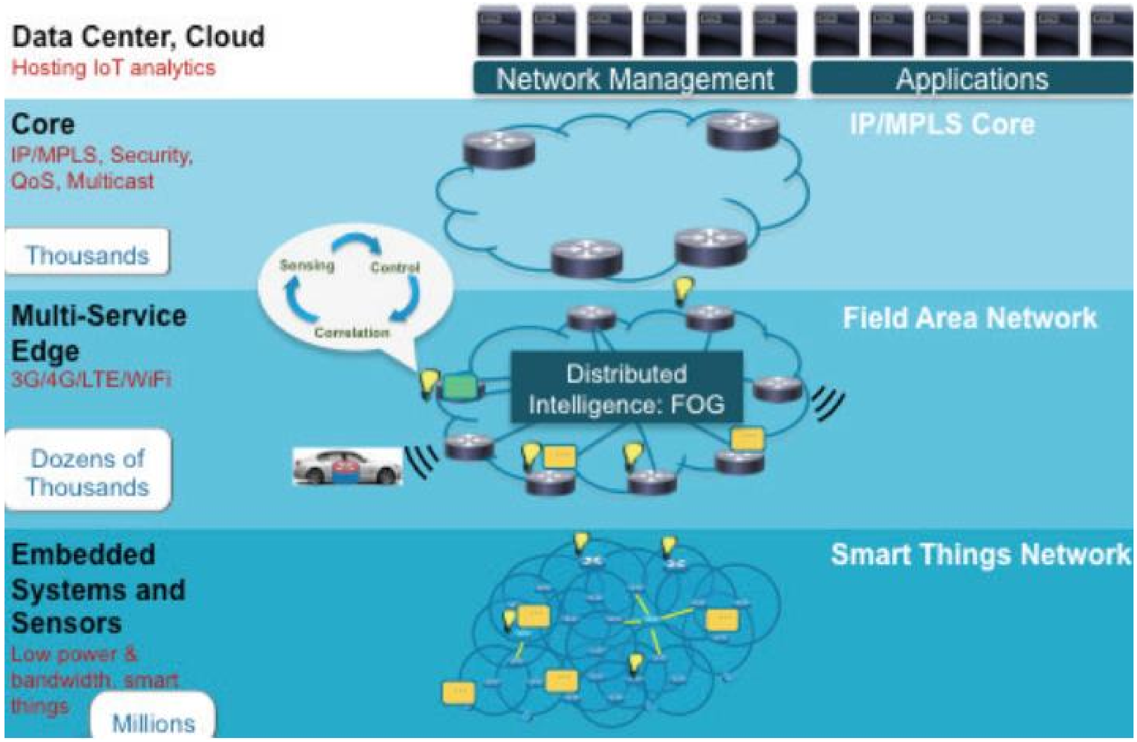


Figure 13. Fog Computing conceptualized architecture [42]

The evolution from local VANETs, to Cloud computing VANETS and IoV to Vehicular Edge Computing, and finally to Vehicular Fog Computing has come a long way in the past two decades.

Of course there are advantages and different use cases better suited depending on the type of applications, but all things must be considered depending on the scenario. Real-time applications need powerful computing and strong communications, and big computing processing must be done as close as possible, giving VEC architectures a complete advantage. Big Data solutions are more suitable to VFC architectures, where multiple processing can be done in different clusters, optimizing computing due to its distributed nature. Lastly, applications which need to preserve data as it is can enter in a more suitable way into a Cloud-Computing environment, in which security of the data can be better taken care of.

Besides the general application, other things can be considered within these three architectures. The amount of storage capacity, context or even application awareness, etc. There is no clear winner in these three architectures in an IoV environment. Depending on the solution desired, any architecture could be better suited. A more detailed comparison is presented in table 2.

Table 2. Comparison Between VCC, VEC and VFC architectures

Features	*VCC*	*VEC*	*VFC*
Location	Remote location	User's proximity	User's proximity, and remote location
Latency	High	Low	Low
Mobility support	Limited	Higher	Highest
Decision Making	Remote	Local	Remote & local
Communication	Constraints in Bandwidth	Real Time	Real Time and asynchronous
Storage Capacity	Highly scalable	limited	Highly scalable, both locally and remotely
Context Awareness	No	Yes	Yes
Device Heterogeneity	Limited	Highly supported	Highly supported
Computing Capability	High	Medium	High
Cost of Development	High	Low	Medium
Cost of Scalability	Low	Medium	Low

With the enormous scalability of vehicular-fog-computing architectures, it must be considered how to scale, create clusters, and take advantages of all the computing processing capability acquired through these types of networks. While Edge Computing needs one good computing appliance or cluster at the border, Fog computing can be built with several low-cost computing devices as nodes of a cluster of clusters, making it more scalable. But, in order to surpass edge computing in terms of benefits, it must be built with the proper amount of nodes and clusters.

The next section describes some state-of-the-art technologies that demonstrate the real benefits of such networks.

2.3 State of the Art

This section is completely focused on the state of the art regarding smart mobility and V2I communications in urban environments. It is divided in 4 main sections:

- V2I communications in Smart Mobility.
- Fog Computing in V2X communications within smart cities.
- Communication technologies regarding V2X.
- Anomaly detection within smart vehicles.

2.3.1 Smart Mobility and V2I applications

When referring to smart mobility, the literature in general refers specifically to smart urban mobility. This distinction is important due to the fact that technology is completely city oriented, since most of V2X network rely heavily on QoS, replication, and are designed for high data transfer/processing volumes.

Even though multiple definitions arise frequently, this dissertation will take as reference the one proposed by [5], defining smart (urban) mobility as: connectivity in towns and cities that is affordable, effective, attractive and sustainable. This is of special value in the context of the following chapters; if by some reason a user cannot afford to connect to the infrastructure, the proposed infrastructure is not smart enough.

Connectivity should be treated only as what it is, a means to an end. This end should be affordable, effective, attractive, and sustainable. This end by no means should strictly be on the technological high-end side, especially if it does not comply with any of the previous definition of smart mobility.

In order to compare the progress, in terms of smart mobility within a city, several authors have tried to condense all the properties of a city that converts them in a smart city, with smart mobility. [43] proposed an indicator, Smart Mobility Indicator (SMI), based on technical infrastructure, information infrastructure, mobility methods and vehicles used for this purpose, and legislation. The technical infrastructure must be able to communicate with different kinds of transport, and the quality of those communications, as well as the implemented applications impact considerably on the “smartness” of the city.

Main V2I applications found in the literature revolve around: traffic solutions, secure communications for privacy of data, and data processing.

Traffic signal algorithms to control flow traffic have been studied by [35], where simulations have showed a decrease in waiting time down to almost 80%, and reductions in travel time of 15%, comparing to non-connected vehicle scenarios.

B. LV et al. developed a LiDAR-Enhanced Connected Infrastructure to provide high-resolution traffic information to both smart vehicles, autonomous or not, and a central computer, providing up to 8 MB of information about real-time traffic in no more than 200ms. [44]

Security in communications vary from safeguarding communications in the physical layer through mathematical models and algorithms with encoded transmissions [45], to using blockchain technology as a way to secure information through V2I communications, with blockchain handling vehicle's authorization [39], [46].

Data processing with this type of architecture takes advantage of artificial intelligence. Autonomous vehicles platooning for time reduction using directional positioning algorithms in a platooning decision making process with multiple communications technologies and sensors has been tested in [47], while Long Short-Term Memory (LSTM) models to reduce interference and enhance data streaming and performance can also be done in this type of networks [48].

2.3.2 Fog computing in V2I infrastructure within smart Cities

There are many reasons of why a fog computing architecture can benefit IoV communications, especially in V2I networks and infrastructures:

The RSU can be physically damaged by some malicious activity or other harsh environments. With appliances exposed to weather, traffic, and all kinds of conditions, nodes of the cluster will eventually go down.

The network connectivity during V2I and I2I communications may be temporarily cut off, affecting immediate data communication.

The vehicular network should be sufficiently scalable to adapt to the increasing number of vehicles. As more and more people adopt vehicles with OBU capacities, or intelligent vehicles go out to market, the amount of nodes within clusters, and clusters themselves must scale accordingly.

Data processing needs to be performed closer to the data sources to minimize network latency [49], and as more roads get smarter, more clusters are needed to cover them in a low latency environment.

Javed et al. proposed a general fault-tolerant V2I framework with a publish/subscribe data communication pipeline using Kubernetes as both a virtualization and fault-tolerant management tool. Allowing the system to survive node failures. They have an 802.11p communication system between vehicles within the 5.850 – 5.925 GHz spectrum. The concept behind Kubernetes and virtual containers is to avoid having a single point of failure throughout the whole architecture.

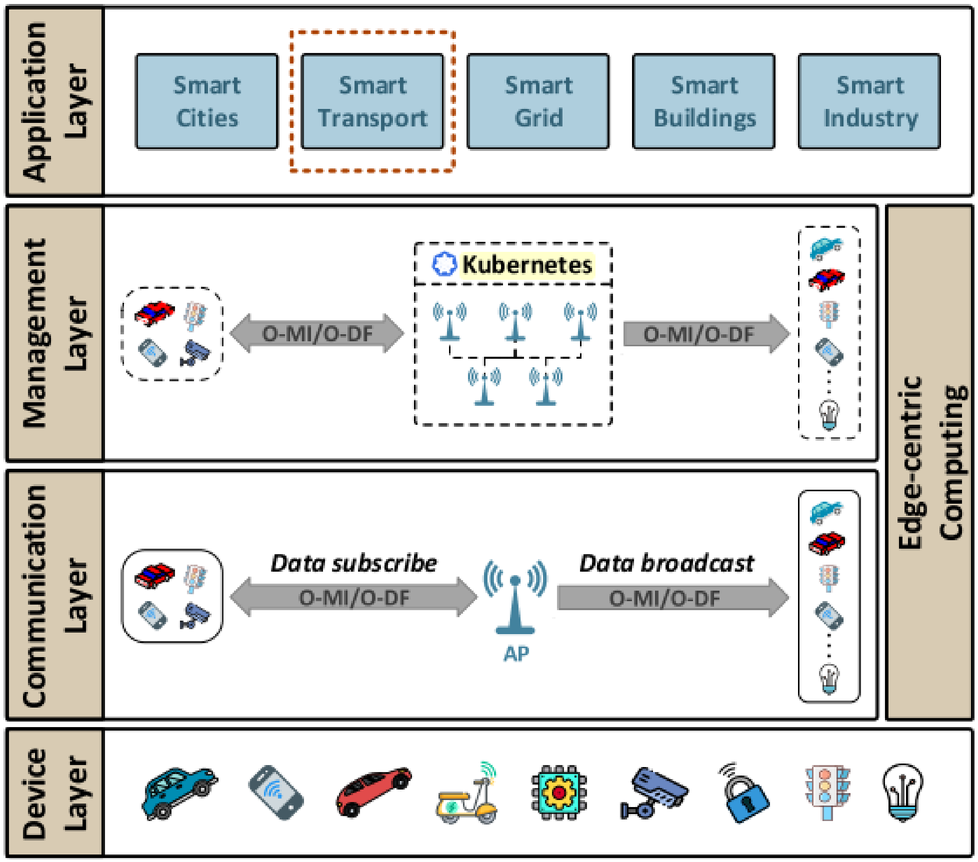


Figure 14. Fault-tolerant distributed network from [49]

In this context of distributed computing within smart cities, Artificial Intelligence can empower edge computing. Dai et al. propose an AI powered computing caching through deep reinforcement learning with a deep deterministic policy gradient that estimates the performance of network policies [50]. The application of Deep Learning in distributed network benefits the entire architecture in terms of improving the actual management of the network, allocating resources more quickly and efficiently.

Fog computing needs multiple components to operate, once the infrastructure allows communications in a V2X network, and the algorithms for data processing, as well as the security challenges are overcome, the distribution of data needs to be considered. FCN operate in a Big Data context, and the Resilience of data is completely necessary. Figure 14 shows a state-of-the-art processing data model for fog computing networks.

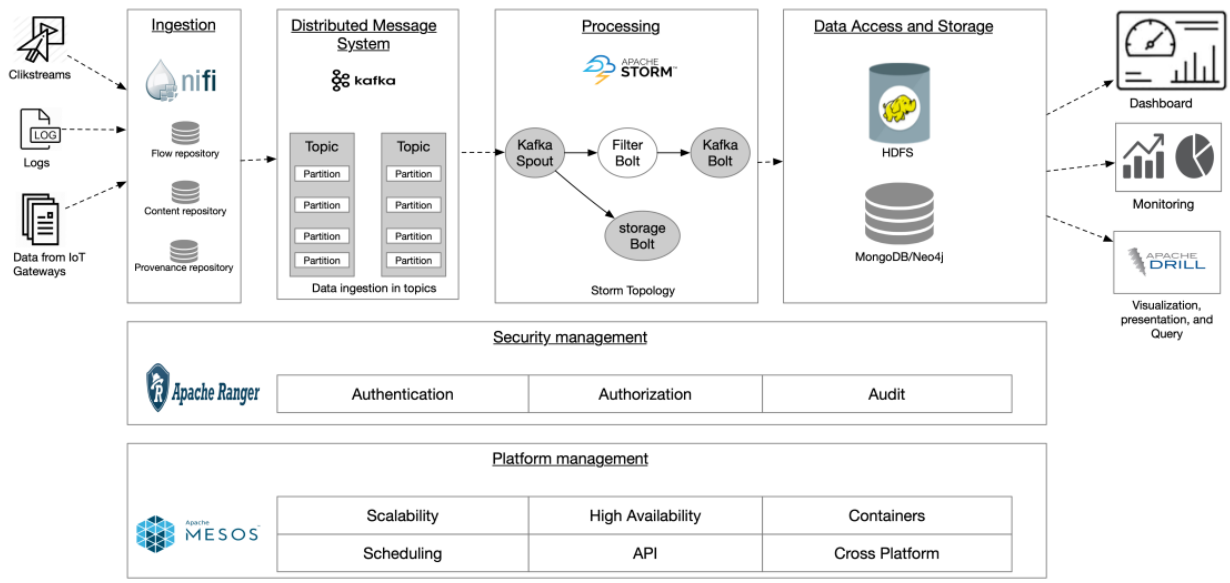


Figure 15. Data Processing model alternative in a Fog Computing Network, from [51]

One of the key points of this alternative is its distributed data access and storage. The combination of Hadoop and the database, which in this case is a noSQL Mongo instance, allow the network to rely in a distributed data environment. In a VFC network, not only the processing of the data must be distributed, but also the data itself.

Apache STORM as the distributed processing software is also a good candidate, considering the entire Apache software ecosystem integration and scalability.

2.3.3 Communication technologies for V2I

Communications can be divided in several ways. However, one of the most effective ways is to divide them by the actual range of distance in which the technologies can operate. Ahangar, Ahmed, Khang et al. did a spectacular state-of-the-art survey regarding communication technologies and divided them by range. The three main classifications can be found in tables 3, 4, and 5.

Table 3. Comparison of short-range technologies from [52]

Standards	Bluetooth	BLE	ZigBee	UWB
Specifications	IEEE 802.15.1	IEEE 802.15.1	IEEE 802.15.4	IEEE 802.15.3
Frequency	2.402–2.481	2.402–2.481	868/902–968 MHz, 2.4 GHz	3.1–10.6 GHz
Bandwidth	1	2	0.3/0.6 MHz, 2 MHz	500 MHz–7.5 GHz
Rate	1–3 Mbps	1 Mbps	20–250 kbps	480 Mbps
Range	10 m	50 m	75–100 m	75 m
Latency (msec)	100	6	30	0.1
Modulation	GFSK	GFSK	BPSK,O-QPSK	BPSK-QPSK
Data Protection	16-bit CRC	24-bit CRC	16-bit CRC	32-bit CRC

Some of the applications of short-range technologies vary from localization to warning signaling. These types of communications can’t carry vast amounts of data and are not suitable candidates for low latency critical applications.

One example of application in short range Ultra-wideband (UWB) technologies is ranging estimates to track the vehicle’s position in an outdoor environment. This could be a more precise alternative to GPS.

Martin et al. built an accurate and reliable positioning solution based on the combination of UWB ranging estimates, together with inertial and odometry data of the vehicle [53]. However, UWB technologies offer a really low range of communications.

Vehicle Identification can be done through low-cost, long battery life ZigBee technology. One proposed application is the classification of vehicles and communication through the 802.15.4 protocol [54].

Zigbee technologies can be used in V2V communications under certain conditions, as long as the bandwidth or throughput needed for the application remains low. The maximum channel bandwidth of Zigbee technology does not exceeds 0.6 MHz, with a maximum data rate of 250 kbps.

One more application regarding short-range technologies is collision warning under low-speed circumstances through Zigbee. [55]

Other applications could include communication of sensors in Wireless Personal Area Networks (WPAN) [56], and low-energy BLE-based fingerprint localization [57]

In closing, short-range communications are suitable for V2X whenever specific conditions are set, or they can help as aide for geo-localization, warning signaling, or identification. More complicated applications that include security, internet connection, or real-time data processing are out of the scope of this type of communications.

Medium range technologies wide out the spectrum of application possibilities. The two main medium range technologies can be compared in Table 4. However, different variants of the 802.11 protocol exist with different bandwidths, modulations, and spectrum ranges.

Table 4. Comparison of medium range technologies, from [52]

Parameters	**Wi-Fi**	**DSRC**
Specifications	IEEE 802.11a	IEEE 802.11p
Width in channel	20	10
Signalling	OFDM	OFDM
Data Rate (Mbps)	upto 54	upto 27
Modulation type	upto 64QAM	upto 64QAM
Symbol duration (micro sec.)	4	8
Guard Time (micro sec.)	0.8	1.6
FFT Size	64	64
FFT period (micro sec.)	3.2	6.4
Preamble duration (micro sec.)	16	32
Sub carrier spacing (MHz)	0.3125	0.15625
Frequency spectrum (GHz)	5	5.9
Latency (msec)	50	100
Range (m)	100	300

Applications of mid-range technologies are generally in the V2I-V2X spectrum, where data safety, management, and some non-maximum security data processing is better suited in comparison of short-range communications, not just in terms of distance, but in data transfer rates, latency, modulation, etc.

Vehicle to vehicle communications in highways were tested in simulation with DSRC, but with high latency and big packet error rates, proving the difficulties these technologies have with this type of communications [58].

Other type of tested applications were vehicle communications in platooning scenarios [59], and maintaining lane distances through vehicle communications [60]. However, the results were never as promising, with higher latencies and packet error rates than expected.

Pedestrian recognition, detection and informing can also be considered within this range of applications, integrating human-vehicle classification through Support Vector Machines (SVM) and radar systems at the infrastructure, while communicating this information to vehicles, bicycle, or even other pedestrians. [61]

One sustainability application found in the literature revolves around eco-driving with fuel optimization with traffic status and v2i communications [34]

The IEEE 802.11 family can be used for information dissemination and Internet access, however, if the goal of the network is to enhance complete autonomous driving, other technologies must be integrated into the system [62].

Studies have been made about the support of applications over Wi-Fi in smart cities, testing access to the internet or voice over IP (VoIP) applications, which can upgrade the user experience of passengers. [63]

As seen on the literature, mid-range technologies can enhance better algorithms, communication between vehicles, and even the integration of Artificial Intelligence. There is a significant improvement in terms of latency, reliability, distance and bandwidth. However, new generation applications such as autonomous driving and real-time processing in V2V networks are still out of the scope of mid-range communications.

This is where long-range technologies come into play. A comparison between mid-range technologies can be seen on table 4. However, the theoretical comparison and reliability of the network still lacks that effectiveness in practical scenarios. Table 5 offers a good comprehension of long-range technologies.

Table 5. Comparison of long range technologies, from [52]

Parameters	C-V2X	5G-NR V2X
Subcarrier spacing (kHz)	15	up to 240
Carrier aggregation	Up to 32	up to 16
Channel Bandwidth (MHz)	20	400
Latency (msec)	<10	<1
Reliability	95–99%	99.9–99.999%
Channel Coding	Turbo	LDPC, Polar
Network Slicing	No	YES
Waveform	SC-FDMA	OFDM
Control and data multiplexing	FDM	TDM
Modulation	16 or 64QAM	256 QAM
Communication Type	Broadcast	Broadcast, multicast and unicast
Re-transmission	Blind	PSFCH
Security and privacy	Basic	Advanced
Positioning accuracy (m)	>1	0.1
Frequency Spectrum	800/1800 MHz	700 MHz/3.6 and 26 GHz
Range	100 m to >5 Km	50 m to >5 Km

Applications of long-range technologies can really achieve complete secure and low latency V2V, and general V2X applications.

Long range technologies can make possible V2X service negotiation and location-aware scheduling to provide the network with insights about the mobility pattern and application requirements of the vehicular equipment, enabling the network to optimize the availability and scalability of automotive applications. [64]

Other important benefit that 5G and C-V2X communications bring is a better Quality of Service (QoS). 5G offers a unique provision of providing computational resources and storage at the network edge. This enables the network to host applications closer to vehicles, reducing greatly latency. Multi-Access Edge Computing (MEC) utilizes the network core and backhaul efficiently, therefore reducing the latency requirement of autonomous vehicles [178]. Since vehicular users are highly mobile, frequent handovers can be supported via resource management at the edge of the access network [64]

Most of these communications technologies operate within the IP protocol, a layer of the network architecture in which the vast majority, if not all of the internet devices operate. Jeong et al. did a survey focused on different IP-based vehicular networks [52, Fig 15.], where several protocols, architectures, and mobility handling techniques were discussed. An IP approach to vehicular communications is essential in a cloud or fog computing environment to interconnect vehicles with the rest of the internet, bringing together all types of software applications.

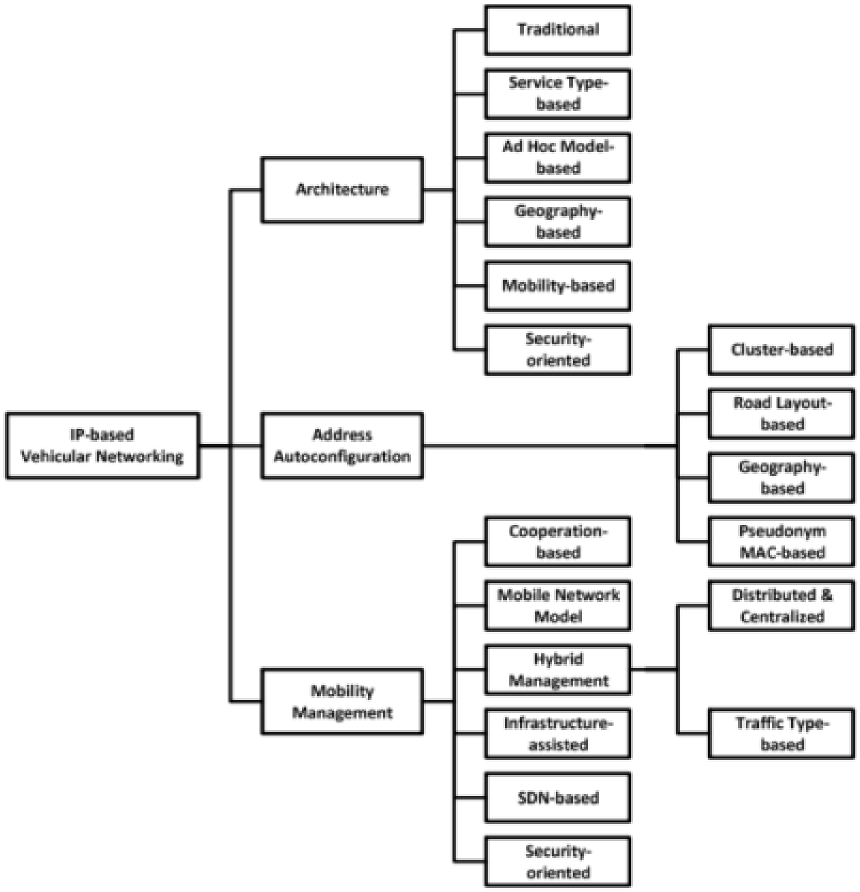


Figure 16. Classification of IP-based vehicular networks by Jeong et al.

In recent years, new clustering approaches over V2I communications have emerged. One of the most recent is a Multi-hop Clustering Approach over Vehicle-to-Internet communication called MCA-V2I to improve local V2I communication performance, allowing vehicles to connect to the Internet via a special infrastructure called a Roadside Unit Gateway (RSU-G) to perform the clustering process. The latter is performed using a Breadth-first search algorithm for traversing the graph and based on a Mobility Rate. The MCA-V2I approach strengthens the cluster's stability, improving latency and traffic handling. [66]

Cluster-based approach in a fog computing environment is desirable, since distributed RSU equipment could communicate effectively and have processing resilience. With all these different technologies, resource management, mobility handling, and clustering processing, could make possible an automated transporting environment. It is possible to have an intelligent transport system in coming years, similar to that of figure 16, where a fully integrated technological environment of vehicle systems can be appreciated.

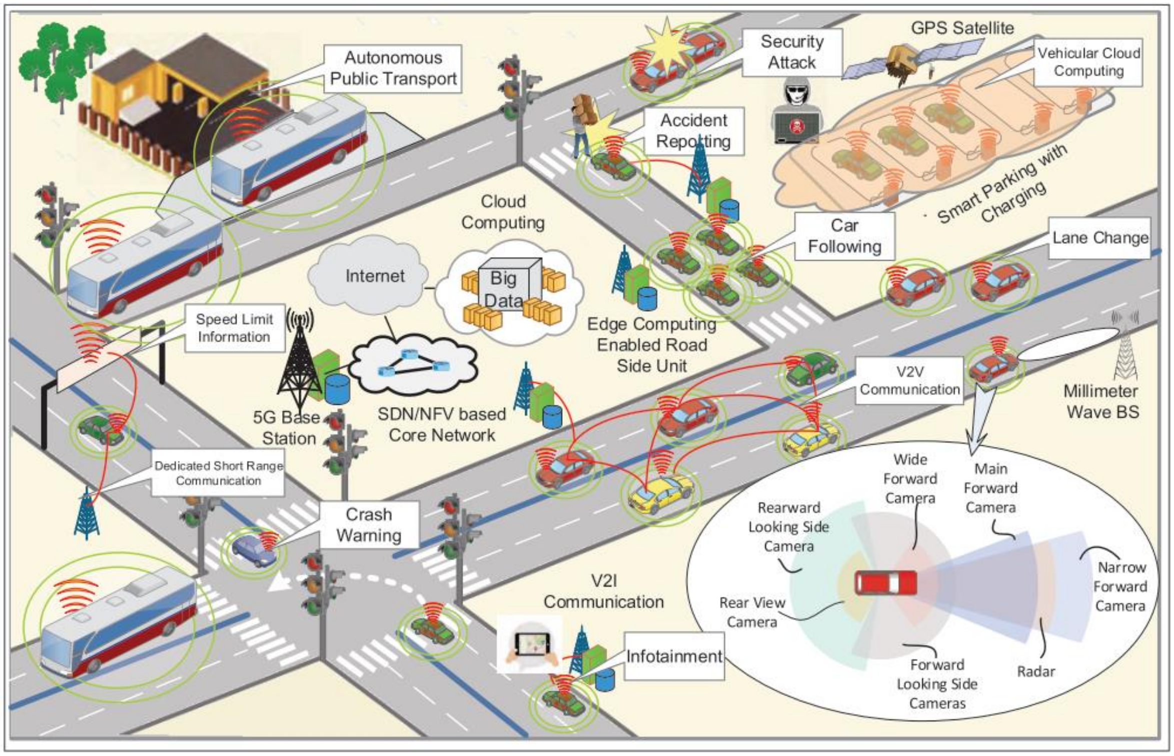


Figure 17. Autonomous driving cars in smart cities and its integration with multiple state-of-the-art technologies, from [62]

Depending on the application, different technologies can enhance the user experience and ensure reliability. Figure 17 offers an example of different types of different-range communications enhancing a complete IoV environment, including safety, warning, driving assistance and much more applications.

Instead of choosing the better alternative, the best alternative is to understand which technologies are better suited to the given task, and creating an environment that maximizes the use of the electromagnetic spectrum is critical to create an IoV environment which can be integrated seamlessly to a smart city environment.

2.3.4 Anomaly detection within smart vehicles

The combination of communication technologies, fog computing, and different types of vehicular communications in a smart urban mobility environment within smart cities bring the possibility of the integration with multiple kinds of Artificial Intelligence that detect, classify, and even predict different types of relevant phenomena in transportation.

One impressive feature that AI enables is driver's behavior detection, including driving under influence detection [67], and the detection of the driver's emotional state through sentiment analysis [68], potentially saving drivers from accidents caused by their temporal inability to drive.

Besides from the driven or the passenger inside vehicles in V2X networks, other anomalies can be detected with the aid of artificial intelligence. Pothole detection can be achieved using Convolutional Neural Networks (CNN) such as the AlexNet [69], or the "You Only Look Once" CNN algorithm [70]. Data anomalies caused by faults or errors in the vehicle's sensors, or even cyberattacks can be detected through a combination of Long Short-Term Memory (LSTM) Neural Networks and CNNs [71].

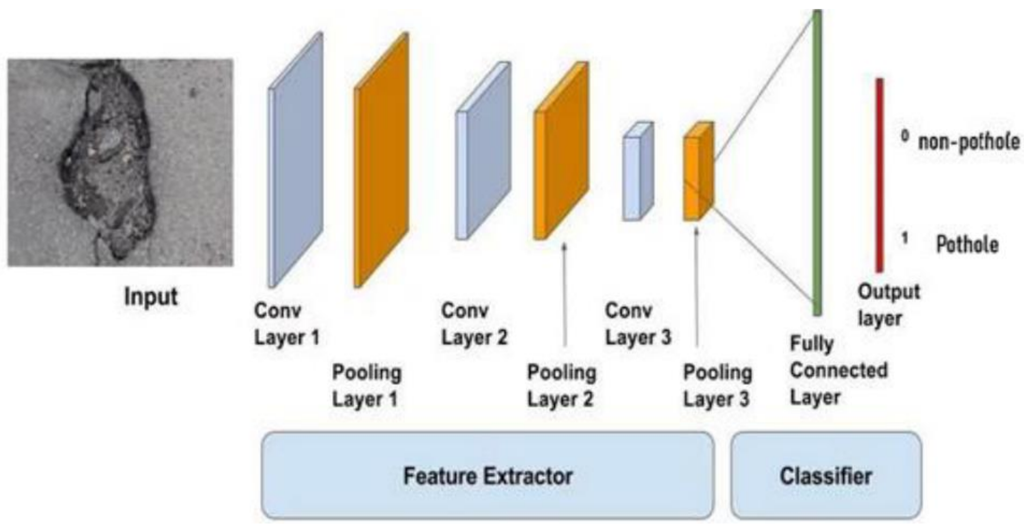


Figure 18. Detection Pothole AlexNet Convolutional Neural Network, from [69]

One of the main problems with the proposed CNN network is that it depends on a clear zenithal plane to detect the pothole. This, in a vehicular mobility environment is non-practical, due to the angle of the camera. It is better suited a front-plane view of the camera to detect more types of anomalies other tan potholes.

There are many approaches to detect anomalies in the streets via image recognition, and there are multiple approaches to sentiment analysis and behavioral classification and prediction, but there was not found in the literature pothole detection through the sensors of an intelligent vehicle, or vice versa, sentiment analysis through image processing and CNNs. While the latter could result in privacy invasion, the former could potentially be a useful application for pothole mapping throughout the city.

Chapter 3. Research Methodology

To present a smart solution that can correctly benefit the mobility context of Mexico City, a great amount of research on different areas must be done. This chapter builds a methodology from the ground up, from preliminary studies to a functional prototype with a data-visualization solution.

The proposed solution, based on the hypothesis presented in chapter 1.5, is that a V2I-Fog computing architecture that integrates different sensors in vehicles, preprocesses information at the edge of the network, and through fog computing detects problems in the state of the streets and brings data visualization through cloud will bring Smart urban mobility solutions in terms of affordability, effectiveness, attractiveness, and sustainability.

Figure 19 presents the solution graphically and broken down into different levels, from the perspective of physical communication, pre-processing, computation, storage, and presentation in the form of a data visualization tool.

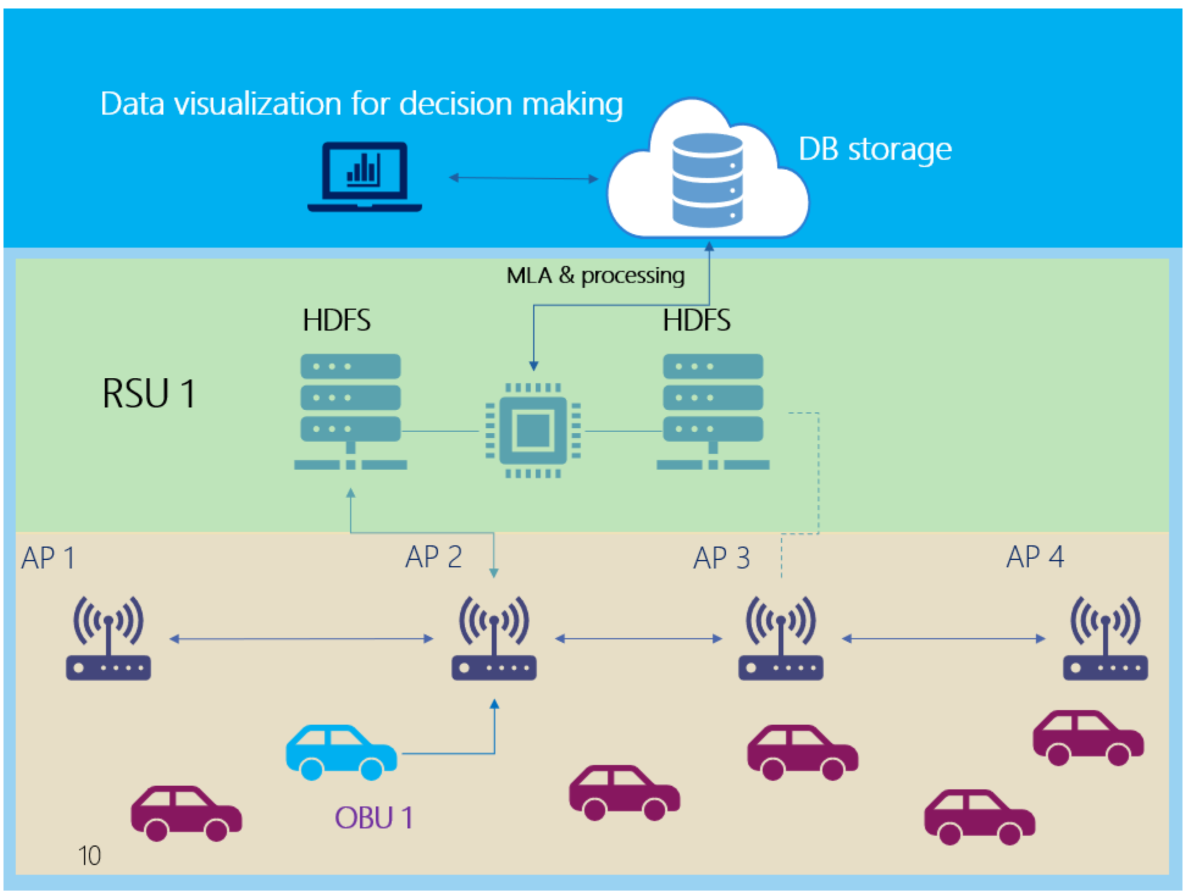


Figure 19. Diagram of the proposed V2I-Fog computing network solution

This solution can be broken-down into several parts. The first part corresponds to the physical communication between vehicles and APs. This corresponds to layer 1, layer 2, and layer 3 metrics, since it is an IP-Based approach. Here, maximum range distance, packet loss, latency and related measures help determine the feasibility of the overall solution.

The second part is related to OBU-RSU communications, and data handling. Since an important part of this project, which is distributed computing can be tested due to external limitations, benchmarks and other research studies are presented to justify the selection of technologies.

The third part corresponds to classification through Machine Learning. In order to process information in a useful way, machine learning algorithms have been proven to be the best alternative multiple times. So, it is needed to come up with useful models that correctly classify the data obtained via the OBU.

The last part of the project corresponds to the visualization of the processed data, and show useful information regarding the city's streets and avenues conditions. This last part proves the true value of the project.

However, before any real implementation of the solution is actually done, a study case of Mexico City must be done beforehand. If the solution presented is not a good fit considering the city's context, or is not aligned with the smart mobility principles presented in chapters 1 and 2, this project lacks purpose.

The solution conceptualized is built and tested in the following sections. To test and verify every aspect of this solution, from conception to the functional prototype, the following methodology is presented:

Section 3.1 Presents a Case Study of CDMX, where population conditions and communications infrastructure are analyzed to compare different alternatives that proposal and its accessibility to the population, both in terms of economic and physical distribution accessibility.

Section 3.2 provides the methodology and analysis of a proposed V2I architecture based on its communications capabilities. Multiple experiments and iterations are presented until the system provides a useful communication system for the project's purposes.

Section 3.3 compares different technologies to implement a faster, and scalable ecosystem for a distributed computing architecture.

Section 3.4 presents integration tests of both the communication system and the computation system in an RSU-OBU environment, with multiple sensors implemented in a standard vehicle, circulating in two different roads and circuits. There are multiple iterations on the system, until the amount of information provided by the OBU is sufficient to detect different anomalies in the pavement of the streets.

Section 3.5 analyses two different ML algorithm to detect anomalous situations on streets and avenues based on the data provided by a vehicle with the OBU integrated technology. Then, show geographical data visualization on the data obtained by previous experiments.

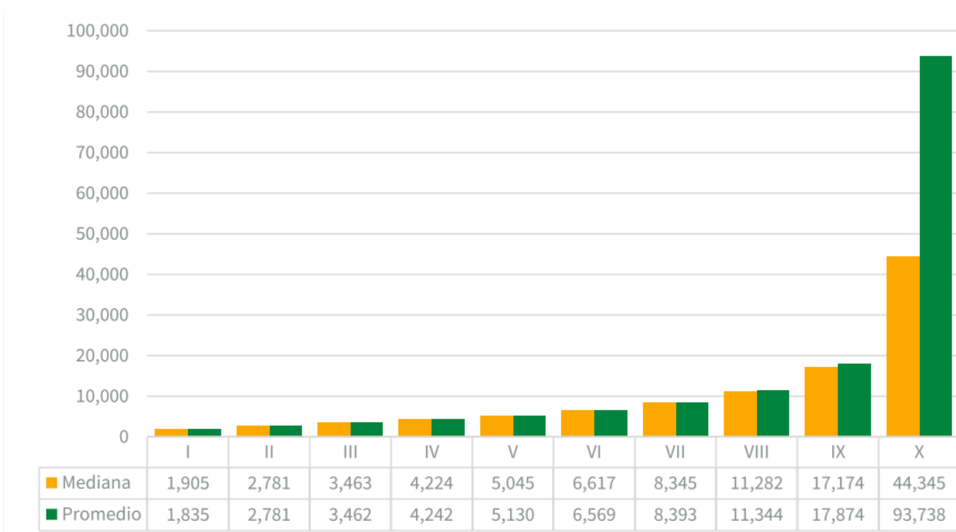
3.1 CDMX Case Study for an IoV communications system

Based on the data provided by the Mexican National Institute of Statistics and Geography (INEGI), Mexico's capital: CDMX (renamed on 2019) is divided in 16 counties, with a total of 1,495 square kilometers, representing the 0.1% of the national territory.

However, being the country's capital, the number of registered inhabitants in 2020 were 9,209,944, representing the 7% of the country's population [13]. This number does not take into consideration the amount of people that travel on a daily basis from the surrounding states to the capital. It is important to notice that in a city with a population density of more than 6000 inhabitants per square kilometer, transportation and commuting is severely challenged.

CDMX represents 17.7% of the total GDP, with the service sector being the most relevant [72]. But even with the amount of productive value Mexico City has, one major problem with Mexico City is the distribution of wealth. Figure 19 shows the average and median income of the population divided by deciles, where the top decile of the population in 2018 earned on average 51 times more than the lowest.

Any solution that plans to benefit the whole city must take into consideration the access to the technology, considering the staggering amount of people with low incomes.



	I	II	III	IV	V	VI	VII	VIII	IX	X
Mediana	1,905	2,781	3,463	4,224	5,045	6,617	8,345	11,282	17,174	44,345
Promedio	1,835	2,781	3,462	4,242	5,130	6,569	8,393	11,344	17,874	93,738

Figure 20. Median (Yellow) and average (Green) income by deciles in Mexico City's population

Out of all the 9.2 million people living in CDMX, 52.5% of the population is considered to be in poverty as of 2018, as shown in figure 20. Considering this situation, it is of utmost importance to bring a solution as low-cost as possible, given that practically half the population will have many problems accessing technology. With smart decisions, technology can bridge the gap between income inequality and technology, or at least increase the average income for every decile within Mexico City.

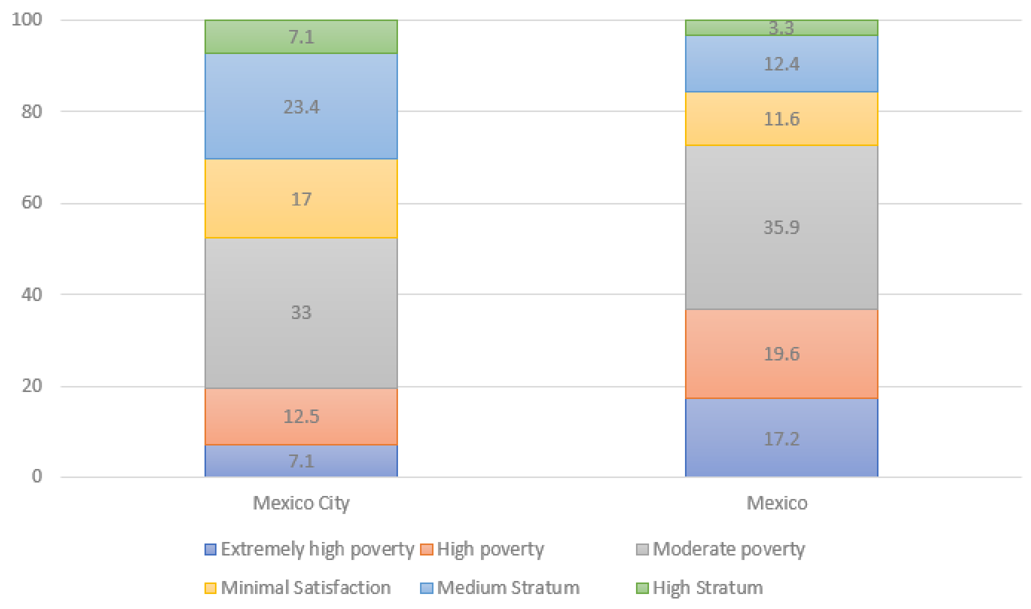


Figure 21. Stacked bar chart of socioeconomic strata in Mexico. [73]

This is all relevant because the distribution and access of technology is normally directly correlated with the distribution of wealth. Early adopters' profile of new technologies generally include higher incomes.

However, another important benefit to the access of Information and Communication Technologies (ICT) can generate social benefits such as solving partially income inequality.

ICTs have been instrumental in generating numerous new employment opportunities as global value chains and outsourcing have shifted production to low and middle-income countries. As ICTs reduce entry barriers into many economic activities, they create new opportunities for self-employment in high and low-income countries. Likewise, the emerging platform economy promises new opportunities of integrating individuals and firms in low-income countries into the global economy. [74]

In order for a smart mobility solution to be considered as smart, the access of technology cannot be restricted by wealth. It should be accessible to as many people as possible.

There are two alternatives for big scale access IoV communications for the capital, given the telecommunications infrastructure of the city: cellular communications or the C5 communication system: a complex distributed technological system that provides the city with video surveillance, seismic alarm, data gathering and compute, and also free 802.11/x Wi-Fi.

There are immediate differences between both alternatives. First of all, every cellular-data frequency band in Mexico can only be used through federal grants. This is the first obstacle in utilizing cellular technologies; there will be cost associated to the internet-access of every vehicle connected to the network, and its access is limited to mobile operators, or virtual operators offering their service through the infrastructure of those operators.

Other obstacle in cellular communications is the country's technological lag. 5G technologies, which are expected to connect the smart vehicles of the future, are far from deployed. Federal grants are still in process, and

Although the Federal Telecommunications Institute (IFT) launched the first public bids for 5G technology frequency bands grants in February 2021 after almost 2 years of analysis [75]–[77]. These late bids will probably delay the deployment of 5G technologies for months or even years, making this technological option infeasible for the short and medium term.

There are currently efforts of the country's main mobile operators to offer dedicated internet access to vehicles. However, the solutions are not adequate for mass scale projects due to the fact that the main target of this solution are luxurious lines, with important modifications of the actual vehicle. The cost for standard vehicles compared with the inhabitant's average income is completely out of proportion.

The other alternative is mounting a distributed computing architecture in the existing C5, or Command, Control, Compute, Communications and Citizen-Contact infrastructure. However, it must be guaranteed that the access of the APs must be distributed equally throughout the city, and an acceptable amount of coverage area is reached.

The first step is to obtain the average income of inhabitants per county, and compare wealth, income, and other factors with the distribution of APs in the city to find any kind of correlation.

Evalúa CDMX is the organization responsible for calculating and analyzing all kinds of social disparities within the capital, and most of the information presented was obtained through the annual 2020 report. Information was obtained as well from CONEVAL, and the open database system of the city.

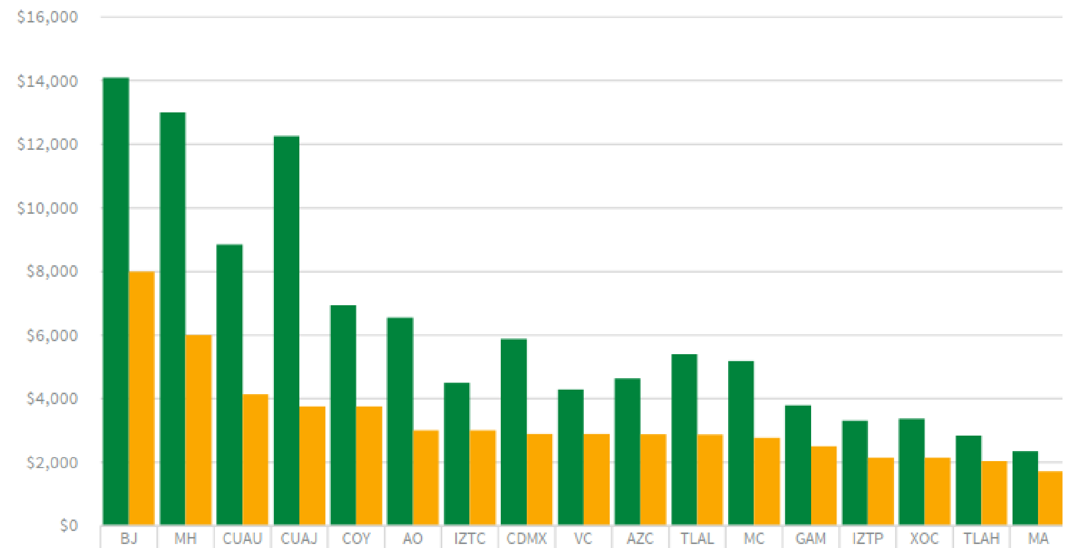


Figure 22. Average income [green] and median income [yellow] of CDMX habitants per county, from [73]

An important measure of economic wellbeing is social development. Its calculation is based on 6 dimensions: Housing quality and space, access to health and social security, educational lag, durable goods, sanitary adaptation, and energy adequacy [78]. It is a dimensionless unit, calculated with the following equation.

$$\mathbf{IDS_K = 1 - HI_K}$$

Where the incidence, or H is calculated by the amount of poor people divided by the total population.

$$\mathbf{(H = q/n)}$$

And the intensity, or I, is calculated as a relative measure of the weighted sum of all 6 dimensions stated earlier (Z), and the distance between the poverty line and the population wellbeing L_j.

$$\mathbf{I_j = (Z - L_j)/Z}$$

One way to obtain information about the correlation of the distribution of access points and the wellbeing or lack thereof, is to get the percentage of high and low social development of each county and measure the correlation to the APs' distribution.

Table 6 Shows information regarding geographical and social conditions that could impact the decisions on how the communication infrastructure would be distributed.

Table 6. Social and Geographical Conditions of Mexico City by County [73], [78], [79]

County	Area [km^2]	Population [%]	Average Income [MXN]	High social development [%]	Low social development [%]	Access points
Benito Juárez	26.03	434,153	14,090	83.2	0	796
Coyoacán	54.4	614,447	6,934	49.6	15.5	907
Miguel Hidalgo	46.99	414,470	12,500	37.9	0.3	926
Tlalpan	312	669,928	5,404	21.8	30.9	1448
Cuauhtémoc	32.4	545,884	8,843	20.8	21.4	
Álvaro Obregón	96.17	759,137	6,548	17.4	31.6	899
Gustavo A. Madero	94.07	1,173,351	3,795	16	17.4	1790
Azcapotzalco	33.6	432,205	4,632	13.9	0.1	707
Venustiano Carranza	33.4	443,704	4,289	11.8	0.6	990
Magdalena Contreras	74.58	247,622	5,187	11.1	38.2	305
Xochimilco	122	442,178	3,365	9.3	37.3	416
Iztacalco	23.3	404,695	4,504	8	0	664
Cuajimalpa de Morelos	80.95	217,686	12,250	6.1	30.7	239
Iztapalapa	117	1,835,486	3,308	4.1	40.7	2100
Tláhuac	85.34	392,313	2,843	0	49.6	540
Milpa Alta	228.41	152,685	2,345	0	100	213

One way to have confidence that this alternative is accessible to the population is that the correlation between the number of Access Points in each county and the size of population is high, while other factors such as income, or level of social development is close to zero. This means that the main factor for the distribution of APs is the population, without any economic bias.

The correlation analysis is presented in Figure 22. After a comparison with several other factors, the only relevant correlation between the access points distribution and any other is in fact the population of the county, with a direct relation given its value (0.89). This preliminary measure ensures that this technology is distributed approximately equally throughout the city depending on the size of the population. This also ensures that the APs are less prone to access saturation.

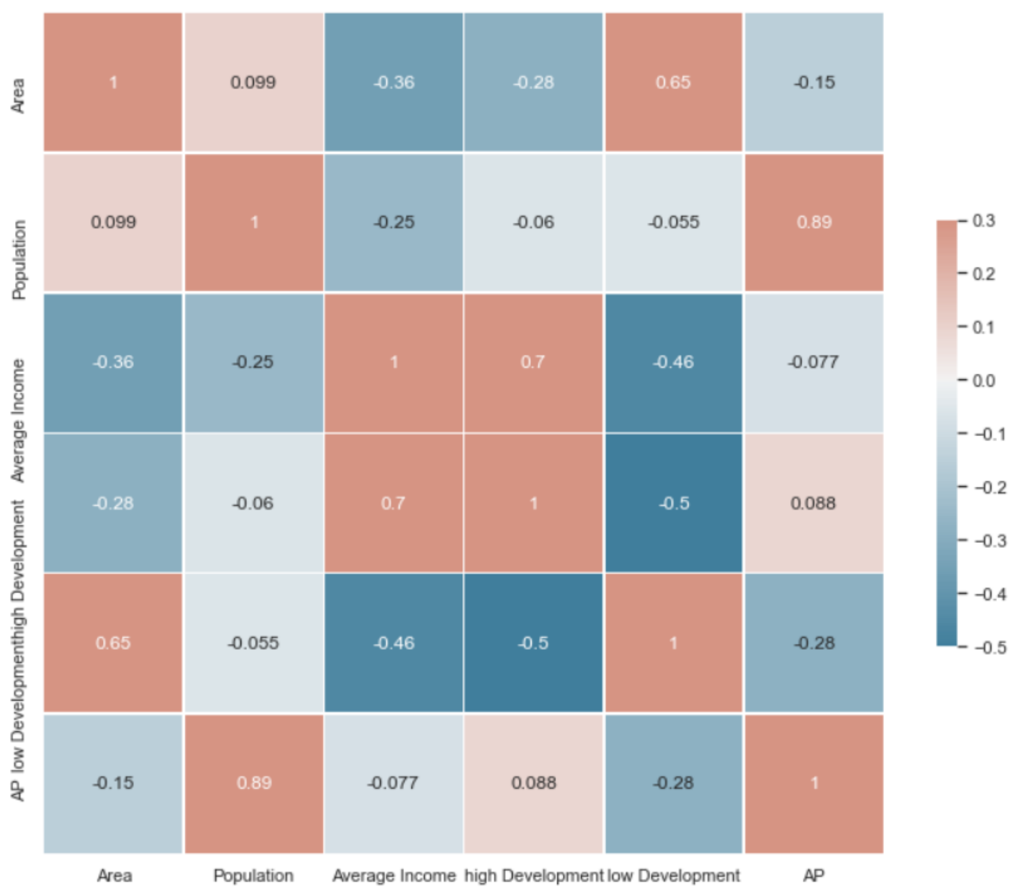


Figure 23. Correlation Matrix between social development, income, and APs in CDMX

The next step is to know the coverage area of the C5 infrastructure. The city's government updates constantly this information in the open database system of the city. There are currently almost 14,000 APs distributed throughout the city. A more visual approach to understanding the distribution is presented in figures 23 and 24.

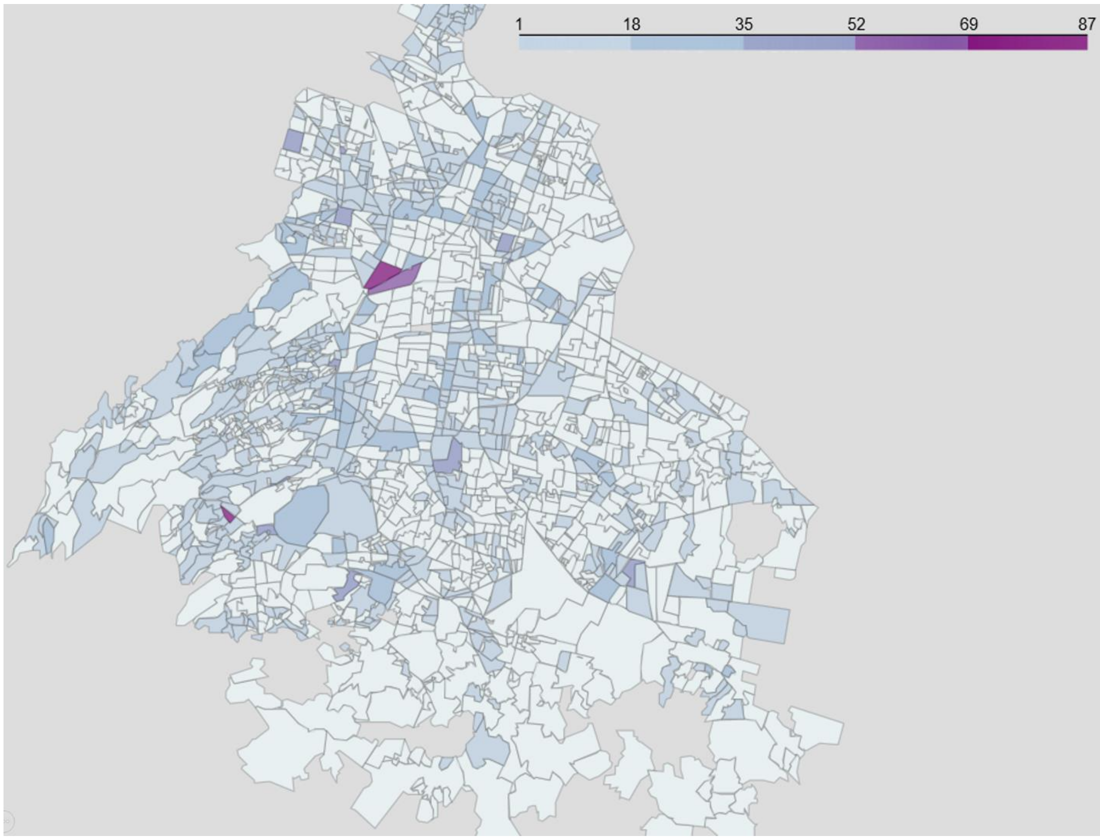


Figure 24. C5 APs distribution cloroplexic map by neighborhood.

Showing the distributed access points by neighborhood, given the correlation with its population, also shows the approximate distribution of the population. However, the size of each neighborhood varies greatly, and so this level of granularity can be deceiving. Also, the southern part of the capital is the less populated and with less urban density, giving the false idea that the Access Points are serving only half the city.

Figure 25 shows, with less granularity, the information about the distributed Access points. This bigger picture shows an AP density that could preliminarily give a scalable solution to the vehicular communication demand throughout the city.

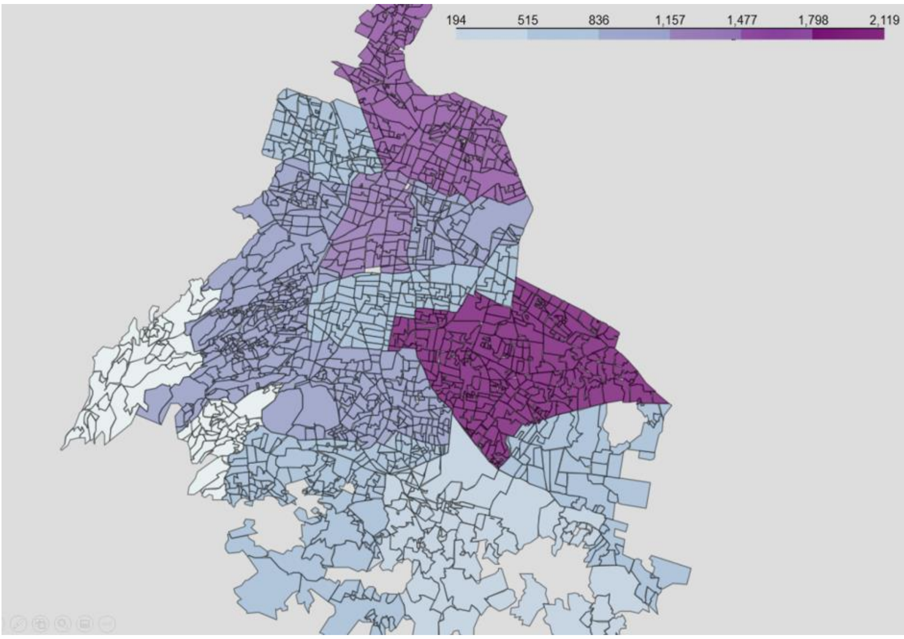


Figure 25. C5-APs distribution cloroplexic Map by County.

In order to have a better understanding and not depend on general statistics, an interactive map to locate Access Points through all the city was programed. A Screenshot of such interactive map is shown on figure 26.

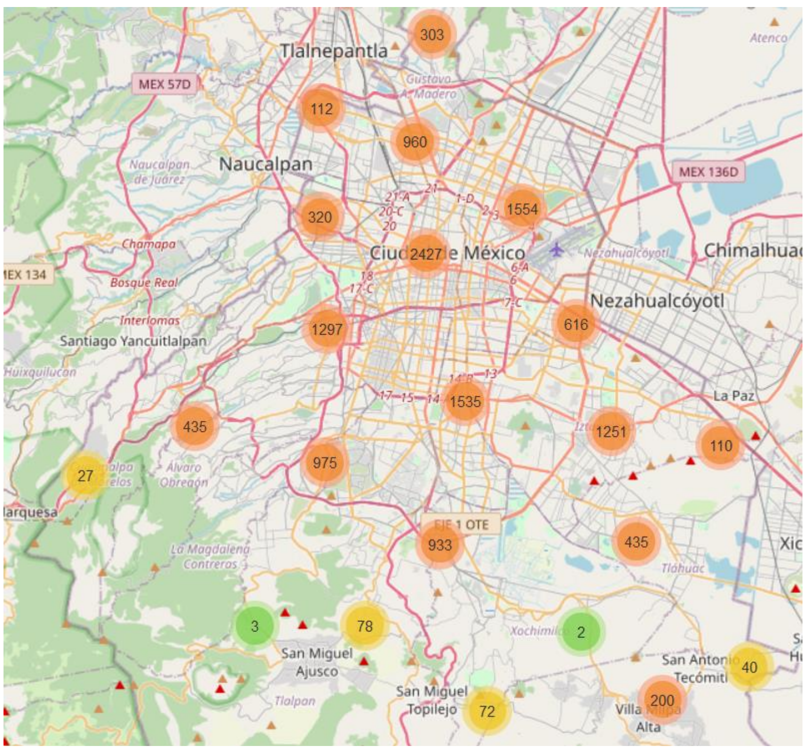


Figure 26. Interactive map for Access Point localization.

Figures 24 through 26 show a bigger picture of the Access Points' distribution. Practically every neighborhood has more than 1 AP, with an average of 7.7 Access Points per neighborhood. This number varies greatly depending on the neighborhood: However, this variation corresponds to the number of inhabitants.

Even though the density of APs could be higher, they are placed where they are most effective.

3.2 Communications

Once we are certain about the feasibility of the project, we need concise metrics regarding the proposed communication system.

For local connection and edge computing data preprocessing, the emulation communication system is proposed with TP-Link modems devices. In terms of specifications, both the TP-Link modems and the C5 Access Points use the same protocol for wireless connection, and thus have approximately the same coverage area and speed.

It is desired that this emulation system can process data with an acceptable level of latency and range. Lacking the possibility to test directly on the C5 equipment due to the absence of legal permits to do so, a local area network is implemented with these devices.

This section focuses on the quality of the connectivity of these emulation devices that operate with the protocol 802.11b/g/n. In this section, several experiments and measures are presented, from maximum distance to LOS vs NLOS scenarios comparisons, as well as different latency cumulative distributions varying the distance in a straight plain avenue.

Subsection 3.2.1 presents all the preliminary measurements to configure the simplest local area network possible, and to test latency, packet loss percentage and signal strength between an AP and an end device.

Subsection 3.2.2 presents a more sophisticated approach to building a LAN with enhanced communication capabilities, bringing physical additions to the network in order to increase its coverage area, and communication between APs was also tested.

3.2.1 Preliminary Measurements

All the physical equipment used in this preliminary experiment can be found in Table C-1. All three devices work with the 802.11b/g/n protocol, as well as 802.11ac. Because this is a test for maximum distance and signal reception, only the

802.11b/g/n protocol will be considered as it has the longest range given its lower frequency band, and is least sensitive to a non-line of sight (NLOS) scenario.

Figure 27 presents the logical connection of these devices to test the reception sensitivity and connectivity depending on the distance and LOS and complete NLOS scenarios. This network is simply to check the layer 1 connection and capacities between the two devices that connect wirelessly. Both the Raspberry Pi 3B +, and the TP-Link model Archer 5 router.

This scenario in which a static local network with only 3 IPv4 addresses with 2 end devices running different operative systems was implemented. A software tool for signal strength checking on the laptop and the built-in Linux tools of the raspberry were used to report the results.

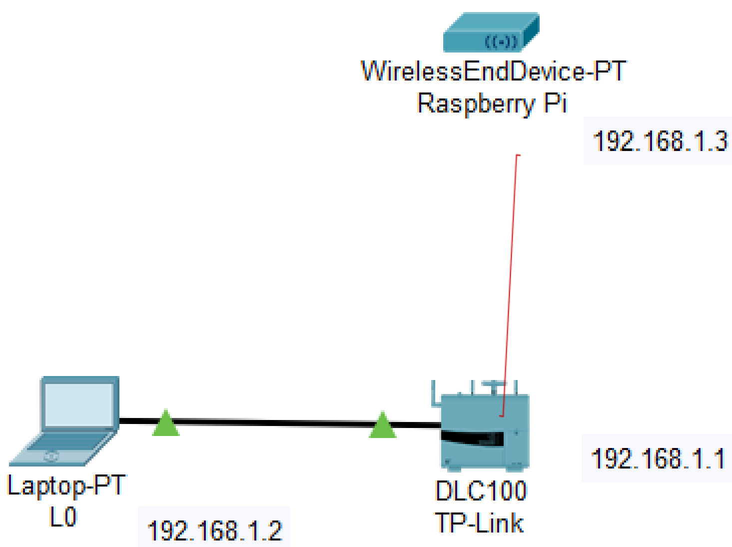


Figure 27. Preliminary logical network.

A completely plain street was chosen to be able to carry out tests with Line of Sight of up to a maximum of 140 meters. Due to the shape of the street, and the low traffic flow, it is a scenario where the tests are completely possible to perform.

For the NLOS scenario, the vehicle's OBU device was hidden behind another vehicle, simulating a situation where a car blocks the line of sight. Figure 28 presents the satellite view of the street with the software measurement of its largest unobstructed distance.

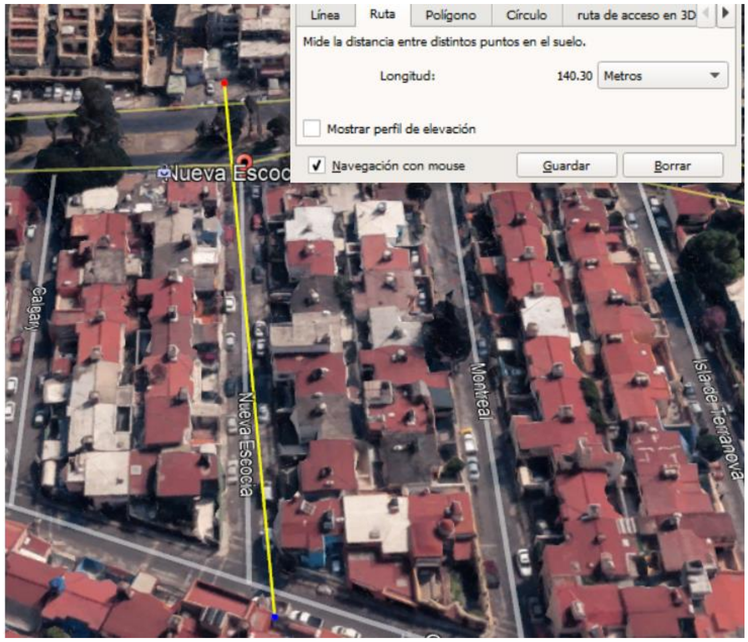


Figure 28. Sattelite view of the chosen street with maximum LOS distance.

All tests were made without any climate inconvenience, and with no vehicle interference whatsoever in LOS testing. The devices were elevated to a distance of approximately 1.8 meters for LOS, and 1.2 meters for NLOS.

It was verified that the signal coming from the router had an adequate intensity, and completely visible by the receiving device. Since Wi-Fi operates within unlicensed frequency bands, it is expected to have a considerable amount of devices interfering with the signal.

Figure 29 presents a plot with the different access points' signal strengths found on the street, showing the variety of devices that share Wi-Fi channels. A total of 14 devices were active, with approximately 10 devices operating on the same channel.

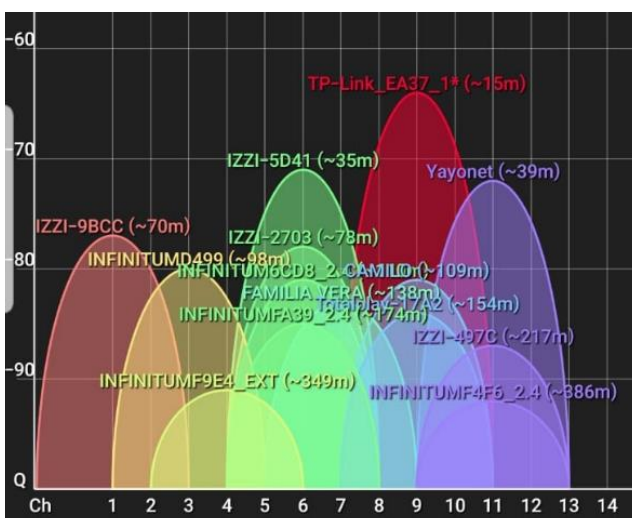


Figure 29. Measurement of signal strengths [dBm] at the street center.

IEEE 802.11a/c or b/g/n have a guaranteed 50 meters connectivity radius with line of sight, looking for its average signal strength. This process was repeated for an NLOS scenario, and the entire process was iterated for 50,60,70,80 and 90 meters. A Wi-Fi strength signal software was used to obtain the sensitivity, and the evidence can be found in Appendix C.

Once the average signal strength was captured, a total of 50 packets with 1KB of information were sent in each position. This experiment was repeated for the following distances: 50,60,70,80,90 meters, both with LOS and NLOS. The results are shown in tables 7 and 8.

Table 7. LOS Experiment Results

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
50	-65	2%	12
60	-69	12%	15
70	-82	16%	152
80	-84	26%	314
90	-86	82%	627

Table 8. NLOS Experiment Results

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
50	-66	2	9
60	-74	18	36
70	-84	24	219
80	N/A	100	N/A
90	N/A	100	N/A

Given the high percentage of packets lost at a distance of 90 meters, it is not feasible to consider it an adequate distance for connectivity. In an ideal scenario, the maximum distance in terms of radio that these communications can support is 80 meters with its current capacities. The packet loss at higher distances is unacceptable for this communication system.

Due to the nature of the Raspberry Pi 3B + equipment, it is possible to connect an antenna with high gain via USB to increase this range, or consider the possibility of repeaters to increase the signal coverage, and an alternative of signal repeaters could also benefit the maximum connectivity radius.

The TP-Link devices have an inbuilt software protocol called WDS, which enables a wireless link between devices, with a shared local area network. All these features and possibilities were tested in the following experiment.

3.2.2 Communication measurements

In this experiment the same scenario was presented, with the main difference that the WDS protocol was enabled to test multiple AP devices, an external antenna was integrated to the OBU device, and signal repeaters were tested as well. All the devices used in these experiments are shown in table C-2.

In the WDS topology, a router acts as an AP connected wirelessly to the first. Making two wireless LANs seamlessly connect between routers, sharing the same subnet, which is the main benefit. The disadvantage of this configuration is the dependency between routers, which means that they cannot be separated big distances from each other, and an important amount of the coverage radius of both is lost because they share partial area of coverage. Figure 30 presents the connection of two TP-Link devices connected through WDS.

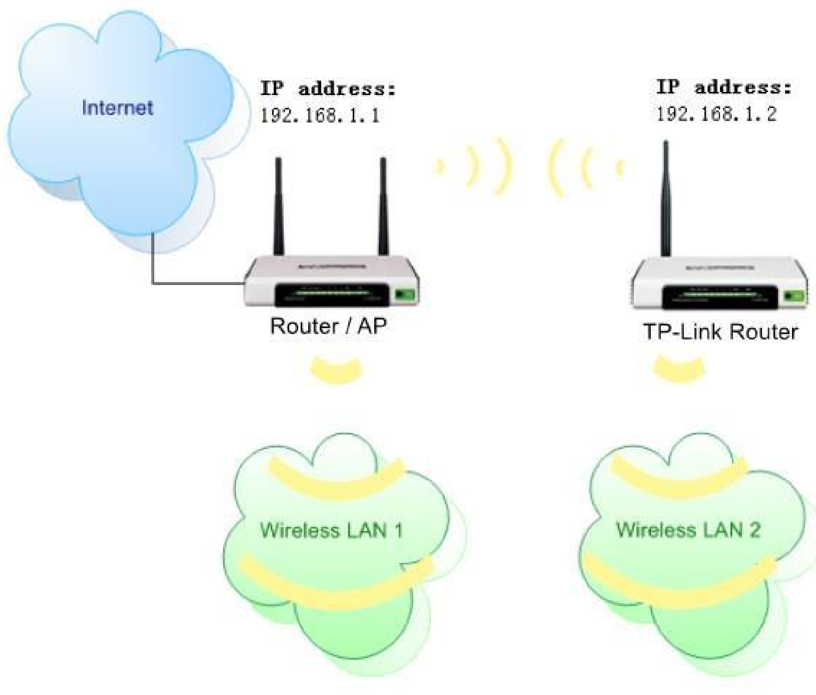


Figure 30. WDS topology with 2 TP-Link devices.

This experiment is divided in two sections. First, the previous experiment was repeated with the new configurations: external antenna and signal repeaters. For the second section of the experiment, the APs were tested on distance and latency, obtaining the cumulative distribution function of packets sent from an end device to a second AP connected through WDS.

In the first section, a total of 50 packets with 1KB of information were sent in each position. This experiment was repeated for the following distances: 50,60,90,100, and 90 meters with both LOS and NLOS tests, comparing the network communication with the external antenna and with signal repeaters. The results are shown in tables 9 through 11.

Table 9. LOS Experiment Results with an External Antenna

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
50	-51	2%	11
60	-47	0%	39
90	-78	12%	81
100	-85	20%	150

Table 10. NLOS Results with an External Antenna

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
50	-62	2%	23
60	-71	10%	48
90	-83	18%	115
100	-86	28%	326

Given the low performance of the signal repeaters, only a LOS scenario was tested. With a 50m radius of distance the percentage of lost packets surpassed 80%, so this implementation was immediately discarded.

Table 11. LOS Experiment results with a Signal Repeater

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
40	-71	4%	11
50	-87	85%	44

Given that the communication between APs for this range had no registered packet loss, more extensive tests regarding its operation were carried out. Table 12 shows how there was no packet loss for 80 and 100 meters in a scenario without line of sight.

Table 12. WDS distance sensitivity

Distance	Signal Strength [dBm]	Lost packet percentage	Average round trip time [ms]
80	-51	0	3
100	-47	0	10

The experiment ran on a real-world physical environment, with two APs connected in WDS mode with distance tests from 70 to 120 m, sending hundreds of packets, with 100 bytes or 1000 bytes of information. Tests were run on a busy avenue at 6pm in the afternoon, and the cumulative distribution was obtained for each package size scenario. In order to find the range of time it took to send the packets between two access points depending on their distance, a python Extract-Transform-Load (ETL) script was generated to read ping logs from text files obtained through a Windows command line, and then each cumulative sum was calculated and plotted.

The end of these measures is to determine the maximum acceptable distance between devices. Figures 31 and 32 show the results for 70 and 120 meters, and the rest of the plots can be seen in Appendix C, Figures C-5 to C-8.

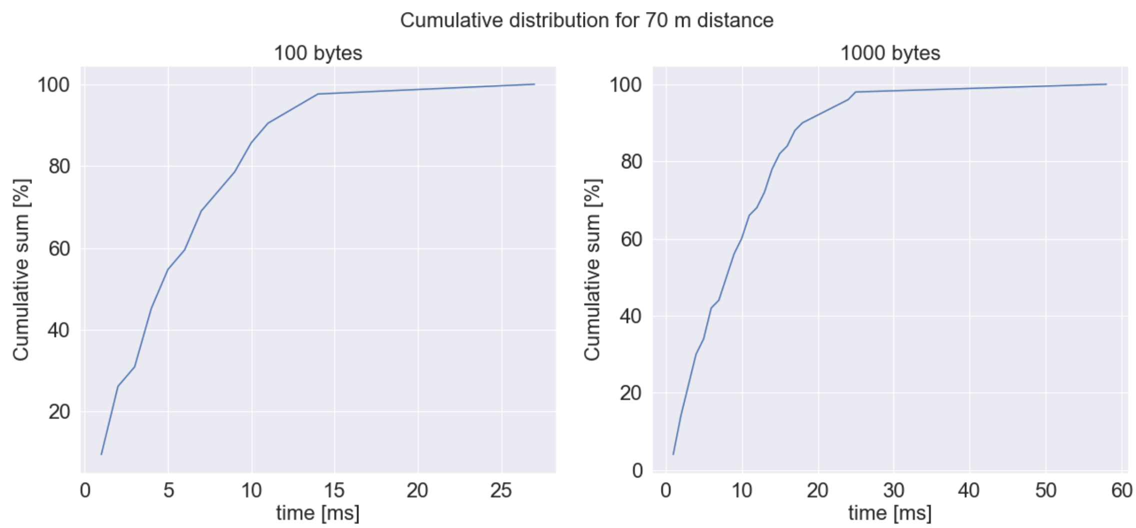


Figure 31. Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance.

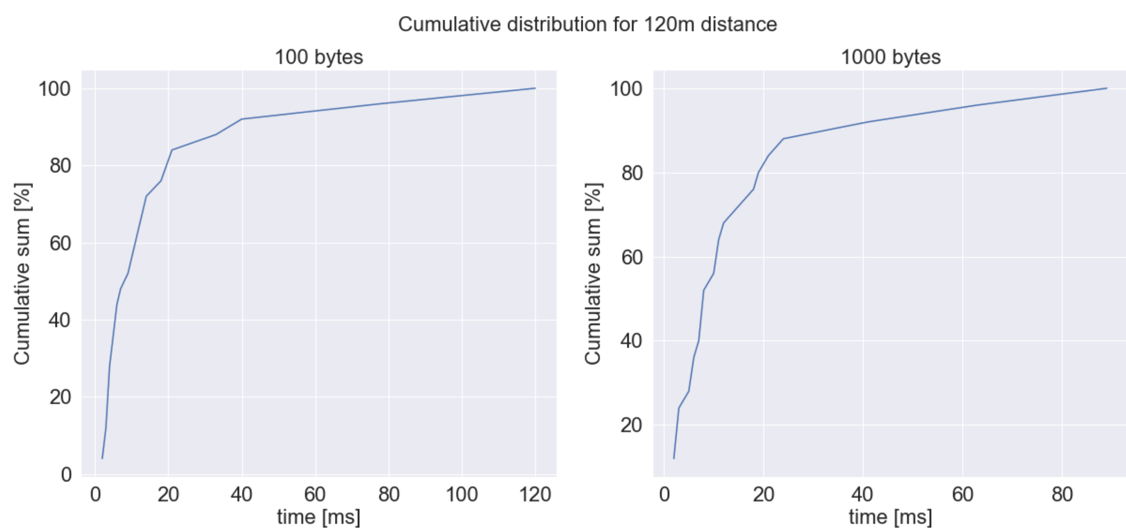


Figure 32. Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance.

The 90% mark can show the maximum time limit where most of the packets will take in a round trip time with 4 logical hops. Table 13 shows the round-trip time for 70,80,90 and 100% of the packets of every trial with 1kB of information sent.

Table 13. Round trip latencies for 1KB packets

Distance	70% round trip time latency [ms]	80% round trip time latency [ms]	90% round trip time latency [ms]	100% round trip time latency [ms]
70	13	15	17	55
80	7	11	13	25
90	11	13	15	33
100	11	12	16	35
110	8	11	18	25
120	16	19	40	90

The 120-distance experiment showed relatively poor connection performance in comparison with lower distances. For the following experiments, the 120-meter radius will be set as the maximum limit in which communications can be done between OBU units and the RSU.

3.3 Distributed Computing and its applicability to Urban Mobility

In order to maintain a low-cost solution, low-cost equipment must be implemented. The main problem with this type of devices is their capacities. Hence, they require techniques of distributed systems, where each computing system has local storage and computation. The benefits of this systems start to show when projects scale.

For this particular dissertation, the amount of data obtained was not enough to encounter Big Data problems. However, if this solution intends to scale up to a City Level, distributed computation systems must be correctly implemented.

Besides virtualization techniques, Hadoop emerged as an alternative to handle distributed databases for its processing, and multiple benchmarks and studies have been done regarding its scalability.

Bajcsy et al. benchmarked hadoop with different CPUs as nodes, with the following characteristics:

Master Node: Intel Core i5-2400 @ 3.1GHz – 10 GB RAM memory

Slave Node 1: Intel Core i7-2600 @ 3.4GHz – 16 GB RAM memory

Slave 2: Intel Core i7-2600 @ 3.4GHz – 8 GB RAM memory

Slave 3: Intel Core i3-540 @ 3.07GHz – 8 GB RAM memory

Hadoop was then tested with a Tera Sort algorithm, which consists in ordering from 1GB, up to Terabytes long array in ascending or descending order. The tests is run varying nodes and array sizes. Figure 33 shows the runtime results with the different nodes described above.

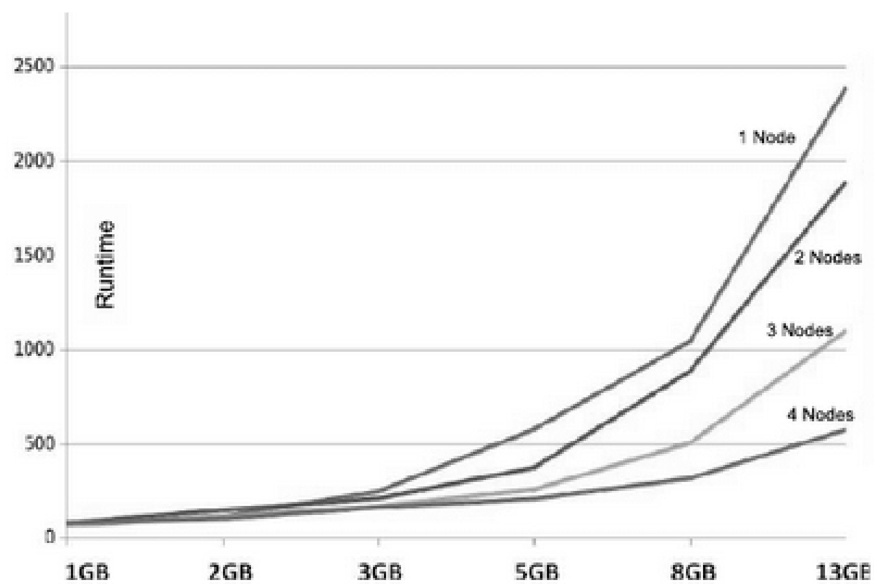


Figure 33. Tera-Sort scalability test runtime on different nodes, from [80].

Tinetti et al. also performed tests on image processing algorithms using hadoop: Flat Field correction, Segmentation based on convolution kernels, and Feature extraction.

These algorithms handle multiple operations to correct, stich, and format different images in TIFF format. The Hardware for each node consisted of 2 to 16 virtual processors, with 4 up to 32GB of RAM with a limit of 800 total nodes. Hadoop was entirely run in a Java environment, within the Linux CentOS flavor.

Figure 34 shows a comparison between the Tera Sort computation algorithm, a hadoop cluster implementation, vs a non-hadoop Raritan cluster with the same algorithm, where it shows a better performance for a Hadoop cluster.

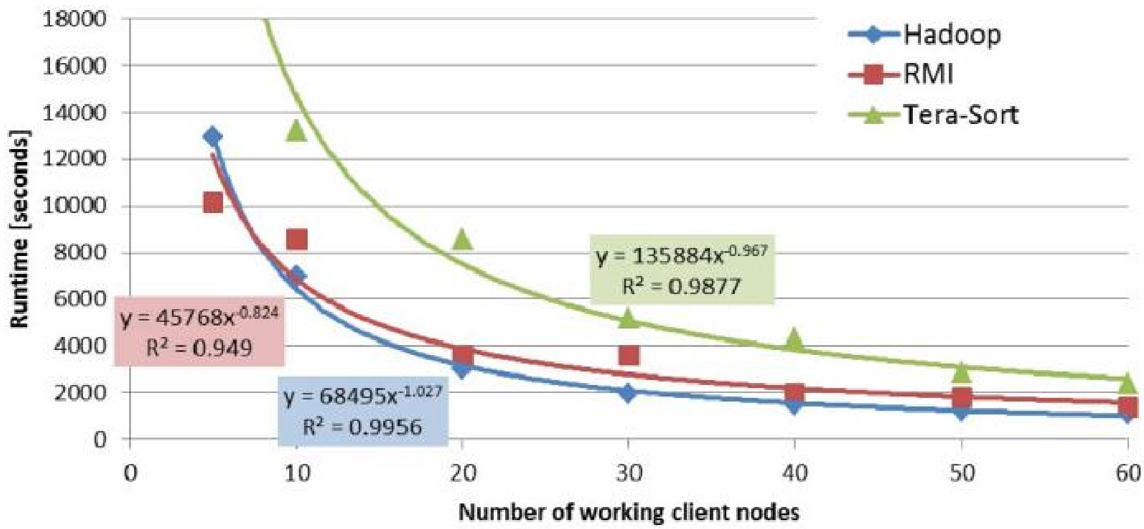


Figure 34. Tera-Sort vs Segmentation algorithm on different working nodes, from [81].

The important part of this plot is to show the major improvement of runtime by the addition of nodes. In both figure 33 and 34, low end computers acted as nodes with Hadoop technology, and in both cases the runtime fell dramatically.

The other piece of software left to test is related to the Machine Learning Algorithms. Since this solution is being entirely built on Python 3.6, two main set of MLA libraries currently can bring the possibility of computation within this whole solution. On one side, the most popular framework for data science: TensorFlow, Pandas, and Sci-kit Learn. On the other hand, the Apache foundation ecosystem that runs on python: Py Spark and MLlib.

Since this solution will be running on low end software, a single cloud instance was created in the IBM platform to test how these libraries compare to each other.

A simple benchmarking test of map-reduce operation was set to compare the computation time for both. It consisted of creating a dataset of random numbers from a N by 2 matrix, and map-reduce it by adding each element of the data frame. The virtual instance consisted of a 3.6 python virtual environment, with 1GB of RAM memory running on a Linux-based Jupyter notebook.

Figure 35 shows a runtime benchmark between both technologies with multiple map-reduce computations.

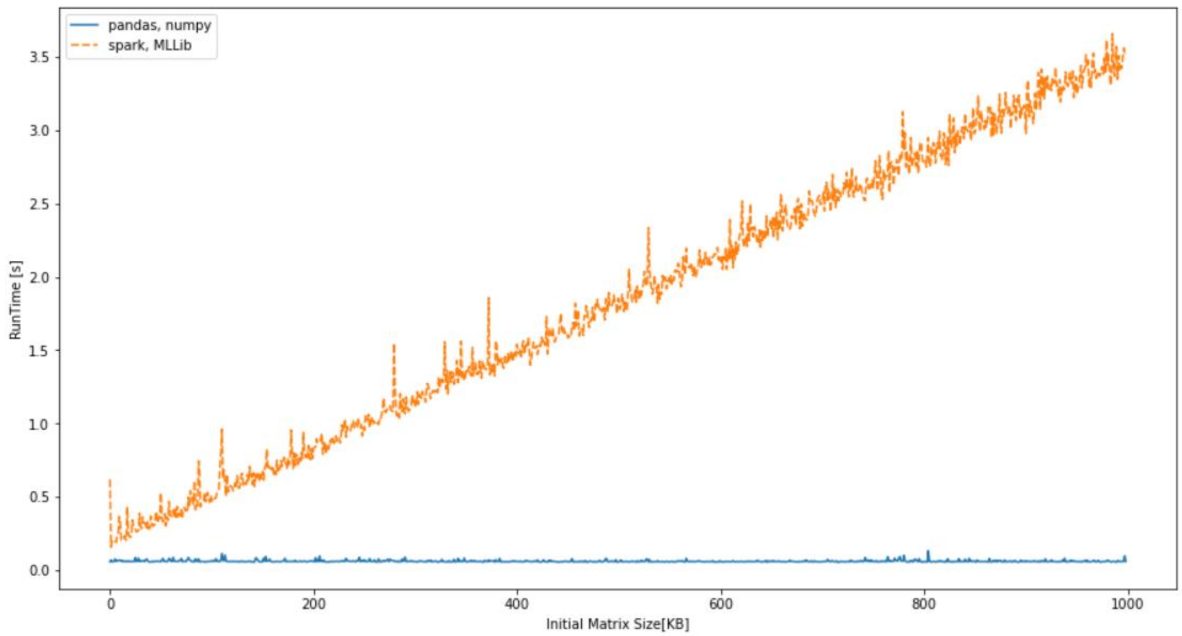


Figure 35. Runtime comparison between Pandas and NumPy Vs PySpark and MLlib.

There are benefits for either set of technologies, however, the runtime proved to be remarkably better for one set technology over the other. Table 14 shows a quick comparison for both set of technologies.

Table 14. Comparison between Data Science libraries and frameworks

	NumPy, Pandas, TensorFlow, Sci-Kit Learn	PySpark, MLlib
Distributed-computing integrated ecosystem	x	✔
Open Source	✔	✔
Size	Big	Small
Runtime over large datasets	Fast	Slow
Vectorized Computation	✔	x

From recent years, several libraries and alternatives have been developed to create faster libraries for data science. However, the most maintained libraries up to date are pandas, sci-kit learn and TensorFlow. Pandas, for example, works with java as an underlay technology, and TensorFlow is currently maintained by Google.

Even though spark and hadoop work in a single ecosystem, there is no library that can run as nearly as fast, with its vectorized algorithms and better programming language decisions.

3.4 RSU-OBU communications and data preprocessing

The goal of this section is to get useful information via RSU-OBU communications emulating the C5 Mexico City system with interconnected mobile Access Points, using as our area of experiment the Tec de Monterrey CCM's nearby streets and avenues.

There are various approaches and possibilities to carry out the experiment. In section 3.3 it was proved that the external LG antennas increase the coverage radius up to 100 meters and up to an absolute limit of 120 meters with an acceptable packet loss and latency, given that none of these communications are for real-time applications. With the proposed experiment scenario, two possible scenarios arise: a series of independent W-LANs, with a coverage diameter of 200 meters, or a shared WLAN through WDS with 120m of separation between each router. The use of repeaters was ruled out due to the short distance they extended to the range (40m maximum), in addition to increasing latency considerably by including more hops between OBU and RSU.

The next question to be solved is whether it is completely necessary to cover 100% of the route, since being an experiment to measure the state of the streets through sensors inside cars, this data can be temporarily stored in the RAM memory of the cars. Raspberry Pi 3B + in which the car is reconnected. Since the communications are based on the IP/TCP protocol, the same mechanism of TCP helps creating a more robust communication.

This section focuses mainly on the sensor integration, and complete testing of an RSU-OBU system that could collect useful information regarding streets and avenues of Mexico City. Based on the architecture presented in section 3.2, and with software built partially during section 3.3, several iterations were done to find errors, optimize software and improve performance. All the network equipment used during the experiments and tests was single vendor, and all code was written in python 3.5+ language. Appendix C shows all the devices used during this section.

Section 3.4.1 presents an initial experiment to understand the network behavior under car movement and communication handover between two different Access Points with the OBU. Also, the preliminary software was tested to find further improvements.

Section 3.4.2 presents a full experiment using a total of 4 Access Points, and all sensors available in the OBU. Several laps were done over two different routes to test the system in a stressed environment, and all the data was saved for further analysis.

Section 3.4.3 presents the recollection of reference data for Machine Learning models training. The criteria for that reference data is described, as well as part of the methodology of data preprocessing.

3.4.1 Initial Experiments

Intelligent vehicles need a huge number of sensors to enable autonomous driving either partially or totally; they need to plan routes, detect pedestrians, potholes, traffic lights, as well as other vehicles to be able to make decisions. This information, in addition to being useful and completely necessary for the vehicle, can become extremely important for a city focused on developing sustainable mobility. The advantages of processing and storing this information outside the vehicle are mainly to inform other autonomous vehicles of the physical state of the road on which they travel, as well as to inform the authorities about the conditions on which their streets and avenues are, in order to make better and more informed decisions regarding the investment of public spaces in order to have better mobility in the city.

To prove if the proposed technology could serve up the purpose stated below, a series of experiments were carried out in two scenarios. The first scenario was a route with a total distance of 711 meters, the second was a route with a distance of 1555 meters, both routes being part of the same avenue, using the 802.11n protocol. Data was collected from a vehicle with a device called OBU with different sensors, to store data from different events during the transfer of the vehicle.

3 different sensors were integrated to recognize images, acceleration in three axes, and the geolocation of a vehicle traveling on the two routes within the avenue. The experiments were carried out with a total of 3 access points, which were strategically placed in different positions on the avenue to cover both routes, based on the max distance determined by previous experiments. All the devices used in these experiments can be found in Appendix C, table C-3.

All the devices were connected using the WDS protocol, in which the routers are wirelessly connected to each other. This keeps any device connected to any of the routers within the same LAN. Figure x presents the AP topology used in both experiment scenarios.

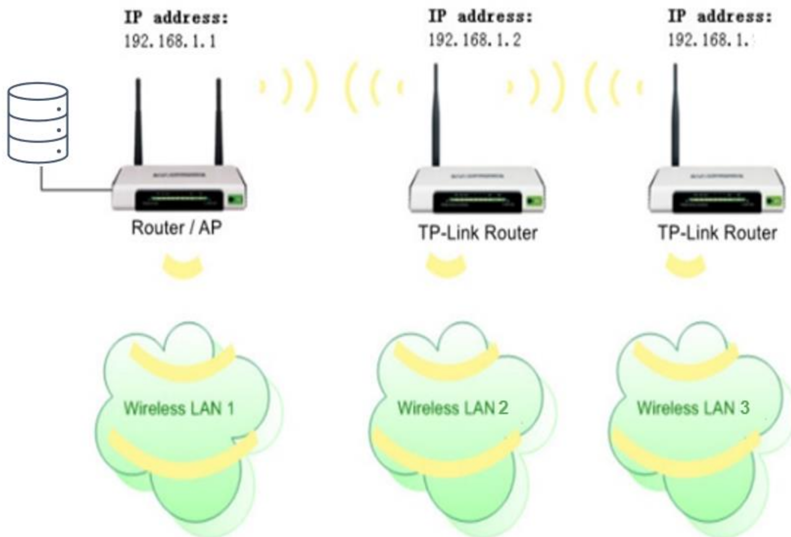


Figure 36. Network topology for 3.4.1 experiment.

The two routes used for the experiment are shown below, as well as the position of the routers on the map. The first route was sought to be within the coverage area for as long as possible, while for the second route the maximum coverage was exceeded maximum coverage area and stress the system to its limits. The coverage radius can be seen in red, and the route taken by the vehicle in yellow.

Figure 37. Route 1, avenue upper section

Figure 38. Route 2, lower avenue section

The avenue in which the experiments were carried out is "Prolongación Canal de Miramontes", around the intersection with street “Calle Puente”, on one side of the campus vicinity. The experiment was carried out on a Friday, between 11:15 am and 1:00 pm, with heavy traffic, close to a construction area. Figures 36 and 37 present a satellite view of the experiment, as well as the placement of the routers in it. During the experiment, light adjustments were made to the placement of the APs to ensure the better connectivity possible.

Taking advantage of the functions of the TCP protocol, and the connectivity via sockets between server and client, packets were sent continuously, with a separation of less than 500ms, containing the GPS position of the OBU with their respective packet number. At the end of the transmission, the packets sent by the OBU and received by the RSU was sent by SSH to a computer, the data were tagged with the respective date, and stored in csv files for further processing.

Using TCP sockets, communications can be kept open and lost messages can be automatically resent. This way it is ensured that the packages that are lost when there is a change of link between the end device and the router, could be resent later.

Socket communication works in a server-client model. The server for the experiment was the RSU, with port 8080 available to receive messages from any OBU that sent information. A dedicated socket was opened for sending GPS data and MPU6050 sensor data.

Images captured by the camera were sent through a different socket. The reason it was implemented in this way was because are only intended to be a visual aid to identify the anomalies detected in the road, while the data to be processed comes from the other sensors and are not completely essential to the experiment.

This system, however, reported a low packet generation. Mainly because software and hardware related performance issues on the OBU. In the second

Table 15. Initial Experiment results

Route	Laps	Sent packets	Received packets [%]	Sensor related packets with error per lap
First	2	150	100%	2%
Second	2	271	58.6%	0%

Despite the improvements made in the past, new areas of opportunity were discovered with the system pushed to a higher stress level.

There was only one major packet loss occasion, however, at least full 50 seconds worth of route information was lost on the second route. This may be due to the asynchronous processing of data collection and sending of images, which saturated the buffer when trying to forward this information to the RSU and not having a connection. The information collected from the experiment shows that the most important areas of opportunity are the following:

• By having asynchronous processing software, the system stops completely when one of its components fails

• The images were taken in high resolution, without any compression algorithm or streaming, thus facilitating delivery.

• Having only two access points, and long-distance routes (693m, and 1293m respectively), led to a point where the current configuration could not hold the stored information for too long.

• There is an important area of improvement in the handling of socket connection with the RSU

Those are the major challenges of the prototype to improve the connection and to be able to have a functional system collecting traffic data, as well as images to improve visualization. Once improvements have been implemented to correct the errors that stopped the experiment, next improvements could point towards the direction of saving processed information in the cloud, create a webapp that connects to the cloud information to visualize the information, create and develop analysis tools for the anomaly detection, and observe the behavior of 2 or more vehicles connected to the network.

3.4.2 Full OBU-RSU V2I Experiment

For this set of experiments, the following improvements were done: image compression, concurrent data collecting in sensors, software improvements, and more APs were set for the experiment,

Figure 40 presents the network topology for the experiment with 4 V-WLANs

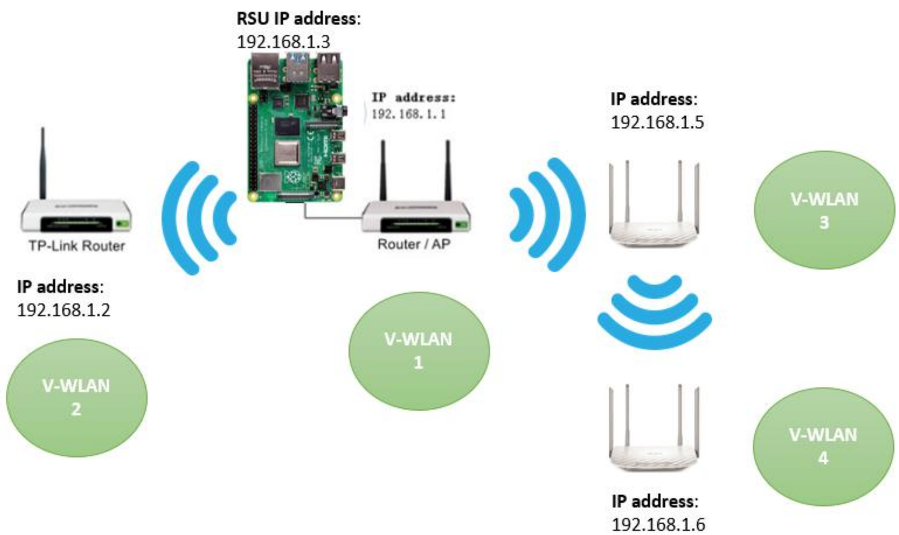


Fig 40. Full Network topology with Raspberry Pi as RSU and TP-Links as APs.

The two routes used for the experiment are shown below, as well as the position of the routers on the map. Due to the limited coverage area of the access points, it was sought to stress the system to find failures in the buffer or in the programming, as well as to measure the capacity to send data in a strained system.

The two routes used for the experiment are shown below, as well as the position of the routers on the map. Due to the limited coverage area of the access points, it was sought to stress the system to find failures in the buffer or in the programming, as well as to measure the capacity to send data in a strained system.

Finally, in Figures 41 and 42 the location where the APs were placed can be seen, each with a coverage radius of maximum 120 m. The numbers on the marks also correspond to the distance they are from the starting point of the avenue at its intersection with Avenida De las Torres.

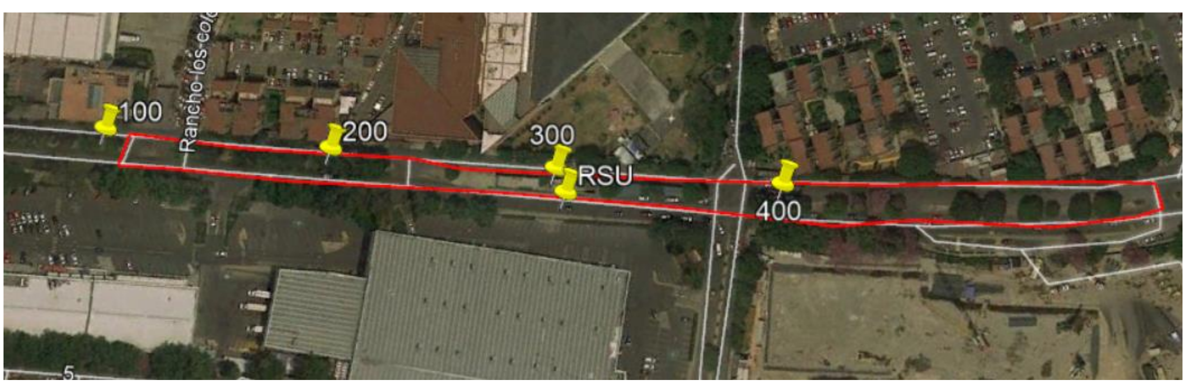


Figure 41. APs marks on the first route of the experiment.

Figure 42. APs marks on second route of the experiment.

Assuming a maximum range of 100m, both routes were carried out with a theoretical coverage area greater than 90%. All the access points were linked in WDS mode and all the information was condensed into a single dataset to be able to process the most important points where there was an acceleration change. 11,985 discretized signal data was obtained from the MPU6050 sensor at approximately 3,595 different points along the paths. Table 16 shows a summary of the data captured in the experiments.

Table 16. Experiment 3.4.2 summary results

Route	Laps	Sent packets	Total distance [m]	Route coverage [%]
First route	5	2182	950	92.6%
Second route	3	1414	880	50%

The software improvements from experiment 3.4.1 to experiment 3.4.1 show a more than significant difference. Figure 43 and 44 present maps of clusters of datapoints. It is immediately obvious that this experiment's data collection was way more effective and with an increased volume of over 900%.

Figure 43. Distribution of packets by area of experiment 3.4.1

Figure 44. Distribution of packets by area of experiment 3.4.2

The experiment demonstrated positive aspects of the network, while showing some areas for improvement. The important thing is to notice the areas of opportunity in the system.

The access points where there was considerably less efficient communication were those corresponding to the second branch, whose IP addresses are 192.168.1.4 and 192.168.1.5. In the first route they were found in the corresponding position at 100 m and 200 m, while in the second at 800 m and 900 m.

Besides being the branch with the greatest physical extension in terms of total distance, there were problems at the electrical / electronic level. The first was related to the battery level. The second, which could have been the product of the first, was a constant deconfiguration of the access points throughout the experiment. It is possible that the communication has had errors due to the stress submitted by the distance between teams.

These situations unleashed an error in the unforeseen code, since the code did not check to have a connection directly to the socket, but rather to have an active connection to send packets. This what caused was that having a connection with access points 3 or 4, but not with the RSU, the packets were lost in the geographical areas where the access points were located.

3.4.3 Reference Data Obtention and Preprocessing

This last procedure consisted in obtaining data from a good reference of both how a street is supposed to be without any anomaly, and how specific anomalies' signals should be. The main reason is to obtain training and test data for ML algorithms.

The street used for reference had very few anomalies, was recently paved, and it is mostly flat in all its extension. *Transistores* street, is located in front of a section of the CCM campus, and it became a suitable candidate because it has the aforementioned characteristics and its proximity to the campus is a great advantage. It was found in this whole street three uneven drains, several speed bumps, and one steep curve therefore, these were the anomalies labeled and used as reference in the training and classification of the visualization system. Figure 45 shows a satellite view of the street, while figure 46 shows an image captured by the OBU and sent to the RSU for storage.

Figure 45. Satellite view of the reference street.

Figure 46. Photograph taken by the OBU unit during experimentation.

For each anomaly detected in this street, multiple runs were made to take samples at different approximate average speeds. The different speeds were 20, 25 and 30 kilometers per hour. Only 2 access points were used since the distance was a controlled situation where it was always measured within the coverage area, and all samples were subsequently labeled by type of anomaly. Table 17 shows the classification of the anomalies taken as reference. In order to have more data inputs, several methods were applied to the raw data obtained. The explanation of that process is found on the next chapter.

Table 17. Classification of obtained data.

Type	unique physical anomaly	total physical samples	total individual sensor samples	total input samples

pothole	3	9	3000	60
speed bump	2	6	4800	95
curve	1	3	2700	54
plain	2	3	4200	84

Lastly, appendix C shows different graphs of the 3 axes obtained by the sensor in the time domain, as well as photographs of different anomalies captured by the OBU during experimentation.

3.5 Data Preprocessing and Machine Learning Algorithm Implementation for anomaly detection

One of the most relevant sections of the entire dissertation is the measurement of ML models, since it proves that all the previous work and data collection does have a clear and practical utility as a final product or service.

Two machine learning models were developed to be able to classify anomalies in Mexico City streets: K-nearest neighbor algorithm and an artificial neural network. Both models have almost similar preprocessing in terms of their input data, and they can accomplish the same sorting task in multiple categories.

This section is completely focused on data processing, training, and an analysis of the performance of the models, which shows if all the previous work has a real applicability.

Subsection 3.5.1 focuses on the pre-processing of the data, where different processes were applied to obtain a larger dataset, the comparison between the time domain and the frequency, mathematical methods applied to the datasets, and the input form and total samples sent as input to both models.

Subsection 3.5.2 corresponds to the breakdown of the models and their characteristics, their training, and their results in terms of precision, and F1-Score, making a visual comparison with scatter plots and confusion matrices.

Finally, subsection 3.5.3 corresponds to the application of these same algorithms to classify the data from experiment 3.4.2 and 3.4.3, to visually show the difference between a newly paved street, and one with multiple complications and anomalies such as was the avenue of experiment 3.4.2.

3.5.1 Data preparation and preprocessing

In order to have a training and test dataset, they agglomerated the different anomalies into 4 different datasets, depending on the type of anomaly: healthy, pothole, bump, curve.

Because each anomaly had a significant number of previous points and subsequent points, multiple time offsets could be made, in order for the system to register the anomaly as if it were a different one.

Taking advantage of the fact that one of the major differences between the speeds at which the route was traveled is the power of the voltage obtained, different samples can be generated with the combination of a time offset and a small increase or decrease in the y-axis of the signal.

Applying these two methods on the data obtained, and putting them all together in different datasets in a random way, was how the number of samples was increased artificially. All datasets were exported as comma separated files for machine learning algorithm training.

Finally, a comparison was made between the signals in the frequency domain vs the time domain to have a visual reference of which signals visually have the greatest difference between them, as a preliminary measure on the preprocessing of the data prior to training. by ML.

Figures 47 and 48 show the difference between different anomalies in the time domain and the frequency domain with just a simple Fast Fourier transformation (FFT), considering only the real part of the FFT. The difference in frequency domain is much more noticeable than time, so it may be an indication that applying a Fourier transform to the signals may be a good step in preprocessing.

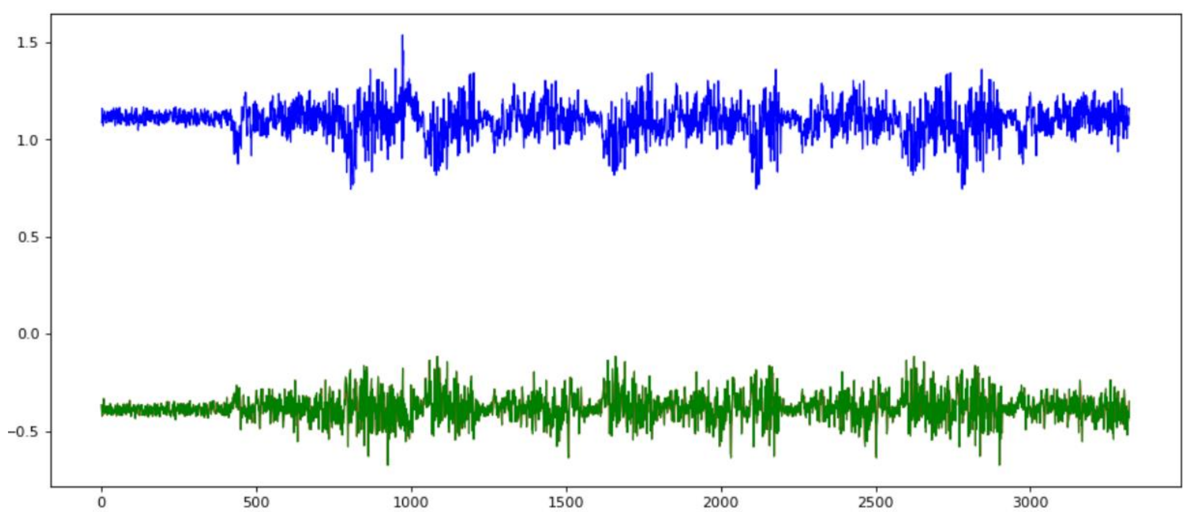


Figure 47. Time domain with multiple anomalies dataset.

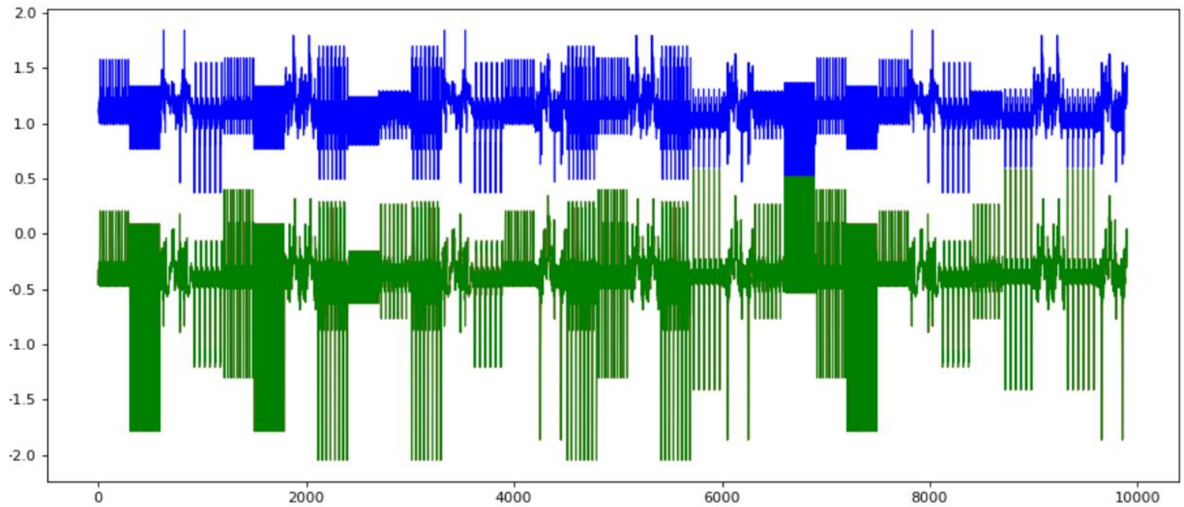


Figure 48: Simple FFT real transform over anomalies' dataset.

It should be noted that the sampling frequency of the MPU6050 sensor is 1KHz, and the frequency of anomalies at a low level of driving speed gives room to maneuver to apply a low-pass filter to eliminate possible noise in the signal and have models with a Higher F1 Score. However, the elimination of high frequencies in the three signals was not carried out in this project. The only frequencies removed were those greater than half the sensor sample rate as they do not have a valid representation due to the Nyquist frequency limit.

There are multiple ways in which the data can be transformed, not only in terms of frequency but in the power of the frequency, phase shifts, etc.

The power spectrum is commonly defined as the Fourier transform of the autocorrelation function. In continuous and discrete notations the power spectrum equation becomes:

$$PS[m] = \sum_{n=1}^{N} r_{xx}[n] e^{-\frac{j2\pi mn}{N}} \qquad m = 0,1,2,3 \ldots N$$

[82]

the power spectrum can be calculated using the FFT algorithm. The time-domain signal x(t) contains N samples, and n refers to the sample number (total sampling time of T = NΔt).

Here, the lower integration limit starts at t = 0 in order to account for causal signal behavior. For any time series, the series' autocorrelation function can be used to calculate the power spectral density in terms of autocorrelation.

$$S(f) = \sum_{k=1}^{N} R(k) e^{-i2\pi fk}$$

$$R(k) = \frac{1}{N} \sum_{n=1}^{N} x(n)x(n-k)$$

Appendix C contains a pseudocode to obtain the power spectrum based on the FFT. There are multiple frameworks and libraries in multiple programming languages. Thus, the FFT will not be discussed in this section.

Figures 49 and 50 show the power spectrum obtained from a signal without anomalies in the street, and the power spectrum from a speed bump. The difference is immediately detectable, not only in terms of the frequency peaks, but in their amplitude and frequency.

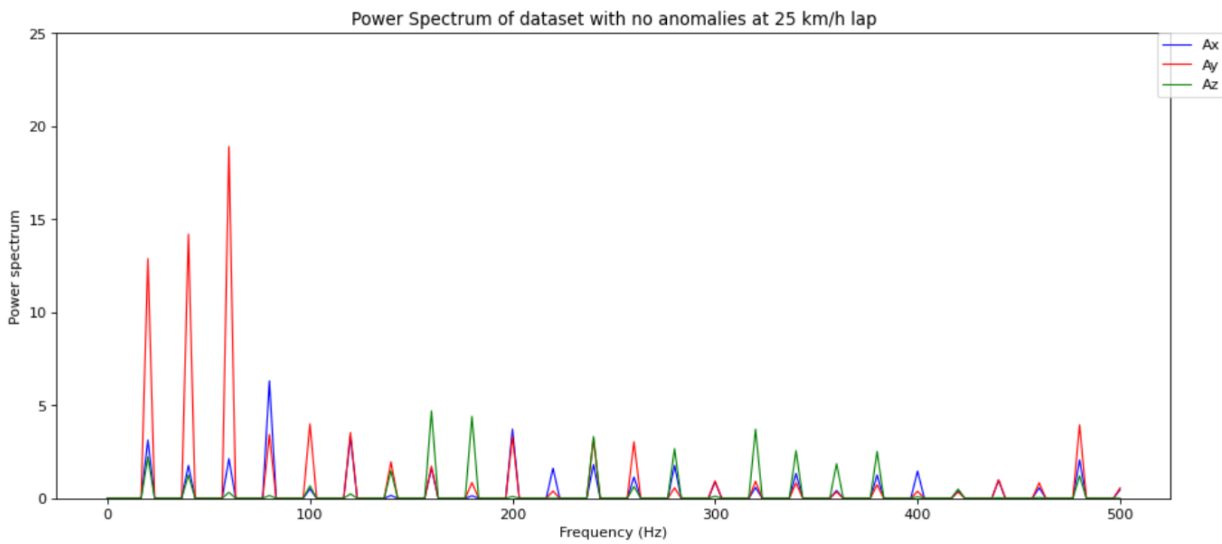


Figure 49. Power spectrum for non-anomalous 50 data points.

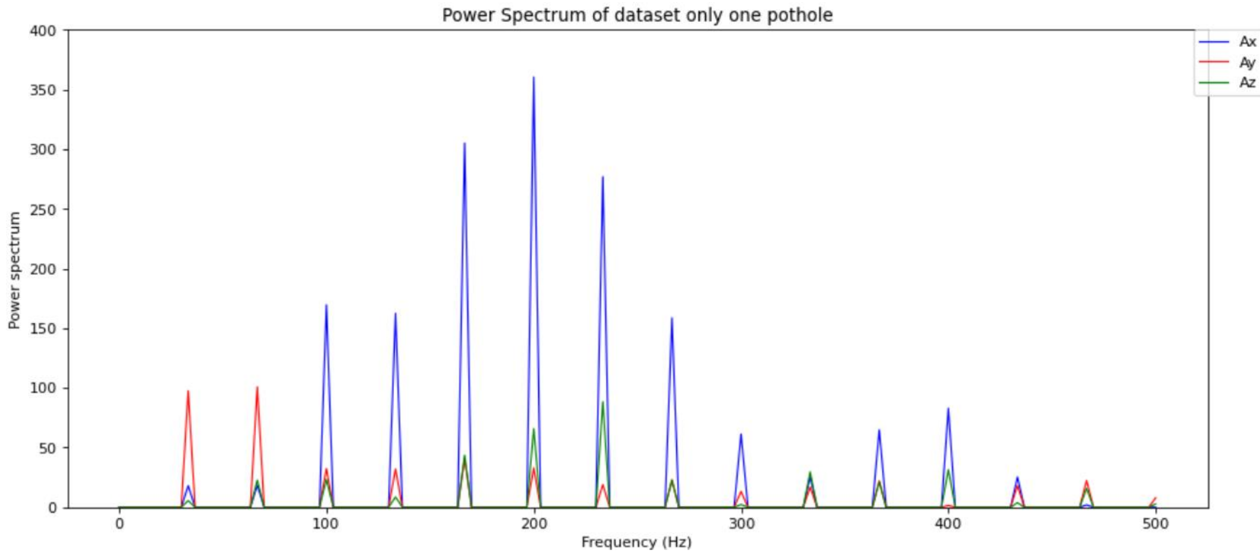


Figure 50. Power spectrum of a pothole detected with 50 data points.

In appendix C, plots for all other examples of anomalies can be found, where it shows big differences every type of anomaly. Not only in terms of amplitude, but in where the spectral peaks are located, as well as the shifts between the peaks.

3.5.2 Models Breakdown and training process

In order to generate a good classification system, two MLA were proposed: K-Nearest Neighbor algorithm, and an Artificial Neural network with a SoftMax layer at the end.

With the classification of table 16, where 4 different types of output are established, the KNN algorithm K constant must be set to 4, while the SoftMax output layer of the ANN must have 4 different outputs.

Regardless of the model, all the data obtained via the OBU must be preprocessed in order to get better classification results. With only a few key differences, the preprocessed data was exactly the same for both algorithms.

Offset removal

Due to a possible offset in the signal, given to the nature of the sensors, as well as the practically constant gravitational force in the z axis of the accelerometer, it was decided to eliminate the offset of the signal by subtracting the mean in every axis. Even though a DC component in a frequency spectrum analysis is irrelevant as it represents a peak around 0 Hz, it represents noise in the input of both models and was decided to be removed.

Time-domain to frequency-domain transformation

Based on the procedure presented on 3.5.1, the signal of every axis was transformed into power spectrum, and every frequency superior to the 500 Hz threshold was ignored.

Data scaling

In case of the artificial neural network, additional scaling was done to improve the training process using a min-max scaler from 0 to 1. This is because by the nature of the artificial neural network, scaled data tends to improve training given the great amount of matrices multiplications. This is the main difference between the input of both models. The KNN algorithm classifies information based on different types of distances: Euclidian distance, cosine squared distance, etc. And the scaling of data does not represent any difference.

Input shape transformation

For every 50x3 data points provided by the OBU, the signal was repeated 5 times to show more easily the power peaks in the signal, the offset removed, the power spectrum obtained, and a single sample with 453 total features is set as the input for both the neural network and the KNN ML algorithm.

KNN architecture

The KNN algorithm is quite simple and direct. The classifier considered 4 different kinds of neighbors, with a uniform weight consideration, meaning that all points in each neighborhood are weighted equally. This algorithm was built with python's library Sci Kit Learn.

ANN architecture

The ANN was built using the TensorFlow library for python. It is a sequential model with four hidden layers, 3 Dense hidden layers with 5, 10 and 5 neurons respectively, and a SoftMax layer with 4 outputs.

After previous trains and tests, it was found that the neural network was rapidly overfitting the data. In response to that behavior, the original number of neurons was reduced, and a Dropout Layer with a rate of 0.15 was added. This ensured that the model behaved in a more accurate manner when presented with unknown data.

To train the network, a batch size of 50 was used, and 15 epochs for back propagation were set. While the batch size number was arbitrary and iterated for performance reasons, the number of epochs was set low given the fact that the signals were clearly identifiable, the training dataset was small, and overfitting was to be avoided. For the same reason of the small dataset, validation data was not provided to the model.

Two optimizers were tested to improve performance. Gradient descent with momentum, and the Adam optimizer, and the differences were abysmal. The Gradient descent with momentum was not able to create an accurate model and thus, was discarded.

Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is appropriate for non-stationary objectives and problems with very noisy and/or sparse gradients, and previous empirical results demonstrate that Adam works well in practice and compares favorably to other stochastic optimization methods. [83]

Given the nature of the data, a Categorical Cross-Entropy loss was used as the parameter to optimize. Figure 51 shows the full architecture of the neural network.

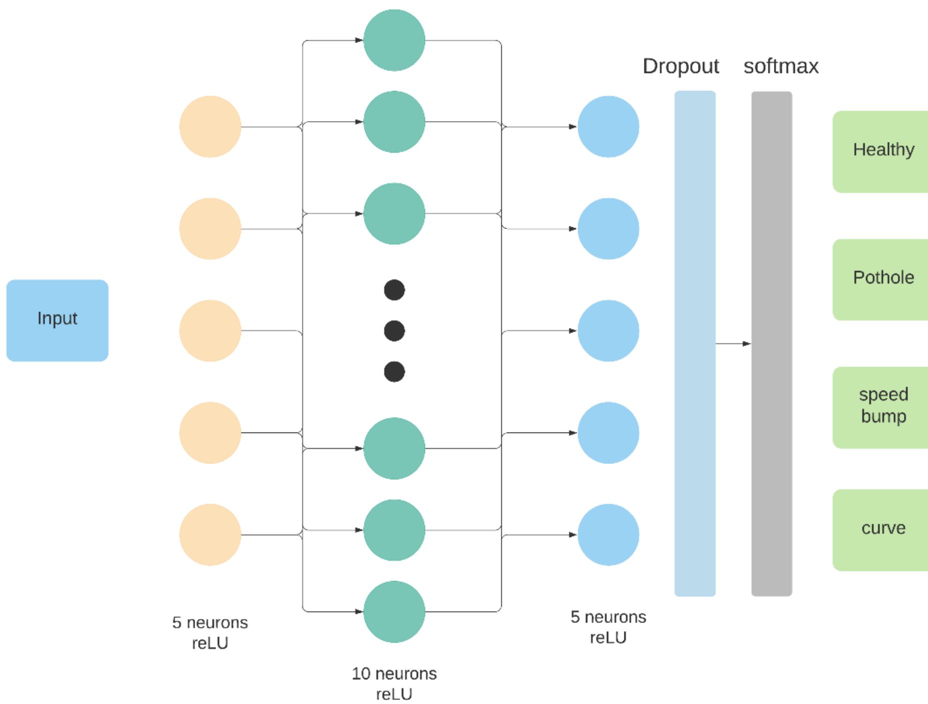


Figure 51. ANN network topology

Lastly, both algorithms were given 85% of the whole dataset to train and 15% to test. These tests consisted of accuracy for training, and F1-Score for test data, as well as a confusion Matrix to determine in which specific classifications they did not perform well.

Comparison

Both algorithms were trained, with highly acceptable results. Both algorithms had over 90% accuracy during training, with the KNN having a 95.55% and the ANN scoring a result of 96.79% over training data.

Figures 52 and 53 show the confusion matrices generated with the test data. While the KNN algorithm did have minor errors in the classification of non-anomalous data and curves, the ANN had trouble classifying only curves. In both confusion matrices, the number 0 represents healthy data, number 1 represents possible potholes, number 2 speed bumps, and number 3 harsh curves.

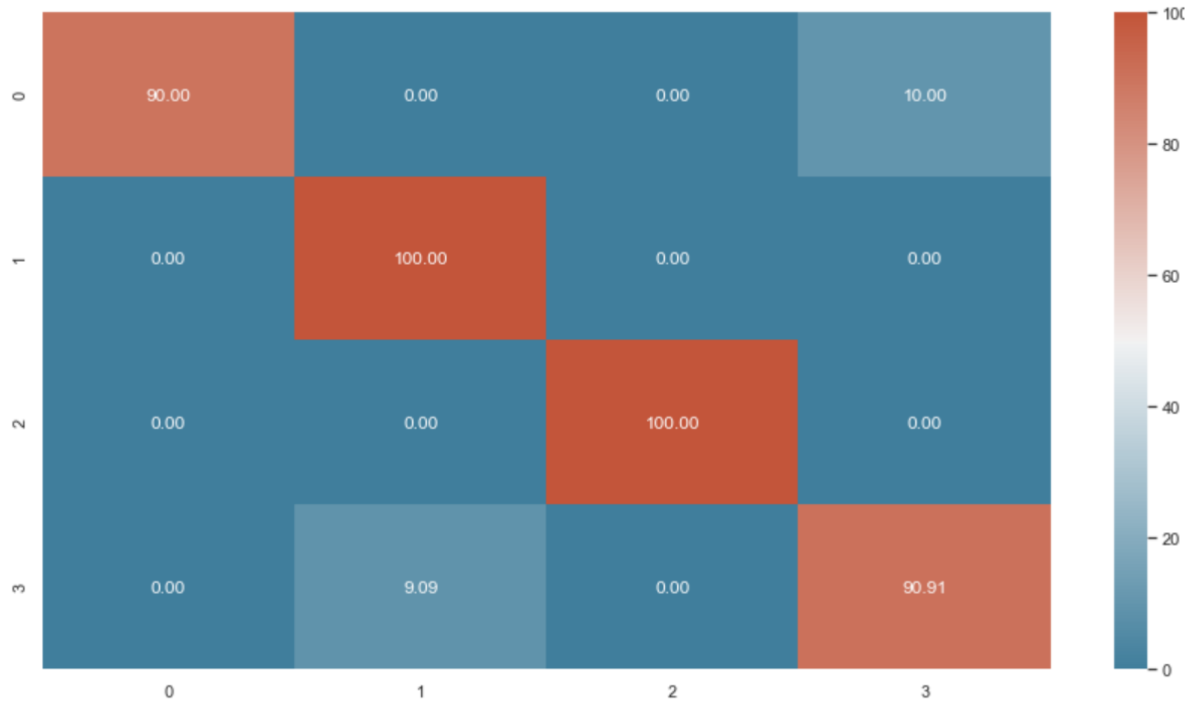


Figure 53. KNN confusion matrix.

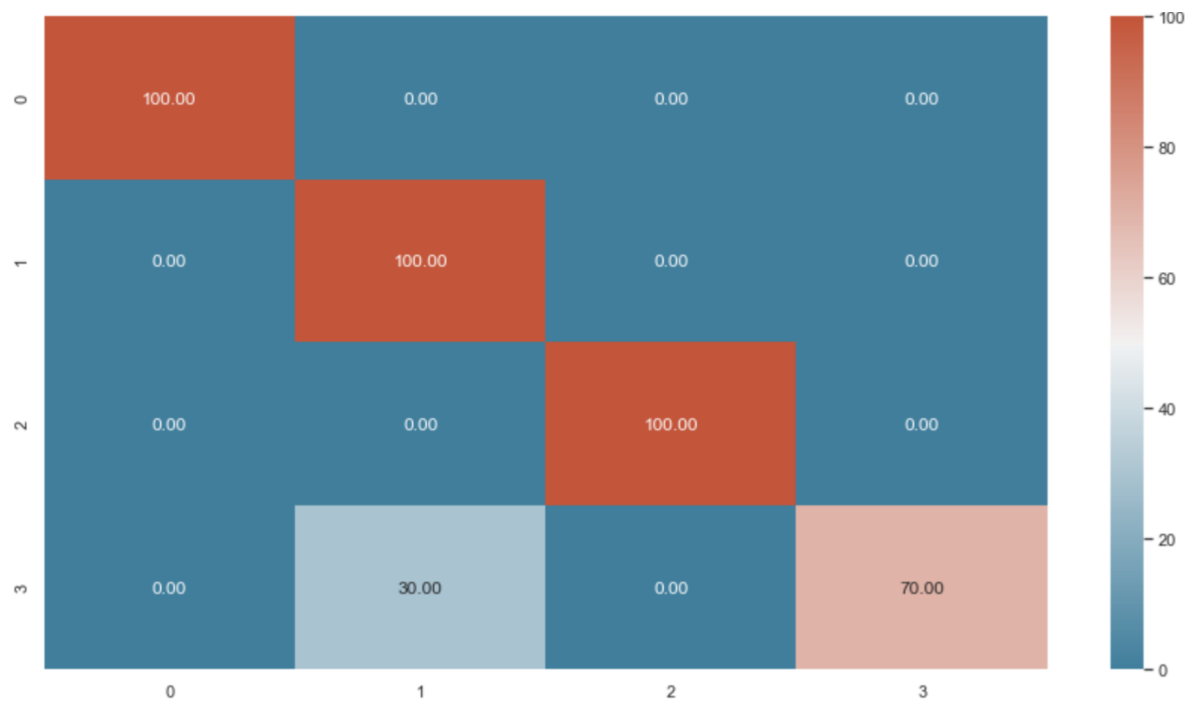


Figure 53. ANN confusion matrix.

The other measurement was the F1 score. This is a measure of complete accuracy, using both precision and recall. Reaches its best value at 1 and worst at 0. The precision refers to the amount of true positives per true positive and false positive, while the recall measures the amount of true positives per true positives and false negatives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$F1\ Score = \ 2 * \frac{Precison * Recall}{Precison + Recall}$$

The results on the test data was coherent with the results of the training. The KNN algorithm scored 97.80%, while the ANN scored 93.33%.

In order to visualize the data presented, 3D scatter plots were done with the labeled test data. Figure 54 shows both scatter plot, done with the mean of the 3 axes of the sensor data, labeled the same as the confusion matrices. There is a complete similarity between both results, and the types of anomalies that differ the most within these visualizations are the speed bumps and healthy data.

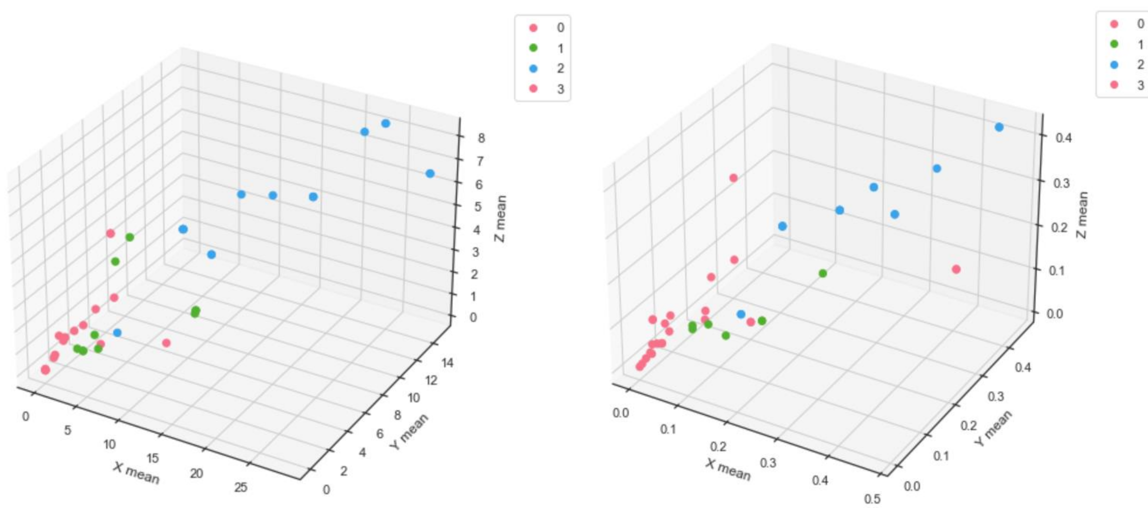


Figure 54. KNN 3D scatter plot (left) vs ANN 3D scatter plot (right).

3.5.3 Classification on Experiment 3.4.2's Data

The same preprocessing for each model was done with two different datasets: one composed exclusively from random sections of the reference dataset, and another with multiple laps from the 3.4.2 experiment. The idea is to visualize completely how the streets would look and be classified with these models.

The KNN classifier worked in a very acceptable way. While the reference avenue was correctly detected, all the data obtained from the 3.4.2 experiment seems approximately correct. The red zones from figure 55 results coherent with the observations, and that avenue was specifically chosen because the amount of anomalies presented.

Figure 55. KNN geo-spatial classification

The ANN on the other hand, labeled incorrectly the street with which the model was trained, and seemed to mislabel some of the information of the 3.4.2 experiment's data. Figure 56 shows significant differences with the KNN classifier.

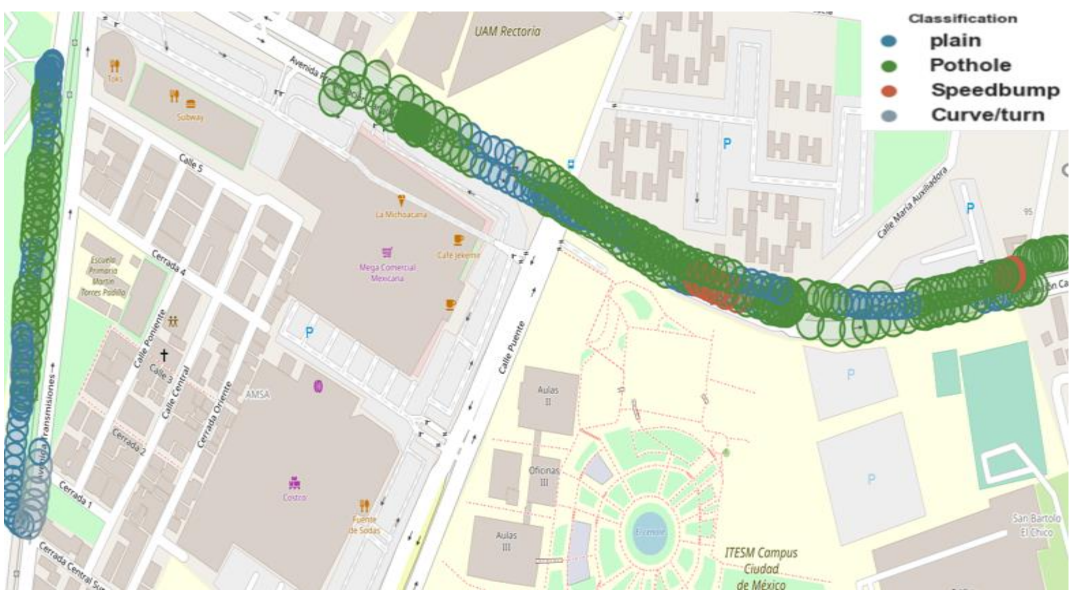


Figure 56. ANN geo-spatial classification

Where the KNN classifier failed the most was with the mislabeling of other classes, or neighbors, into curves. However, with only one real physical example, any kind of data that its power spectrum varied more, it was more probable that it fell into that category.

The ANN SoftMax classifier failed to accurately detect a great amount of the data without distinction. One of the main problems of the classifier could possibly be the sensitivity to scaling.

SoftMax showed not too convincing probability for each of the classes in most of the cases, and scaling could potentially be the main issue in the classification. Since the anomalies were more frequent, greater, and different to the training dataset, the amplitude of the signal of bigger anomalies probably caused troubles translating into correctly scaled signals for the ANN to recognize.

Even though both algorithms were trained in a completely similar form, and the test of both algorithms resulted in great performance results, the data obtained from the street reference was not completely similar to the experiment's data. This showed clearly in the result of the classification. A deeper analysis of this situation is presented in chapter 5, but there are several reasons as to why this happened, including labeling, human error, lack of training data, etc.

4. Results

This project has results in multiple levels, from layer 1 up to layer 7 of the OSI model. Thus, the results are divided by the different sections of chapter 3.

Section 3.1 Results

The analysis showed a strong correlation of 0.89 between inhabitants per county and AP density, meaning that the distribution of the population matches the distribution of Access Points. However, Mexico City lacks sufficient APs to cover completely the city.

This system would not cover the entirety of the capital's area; however, it will cover the city evenly, with 7.7 APs and a total of 292,702 square meters per neighborhood.

Section 3.2 Results

Considering an acceptable threshold level for latency of 15ms in a 2-hop communication between virtual LANs, the major coverage radius for the APs was set to 110 m. In ideal conditions, without any major interference such as moving vehicles blocking de signal, 120-meter radius would be considered acceptable. This sets the maximum distance of connection for the WDS protocol between 2 TP-Link routers.

The WDS protocol proved to be reliable in that network setup, with no major packet losses. However, the system lacked an internet direct access due to the mobility conditions.

Section 3.3 Results

Hadoop technology has been tested over more than a decade, and it has proved several times its efficiency handling Big Data. The benchmarks of section 3.3 proved its significant improvement in runtime while adding different nodes. One important thing to notice is that the hardware of the majority of Hadoop slave nodes used in both cited papers were not high-end technology. This couples with the idea of fog computing, where low-end software computes and processes data in a distributed way, lowering costs an making it a more affordable solution.

The other technology decision was about libraries and framework for data processing. Python was the set language for the project given its facilities and native support from both OBU and RSU technologies, so the data processing decision was between the set of Pandas, NumPy, TensorFlow and sci-kit learn technologies versus Apache PySpark and MLlib. While the former were developed and are maintained independently, the latter are part of the Apache ecosystem, the same organization that currently maintains Hadoop.

The benchmark on map-reduce operations, proved not only that the first technology is faster, but escalates better. This could be due to several reasons, including the underlying software, and that most of that technology is maintained by companies such as Google, and that those become the most popular libraries for Data Science, and have been supported more widely.

Section 3.4 Results

The main result of this section is the V2I network itself. The OBU and RSU were designed and tested in a real environment, with experiments during working hours and major traffic, NLOS scenarios, multiple APs, and moving vehicles with speed limits of 35 km/h.

Major improvements were done with the software of the OBU device, with concurrent script-running, and parallelizing sensor reading instead of waiting on GPS sensor to actually read the position.

The difference between experiment 3.4.1 and 3.4.2 is abysmal in terms of data acquisition, obtaining around 9.1 times more data in the second experiment.

Section 3.4.3 was done in a more controlled environment, with a more stable, plain, recently paved, and straight avenue, with only a few anomalies to measure. As much data as possible was obtained in a period of 1 hour, and that data was labeled and saved for the following section.

Two datasets with accelerometer and GPS data were obtained thanks to the V2I network developed thanks to sections 3.2, 3.3 and 3.4. With different processing techniques, a total of 293 labeled samples were obtained for further model training.

Section 3.5 Results

This section's results are completely tied to the previous, since everything obtained from this section corresponded to experiment 3.4.2 and 3.4.3. Two different datasets were obtained, one labeled and one unlabeled, based on experiments with the V2I network developed in previous sections.

The two different models were trained with acceptable levels of training accuracy and F1-Scores on test. After several iterations and different data preprocessing techniques, Power Spectrum transformation over time-domain signals proved to be the most accurate way of preprocess input features to both models.

The KNN algorithm was trained with a result of 95.55% in accuracy during training, and an F1-Score of 0.978. Comparing and contrasting the classification of the unlabeled dataset, and comparing the classification with human observation, the system proved to be consistent in the classification of both datasets.

The ANN algorithm was trained with a result of 100% in accuracy during training, and an F1-Score of 1 on testing, outperforming the KNN model. However, the classification of the unlabeled dataset was completely inconsistent in comparison with human observation. This model had major problems corresponding to data scaling, and probably overfitting.

Geo-spatial visualization was done to compare and choose a better model for classification of anomalies on the city's streets. The main reason why KNN was a better fit is that it did not require data scaling for its inputs, and thus, handled better a street with mode deep potholes, speed bumps without any paint or signaling, and a generally more chaotic avenue that generated bigger sensor signals.

Maps generated with the folium library can actually be exported, and since they run a technology called leaflet as the underlying software, they can be exported, or even generated with front-end technologies, libraries, and framework such as React, Angular, Vue, or any technology that runs with the Node runtime environment.

All the codebase, Jupyter Notebooks, and exported data in csv format can be found in a public repository made with the purpose of sharing the whole project's internals at [84]

5. Analysis and Discussion

During section 3.1, where the city analysis about was done, mobile alternatives were discarded because of the costs it represents nowadays. However, this project is based on the premise that cellular technologies are not currently available. 5G technologies were delayed due to the Covid-19 global pandemic. And one of its many side effects was the delay of 5G deployment in Mexico. This alternative should not be discarded, as it many countries have started its deployment, and it will become the global standard for IoT communications.

Several Cloud companies such as Azure, Amazon Web Services, or Google Cloud Platform are starting to offer Edge Computing solutions based on 5G communications, and it this technology will eventually arrive to every corner of most cities.

While this happens, mobile alternatives such as 4G LTE or C-V2X could fill the gap that the C5 system does not cover. Mobile service providers do have a much better coverage area for their communications systems and that alternative should be considered.

Section 3.2 to section 3.4 show similar areas of improvement. Starting with the emulation system. The WDS protocol was able to create stable communications but with a low number of APs. During the 3.4.2 experiment, the main obstacle and problem was the communication between APs. Constant reconfigurations had to be made, and since it was an emulated mobile network, problems related to the energy supply also arose. However, those problems would not exist while having a constant energy supply, which is the case of the C5 system.

One major unsolved issue during V2I communication handover. There are visible areas without coverage on Figure 44. This is because of the static IP handling. Other alternatives for a better performance should be considered, as well as software optimization in case of flickering connectivity, or when the communication is changed from AP, even if they are on the same virtual LAN.

For further experiments, multiple RSU clusters should be integrated, with multiple nodes to really measure the benefits of this distributed architecture, since these experiments were done at low scale due to budget constraints.

The C5 system does also have integrated cameras and sensors. The combination of OBU devices with the C5 sensors could also potentially bring new opportunities for distributed computation systems.

On distributed computation systems, multiple alternatives were not explored. Kubernetes, Docker containers, and other virtual machine technologies could potentially help lower costs and benefit greatly the usage of hardware in the RSU.

Data acquisition was sufficiently acceptable to adequately train two ML models, however, a speed limit was set in order to obtain those signals. This means that the vehicles with this integrated technology would only provide useful information when they drive at low speeds.

KNN showed a considerable better performance with unknown data with different amounts of signal change, which was the case of experiment 3.4.2's data. The main reason is that this algorithm is quite simple and does not require data scaling, or even labeled data to begin with. Since there are only 2 non trainable parameters on the KNN model, it is easier to train and shown better results.

ANN is too sensitive to the amplitude, and Deep Learning models rely on bigger training datasets to outperform other ML models. Two alternatives for the improvement of the network are proposed. First, increase the amount of labeled, not just in terms of volume, but with all kinds of anomalies, so the scaling of the whole dataset would correctly represent all the different amplitudes. The second is to fine tune the scaling system, in such a way that the amplitude does not affect excessively the model.

Other problem to consider with ANN is the overfitting of the data. Since the dataset was not big enough to train with validation data, the overfitting of the model could not be correctly detected. However, this problem should also be solved with the proposed improvement.

In addition to the previous improvements, other considerations within the ANN network topology could be improved for overfitting, with dropout layers included in the hidden layers of the model.

Fine tuning must be considered in both classifiers, to correctly detect any anomaly. This include a more thorough classification, data obtention and training. However, the models seem to be a good fit. Is just a matter of the data used to train, test, and when the dataset is sufficiently big, create a validation set to detect if further improvements are needed.

6. Conclusion and Further Research

This whole system proved to be able to detect anomalies and create geo-spatial data visualization to represent the streets condition on two tested avenues. While the proposed research hypothesis proved to be true, this project has vast areas of opportunity, improvements, and further research in order to scale to a system to a complete city level.
One of the untouched aspects of this project is the cloud solution. Design and implementation of cloud services to host the website are needed for the visualization tools, as well as the creation and maintenance of a database, relational or not, and complementary systems necessary for the operation of the proposed data recovery and analysis platform.

More data acquisition should be done, as well as the generation of more datasets with more labels, as well as more precise labels for anomaly detection, including more anomalies, such as culverts, transverse and longitudinal cracks, stoplights, speed bumps, dents, etc. This could be done in the periphery of Tecnológico de Monterrey CCM before expanding to other neighborhoods or counties.

With more data, better labeling, and a fog-cloud solution, experiments, training, and more tests should be done on the MLA proposed in this dissertation.

This project proposed a simple geo-spatial visualization tool for anomaly detection. However, it is yet to be determined the most appropriate, efficient, and useful type of visualizations for the observation of relevant information, detected anomalies and areas of opportunity found in the data retrieved by the OBUs.

Finally, this project could be taken a step further by automatically express recommendations and conclusions regarding urban mobility to the corresponding authorities, besides only showing visualization.

Other recommendations for further research include the analysis of the deployment 5G in upcoming years, join different network alternatives for different types of traffic, software defined networks, and research of other virtualization tools for a better V2I network scalability.

Appendix A

Abbreviations and acronyms

	Description
5G	**Fifth Generation Technology Standard for Broadband networks**
IoT	**Internet of Things**
IoV	**Internet of Vehicles**
DSRC	**Dedicated Short Range Communications**
ICT	**Information and Communications technology**
V2I	**Vehicle to Infrastructure**
V2V	**Vehicle to Vehicle**
V2X	**Vehicle to Everything**
C-V2X	**Cellular-V2X**
OBU	**Onboard Unit**
RSU	**Roadside Unit**
PER	**Packet Error Rate**
BER	**Bit Error Rate**
Wi-Fi	**Wireless Fidelity**
AP	**Access Point**
VANET	**Vehicular Ad-Hoc Network**
LOS	**Line of Sight**
NLOS	**No Line of Sight**
ITS	**Intelligent Transport System**
AI	**Artificial Intelligence**
SVM	**Support Vector Machines**

Appendix A

GPRS	General Packet Radio Services
CDF	Cumulative Distribution Function
FCN	Fog Computing Networks
CCN	Cloud Computing Networks
VCC	Vehicular Cloud Computing
VEC	Vehicular Edge Computing
VFC	Vehicular Fog Computing
YOLO	You Only Look Once
AI	Artificial Intelligence
MCA	Multi-hop Clustering Approach
MEC	Multi-access Edge Computing
RSU-G	Roadside Unit Gateway
IP	Internet Protocol
QoS	Quality of Service
IEEE	Institute of Electrical and Electronics Engineers
VoIP	Voice over IP
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
ETL	Extract-Transform-Load
FFT	Fast Fourier Transform
MLA	Machine Learning Algorithm
KNN	K-nearest neighbors
ANN	Artificial Neural Network

Appendix B

Variables and Symbols

Variable	Description	Units
W	power or radiant flux	$kg*m^2*s^{-3}$
Hz	*Hertz. Unit of frequency*	s^{-1}
BW	*Maximum data transfer rate*	*Mbps*
EbNo	*Energy per bit to noise power*	*dB*

Appendix C

Additional Plots, Tables, and Data

In this Appendix, all the plots, tables, and other additional information is presented. It is divided by subsection, and the only subsections referenced here correspond to previous ___ on its corresponding subsection, section, or chapter.

Section 3.2.1

Table C-1 Preliminary measures devices

Raspberry Pi 3B+	
TP-Link router model Archer A5	
Lenovo laptop	

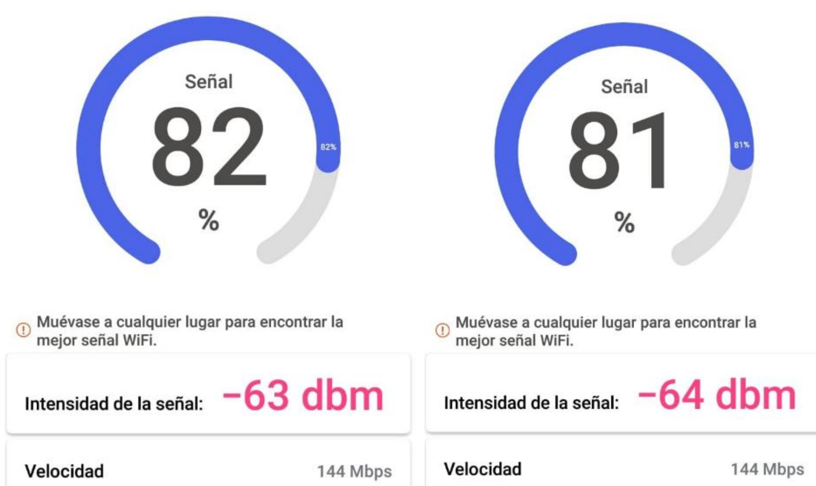


Figure C-1. Maximum and minimum signal level at a distance of 50 meters with LOS.

Appendix C

Figure C-2. Maximum and minimum signal level at a distance of 50 meters with NLOS.

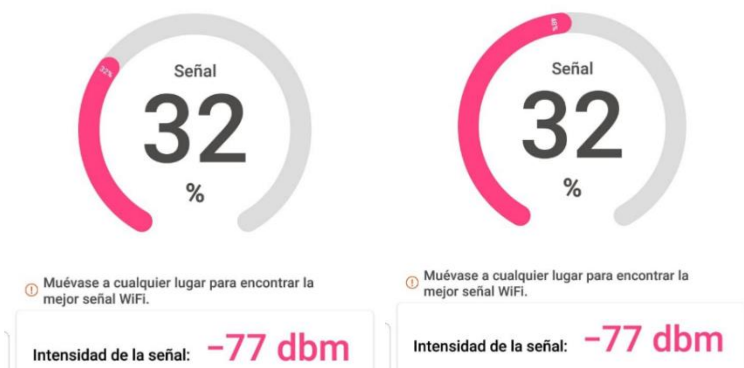


Figure C-3. Maximum and minimum signal level at a distance of 60 meters with LOS.

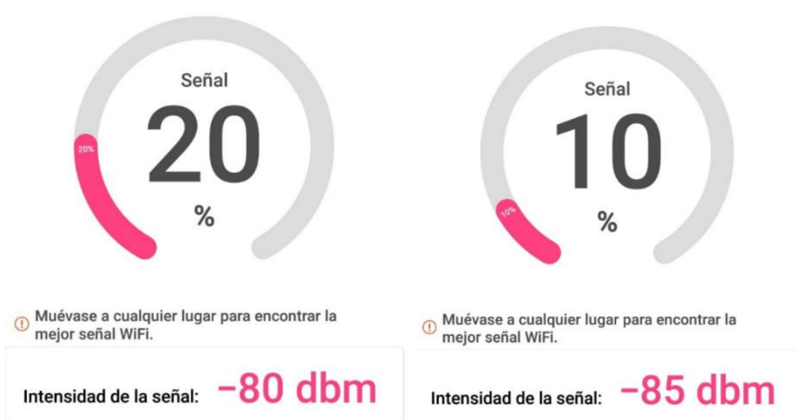


Figure C-4. Maximum and minimum signal level at a distance of 60 meters with NLOS.

Appendix C

Section 3.2.2

Table C-2

Raspberry Pi 3B+	
TP-Link router model Archer A5 (2)	
Lenovo laptop	
Green Leaf external antenna	
Repetidor COM-8200	

Appendix C

Subsection 3.2.3

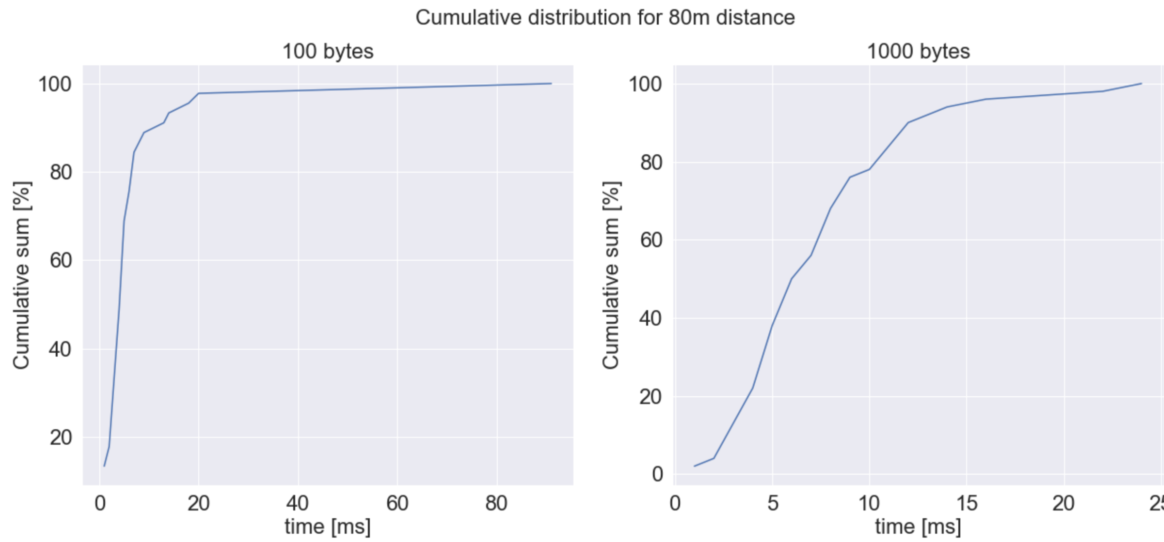


Figure C-5 Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance

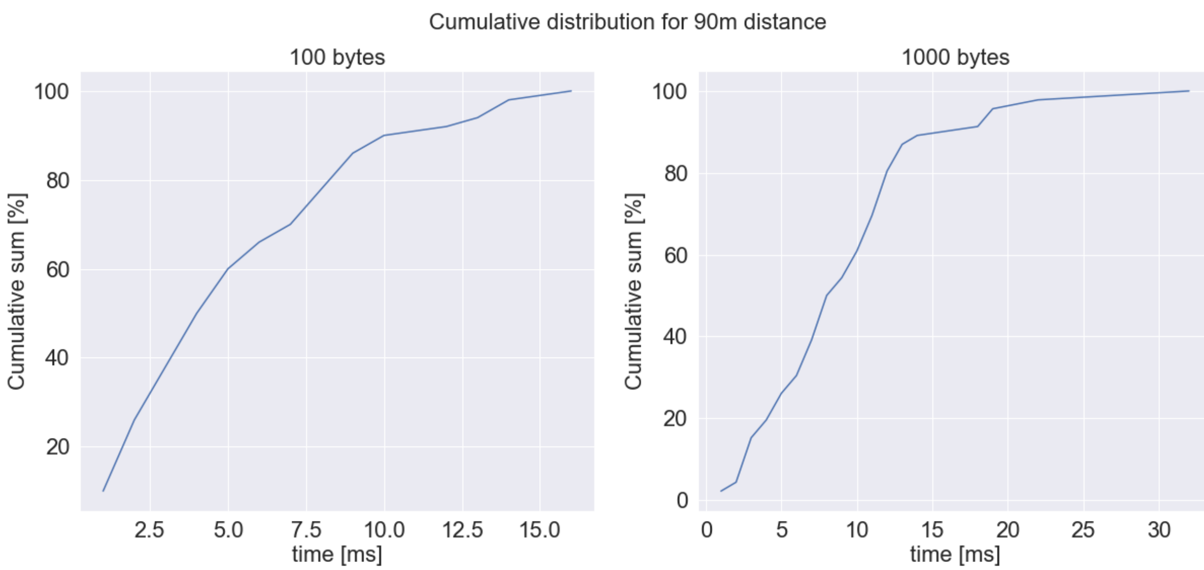


Figure C-6 Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance

Appendix C

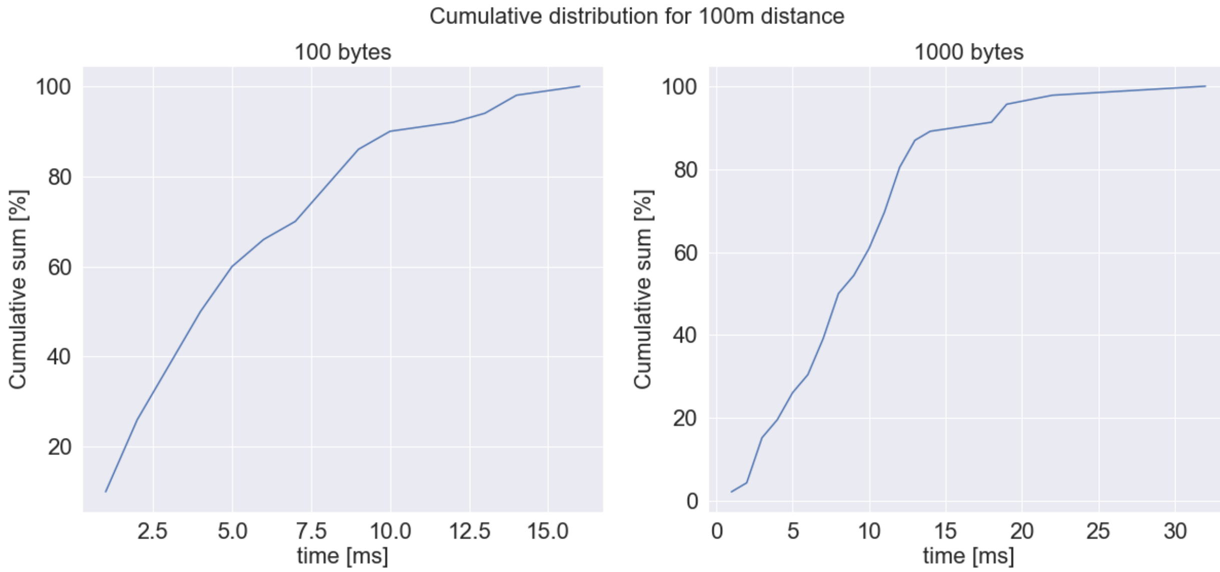


Figure C-7 Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance

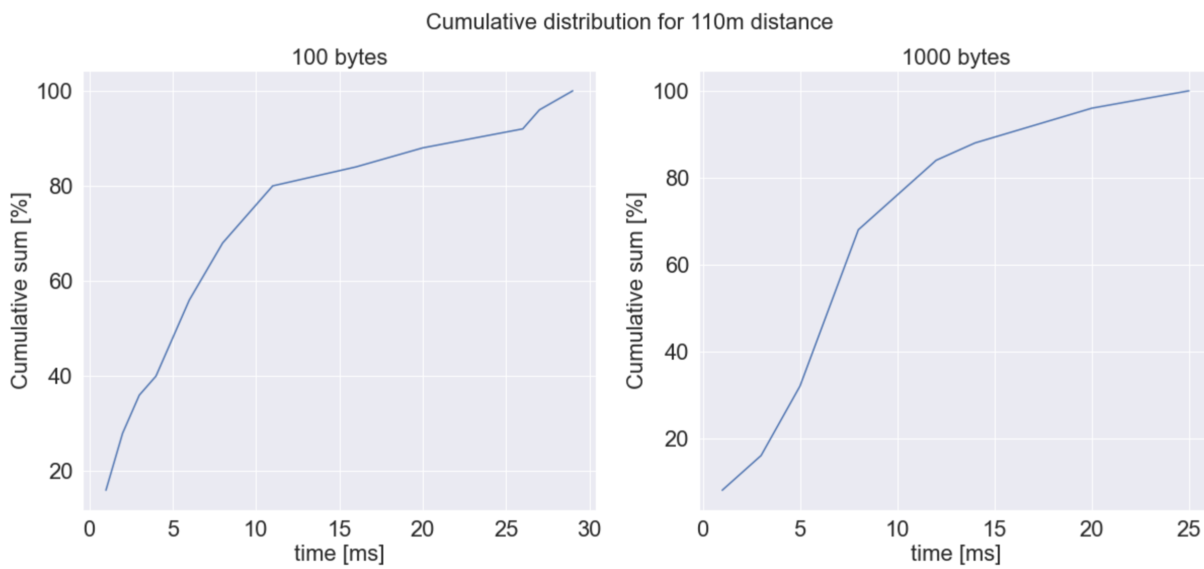


Figure C-8 Cumulative distribution function for 100 and 1000 bytes between 2 APs at 80 mts distance

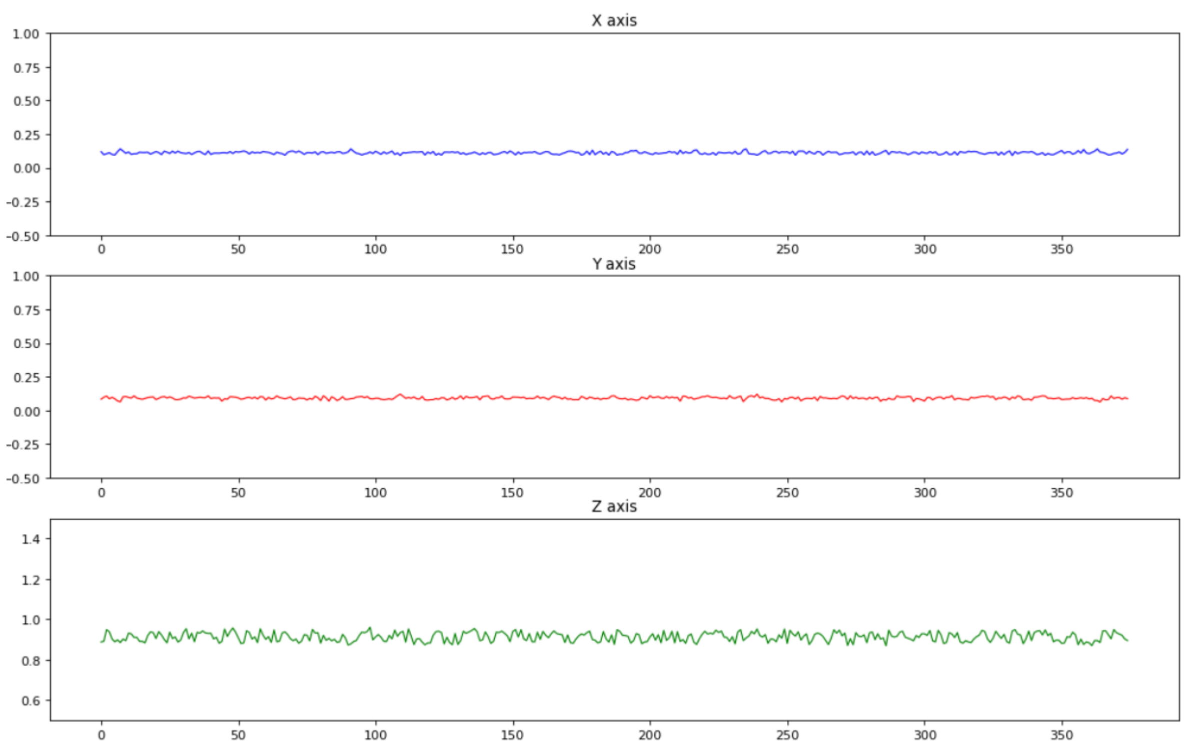
Appendix C

Section 3.4

Table C-3. Equipment required for experiments 3.4.1 to 3.4.3

Device	Image	Specs
Raspberry Pi 3B+		GPU: Broadcam Videocore-IB RAM:1GB SSDRAM, SoC: ARMb8 @ 1.4Ghz. Memory: 32 GB SD
TP-Link router modelo Archer A5		802.11a/b/g/n, WDS protocol 2 antennas
TP-Link C50		802.11a/b/g/n, WDS protocol 4 antennas
Lenovo laptop		Intel core i5-8300H 2.3 GHz 8 GB RAM, 1TB ROM Ethernet, 802.11a/b/g/n Windows 10 Home
Rompe muros Green Leaf		28dBm gain Frequency band: 2.4-2.4835
Sensor neo 6m Ublox		Communication protocol: NMEA, RTCM Baud Rate: 4800-340400, Cool start: 27s, hot start: 1s, Sensitivity: -161dBm
Raspberry Pi Camera		12.3Mega Pixels Max 90FPS
Accelerometer - Gyroscope MPU6050		Sample rate: 1KHz. I2C interface Tolerance: +-3% Measuring ranges: +-2g up to +-16g

Section 3.5



Appendix C

Figure C-9. Parked signal reference

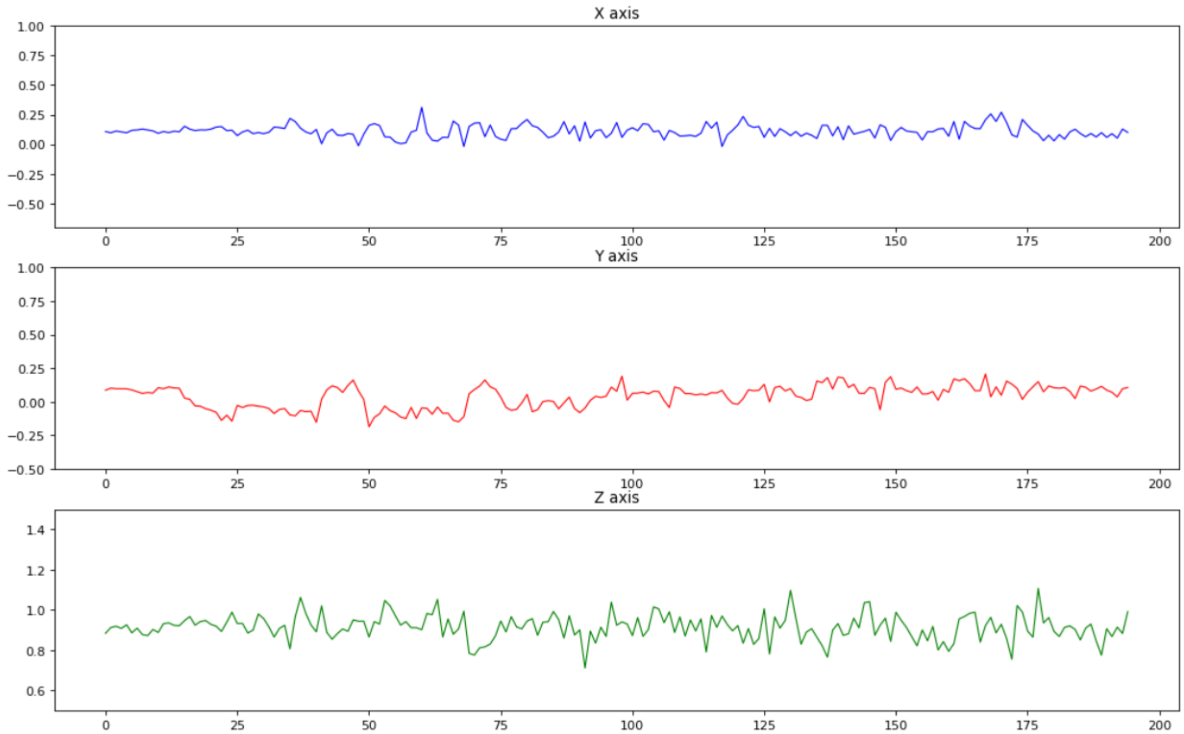


Figure C-9. Plain route signal at 20km/h

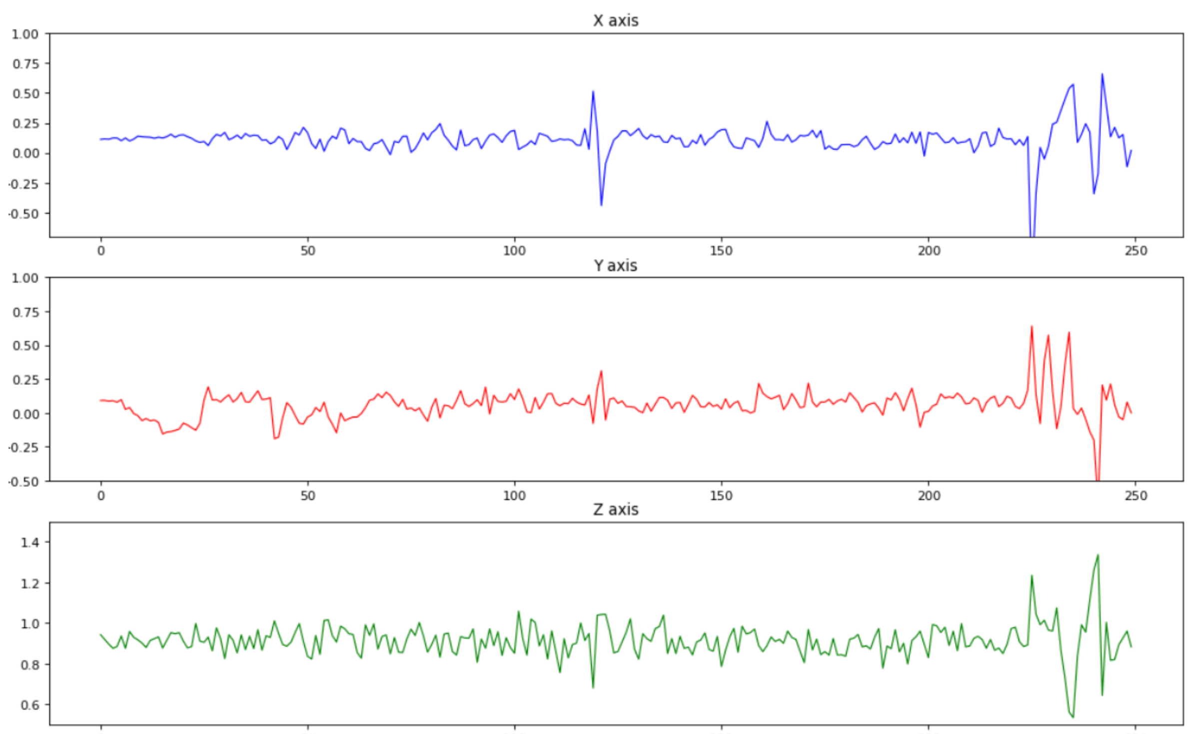


Figure C-10. Street with pothole and speedbump at 20km/h on average

Appendix C

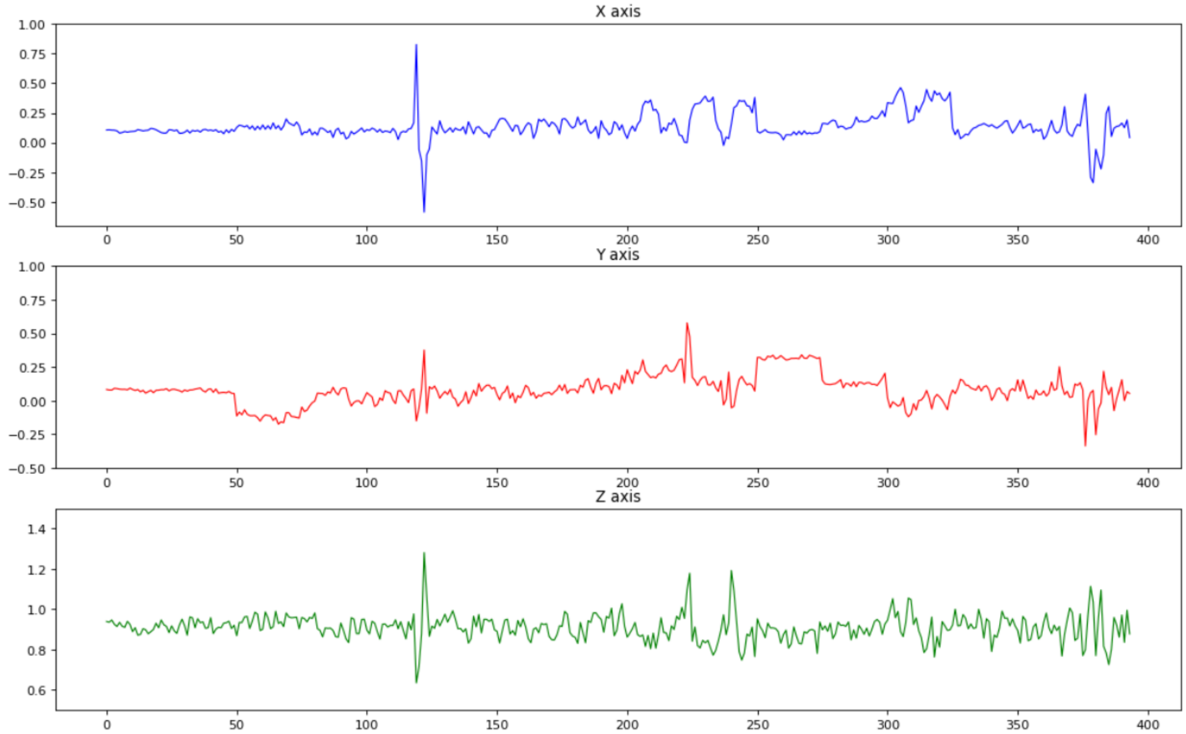


Figure C-11. Street with 3 different potholes at 30km/h on average.

Figure C-12. OBU photograph before a pothole and a speed bump.

Appendix C

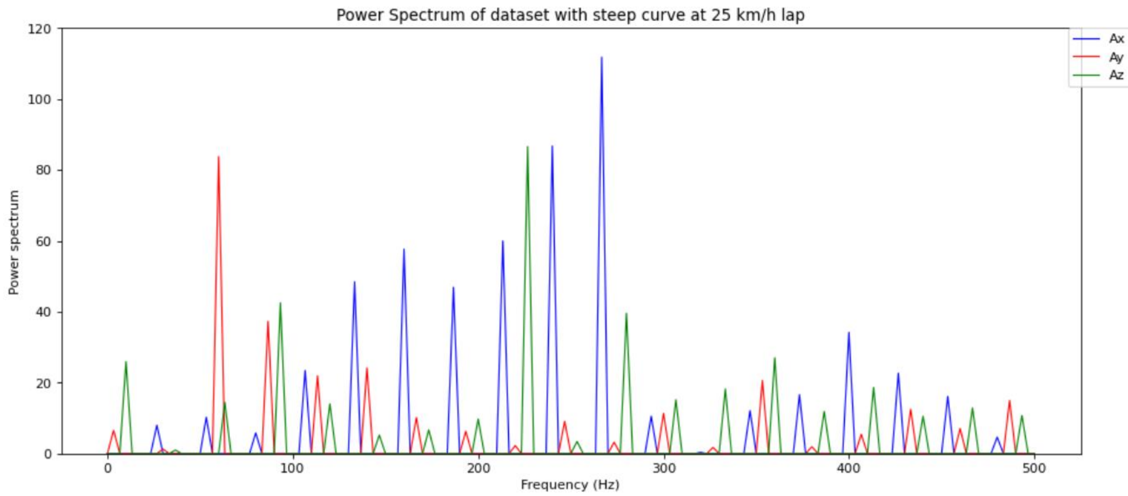


Figure C-13. Power Spectrum of a steep curve/turn at 25 km/h

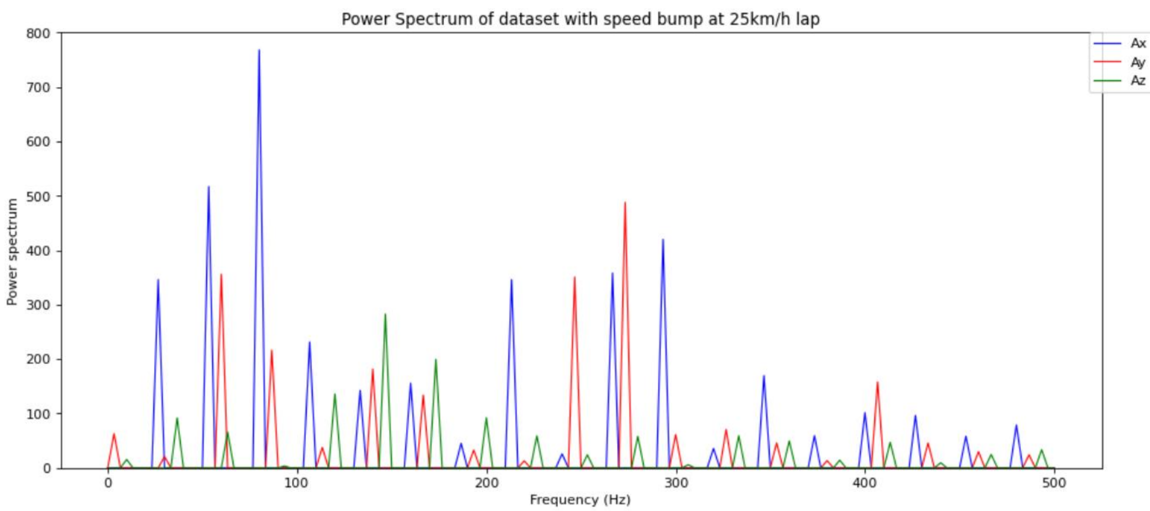


Figure C-14. Power Spectrum of a speed bump at 25 km/h

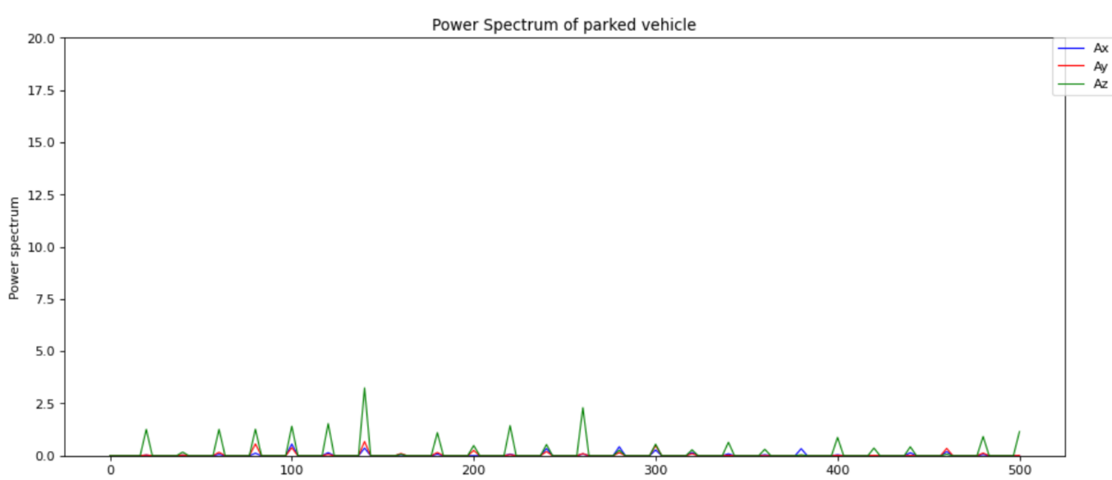


Figure C-15. Power Spectrum of the oscillations of a vehicle turned on without moving.

Appendix C

Pseudocode for Power spectrum calculation.

```
time_domain_signal = getTimeDomainSignal(source);
Fs=Sampling_Rate
N= length of(time_domain_signal);
xdft = FFT(time_domain_signal);
xdft = xdft(1:N/2+1);
psdx = (1/(Fs*N)) * abs(xdft).^2;
psdx(2:end-1) = 2*psdx(2:end-1);
freq = 0:Fs / length(time_domain_signal):Fs/2;

#plotting
plot(freq, psdx)

title('Periodogram Using FFT')
xlabel('Frequency (Hz)')
ylabel('Power [V^2]')
```

Section 3.5.3

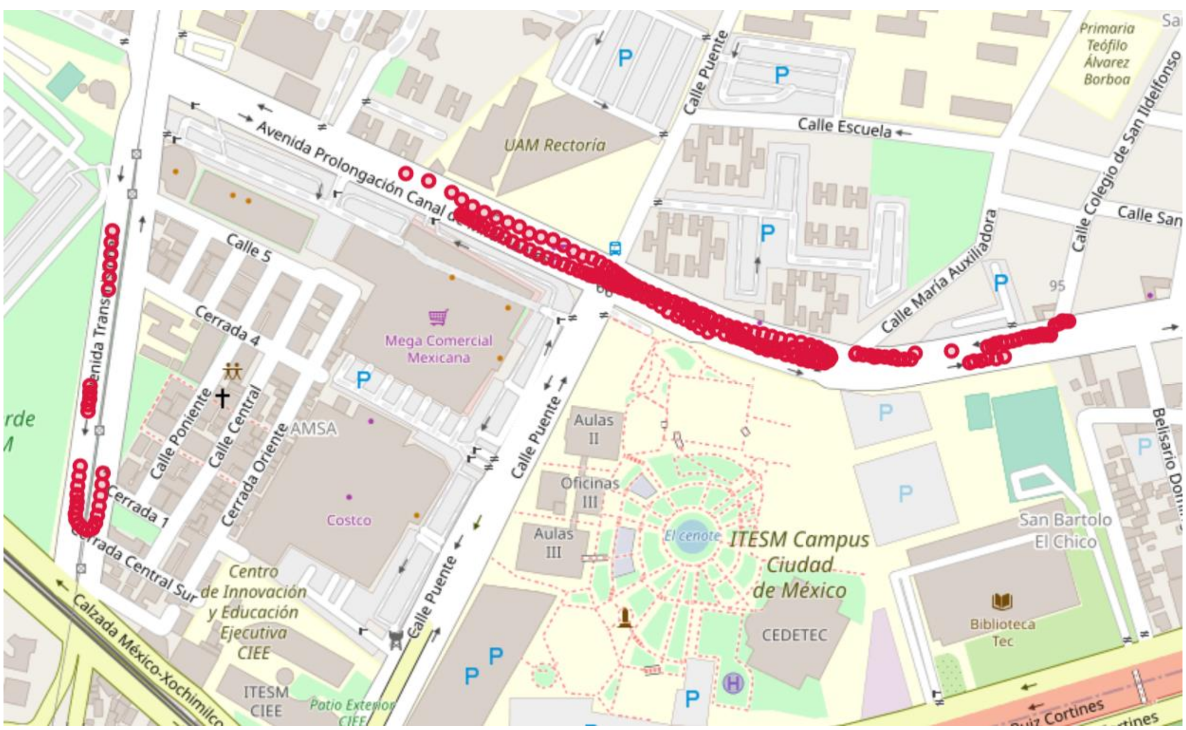


Figure C-16. KNN detected anomalies in both avenues.

Appendix C

Figure C-17. KNN healthy street points.

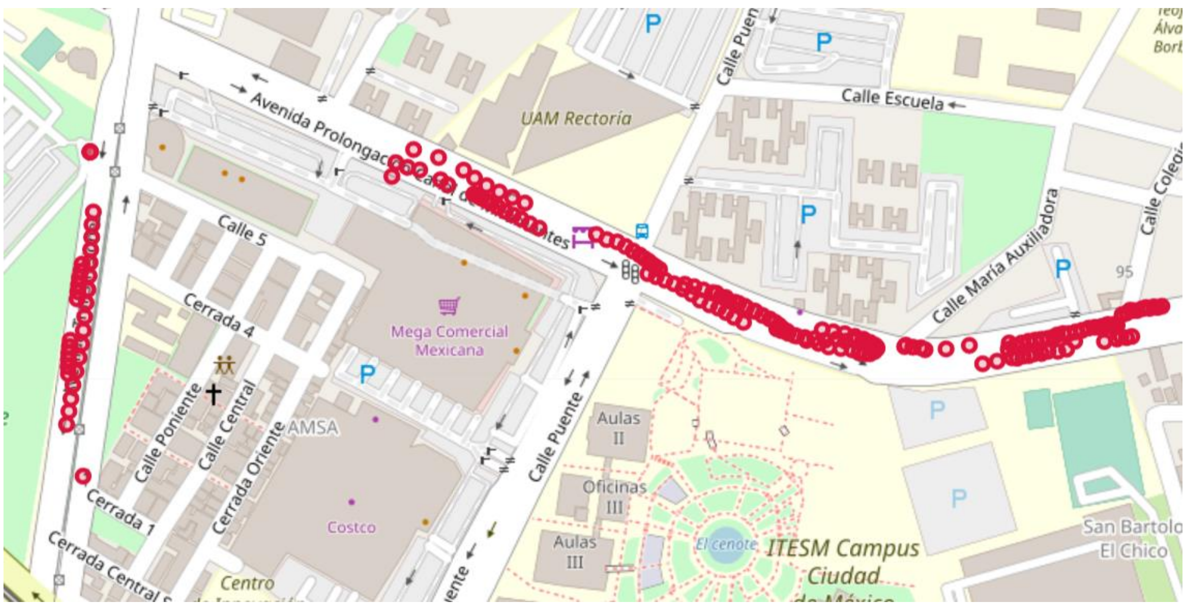


Figure C-18. ANN detected anomalies in both avenues.

Appendix C

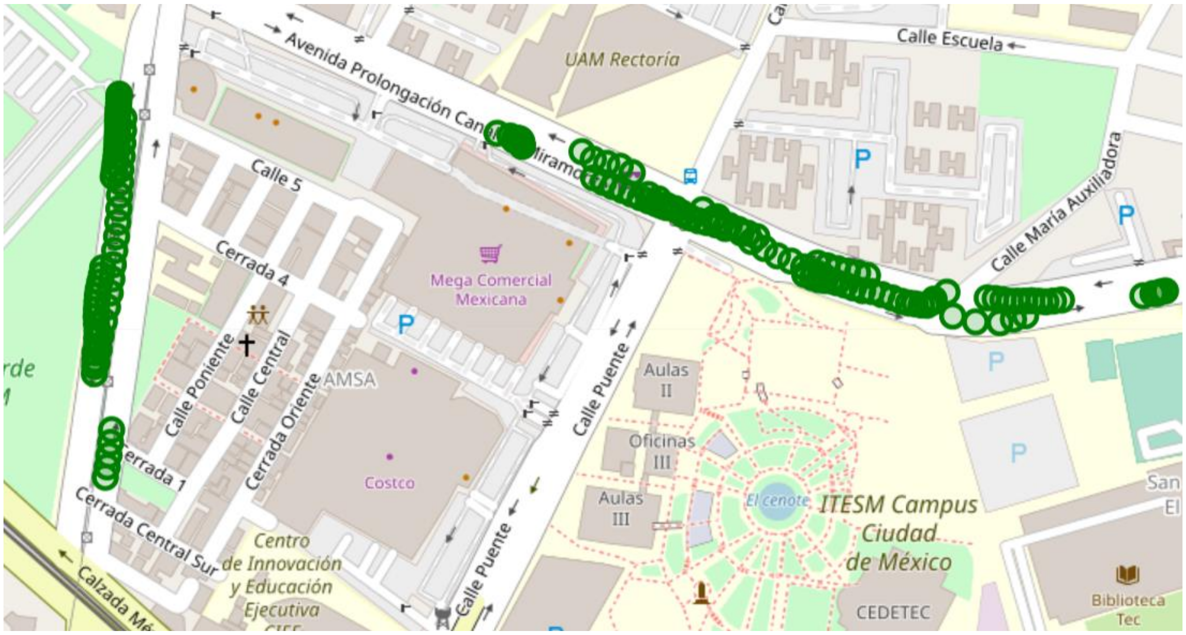


Figure C-19. ANN healthy street points.

Bibliography

[1] Statista, "Number of Internet of things (IoT) connected devices worldwide in 2018,2025 and 2030," 2021. [Online]. Available: https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/.

[2] B. Y. R. Khatoun and S. Zeadally, "Smart Cities : Concepts, Architectures, Research Oportunities," *Commun. ACM*, vol. 59, no. 8, pp. 46–67, 2016.

[3] H. Ahvenniemi, A. Huovila, I. Pinto-Seppä, and M. Airaksinen, "What are the differences between sustainable and smart cities?," *Cities*, vol. 60, pp. 234–245, 2017.

[4] M. Batty *et al.*, "Smart cities of the future," *Eur. Phys. J. Spec. Top.*, vol. 214, no. 1, pp. 481–518, 2012.

[5] G. Lyons, "Getting smart about urban mobility – Aligning the paradigms of smart and sustainable," *Transp. Res. Part A Policy Pract.*, vol. 115, pp. 4–14, 2018.

[6] G. Dimitrakopoulos, *Current Technologies in Vehicular Communication*. 2017.

[7] J. Garcia, E. Simo, X. Masip-Bruin, E. Marin-Tordera, and S. Sanchez-Lopez, "Do we really need cloud? estimating the fog computing capacities in the city of Barcelona," *Proc. - 11th IEEE/ACM Int. Conf. Util. Cloud Comput. Companion, UCC Companion 2018*, pp. 290–295, 2019.

[8] H. J. Hong, "From cloud computing to fog computing: Unleash the power of edge and end devices," *Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom*, vol. 2017-Decem, pp. 331–334, 2017.

[9] Statista, "Global cloud data center IP traffic from 2015 to 2021," 2018. [Online]. Available: https://www.statista.com/statistics/227267/global-cloud-ip-traffic-growth-by-segment/.

[10] K. Dolui, "Comparison of Edge Computing Implementations : Fog Computing , Cloudlet and Mobile Edge Computing," 2017.

[11] A. Yousefpour *et al.*, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, 2019.

[12] Industrial Internet and Open Fog Consortium, "Openfog reference for fog computing and edge computing," 2021. [Online]. Available: https://www.iiconsortium.org/IIC-OF-faq.htm.

[13] INEGI, “CENSO DE POBLACIÓN Y VIVIENDA 2020,” 2021.

[14] INEGI, “Encuesta Origen-Destino En Hogares,” p. 80, 2017.

[15] IMCO, “El costo de la congestión,” 2019.

[16] C. A. Garcia-Huerta, R. M. Bustamante-Bello, A. Garcia-Barba, and J. Izquierdo-Reyes, “Emulación de comunicación vehicular V2X,” 2019.

[17] L. A. Curiel-Ramirez, J. Izquierdo-Reyes, M. R. Bustamante-Bello, R. A. Ramirez-Mendoza, and A. Garcia-Barba, “A Simulation Approach of the Internet of Intelligent Vehicles for Closed Routes in Urban Environments,” *2019 42nd Int. Conf. Telecommun. Signal Process.*, pp. 672–680, 2019.

[18] S. Damien-LHUSSIer, “ETSI - Mobile Technologies - 5g, 5g Specs | Future Technology,” 2019. [Online]. Available: https://www.etsi.org/technologies/5g?jjj=1569967725903. [Accessed: 01-Oct-2019].

[19] A. G. Haddad, A. Takiddeen, A. Obeid, and A. Sleptchenko, “Traffic Optimization by Simultaneous Control of Vehicles Speeds and Routes,” *2019 IEEE 6th Int. Conf. Ind. Eng. Appl. ICIEA 2019*, pp. 786–790, 2019.

[20] P. Zechel, R. Streiter, K. Bogenberger, and U. Gohner, “Pedestrian Occupancy Prediction for Autonomous Vehicles,” *Proc. - 3rd IEEE Int. Conf. Robot. Comput. IRC 2019*, pp. 230–235, 2019.

[21] B. Ulmer, “VITA II-active collision avoidance in real traffic,” in *Proceedings of the Intellingent Vehicles ’94 Symposium*, 2002, pp. 1–6.

[22] P. Furgale *et al.*, “Toward automated driving in cities using close-to-market sensors: An overview of the V-Charge Project,” *IEEE Intell. Veh. Symp. Proc.*, no. Iv, pp. 809–816, 2013.

[23] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, “An overview of Internet of Vehicles,” *China Commun.*, vol. 11, no. 10, pp. 1–15, 2014.

[24] R. Iqbal, T. A. Butt, M. O. Shafique, M. W. A. Talib, and T. Umer, “Context-aware data-driven intelligent framework for fog infrastructures in internet of vehicles,” *IEEE Access*, vol. 6, pp. 58182–58194, 2018.

[25] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibanez, “Internet of Vehicles: Architecture, Protocols, and Security,” *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3701–3709, 2018.

[26] F. Cunha *et al.*, “Data communication in VANETs: Protocols, applications and challenges,” *Ad Hoc Networks*, vol. 44, pp. 90–103, 2016.

[27] Nokia, “5G Highlights - Potential Technology for 3GPP,” 2016.

[28] J. A. Guerrero-Ibáñez, C. Flores-Cortés, and S. Zeadally, "Vehicular Ad-hoc Networks (VANETs): Architecture, Protocols and Applications," pp. 49–70, 2013.

[29] J. T. Isaac, S. Zeadally, and J. S. Cámara, "Security attacks and solutions for vehicular ad hoc networks," *IET Commun.*, vol. 4, no. 7, pp. 894–903, 2010.

[30] S. Zeadally, R. Hunt, Y. S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): Status, results, and challenges," *Telecommun. Syst.*, vol. 50, no. 4, pp. 217–241, 2012.

[31] G. Dimitrakopoulos, "Intelligent transportation systems based on internet-connected vehicles: Fundamental research areas and challenges," *2011 11th Int. Conf. ITS Telecommun. ITST 2011*, pp. 145–151, 2011.

[32] P. Rahimian, E. E. O. Neal, S. Zhou, J. M. Plumert, J. K. Kearney, and I. City, "Harnessing Vehicle-to-Pedestrian (V2P) Communication Technology : Sending Traffic Warnings to Texting Pedestrians," 2018.

[33] J. Harding *et al.*, "Vehicle-to-vehicle communications: Readiness of V2V technology for application. (Report No. DOT HS 812 014)," no. August, 2014.

[34] H. Rakha and R. K. Kamalanathsharma, "Eco-driving at signalized intersections using V2I communication," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 341–346, 2011.

[35] A. M. Pereira, "Traffic signal control for connected and non-connected vehicles," *2018 Smart Cities Symp. Prague, SCSP 2018*, pp. 1–6, 2018.

[36] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero Ibáñez, "A seven-layered model architecture for Internet of Vehicles," *J. Inf. Telecommun.*, vol. 1, no. 1, pp. 4–22, 2017.

[37] T. Tharang and D. Vallidevi, "Internet of Vehicles (10 V) for traffic management," no. 10 V, 2017.

[38] S. Raza, S. Wang, M. Ahmed, and M. R. Anwar, "A Survey on Vehicular Edge Computing : Architecture , Applications , Technical Issues , and Future Directions," vol. 2019, 2019.

[39] J. Kang *et al.*, "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4660–4670, 2019.

[40] Z. Ning, P. Dong, X. Wang, J. J. P. C. Rodrigues, and F. Xia, "Deep reinforcement learning for vehicular edge computing: An intelligent offloading system," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 6, 2019.

[41] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016.

[42] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," *MCC'12 - Proc. 1st ACM Mob. Cloud Comput. Work.*, pp. 13–15, 2012.

[43] A. Orlowski and P. Romanowska, "Smart Cities Concept: Smart Mobility Indicator," *Cybern. Syst.*, vol. 50, no. 2, pp. 118–131, 2019.

[44] B. Lv *et al.*, "LiDAR-Enhanced Connected Infrastructures Sensing and Broadcasting High-Resolution Traffic Information Serving Smart Cities," *IEEE Access*, vol. 7, pp. 79895–79907, 2019.

[45] M. Rice *et al.*, "Physical-layer security for vehicle- to-everything networks," no. September, pp. 2–10, 2020.

[46] P. Singh, P. Khanna, and S. Kumar, "Communication Architecture for Vehicular Ad Hoc Networks , with Blockchain Security," 2020.

[47] J. Rezgui, E. Gagne, G. Blain, O. St-Pierre, and M. Harvey, "Platooning of Autonomous Vehicles with Artificial Intelligence V2I Communications and Navigation Algorithm," *2020 Glob. Inf. Infrastruct. Netw. Symp. GIIS 2020*, 2020.

[48] P. Bhadauria, R. Kumar, and S. Shamra, "Performance Dependency of LSTM and NAR Beamformers With Respect to Sensor Array Properties in V2I Scenario Thapar Institute of Engineering and Technology , Patiala , Punjab , India ."

[49] A. Javed, A. Malhi, and K. Framling, "Edge Computing-based Fault-Tolerant Framework: A Case Study on Vehicular Networks," *2020 Int. Wirel. Commun. Mob. Comput. IWCMC 2020*, pp. 1541–1548, 2020.

[50] Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang, "Artificial Intelligence Empowered Edge Computing and Caching for Internet of Vehicles," *IEEE Wirel. Commun.*, vol. 26, no. 3, pp. 12–18, 2019.

[51] E. Badidi, "Fog Computing for Smart Cities ' Big Data Management and Analytics : A Review," pp. 1–28, 2020.

[52] M. N. Ahangar, Q. Z. Ahmed, F. A. Khan, and M. Hafeez, "A survey of autonomous vehicles: Enabling communication technologies and challenges," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–33, 2021.

[53] J. S. Martín, A. Cortés, L. Zamora-Cadenas, and B. J. Svensson, "Precise positioning of autonomous vehicles combining UWB ranging estimations with on-board sensors," *Electron.*, vol. 9, no. 8, pp. 1–21, 2020.

[54] S. S. Dorle, "Vehicle Classification and Communication Using Zigbee Protocol," pp. 2–5, 2010.

[55] Y. Lei and J. Wu, “Study of applying ZigBee technology into foward collision warning system (FCWS) under low-speed circumstance,” *WOCC 2016 - 25th Wirel. Opt. Commun. Conf. Jointly held with Photonics Forum Chiao-Tung Univ.*, pp. 0–3, 2016.

[56] P. Cheolhee and T. S. Rappaport, “Short-range wireless communications for next-generation networks: UWB 60 GHz millimeter-wave wpan, and ŹigBee,” *IEEE Wirel. Commun.*, vol. 14, no. 4, pp. 70–78, 2007.

[57] C. Yiwen, L. Xinglong, Z. Zhongua, J. Xiaujun, Y. Jian, and P. Xiaofeng, “A Comparative Study of BLE-based Fingerprint Localization for Vehicular Application,” vol. 6.

[58] S. Biswas, R. Tatchikou, and F. Dion, “Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety,” *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 74–82, 2006.

[59] S. Gao, A. Lim, and D. Bevly, “An empirical study of DSRC V2V performance in truck platooning scenarios,” *Digit. Commun. Networks*, vol. 2, no. 4, pp. 233–244, 2016.

[60] K. Abboud, H. A. Omar, and W. Zhuang, “Interworking of DSRC and Cellular Network Technologies for V2X Communications : A Survey,” 2016.

[61] S. Lee, Y. J. Yoon, J. E. Lee, and S. C. Kim, “Human-vehicle classification using feature-based SVM in 77-GHz automotive FMCW radar,” *IET Radar, Sonar Navig.*, vol. 11, no. 10, pp. 1589–1596, 2017.

[62] I. Yaqoob, L. U. Khan, S. M. A. Kazmi, M. Imran, N. Guizani, and C. S. Hong, “Autonomous Driving Cars in Smart Cities: Recent Advances, Requirements, and Challenges,” *IEEE Netw.*, vol. 34, no. 1, pp. 174–181, 2020.

[63] A. Balasubramanian, R. Mahajan, A. Venkataramani, B. N. Levine, and J. Zahorjan, “Interactive wifi connectivity for moving vehicles,” p. 427, 2008.

[64] M. Condoluci *et al.*, “5G V2X system-level architecture of 5GCAR project,” *Futur. Internet*, vol. 11, no. 10, 2019.

[65] J. Jeong *et al.*, “A comprehensive survey on vehicular networks for smart roads: A focus on IP-based approaches,” *Veh. Commun.*, vol. 29, p. 100334, 2021.

[66] O. Senouci, Z. Aliouat, and S. Harous, “MCA-V2I: A Multi-hop Clustering Approach over Vehicle-to-Internet communication for improving VANETs performances,” *Futur. Gener. Comput. Syst.*, vol. 96, pp. 309–323, 2019.

[67] A. Reyes-Muñoz, M. C. Domingo, M. A. López-Trinidad, and J. L.

Delgado, "Integration of body sensor networks and vehicular Ad-hoc networks for traffic safety," *Sensors (Switzerland)*, vol. 16, no. 1, pp. 1–29, 2016.

[68] N. Ding, H. Ma, C. Zhao, Y. Ma, and H. Ge, "Driver's emotional state-based data anomaly detection for vehicular Ad Hoc networks," *Proc. - 2019 IEEE Int. Conf. Smart Internet Things, SmartIoT 2019*, pp. 121–126, 2019.

[69] G. Srinidhi and R. D. S. M, "Available on : SSRN Pothole Detection using CNN and AlexNet," 2020.

[70] P. Ping, X. Yang, and Z. Gao, "A Deep Learning Approach for Street Pothole Detection," *Proc. - 2020 IEEE 6th Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2020*, pp. 198–204, 2020.

[71] A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, "Anomaly Detection in Automated Vehicles Using Multistage Attention-Based Convolutional Neural Network," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–10, 2020.

[72] SEDECO, "Reporte Económico de la Ciudad de México," 2020.

[73] E. CDMX, "Diagnóstico de la desigualdad socio-territorial," 2020.

[74] J. M. Bauer, "The Internet and income inequality : Socio-economic challenges in a hyperconnected society," *Telecomm. Policy*, no. May, pp. 1–11, 2017.

[75] IFT, "Cuadro Nacional de Atribución de Frecuencias," 2021.

[76] IFT, "Panorama del espectro radioeléctrico en México para servicios móviles de quinta generación." 2019.

[77] DOF, "Convocatoria Licitación Pública IFT-10." 2021.

[78] Evalúa CDMX, "Índice de Desarrollo Social De las Unidades Territoriales Del Distrito Federal: Delegación, Colonia y Manzana," 2011.

[79] Gobierno de la Ciudad de México, "Ubicación de puntos de acceso gratuito a internet WiFi vía infraestructura C5." CDMX, 2021.

[80] F. G. Tinetti, I. Real, R. Jaramillo, D. Barry, and F. De Informática, "Hadoop Scalability and Performance Testing in Heterogeneous Clusters," p. 441, 2015.

[81] P. Bajcsy, A. Vandecreme, J. Amelot, P. Nguyen, J. Chalfoun, and M. Brady, "Terabyte-sized image computations on Hadoop cluster platforms," *Proc. - 2013 IEEE Int. Conf. Big Data, Big Data 2013*, pp. 729–737, 2013.

[82] J. Semmlow, "The Fourier Transform and Power Spectrum," *Signals*

Syst. Bioeng., pp. 131–165, 2012.

[83] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[84] A. Garcia-Barba, “V2I CDMX Smart Mobility Code Repository,” 2021. [Online]. Available: https://github.com/AlecGarBarba/V2XSmartMobility.

Published Papers

A Simulation Approach of the Internet of Intelligent Vehicles for Closed Routes in Urban Environments

Luis A. Curiel-Ramirez, Javier Izquierdo-Reyes, M. Rogelio Bustamante-Bello, Ricardo A. Ramirez-Mendoza and Alec Garcia-Barba
Escuela de Ingeniería y Ciencias
Tecnologico de Monterrey
Mexico City, Mexico
Email: lacuriel55@gmail.com, jizquierdo.reyes@tec.mx, rbustama@tec.mx, ricardo.ramirez@tec.mx and alecbarba@gmail.com

***Abstract*—In recent years, through the evolution of intelligent transport systems (ITS), there has been an increase in components to monitor different variables in the vehicles (such as driver monitoring, tire pressure, oil pressure, vehicle speed, acceleration, among others). Internet of Things (IoT) is enabling the collection of a variety of information of all types of components for monitoring the environment and systems. Also, as part of this evolution, the original concept of Vehicular Ad-hoc networks (VANETs) is being transformed into a new concept, the Internet of Vehicles (IoV), which involves Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I) communications, and other variations of these (V2X). In this paper, a review of the state of the art of the technologies and architectures generated around the Internet of the vehicles (IoV) is carried out and with this, an implementation of the simulation of intelligent vehicular communications through the IEEE 802.11p protocol and one of its applications. An algorithm for the generation of vehicular convoys to improve vehicle flow application through V2X communication is proposed and implemented in the simulation, and the communication quality is validated in a closed route in an urban environment during the simulations, omitting aspects of security in communications.**

***Keywords*—Automotive Vehicles, Autonomous Vehicles, Internet of Things, Internet of Vehicles, Semi-autonomous Vehicles**

I. Introduction

The concept of Vehicular Ad Hoc Networks (VANETs) was conceived over a decade ago and it has been a very active research area, both in academia and industry [1] [2] [3] [4] [5]. However, as the number of vehicles connected to the Internet of Things increases, new requirements of VANETs are emerging [1]. The basic principle of a VANET is that a vehicle is a mobile node that enables the connection with other vehicles thereby creating a network [1]. One of the main problems of VANETs is its limited capacity for processing all the information that is collected by themselves and other actors (such as sensors and mobile devices) around the environment [1]. In this context, vehicles must evolve into 'smart' objects equipped with a multi-sensory platform, including a set of communication technologies, robust computational units, IP-based connectivity to the Internet, and a direct or indirect connection to other vehicles and with all devices around. This is where the Internet of Vehicles emerges as an evolution to this technology, where the IoV integrates two technological visions: vehicles networking and vehicles intelligence [6] and focuses on the integration of objects such as humans, vehicles, things, networks and environments to create an intelligent network based on computing and communication capabilities that supports services for large cities or even a whole country [1].

Some studies have predicted that 25 billion devices will be connected to the internet by...

Curriculum Vitae

Alec García Barba was born in Distrito Federal, México, on February 17, 1996. He earned the Telecommunications and Electronic Systems Engineering degree from the *Instituto Tecnológico y de Estudios Superiores de Monterrey*, Monterrey Campus Ciudad de México in May 2019. He was accepted in the graduate programs in Engineering Sciences in June 2019.

During his undergraduate program, he completed 2 research stages at ITESM CCM. This research program ended up with the publication of the article “A Simulation Approach of the Internet of Intelligent Vehicles for Closed Routes in Urban Environments”.