



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

Análisis de protocolos de comunicación con aprendizaje de máquina para ciudades inteligentes.

P R E S E N T A:

Edgar Romo Montiel

Directores de tesis:
Dr. Mario Eduardo Rivero Ángeles
Dr. Ricardo Menchaca Méndez

Ciudad de México

Junio 2022





INSTITUTO POLITÉCNICO NACIONAL

SECRETARIA DE INVESTIGACIÓN Y POSGRADO

SIP-13
REP 2017

ACTA DE REGISTRO DE TEMA DE TESIS Y DESIGNACIÓN DE DIRECTOR DE TESIS

Ciudad de México, a 7 de abril del 2022

El Colegio de Profesores de Posgrado del Centro de Investigación en Computación en su Sesión
(Unidad Académica)

Extraordinaria No. 31 celebrada el día 1 del mes diciembre de 2021 conoció la solicitud presentada por el (la) alumno (a):

Apellido Paterno:	Romo	Apellido Materno:	Montiel	Nombre (s):	Edgar
--------------------------	-------------	--------------------------	----------------	--------------------	--------------

Número de registro: A 1 8 0 4 8 4

del Programa Académico de Posgrado: Doctorado en Ciencias de la Computación

Referente al registro de su tema de tesis; acordando lo siguiente:

1.- Se designa al aspirante el tema de tesis titulado:

"Análisis de protocolos de comunicación con aprendizaje de máquina para ciudades inteligentes"

Objetivo general del trabajo de tesis:

Analizar el desempeño de protocolos de comunicación de la capa de control de acceso al medio en redes inalámbricas para entornos de ciudades inteligentes y mejorarlos mediante técnicas de aprendizaje de máquina para la elección de parámetros óptimos de transmisión.

2.- Se designa como Directores de Tesis a los profesores:

Director: Dr. Mario Eduardo Rivero Ángeles 2° Director: Dr. Ricardo Menchaca Méndez
No aplica:

3.- El Trabajo de investigación base para el desarrollo de la tesis será elaborado por el alumno en:

El Laboratorio Redes y Ciencia de Datos

que cuenta con los recursos e infraestructura necesarios.

4.- El interesado deberá asistir a los seminarios desarrollados en el área de adscripción del trabajo desde la fecha en que se suscribe la presente, hasta la aprobación de la versión completa de la tesis por parte de la Comisión Revisora correspondiente.

Director(a) de Tesis

Dr. Mario Eduardo Rivero Ángeles

Alumno(a)

Edgar Romo Montiel

2° Director de Tesis

Dr. Ricardo Menchaca Méndez
Presidente del Colegio de Profesores de Posgrado
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
DIRECCIÓN IPN-CIC

Dr. Francisco Hiram Calvo Castro



INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de siendo las horas del día del mes de del se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Posgrado de para examinar la tesis titulada:

del (la) alumno (a):

Apellido Paterno:	Romo	Apellido Materno:	Montiel	Nombre (s):	Edgar
-------------------	------	-------------------	---------	-------------	-------

Número de registro:

Aspirante del Programa Académico de Posgrado:

Una vez que se realizó un análisis de similitud de texto, utilizando el software antiplagio, se encontró que el trabajo de tesis tiene 5% de similitud. **Se adjunta reporte de software utilizado.**

Después que esta Comisión revisó exhaustivamente el contenido, estructura, intención y ubicación de los textos de la tesis identificados como coincidentes con otros documentos, concluyó que en el presente trabajo SI NO **SE CONSTITUYE UN POSIBLE PLAGIO.**

JUSTIFICACIÓN DE LA CONCLUSIÓN: *(Por ejemplo, el % de similitud se localiza en metodologías adecuadamente referidas a fuente original)*

El porcentaje de similitud es bajo y se localiza en frases cortas y comunes, así como en nombres propios y textos adecuadamente referenciados, lo cual no representa plagio. Lo anterior se puede constatar en el resumen reportado por el software Turnitin.

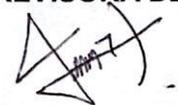
****Es responsabilidad del alumno como autor de la tesis la verificación antiplagio, y del Director o Directores de tesis el análisis del % de similitud para establecer el riesgo o la existencia de un posible plagio.**

Finalmente y posterior a la lectura, revisión individual, así como el análisis e intercambio de opiniones, los miembros de la Comisión manifestaron **APROBAR** **SUSPENDER** **NO APROBAR** la tesis por **UNANIMIDAD** o **MAYORÍA** en virtud de los motivos siguientes:

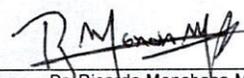
Cumple con los requisitos de una tesis de Doctorado en Ciencias de la Computación.

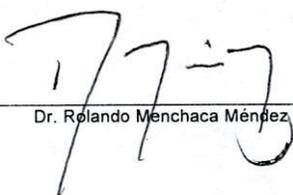
COMISIÓN REVISORA DE TESIS

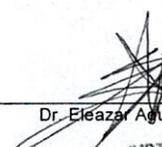

Dr. Mario Eduardo Rivero Angeles
Director de Tesis


Dr. Marco Antonio Moreno Armendáriz


Dr. Noé Torres Cruz


Dr. Ricardo Menchaca Méndez
2° Director de Tesis (en su caso)


Dr. Rolando Menchaca Méndez


Dr. Francisco Hiram Calvo Pantoja
PRESIDENTE DEL COLEGIO DE PROFESORES IPN-EIC





INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA DE AUTORIZACIÓN DE USO DE OBRA PARA DIFUSIÓN

En la Ciudad de México el día 20 del mes de Junio del año 2022, el (la) que suscribe Edgar Romo Montiel alumno(a) del programa Doctorado en Ciencias de la Computación con número de registro A180484, adscrito(a) a Centro de Investigación en Computación manifiesta que es autor(a) intelectual del presente trabajo de tesis bajo la dirección de Dr. Mario Eduardo Rivero Ángeles y Ricardo Menchaca Méndez y cede los derechos del trabajo intitulado Análisis de protocolos de comunicación con aprendizaje de máquina para ciudades inteligentes, al Instituto Politécnico Nacional, para su difusión con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expresado del autor y/o director(es). Este puede ser obtenido escribiendo a las siguiente(s) dirección(es) de correo. edgar.romo.montiel@gmail.com. Si el permiso se otorga, al usuario deberá dar agradecimiento correspondiente y citar la fuente de este.


Edgar Romo Montiel

Nombre completo y firma autográfica del (de la)
estudiante

Resumen

Sistemas de comunicación inalámbrica. Los sistemas de comunicaciones inalámbricas permiten que diferentes dispositivos se conecten a una red con la característica principal de tener facilidad de movilidad, permitiendo al usuario ubicarse en distintos sitios y poder acceder, en la mayoría de los casos, a algún servicio de Internet.

Hoy en día los dispositivos que acceden por algún medio inalámbrico ha crecido significativamente. Es común ver teléfonos celulares, relojes inteligentes, computadoras portátiles, e incluso vehículos, conectados a Internet. Esto ocasiona que muchos dispositivos demanden recursos de un mismo canal inalámbrico y, al crecer el número de usuarios, genere ambientes hostiles de comunicación. Por otro lado, las técnicas de inteligencia artificial han tomado relevancia en la mejora de distintas áreas de las ciencias, y las comunicaciones no son la excepción. La alta capacidad que ha mostrado en tareas de predicción y de control ha impulsado la autonomía de múltiples sistemas.

En este estudio se presenta el análisis de protocolos de comunicación de la capa de control de acceso al medio conocida como MAC para la asignación de recursos de canal a una red inalámbrica. Se estudia el caso de una red Vehicular (VANET) conectada a una infraestructura de telefonía celular que posee un canal ALOHA ranurado (S-ALOHA) para la transmisión de información, principalmente de telemetría, de los automóviles. El contexto considerado es en ciudades inteligentes (Smart Cities), donde el flujo de información es alto y es preciso tener parámetros óptimos que maximicen el desempeño de las redes de comunicación.

Como tal, se propone una red asistida con un módulo de aprendizaje de maquina que sea capaz de observar el estado de la red, inferir las condiciones futuras y hacer ajustes a los parámetros críticos para hacer eficiente la transmisión de información de los vehículos empleando un protocolo de baja complejidad basado en S-ALOHA.

Abstract

Wireless communication systems. The wireless communication systems allow different devices to connect to a network with the main feature of having ease of mobility. It enables users to be in multiple places, and they can access, in the majorly of cases, some Internet service.

Today, devices connecting by wireless means have grown up significantly. It is usual to see cell phones, smartwatches, laptops even vehicles connected to the Internet. This fact causes many devices to demand resources from the same wireless channel, and, as the number of users grows, it creates unfriendly environments for communication system. On the other hand, artificial intelligence techniques have taken importance in different science areas, and communications is not an exception. The high capacity that they have shown in prediction and control tasks has impuled autonomy in different systems.

This study analyses communication protocols of the media access control (MAC) for the assignation of channel resources to a wireless network. It considers the case of a Vehicular network (VANET) connected to a cellular telephony infrastructure that possesses a Slotted ALOHA channel (S-ALOHA) for data transmission, mainly telemetric, from automobiles. The considered context is Smart Cities, where data flow is high, and it is precise to have optimal parameters that maximize communication network performance.

As such, it proposes an assisted network with a machine learning module able to observe the network state, infer future conditions, and adjust critical parameters to make an efficient data transmission from vehicles using a low complexity protocol based on S-ALOHA.

Agradecimientos

Agradezco al Consejo Nacional de Ciencia y Tecnología, CONACYT, por la beca recibida durante mis estudios de doctorado, así como el apoyo brindado para realizar estancia de investigación y los recursos informáticos proporcionados.

Agradezco al Centro de Investigación en Computación del Instituto Politécnico Nacional, CIC-IPN, por brindarme el entorno y herramientas para llevar a cabo mis estudios.

Agradezco a mis directores de tesis, el Dr Mario Eduardo Rivero Ángeles con quién he tenido la oportunidad de trabajar de cerca a lo largo de mi formación académica de una manera satisfactoria y enriquecedora y al Dr Ricardo Menchaca Mendez por compartirme sus conocimientos e impulsarme a aprender nuevas cosas y poner en práctica los conocimientos adquiridos, además de brindarme la oportunidad de integrarme en un ambiente de trabajo ameno en convivencia con los compañeros del laboratorio.

Agradezco a todos y cada uno los miembros del jurado de evaluación y al Laboratorio de Redes y Ciencia de Datos del CIC-IPN.

Agradezco a la vida por darme la oportunidad de haber llegado hasta este punto y permitirme realizar este deseo de convertirme en Doctor en Ciencias de la Computación.

Agradezco a mis padres Roque y Sara por su apoyo y comprensión a lo largo de mi vida y que me ha permitido llegar hasta este punto y terminar con esta etapa tan importante, gracias por enseñarme a salir adelante y a esforzarme por cumplir mis sueños, y más que sólo un agradecimiento por su apoyo, comparto con ellos este logro.

Agradezco a mis hermanos Anabel y Oscar, quienes me han brindado su apoyo y compañía, así como haberme alentado a continuar adelante en mis estudios y con quienes comparto momentos especiales y únicos.

Agradezco a Maricela por su compañía a lo largo de esta etapa, su comprensión, motivación y apoyo en cada momento han sido muy importantes en este proceso de formación, tanto académico como personal.

Agradezco a los compañeros y amigos con quienes he compartido momentos de estrés y alegría a lo largo de esta estancia y que gracias a ellos he podido mejorar día a día.

Acrónimos, abreviaturas y siglas

En este apartado se listan los acrónimos, abreviaturas y siglas usadas en este trabajo. Se incluye la respectiva traducción en los casos que así lo requieren y la página donde aparece citada por primera vez a fin de que el lector pueda conocer el contexto en el que fue enunciada.

3GPP 3rd Generation Partnership Project, es una colaboración de grupos de asociaciones de telecomunicaciones, conocidos como miembros organizativos, página 27

4G Tecnología de comunicación celular de cuarta generación, página 2

5G Tecnología de comunicación celular de quinta generación, página 2

6G Tecnología de comunicación celular de sexta generación, página 2

ACK Acuse de recibido, página 15

ALOHA Protocolo de comunicación aleatorio en redes de computadoras, página 1

ANN Artificial Neural Networks, (Redes neuronales artificiales), página 18

Batch Lote o sección de un conjunto de datos, página 52

Bluetooth Estandar industrial para redes inalámbricas de área personal, página 9

BTS Base Transceiver Station, (Estación base), página 16

CNN Convolutional Neural Network, (Red neuronal convolucional), página 20

CSMA Carrier Sense Multiple Access (Acceso múltiple con detección de portadora), página 5

DL Deep Learning, (Aprendizaje profundo), página 18

eNodeB Evolved Node B, componente que permite que los equipos de usuario conectarse a la red 4G, página 16

Feed-Forward Retroalimentación hacia adelante, página 25

GPU Graphic Process Unit, (Unidad de procesamiento gráfico), página 36

HSPA High-Speed Packet Access, es un estándar de internet móvil definido en la versión 7 de 3GPP, página 27

IFT Instituto Federal de Telecomunicaciones, página 9

IoT Internet of the Things, (Internet de las cosas), página 26

ITS Intelligent Transportation Systems, (Sistemas de transportación inteligente), página 16

LAN Local Area Network (Red de área local), página 14

LiFi Light Fidelity, Tecnología de comunicación inalámbrico por luz, página 9

LSTM Long Short Term Memory, (Memoria de largo y corto plazo), página 20

LTE Long Term Evolution, (Evolución a largo plazo), es un estándar de transmisión de datos móviles , página 27

LTU Linear Threshold Unit, Unidad que calcula una combinación lineal de pesos y entradas, página 19

LuST Luxemburg SUMO Traffic, simulación del tráfico con simulador SUMO para la ciudad de Luxemburgo, página 45

MAC Media Access Control (Control de Acceso al Medio), página 3

ML Machine Learning, (Aprendizaje de máquina), página 18

MLP Multi Layer Percptron, (Perceptron Multi Capa), es una red neuronal formada por múltiples capas, página 19

NFC Near Field Communication (Comunicación de campo cercano), página 10

OSI Open Systems Interconnection, modelo de interconexión de sistemas abiertos que permite que diversos sistemas se comuniquen usando protocolos estándar, página 12

RNN Recurrent Neural Network, (Red neuronal recurrente), página 20

RSU Road Side Unit, (Unidad a lado del camino), página 34

S-ALOHA Protocolo de comunicación de acceso aleatorio segmentado en el tiempo, página 2

Seq2Seq Secuencia a secuencia, página 23

SUMO Simulation of Urban Mobility (Simulación de movilidad urbana), página 7

TCP/IP Transmission Control Protocol/Internet Protocol (Protocolo de control de transmisión/Protocolo de Internet, página 12

TDMA Time-Division Multiple Access, (Acceso múltiple por división de tiempo), página 29

Throughput Tasa de transferencia efectiva, página 2

Transceptor Dispositivo que cuenta con un transmisor y un receptor y que comparten la circuitería, página 13

V2I Vehicle to Infrastructure, (Vehículo a Infraestructura), página 16

V2V Vehicle to Vehicle, (Vehículo a Vehículo), página 16

VANET Vehicular Ad Hoc Network, (Red Ad Hoc Vehicular), página 2

WiFi Wireless Fidelity, Tecnología que permite la comunicación inalámbrica entre dispositivos electrónicos, página 9

Índice general

Resumen	I
Abstract	II
Acrónimos, abreviaturas y siglas	IV
Índice de figuras	X
1. Introducción	1
1.1. Planteamiento del problema	2
1.2. Hipotesis	4
1.3. Objetivos	4
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos	4
1.4. Justificación	5
1.4.1. Beneficios esperados	7
1.4.2. Alcances y limites	7
1.4.3. Organización de la tesis	8
2. Protocolos de comunicación con aprendizaje de máquina	9
2.1. Redes inalámbricas	9
2.2. Protocolos de comunicación	11
2.2.1. Protocolos de Control de Acceso al Medio	13
2.3. Protocolo ALOHA	13
2.3.1. Pure ALOHA	14
2.4. S-ALOHA	15
2.5. VANETs	16
2.6. Aprendizaje de máquina	18
2.6.1. Long-Short Term Memory	21
2.6.2. Transformadores	23
2.7. Estado del Arte	26
3. Diseño y análisis de la propuesta	31
3.1. Propuesta de estudio	32
3.2. Red vehicular V2I	33
3.3. Protocolo S-ALOHA	34

3.4.	Estimación de τ	38
3.4.1.	Esquema fijo	40
3.4.2.	Esquema con media	40
3.4.3.	Esquemas basados en aprendizaje profundo	42
3.5.	Obtención de datos	45
3.6.	Estimación basada en transmisiones	48
4.	Experimentos y resultados de modelos de aprendizaje profundo en los protocolos de comunicación	54
4.1.	Esquema óptimo	55
4.2.	Esquema fijo	56
4.3.	Esquema con valor medio	58
4.4.	Esquema basado en redes LSTM	60
4.4.1.	Estimación por transmisiones detectadas, redes LSTM	62
4.5.	Esquema basado en Transformadores	64
4.5.1.	Estimación por transmisiones recibidas, Transformadores	66
4.6.	Comparación de resultados	68
5.	Conclusiones y trabajos a futuro	77
5.1.	Trabajos futuros	78
	Bibliografía	80

Índice de figuras

2.1. Sistemas comunicados mediante redes inalámbricas.	10
2.2. Clasificación de las redes inalámbricas con base en su cobertura [2].	11
2.3. Pila de protocolos para los sistemas de comunicaciones.	12
2.4. Clasificación de los protocolos en capa MAC.	14
2.5. Redes VANET [5].	17
2.6. Diagrama de una neurona McCulloch-Pitts.	19
2.7. Diagrama de red neuronal MLP [8].	19
2.8. Diagrama de red neuronal convolucional [10].	20
2.9. Diagrama de red neuronal recurrente [11].	21
2.10. Diagrama de una celda LSTM [11].	22
2.11. Diagrama de una Transformador [13].	24
2.12. Diagrama de una el modulo de atención (izq.) y la cabeza de atención múltiple (der.) [13].	25
3.1. Diagrama del sistema propuesto	33
3.2. Diagrama del proceso de entrenamiento	33
3.3. Diagrama de cobertura de las estaciones base	34
3.4. Representación de la organización del canal	36
3.5. Comparación de probabilidad de ranuras exitosas con diferentes valores de τ	39
3.6. Cantidad máxima, mínima y media de vehículos en una zona, así mismo varianza, desviación estándar de diferentes observaciones	41
3.7. Entrada a los modelos de predicción y salida para un tiempo k posterior.	42
3.8. Arquitectura modelo basado en red LSTM.	43
3.9. Arquitectura modelo basado en Transformadores.	44
3.10. Colocación de las estaciones base en la ciudad de Luxemburgo.	46
3.11. Conjunto de datos utilizado en el proyecto.	47
3.12. Histogramas del número de vehículos dentro de cada estación base a lo largo de las 24 hora.	48
3.13. Histogramas del número de vehículos dentro de cada estación base de las 4:00 a las 23:00 horas.	49
3.14. Comparación de número de vehículos contra número de ranuras necesarias para que al menos cada vehículo transmita una vez.	51
3.15. Comparación de número de vehículos (N) contra el número de vehículos detectados por la transmisiones recibidas en la estación base (\hat{N}).	52
3.16. Proceso de generación de instancias y entrenamiento de modelos de Aprendizaje de Máquina.	53

4.1. Métricas de desempeño del esquema con un valor óptimo de τ	56
4.2. Métricas de desempeño del esquema con un valor fijo de τ para cada hora del día	57
4.3. Métricas de desempeño del esquema con un valor de τ basado en la media de distintas observaciones para cada segundo del día	59
4.4. Métricas de desempeño del esquema con un valor de τ basado en redes neuronales LSTM de distintas observaciones para cada segundo del día.	61
4.5. Métricas de desempeño del esquema con un valor de τ basado en redes neuronales LSTM de distintas observaciones para cada segundo del día.	63
4.6. Métricas de desempeño del esquema con un valor de τ basado en Transformadores para cada segundo del día.	65
4.7. Métricas de desempeño del esquema basado en Transformadores que es entrenado con un oráculo y en la implementación recibe de entrada el número de vehículos detectados por la estación base.	66
4.8. Métricas de desempeño del esquema basado en Transformadores que es entrenado e implementado con entradas del número de vehículos detectados por la estación base.	67
4.9. Comparación de resultados obtenidos por los diferentes esquemas para la probabilidad de ranura con transmisión exitosa.	69
4.10. Comparación de resultados obtenidos por los diferentes esquemas para la probabilidad de ranura con colisiones.	70
4.11. Comparación de resultados en similitud respecto a los valores óptimos.	71
4.12. Comparación de resultados en el retardo de paquetes en esquema de actualización de paquetes por ranura.	72
4.13. Comparación de resultados en el retardo de paquetes en esquema de actualización de paquetes por transmisión.	73
4.14. Comparación del número de vehículos dentro de la zona de cobertura en un día normal contra un día en el que ha ocurrido un conflicto vial.	74
4.15. Métricas de desempeño en una zona de conflicto.	74
4.16. Estimación de paquetes actualizados en la zona de observación bajo una política de actualización por ranura.	75
4.17. Estimación de paquetes actualizados en la zona de observación bajo una política de actualización por transmisión de paquete previo.	76

Capítulo 1

Introducción

En este documento se presenta el proyecto realizado a lo largo del trabajo doctoral en el cual se estudió el uso de algoritmos de aprendizaje de máquina, o comúnmente conocido como *Machine Learning* por su traducción del inglés, en los protocolos de comunicación de sistemas inalámbricos. En especial se retoman protocolos que son de baja complejidad a fin de mejorar su desempeño y que puedan ser utilizadas en las ciudades inteligentes.

Este proyecto ha sido pensado para brindar alternativas de comunicación a los millones de dispositivos que hoy en día se conectan a los múltiples servicios ofrecidos. En particular se aborda sistemas vehiculares actuando como una red de sensores que reporta continuamente el estado del vehículo y las condiciones generales del tráfico.

Las redes de sensores se han convertido en herramientas vitales para monitorear y controlar diferentes fenómenos físicos. Su alto uso en estos dispositivos permite a otros sistemas tomar decisiones con base a las señales que detecta algún sensor, por ejemplo, un sensor que monitorea el clima exterior, será la señal base para controlar el sistema de clima automático en una casa o automóvil. De esta forma, la información recolectada por los vehículos es viable para ser ampliamente utilizada para las aplicaciones de ciudades inteligentes.

Por otro lado, la alta movilidad de los usuarios finales ha evolucionado al punto en que los dispositivos por sí mismos necesitan conectarse de forma transparente a las distintas redes que interactúan y se comunican a lo largo y ancho del planeta. Esto hace que hoy en día la comunicación se realice de una forma dinámica creando topologías que cambian constantemente y que, a la vez, demandan conexiones aún más eficientes y estables.

Un ejemplo de estas redes son las vehiculares, en las que constantemente se generan cambios en la topología, la demanda de recursos y la cantidad de usuario. Estos cambios constantes obligan a ingenieros, investigadores y administradores de red a proponer e implementar algoritmos y protocolos que permita adaptarse a estos cambios. En estos casos se recurre al uso de protocolos aleatorios que permite destinar recursos de canal de forma dinámica a los usuarios de una red. La importancia de los protocolos aleatorios recae en que facilita el intercambio de paquetes cortos para compartir información de control, estado de la red, estado de los dispositivos o solicitud de recursos fijos de canal, por lo que su uso continúa vigente.

Sin embargo, el uso de protocolos aleatorios impacta en un bajo desempeño de la red. Contrario a los protocolos programados, los aleatorios de baja complejidad tienden a presentar un bajo nivel de uso del canal, por ejemplo, específicamente hablando en el caso de ALOHA

ranurado (S-ALOHA), éste alcanza un *Throughput* del orden de 36.5 %. A lo largo de los años se han propuesto mejoras que han conducido a un mayor desempeño, en las que se realizan distintos ajustes para lograr este objetivo. Estos algoritmos para mejorar protocolos aleatorios tienen inmersos procesos de contención en los que los dispositivos intentarán transmitir su paquete información, y uno de los parámetros clave a ser ajustado es la probabilidad con la que los dispositivos intentan transmitir.

Mediante técnicas de aprendizaje de máquina se propone intuir el comportamiento que tienen los dispositivos y la propia red para adaptar dicha probabilidad de transmisión a las demandas actuales de conexión. La teoría que se plantea es utilizar estas técnicas para, lejos de pretender superar el desempeño de S-ALOHA, encontrar el valor para el parámetro conocido como probabilidad de transmisión de los nodos de tal forma que sea lo más cercano al óptimo teórico. Adicionalmente se parte del hecho de que las transmisiones recibidas en una estación base son la única manera de conocer la cantidad de usuarios conteniendo por recursos del canal dentro de una área determinada, por lo que es susceptible a que los modelos que estiman dicha cantidad estén recibiendo información con ruido. Con esta mejora se desea brindar usos del canal lo más cercanos al óptimo teórico al momento de ser implementado, y con ello ofrecer un canal capaz de permitir una comunicación más estable para los usuarios de la red.

Hoy en día los dispositivos inalámbricos están constantemente en comunicación con Internet a fin de descargar y subir información de toda clase, ya sea actualización de software y firmware, o incluso para acceder a contenido multimedia. De tal modo que la tarea de predicción de usuarios en la zona de interés tiene una aportación directa y es empleada en cada segundo en la planeación del protocolo de comunicación.

El caso particular que se aborda es para una red vehicular VANET (Vehicular Ad Hoc Network), en el cual los vehículos se conectan a una infraestructura de telefonía celular 4G o 5G en el contexto de ciudades inteligentes. Incluso se puede expandir al diseño de las emergentes redes 6G. La aplicación primordial con la que se planea este estudio es para el caso en donde los vehículos comparten información de tipo telemetría para la asistencia a la conducción autónoma, reporte de tráfico, reporte de contaminación, canales de acceso aleatorio para la solicitud de recursos fijos o alguna otra aplicación de asistencia en ambientes de ciudades inteligentes donde el flujo de información es de tal magnitud que es preciso utilizar la mayor cantidad de técnicas de comunicación posibles con el fin de maximizar el uso del canal.

1.1. Planteamiento del problema

Debido a la creciente demanda de recursos para comunicación inalámbrica, es necesario contar con protocolos de comunicación que sean más intuitivos y ofrezcan una comunicación estable. Muchos de los problemas de comunicación se debe a la escasez de canales en los que los nodos, llamase sensores, automóviles, teléfonos celulares, etc, transmiten su información. Sin embargo el problema principal es que no se puede incrementar deliberadamente la cantidad de canales, pues esto implica ocupar un ancho de banda más grande, o una utilización del canal por mayor tiempo para cada usuario. Además, la dinámica de una red inalámbrica limita la asignación de recursos de canal fijos a cada uno de los nodos.

Es por ello que se recurre a la capa MAC para establecer reglas de acceso al canal y que permita una óptima comunicación de los nodos de una red. Aquí existen dos tipos de mecanismos, los protocolos libres de contención y los protocolos de contención. Los protocolos de contención son ampliamente usados en redes inalámbricas, incluso son utilizados previo a una asignación fija de recursos por lo que su utilización continúa vigente.

En un protocolo de contención los nodos tratarán de comunicarse compitiendo por transmitir a su destino sin colisionar con la transmisión de otro miembro de la red. El mayor reto es establecer un mecanismo que garantice que los nodos no intervendrán en la transmisión de sus vecinos, y esto se logra mediante diferentes técnicas, por ejemplo, monitoreando el canal estableciendo probabilidades de transmisión o generando una sincronización con todos los miembros de la red.

Sin embargo, un gran problema es conocer cuántos usuarios están intentando transmitir, con esta información el protocolo podría operar en una forma muy cercana a la óptima teórica debido a que se brindaría la oportunidad de tener recursos de forma equitativa entre todos los usuarios. El problema principal recae en la forma en que se conoce el número real de usuarios que quieren utilizar el canal de comunicación. En ocasiones se utilizan las mismas transmisiones para conocer los identificadores o direcciones MAC de los dispositivos dentro de la red y con ello estimar el número total de nodos intentando transmitir. Pero esta técnica no siempre da ese valor real.

Por lo tanto, contar con un módulo que permita hacer una inferencia precisa del número de usuarios a partir de las transmisiones que se reciben es de gran ayuda para establecer protocolos de transmisión con mejor desempeño. En el caso de los vehículos se pueden utilizar elementos externos al sistema de comunicaciones como sensores de conteo de tráfico, cámaras que permite identificar los vehículos en una zona, entre otros. Desafortunadamente el contar con módulos externos puede conllevar un retardo en la actualización de los parámetros críticos. Es por ello que es necesario contar con un protocolo que posea de forma interna el mecanismo de predicción del número de usuarios para poder ajustar los parámetros críticos que ofrezca una mejor transmisión para los nodos. Este proceso de ajuste, en realidad es la capacidad del protocolo para adaptarse a las condiciones de los usuarios y las demandas del canal.

Con la inclusión de técnicas de inteligencia artificial es posible incluir este módulo de forma interna dentro del sistema computacional que aloja al protocolo de comunicación, y que con base en las transmisiones que se reciben en un determinado periodo de tiempo, puede estimar cuántos usuarios habrá en un futuro dado. Esto pensando en estaciones base con capacidades de procesamiento mínimas para la implantación de técnicas de inteligencia artificial. Así mismo, es importante que el sistema computacional esté dentro de la estación base con la intención de reducir el tiempo de predicción por el retardo inminente que representaría un módulo alojado en la nube o una instancia externa.

Cabe resaltar que este mecanismo es pensado en ambientes de contención donde los nodos necesitan acceder a la red para enviar paquetes cortos, principalmente de telemetría, y el tiempo de permanencia en una zona es corto, tal que asignar recursos fijos es poco viable o demanda forzosamente un periodo previo de contención. Dentro de la información que puede compartir un vehículo está su velocidad, su posición, la cantidad de emisiones que desprende etc. Por tal motivo el uso de protocolos de baja complejidad como S-ALOHA con parámetros adaptables a las condiciones y demandas de canal es una solución para

brindar acceso a dispositivos que comparten paquetes de información pequeños en ambientes altamente dinámicos como es el caso de una red VANET.

1.2. Hipotesis

El creciente uso de técnicas de inteligencia artificial en diferentes áreas de las ciencias ha mostrado que su uso mejora el rendimiento de los sistemas en los que se implementa. En el área de las comunicaciones inalámbricas, conocer el número real de usuarios permite ajustar los parámetros de transmisión de los nodos para optimizar el uso de un canal inalámbrico compartido. Es por ello que se tiene la hipótesis de que con el uso de técnicas de aprendizaje de máquina para la predicción de valores futuros, se podrá inferir el número de usuarios en la red para un tiempo futuro y, con ello, ajustar los parámetros de transmisión en un protocolo de baja complejidad basado en S-ALOHA para que opere de forma cercana al óptimo teórico.

Adicionalmente, el uso de estas técnicas permite crear un sistema que sea independiente de herramientas externas a la hora de ser desplegado para su funcionamiento, es decir que así se pueda inferir el número de vehículos accedendo a la red basado únicamente en las transmisiones que se detectan en la estación base. Dado que en múltiples ocasiones hay nodos que no logran transmitir, estimar la cantidad real de usuarios se vuelve una tarea complicada. Por ello la importancia de contar con un sistema de predicción que sea capaz de proporcionar dicha información al protocolo de comunicación.

1.3. Objetivos

A continuación se describe el objetivo general y los objetivos específicos del trabajo doctoral planteada en este documento.

1.3.1. Objetivo general

Analizar el desempeño de protocolos de comunicación en la capa de control de acceso al medio (MAC) para redes inalámbricas destinadas al monitoreo y envío de información de forma continua como lo es una red tipo VANET y que sean mejorados mediante técnicas de aprendizaje de máquina para la elección de parámetros de transmisión por parte de los dispositivos móviles, a fin de incrementar el *throughput* del canal y permitir que se optimice el desempeño de la red de acuerdo a las condiciones y demandas del canal en donde se realizan las transmisiones.

1.3.2. Objetivos específicos

- Identificar las características de una red inalámbrica que comprometen su desempeño en ambientes altamente dinámicos.
- Proponer un esquema de acceso al medio por protocolos de baja complejidad con políticas de transmisión basadas en la cantidad de dispositivos solicitando recursos de canal.

- Identificar los parámetros que mejoren el desempeño de la red y que permita incrementar la eficiencia de reporte al incrementar la posibilidad de que los dispositivo intenten transmitir en los periodos en que haya recursos de canal disponibles.
- Seleccionar las técnicas de aprendizaje de máquina que se adapten a los requerimientos para ajustar aquellos parámetros que ofrezcan una mayor utilización del canal en redes inalámbricas tipo VANET.
- Implementar las técnicas de aprendizaje de máquina en el ajuste de parámetros en el protocolo de comunicación con la finalidad de incrementar el desempeño de las redes inalámbricas al mejorar el uso del canal de transmisión.
- Generar simulaciones y modelos para visualizar el desempeño de la red inalámbrica con técnicas de aprendizaje de máquina.
- Comparar el desempeño de la red inalámbrica utilizando técnicas de aprendizaje de máquina para observar su comportamiento respecto al óptimo teórico.

1.4. Justificación

Con el surgimiento de nuevas tecnologías y conceptos como el de ciudades inteligentes (Smart Cities), las comunicaciones inalámbricas se han convertido en medios vitales y son donde se transmite un alto flujo de información inmerso en condiciones adversas. Es conocido que las comunicaciones inalámbricas permite a los usuarios libertad de movimiento, sin embargo, esto causa que las redes sean altamente dinámicas debido a que constantemente los vehículos entran, salen y permanecen conectados a la red por periodos de tiempo no definidos.

Por esta razón, los usuarios de comunicaciones inalámbricas normalmente no poseen un canal asignado en el que puedan transmitir libremente y se ven obligados a competir entre ellos por dichos recursos, o en su caso, esperar a que un agente administrador les asigne recursos a petición. A pesar de que hoy en día existen técnicas de compresión de datos, redundancia en los paquetes de información, técnicas de codificación bastante robustas e incluso predicción de datos perdidos, los recursos de canal de las capas bajas, llámese física y acceso al medio, continúan siendo limitados por los protocolos de comunicación de dichos niveles debido a la propia naturaleza finita de los canales de comunicación.

Este estudio se enfoca en abordar los protocolos de comunicación de la capa de control de acceso al medio, conocida como MAC. Dentro de los protocolos más conocidos se encuentran los protocolos ALOHA, en su versión pura y ranurado (S-ALOHA) y los protocolos de acceso al medio con detección de portadora conocidos como CSMA del inglés (Carrier Sense Multiple Acces) en sus distintas versiones; los cuales han sido la base para el desarrollo de diferentes estándares de comunicación. Actualmente, existen múltiples versiones de los protocolos clásicos de esta capa que han logrado y demostrado una mejora al rendimiento presentado en los protocolos base, y continuamente se proponen nuevas estrategias de acuerdo a las tecnologías que emergen año con año. Además, su uso continúan siendo vigentes para determinadas aplicaciones, siendo incluso necesarios para la solicitud de recursos de canal de ancho de banda para el envío de un alto volumen de información, como es el caso de los canales de control y de acceso aleatorio de los sistemas celulares.

Por otra parte, un problema para establecer parámetros similares al óptimo teórico es que en los sistemas inalámbricos es difícil conocer la cantidad de usuarios que realmente están intentando acceder a recursos del canal, esto hace que la información de entrada a modelos de estimación del número de usuarios sea ruidosa y estime valores incorrectos. Por lo tanto, contar con una mecanismo que permita hacer esta estimación de una forma más precisa originará que sea factible establecer parámetros que permitan al sistema lograr un desempeño cercano al óptimo.

Como anteriormente se mencionó, debido al vasto número de dispositivos conectados a alguna red inalámbrica, el contar con estos protocolos permite la comunicación de nodos o usuarios que envían información en entornos altamente cambiantes. Tal es el caso de una red Vehicular conocida como VANET, en la que los autos viajan a distintas velocidades, con distintas direcciones y permanecen conectados por periodos de tiempo heterogéneos. Aunado a ello, la información que comparten en la red va desde datos de telemetría, hasta contenido multimedia, donde las prioridades de los paquetes cambia de acuerdo a la aplicación empleada.

Por ello, es importante continuar realizando investigación entorno a los protocolos de comunicación desde sus formas más simples. Hoy en día la introducción de técnicas de inteligencia artificial han tenido mayor fuerza y esto ha sido el impulsor del desarrollo de nuevas aplicaciones. Gracias a estas técnicas se ha podido dar autonomía a distintos sistemas que van desde la detección de objetos, predicción de eventos futuros, tareas de procesamiento de lenguaje natural, entre otros. En el ámbito de las telecomunicaciones existen múltiples campos de uso, tales como reconstrucción de señales, predicción de tráfico, entre los más principales.

Tal como se ha observado en las diversas áreas mencionadas, el uso de inteligencia artificial es un hecho prometedor para mejorar el desempeño de los sistemas de comunicaciones inalámbricas. Con su uso se desea contar con herramientas más robustas que permita crear sistemas autónomos que sean capaces de tomar acciones correctas basados en las condiciones observadas dentro de las redes de comunicaciones.

Por otra parte, a pesar de que una misma técnica de aprendizaje de máquina puede ser aplicada a diferentes problemas, por ejemplo, la predicción de acciones de bolsa con redes recurrentes y la predicción de ventas de vuelos con un modelo similar, es necesario hacer una implementación correcta dependiendo del problema que se trata de resolver en la que se atiendan situaciones como la imprecisión de los datos, el sesgo de la información, contar con datos que sean significativos para la tarea que se desea solucionar, entre otras. En el caso de emplear aprendizaje de máquina para sistemas de comunicación inalámbrica que operan bajo esquemas aleatorios, un reto es conocer la cantidad exacta de usuarios dentro de un área de cobertura, ya que los nodos sólo son descubiertos en la red al momento de compartir un paquete de información. En este sentido, el proceso de entrenamiento del modelo de aprendizaje de máquina requiere de considerar esta situación para hacer una mejor generalización a la hora ser implementado. De este modo, aunque no se tenga la información exacta de cuantos nodos compiten por un canal, ajustar los parámetros de transmisión por medio de aprendizaje de máquina brinda la oportunidad de transmitir de forma equitativa a cada uno de ellos.

1.4.1. Beneficios esperados

Con esta tesis se pretende introducir modelos de predicción a los protocolos de comunicación inalámbricos de baja complejidad de capa MAC con el fin de mejorar el uso del canal por parte de los usuarios. De esta forma, el modelo de predicción basado en aprendizaje de máquina permitirá ajustar parámetros críticos para la transmisión de los nodos de tal forma que sea más factible alcanzar desempeños cercanos al óptimo teórico.

Al mismo tiempo, se desea generar modelos de aprendizaje de máquina capaces de predecir el número real de usuarios conteniendo por recursos de canal basado en las transmisiones exitosas que se reciben en la estación base a la que se comunican de forma directa los vehículos dentro de una red. Con la firme convicción de que conocer este dato permitirá manejar la transmisión de los nodos de una forma cercana al óptimo.

Con estas propuestas se analizará el desempeño y las mejoras observadas en el protocolo de comunicación comparando el uso de técnicas de aprendizaje de máquina contra las técnicas clásicas. De este modo, se comprobará su utilización en diferentes entornos con la intención de verificar tal desempeño particular para cada uno de los distintos escenarios.

1.4.2. Alcances y límites

En este estudio se propone el uso de técnicas de aprendizaje de máquina para la estimación de usuarios dentro de un área de cobertura y, basado en ello, ajustar el parámetro de transmisión en los protocolos de comunicación inalámbrica a fin de mejorar el desempeño del canal desde la capa de acceso al medio. El estudio se plantea para el caso especial en una red vehicular operando bajo un protocolo de contención basado en S-ALOHA

Se realizó un análisis previo de la información existente en una simulación vehicular para observar en que forma se puede abordar la implementación de aprendizaje de máquina, concluyendo que el problema se aborda desde el punto de vista de series de tiempo para estimar la cantidad de vehículos que habrán de contender en tiempos futuros, y con base en esa estimación ajustar la probabilidad de transmisión de los nodos operando en un protocolo de baja complejidad, como es el caso S-ALOHA, y éste es utilizado para que los vehículos transmitan continuamente paquetes cortos conteniendo información de telemetría dentro de una red Vehicular tipo Vehículo-a-Infraestructura.

Existe la limitante de la dificultad de contar con información real del número de vehículos conectados a una red VANET en la vida real, por lo tanto, se utilizan datos obtenidos del simulador de tráfico SUMO con lo cual es posible acceder a un modelo operando en el óptimo teórico en donde la información es precisa, y que sirve de comparación para el desempeño obtenido con la propuesta hecha en esta tesis. Adicionalmente se simula el protocolo de contención. Del simulador de tráfico vehicular y en conjunto con el simulador del protocolo S-ALOHA se obtienen secuencias para entrenar los modelos de aprendizaje de máquina. El desempeño del protocolo se mide mediante la simulación del protocolo S-ALOHA con ajuste de parámetros de transmisión con base a las estimaciones de los modelos de aprendizaje de máquina.

El estudio se limita a obtener resultados teóricos y se compara el desempeño con respecto a resultados de S-ALOHA. Este desarrollo teórico servirá como parteaguas para añadir mejoras a otros protocolos de comunicación e incluso a las demás capas de comunicación a fin de crear

sistemas más robustos que garanticen procesos de transmisión más eficientes.

Para asegurar la visualización de los resultados obtenidos en cualquier dispositivo, el trabajo no considera unidades de consumo de energía por dispositivos comerciales, pues se basa en obtener el número de transmisiones y recepciones que realizan los vehículos, además, las ranuras de tiempo para que los nodos transmitan su paquete de información son definidas de acuerdo a los trabajos encontrados en el estado del arte. Este último valor juega un papel clave ya que afecta directamente la información con la que operan los modelos de aprendizaje de máquina.

La generalización de esta propuesta a otros protocolos no se comprueba en este estudio, por lo que de ser necesaria su implementación en un esquema de transmisión diferente, es necesario replantear el análisis correspondiente al nuevo sistema para que la utilización de aprendizaje de máquina sea la correcta para ajustar los parámetros correspondientes.

1.4.3. Organización de la tesis

Capítulo 1. Contiene la introducción y la motivación que dio origen a este trabajo, en la cual se habla de las generalidades del trabajo, se plantean los objetivos, justificación así como, beneficios esperados, alcances y límites.

Capítulo 2. Contiene las bases teóricas necesarias para abordar el trabajo reportado en esta tesis así como el estado del arte.

Capítulo 3. Describe el diseño de la propuesta hecha para dar solución al problema planteado en el Capítulo 1, así mismo se describen los escenarios sobre los cuales se estipula y evalúa la solución propuesta.

Capítulo 4. Presenta las pruebas hechas a los modelos y al protocolo mediante su simulación con sus respectivos resultados y análisis.

Capítulo 5. Contiene las conclusiones del trabajo realizado, así como los trabajos futuros derivados de la tesis.

Capítulo 2

Protocolos de comunicación con aprendizaje de máquina

En este capítulo se presenta el fundamento teórico necesario para desarrollar esta tesis, así como el estado del arte que lo inspiró. Dentro de los temas que se tratan en el capítulo, se encuentra las redes vehiculares VANET, el protocolo de comunicación S-ALOHA y las técnicas de aprendizaje de máquina para tareas de predicción. Especialmente se retoman aquellas empleadas para predicción de secuencias o series de tiempo.

2.1. Redes inalámbricas

Las redes inalámbricas son aquellos sistemas de comunicación compuestos por dispositivos que se conectan, tanto a un concentrador, como entre ellos mediante enlaces inalámbricos. Estas comunicaciones se realizan utilizando el espectro radio eléctrico, y cada sistema accede mediante diferentes frecuencias, es por ello que múltiples sistemas conviven y comparten información al mismo tiempo, además de otras técnicas como la división espacio-temporal y por código.

El espectro electromagnético es un recurso limitado y regulado por una determinada institución en su país correspondiente. Este instituto, en el caso de México, el Instituto Federal de Telecomunicaciones IFT [1] es el organismo que asigna y concede las diferentes bandas de frecuencia a las empresas y demás usuarios. Además, existen bandas que son asignadas como de libre uso, a las cuales los usuarios pueden acceder de forma gratuita. Estas bandas, son de las más congestionadas y típicamente son empleadas para comunicaciones locales, tal es el caso de WiFi o Bluetooth. Adicionalmente, el uso de nuevas tecnologías como LiFi, la cual utiliza luz visible para comunicar dos dispositivos de forma directa, son un ejemplo más de un sistema de comunicación inalámbrica.

Sin embargo, existen sistemas que se comunican mediante enlaces inalámbricos que no son abiertos al público en general y ocupan determinadas bandas de frecuencia dependiendo de la concesión a que han sido asignadas, tal es el caso de las redes satelitales, las redes de microondas, o las redes de telefonía celular, las cuales son operadas por empresas a las que se le otorgan espacios específicos dentro del espectro radio eléctrico para su transmisión.

En general las redes de comunicaciones inalámbricas se enfrentan a diferentes problemas

tener casi contacto entre los dispositivos. Posteriormente existen las redes de área personal en donde un claro dominante es Bluetooth, seguidas de las redes de área local donde un principal representante de estas redes son las basadas en WiFi. Para redes de mayor cobertura, existen aquellas de área a nivel vecindario e incluso las redes de área amplia, que ya alcanzan distancias de 5km hasta más de 100km. En la Figura 2.2 se presenta un diagrama de la clasificación de las redes respecto al área de cobertura

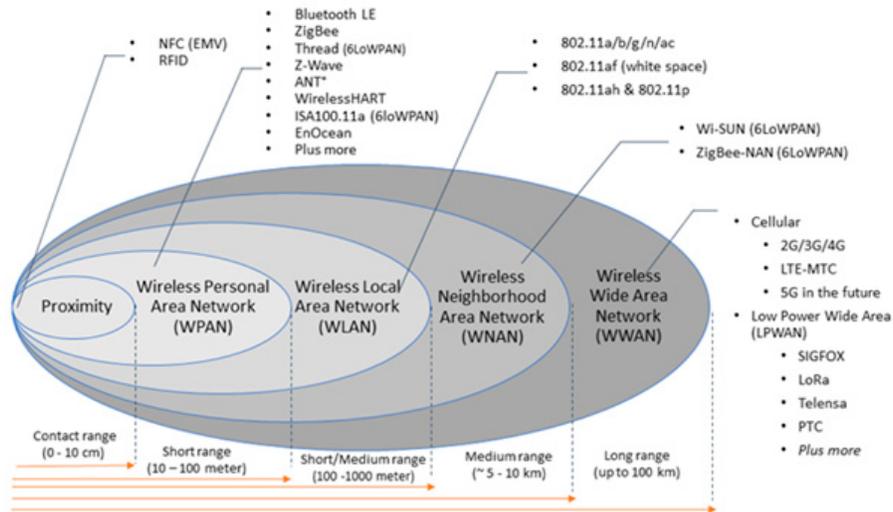


Figura 2.2: Clasificación de las redes inalámbricas con base en su cobertura [2].

Independientemente del tipo de red inalámbrica que se trabaje, en cualquier caso es necesario contar con un determinado protocolo de comunicación. A continuación se presenta en la siguiente sección los protocolos de comunicación utilizados en redes inalámbricas.

2.2. Protocolos de comunicación

Para poder establecer una comunicaciones de calidad, los dispositivos dentro de una red requieren de reglas y mecanismos que generen una coordinación para evitar conflictos en las transmisiones y así lograr el correcto intercambio de información.

Con la intención de dividir las tareas dentro del sistema, se han establecido ciertas estructuras en la que se delegan tareas a diferentes entidades. Estas estructuras son conocidas como pila de protocolos y está dividida en capas con determinadas funciones para realizar ajustes a las tareas de comunicación en cada una de ellas. Los sistemas más simples constan de al menos tres capas enlistadas a continuación.

- Capa Física
- Capa de Enlace de Datos
- Capa de Red

Para aquellos sistemas más complejos se consideran dos capas adicionales que son:

- Capa de Transporte
- Capa de Aplicación

En la Figura 4.15 se presenta un diagrama de esta pila de protocolos.

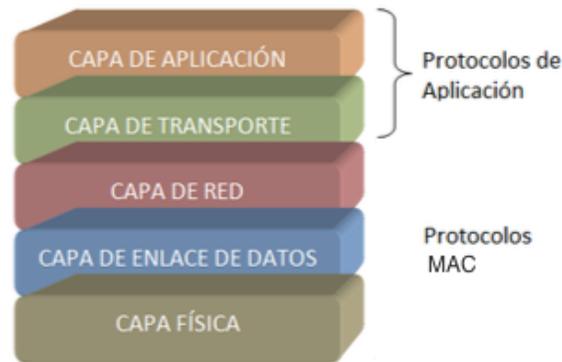


Figura 2.3: Pila de protocolos para los sistemas de comunicaciones.

La capa Física es encargada de seleccionar la frecuencia de transmisión, los esquemas de modulación así como la potencia con la que los transmisores y receptores operarán. Adicionalmente cumple funciones de detección de las señales portadoras e incluso encriptación de la información.

La capa de Enlace de Datos se responsabiliza de la multiplexación en el canal para el flujo información, así mismo es encargada de detectar los inicios y finales de las tramas de datos. Esta capa es conformada a la vez por una par de subcapas, la sub capa de Control de Acceso al Medio (Media Access Control, MAC) y la sub capa de control de enlace lógico (Logical Link Control). En algunas ocasiones la capa Física y la capa de Enlace de Datos se fusionan en una sola capa, por ejemplo, en el caso de la pila de protocolos TC/IP.

La capa de Red se encarga de elegir con que nodos se comunica, es decir a quien va dirigida la información y con ello se establece una ruta para la transmisión de información considerando las condiciones de la red y las prioridades en los servicios requeridos.

Las capa de Transporte se encarga del manejo y segmentación de los paquetes, a fin de recuperar mensajes completos sin error y en la secuencia correcta, incluso cuando hay duplicidad de paquetes recibidos.

Para finalizar la descripción de esta pila de protocolos se menciona la capa de Aplicación, la cual actúa como un agente de comunicación entre la aplicación del sistema y la red de comunicación, aquí se procesa toda la información recolectada y se almacena o presenta en el formato que la aplicación por si misma lo requiera.

Cabe notar que el modelo de Interconexión de Sistemas Abiertos (OSI) define cómo tal 7 capas, incluyendo entre la capa de aplicación y de transporte a la capa de Sesión y la capa de Presentación. Sin embargo, es común observar la fusión o mezcla de diferentes capas como en el caso de los protocolos de red TCP/IP que incluye estas dos últimas en la capa de aplicación, obteniendo así, la pila descrita previamente y que es la que se considera en este trabajo.

En particular, en esta tesis se estudia únicamente protocolos de la sub capa MAC, en específico el protocolo S-ALOHA

2.2.1. Protocolos de Control de Acceso al Medio

El proceso de acceder al medio de comunicación dentro de una red inalámbrica se lleva a cabo por la sub capa MAC, para ello se atienden dos premisas importantes que es el correcto establecimiento de enlaces de comunicación y un reparto justo de los recursos de canal para garantizar el intercambio de información con los dispositivos miembros de la red. De este modo, la capa MAC cuenta con la habilidad de controlar cuando se utilizará el transceptor para enviar información, cuya tarea es de alto consumo de energía y existirán sistemas en los que este aspecto repercute altamente, como es el caso de las redes de sensores WSN. Aunque actualmente en vehículos automotores, este hecho se puede considerar despreciable, ya que los autos poseen un generador que mayormente es impulsado por un motor de combustión interna, un protocolo MAC eficiente en energía será altamente benéfico en un futuro no muy lejano cuando los vehículos eléctricos requieran economizar al máximo la energía.

Dentro de los protocolos de MAC existen diferentes tipos y típicamente son clasificados como programados y de contención. En los protocolos programados los dispositivos de una red cuentan con recurso de canal asignados de forma fija y por tanto son libres de contención; esta clase de protocolos demandan de una correcta sincronía entre emisores y receptores y los recursos de canal son asignados por medio de frecuencias, tiempos, una mezcla de ambas, o incluso por códigos. Generalmente estos protocolos se usan en redes que no cambian en su topología o son de baja dinámica, esta clase de protocolos también son conocidos como protocolos libres de contención. Por otro lado los protocolos de contención son aquellos en los que los dispositivos están constantemente en un proceso de envío de paquetes intentado que éste llegue al destino de forma íntegra, aquí ningún nodo cuenta con recursos asignados, por lo tanto es susceptible a presentar colisiones y tiempos de uso de canal desperdiciados. Adicionalmente, se han realizado mejoras a los protocolos, por ejemplo, el uso de ranuras de tiempo en los que los nodos pueden transmitir. Este aspecto demanda contar con sincronía para saber en que momento comienza y termina la ranura, sin embargo, logra duplicar el desempeño de la red respecto al uso del canal, como es el caso de ALOHA puro, contra S-ALOHA, de los cuales se hablará más adelante.

Otra de las mejoras observadas a los protocolos de comunicación es la fusión de esquemas de contención seguidos de esquemas programados, este tipo de protocolos son conocidos como protocolos híbridos, en los que después de un proceso de contención, los dispositivos son asignados a recursos fijos para mejorar el uso del canal e incrementar las tasas de transmisión. Este hecho permite la auto-organización de la red, así como tolerancia a una mayor dinámica en la red. En la Figura 2.4 se presenta un diagrama conceptual de los tipos de protocolos MAC con el ejemplo de ellos.

2.3. Protocolo ALOHA

Es un protocolo de acceso por medio de contención, también definido como de acceso aleatorio, el cual permite la transmisión de múltiples dispositivos en un canal de comunica-

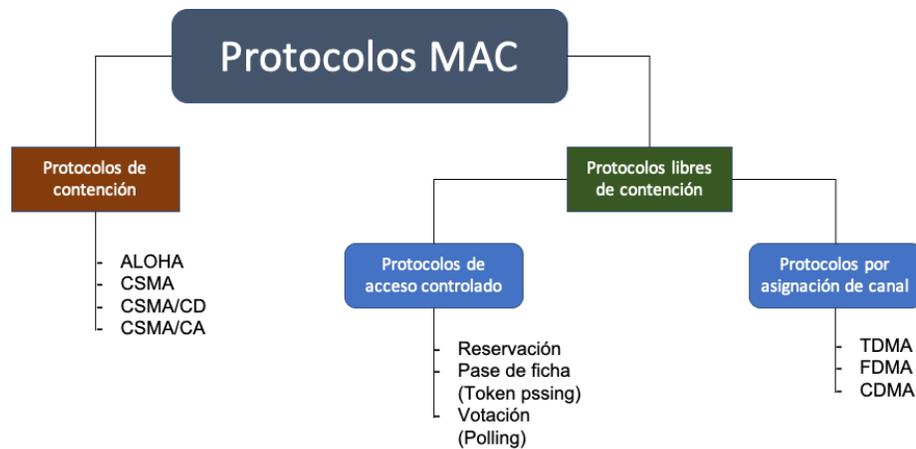


Figura 2.4: Clasificación de los protocolos en capa MAC.

ciones común y compartido. Es considerado un protocolo simple donde cada uno de los nodos transmisores tienen la misma prioridad a la hora de intentar acceder al canal y los mismo operan de manera independiente, lo cual lo hace viable para operar en sistemas inalámbricos en donde la dinámica de la red es alta al no asignar recursos de canal de forma fija, sin embargo, es susceptible a presentar problemas de colisión ya que cada nodo transmite su información sin considerar a los demás nodos.

Este protocolo fue propuesto en el año 1968 en la Universidad de Hawaii. Originalmente este protocolo fue pensado para redes inalámbricas LAN y posteriormente ha ganado popularidad a lo largo del mundo de las redes que comparten un canal inalámbrico [3]. La principal ventaja que caracteriza a ALOHA es ser un protocolo económico ya que no demanda equipamiento muy sofisticado para lograr comunicación entre dispositivos inalámbricos.

Existen dos versiones del protocolo ALOHA, el primero ALOHA puro (Pure ALOHA) y la versión que considera un canal segmentado en ranuras de tiempo ALOHA ranurado (Slotted ALOHA, S-ALOHA)

2.3.1. Pure ALOHA

Es el protocolo base de ALOHA, el cual es un protocolo de contención en donde los paquetes de información provienen de diferentes nodos y son transmitidos a través de un canal común con base en la demanda de los usuarios, es decir, cuando un usuario tenga un paquete que transmitir, lo realizará sin importar que algún otro usuario esté transmitiendo. Esto da origen a un alto número de colisiones entre paquetes. Un aspecto importante es que la colisión se puede dar en cualquier instante, por lo tanto, una transmisión exitosa únicamente se dará cuando el paquete ha alcanzado su destino.

Cualquier colisión conllevará a una pérdida de los paquetes involucrados. Dado a que los nodos transmiten sin considerar si el canal está ocupado, durante todo tiempo en que el paquete está viajando en el canal se vuelve susceptible a colisionar, ya sea que algún nodo más de la red comience su transmisión al principio, en medio o al final del paquete que se encuentra ya ocupando el canal.

Posterior a su transmisión, los nodos saben si han alcanzado su destino mediante mensajes de acuse de recibido conocidos como *acknowledge* (ACK). En el caso de colisión, los nodos involucrados no detectarán este ACK, por lo que el nodo sabe de esta situación y los paquetes involucrados de los cuales no se recibe ACK necesitaran ser reenviados.

ALOHA incorpora una política de retransmisión para los paquetes que no se enviaron de forma exitosa debido a una colisión. En este caso, las nodos involucrados esperarán por un tiempo aleatorio independiente para volver a retransmitir. Este tiempo es conocido como *back-off time* y después de ese tiempo el nodo enviará nuevamente el paquete del que no recibió su ACK. La implementación de este tiempo *back-off* con valores aleatorios para cada nodo, brinda una opción para evitar nuevamente colisión entre dichos nodos.

A pesar de estas implementaciones, el protocolo ALOHA es poco eficiente, ya que apenas alcanza un 18.4% de eficiencia, esta medición corresponde al *throughput* máximo del protocolo.

2.4. S-ALOHA

El protocolo S-ALOHA es la mejora propuesta al protocolo Pure ALOHA con la intención incrementar la eficiencia observada previamente. Dado a que los nodos transmiten de forma libre en cualquier instante de tiempo, la presencia de colisiones se puede dar en cualquier momento mientras un mensaje se está transmitiendo, creando una inmensa cantidad de posibilidades para que esto ocurra. Para solucionar esto, en S-ALOHA el canal de comunicación se segmenta en múltiples ranuras a lo largo tiempo lo cual crea un canal donde los nodos transmiten de forma compartida pero ahora es discreto en el tiempo [4].

El canal ahora está dividido en el tiempo con pequeñas ranuras de duración t_{slot} para cada una de ellas. Cada dispositivo también tiene la posibilidad de transmitir cuando tiene un paquete de información que enviar, pero para hacerlo debe esperar a que sea el inicio de la ranura de tiempo, es decir, las transmisiones solo se pueden comenzar al inicio de alguna ranura de tiempo y en consecuencia las colisiones solo han de ocurrir al inicio de la ranura de tiempo. De esta forma se elimina la posibilidad de que exista colisiones intermedias en la duración de la ranura de tiempo. Con esto se aprecia que si una colisión ocurre, el o los paquetes involucrados colisionan completamente y no una parte de ellos.

Por lo tanto la transmisión exitosa de información únicamente se dará cuando sólo una transmisión se realiza de forma individual en una ranura de tiempo. En el caso en que ningún nodo intente transmitir en una cierta ranura de tiempo, a ésta se le conoce como ranura inactiva (en Inglés, idle slot) o ranura vacía.

Similar a Pure ALOHA, los dispositivos conocen que su paquete de información se ha enviado de forma exitosa tras recibir un ACK y en caso de no recibir esta confirmación, los nodos intentarán transmitir nuevamente en la ranura correspondiente al *back-off time* que obtengan de forma aleatoria.

Con el canal ranurado a lo largo del tiempo, el protocolo S-ALOHA alcanza una eficiencia del doble al obtenido en Pure ALOHA, por lo que en este caso su eficiencia máxima es del 36.8%. Sin embargo, el protocolo S-ALOHA no logra eliminar completamente la presencia de colisiones, y a lo largo de la evolución de los sistemas de comunicaciones, se han propuesto múltiples protocolos. Una opción son la serie de protocolos de Acceso al Medio por Detección

de Portadora, conocidos como CSMA, que escuchan el canal antes de comenzar a transmitir y cuentan con una versión con prevención de colisiones, conocida como CSMA/CA. Este tipo de protocolos no se abordan en esta tesis

2.5. VANETS

Las redes vehiculares, conocidas como VANETs por su traducción del Inglés Vehicular Ad Hoc Networks son redes creadas por vehículos en movimiento y que se conectan de forma inalámbrica para compartir información entre ellos. Este tipo de redes han emergido y son altamente atractivas gracias a su bajo costo, su alta capacidad de monitoreo de información, e incluso gozan de tolerancia a errores. Normalmente se utilizan para comunicación en los vehículos que se encuentran en autopistas o en zonas urbanas. Su principal uso es entorno a la seguridad vial y el reporte de condiciones críticas que comprometan la integridad de las personas, así como para servicios externos a la red como estadísticas de movilidad, aunque hoy en día también se estudia su utilización para brindar acceso multimedia a los ocupantes de los vehículos a fin de contar con servicios de entretenimiento.

Una VANET es considerada una red libre en donde los nodos pueden ingresar o salir en cualquier momento, por lo que tiene una dinámica alta tanto de información como de los usuarios que interactúan en la red. Cada vehículo cuenta con la autonomía de decidir que información es útil para el resto de los vehículos (estos son los nodos de una VANET) y gracias a que hoy en día los automóviles cuentan con más y sofisticados sensores, el mar de información que puede generarse dentro de una VANET puede ser utilizado en múltiples aplicaciones que van desde el monitoreo de condiciones ambientales o de condiciones de tráfico, hasta información de apoyo a la conducción autónoma de vehículos, esto sin descartar servicios Web y multimedia.

Para lograr su comunicación existen diferentes opciones, en algunos casos se equipan con componentes dedicados para comunicación de corto alcance, transmisores y receptores basados en WiFi, equipos de radiocomunicación con estaciones celulares e incluso sistemas sistemas de comunicación satelital. Gracias a esta amplia variedad para comunicarse de forma independiente sin necesitar la intervención de las personas, estas redes se han convertido en una parte vital para los sistemas de transportación inteligente (Intelligent Transportation Systems, ITS). Así mismo, la red es bastante robusta pues permite comunicar a los vehículos entre ellos mismos o comunicarse a un punto de acceso lo que brinda la opción de buscar rutas alternas cuando existen puntos sin conexión.

Las topologías posibles en una red VANET son dos: las Vehículo a Vehículo (Vehicle to Vehicle, V2V) y las Vehículo a Infraestructura (Vehicle 2 Infrastructure, V2I). En la V2V los automóviles tiene la capacidad de mandar información directamente a otros automóviles en la red, incluso alcanzar autos que se encuentren dentro de la red en extremos opuestos mediante multi-saltos. Por su lado, la topología V2I es aquella en la que los nodos se comunican a un punto de acceso, ya sea un punto de acceso dedicado, algún punto WiFi o una estación base celular (BTS, eNodeB, etc, dependiendo la generación de telefonía celular). Opcionalmente ambas topologías pueden trabajar en conjunto para brindar mayor robustez y tolerancia a fallos de comunicación, ya sea entre vehículos o con el punto de acceso. En la Figura 2.5 se presenta un diagrama típico en la literatura donde se ejemplifican la VANETs.

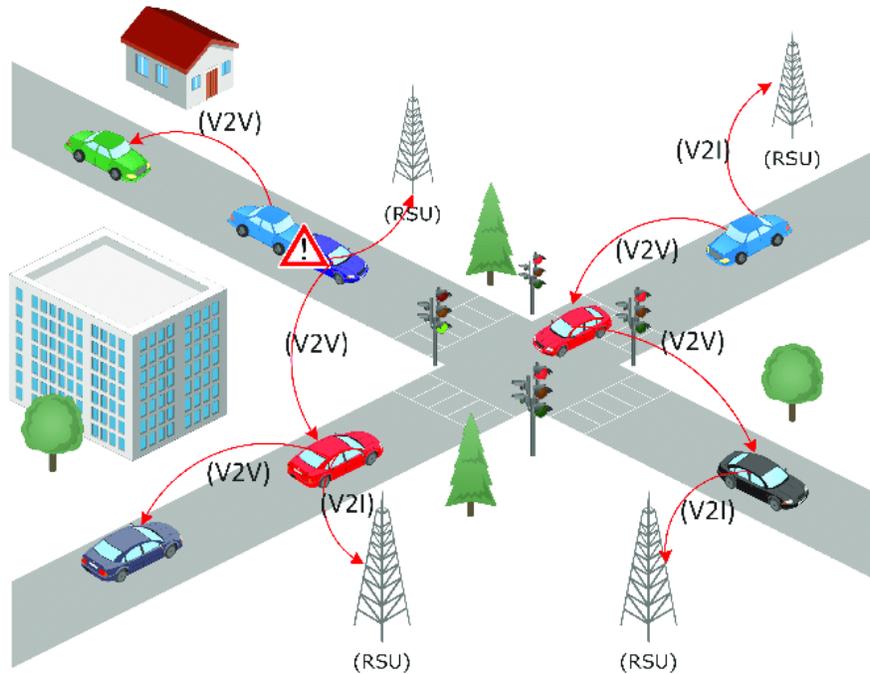


Figura 2.5: Redes VANET [5].

Esta clase de redes poseen diferentes características que las hacen interesantes y atractivas para su investigación. En primer lugar poseen una alta dinámica que causa cambios constantemente en la topología de la red. Sin dejar de lado que sea en forma V2V, la movilidad de los vehículos a diferentes velocidades y de forma aleatoria, cambia la posición de los nodos constantemente y en consecuencia la topología de la VANET se está modificando por cada cambio. Para el caso de una topología V2I sucede algo similar ya que diferentes nodos se pueden estar conectando y desconectando del área de cobertura del punto de acceso. Por otro lado, este mismo hecho provoca que el tamaño de la red no esté totalmente acotado, ya que la red podrá expandirse mientras haya un nuevo nodo que agregar, pese a que el área de cobertura de un punto de acceso sea un impedimento físico para la cantidad de nodos que pueden agregarse, es importante recordar que tanto los vehículos como el punto de acceso tienen posibilidad de conectarse a otros dispositivos externos de la red inicial con el objetivo de lograr una mayor expansión.

Por otro lado, es importante señalar que el flujo de información en las VANETs es muy alto, ya que al existir multi-saltos para alcanzar nodos distantes, los vehículos se ven en la necesidad de recolectar información de un gran número de vehículos. Esto obliga a que los nodos deban poseer una capacidad de almacenamiento suficientemente grande para tratar tal contenido, así como fuentes de energía que no tengan problemas con el alto número de transmisiones y recepciones que han de realizar los vehículos. Debido a que los automóviles contienen múltiples sensores, estos también son fuentes de información que en algún momento será necesaria transmitir para aplicaciones, tanto de la VANET como externas por lo que los retos en el manejo de información y energía son características con las cuales lidiar en las VANETs.

Actualmente los retos a los que se enfrentan las VANETs motiva a los investigadores a buscar mejoras en diferentes aspectos, los cuales abarcan seguridad de la información en la VANET, garantías en calidades de servicio, enrutamiento, así como aspectos técnicos como los son el manejo de energía en la red, control de congestión y colisión de información, diseño de protocolos MAC, e incluso su impacto ambiental.

Cómo mejora a los protocolos en distintos sistemas de comunicaciones inalámbricas se propone el uso de técnicas de aprendizaje de máquina que permita lograr tal reto de establecer nuevos protocolos de capa MAC por lo que en las siguiente sección se plantean los mecanismos de aprendizaje de máquina para comprender su funcionamiento.

2.6. Aprendizaje de máquina

El aprendizaje de máquina, que comúnmente se le conoce Machine Learning (ML), es considerada la ciencia de crear programas computacionales que permitan a un sistema aprender de los datos, de esta forma, el programa adquiere la habilidad de inferir o detectar patrones claves con únicamente tener grandes cantidades de información o ejemplos del problema que desea resolver. En la literatura existen diferentes definiciones de acuerdo al autor, la más orientada a una definición de ingeniería es la que dió Tom Mitchell en 1997:

"Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y alguna métrica de desempeño P , si su desempeño en T medido por P , mejora con la experiencia E "

Por lo tanto, los aspectos claves para utilizar ML, es contar con un amplio número de datos que permita al programa tomar experiencia para poder realizar de mejor manera una tarea determinada. Dentro de las tareas más populares en ML se encuentra la detección y clasificación de correos no deseados (Spam) en servidores de correos, la aprobación de créditos bancarios, sistemas de recomendación de videos, música, entre otros; o la predicción de valores futuros en distintas series temporales. Por igual, el mundo de ML ha incursionado en tareas más complejas como reconocimiento facial o tareas de lenguaje natural, las cuales ocupan modelos más complejos que suelen clasificarse como Aprendizaje profundo o conocidos como Deep Learning (DL).

Uno de los algoritmos altamente usados son las Redes Neuronales Artificiales (ANN), incluso son la parte medular del Deep Learning. La Redes Neuronales fueron introducidas en el año 1943 por el neurofisiólogo Warren McCulloch y el matemático Walter Pitts [6]. Éste era un modelo simple inspirado en la habilidad de las neuronas biológicas para trabajar juntas, con el cual se realizaron cálculos computacionales complejos. La neurona artificial es un modelo que simplemente activa su salida cuando un cierto número de entradas son activadas. Ellos mostraron que con una red de simples neuronas se podía calcular cualquier preposición lógica. En un inicio el modelo empleaba valores lógicos de 1 y 0 y sus salidas eran del mismo tipo, con el tiempo se fue perfeccionando para operar con otro tipo de entrada. En la Figura 2.6 se presenta un esquema de la neurona McCulloch-Pitts donde ya se consideran distintas entradas s que permitirá que se active una función con base al valor del umbral θ [7].

Una de las arquitecturas más simples de las redes neuronales artificiales fue el perceptron introducido en 1957 por Frank Rosenbaltt. Su modelo fue conocido como Unidad de Umbral

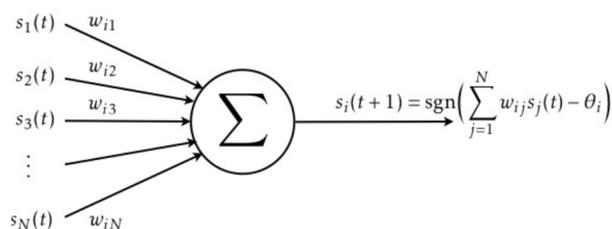


Figura 2.6: Diagrama de una neurona McCulloch-Pitts.

Lineal (Linear Threshold Unit) donde las entradas y salidas eran números y no sólo valores lógicos, cada una de las entradas es asociada a un determinado peso con el cual era ponderado. La LTU calculaba una suma ponderada de las entradas para posteriormente aplicar una función paso de dicha suma, ésta es la salida de la neurona. Luego entonces, el perceptron es simplemente una capa compuesta por diferentes LTUs que comparten todas las entradas. A pesar de ello, estas redes tenían ciertas limitaciones, por ejemplo, no podían resolver el problema XOR. Para ello se apilaron múltiples unidades LTUs para generar más de una capa de neuronas, dando origen a un perceptron multicapa, conocido como Multi-Layer Perceptron (MLP). En la Figura 2.7 se presenta un esquema de una red MLP, donde se observa una capa de entrada (\mathbf{X}), un conjunto de capas ocultas y finalmente una capa de salida (\mathbf{Y}), cada capa compuesta por un conjunto de neuronas tipo perceptron.

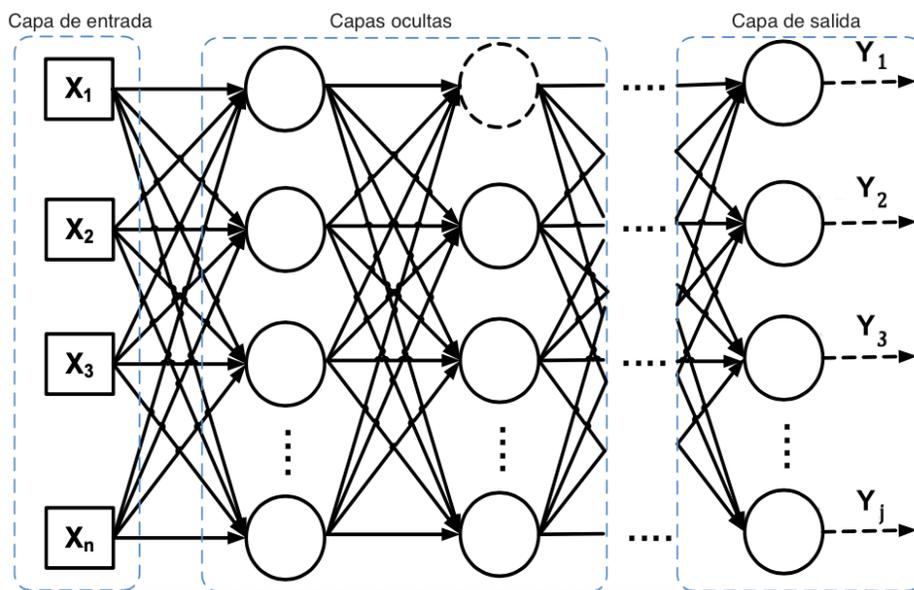


Figura 2.7: Diagrama de red neuronal MLP [8].

A partir de esta idea se comenzó a observar que los modelos con múltiples capas son capaces de realizar tareas bastante complejas, dando paso a lo que hoy se conoce como aprendizaje profundo o Deep Learning (DL). Deep Learning comprende modelos con múltiples capas ocultas de redes neuronales que, a su vez, cada capa puede contener de cientos e incluso miles de neuronas. En estos modelos las capas poseen una cantidad muy grande de neuronas interco-

nectadas y que normalmente poseen diferentes funciones de activación que no necesariamente son lineales, como lo eran en el caso del MLP, con lo cual obtienen la capacidad de realizar tareas muy complejas para una computadora con desempeños muy similares a los que obtendría una persona.

Con la introducción de los modelos de Deep Learning, las computadoras adquieren habilidades para tareas complejas como detectar y clasificar rostros dentro de una fotografía, identificar objetos en un video, generar imágenes artísticas a partir de un boceto, analizar un texto escrito en una carta o en una red social, predecir el movimiento de un automóvil en la calle, hacer una predicción bastante cercana para las acciones bursátiles, predecir el clima dadas las condiciones actuales, o incluso hacer traducción a distintos lenguajes [9].

Dentro de los modelos de Deep Learning más populares, encontramos a las redes neuronales convolucionales conocidas como CNN [10], y las redes neuronales recurrentes conocidas del inglés como RNN [11]. Las CNN son ampliamente usadas en el análisis de imágenes, mientras que las RNN se emplean para análisis de secuencias y la predicción de valores futuros. Nótese que las secuencias pueden ser de diferentes tipos, desde secuencias de números hasta secuencias de imágenes para el análisis de audio, voz, lenguaje y videos; en este último caso, el modelo es un híbrido de CNN que alimenta a las RNN. En las Figuras 2.8 y 2.9 se presentan los diagramas de los autores previamente citados para representar a las redes neuronales convolucionales y recurrentes, respectivamente.

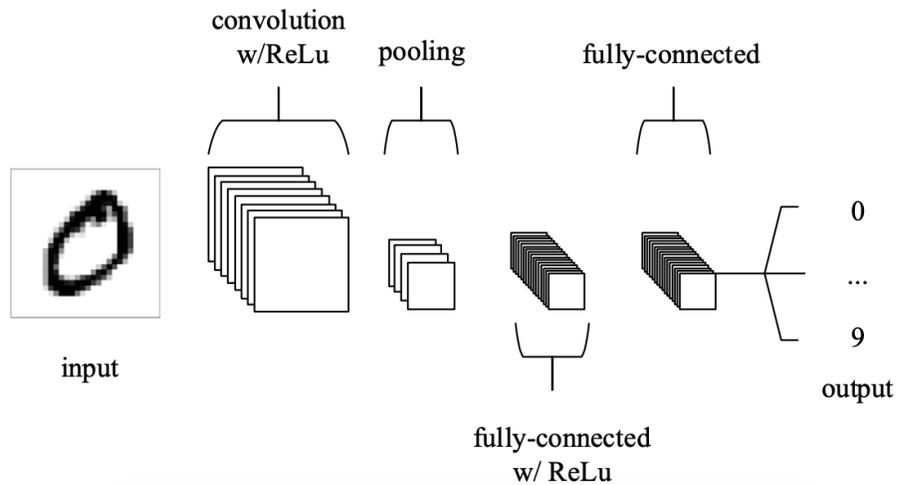


Figura 2.8: Diagrama de red neuronal convolucional [10].

En este proyecto se emplean modelos de Deep Learning para resolver tareas de secuencia de números, en este caso se abordan secuencias con el número de vehículos dentro de un área determinada, de la cual se dará mayor detalle en el siguiente capítulo. Por lo tanto uno de los modelos que se propone está basado en Redes Neuronales Recurrentes, específicamente se aborda un modelo que ha mostrado buenas precisiones en tarea de predicción de secuencias, éste es conocido como Long Short Term Memory o LSTM. Así mismo se considera un modelo conocido como Transformadores o Transformers.

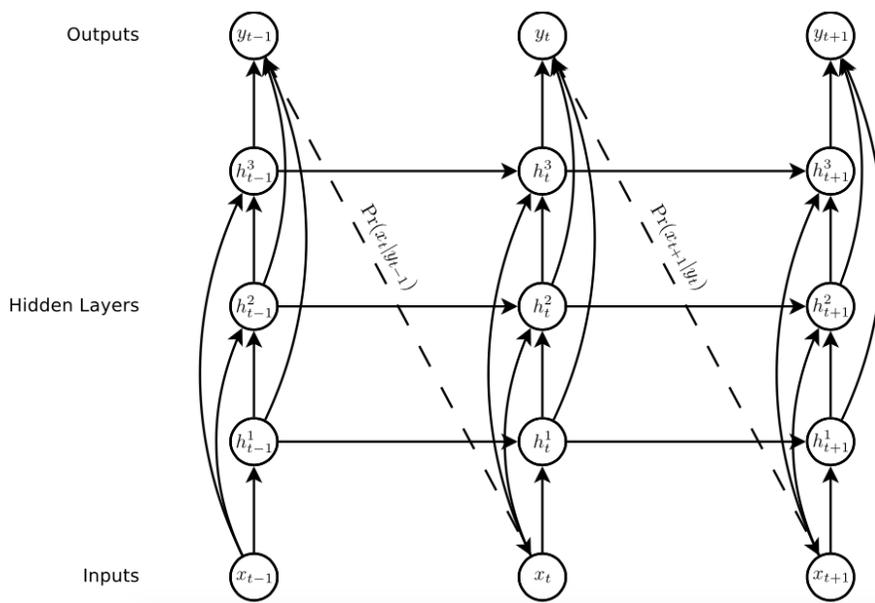


Figura 2.9: Diagrama de red neuronal recurrente [11].

2.6.1. Long-Short Term Memory

Una de las arquitecturas más comunes en las redes recurrente neuronales son las redes Long-Short Term Memory, en realidad este es un tipo especial de RNN que utiliza neuronas especiales que incluyen un celda de memoria para mantener información por largos periodos de tiempo. Esta celda de memoria les permite aprender dependencias hacia valores de largo término.

Las redes LSTM tratan el problema de desvanecimiento y explotación del gradiente al introducir nuevas compuertas. Este problema ocurre debido a la manera en la que los pesos de las redes neuronales son actualizados. Para ello se emplea el algoritmo de propagación hacia atrás conocido como algoritmo de *backpropagation* el cual trabaja desde la capa de salida hacia la capa de entrada propagado el error del gradiente y, una vez que se ha calculado el gradiente de la función de costo con respecto a cada parámetro en la red, este gradiente se utiliza para actualizar los parámetros (pesos) de la red.

El problema sucede que frecuentemente los gradientes se vuelven más y más pequeños conforme el algoritmo va a las capas más bajas. Esto genera que la capa más baja tenga conexiones muy débiles debido a que no hay cambios significativos en sus pesos, ocasionando que el entrenamiento no converja a un buen resultado. Este problema es conocido como el desvanecimiento del gradiente.

En ocasiones puede suceder un caso contrario al desvanecimiento, en el que el gradiente comienza a crecer y es más y más grande causando que varias capas tengan pesos inusualmente grandes pesos y que el algoritmo se vuelva inestable y difícil de aprender los datos de entrenamiento. Este problema es conocido como explotación del gradiente.

Como se había mencionado anteriormente, las Redes Recurrentes Neuronales suelen ocuparse para tareas de predicción en secuencias. Una Red Recurrente tiene una forma muy similar a la de una red neuronal común, sin embargo la principal diferencia es que la Red

Recurrente posee conexiones que hacia atrás es decir, recibe la misma salida de un tiempo anterior como parte de la entrada en el tiempo actual. En la literatura normalmente se emplea una representación desenrollada a lo largo del tiempo en donde la salida en el tiempo $t - 1$ se convierte en parte de la entrada del tiempo t

El proceso de entrenar con secuencias para una Red Neuronal Recurrente consiste, básicamente, en tomar una secuencia de valores de la serie de tiempo y usarlas cómo entrada a la RNN para predecir un valor futuro de esa serie. Por lo tanto, la red contendrá una cantidad dada de neuronas por capa y a la vez se habrá de desenrollar una cantidad de veces equivalente a la longitud de la secuencia. Así, entrenar secuencias largas en una RNN necesita correr este proceso por varias ocasiones, creando que la RNN ya desenrollada sea bastante profunda. Consecuentemente, los problemas de desvanecimiento y explotación del gradiente se harán presentes.

Algunas de las técnicas más conocidas para detener este problema abarca desde una inicialización controlada de los parámetros, utilización de funciones de activación que no se saturan, normalización de los lotes de entrenamiento e incluso de tener el proceso de desenrollar la red después de un determinado paso de tiempo. Adicional al largo tiempo de entrenamiento de una red RNN, el alto consumo de memoria se convierte en un problema adicional presente en este tipo de redes. Inminentemente, estos problemas degradan el desempeño del sistema inteligente, causando que incluso valores o aspectos claves se pierdan a lo largo del tiempo y que traiga consigo una mala precisión cuando el sistema está operando.

Para resolver esta serie de inconvenientes se han introducido diferentes tipos de células de memoria de largo plazo y una de ellas es precisamente las utilizadas en LSTM. Las redes neuronales LSTM fueron propuestas en 1997 por Sepp Hochreiter y Jürgen Schmidhuber [12]. En la Figura 2.10 se muestra un diagrama de una celda de memoria simple LSTM

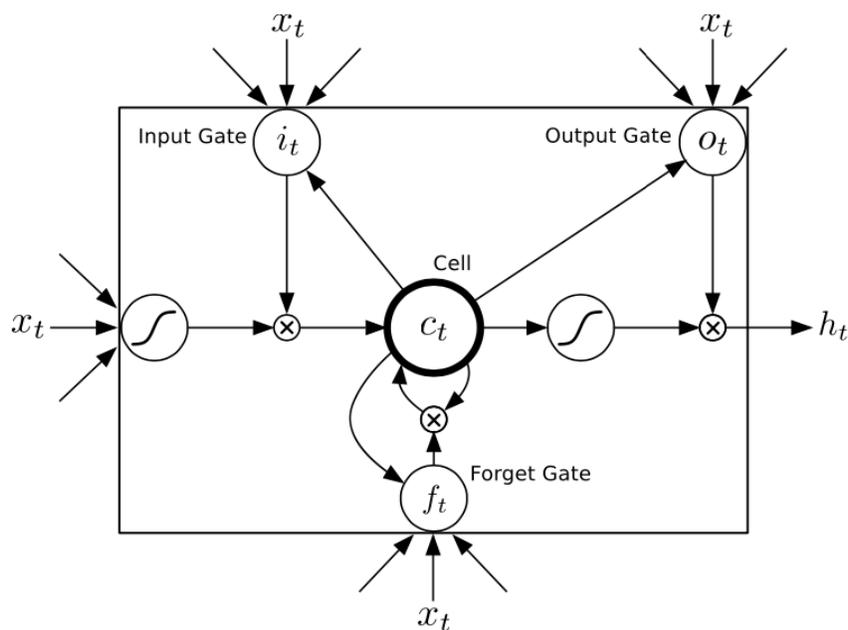


Figura 2.10: Diagrama de una celda LSTM [11].

Una célula LSTM divide su estado en dos vectores, $h(t)$ para el estado de corto plazo

y $c(t)$ para el estado de largo plazo. Con esto se tiene la idea principal de almacenar tanto el estado a un corto y largo plazo, permitiendo discernir qué parte de ellos se conservará y qué parte de ellos se eliminan. Por lo tanto, el estado $c(t-1)$ viaja en la neurona pasando primero por un filtro que elimina parte de la memoria adquirida en pasos anteriores, este filtro es conocido como una compuerta de olvido (Forget gate). Posteriormente se agrega nueva información con una compuerta de entrada (Input gate) la cual contiene información tanto de la memoria de corto plazo del instante anterior como de la entrada x actual. Hasta este punto se tiene el valor del estado $c(t)$ sin considerar más modificaciones y únicamente se eliminaron algunos valores y se agregaron nuevos al estado de largo plazo.

Posteriormente, este valor es copiado y se aplica una función Tangente para ser operado en la compuerta de salida, con este nuevo valor se obtiene el estado de corto plazo $h(t)$ el cual es el mismo valor que tendrá la salida $y(t)$. Nótese que la entrada a las capas de neuronas son alimentadas por los vectores $x(t)$ y $h(t-1)$. Una de estas capas analiza la entrada actual de $x(t)$ y el estado previo de corto plazo $h(t-1)$, la salida de esta capa es $g(t)$ y es parcialmente almacenada en el estado de largo plazo.

Por otra parte, el resto de las capas son compuertas de control ya que poseen una función de activación logística que entrega valores entre 0 y 1, con ello dejará pasar o limitará la memoria previa a los nuevos estados de la célula. La compuerta de olvido, denotada por $f(t)$ es encargada de controlar que tanto ha de olvidarse de la memoria a largo plazo. La compuerta de entrada, denotada por $i(t)$ controla que parte de $g(t)$ ha de ser añadida al estado de largo plazo. Finalmente, la compuerta de salida $o(t)$ controla que parte del estado de largo plazo será la salida de la red LSTM en el momento actual, es decir, $h(t)$ y $y(y)$.

En resumen una célula LSTM es capaz de aprender que elementos de la entrada actual son importantes, almacenarlos en el estado de largo plazo, así como que tanto tiempo preservar dichos elementos y extraerlos cuando sea necesario. Gracias a estas características, las redes tipo LSTM pueden aprender de largas secuencias ya sea como series de tiempo, texto, entre otras.

A pesar de ello, en los recientes años se han implementado nuevos modelos que han demostrado un mejor desempeño para largas secuencias, tomando en cuenta dependencias de los elementos dentro de la secuencia. Este tipo de modelos son los Transformadores, los cuales se abordarán en la siguiente sección.

2.6.2. Transformadores

Los Transformadores son una arquitectura diseñada para la traducción de secuencias que dan por salida, de igual forma, una secuencia. Esta arquitectura es conocida como secuencia a secuencia, Seq2Seq. Los Transformes fueron presentados en el año 2017 el artículo Attention Is All You Need [13]. Esencialmente, es una red que transforma una secuencia de entrada en una secuencia de salida. En el artículo original, las secuencias son palabras dentro de una oración que habrían de ser traducida de un idioma a otro.

Este tipo de arquitecturas Seq2Seq han mostrado una ventaja particular para tareas de traducción. Tal arquitectura Seq2Seq consiste de dos elementos principales, un codificador (encoder) y un decodificador (decoder). A groso modo, el codificador toma la entrada y aplica determinadas operaciones que la mapean a una dimensión mayor, la salida del codificador se convierte en la entrada al decodificador para que este último dé una nueva secuencia de

salida.

Como el artículo original lo menciona en su título, el punto clave de esta arquitectura está en los mecanismos de atención. Este mecanismo tiene una función muy importante que consiste en tomar un elemento de la secuencia de entrada y conocer con qué otras partes de la secuencia tiene una alta relación o, mejor dicho, cual es su relación con cualquier otro elemento dentro de la secuencia, este proceso aporta información contextual de los elementos para poder realizar las operaciones necesarias de forma directa. Además, otra de las grandes ventajas de los Transformadores es la posibilidad de operar de forma paralela la secuencia de entrada y no depender de un procesamiento secuencial que va recibiendo elemento por elemento de la entrada, como en el caso de las redes LSTMs. Esto se debe a que un Transformador recibe una entrada embebida que contienen una codificación del elemento original, por ejemplo, para el caso de palabras vectores de codificación, que a su vez están enriquecidos con la posición dentro de la secuencia, permitiendo paralelizar el procesamiento de la secuencia. A continuación se presenta y detallan los elementos de los Transformadores, en la Figura 2.11 se presenta el diagrama utilizado en el artículo original de un Transformador.

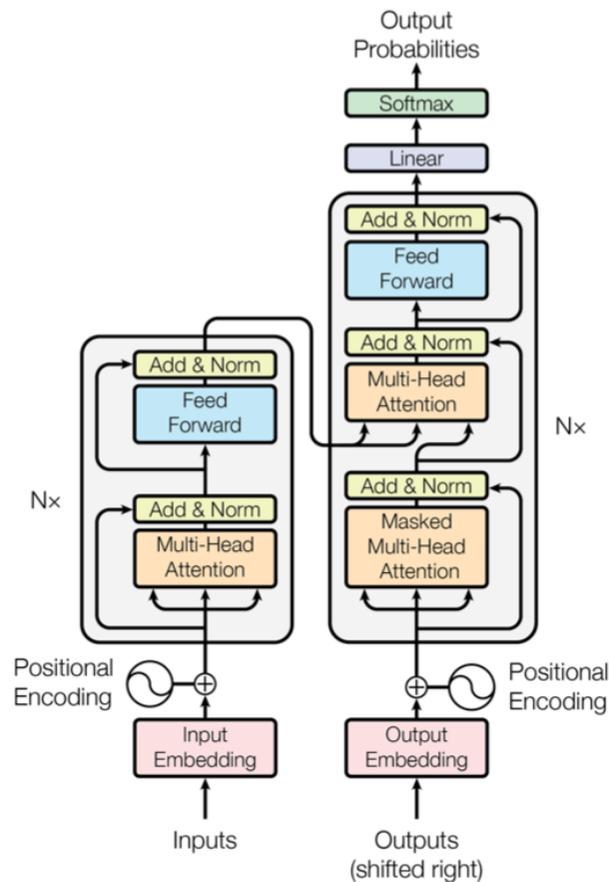


Figura 2.11: Diagrama de una Transformador [13].

El codificador está compuesto por una pila de N capas idénticas, el artículo original menciona 6 capas. Cada una de las capas está compuesta a su vez por dos subcapas que contienen el mecanismo de auto-atención de múltiples cabezas, en la primera subcapa y en la segunda

sub-capa hay una red simple completamente conectada de tipo Feed-Forward respecto a la posición. Se emplean conexiones residuales alrededor de ambas subcapas seguidas de una capa de normalización.

Por su parte, el decodificador esta compuesto por el mismo número N de capas idénticas. Sin embargo, este elemento posee tres subcapas en lugar de sólo dos como en el caso del codificador. Esta tercer subcapa se añade entre la subcapa de más abajo y la de más arriba y es encargada de realizar la atención de múltiples cabezas sobre la salida de la pila del codificador. En el caso de la subcapa de auto-atención del decodificador, ésta debe garantizar que la predicción para la posición del elemento i dependerá sólo de las salidas conocidas a posiciones previas a i

Como se mencionó anteriormente, la característica principal de los Transformadores recae en la atención. La función de atención se puede describir como un mapeo de un conjunto de vectores abstractos generados para ayudar en el proceso de atención denominados *query*, *keys* y *values*. Estos vectores son obtenidos por la multiplicación de una entrada embebida por las matrices de pesos correspondientes a Q , K y V . A groso modo la función de atención recae en hacer una consulta (*query*) de similitud para las diferentes claves (*keys*) dentro de una base de datos (palabras o elementos posibles en la secuencia) para generar una combinación ponderada de los valores (*value*) y al final obtener el valor correspondiente con el que posee mayor similitud para cada elemento de la entrada. Este proceso es únicamente para una cabeza de atención, cabe recordar que el proceso de atención que se realiza es de cabeza múltiple, por lo que este proceso se realiza en múltiples cabezas similares, cada una con sus respectivos pesos independientes y la salida de cada cabeza de atención se concatena con las demás para obtener una única salida que considere la aportación de cada cabeza de atención. Para ello la salida de cada cabeza es ponderada por la matriz de pesos W^O al momento de concatenarlas. En la Figura 2.12 se presentan los diagramas para la atención que el autor emplea en el artículo original.

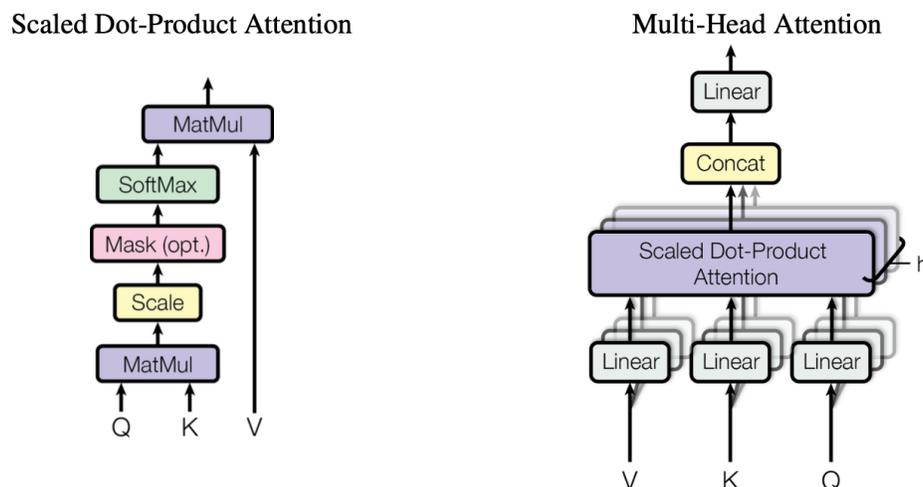


Figura 2.12: Diagrama de una el modulo de atención (izq.) y la cabeza de atención múltiple (der.) [13].

De esta forma el mecanismo de atención de cabeza múltiple puede entenderse mediante

la siguiente ecuación:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O \quad (2.1)$$

donde

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2.2)$$

En el caso de la pila de codificadores, únicamente el codificador más profundo o el primero es el que recibe la secuencia original a transformar, y los codificadores apilados en lo sucesivo, reciben la salida del codificador previo. Para el caso del decodificador, los mecanismos de auto-atención operan de forma similar, sin embargo, en este caso cada uno de los decodificadores de la pila reciben la salida del último codificador. Adicionalmente, se puede observar que tanto la pila de codificadores como decodificadores poseen conexiones residuales dentro de las subcapas que lo componen, esto es con la intención de conservar información original de la secuencia de entrada para prevenir pérdidas en caso de un mal funcionamiento dentro de las subcapas. Por otra parte, la salida del Transformador es nuevamente alimentada a él pero únicamente a través de la pila de decodificación.

Finalmente, el conocimiento de la posición de cada elemento dentro de las secuencias es añadido a la entrada embebida como un vector de posición, lo que posibilita realizar las operaciones de forma paralela y a la vez conocer que el contexto de cada elemento en la secuencia, por ejemplo, en el caso de traducción, una misma palabra puede tener un significado diferente dependiendo de la posición en la que se encuentre, ya que puede fungir como sustantivo o como verbo.

En esta tesis la secuencia de entrada tanto a los Transformadores como a las redes LSTM, es fragmentos de una serie de tiempo con el número de autos que hay dentro de un lugar, más aún, detectar el número de autos en una determinada zona de forma autónoma no es una tarea sencilla. Por ello se plantea que la secuencia de información propia del sistema de comunicación sea el número de identificadores únicos que detecta la estación base debido a la recepción exitosa de paquetes de información, para transformarla en una secuencia con el número real de usuarios o vehículos que están dentro de una zona dada intentando acceder a recursos de canal para transmitir su información.

2.7. Estado del Arte

En las secciones anteriores se ha descrito el marco teórico necesario para realizar la tesis correspondiente a este documento así como el planteamiento del problema y los objetivos esperados. En esta sección se aborda el estado del arte en torno a la propuesta hecha.

Actualmente, el desarrollo de sistemas de comunicación inalámbrica está afrontando múltiples retos, desde años anteriores se ha pronosticado el alto crecimiento de su uso que actualmente se observa. En el año 2013, los autores de [14] pronosticaban un crecimiento bastante significativo en el flujo de información que viajaría en medios de comunicación inalámbrica, en ese entonces esperaban que para el año 2020 el tráfico inalámbrico incrementará en mil veces con respecto al que existía en 2010. Una de las razones que consideran para este incremento es el incremento de comunicación entre dispositivos debido a los sistemas de internet de las cosas (IoT). Pese a que no se han masificado estos sistemas, al menos en México, sí

se ha generado un ambiente suficientemente denso de comunicaciones inalámbricas. Algunos dispositivos del tipo IoT que se han vuelto común observar en los hogares son referentes a domótica como focos inteligentes, cerraduras inteligentes, cámaras de seguridad, televisores o bocinas, principalmente.

Los mismo autores enfatizan (en ese año) que las tecnologías HSPA y LTE serán las que dominarán las comunicaciones inalámbricas, e incluso apuntan a su uso para diferentes sistemas. Abordan el aspecto de la comunicación entre vehículos como una tecnología que toma auge en la década de 2020, señalando que las comunicaciones de corto alcance, vehículo a vehículo (V2V), y las de vehículos a infraestructura (V2I) son las comunes. Con esta creciente, los autores enlistan una serie de retos necesarios de abordar, haciendo hincapié en la necesidad de establecer nuevos enfoques para el acceso canales de comunicación, así como un mejor uso del espectro radioeléctrico.

Sin embargo, en los últimos años la emergente red 5G e incluso 6G, promete dominar las telecomunicaciones, al ser considerada una red heterogénea que permite la integración de distintas redes a altas velocidades de transferencia así como la integración de soluciones por software, la utilización de 5G como infraestructura en redes V2I y V2X han conducido a su investigación y se ha incluido dentro de los estándares de la estandarización 3GPP [15]. Con nuevas propuestas en los dispositivos, es posible tomar mayor ventaja de las características de 5G, por ejemplo en [16] los autores proponen el uso de antenas fractales para la comunicación de una VANET a 5G para dar acceso a Internet a los vehículos, lo cual abre el mundo de aplicaciones de un VANET y permite mayores tasas de transferencia requeridas en aplicaciones multimedia. Así mismo, los autores de [17] analizan la factibilidad de la implementación de comunicaciones vehiculares sobre redes 5G, concluyendo que su implementación cubre gran parte de los requerimientos de latencia, cobertura, retardo y velocidad de transferencia para la aplicaciones VANET, a la vez, mencionan su prominente desempeño en aplicaciones futuras.

En [18] los autores plantean el creciente interés por el uso de técnicas de inteligencia artificial para el diseño de modelos de comunicaciones, en particular ellos abordan aspectos importantes en el diseño de modelos de aprendizaje que permita brindar mayor seguridad a la información. Uno de las consideraciones que hacen es el tratar la información lo más cerca a la fuente que genera los datos, permitiendo así, una descentralización de los modelos que comprometa la integridad de la información debido al largo viaje que hace desde su fuente generadora hasta alcanzar los centros de procesamiento. Con esta propuesta refuerza la propuesta de hacer modelos de aprendizaje distribuido, o en su defecto, modelos que trabajen en instancias cercanas a los dispositivos. Este concepto es conocido como Edge Computing. Una aplicación más que propone utilizar modelos de aprendizaje en las cercanías de los usuarios es la conducción autónoma. En estos contextos, cada uno de los elementos de la red es capaz de entrenar un modelo para una aplicación determinada, por ejemplo, una entidad dada puede aprender cual es la información más demandada para vehículos autónomos y almacenar ésta de forma local para su uso con un menor retardo. En síntesis, de este documento se puede constatar que los modelos de aprendizaje de máquina se han convertido en elementos importantes para hacer más eficientes las comunicaciones.

A pesar de que en esta tesis no se consideran modelos distribuidos en los usuarios finales, vehículos, se asumen modelos de aprendizaje de máquina que son colocados dentro de las estaciones base y que aprenden de la información que reciben de forma local de los dispositivos

conectados a cada una de estas estaciones. De tal forma que cada uno de las estaciones base ajusta parámetros de forma inmediata, sin exponer la información de los usuarios por el hecho de compartirla a una entidad global cuando no es necesario. Así el estudio realizado en esta tesis aporta al beneficio de utilizar modelos de aprendizaje de máquina para la planificación de los recursos de forma local en puntos de acceso a la red.

En [19] los autores introducen el uso de técnicas de aprendizaje reforzado, para mejorar el uso de canal para protocolos S-ALOHA, en sus resultados ellos reportan usos de 70%, lo cual supera por mucho al obtenido por S-ALOHA en su versión original que es inferior a 37%. Sin embargo, esta propuesta tiene ciertas implicaciones, que si bien son factibles en redes Ah Doc tradicionales o de baja dinámica, para el caso de redes VANET, se volverán inconvenientes para su aplicación. De forma resumida, consideran un sistema en el que cada uno de los nodos, de forma independiente, toma la decisión de transmitir información o no en cada ranura de tiempo. Las ranuras son catalogadas como exitosas, colisionadas o vacías dependiendo de cuántos nodos han intentado transmitir. En tal trabajo se señala que los mejores desempeños se obtienen con un valor de nodos menor a 50, así mismo, necesitan de alrededor de 100 ranuras para comenzar a aprender una política óptima. Otra limitante más es que no es práctico de implementar en sistemas con un canal simple. Este último detalle también es presente en su previo trabajo [20], en el cual los resultados alcanzan valores altos de utilización del canal. Nótese que en ambos casos esta utilización se logra a la larga distancia es decir una vez transcurrido un periodo de 20 o más bloques de 100 ranuras de tiempo. Por la alta dinámica de las VANETs este hecho lo vuelve no práctico, ya que ese valor representa 2000 ranuras de tiempo, y si, por ejemplo, una ranura es de entre 2 y 5 milisegundos, eso significa 10 segundos de tiempo para alcanzar dicho porcentaje de uso. En el caso de la VANET, en 10 segundos es susceptible a presentar un cambio totalmente en la arquitectura de la red.

Años atrás los autores de [21] introdujeron el algoritmo Q-Learning para mejorar el desempeño de una red de sensores, con ello obtienen uso del canal cercano a 50%, nuevamente el problema que se presenta es que toma un alto tiempo para poder converger y éste se incrementa rápidamente conforme el número de nodos es mayor.

De forma específica para redes VANETs, se han utilizado diferentes técnicas de aprendizaje de máquina para diversas tareas o en diversos enfoques. En [22] los autores hacen un estudio de las diferentes áreas donde las técnicas de aprendizaje de máquina contribuyen a mejorar el funcionamiento de las VANETs, ellos abarcan numerosas aplicaciones, como control de congestión, en la predicción del flujo vehicular, establecimiento de rutas, etc. en sus sugerencias propone utilizar modelos de aprendizaje profundo como aquellos basados en LSTM.

Normalmente el uso de aprendizaje de máquina se orienta a la predicción del flujo vehicular y su relación con el flujo de datos generados y la asignación de rutas, tanto para vehículos como para los paquetes de información en multi-saltos, como es el caso de [30] donde los autores hacen un estudio referente a distintas técnicas utilizadas para la predicción de flujo vehicular y consideran que aquellas basadas en aprendizaje de máquina son las más flexibles. En [24] proponen un esquema basado en la predicción de tasas de información presentes en el canal a un futuro dado para hacer eficiente la asignación de recursos en comunicaciones Vehículo-Cloud, mediante los tiempos de transmisión de la información, considerando que los vehículos son fuentes masivas de información por sus múltiples sensores.

Los autores de [25] utilizan un arreglo de una Red Convolutiva, una Red Recurrente basada en LSTM e información contextual de las historias del tráfico para encontrar patrones dentro y fuera del flujo vehicular de una zona en un determinado día, y así, hacer una mejora a la predicción del flujo vehicular, logrando una precisión encima del 90 %. Este estudio está enfocado únicamente a la predicción del flujo vehicular.

En [26] los autores presentan un modelo basado en auto-atención de escala múltiple para predecir el flujo vehicular en tiempos futuros basado en las observaciones previas. Aquí proponen muestreos de la cantidad de vehículos en una ciudad con resoluciones de 30 minutos en adelante. Su aportación recae en aprender la dependencia espacio-temporal respecto a otras regiones con la finalidad de obtener mejores predicciones. En este caso, uno de los inconvenientes es la resolución de tiempo que se maneja, ya que para aplicaciones a nivel de protocolo de comunicación, es una resolución muy alta que no aporta gran mejora al desempeño de una red inalámbrica.

En [27] los autores proponen un modelo de predicción del flujo vehicular utilizando una red neuronal profunda de tipo Convolutiva basada en la arquitectura de redes residuales. Aquí, el modelo es alimentado con una matriz del flujo del tráfico y con información externa para brindar un modelo más robusto, en este caso los autores utilizan una resolución de tiempo más favorable para los protocolos de comunicación, ya que para sus estudios hacen muestreo en espacios de tiempo de 60 segundos. Sin embargo, el estudio se limita únicamente a la predicción del flujo y no consideran su aportación a los protocolos de comunicación.

Por otro lado [28] provee un estudio de los modelos de aprendizaje profundo utilizados en las distintas tareas de predicción de tráfico, describe las aplicaciones en las que se utiliza la predicción del tráfico como lo es el propio flujo vehicular, velocidad esperada de los automóviles, el número de peticiones de información vehicular, los tiempos de viaje, así como la ocupación de los caminos. A pesar de que el estudio revisa diferentes publicaciones, no reportan alguna que presente de forma explícita las afectaciones en el canal de comunicación inalámbrico de una VANET con la predicción del flujo vehicular. Una situación similar se presenta en [29], [30], donde los estudios presentados se centran sólo en la predicción del tráfico, siendo los modelos de atención y los recurrentes, o la mezcla de ambos, los que mejor desempeño ofrecen.

Finalmente, dentro de los protocolos de comunicación en VANETs, existen diferentes tipos de ellos, los cuales son clasificados en libres de contención y los de contención. En los protocolos libres de contención se utilizan recursos destinados para cada uno de los nodos, con los cuales cada vehículo puede transmitir sus paquetes de información sin que algún nodo dentro del área de cobertura interfiera con dicha transmisión. Contrariamente, los protocolos de contención transmiten cada que requieren enviar un paquete de información, esta transmisión es realizada esperando que ningún otro nodo interfiera con la transmisión. En los protocolos libres de contención el estado del arte obtiene mejores resultados para aquellos basados en TDMA con asignación de recursos de forma programada, logrando un uso eficiente del canal cercanos al 100 %. En el caso de los protocolos de contención el terreno más explorado es CSMA, sin embargo, en aplicaciones de *broadcast* y *multicast* o aquellas que están basadas en un red meramente como un sistema Ad Hoc, CSMA llega a presentar menor fiabilidad que aquellos basados en ALOHA [31].

Bajo el estándar 802.11p y WAVE la comunicación primordialmente se basa en prevención de colisiones mediante el protocolo CSMA/CA tanto para la comunicación de vehículo

a vehículo como la comunicación vehículo a infraestructura; sin embargo, en distintos protocolos se conserva el enfoque del protocolo ALOHA [32]. Por ejemplo, el protocolo ADHOC MAC presentado en [34] fue pensado para solventar los problemas de terminal expuesta y terminal oculta dentro de la comunicación entre vehículos, los autores basan su propuesta en el protocolo RR-ALOHA [33] que habilita a los nodos para reservar un slot dentro de cada trama cuando se logra una transmisión exitosa. Dentro de los protocolos más recientes, se ha estudiado aquellos basados en TDMA y CSMA/CA con técnicas como el cambio de la longitud para los intervalos de los canales de control, como lo proponen los autores de [35]. Mientras que en [36] los autores utilizan TDMA para compartir mensajes en saltos múltiples eligiendo los nodos de retransmisión mediante la difusión de la longitud de sus buffers. En general se atienden las problemáticas en la comunicación vehículo a vehículo, sin embargo, en la comunicación hacia la infraestructura la comunicación tiende a limitarse a lo establecido en 802.11p.

En el siguiente capítulo se presenta el diseño de la propuesta y su análisis, se explican las consideraciones tomadas para el caso de estudio que se aborda así como la forma en que se aplican las técnicas necesarias en la propuesta de solución.

Capítulo 3

Diseño y análisis de la propuesta

En esta sección se describe el conjunto de consideraciones tomadas para dar las soluciones propuestas al protocolo de comunicación para VANETs mejorados con la ayuda de técnicas de inteligencia artificial. Cabe recordar que se propone un protocolo de comunicación basado en S-ALOHA, con el cual los nodos, en este caso vehículos, de una red vehicular acceden de forma aleatoria para compartir información relevante a aplicaciones de *Smart Cities*.

Los paquetes de información son cortos y contienen únicamente información de telemetría de cada vehículo, como lo es su velocidad, su posición, la cantidad de contaminación, entre otros. Así mismo, el reporte de algún incidente de tráfico también puede ser enviado por medio de este tipo de mensajes. Por lo tanto, el enfoque que tiene esta red es similar a la de una red de sensores inalámbricos, la cual no es estrictamente considerada para manejar contenido multimedia. Adicionalmente, un protocolo de contención precede a un esquema con recursos fijos, por lo que en canales de control o para solicitud de recursos se emplea este tipo de protocolos de acceso aleatorio.

Sin embargo, este tipo de redes posee una característica muy especial que las limita al utilizar protocolos convencionales con alto índice de uso del canal, siendo su principal limitante la alta dinámica presente en estos sistemas, en los cuales la entrada y salida de los vehículos en una zona específica, así como la velocidad de los vehículos, hace que utilizar protocolos de contención brinde comunicación sólo a los nodos activos y que estos puedan acceder a los recursos del canal cuando así lo requieran. Cabe señalar que los protocolos programados obtienen usos de los recursos muy cercanos al 100 %, sin embargo, para ello se necesita conocer la topología de la red, así como tener un mapeo de los usuarios que utilizarán el canal a fin de establecer periodos de comunicación para cada uno de los usuarios. En este caso por la alta dinámica del sistema, los nodos son susceptibles a abandonar la red de forma no uniforme para cada vehículo, causando que los recursos previamente asignados a un usuario u_1 sean negados a un usuario u_x cuando $x \neq 1$, donde x es sólo un identificador temporal asignado a cada usuario durante su estancia en alguna zona de cobertura. A continuación se describe el escenario de análisis

3.1. Propuesta de estudio

El sistema que se propone integra el uso de técnicas de aprendizaje de máquina con una red vehicular tipo V2I, donde los nodos son los vehículos que se encuentran dentro de la zona de cobertura de una estación base de telefonía celular, la cual funge como infraestructura de comunicación a la VANET. Cada estación cuenta con un área de cobertura para brindar servicios de comunicación a los vehículos por medio de un canal de acceso aleatorio ranurado en el tiempo y que brinda acceso a los vehículos en un esquema de contención basado en S-ALOHA

En cada una de las estaciones base se asume que existe un módulo con capacidades de cómputo para desempeñar las funciones de predicción para estimar el número de usuarios en tiempos futuros. Debido a que se desea la mayor autonomía posible del sistema, la estación base será la encargada de contabilizar el número de vehículos dentro de su área de cobertura, utilizar un modelo de predicción basado en aprendizaje de máquina para estimar la cantidad de vehículos en un futuro dado, para posteriormente establecer la probabilidad de transmisión con la que los vehículos intentarán acceder al canal. Por otra parte, la actualización de la probabilidad de transmisión también será enviada por la estación base en un mensaje de difusión en intervalos de tiempo periódicos a fin de adaptarse lo mejor posible a los cambios de la cantidad de usuarios en su zona de cobertura.

El sistema de predicción es local para cada estación base a fin de evitar los efectos de retardo que conlleva tener un sistema centralizado en la nube. En un caso de estudio fuera del presentado en esta tesis, el sistema se puede visualizar como distribuido, en donde cada estación base procesa la información correspondiente a su zona de cobertura y la comparte a las demás estaciones base con la intención de hacer más robusta la predicción de la cantidad de usuarios. Este último caso, no se aborda en esta tesis. Adicionalmente se utiliza la simulación de SUMO para desplegar los vehículos en la ciudad y conocer la cantidad real de ellos dentro de cada una de las zonas de cobertura de las estaciones base.

En la Figura 3.1 se presenta un esquema de como está compuesto el sistema propuesto, donde se visualiza los planteamientos descritos previamente, además se observa que la entrada al sistema de predicción es una serie de tiempo con la cantidad de vehículos detectados dentro de la zona de cobertura de la estación base en los momentos correspondientes a cada periodo de actualización de la probabilidad de transmisión. La estación base recibe paquetes de información proveniente de los vehículos, los cuales cuentan con identificadores únicos de cada vehículo, cómo puede ser la dirección MAC, de esta forma la misma estación base es capaz de hacer un conteo sobre los identificadores únicos y saber que cantidad de usuarios lograron conectarse. Obligadamente la cantidad de vehículos que transmiten a la estación base siempre es menor o igual que la cantidad de vehículos en dentro de zona de cobertura.

En primera instancia, el entrenamiento de los modelos de aprendizaje se propuso en un modo fuera de línea, es decir, por medio de datos obtenidos por el simulador SUMO se entrenan los modelos, de tal forma que la estación base opera únicamente con un modelo previamente entrenado para estimar la cantidad de vehículos que habrá en un instante de tiempo futuro. Sin embargo, uno de los problemas observados es que la alta dinámica de los vehículos ocasiona que no todos los vehículos dentro de la zona de cobertura alcancen a transmitir dentro del periodo de actualización de la probabilidad de transmisión. Con ello la estación base no podrá contar de forma precisa la cantidad de vehículos que hay

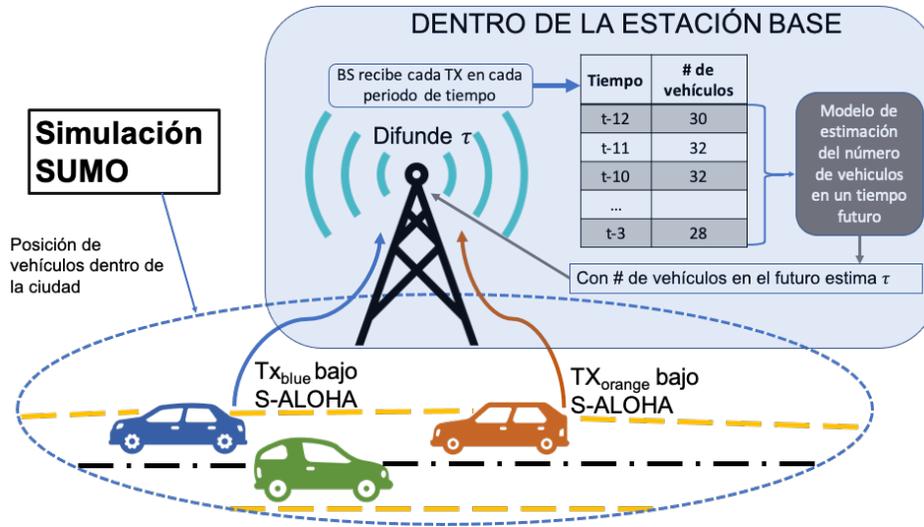


Figura 3.1: Diagrama del sistema propuesto

dentro de la zona, y consecuentemente el estimado para tiempos futuros no será preciso. Realmente el conteo que hace la estación base se puede observar como una señal con ruido y los modelos de estimación requieren ser capaces de sobrellevar dicha situación. En este sentido, es necesario simular la operación de la red vehicular con el protocolo S-ALOHA operando para que los modelos de aprendizaje de máquina reciben como entrada la cantidad de vehículos que transmiten en un lapso de tiempo determinado así como los valores de la simulación en SUMO para que sean número real que debe estimar los modelos de predicción (valor objetivo o *target*), esto para la fase de entrenamiento de los modelos de aprendizaje de máquina permitirá dar mayor generalidad a la predicción de valores futuros. En la Figura 3.2 se ejemplifica este proceso



Figura 3.2: Diagrama del proceso de entrenamiento

En las siguientes secciones se detalla en mayor medida cada una de las consideraciones en este estudio, incluido la forma de obtener los datos, los pormenores de la red y el proceso de entrenamiento de los modelos de aprendizaje de máquina.

3.2. Red vehicular V2I

La red vehicular para la que se desarrolla el protocolo de comunicación con ayuda de algoritmos de inteligencia artificial sigue una arquitectura V2I, en esta arquitectura los vehículos se comunican en un sólo salto hacia una estación base (BS) mientras se encuentran en sus

zona de cobertura. Las estaciones base pueden ser diferentes dispositivos, por ejemplo, una unidad situada a lado del camino (*Road Side Unit*, conocida como RSU), la estación base de una celda celular de tecnología 3g, 4g o 5g (BTS, NodeB, eNodeB), incluso algún punto de acceso WIFI.

En tales casos, los radios de cobertura de las estaciones base son limitados en distancia, típicamente se habla de radios del orden de cientos de metros, para fines prácticos, en este proyecto se propone un radio de cobertura de 200 metros, este radio es fácil de alcanzar en sistemas celulares [37], [38] incluso las microceldas pueden duplicar fácilmente ese radio de cobertura. Sin embargo, en ciudades con una alta densidad de usuarios, el uso de pico celdas es más frecuente para ofertar una mayor cantidad de canales a los usuarios. Por otro lado, las RSUs normalmente operan con radios de cobertura al rededor de 300 metros [38]. Además, las infraestructuras WiFi suelen tener menor cobertura debido a su alta congestión de usuarios, recordando que WiFi opera en bandas del espectro electromagnético libre, a mayor distancia la señal original se percibe más débil por la interferencia con una mayor cantidad de señales de sistemas externos operando en las mismas bandas de frecuencia.

Mientras un vehículo se encuentre dentro del área de cobertura de la estación base, el vehículo podrá compartir su información de telemetría en el canal asignado para dicha tarea. El canal es compartido para todos los nodos dentro del área de cobertura, por lo que la transmisión de los paquetes se realiza bajo un esquema de contención basado en S-ALOHA, considerando que cada nodo con información por enviar competirá con por acceder con una probabilidad de transmisión $\tau = 1/N$ donde N es el número de vehículos por transmitir. En la Figura 3.3 se ejemplifica la arquitectura de una red V2I, donde la Infraestructura es una estación base de telefonía celular.

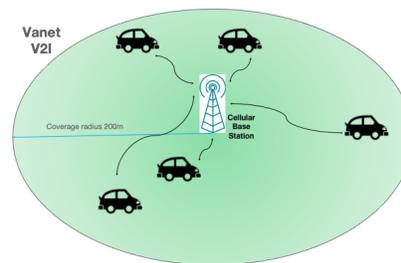


Figura 3.3: Diagrama de cobertura de las estaciones base

3.3. Protocolo S-ALOHA

En los esquemas de contención, la transmisión de los paquetes de información comienza cuando el nodo tiene información para enviar, en este contexto se pueden asumir dos enfoques, uno en donde los nodos están compartiendo la información más reciente a cada instante, o que reporten eventos que se suscitan con una cierta probabilidad P_{ev} . En este estudio se toma el primer enfoque mencionado, en el que los vehículos trataran de enviar su paquete de información cada determinado periodo de tiempo, cuyo paquete de información contendrá sus datos telemétricos más actuales. Al utilizar como base el protocolo S-ALOHA, se entiende

que el canal de comunicación es ranurado en el tiempo, por lo tanto se asume que el periodo de tiempo para que el nodo actualice su paquete de información es igual al tiempo de duración de cada una de las ranuras. Se define a la ranura s_j como la j -ésima ranura dentro del canal. El algoritmo propuesto para cada uno de los vehículos dentro del área de cobertura de la estación base se describe a continuación.

```

while vehículoi está en la zona de cobertura do
  for cada ranura  $s_j$  do
    el vehículoi obtiene variable  $0 \leq U \leq 1$ 
    el vehículoi intenta transmitir siguiendo la condición
    if  $U \leq$  then
      El vehículoi envía su paquete de información
    else
      El vehículoi no envía su paquete de información
      el vehículoi actualiza su paquete de información
    end if
  end for
end while

```

De esta forma se considera que los vehículos tienen la capacidad de descartar aquellos paquetes que no se intentaron transmitir, por lo tanto, se espera que la tasa de paquetes descartados sea alta. Por otro lado, el automóvil podría utilizar técnicas adicionales, para reducir esta tasa de paquetes descartados, por ejemplo, utilizar funciones de correlación para saber que tanto cambiaron las variables contenidas en el paquete de un periodo a otro, de alcanzar un cambio significativo en los nuevos valores con respecto a los encontrados en el paquete previo, se descarta el paquete actual y se crea uno nuevo con los valores más actuales. Con ello los vehículos colaboran en utilizar de mejor forma el canal para compartir solo información más actualizada. Así mismo, también se puede utilizar una política en la que los nodos intentan transmitir en cada ranura de tiempo su paquete de información y una vez que logran transmitir el paquete actual, generan un nuevo paquete con la información más reciente que intentará transmitir en las siguientes ranuras de tiempo, en los tres casos, se sigue considerando que los N paquetes contienden por el canal en todo momento.

Respecto a la capacidad de los vehículos de transmitir a la estación base, se asume que los vehículos transmiten con la potencia suficiente para que la señal alcance la estación base desde cualquier punto dentro del radio de cobertura previamente establecido. Consiente de que la interferencia con estaciones base vecinas puede ser posible, se asume que las estaciones base operan bajo el paradigma de reuso de frecuencia, de forma similar a las de telefonía celular, en las que celdas vecinas utilizan un conjunto de frecuencia diferentes para prevenir dicha interferencia.

Por otra parte, se establece una política en la que la estación base recibe las transmisiones y después de un periodo de tiempo T , mediante un mensaje de difusión, envía el valor de la probabilidad de transmisión con la que cada vehículo intentará acceder al canal desde ese momento hasta el siguiente periodo de tiempo de actualización. El periodo de tiempo de actualización se propone que sea una variable fija. En los protocolos de contención, la probabilidad de transmisión de los vehículos es un parámetro crítico, por lo tanto es primordial tener un tiempo de actualización que sea corto para que los vehículos se adapten a los cambios

que se presentan en la red, pero también es necesario que no sea tan repetitivo el envío del mensaje de difusión en el que se envía la nueva probabilidad de transmisión ya que, con ello, también se causaría un uso ineficiente del canal. Se propone que el periodo de actualización sea de un segundo. Por lo que en cada segundo es preciso hacer la nueva estimación con la cantidad de vehículos que se ha detectado teniendo como consecuencia que en el canal de transmisión un segundo abarca 375 *slots* de igual tamaño en el que los nodos transmiten hacia la estación base y se propone que al inicio de cada segundo exista un tiempo de menor duración en el que la estación base envía el valor actual de τ con el que intentará transmitir cada vehículo para ese segundo en particular.

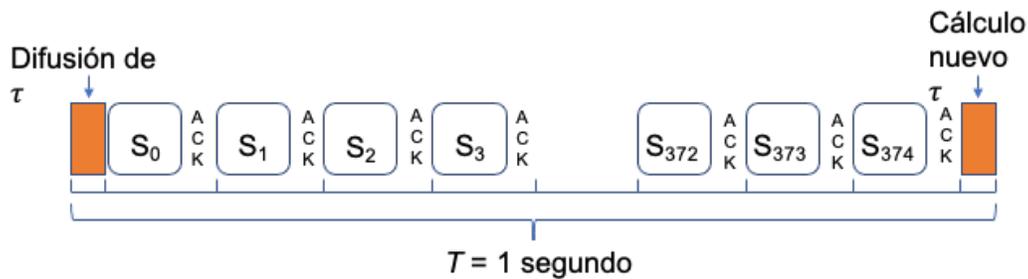


Figura 3.4: Representación de la organización del canal

En la Figura 3.4 se presenta la organización del canal que se ha propuesto para este estudio, aquí se puede observar que la estación base confirma de recibido el paquete de información mediante un mensaje tipo *acknowledge* (ACK), con el cuál el vehículo sabe que ha logrado transmitir de forma exitosa su información. Por otro lado, se observa que al término del segundo actual existe un intervalo de tiempo considerado para que se estime el próximo valor de τ , mismo que será enviado al inicio de cada segundo. Cabe notar que uno de los aspectos a considerar en los canales ranurados es la necesidad de una sincronía entre la estación base y los nodos que se comunican, sin entrar en mayor detalle, ya que éste puede ser una estudio completo, se propone que la estación base envíe a los vehículos el inicio de ranura para lograr la sincronización necesaria.

Para el proceso de actualización de τ , se considera que la estación base cuenta con el *hardware* necesario para poder realizar estimaciones del valor que considera más adecuado de acuerdo a la cantidad de vehículos dentro de su zona de cobertura para ese instante de tiempo mediante diferentes técnicas, incluidas las de aprendizaje de máquina que pueden requerir de unidades de procesamiento gráfico (GPUs). En el caso de los Transformadores, el uso de GPUs hace que su implementación sea bastante eficaz en tiempo, permitiendo hacer entrenamientos y predicciones bastante rápidas, sin embargo, implementaciones sin GPUs son posibles lo que permite que puedan desplegarse en sistemas convencionales.

Cada una de las ranuras puede presentar únicamente uno de los tres eventos posibles, los cuales son mutuamente excluyentes y se describen a continuación.

Ranura de tiempo exitosa. Las ranuras de tiempo exitosas ocurren cuando únicamente uno de los nodos transmite en dicho espacio de tiempo, en este caso, una ranura exitosa se presentará cuando únicamente un vehículo dentro de la zona de cobertura intenta transmitir

mientras que el resto de los vehículos permanecen en un estado de espera, ya sea con los transceptores apagados o en escuchando lo que ocurre en el canal con el receptor encendido.

Este evento ocurre con la siguiente probabilidad.

$$P_{suc} = \binom{N}{1} \tau (1 - \tau)^{N-1} = N \tau (1 - \tau)^{1-N} \quad (3.1)$$

Las ranuras de tiempo exitosas son los eventos que se desea sea mayormente presentes, En el proyecto se toma esta métrica para medir el desempeño de la red, y evaluar cuál de los diferentes esquemas que se evalúan en este proyecto tienen mejor desempeño. La presencia de una ranura exitosa representa que el canal se está utilizando de forma correcta, por ende, esta métrica indica el uso del canal y a mayor número de ranuras exitosas existentes, se estará haciendo un mayor uso correcto del canal.

Cabe notar que al considerarse un canal libre de errores, la probabilidad de que ocurra una transmisión exitosa depende únicamente de la cantidad de nodos que intentan transmitir en esa determinada ranura tiempo dentro de la zona de cobertura de la estación base. Sin embargo, considerando que la transmisión de los vehículos es omnidireccional, aquellos vehículos en las orillas de la zona de cobertura de cada estación base son propensos a transmitir con potencia suficiente para ser escuchados en las zonas de cobertura adyacentes, por lo tanto, el uso de frecuencias y/o códigos diferentes en las estaciones base, previene la interferencia de vehículos que no están dentro de la zona, logrando así un canal libre de errores. Esta consideración se toma para todas las ranuras de tiempo sin importar que tipo de evento presenten.

Ranura de tiempo libre. La ranura de tiempo libre es aquella en la que no se realiza ninguna transmisión, también puede ser considerada como una ranura de tiempo vacía. En este caso ningún vehículo intento transmitir en dicha ranura de tiempo, causando con ello un mal uso del canal debido a que durante ese tiempo no hay información viajando en él. De forma general, esto se puede deber a dos situaciones, una es que los nodos no tienen información para compartir o que la probabilidad de transmisión con la que operan los nodos intentarán transmitir es baja y los limita a enviar su información. Adicionalmente, aunque es haya un valor de τ alto, al ser un proceso aleatorio es susceptible a presentarse ranuras de tiempo vacías En el caso particular de este estudio los vehículos en cada ranura de tiempo tienen información que compartir, y se puede presentar una ranura libre debido a que la probabilidad de transmisión es muy baja para el número de nodos actuales por transmitir o por que en el proceso aleatorio ninguno transmitió. Así mismo es intuitivo notar que una ranura libre se da cuando no hay vehículos dentro de la zona de cobertura.

Este evento ocurre con la siguiente probabilidad.

$$P_{idle} = \binom{N}{0} \tau^0 (1 - \tau)^N = (1 - \tau)^N \quad (3.2)$$

Las ranuras de tiempo libres son una métrica de bajo desempeño, es decir que entre mayor número de ranuras de tiempo libre, menor será el desempeño de la red. Por lo tanto, es deseable que la probabilidad de transmisión con la que operan los vehículos dentro de la red sea tal que reduzca la cantidad de ranuras de tiempo libre, para así mejorar el uso del canal.

Ranura de tiempo con colisión. Debido a que en los protocolos de contención más de sólo un nodo pueden intentar transmitir en la misma ranura de tiempo, la presencia de colisiones es posible. Este evento ocurre debido a que se presenta más de una transmisión (se entiende que provienen de diferentes nodos) en la misma ranura de tiempo, ocasionando que la información contenida en los paquetes se destruya y ninguna alcance la estación base.

Un evento de colisión se genera desde dos intentos de transmisión hasta N intentos de transmisión al mismo tiempo. Por lo tanto, se define que la probabilidad de que transmitan i nodos es:

$$P_{tx}(i) = \binom{N}{i} \tau^i (1 - \tau)^{N-i} \quad (3.3)$$

Nótese que esta misma ecuación fue aplicada para calcular la probabilidad de una ranura exitosa y la de una ranura libre, para el primer caso $i = 1$ y para el segundo caso $i = 0$. En el caso de la probabilidad de colisión, ésta se calcula como la suma de probabilidades para $i = 2, 3, 4, \dots, N$, por lo tanto, la siguiente ecuación indica la probabilidad de una ranura con colisión.

$$P_{col} = \sum_{i=2}^N P_{tx}(i) \quad (3.4)$$

Observese que siguiendo las propiedades de las probabilidades, $\sum_{i=0}^N P_{tx}(i) = 1$ y en la ecuación 3.4 únicamente no se considera el caso de $i = 0$ e $i = 1$, por lo tanto la probabilidad de colisión se puede reescribir de la siguiente forma.

$$P_{col} = 1 - P_{suc} - P_{idle} = 1 - N\tau(1 - \tau)^{N-1} - (1 - \tau)^N \quad (3.5)$$

Al igual que la ranuras libres, las ranuras con colisión son métricas de mal desempeño, por lo tanto a mayor número de ranuras con colisión, menor será el uso del canal y por ende el desempeño de la red. Más que sólo causar que una ranura de tiempo no sea utilizada para enviar información, en el caso de una colisión también hay un efecto negativo ya que se entiende que los vehículos que intentaron transmitir consumieron energía y en el caso de dispositivos alimentados por baterías o donde el consumo de energía sea un aspecto crítico, una colisión tiene un impacto alto en el rendimiento energético, además de un mal uso de recursos del canal de comunicación.

Una vez definidos los eventos posibles en una ranura se concluye que es necesario que la probabilidad de transmisión con la que operan los vehículos debe de ser tal que maximice la cantidad de ranuras exitosas y que a la vez minimice las ranuras libres y la probabilidad de colisión.

3.4. Estimación de τ

El tiempo al final de cada segundo es utilizado para calcular el valor de τ con el que los vehículos tratarán de transmitir durante el próximo segundo. Como se ha mencionado anteriormente, contar con una probabilidad de transmisión adaptable a las condiciones actuales del sistema, permite mejorar el desempeño de los protocolos de contención, logrando en mayor medida que se haga un uso eficiente del canal.

En este caso se considera un canal libre de error, por lo que el valor de la probabilidad de transmisión estará en función de la cantidad de usuarios que hay dentro de la zona de cobertura de una estación base. En particular para S-ALOHA, se ha mostrado que el valor óptimo de la probabilidad de transmisión es $1/N$ [39], por lo tanto el principal reto es contar con mecanismos que permitan conocer el valor de N en cada segundo con la intención de establecer el valor óptimo de τ . A continuación se presentan las propuesta consideradas para la estimación de τ a instantes de tiempo futuros.

La gráfica de la Figura 3.5 muestra una comparación de la probabilidad de tener una ranura exitosa para diferente número de nodos en la red y escalando el el valor de τ para visualizar el comportamiento de ese parámetro en el desempeño de la red. Adicionalmente, en ella se observa que la probabilidad de un evento de éxito P_{suc} disminuye conforme mayor número de usuarios existe, acotado a 36.8% (.368) cuando el valor de N comienza a crecer, dicha cota es el valor estándar de la tasa de transferencia efectiva de S-ALOHA.

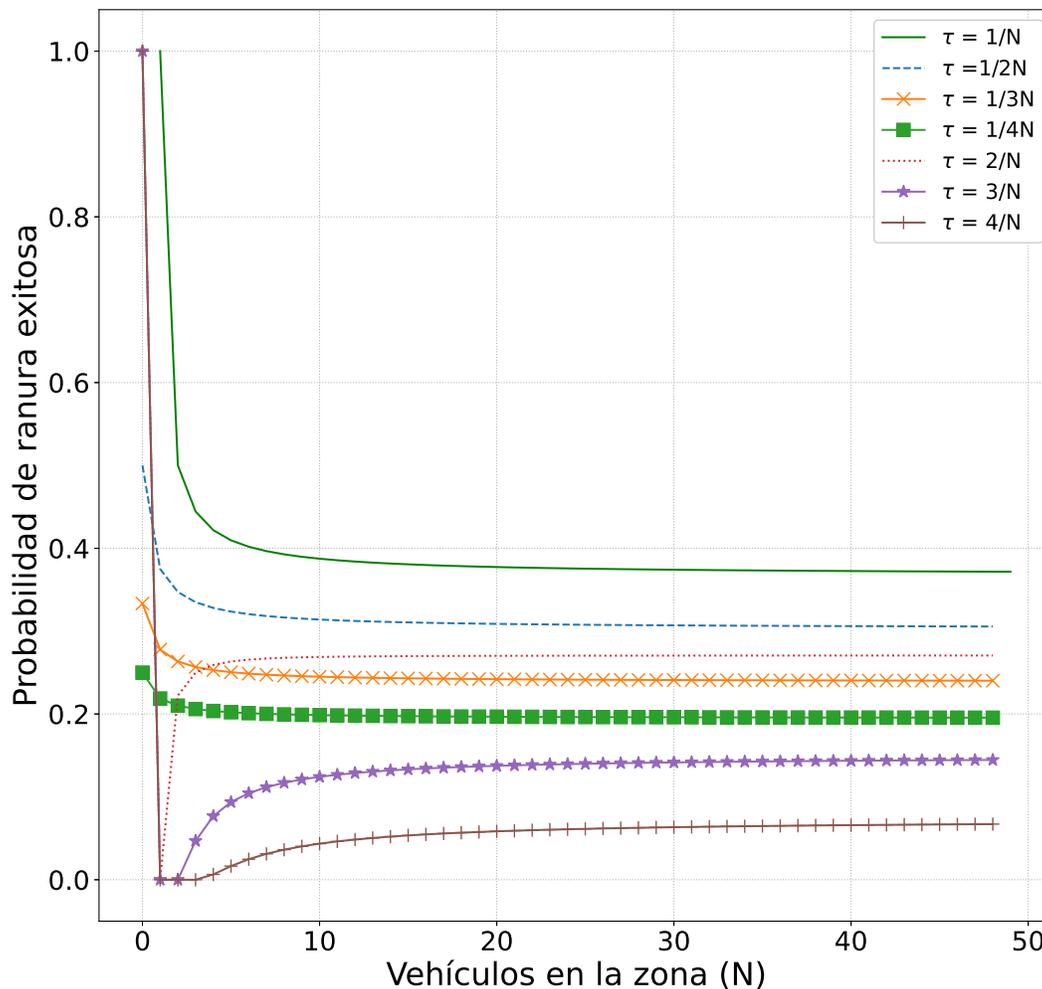


Figura 3.5: Comparación de probabilidad de ranuras exitosas con diferentes valores de τ

En las siguientes subsecciones se describen los esquemas elegidos para la selección de τ , se presenta el esquema fijo, que a pesar de no adaptarse a las condiciones del canal, es

considerado en este estudio para conocer la ventaja del uso de esquemas adaptativos con respecto a los esquemas fijos. Dentro de los esquemas adaptativos, se presentan diferentes esquemas, desde el uso de valores medios obtenidos por observaciones parciales a lo largo del día, como aquellos que utilizan técnicas de aprendizaje de máquina para encontrar la probabilidad de transmisión basada en el número de vehículos esperados para un futuro determinado.

3.4.1. Esquema fijo

Este esquema es la solución más simple y comúnmente utilizada, en el que la probabilidad de transmisión realmente no cambia a lo largo del tiempo, es decir, no se adapta a las condiciones del canal ni a la cantidad actual de usuarios dentro del sistema. Por lo tanto, los vehículos que se encuentren en la zona de cobertura de una determinada estación base transmitirán en todos los segundos del día con la misma probabilidad previamente establecida.

Normalmente el valor es asignado mediante una búsqueda a prueba y error de los valores que mejoran la desempeño del sistema. La principal desventaja es que no se adapta a los cambios presentes en la red y puede causar, incluso, bloqueos en el sistema. En este esquema el tiempo destinado para la estimación de τ no es para dicha función, sin embargo, esto no sucede con el periodo de tiempo para compartir el valor de τ pensando que las zonas en las que son colocadas las antenas cuentan con probabilidades diferentes obtenidas por estudios previos, ese espacio de tiempo permite a los vehículos entrantes a la zona de cobertura conocer la probabilidad con la que deben iniciar su transmisión.

Aunque se buscan probabilidades de transmisión fijas que maximicen las métricas de un correcto uso del canal, en el esquema fijo, es susceptible a presentar una tasa alta de las colisiones y los espacios de tiempo desperdiciados, los cuales son eventos no deseados. Por ende, es de esperarse que el desempeño general de la red sea bajo comparado con el de los esquemas adaptativos.

3.4.2. Esquema con media

La idea de contar con un sistema adaptativo es que éste permita operar a los vehículos con diferentes valores para la probabilidad de transmisión dependiendo de cómo cambia la cantidad usuarios dentro de la red. Una forma simple es utilizar valores medios de la cantidad de usuarios a lo largo del día para la zona específica donde se desea emplear el sistema adaptativo. De antemano, el valor medio supone que el número de autos dentro del área de cobertura sigue un patrón que se repite en determinados periodos de tiempo, por ejemplo, pensar que el día lunes en el horario de 10:00 a las 10:10 horas en la estación base 3 de la zona 1 se encuentran, en promedio, 10 vehículos por segundo. Es normal que no sea siempre el número exacto, por lo que habrá ocasiones en las que haya una cantidad mayor o menor de vehículos.

Con los valores medios para cada uno de los segundos a lo largo del día se estima el valor de τ y éste habrá de adaptarse de acuerdo a la media obtenida de las distintas observaciones realizadas, de tal manera que el valor de τ estará determinado de la siguiente forma:

$$\tau_t = \frac{1}{\overline{N_t}} \quad (3.6)$$

Donde t es el tiempo de observación que toma valores de 0 a 86399, lo cual corresponde a la cantidad de segundos dentro de las 24 horas del día, de esta forma la adaptación tiende a ser más cercana a la dinámica de usuarios en la zona, sin embargo existen algunas implicaciones.

En la Figura 3.6 se presentan gráficas representativas a partir de diferentes observaciones en la misma estación base desde las 10:00 hasta las 10:10, en donde se muestra la media de diez observaciones independientes en la misma zona y misma hora, el valor mínimo y máximo del número de vehículos obtenidos, así como la varianza y la desviación estándar de las observaciones. Un punto a destacar es que los valores máximos y mínimos de las observaciones en múltiples ocasiones generan un rango amplio de valores, por ejemplo, en las muestras tomadas a las 10:08:10 donde el número mínimo es 15, mientras que el valor máximo es de 30; el valor medio para ese punto es alrededor de 23. En este sentido, el valor de τ es mayor al ideal para el caso de tener el máximo valores de usuarios. Como se presentó en la Figura 3.5, valores más grandes o más pequeños de τ degradan el uso correcto de la canal al generar menos ranuras de tiempo utilizadas para la transmisión exitosa de paquetes de información.

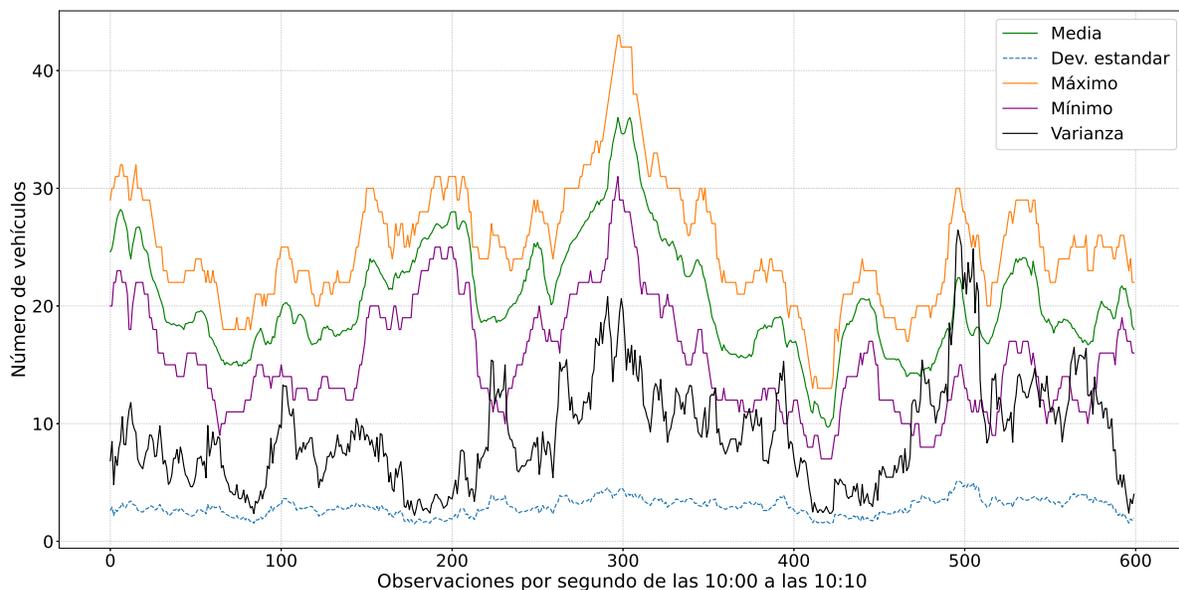


Figura 3.6: Cantidad máxima, mínima y media de vehículos en una zona, así mismo varianza, desviación estándar de diferentes observaciones

Además de la falta de precisión en el esquema adaptativo por valores medios, en este esquema es necesario contar con sistemas externos que permitan observar y actualizar constantemente el tráfico en las distintas zonas, esto con el fin de obtener valores más precisos de la media \overline{N} y, en consecuencia, valores de τ más adecuados de acuerdo a la cantidad de autos en la zona. A continuación se describe el funcionamiento de los esquemas basados en técnicas de aprendizaje profundo.

3.4.3. Esquemas basados en aprendizaje profundo

Los esquemas basados en aprendizaje profundo están pensados para utilizar información del pasado y con ella predecir valores futuros, en este caso, estos mecanismos buscan hacer una estimación del número de vehículos en tiempos futuros denotado por la variable N'_t , donde t es una variable para indicar el tiempo que se desea predecir y que siempre es mayor al instante de tiempo de la última observación hecha. Para este estudio t está en el orden de segundos.

Con las estimaciones de cada N' , el sistema será capaz de obtener el valor de τ de la siguiente manera

$$\tau_{t+k} = \frac{1}{N'_{t+k}} \quad (3.7)$$

Donde N'_t es calculada de la siguiente forma.

$$N'_{t+k} = f([N_{t-l+1}, N_{t-l+2}, \dots, N_{t-2}, N_{t-1}, N_t]) \quad (3.8)$$

Es importante mencionar que t y l son variables enteras con valores iguales o mayores a 0, además, se debe de cumplir que $k \geq 1$. De este modo, se observa que el estimado de N' para todo momento en el futuro, k segundos después del instante de tiempo t , es una función dependiente de las l observaciones previas, e incluyendo, al tiempo t . Por ende, se entiende que cada una de las l observaciones constituyen un vector que será la entrada a los modelos de predicción basados en aprendizaje profundo, y la salida de dichos modelos es el estimado N' .

Aunque en este estudio se considera N' como una salida de una dimensión, realmente la salida puede ser un vector de dimensión mayor a 1. Lo que representa un vector de salida de dimensión $d > 1$ es que el modelo está estimando la cantidad de vehículos para diferentes segundos en un futuro. Por lo tanto, para una salida que contemple más de una dimensión se puede definir como $[N'^t + k - 1, N'^t + k, N'^t + k + 2, \dots, N'_{t+k+d-1}]$ donde $d > 1$. La Figura 3.7 ejemplifica el vector de entrada y de salida de los modelos basados en técnicas de aprendizaje profundo.

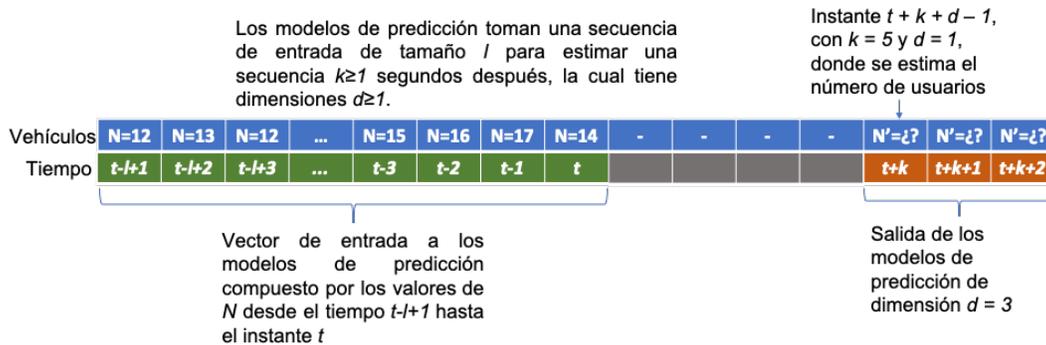


Figura 3.7: Entrada a los modelos de predicción y salida para un tiempo k posterior.

Los modelos de aprendizaje profundo requieren de pares entrada-salida para ser entrenados y de los cuales aprender los patrones que le permitirá estimar valores para entradas

futuras que no han sido vistas previamente. Principalmente se utilizaron modelos enfocados a la predicción en series de tiempo, como es el caso de las redes tipo LSTM y los Transformadores, donde la serie de tiempo es la secuencia con el número de vehículos que se observan en la zona de cobertura de cada estación base en cada segundo del día. A continuación se describen los modelos utilizados.

Redes LSTM

El modelo basado en redes recurrentes tipo LSTM está compuesto por una red profunda con capas integradas por unidades LSTM y una capa densamente conectadas. Específicamente, es un modelo secuencial de cuatro capas LSTM, cada una seguida de una capa de *Dropout* al 20%, al final se encuentra una capa densamente conectada de dimensiones igual a las dimensiones d del vector de salida, para el caso de la predicción de un solo valor N' esta capa será de tamaño de salida $d = 1$.

La primer capa de la red está compuesta por un número de unidades igual a l que es el tamaño la secuencia previa de donde se desea estimar el valor de N' , a la salida de esta capa se obtiene nuevamente una secuencia. Las siguientes dos capas, ambas, están compuestas cada una por un mayor número de unidades LSTM, en el orden de 10 veces la longitud de la secuencia de entrada. La cuarta capa de la red, está compuesta por unidades LSTM con una dimensión equivalente al tamaño de la secuencia de entrada l , la cual alimentará la última capa que consta de unidades densamente conectadas.

El proceso de entrenamiento de esta red es mediante el optimizador Adam, la pérdida es considerada con el error cuadrático medio, mientras que la métrica en el conjunto de validación es con el error absoluto medio y los valores en las secuencias son escalados mediante el método MinMaxScaler para normalizar en un intervalo cerrado entre 0 y 1. En la Figura 3.8 se presenta la arquitectura de la red, cabe mencionar que se variaron diferentes hiperparámetros por medio de búsqueda en cuadrícula para encontrar aquellos que ofrecieran los mejores resultados.

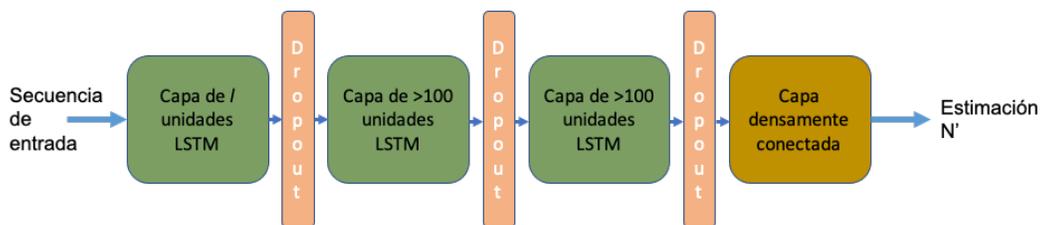


Figura 3.8: Arquitectura modelo basado en red LSTM.

Tiempo atrás las redes LSTM encabezaban el estado del arte en tareas de predicción para series de tiempo, por ello se retoman en este trabajo, sin embargo, por su funcionamiento tiene diferentes limitaciones, principalmente durante el proceso de entrenamiento. Esto se debe a que cada elemento de la secuencia es procesado de uno en uno en forma secuencial sin brindar la opción de utilizar computo paralelo o tarjetas gráficas que hacen más eficaz dicho proceso. Por otro lado, los Transformadores recientemente mejoraron al desempeño ofrecido por los modelos LSTM y permiten utilizar operaciones simultáneas en las secuencia con lo

que se logra, además, tiempos de entrenamiento más cortos. En la siguiente subsección se describe la arquitectura utilizada para los modelos basados en Transformadores.

Transformers

Los Transformadores son modelos utilizados en tareas de predicción de secuencias, normalmente se utilizan para la traducción de textos y la generación de lenguaje natural, sin embargo, dada su naturaleza de trabajar con secuencias, en este trabajo se abordan a los Transformadores como un modelo para la predicción de series de tiempo. Una de las ventajas de este modelo es la rapidez con la que entrenan ya que toman y procesan todos los elementos de la secuencia de entrada al mismo tiempo, esto permite una mejor utilización de *hardware* ya que es posible sacar provecho de las operaciones paralelas de las GPUs, incluso en el caso de que estén utilizando únicamente el CPU, el tiempo de procesamiento y entrenamiento de los modelos es más rápido que en LSTM. Sin embargo, el aspecto más destacable de los modelos basados en Transformadores es que mejoran el desempeño obtenido por las redes neuronales LSTM.

Al igual que las redes neuronales de tipo LSTM, los Transformadores son alimentados por una secuencia de entrada de tamaño l compuesta por el número de autos en los segundos previos a los que se desea estimar N' . Se emplea el modelo genérico compuesto por una pila de codificadores alimentados por las entradas embebidas en un vector de dimensiones superiores a la longitud de la secuencia l y cuya salida servirá para alimentar a una pila de decodificadores. La salida de esta última pila alimenta a una capa de red neuronal lineal la cual dará la salida con la estimación de N' .

Se contemplaron y evaluaron diferentes configuraciones de hyper-parámetros, obteniendo buenos resultados con la siguiente arquitectura, la cual es representada en la Figura 3.9.

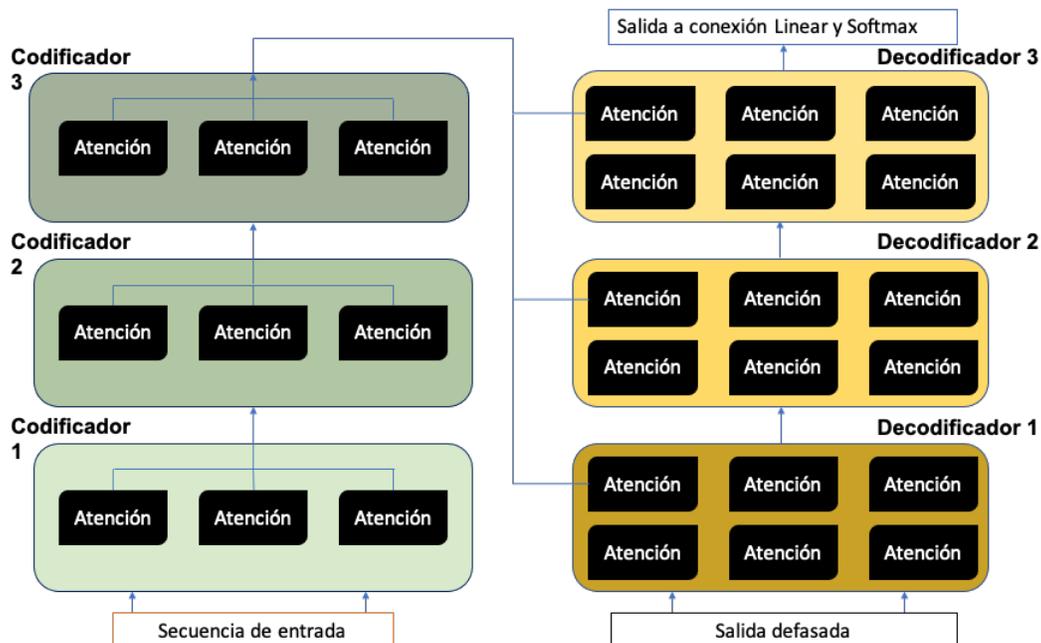


Figura 3.9: Arquitectura modelo basado en Transformadores.

La entrada al modelo es una secuencia de longitud l cuyos elementos son la cantidad de vehículos dentro de una determinada zona en cada segundo, cada elemento es expandido para generar una entrada embebida $l \times 10$. La pila de codificadores contiene 3 codificadores, cada uno de los codificadores contiene, a su vez, 3 módulos de atención, la salida de los módulos de atención pasa a través de una capa lineal para normalizar y añadir la información original con la intención de preservarla a lo largo del procesamiento de la misma y evitar su pérdida por causa anormales.

Por otro lado, la pila de decodificadores es alimentada por la salida de la pila de codificadores y por la salida previa de la secuencia en una forma enmascarada, esta pila está compuesta de 3 decodificadores con 2 bloques de multi-atención, cada bloque de multi-atención contiene 3 módulos de atención cada uno, el primer bloque de multi-atención es para enmascarar la salida previa de la secuencia, por lo tanto recibe como entrada dicho elemento en forma embebida, mientras que el segundo bloque de multi-atención recibe la salida del primer bloque y la salida de la pila de codificadores.

Similar a las redes LSTM, los valores de la secuencia de entrada son escalados mediante el método MinMaxScaler para obtener valores entre 0 y 1. Cada uno de los modelos es entrenado con el mismo conjunto de datos que fueron obtenidos por medio de la simulación en la ciudad de Luxemburgo en el simulador LuST. La obtención de datos es descrita a continuación.

3.5. Obtención de datos

Uno de los principales retos en el diseño de modelos de aprendizaje de máquina, incluidos los modelos de aprendizaje profundo, es la necesidad de contar con datos con los cuales entrenar los modelos y que con dicho entrenamiento puedan operar con datos nunca antes vistos, adicionalmente, en este proyecto también es necesario contar con datos con los que se pueda probar las propuestas de comunicación en la red vehicular. Debido a la alta dinámica existente en las redes vehiculares, los comportamientos del tráfico varían de un punto a otro y es complicado encontrar dependencias explícitas que puedan predecir el comportamiento de los automóviles.

Mediante el simulador SUMO se obtienen diferentes conjuntos de datos para el entrenamiento y prueba de los modelos de aprendizaje profundo, así como para probar el desempeño de la red vehicular a nivel del protocolo de comunicación para verificar que esquema utilizado para llevar a cabo la estimación de la probabilidad de transmisión es el que mejores resultados da. El proyecto LuST (Luxembourg SUMO Traffic) es una representación de la ciudad de Luxemburgo en el cual se imitan los patrones de movilidad vehicular en esa ciudad. Los autores de dicho proyecto lo plantean como una herramienta para desarrollar estudios referentes a los sistemas vehiculares [40]. El conjunto de datos generado en la simulación incluye el identificador de los vehículos, su posición espacial, el carril en que se encuentran, la velocidad a la que viajan, así como el tipo de vehículo, entre otros.

El archivo generado por SUMO contiene el mapeo completo de los vehículos a lo largo de las 24 horas del día, con los datos de cada uno de los automóviles dentro de la ciudad en cada segundo. El archivo ronda en un tamaño de 30GB, sin embargo, para este estudio mucha de la información contenida no es esencial. Cabe recordar que el problema se está abordando como una serie de tiempo que contiene la cantidad de usuarios solicitando acceso a una estación

base con cobertura en un determinado punto. Para ello, únicamente es necesario conocer la posición de los vehículos en cada segundo y contar la cantidad de ellos dentro de la zona de interés.

Se proponen tres zonas para la ubicación de las estaciones base considerando la distancia con respecto al centro de la ciudad. La primera zona está compuesta por un conjunto de estaciones base colocadas a una distancia de 500 metros del centro de la ciudad, la segunda zona comprende a las estaciones base situadas a 1000 metros del centro de la ciudad, mientras que la tercera y última zona cuenta con estaciones base a 2000 metros de distancia del centro de la ciudad de Luxemburgo. Se propone que las estaciones base de todas las zonas cuenten con un radio de cobertura de 200 metros. En la Figura 3.10 se presenta la distribución de las estaciones base.

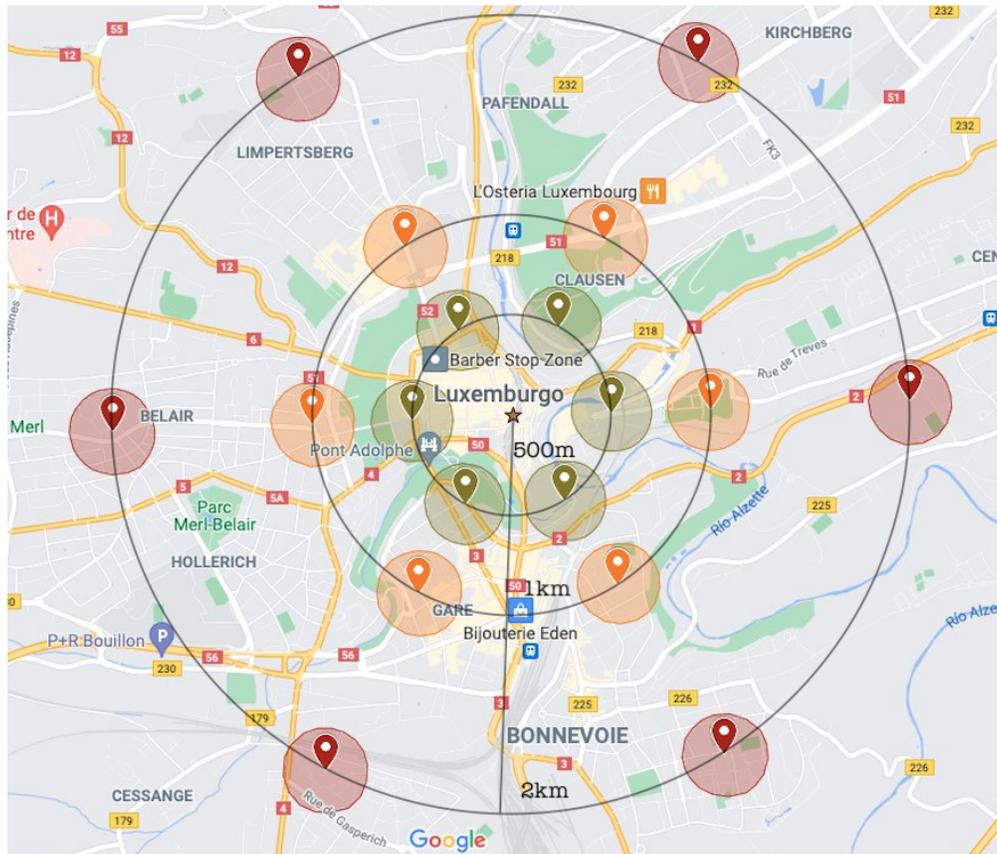


Figura 3.10: Colocación de las estaciones base en la ciudad de Luxemburgo.

Como se observa, la posición de las estaciones sigue un orden de separación entre cada una de ellas de 60° , iniciando en el punto 0° a 500 metros para la primera zona, 0° a 1000 metros para la segunda zona y 0° a 2000 metros para la última zona. Éstas son estaciones base ficticias que sirven como infraestructura para la VANET a las cuales los vehículos se conectarán, si bien no están exactamente a lado de la carretera su ubicación puede representar la posición de una estación base celular colocada dentro de un predio en la ciudad.

Finalmente el conjunto de datos con el que se trabaja en este proyecto, y que se obtiene después de contabilizar los vehículos reportados por la simulación de SUMO, está compuesto

por los segundos del día, empezando del segundo 0 al segundo 86399, que corresponde desde las 0:00:00 horas hasta las 23:59:59 horas, y por la cantidad de vehículos dentro de la zona de cobertura de cada una de las estaciones base. En la Figura 3.11 se ejemplifica la estructura del conjunto de datos utilizado en el proyecto.

Segundo	Z1_0	Z1_1	Z1_2	Z1_3	Z1_4	Z1_5	Z2_0	Z2_1	Z2_2	Z2_3	Z2_4	Z2_5	Z3_0	Z3_1	Z3_2	Z3_3	Z3_4	Z3_5
00:00:00	1	1	8	10	6	4	1	2	7	5	4	3	1	1	1	1	0	2
00:00:01	1	1	8	10	6	4	1	2	7	5	4	3	1	1	1	1	0	3
00:00:02	2	1	8	10	5	4	1	2	7	4	4	3	1	0	1	1	0	4
...																		
23:59:58	1	1	8	10	6	5	1	2	7	5	4	4	1	1	1	2	0	2
23:59:59	1	1	8	10	6	4	1	2	7	5	4	4	1	1	1	1	0	2

Figura 3.11: Conjunto de datos utilizado en el proyecto.

Mediante la opción de desplazamientos aleatorios de los vehículos dentro del simulador, cada que se ejecuta una nueva simulación se obtiene un nuevo conjunto de datos para finalmente contar con una tabla similar a la mostrada en la Figura 3.11 pero con valores diferentes. La columna Z1_0 representa la historia de vehículos en la estación base 0 de la zona 1 que se encuentra situada a 500 metros del centro de la ciudad. De tal forma que los modelos reciben como entrada una secuencia de la historia para hacer una predicción acerca de cuántos vehículos habrá en un determinado instante en el futuro en esa misma estación base.

Se obtuvieron 10 conjuntos de datos ya procesados en el formato descrito en la Figura 3.11, los primeros 9 fueron utilizados para entrenar los modelos de aprendizaje de máquina, mientras que el último de ellos fue empleado para probar el desempeño tanto de los propios modelos como de la red vehicular utilizando el protocolo de comunicación basado en S-ALOHA. En particular se evalúa si las predicciones de los modelos de aprendizaje profundo ofrecen un mejor desempeño al hacer mayor uso de las ranuras de tiempo para transmitir de manera exitosa.

En las Figuras 3.12 y 3.13 se presentan los histogramas del número de vehículos dentro de cada una de las zonas propuestas para el caso de las 24 horas del día y para el caso en que ya hay mayor actividad vehicular en un horario de las 4:00 a las 23:00 horas respectivamente, esto debido a que en un horario de las 00:00 a las 4:00 horas, existe un número alto de segundos con 0 vehículos en las distintas zonas. En el caso de estimar un valor medio para mantener fija la probabilidad de transmisión, esta característica impacta el promedio de vehículos intentando conectarse y es abordado en las secciones siguientes. El total de los histogramas corresponde a una simulación en particular, sin embargo, los 10 conjuntos de datos obtenidos presentan un comportamiento bastante similar.

Adicionalmente, se desea que el sistema sea autónomo y dependa en la menor medida de sensores o sistemas adicionales para estimar la cantidad de vehículos en la zona de interés, por lo que se propone que las transmisiones recibidas en la estación base sean la entrada a los modelos de aprendizaje de máquina para estimar el número real de automóviles en dicha zona.

Debido a que el paquete de información que envían los vehículos contiene un identificador

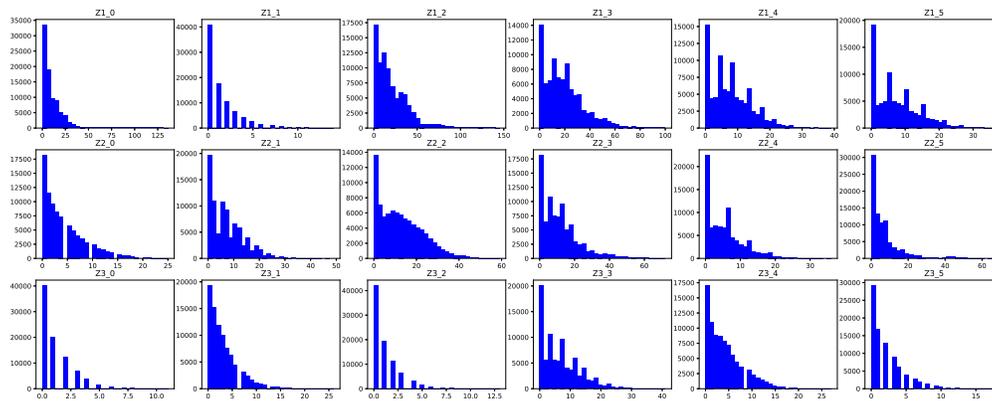


Figura 3.12: Histogramas del número de vehículos dentro de cada estación base a lo largo de las 24 hora.

único para cada uno de los usuarios, similar al caso de las direcciones MAC, contabilizar los automóviles que transmiten en cada segundo a una estación base es una tarea relativamente sencilla, sin embargo, es importante reconocer que este método de conserva la deficiencia de que exista la posibilidad que algunos de los vehículos no hagan una transmisión exitosa y, consecuentemente, la secuencia de entrada a los modelos de aprendizaje de máquina posea cierto ruido implícito. En la siguiente subsección se explica esta situación.

3.6. Estimación basada en transmisiones

Con la intención de generar un sistema autónomo, se propone que se utilicen las transmisiones detectadas en la estación base como entrada a los modelos de aprendizaje profundo para estimar el número de vehículos dentro de la zona de cobertura de la misma. De esta forma se evita la dependencia de sensores, cámaras o dispositivos que ayuden al conteo de vehículos. Sin embargo, se debe de tener en cuenta algunos hechos para poder implementar esta estrategia, los cuales son mencionados a continuación.

- Se asume que todos los vehículos dentro de la zona de cobertura de cualquiera de las estaciones base cuentan con el hardware necesario para conectarse a una infraestructura VANET.
- En un caso más general en el cual exista algún vehículo que no cuente con capacidades de comunicación con la VANET, éste no interfiere en el protocolo, debido a que no buscará conexión con la estación base. Para este estudio, todos los vehículos son usuarios de la VANET.
- La estimación del número de vehículos dentro de la zona de cobertura de la red se realiza basado en las transmisiones recibidas en la estación base.
- Todos los vehículos tienen posibilidad de transmitir en cada una de las ranuras de tiempo.

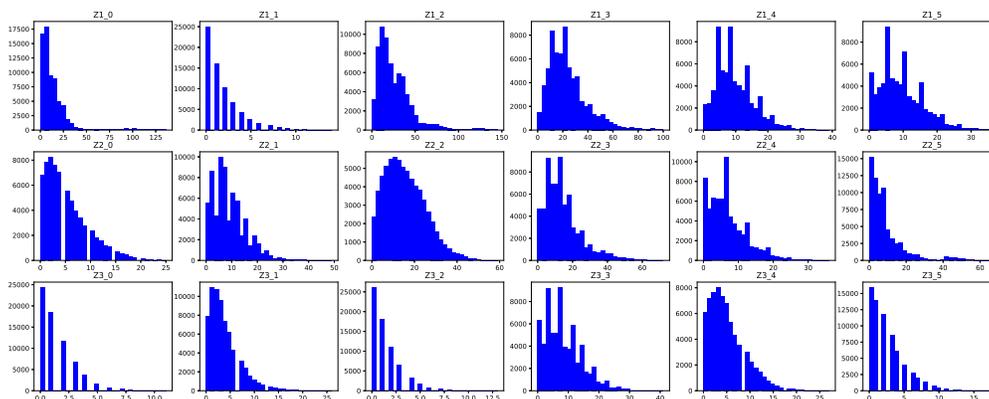


Figura 3.13: Histogramas del número de vehículos dentro de cada estación base de las 4:00 a las 23:00 horas.

- El número de ranuras de tiempo, $s = 375$, en las cuales los vehículos intentan transmitir antes de que se actualice el valor de τ es fijo. Esa cantidad de ranuras componen un trama de periodo $T = 1$ *segundo*.

Bajo estos principios se observa que los vehículos tienen s oportunidades de transmitir información antes de que se haga una actualización al valor de τ , sin embargo, dada las condiciones de contención y como se mencionó previamente, no todas las ranuras presentan una transmisión exitosa por lo que el valor de s juega un papel importante. Debido a que se desea que todos los vehículos puedan transmitir a la estación base al menos en una ocasión a lo largo de la trama, con la intención de tener un paquete de información de telemetría de cada vehículo, la comunicación de los vehículos a la estación base puede ser vista como problema del coleccionista de cupones.

En el problema del coleccionista de cupones se presenta un juego en el que hay un cupón identificable dentro de una caja, cada caja contiene un cupón que se puede repetir en una caja diferente. El reto es conocer cuántas cajas es necesario abrir hasta juntar al menos un ejemplar de todos los cupones. En el caso de la VANET, se plantea conocer el número de ranuras necesarias para que los vehículos transmitan al menos una ocasión, siendo ambos casos un problema aleatorio con reemplazo y que ha sido estudiado ampliamente en el análisis y diseño de algoritmos. De acuerdo a su análisis se sabe que si se tuviera conocimiento de donde está cada cupón se requiere de al menos N cajas para que los N cupones sean encontrados, en el caso de los vehículos se necesita de al menos N ranuras de tiempo para que los N vehículos puedan transmitir. Éste podría ser el caso particular de un protocolo programado como TDMA en donde cada nodo tiene su ranura asignada. Sin embargo, cuando el proceso es aleatorio, el valor esperado de intentos hasta que todos los vehículos hayan transmitido al menos una vez será similar a la expresión del problema del coleccionista de cupones expresado por la siguiente ecuación [41]:

Sea X el número de intentos hasta que los N vehículos transmitan al menos una vez,

$$E[X] = NH(N) \quad (3.9)$$

donde $H(N)$ es el armónico de N , expresado como

$$H(N) = \sum_{i=1}^N \frac{1}{i} \quad (3.10)$$

De acuerdo a esta expresión se puede calcular el número de ranuras necesarias para determinada cantidad de vehículos. Sin embargo, es necesario considerar que la comunicación de los vehículos es mediante un protocolo aleatorio, el cual presenta tres eventos diferentes y únicamente en las ranuras en que se existe una transmisión exitosa es posible que alguno de los N automóviles logre transmitir, por tanto esto es necesario de tomarse en cuenta. Dicho de otra forma el valor obtenido por $E[X]$ es el número de ranuras exitosas necesarias para que cada vehículo logre transmitir al menos en una ocasión. La probabilidad de una ranura exitosa fue explicada en la ecuación 3.1, por lo tanto, se propone que la expresión para calcular el valor de ranuras necesarias para que los vehículos transmitan al menos una vez, incluya a dicha probabilidad, obteniendo entonces la siguiente expresión:

$$E[X]_{ALOHA} = NH(N) \frac{1}{P_{suc}} \quad (3.11)$$

En el caso de contar con una probabilidad de transmisión ideal para los vehículos con $\tau = \frac{1}{N}$

$$E[X]_{ALOHA} = N \left(\sum_{i=1}^N \frac{1}{i} \right) \frac{1}{\left(1 - \frac{1}{N}\right)^{N-1}} \quad (3.12)$$

En la Figura 3.14 se presentan las gráficas que describen el comportamiento de la cantidad de ranuras necesarias basada en el número de vehículos dentro de la zona para que cada uno de ellos realice al menos una transmisión exitosa. La gráfica en color verde muestra el número de ranuras exitosas necesarias para que todos los vehículos transmitan al menos en una ocasión, mientras que la gráfica en color naranja representa el número de ranuras utilizando el protocolo ALOHA, considerando que en promedio de cada 3 ranuras de tiempo en el canal, únicamente 1 de ellas es una ranura exitosa. Adicionalmente, se grafica el número de ranuras propuestas, con la intención de visualizar a partir de qué cantidad de vehículos es menos probable de que todos ellos transmitan exitosamente, observando que a partir de 35 vehículos en la zona de cobertura, será menos probable que todos ellos envíen un primer paquete de información.

Por otro lado, observando el comportamiento del número de vehículos en la estación base 3 de la zona 1 es posible notar que existirán momentos en los que no todos los vehículos lograrán hacer al menos una transmisión exitosa, por lo que la estación base detectará una menor cantidad de vehículos. Se define al total de vehículos detectados por medio de las transmisiones mediante la variable \hat{N} .

Sin embargo, es importante notar que la aleatoriedad del protocolo de comunicación causa que, aunque exista una cantidad de ranuras mayor a la cantidad de vehículos, esto no garantiza que todos los nodos podrán realizar su transmisión. En la Figura 3.15 se observa como hay periodos de tiempo donde la cantidad de vehículos en la zona de cobertura es menor a 10 y la mitad de ellos no logran hacer una transmisión exitosa, por otra parte, el

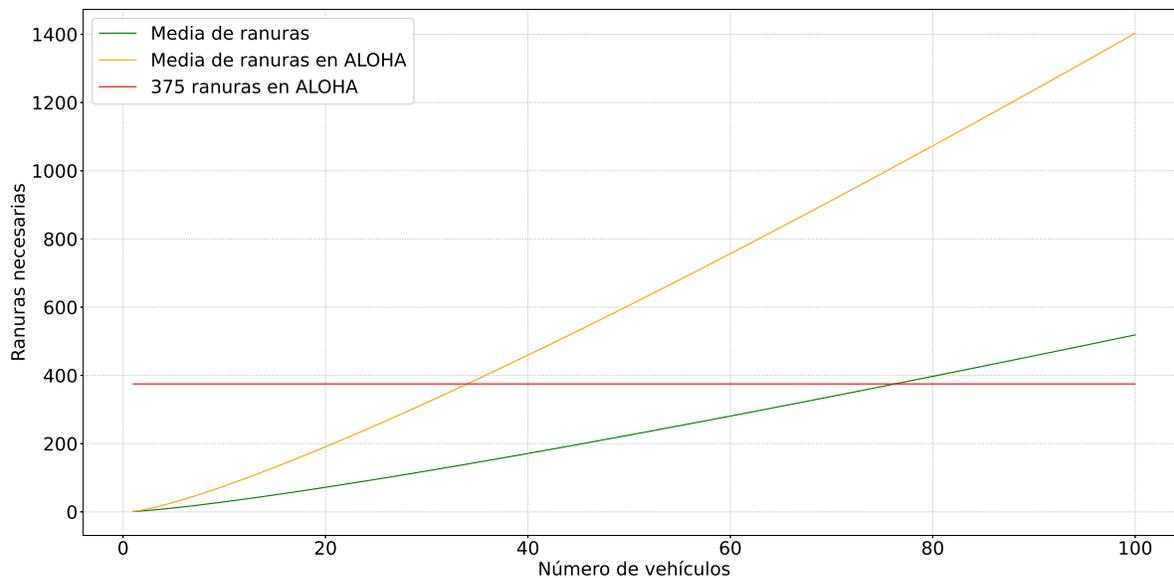


Figura 3.14: Comparación de número de vehículos contra número de ranuras necesarias para que al menos cada vehículo transmita una vez.

periodo con mayor cantidad de nodos que no transmiten es entre el segundo 64800 a 70000, que representa desde las 18:00hrs hasta al rededor de las 19:26hrs, con un valor promedio de $N > 50$ vehículos alcanzando un valor máximo de $N < 110$ vehículos, la cantidad de vehículos que no logran transmitir es menor a 40, lo que representa un 36.4% de vehículos no detectados. Ese es el escenario en el peor de los casos mientras que la mayoría de ellos está con una diferencia menor a 15 vehículos significando un valor porcentual del 20%.

La cantidad de vehículos que logran transmitir estará relacionada por dos variante, la primera y que se estudió en esta sección es la cantidad de ranuras de tiempo con respecto a la cantidad de vehículos en la zona de cobertura, mientras que la otra variante es la probabilidad de transmisión con la que los vehículos compiten por los recursos del canal. En este caso, se ha elegido dejar el número de ranuras fijas por cada segundo e introducir los modelos predictivos para estimar el número de vehículos N' basado en la cantidad de vehículos detectados \hat{N} con la intención de mejorar el desempeño de un protocolo de comunicación simple basado en S-ALOHA, considerando que la entrada a los modelos predictivos es una señal con información no precisa del número real de vehículos en la zona de cobertura, ya que el valor de $\hat{N} \leq N$, y a pesar de esa entrada, el valor de la predicción N' necesita ser bastante cercana al valor real del número de vehículos en la zona de cobertura de la estación base en cuestión (N) debido a que la probabilidad de transmisión será determinada como $\frac{1}{N}$.

En la Figura 3.2 se observó que el proceso de entrenamiento de los modelos de predicción basados en aprendizaje de máquina, ahora definidos como redes LSTM y Transformadores, implica realizar la simulación de la red con el protocolo S-ALOHA implementado. A continuación, se presenta de forma explícita como se componen los elementos involucrados en el proceso de entrenamiento partiendo del hecho de que las secuencias de entrada a los modelos son la cantidad de vehículos que logran transmitir en al menos una ocasión hacia la estación base y por lo tanto son detectados por esta última. Al presentarse una entrada

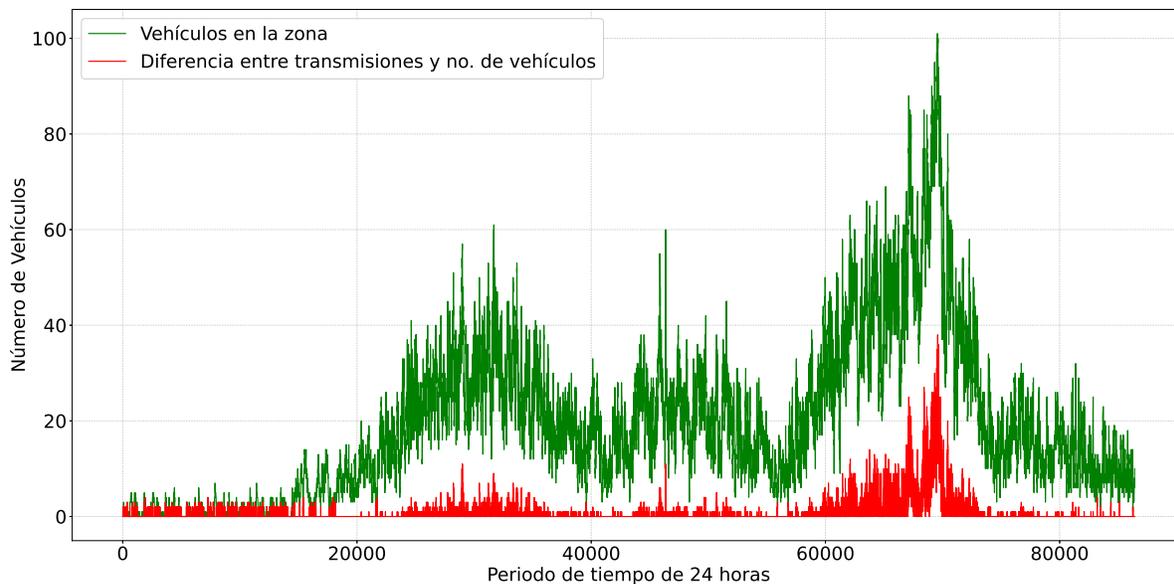


Figura 3.15: Comparación de número de vehículos (N) contra el número de vehículos detectados por la transmisiones recibidas en la estación base (\hat{N}).

con ruido, el proceso de entrenamiento es considerado como un proceso con regularización en donde la información de entrada no es perfecta y los modelos de ML necesitan aprender de esa información para predecir valores que sean representativos y cercanos al número real de vehículos dentro de la zona de cobertura.

El conjunto de entrenamiento es compuesto por un lote pequeño (*Batch*) de instancias en pares ordenados (X, Y) donde X es un vector de valores de longitud l con el número de vehículos correspondientes a los instantes de tiempo previos al que se desea estimar, obtenidos de la serie de tiempo con el número de vehículos detectados en cada zona de la estación base durante la simulación de S-ALOHA. Mientras que Y es el valor real del número de vehículos que se desplegaron en la zona de la estación base durante la simulación de SUMO para el instante de tiempo que se desea predecir. Así cada *Batch* contiene k instancias con los vectores X y Y correspondientes a diferentes instantes de tiempo en los que se desea predecir. En la Figura 3.16 se presenta el proceso de entrenamiento desglosado.

Cabe notar que τ tendrá un efecto en la estimación hecha por los modelos de predicción, ya que cuando su valor sea más pequeño al óptimo, será susceptible a que una menor cantidad de nodos hagan una transmisión exitosa, en caso contrario un valor de τ elevado, lejos de lograr más transmisiones, es susceptible a una cantidad alta de colisiones que en consecuencia genera también menor cantidad de transmisiones correctas. Por ende, la entrada a los modelos de predicción reciben información imprecisa y es necesario ejecutar ciclos de entrenamiento con las secuencias obtenidas por S-ALOHA como entrada a los sistemas, a fin de ajustar en mejor medida las secuencias con las que entrenan los modelos de predicción. En caso de contar con el valor real de vehículos dentro de la zona de cobertura, el entrenamiento se lleva a cabo directamente con los valores óptimos suponiendo que el comportamiento de los vehículos en implementación siguen una distribución similar a los datos de entrenamiento.

En el siguiente Capítulo se presentan los resultados obtenidos en el estudio de los distintos

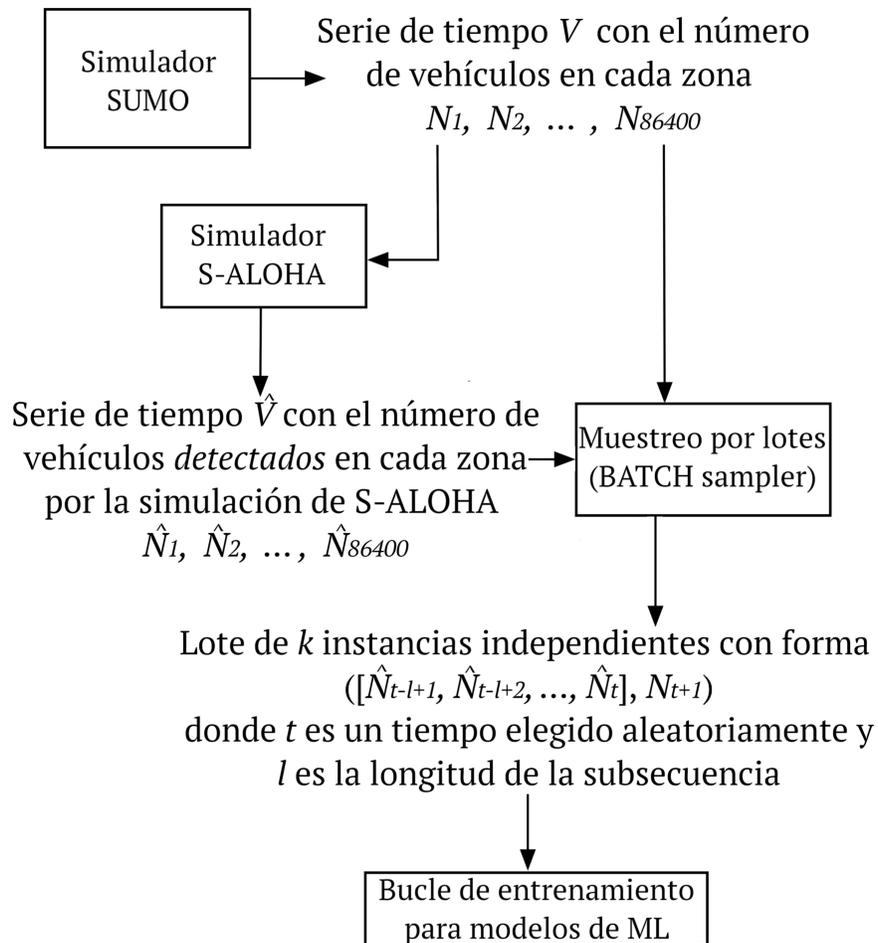


Figura 3.16: Proceso de generación de instancias y entrenamiento de modelos de Aprendizaje de Máquina.

esquemas y propuestas presentadas, en donde se aprecian los beneficios de emplear modelos de aprendizaje profundo para estimar la cantidad de vehículos en una zona determinada y con ello dar un valor a la probabilidad de transmisión τ de los nodos que optimice el uso del canal, con la intención de que la mayoría de las ranuras de tiempo contengan eventos de transmisión exitosa.

Capítulo 4

Experimentos y resultados de modelos de aprendizaje profundo en los protocolos de comunicación

En este capítulo se presentan los experimentos realizados y los resultados obtenidos al evaluar el desempeño de la red VANET utilizando los distintos esquemas para la asignación de la probabilidad de transmisión de los vehículos, τ , con la cual se busca mejorar el uso del canal bajo un protocolo de contención basado en S-ALOHA. La métrica principal que se emplea y la cual se desea maximizar es el porcentaje de ranuras de tiempo exitosas, ya que a mayor cantidad de ranuras exitosas, mayor cantidad de información útil estará fluyendo en la red. El modelo base que se toma es el modelo S-ALOHA con una probabilidad de transmisión ideal.

Adicionalmente, en estudios previos se encontró que es necesario acotar esta probabilidad cuando la red es susceptible a errores en el canal, como es el caso de ruido o interferencias que puedan ocurrir por sistemas vecinos. Como se planteó anteriormente, la presencia de interferencias de sistemas externos es mínima al considerar como infraestructura a las estaciones base de un sistema celular, en el cual las bandas de frecuencia son licitadas, sin embargo, el valor estimado N' está en función de una entrada con ruido que es la secuencia de \hat{N} causando que, a pesar de contar con $\hat{N} \leq N$, sólo se puede asegurar que $N' = N \pm \epsilon$ donde se busca que ϵ sea lo más cercano a cero. El principal problema es cuando $N' = 1$ pero en realidad $N \geq 2$ ya que el modelo está estimando que hay únicamente un vehículo intentando transmitir con $\tau = 1$, lo que representaría que los nodos en la red intentarían transmitir con probabilidad 1 en todas las ranuras de tiempo debido a que la estación base envía a todos los nodos esa probabilidad de transmisión y esto causará que el canal presente colisiones todo el tiempo ya que en realidad habría más de un vehículo con probabilidad de transmisión del cien por ciento.

Para unificar el estudio en los distintos esquemas se toma un umbral de transmisión máximo para que los nodos envíen su información, el cual se define como $\tau_{th} = 0.7$ debido a análisis previamente realizados [42]. Por lo tanto el valor de τ será definido por la siguiente expresión.

$$\tau = \begin{cases} \frac{1}{N'} & \text{si } N' > 1 \\ 0.7 & \text{otro caso} \end{cases} \quad (4.1)$$

Para el caso del esquema de elección fijo, el valor de τ se define sin cambios independiente del valor de N , en el caso del esquema con media, N' será igual a la media de vehículos observada por las simulaciones siguiendo la misma condición del umbral, y para el esquema ideal donde N es conocido, el cual será la cota superior del mejor desempeño esperado, se N toma en lugar de N' con la misma condición establecida en la ecuación 4.1. Las gráficas de este capítulo se obtienen a partir de visualizar la cantidad de eventos presentes en las ranuras de tiempo en lapsos de 10 minutos a lo largo de diferentes horas del día.

4.1. Esquema óptimo

En el esquema óptimo, los vehículos intentan transmitir con un valor de $\tau = \frac{1}{N}$ el cual se ha demostrado que es el valor óptimo con el que el protocolo de S-ALOHA obtiene el mejor desempeño. Este modelo es un ideal teórico que requiere del conocimiento exacto del número de vehículos, el cual en ocasiones es complicado de tener. En este estudio, se considera este esquema como aquel que ha de dar la cota superior para el desempeño esperado en los modelos descritos en el capítulo anterior.

En la Figura 4.1 se presentan las métricas de ranuras con transmisión exitosa, las ranuras libres, así como las ranuras colisionadas obtenidas para el modelo óptimo. A pesar de que la métrica principal es el número de transmisiones exitosas, también se puede intuir que reducir el número de ranuras colisionadas es una métrica de importancia ya que afecta el consumo de energía, por lo que, a mayor número de ranuras colisionadas, el consumo de energía por transmisiones realizadas sin éxito será mayor; por lo tanto un modelo de transmisión con buen desempeño también se verá reflejado en aquel con menor número de colisiones en el canal. Se puede observar que el esquema óptimo contiene una cantidad de ranuras exitosas superior al 37% (.37), siendo incluso superior al 40% cuando la cantidad de vehículos en la zona es menor. Cabe recordar que este comportamiento se observa en la Figura 3.5, e incluso con un solo nodo se alcanza el 100% de ranuras exitosas.

Otra situación importante a denotar es que la cantidad de colisiones se mantiene casi constante a lo largo del día, siendo únicamente más bajo cuando la cantidad de vehículos es realmente baja, es decir, en el horario de 4:00 a 5:00 horas. Consecuentemente, se observa que utilizar un protocolo de contención que opere con un valor óptimo de τ y que éste se adapte a la cantidad de nodos en la red, permite disminuir la cantidad de colisiones a un valor constante y maximizar la cantidad de ranuras con transmisiones exitosas. Por otra parte, la métrica de ranuras libres, se mantiene en todo momento por debajo de la cantidad de las ranuras con transmisión exitosa, con ello también se contribuye tanto a un mayor uso del canal como a una menor cantidad de consumo de energía ya que en las ranuras libres ningún nodo intenta transmitir y a nivel electrónico indica, generalmente, que la unidad de transmisión-recepción de los vehículos se encuentran en un modo dormido o de bajo consumo de energía.

En las siguientes secciones se muestra el desempeño obtenido por los esquemas descritos en el Capítulo anterior, observando que, si bien no se logra obtener desempeños superiores al óptimo, los modelos apoyados por aprendizaje profundo muestran métricas muy cercanas a las óptimas.

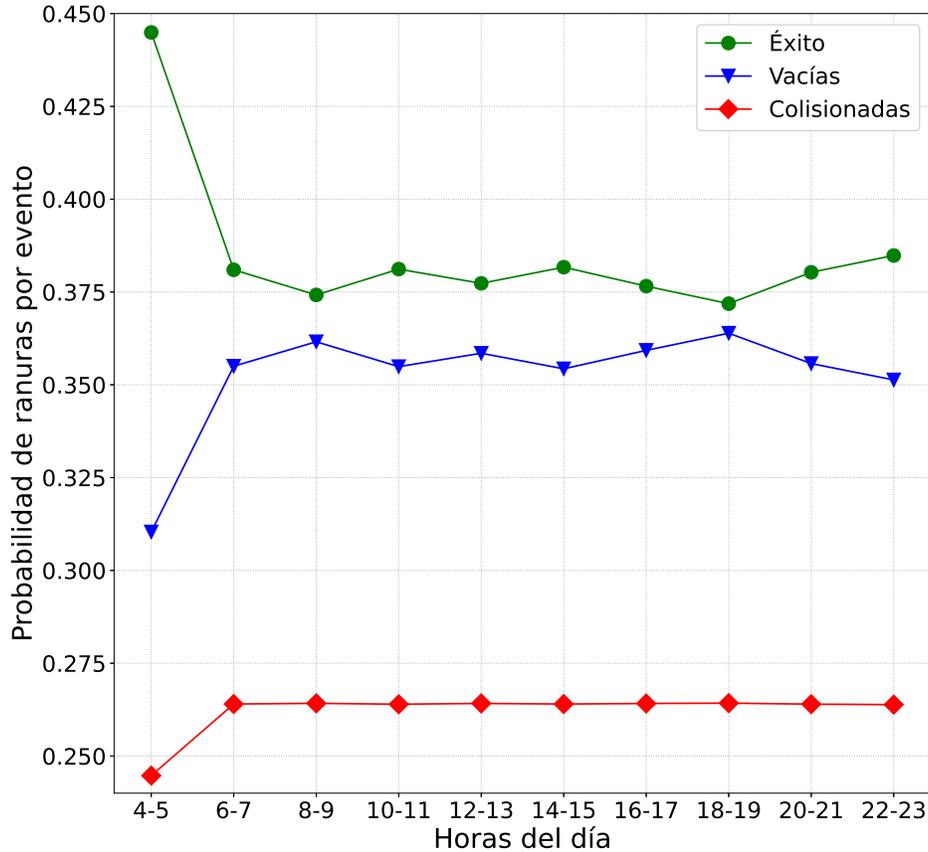


Figura 4.1: Métricas de desempeño del esquema con un valor óptimo de τ

4.2. Esquema fijo

El esquema fijo, es el más simple en el que se define una probabilidad de transmisión para todos los vehículos que tiene el mismo valor a lo largo del día. La ventaja de este esquema es que puede operar en cualquier instante sin hacer actualizaciones del valor de τ , por lo tanto no se adapta a las condiciones del canal y esto trae como consecuencia que su desempeño no sea el mejor de todos los esquemas.

En la Figura 4.2 se presenta el porcentaje de ranuras con transmisión exitosa, el porcentaje de ranuras vacías y el porcentaje de ranuras que presentaron colisión cuando el protocolo de transmisión utiliza una probabilidad de transmisión igual para toda hora del día. Después de revisar diferentes propuestas para el valor fijo que maximizaran el desempeño de la red, se observó que el valor que mejores resultados ofrece es aquel que se acerca al valor medio de vehículos observados en cada hora del día, es decir, para cada una de las horas se observó un comportamiento diferente respecto a los valores que toma τ .

Por lo tanto, el valor de τ es fijo respecto a la hora de observación, es decir

$$\tau_0 = \tau_1 = \tau_2 = \dots = \tau_{3597} = \tau_{3598} = \tau_{3599} = \frac{1}{\bar{N}} \quad (4.2)$$

Donde el subíndice de τ al segundo de cada hora y que es compuesta por un total de 3600

segundos, para este caso en particular \bar{N} es el valor medio observado a lo largo de esa hora en particular, por ejemplo de las 10:00 a las 11:00 horas. Por lo tanto

$$\bar{N} = \frac{N_0 + N_1 + N_2 + \dots + N_{3597} + N_{3598} + N_{3599}}{3600} \quad (4.3)$$

Y consecuentemente,

$$\tau_0 = \tau_1 = \tau_2 = \dots = \tau_{3598} = \tau_{3599} = \frac{3600}{N_0 + N_1 + N_2 + \dots + N_{3598} + N_{3599}} \quad (4.4)$$

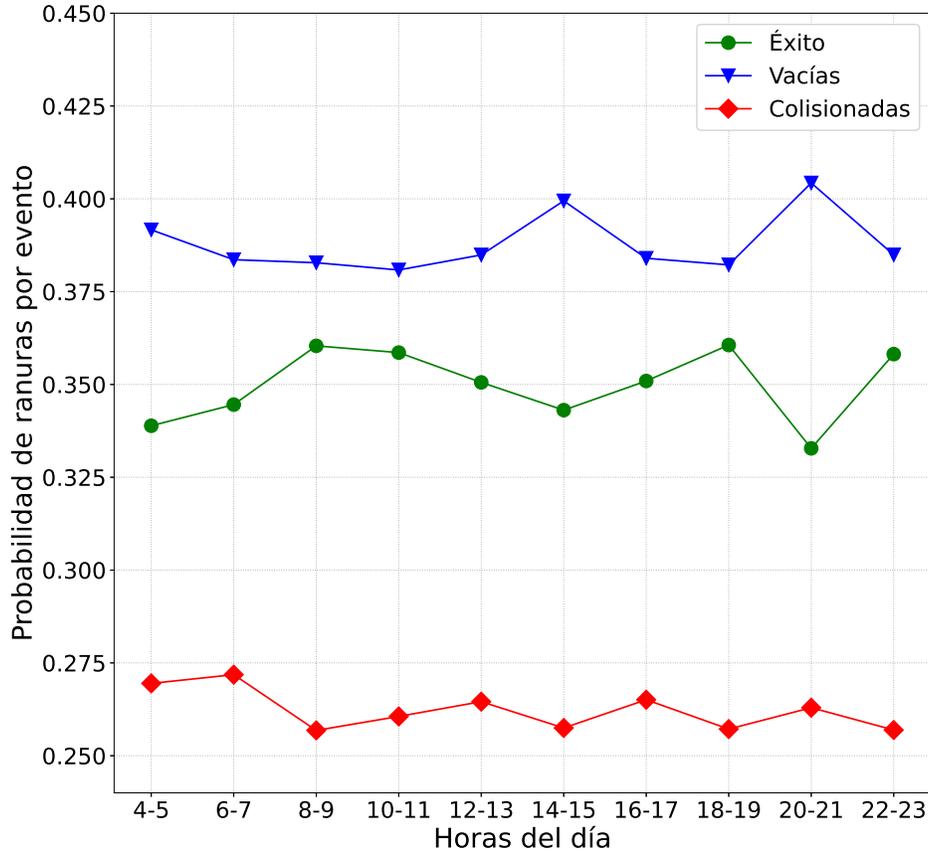


Figura 4.2: Métricas de desempeño del esquema con un valor fijo de τ para cada hora del día

Intuitivamente, este comportamiento es reflejo de que al menos por espacios de una hora es necesario tener una probabilidad de transmisión diferente. Esto se debe a que el valor medio a lo largo del día es representado por 19.44 vehículos, mientras que los valores máximos y mínimos son del orden de 100 y 0 autos, respectivamente. Además, el valor que más se repite, o sea la moda estadística, es 1. Por lo que operar a lo largo del día con sólo el valor de $\tau = \frac{1}{19} = 0.052$ causará un alto valor de ranuras vacías y a la vez generará un gran número de colisiones cuando el número de vehículos supere a los 20 existentes dentro de la zona de cobertura de la estación base.

Sin embargo, el desempeño obtenido para este modelo es, aún así, bastante pobre. Como se puede apreciar en las gráficas de esta Figura 4.2 se nota claramente la disminución porcentual del número de ranuras exitosas, mientras que se contiene en cierta medida la presencia

de colisiones con valores bastante cercanos al esquema óptimo, e incluso en algunos casos con menor número de colisiones. Por otro lado, la presencia de ranuras libres es mayor que el óptimo en todo momento, superando el 37.5%, mientras que el esquema óptimo permanece siempre por debajo de ese valor. Esto se refleja nuevamente que el esquema para este protocolo de contención necesita ajustarse en una granualidad más fina para obtener mejores desempeños. Adicionalmente, este esquema requiere de tener un control y exactitud del número medio de vehículos en la zona de cobertura de interés, al menos en lapsos de una hora a lo largo del día.

Por está razón se analizó un esquema en el que la probabilidad de transmisión sea calculada con un mecanismo de granualidad más fino. El esquema propuesto y que se analiza en la siguiente sección es el esquema con probabilidad adaptativa para cada segundo del día basado en la media de muestras previas.

4.3. Esquema con valor medio

Con los resultados obtenidos en la sección anterior se pudo observar que el uso de valores de τ con respecto a la media del número de vehículos presentes en una zona a lo largo de un periodo de tiempo permite obtener mejores resultados que utilizar únicamente un valor fijo a lo largo del día. Por ello en esta sección se presenta un esquema que basa el valor de τ en el valor medio de diferentes observaciones, en este caso simulaciones, para cada una de los segundos del día. Con esta propuesta, en lugar de tener un valor de τ fijo a lo largo de una hora en el día, el valor de τ será actualizado en cada segundo a lo largo del día, permitiendo una granualidad de adaptación más fina.

A pesar de que el número de vehículos para un segundo en particular cambia de una simulación a otra, el cambio no es tan alto como en el caso de una hora del día a otra. En este sentido, el valor de \bar{N} estará calculado para cada segundo t del día, con $t = 0, 1, 2, \dots, 86398, 86399$, de la siguiente manera.

$$\bar{N}_t = \frac{N_t(\text{Sim}0) + N_t(\text{Sim}1) + \dots + N_t(\text{Sim}7) + N_t(\text{Sim}8)}{9} \quad (4.5)$$

Se adopta la nomenclatura $N_t(\text{Sim})$ para hacer referencia a que se toma el valor de cada una de las simulaciones para ese instante t en particular. Nótese que el valor de Sim parte de Sim0 y llega hasta Sim8 ya que de las 10 simulaciones realizadas para obtener conjuntos de datos, la última de ellas es utilizada como un conjunto de validación para hacer las pruebas de desempeño tanto de los modelos productivos como del protocolo de comunicación.

A partir de esto, el valor de τ es calculado de la siguiente manera, siempre y cuando \bar{N}_t sea mayor a 1

$$\tau_t = \frac{1}{\bar{N}_t} = \frac{9}{N_t(\text{Sim}0) + N_t(\text{Sim}1) + \dots + N_t(\text{Sim}7) + N_t(\text{Sim}8)} \quad (4.6)$$

En la Figura 4.3 se presentan las gráficas de los resultados obtenidos para este esquema, en donde se puede observar que la probabilidad de encontrar una ranura exitosa incrementa sustancialmente respecto al esquema de τ fija. El comportamiento mostrado de la presencia de ranuras exitosas es ahora similar al del esquema óptimo, donde una baja cantidad de

vehículos permite alcanzar probabilidades superiores al 40 % mientras que la mayoría de las horas del día presentan un desempeño superior al 37.5 %. Por otro lado el número de colisiones continúa por debajo del 27.5 % (.2738 el mayor de los casos de 6:00 a 7:00 horas) sin embargo, éste no es constante como lo es el esquema óptimo y en tres ocasiones la cantidad presencia de colisiones es ligeramente mayor que en el caso del esquema fijo, específicamente en los horarios de 12:00 a 13:00 horas, de 18:00 a 17:00 horas y de 20:00 a 21:00 horas, donde, a lo más incrementa en 0.1 % respecto al esquema fijo.

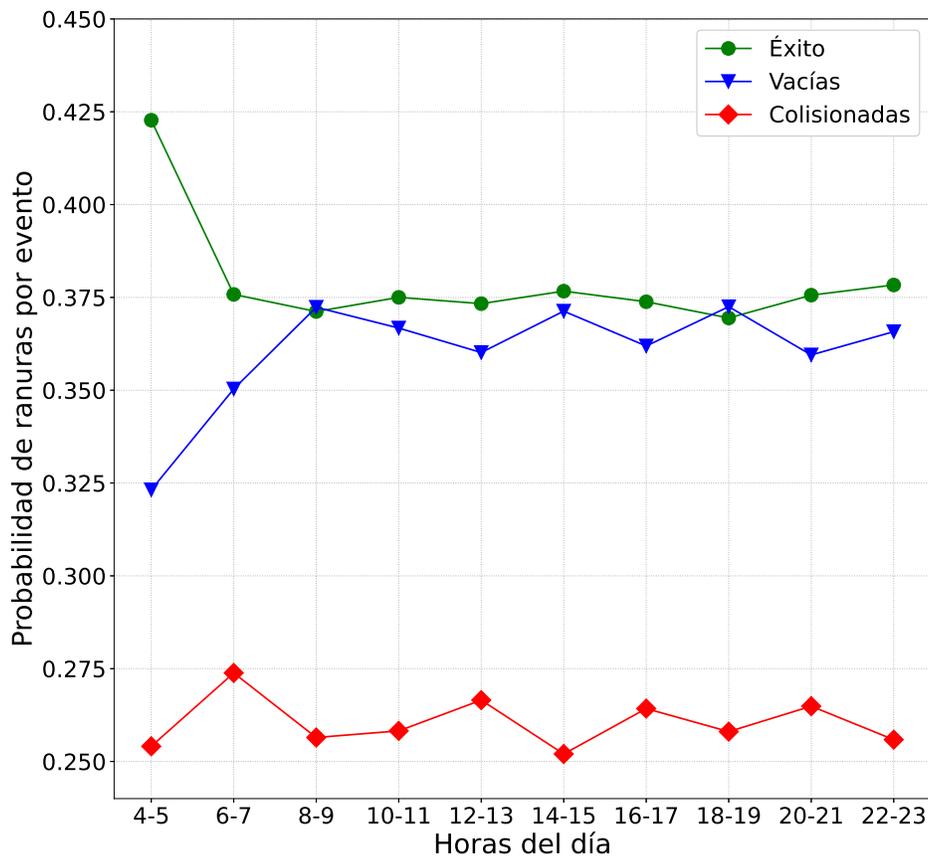


Figura 4.3: Métricas de desempeño del esquema con un valor de τ basado en la media de distintas observaciones para cada segundo del día

Para el caso de las ranuras libres, el esquema presentado en esta sección ofrece una gran mejora respecto al esquema fijo, ya que la probabilidad de ranuras libres se reduce significativamente con una presencia menor al 37.5 % para todas las horas analizadas, sin embargo, en un par de ocasiones la probabilidad de ranuras vacías o libres, supera la probabilidad de ranuras exitosas, observado que de 8:00 a 9:00 horas y de 18:00 a 19:00 horas, es más probable observar una ranura vacía que una exitosa. Cabe señalar que estos casos se presentan cuando hay una mayor cantidad de vehículos que la media, por lo que en horas pico el valor medio de distintas observaciones ofrece una probabilidad de transmisión ligeramente menor que la óptima.

En general una posible solución es basar la probabilidad de transmisión de los vehículos con el valor medio de varias observaciones en cada instante de tiempo. El principal problema

recae en que la estación base necesitaría tener el historial de observaciones para poder calcular el valor de τ , creando la necesidad de contar con sistemas auxiliares, como lo es el uso de sensores para el conteo de automóviles o cámaras que permita obtener la cantidad de vehículos en cada instante. Además, existirán días en los que no se presente ese mismo comportamiento en el tráfico vehicular, por ejemplo, en día lunes no se presenta el mismo número de autos en que el día sábado, por lo que este valor medio es susceptible a englobar valores atípicos dentro de un valor medio.

Con base en esta premisa, se plantea utilizar los modelos de aprendizaje de máquina, en especial aquellos de aprendizaje profundo que permita elegir un valor óptimo con base en los valores observados en una ventana de tiempo previa. En las siguientes secciones se describen los resultados obtenidos por los esquemas apoyados por modelos de aprendizaje profundo.

4.4. Esquema basado en redes LSTM

El esquema basado en LSTM está pensado para estimar el valor de N en un futuro dado, a través de los esquemas previos se ha observado que el principal problema es conocer el número actual de vehículos en la zona para poder establecer un valor de τ . En este sentido, el uso de aprendizaje profundo es la herramienta con la que se obtiene un valor estimado N' del valor real de vehículos en la zona, con la intención de que N' sea bastante similar a N . Con dicho estimado, se calcula el valor de τ para ese instante en el futuro.

Adicionalmente, se constató que calcular el valor de N con las media de diferentes observaciones mejora el desempeño de la red, pero como tal no captura la dinámica del flujo vehicular y eso es propenso a ocasionar valores de τ que no son óptimos. Con el uso de redes tipo LSTM se desea que, durante el proceso de entrenamiento, ésta capture la dinámica del flujo vehicular al verlo como una secuencia de tiempo y, así, estimar el número de vehículos para un tiempo futuro basado en la cantidad de automóviles presentes en un lapso de tiempo anterior.

En este caso se observó que el mejor desempeño fue obtenido cuando la secuencia de entrada a la red LSTM es de tamaño 10, y la salida o predicción para el instante de tiempo siguiente es de una longitud de 1 segundo. Con esta longitud de secuencia, la red neuronal logra aprender el comportamiento de la secuencia que representa a la dinámica interna del flujo vehicular dentro de la zona de cobertura. De esta forma la única información necesaria para hacer una estimación útil de N' es la cantidad de vehículos en la zona de cobertura de la estación base en los diez segundos previos. Una vez que se cuenta con el valor de N' y asumiendo que es muy similar al valor de N para ese instante de tiempo en que se ha estimado, el valor de τ es simplemente la inversa de N' .

En la Figura 4.4 se presentan las métricas de desempeño obtenidas por el esquema basado en LSTM, en donde se observa que este esquema supera ligeramente al esquema de basado en valores medios de cada segundo. La mejora se obtiene tanto en la probabilidad de ranuras con éxito con un valor notoriamente mayor a lo largo del día pero con menor probabilidad de éxito en el horario de 4:00 a 5:00 horas, en donde el desempeño cae por debajo del 37.5% contra el más de 40% del esquema con valor medio de segundos. Por su parte, la gráfica representando el porcentaje de ranuras libres se redujo considerablemente, siendo menor en todo momento a los porcentajes presentados en el esquema de valores medios. Sin embargo,

el costo a pagar es una mayor cantidad de colisiones.

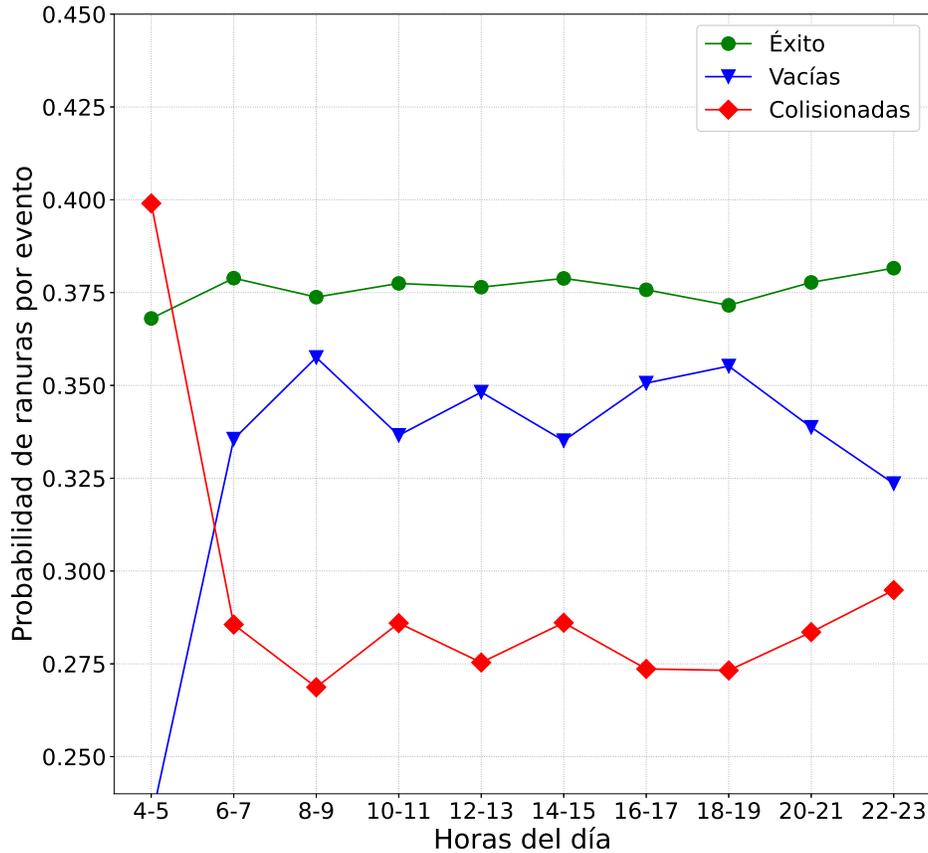


Figura 4.4: Métricas de desempeño del esquema con un valor de τ basado en redes neuronales LSTM de distintas observaciones para cada segundo del día.

Al contar con un valor mayor de colisiones se deduce que el valor de τ está siendo calculado de una forma no óptima y éste es más grande para la cantidad real de vehículos que están presentes dentro de la zona. Además, es importante mencionar que este esquema depende de conocer el número de vehículos de forma precisa en cada uno de los segundos previos para poder estimar el valor del número de automóviles que intentarán conectarse a la red en un tiempo futuro. Como se mencionó previamente, para que el modelo de estimación pueda predecir el valor de N' es necesario conocer l valores reales de N de forma previa y la estación base necesita de algún elemento externo para conocer dichos valores. En este caso en particular se da por entendido que existe un oráculo que en cada segundo da el valor previo de la cantidad de autos que estuvieron presentes en la zona.

Tal implementación del oráculo fue posible debido a las simulaciones obtenidas del proyecto LuST, sin embargo, en una implementación real no siempre será posible contar con esa información. En el caso especial de las redes inalámbricas operando bajo un protocolo de contención, es complicado conocer el valor del número de usuarios que están compitiendo por recursos del canal, si bien cada vehículo contiene un identificador que lo hace único e identificable en la red, al existir un protocolo de contención no se garantiza que todos los nodos puedan transmitir un paquete a la concertadora para darse a conocer en la estación

base y que ésta pueda contar el valor real de vehículos dentro de su zona de cobertura.

A continuación se presentan los resultados del esquema basado en redes LSTM utilizando las transmisiones detectadas para estimar la siguiente cantidad de vehículos dentro de la zona de cobertura con la intención de resolver el problema de depender de un oráculo para la entrada de la red neuronal que ayudará a estimar el valor de τ para el instante de tiempo siguiente.

4.4.1. Estimación por transmisiones detectadas, redes LSTM

Como se explicó en la sección 3.6, la cantidad de vehículos calculados en la zona de cobertura debido a las transmisiones detectadas se define como \hat{N} y esta cantidad para cada uno de los l previos segundos es utilizada en la estimación del nuevo valor de N' . Con ello se obtiene el beneficio de contar con un sistema independiente que lleve el conteo del número de vehículos presentes en los instantes anteriores. Sin embargo, al ser un protocolo de contención, y como se observó en la Figura 3.15, los identificadores de vehículos detectados por las transmisiones recibidas no siempre refleja el número de vehículos dentro de la zona de cobertura. En una implementación real esto se puede deber a colisiones, ruido, o incluso que el vehículo no intenta transmitir en un lapso de tiempo dado.

Por lo tanto, la red neuronal LSTM fue nuevamente entrenada para obtener un modelo diferente para que sea capaz de estimar el número de vehículos N' con base en la cantidad de vehículos que la estación base logra detectar por medio de las transmisiones recibidas \hat{N} , considerando que en todo momento $\hat{N} \leq N$. En este caso, es importante notar que existe la posibilidad de que la estación base detecte menos cantidad de vehículos a los que realmente están intentando conseguir recursos del canal, lo que causaría que el valor de τ que se asigne al instante de tiempo siguiente será más alto al que tendría que poseer causando una tasa de colisiones mayor y degradando el desempeño del uso del canal.

El modelo de redes neuronales LSTM se reajustó para que a la entrada reciba una secuencia de longitud l con los valores \hat{N} de cada uno de los l segundos previos, para que a la salida el modelo de un estimado N' . Nótese como en este caso el sistema no depende del oráculo para conocer los valores previos de N y el *hardware* de la estación base sería suficiente para detectar a los vehículos y hacer las estimaciones. El oráculo, o valores de N de la simulación, son utilizados únicamente en el proceso de entrenamiento para constituir las instancias de la siguiente manera.

$$X_i = [N_{t-l+1}, N_{t-l+2}, \dots, N_{t-2}, N_{t-1}, \hat{N}_t] \quad (4.7)$$

Donde X_i es la i esima entrada al modelo predictor LSTM para estimar la salida N'_i correspondiente a algún instante de tiempo t , y que es entrenada con el valor de salida esperado cuya nomenclatura es Y_i , con

$$Y_i = N_{t+1} \quad (4.8)$$

De tal forma un conjunto de pares (X, Y) son utilizados para entrenar el modelo de red LSTM, con lo que se pretende aprender que número real de vehículos están intentando transmitir dado que se detectaron una cantidad vehículos que realizaron transmisiones exitosas en

momentos previos hacia la estación base. A pesar de que el modelo desplegado ya en implementación no necesita el número real de vehículos en tiempos anteriores dentro de la zona de cobertura para la estimación del futuro valor de τ , el proceso de entrenamiento continua presidiendo de ellos.

En las gráficas de la Figura 4.5 se presentan los resultados obtenidos al entrenar el modelo de red neuronal LSTM con las consideraciones mencionadas previamente, donde se observa que el desempeño se mejora significativamente. Claramente el porcentaje de éxito se mejora con respecto al modelo dependiente del oráculo, en especial al principio y al final del día de 4:00 a 7:00 y de 20:00 a 23:00 horas cuando la cantidad de vehículos no es tan alta. Por otro lado, el caso de la probabilidad de ranuras libres incrementó considerablemente ofreciendo porcentajes mayores al 35% en cada momento, incluso, de las 4:00 a las 5:00 horas la probabilidad de tener ranuras libres es mayor a la de ranuras con transmisión exitosa. Este hecho más que reducir el desempeño del canal por la cantidad de ranuras vacías, realmente representa una mejora respecto al esquema LSTM con oráculo debido a que la cantidad de colisiones disminuye considerablemente, ya que la probabilidad de que exista este último evento, cae por debajo del 27.5% en todo momento, significando que la probabilidad de transmisión con la que transmiten los vehículos permite incrementar la probabilidad de éxito mientras que mantiene en un nivel bajo la cantidad colisiones.

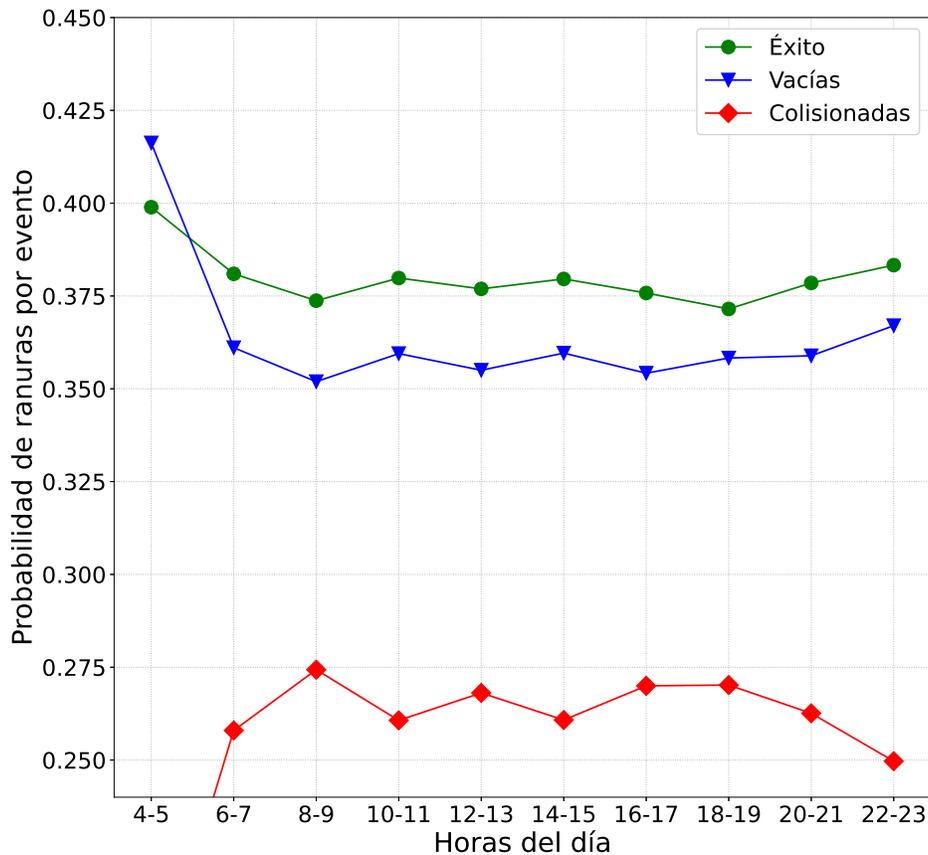


Figura 4.5: Métricas de desempeño del esquema con un valor de τ basado en redes neuronales LSTM de distintas observaciones para cada segundo del día.

A pesar de que el porcentaje de éxito, es bastante cercano al obtenido con un modelo óptimo, aún es más elevada la probabilidad de presentar colisiones en los momentos en que hay una cantidad de tráfico alto, e incluso medio. Por otro lado, con estos resultados se corrobora que el modelo que no depende de un oráculo da resultados cercanos al óptimo y es capaz de realizar una estimación certera sin depender de un oráculo, lo cual lo hace más factible a ser implementado en un entorno real.

En la siguiente sección se presentan los resultados concernientes al esquema basado en Transformadores. Al igual que en el esquema de esta sección, se considera el caso en que se cuenta con un oráculo donde se conoce el número real de vehículos en tiempos previos y aquel en donde la cantidad de automóviles se contabiliza con respecto a las transmisiones recibidas en la estación base.

4.5. Esquema basado en Transformadores

Los Transformadores son un modelo de aprendizaje profundo que actualmente ofrecen un alto desempeño en tareas de predicción de secuencias. En este caso se utilizaron los Transformadores para estimar el valor de N' en un instante de tiempo futuro. Al igual que en el esquema apoyado por redes LSTM, la entrada a los Transformadores es una secuencia de longitud $l = 10$ conteniendo la cantidad de vehículos para cada uno de los l previos segundos para que a la salida se obtenga el valor estimado de N' representando el número de automóviles para un segundo después del que se hace la estimación.

En el primero de los casos se asume que la entrada al Transformador es un vector compuesto por el número real de vehículos dado por un oráculo que, de alguna manera, conoce dichos valores, y esto es tanto para el proceso de entrenamiento como para el sistema ya operando en la estación base, es decir cuando ya se encuentra implementado para estimar N' dentro de la zona de cobertura. Nuevamente, con el valor que se estima de N' por medio del Transformador, se calcula el valor de la probabilidad de transmisión $\tau = \frac{1}{N'}$ permitiendo que el valor se ajuste al estimado de vehículos dentro del área de cobertura de la estación base.

En la Figura 4.6 se presentan los resultados obtenidos por el esquema basado en Transformadores, en la cual se puede observar que el comportamiento de las tres métricas es bastante cercano al del modelo óptimo. Además de obtener un comportamiento o tendencia de las gráficas muy similar al óptimo, los valores también son muy cercanos ya que en los horarios de las 6:00 a las 21:00 horas la diferencia de porcentaje de éxito es de a lo más 0.06 % mientras que en los horarios de los extremos que es cuando hay una cantidad baja de vehículos, difiere en 0.35 % para el caso de las 6:00 a las 7:00 horas mientras que de las 22:00 a las 23:00 horas, la diferencia es de 0.12 %. Por otra parte el porcentaje de la presencia de ranuras libres se mantiene en todo momento por debajo del 37 % (0.37) y en ninguno de los casos supera a la probabilidad de ranuras con transmisión exitosa.

Además, el porcentaje del número de colisiones se mantiene en todo momento por debajo del 27 % y se observa que el comportamiento de la gráfica tiende a ser constante de las 6:00 a las 23:00 horas, tal como ocurre en el caso del esquema óptimo observado en la Figura 4.1. Lo cual significa que el modelo basado en Transformadores logra predecir en una forma muy certera la cantidad de vehículos en un instante de tiempo siguiente y eso le permite elegir un valor de τ muy cercano al óptimo, en secciones posteriores se presenta la comparación de los

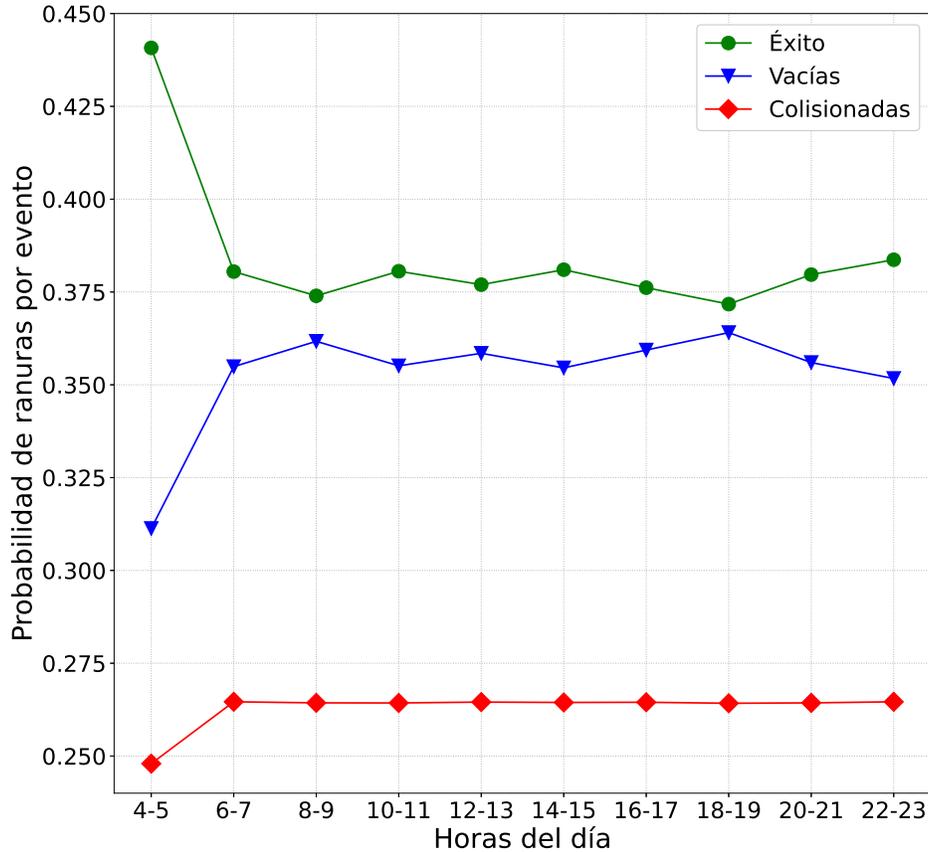


Figura 4.6: Métricas de desempeño del esquema con un valor de τ basado en Transformadores para cada segundo del día.

modelos para visualizar mejor dicho comportamiento.

Sin embargo, como se planteó desde el principio, este modelo depende de un oráculo que no siempre es posible tener a la mano, en otras palabras, depende de la información precisa de cuántos vehículos hay en la zona. En el caso en que se quiera utilizar el modelo ya entrenado en una implementación en la que la cantidad de vehículos sea estimada por la estación base por medio de las transmisiones recibidas, el desempeño del uso del canal decae ligeramente cuando la cantidad de vehículos en la zona de cobertura es alta, generando una menor probabilidad de ranuras con transmisión exitosa e incrementando la probabilidad de presentar colisiones. Este comportamiento se puede observar en la Figura 4.7

A pesar de que los resultados son muy cercanos al óptimo cuando se emplea un esquema basado en Transformadores y un oráculo da el número real de vehículos dentro de la zona de cobertura en los instantes anteriores a la estimación de N' , la implementación de dicho modelo resulta complicada en un mundo real, y al observar las gráficas del desempeño cuando la red ha sido entrenada con el oráculo y se implementa utilizando las transmisiones recibidas en la estación base, es claro notar que para continuar obteniendo un desempeño similar al óptimo es necesario volver a entrenar el modelo, por lo que en la siguiente subsección se presentan las modificaciones hechas al esquema.

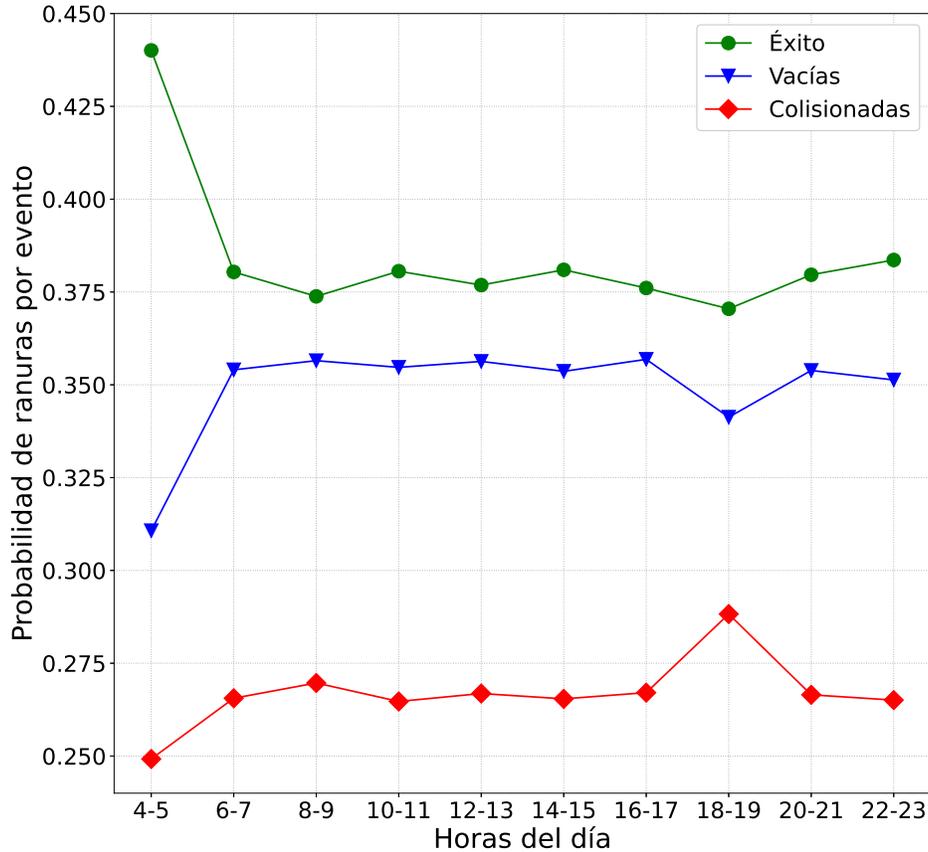


Figura 4.7: Métricas de desempeño del esquema basado en Transformadores que es entrenado con un oráculo y en la implementación recibe de entrada el número de vehículos detectados por la estación base.

4.5.1. Estimación por transmisiones recibidas, Transformadores

Con la intención de diseñar un esquema que no sea dependiente de un oráculo para saber la cantidad de vehículos en instantes previos y que sólo sea autónomo para determinar la cantidad de éstos por medio de las transmisiones recibidas en la estación base se entrena un nuevo modelo de Transformador en el que la entrada está compuesta por una serie de longitud $l = 10$ con la cantidad de vehículos que la estación base detecta en l segundos previos para estimar N' del instante de tiempo siguiente.

Similar a la estimación propuesta en la sección 4.4.1, las instancias de entrenamiento están conformadas por un par ordenado (X, Y) donde:

$$X_i = [N_{t-l+1}, N_{t-l+2}, \dots, N_{t-2}, N_{t-1}, \hat{N}_t] \quad (4.9)$$

X_i es un vector de longitud l que contiene la cantidad de vehículos que ha detectado la estación base en cada segundo previo a la estimación, considerando que en los paquetes que envía cada uno de los automóviles contiene un identificador único que permite conocer cuantos nodos transmitieron de forma exitosa. Por otro lado, $Y_i = N_{t+1}$.

Cabe resaltar que esta forma de la instancia es únicamente en el proceso de entrenamiento

del Transformador, una vez que se concluye tal proceso, el Transformador continua recibiendo como entrada para la predicción el mismo formato de X_i con la cantidad de vehículos detectados por la estación base \hat{N} , pero la salida de él será la estimación de vehículos dentro de la zona de cobertura N' , pretendiendo que la estimación sea bastante cercana a N . Posteriormente, la probabilidad de transmisión de los automóviles se calcula como $\tau = \frac{1}{N'}$

En la Figura 4.8 se presentan los resultados obtenidos por esta modificación al esquema basado en Transformadores en donde se observa que ahora el comportamiento de la probabilidad de éxito mejora en el caso de horas con tráfico alto, esto respecto al modelo que no se entrena con el oráculo y se implementa con la detección de vehículos por transmisiones.

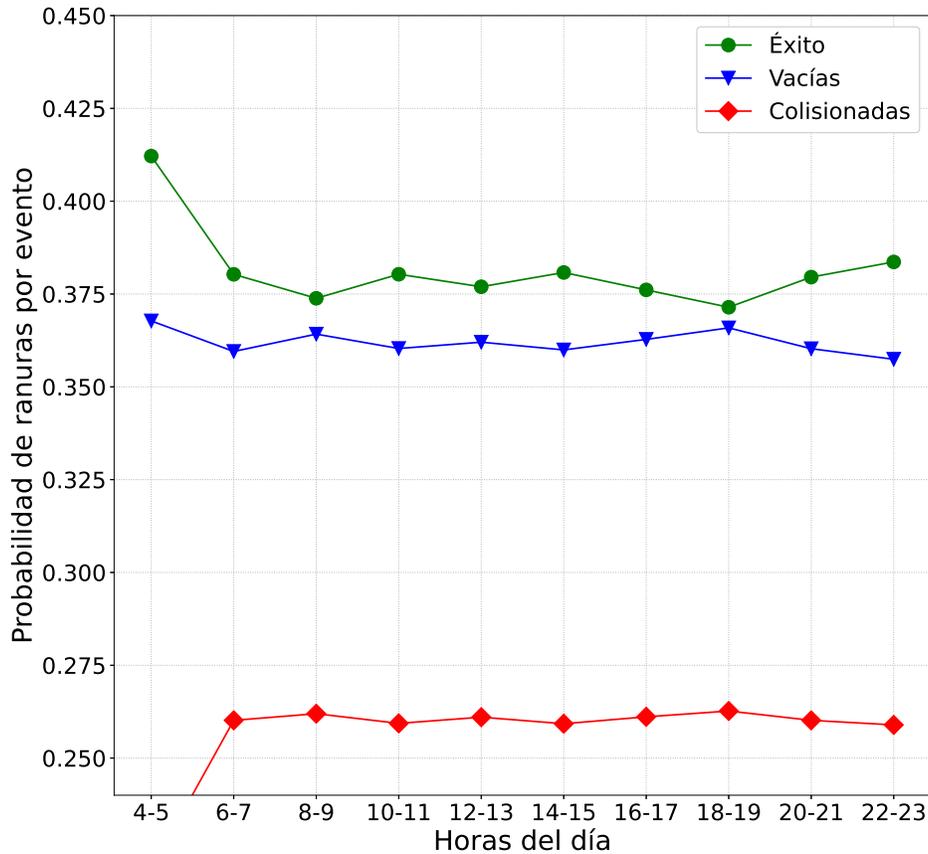


Figura 4.8: Métricas de desempeño del esquema basado en Transformadores que es entrenado e implementado con entradas del número de vehículos detectados por la estación base.

En el caso de la probabilidad de éxito de 4:00 a 5:00 horas, el desempeño se reduce en un 5% respecto al modelo de oráculo, sin embargo la mayor mejora se observa en el horario de las 18:00 a las 17:00 horas donde la cantidad de colisiones reduce significativamente. Esto, además, significa que la técnica de Transformadores opera con resultados cercanos al óptimo en un entorno más cercano a la realidad.

Por otra parte, este proceso de entrenar los modelos de aprendizaje profundo con el número de vehículos detectados representa una acción similar a la de agregar ruido a la entrada para generalizar en la entrada del sistema y mitigar problemas como sobrentrenamiento. Esta técnica es conocida en el mundo del aprendizaje de máquina como regularización, y en

este proyecto se ha podido constatar que los modelos de aprendizaje de máquina, si bien no superan el óptimo teórico, se acercan en una gran medida a los desempeños esperados de forma teórica y que son los ideales para los protocolos de comunicaciones.

4.6. Comparación de resultados

En esta sección se presenta una comparación de los resultados obtenidos por cada uno de los esquemas previamente presentados. Las métricas que se evalúan son el porcentaje de ranuras exitosas y el porcentaje de ranuras que presentan una colisión. Ambas métricas son consideradas las más importantes ya que el porcentaje de transmisiones exitosas refleja el uso del canal, conocido también como *throughput* y que representa el uso efectivo del canal, mientras que las colisiones se pueden relacionar con un mal manejo en consumo de energía, debido a que en esos eventos los nodos transmiten y gastan recursos sin poder compartir información útil a la red. En esta ocasión sólo se presentan los resultados desde las 6:00 a las 23:00 horas, ya que es en donde mayor cantidad de vehículos se encuentran en la zona de cobertura y es aquí en donde se desea ver con mayor claridad la diferencia en el desempeño de cada esquema. También es importante señalar que sólo se presentan los resultados de los esquemas óptimo (Óptimo), de valores medios (Valores medios), LSTM basado en transmisiones (Redes LSTM B/TX) y los del Transformador basado en las transmisiones recibidas (Transformador B/TX), ya que son los esquemas más factibles de ser implementados de forma directa. A pesar de que el esquema fijo también es factible de implementar, no se considera debido su inminente bajo desempeño observado previamente.

La Figura 4.9 muestra la comparación de la probabilidad de una ranura exitosa, en donde se observa que el esquema Transformador B/TX presenta el mejor desempeño en la mayoría de los casos, siendo únicamente de las 6:00 a las 7:00 horas y de las 18:00 a las 19:00 horas cuando es superado por el esquema Redes LSTM B/TX, sin embargo, en la Figura 4.10 se puede constatar que en el caso de las 18:00 a 19:00 horas la cantidad de colisiones se incrementa significativamente, además en esta última figura se aprecia que el esquema Transformador B/TX presenta una cantidad menor de colisiones a lo largo de todos los horarios, mientras que los demás modelos tienen un comportamiento oscilante con valores más altos y más bajos que los esperados por un esquema óptimo.

Por otra parte, se puede observar que el modelo utilizando valores medios es el que presenta un menor desempeño respecto al porcentaje de ranuras exitosas y es oscilante en el caso de las colisiones. Lo cual orilla a analizar el caso en que la cantidad de vehículos se aleja completamente de la media debido a una alteración en el tráfico, como puede ser la presencia de algún accidente, alguna calle cerrada, o algún bloqueo en las cercanías que congestione las vialidades. En tal caso, el valor medio puede moverse a los extremos de ser o muy bajo, o muy alto respecto al número actual de vehículos y consecuentemente causar una mayor cantidad de colisiones y ranuras libres que ranuras con transmisión exitosa.

Adicionalmente se valuó el porcentaje de similitud con el modelo óptimo, que es el que operaría de forma ideal con el valor real de N para tener un $\tau = \frac{1}{N}$ en donde se observa que el esquema basado en transformadores, en específico Transformador B/TX posee una similitud superior al 99.75 % para la mayoría de los horarios y sólo en el caso de las 22:00 a las 23:00 horas, la similitud baja al 99.68 %. En el caso del modelo basado en redes LSTM B/TX el

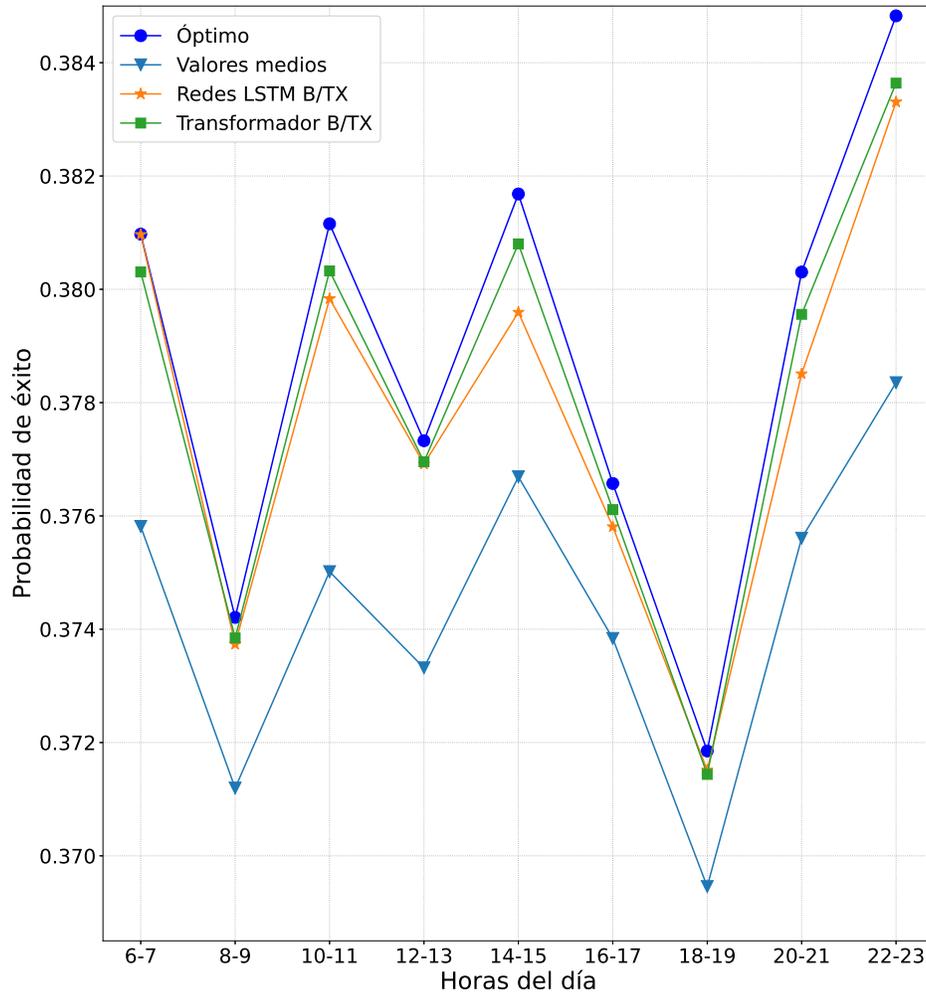


Figura 4.9: Comparación de resultados obtenidos por los diferentes esquemas para la probabilidad de ranura con transmisión exitosa.

grado de similitud se acerca al 100% para el caso de las 6:00 a las 7:00 horas, sin embargo para tres horas dentro del día, la similitud presentada es inferior a la mínima mostrada por los transformadores, y únicamente a las 18:00 horas, la similitud es ligeramente mayor a la mostrada por el modelo Transformador B/TX. Por otro lado el modelo con valores medios, presenta la menor similitud a lo largo de todas las horas de día, presentando como máximo un 99.35%. En la Figura 4.11 se presenta la gráfica de esta comparación.

Así mismo dentro del conjunto de comparaciones realizadas, en la gráfica de la Figura 4.12 se presentan los resultados observados desde la perspectiva del retardo de los paquetes, en este caso se muestra el retardo promedio en que un paquete dentro de los vehículos demora en ser enviado. En este caso, un valor más pequeño denota un mejor desempeño, siendo el esquema óptimo aquel que tiene el menor retardo a lo largo del día. El esquema basado en valores medios es el que peor desempeño presenta con retardos más altos que supera los $100\mu s$. En el caso de los esquemas basados en técnicas de aprendizaje de máquina se observa desempeños muy cercanos al óptimo, donde la mayor diferencia es para el esquema basado en LSTM con

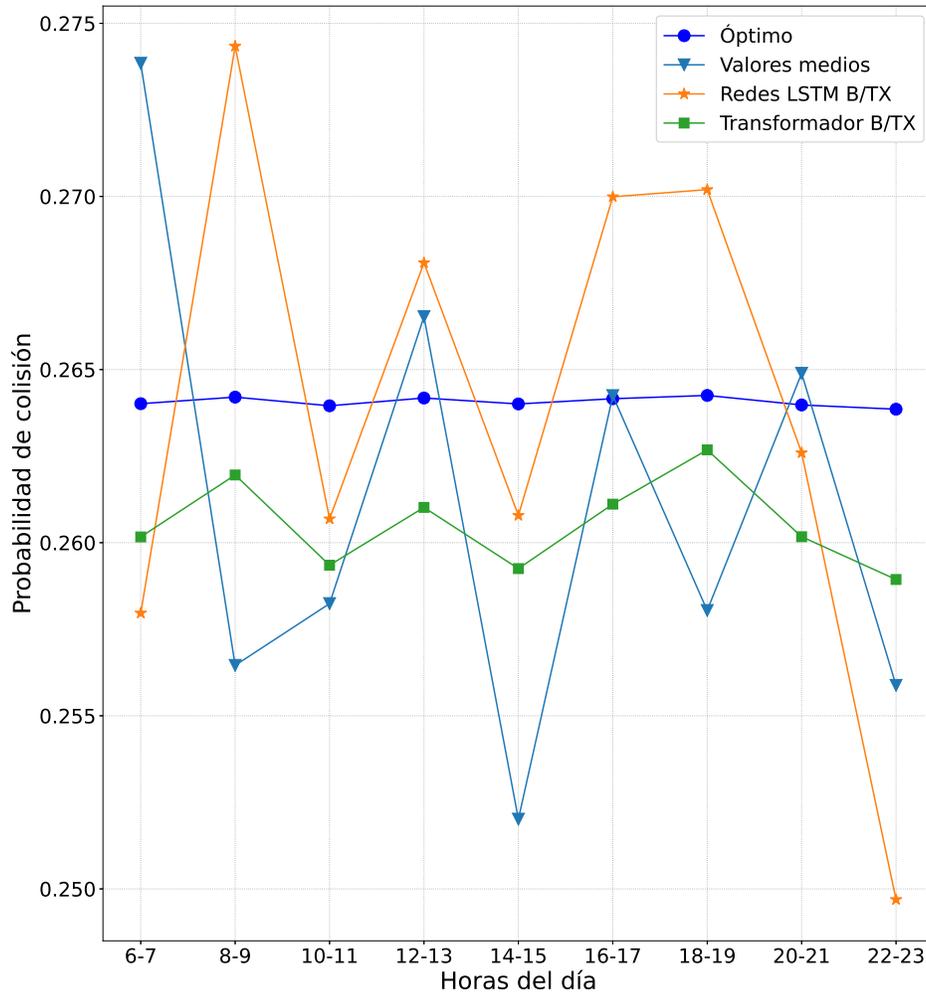


Figura 4.10: Comparación de resultados obtenidos por los diferentes esquemas para la probabilidad de ranura con colisiones.

una diferencia de $30\mu s$ de las 14:00 a las 15:00 horas, mientras que el Transformador B/TX la diferencia mayor es $10\mu s$ en un horario de 22:00 a 23:00 horas. Además se puede observar que los horarios que presentan una menor diferencia en el retardo es cuando hay un mayor flujo de tráfico debido a que los modelos tienden a ajustarse mejor cuando hay una alta tasa de vehículos en las zonas de cobertura de las estaciones base.

En la Figura 4.13 se presenta la comparación del retardo en que un paquete es generado y alcanza la estación base, en este caso se considera que los paquetes persisten dentro de la cola de mensajes a enviar en los nodos y se actualizan hasta que se realiza una transmisión exitosa. Claramente se observa que el retardo incrementa debido a que el tiempo que pasa desde que se genera el paquete de información hasta que el nodo transmita de forma correcta es mayor, dado el proceso aleatorio hace que no sea posible transmitir de forma inmediata. Por otra parte, en gráficas siguientes se observa el comportamiento respecto al porcentaje de paquetes perdidos, en donde se nota que a pesar de que el retardo es mayor, el porcentaje de pérdidas es, por mucho, significativamente menor que en el esquema en que los paquetes se

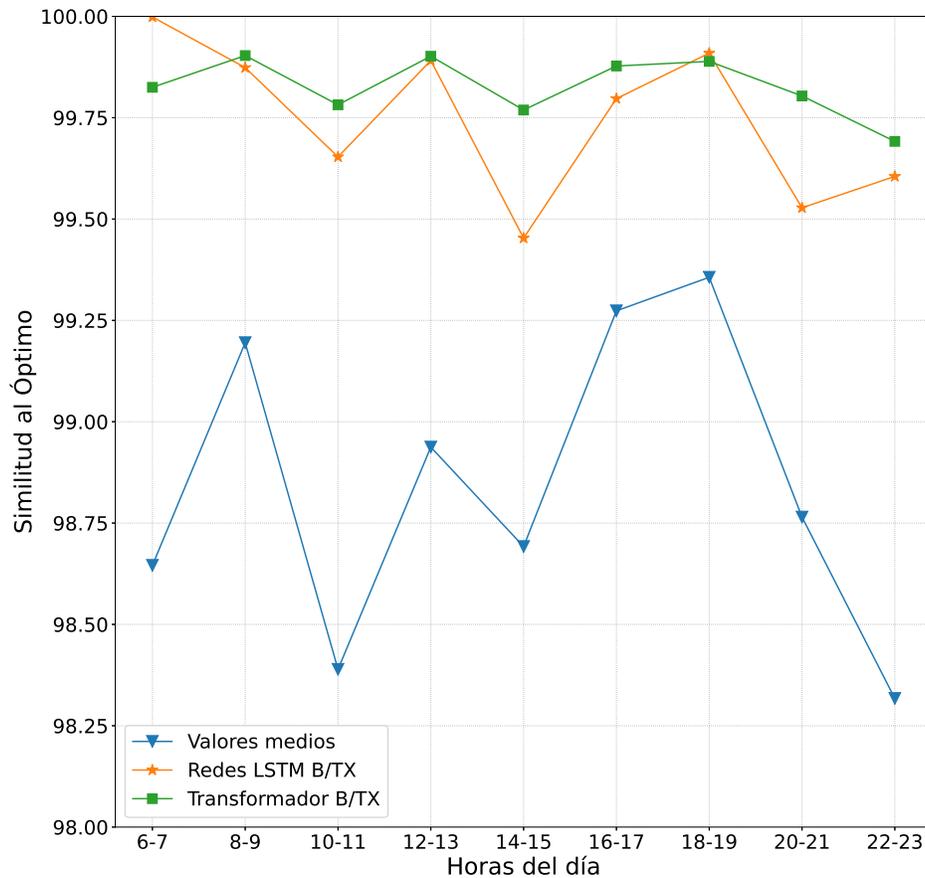


Figura 4.11: Comparación de resultados en similitud respecto a los valores óptimos.

actualizan en cada ranura.

Con los modelos basados en aprendizaje profundo, se espera que en general estén aprendiendo la dinámica del tráfico, de tal forma que sean capaces de adaptarse a las situaciones adversas y, aún así, brindar probabilidades de transmisión para los vehículos que maximicen el uso del canal. A través de las simulaciones, se generó un evento caótico que afecta las vialidades dentro del radio de cobertura de la estación base número 4 de la Zona 3, ubicado en el sur este de la ciudad. El conflicto se presenta a partir de las 20:00 y es donde se evalúan los esquemas con los entrenamientos hechos previamente, esto con la intención, además, de evaluar qué tan permisible es utilizar modelos previamente entrenados en diferentes zonas de la ciudad. En la Figura 4.14 se presenta la evolución del número de vehículos debido a la presencia de un conflicto vial en la que las avenidas de la zona mencionada previamente se saturan.

Donde claramente se observa que es un comportamiento atípico que causa la presencia de un mayor número de vehículos dentro de la zona, el efecto de esta situación es que los vehículos deben de intentar transmitir con una probabilidad menor a la que habitualmente utilizan. En el caso específico del esquema con valores medios, el valor de τ no se ajusta de forma correcta ya que al presentarse en menos ocasiones el conflicto, el valor de τ será considerado con un valor menor de vehículos que los que realmente hay en la zona de cobertura, dando

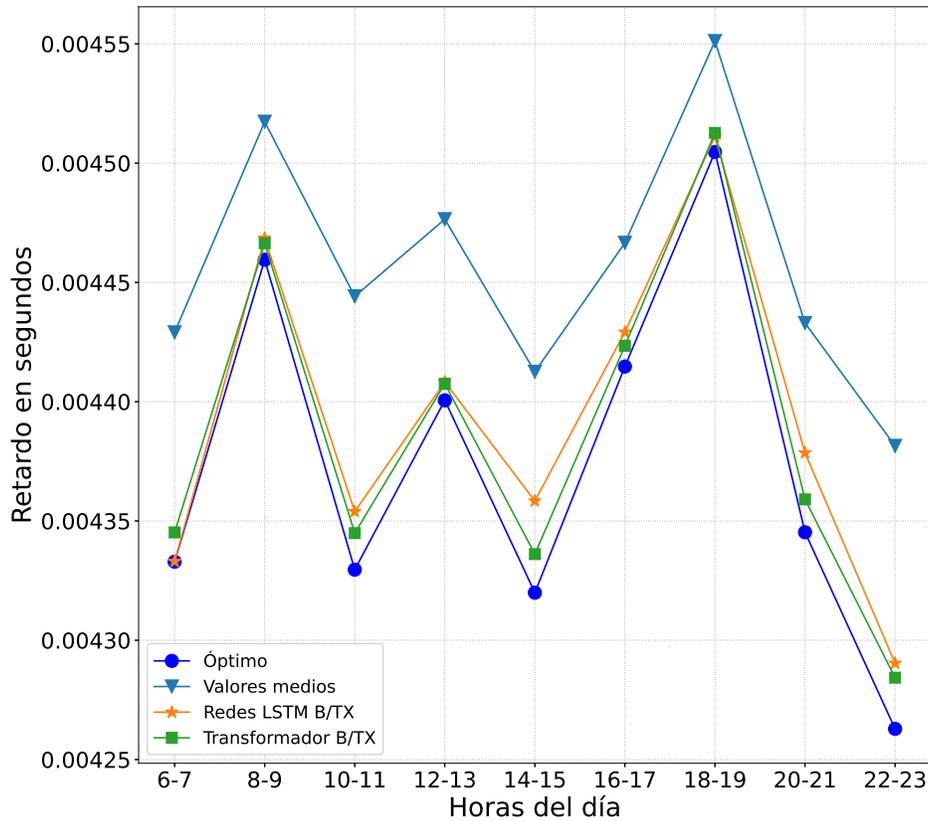


Figura 4.12: Comparación de resultados en el retardo de paquetes en esquema de actualización de paquetes por ranura.

como resultado un probabilidad de transmisión alta para los vehículos que causará un alto índice de ranuras colisionadas. En la Figura 4.15 se presentan las métricas de desempeño obtenidos por los esquemas Óptimo, con valores medios y Transformador B/TX, con la finalidad de mostrar la ventaja de utilizar el esquema basado en aprendizaje profundo para estimar el número de vehículos y con él otorgar un valor cercano al ideal para la probabilidad de transmisión. En las gráficas de dicha figura se observa que el esquema de aprendizaje profundo obtiene una probabilidad de ranuras exitosas cercanas al 40 %, mientras que el esquema con valores medios no supera el 23 % en esta métrica. Por otro lado, el número de colisiones en el esquema Transformador B/TX reduce el número de colisiones al rededor de 24 %, mientras que el esquema con valores medios presenta una probabilidad de colisión superior al 52 %.

Cabe señalar que la probabilidad de una transmisión exitosa, se observa ligeramente mayor que en los observados en situaciones donde no hay conflicto, esto se debe a que al principio del periodo de tiempo existe una baja cantidad de vehículos, lo que permite que se genere mayor probabilidad de éxito, además, la dinámica del sistema tiende a estabilizarse conforme las vialidades se saturan, por lo que las bajas variaciones del número de vehículos genera que la probabilidad tienda a ser constante y con ello lograr equilibrar las ranuras entre eventos de transmisión exitosa y eventos de ranuras libres. Con ello se corrobora que los esquemas basados en aprendizaje de máquina permite al sistema operar con desempeños

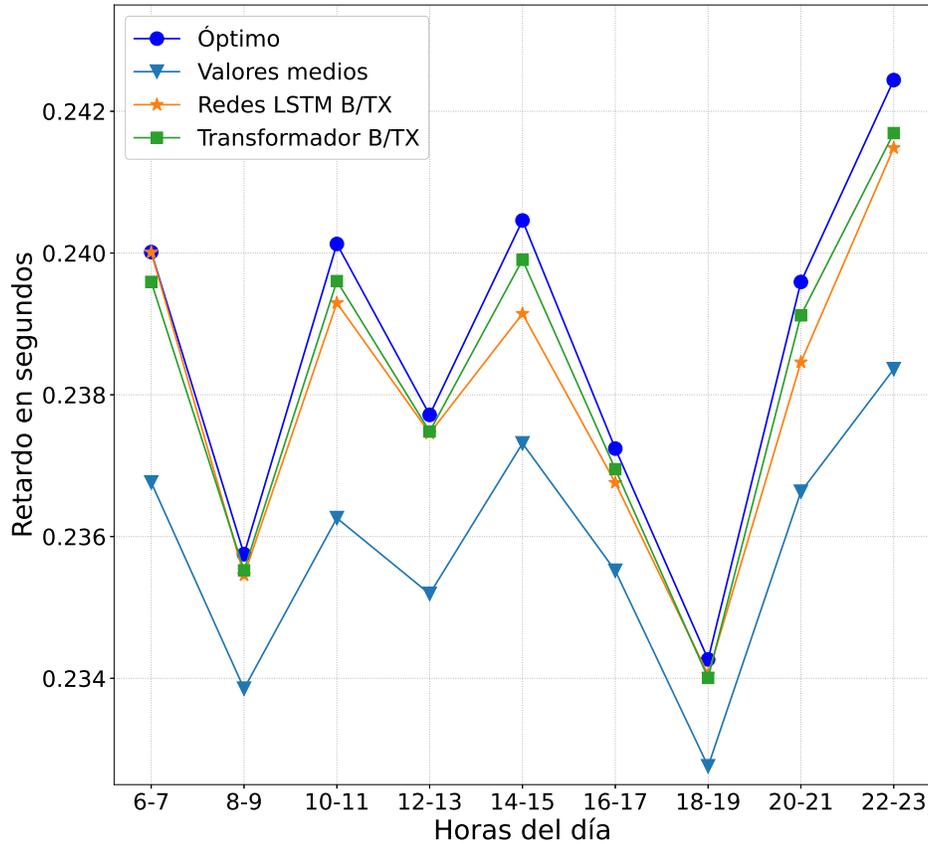


Figura 4.13: Comparación de resultados en el retardo de paquetes en esquema de actualización de paquetes por transmisión.

cercanos al óptimo, en particular el esquema basado en Transformadores ofrece la ventaja de obtener resultados aún cuando las transmisiones que se detectan en la estación base son el único medio para estimar el valor de τ , permitiendo que su implementación sea más factible directamente en la estación base con un buen desempeño en su operación.

Por último, es importante recordar que en este estudio se considera que los vehículos actúan de forma egoísta y ellos intentarán transmitir en cada ranura de tiempo, por lo que la cantidad de paquetes que generan a lo largo del día es equivalente a la cantidad de ranuras de tiempo que hay en el canal para cada uno de los automóviles. Esta situación causa que la relación de la cantidad de paquetes que se entregan respecto a la cantidad de paquetes generados en la red sea muy baja. Por ejemplo, en la Figura 4.16 se presenta el porcentaje de paquetes perdidos pensando que es actualizado el paquete en el buffer de salida y se pierde dentro del nodo a lo largo del día, es decir, de la cantidad de paquetes que se generan en la red por los vehículos a lo largo de cada hora completa de evaluación, y cuantos de ellos no se logran transmitir.

Por otra parte, en el caso de una política de generación y transmisión de paquetes diferente, el impacto en la cantidad de paquetes perdidos se reduce drásticamente, por ejemplo, si cada uno de los vehículos actualiza su información en un periodo de tiempo equivalente a la duración de una ranura de tiempo, la cantidad de paquetes perdidos sería a lo más la

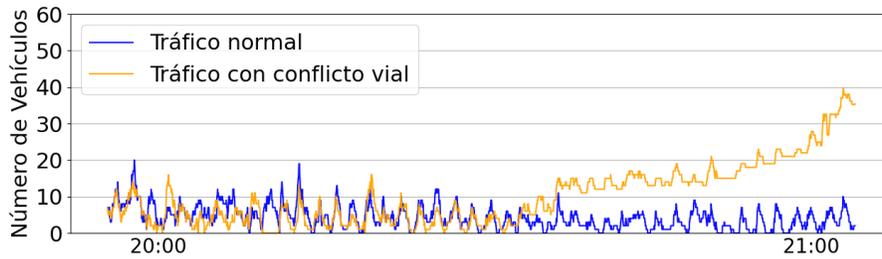


Figura 4.14: Comparación del número de vehículos dentro de la zona de cobertura en un día normal contra un día en el que ha ocurrido un conflicto vial.

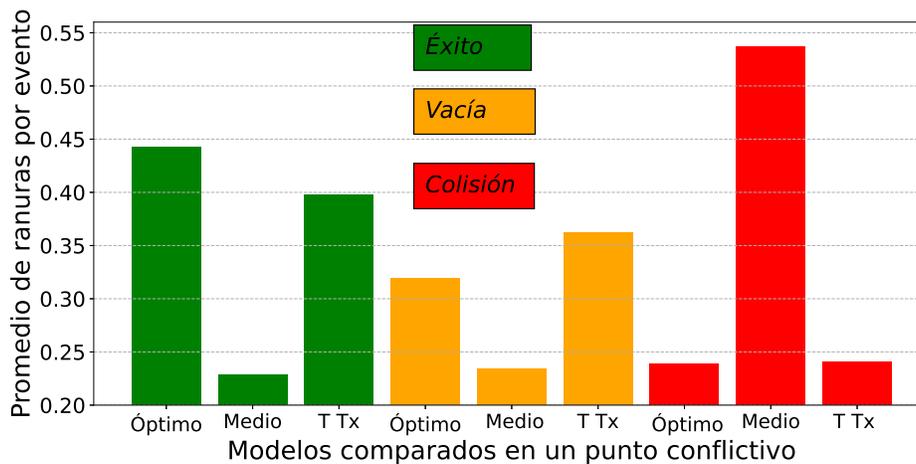


Figura 4.15: Métricas de desempeño en una zona de conflicto.

cantidad de vehículos en la zona, en el remoto caso de que ningún nodo logre transmitir. En otro escenario más realista, una política alternativa para que los vehículos compartan información es que estos actualicen su paquete de a enviar una vez que han transmitido el paquete previo, de esta forma, los nodos envían sus datos sin una pérdida tan alta de información. En la Figura 4.17 En tales casos, el uso del canal continuaría funcionando de la misma forma ya que, aún, todos los nodos dentro del área de cobertura, entrarán en proceso de contención para intentar transmitir en cada una de las ranuras de tiempo.

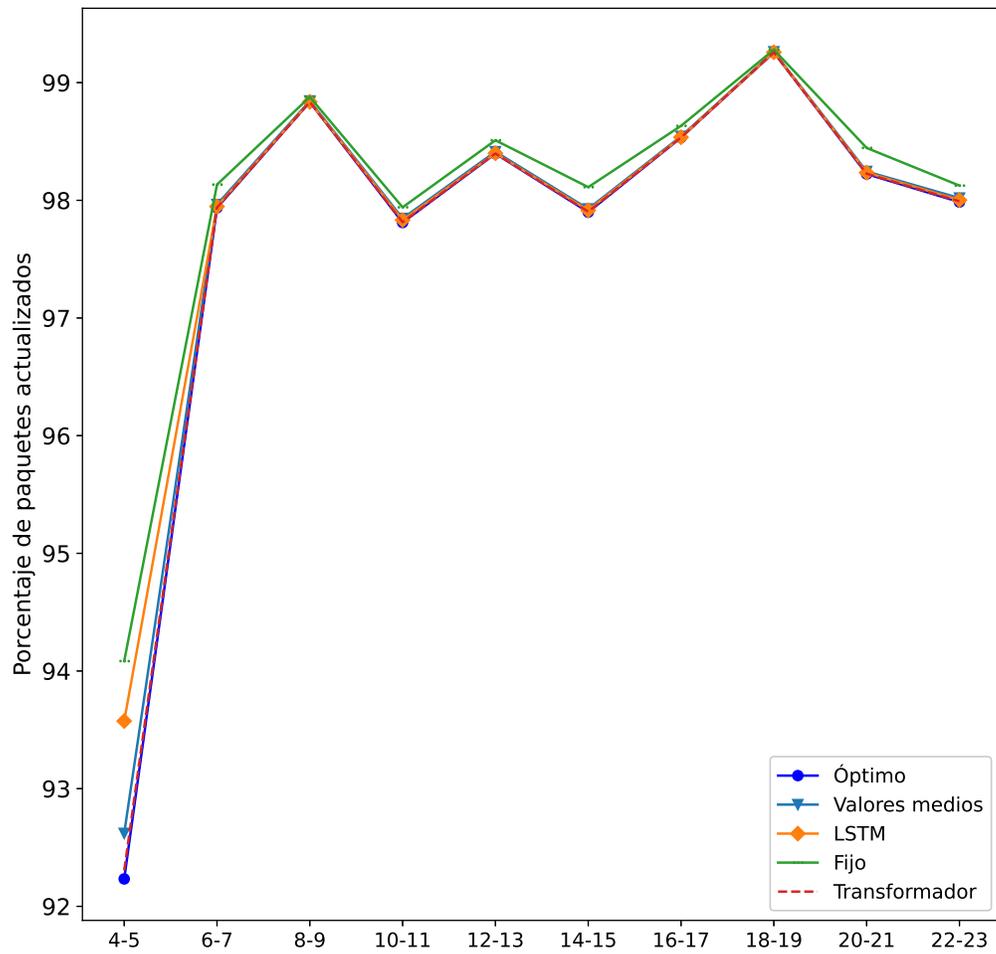


Figura 4.16: Estimación de paquetes actualizados en la zona de observación bajo una política de actualización por ranura.

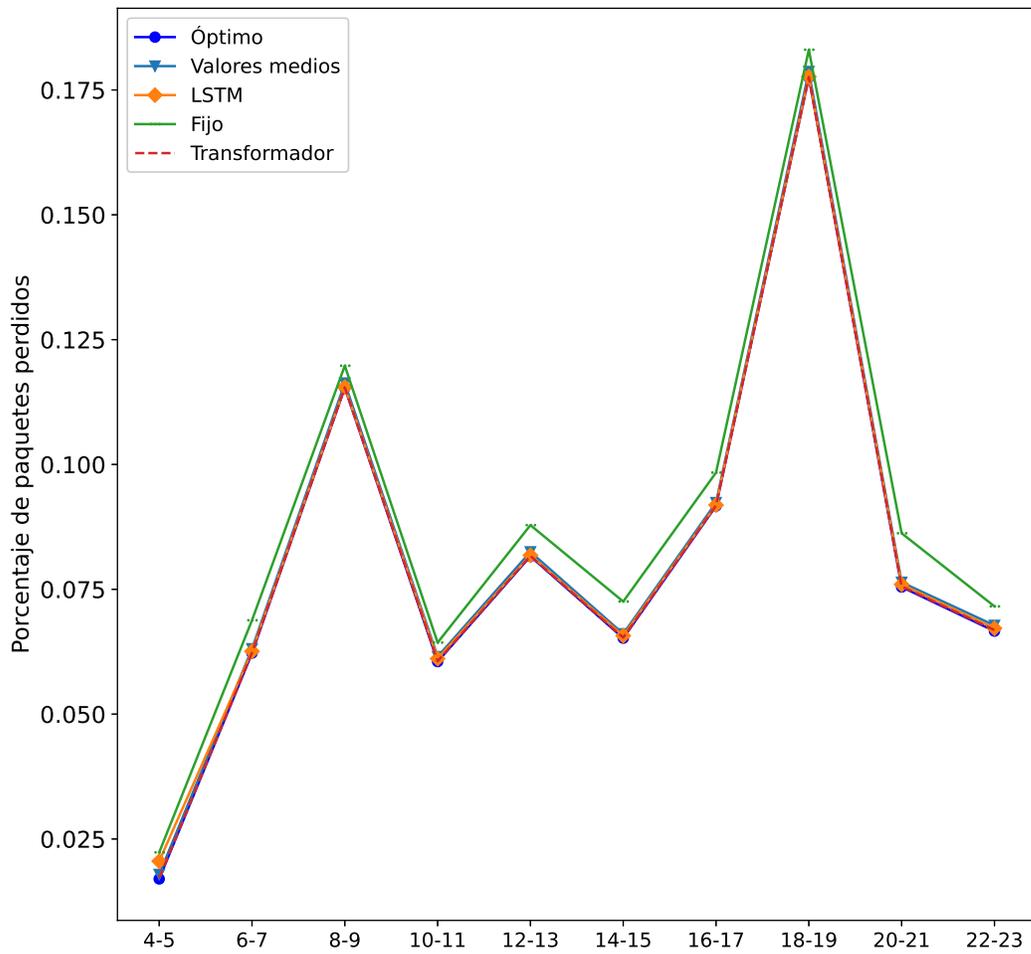


Figura 4.17: Estimación de paquetes actualizados en la zona de observación bajo una política de actualización por transmisión de paquete previo.

Capítulo 5

Conclusiones y trabajos a futuro

A lo largo de este documento se ha analizado y propuesto el uso de técnicas de aprendizaje de máquina en la selección de la probabilidad de transmisión para los nodos de una red inalámbrica compuesta por un conjunto de vehículos que viajan dentro de una determinada área. Dicha red es conocida como VANET, en la que los usuarios o nodos, son los vehículos que buscan conectarse a una estación base en una topología Vehículo a Infraestructura o comúnmente llamadas V2I.

Los vehículos acceden al sistema mediante un proceso de contención basado en el protocolo S-ALOHA para la transmisión de mensajes cortos en las distintas ranuras de tiempo en que el canal es segmentado. Dichos mensajes contienen, principalmente, la propia información telemétrica de cada vehículo dentro de un área de cobertura, y con la información de todos los automóviles se pueden generar estadísticas de movilidad para la transportación dentro de una ciudad inteligente. Adicionalmente, el uso de protocolos de contención continua vigente debido a que son predecesores y necesarios en la solicitud y la asignación de recursos de canal fijos en protocolos organizados para disponer de anchos de banda superiores

Es bien conocido que dentro de los protocolos de contención la utilización del canal se vuelve ineficiente ya que es susceptible a presentar colisiones y espacios libres, por lo que, con el uso de mecanismos de aprendizaje de máquina se ajusto el valor de la probabilidad de transmisión para incrementar el uso del canal. A través de los estudios observados se encontró que el tráfico vehicular presenta una dinámica particular que hace sensible a la red a los cambios en el número de vehículos dentro del área de cobertura y que la probabilidad de transmisión necesita ser ajustada dependiendo de cuantos nodos están presentes e intentan transmitir en cada instante de tiempo.

Debido a esta dinámica, es preciso contar con esquemas adaptativos a las condiciones del canal y los esquemas basados en aprendizaje de máquina obtuvieron los mejores desempeños al incrementar la cantidad de ranuras con transmisiones exitosas y reducir la cantidad de colisiones en la red. En particular los Transformadores presentan una alta precisión al momento de estimar el número de vehículos dentro de la red, y en el caso del esquema Transformador B/TX, la estimación es capaz de llevarse a cabo considerando únicamente la cantidad de vehículos detectados por medio de la transmisión de paquetes a la estación base. Este hecho da cabida en mayor medida a una implementación real. Por otro lado, se observó que la estimación dentro de eventos conflictivos incrementa significativamente el desempeño de la red con una mayor probabilidad de ranuras con transmisión exitosa, y que, modelos clásicos

como el de estimar la probabilidad de transmisión de los vehículos por los valores medios, obtienen un desempeño bastante pobre.

En este estudio se adoptó una política en la que los vehículos transmiten constantemente y renuevan su paquete de información a transmitir en cada una de las ranuras de tiempo, lo cual puede generar una idea de que hay un alto índice de paquetes perdidos. Sin embargo este comportamiento se presentará en los distintos esquemas, ya que la cantidad de paquetes generados es sumamente mayor.

Finalmente, en la literatura se observa que existen distintos protocolos utilizados en redes vehiculares que muestran desempeños superiores, sin embargo, en dichos casos se utiliza un sistema basado en protocolos organizados, los cuales generalmente ofrecen usos del canal cercanos al 100 %, normalmente un protocolo organizado en redes inalámbricas dinámicas requiere un mecanismo que asigne los canales fijos a los nodos, tal mecanismo se basa en protocolos de contención en los que los nodos envían solicitudes de conexión. Por lo tanto, se concluye que prestar atención a los protocolos de contención es de vital importancia para mejorar el desempeño general de la red y en este trabajo se ha comprobado que utilizar técnicas de aprendizaje de máquina permite ajustar parámetros críticos en redes inalámbricas con una alta dinámica como lo son las VANETS.

5.1. Trabajos futuros

Para continuar con este proyecto se presentan diferentes líneas de investigación para abordar con el fin de mejorar el desempeño de las redes vehiculares y hacerlas más robustas de acuerdo a las aplicaciones para las que sean desplegadas. En un principio se planteó este proyecto como un sistema de comunicación de acceso aleatorio en el cual los nodos puede compartir mensajes cortos para una aplicación de ciudades inteligentes en donde, de acuerdo con la información enviada por los automóviles, se generan, por ejemplo, reportes de contaminación o flujo vehicular. Sin embargo, los esquemas de contención también son utilizados para reservar canales dedicados a un usuario en particular.

Como trabajo a futuro en el corto plazo se propone estudiar el uso de técnicas de inteligencia artificial para la asignación de recursos de canal que no sean de contención y permitir a los usuarios la transmisión de información de forma continua. El reto observado en este estudio es que en un mecanismo de contención, cabe la posibilidad de que no todos los nodos logran hacer una transmisión exitosa en una trama de tiempo. Esto abre un enfoque en el que los modelos de aprendizaje profundo contribuyan a la elección de la cantidad de ranuras disponibles por trama de tiempo en el canal de contención y que apoyado con una probabilidad de transmisión óptima, se den garantías para que la mayoría e incluso todos los nodos dentro de un área de cobertura logren transmitir al menos en una ocasión en la trama.

Así mismo, un análisis empleando distintos protocolos de comunicación pueden ser considerados y estudiados, con la finalidad de crear sistemas más robustos y tolerantes a las posibles fallas dentro del sistema. Incluso saber si son igualmente compatibles con el uso de modelos de aprendizaje de máquina como en este estudio.

Derivado del desarrollo del trabajo de asignación dinámica para el número de ranuras, las métricas de desempeño recaen en la cantidad de vehículos que no logran transmitir, que puede ser interpretado como porcentaje de denegación de servicio; así como la cantidad de

solicitudes de recursos fijos por periodo de tiempo, retardo en la petición de mensajes, entre otras.

Un trabajo en el mediano plazo es el diseño de un sistema híbrido que considere el enfoque anterior y añada la estimación de recursos de canal necesarios para que los nodos que soliciten canales en un protocolo programado puedan tener acceso al momento de su solicitud. Mediante modelos de aprendizaje profundo, la estación base detecta la cantidad de vehículos, detecta el tipo de vehículo, y basado en la experiencia previa, define anchos de banda para canales de acceso programado a fin de brindar acceso a todas las futuras solicitudes, de tal forma que la cantidad y capacidad de los canales de acceso programado serán dinámicos en función de la demanda estimada. En este enfoque, las métricas principales serían la probabilidad de denegación de servicio, los anchos de banda promedio, así como el uso del canal, ya que aunque se empleen modelos programados, es susceptible a que determinados canales no sean utilizados debido a una estimación poco precisa.

También en el mediano plazo se plantea la implementación del sistema en módulos tipo Edge Computing para verificar su compatibilidad. A pesar de que hoy en día existen dispositivos lo suficientemente capaces para ejecutar modelos de aprendizaje de máquina, hay que recordar que estas operaciones pueden demandar una cantidad importante de recursos y esto puede mermar su factibilidad, haciendo necesario realizar nuevas consideraciones para los esquemas estudiados.

Como trabajo final y en un largo plazo, se propone el despliegue de la red, diseño de dispositivos necesarios para su implementación y una mesa de pruebas en donde se conjunte el trabajo. Cabe mencionar que en este trabajo es necesario la intervención de distintas disciplinas, como expertos electrónicos, desarrolladores de software entre otros. En esta fase de trabajo futuro el uso de inteligencia artificial se extiende al almacenaje, depuración y análisis de datos recabados por la estación base. Adicionalmente, en esta fase de trabajo, nuevas aplicaciones pueden ser consideradas y agregadas como trabajos a futuro.

Bibliografía

- [1] IFT, (01 febrero 2022), *INSTITUTO FEDERAL DE TELECOMUNICACIONES*, IFT, INSTITUTO FEDERAL DE TELECOMUNICACIONES <http://www.ift.org.mx>
- [2] Tecnología para los negocios, (01 febrero 2022), *Caminar con éxito hacia la Industria 4.0: Capítulo 11 – Infraestructuras (I) Redes Inalámbricas*, Portal de las tecnologías de Cámara Valencia <https://ticnegocios.camaravalencia.com/servicios/tendencias/caminar-con-exito-hacia-la-industria-4-0-capitulo-11-infraestructuras-i-redes-inalambricas/>
- [3] N. ABRAMSON, 1970, *THE ALOHA SYSTEM—Another alternative for computer communications*, Association for Computing Machinery. New York, NY, USA, DOI: [10.1145/1478462.1478502](https://doi.org/10.1145/1478462.1478502)
- [4] G. R. Lawrence, April 1975, *ALOHA packet system with and without slots and capture*, SIGCOMM Comput. Commun. Rev. 5, 2, pages 28–42. DOI:<https://doi.org/10.1145/1024916.1024920>
- [5] S.S. Shah, A. Malik, A. U. Rahman, S. Iqbal, S. U. Khan, (2019). *Time Barrier-Based Emergency Message Dissemination*, Vehicular Ad-hoc Networks, IEEE Access. pages 1-1.
- [6] W.S. McCulloch and W. Pitts, (1943), *A logical calculus of ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, Vol 5, pages 115 -133.
- [7] B. MEHLIG, (2021) *Machine learning with neural networks*, Department of Physics, UNIVERSITY OF GOTHENBURG.
- [8] A. N. Pérez-García, G. M. Tornez-Xavier, L. M. Flores-Nava, F. Gómez-Castañeda and J. A. Moreno-Cadenas, *Multilayer perceptron network with integrated training algorithm in FPGA*, (2014), 11th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 2014, pages 1-6.
- [9] Y. LeCun, Y. Bengio, G. Hinton, (2015) *Deep learning*, Nature, Vol 521, pages, 436–444, doi: <https://doi.org/10.1038/nature14539>
- [10] K.O’Shea, R. Nash, (2015), *An Introduction to Convolutional Neural Networks*, CoRR, Vol abs/1511.08458.

- [11] A Graves, (2013) *Generating Sequences With Recurrent Neural Networks*, CoRR, Vol abs/1308.0850.
- [12] S. Hochreiter, J. Schmidhuber, (15 Nov. 1997) *Long Short-Term Memory*, Neural Computation, vol. 9, no. 8, pages. 1735-1780.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, (2017), *Attention Is All You Need*, arXiv.
- [14] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M. Uusitalo, M. Schellmann, (2013). *The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions*, pages 1-5.
- [15] R. Hussain, F. Hussain, S Zeadally, (2019), *Integration of VANET and 5G Security: A review of design and implementation issues*, Future Generation Computer Systems, vol. 101, pages 843-864
- [16] A. Rahim, P.K. Malik, V.A. Sankar Ponnappalli, (2020), *State of the Art: A Review on Vehicular Communications, Impact of 5G, Fractal Antennas for Future Communication*, Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)
- [17] M. N. Tahir, M. Katz, U. Rashid, (2021), *Analysis of VANET Wireless Networking Technologies in Realistic Environments*, IEEE Radio and Wireless Symposium (RWS), pages. 123-125
- [18] S. Niknam, H. S. Dhillon, J. H. Reed, (June 2020) *Federated Learning for Wireless Communications: Motivation, Opportunities and Challenges*, IEEE Communications Magazine, vol. 58, No. 6, pages 46-51
- [19] M. Zhang, L. de Alfaro, J.J. Garcia-Luna-Aceves, (2020) *Using Reinforcement Learning in Slotted Aloha for Ad-Hoc Networks*, Association for Computing Machinery, New York, NY, USA, Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pages 245–252, MSWiM '20.
- [20] L. de Alfaro, M. Zhang, J. J. Garcia-Luna-Aceves, (2020), *Approaching Fair Collision-Free Channel Access with Slotted ALOHA Using Collaborative Policy-Based Reinforcement Learning*, 2020 IFIP Networking Conference (Networking), pages 262-270.
- [21] Y. Chu, S. Kosunalp, P. D. Mitchell, D. Grace, T. Clarke, (2015), *Application of reinforcement learning to medium access control for wireless sensor networks*, Engineering Applications of Artificial Intelligence, Vol. 46, Part A, pages 23-32.
- [22] L. Liang, H. Ye, G. Y. Li, (Feb. 2019) *Toward Intelligent Vehicular Networks: A Machine Learning Framework*, IEEE Internet of Things Journal, Vol. 6, no. 1, pages. 124-135.
- [23] A. Boukerche, Y. Tao, P. Sun, (2020) *Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems*, Computer Networks, Vol. 182.

- [24] B. Sliwa, R. Falkenberg, T. Liebig, N. Piatkowski, C. Wietfeld, (Aug. 2020), *Boosting Vehicle-to-Cloud Communication by Machine Learning-Enabled Context Prediction*, IEEE Transactions on Intelligent Transportation Systems, Vol. 21, no. 8, pages. 3497-3512.
- [25] D. Ma, X. Song, P. Li, (May 2021), *Daily Traffic Flow Forecasting Through a Contextual Convolutional Recurrent Neural Network Modeling Inter- and Intra-Day Traffic Patterns*, IEEE Transactions on Intelligent Transportation Systems, Vol. 22, no. 5, pages. 2627-2636.
- [26] X. Zhang, C. Huang, Y. Xu, L. Xia, P. Dai, L. Bo, J. Zhang, Y. Zheng, (2021), *Traffic Flow Forecasting with Spatial-Temporal Graph Diffusion Network*, arXiv.
- [27] S. Sun, H. Wu, L. Xiang, (2020), *City-Wide Traffic Flow Forecasting Using a Deep Convolutional Neural Network*, Sensors, 20(2), page 421.
- [28] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi and B. Yin, (Feb. 2021) *Deep Learning on Traffic Prediction: Methods, Analysis and Future Directions*, IEEE Transactions on Intelligent Transportation Systems, pages 1-17.
- [29] X. Luo , D. Li, Y. Yang, S. Zhang, (2019), *Spatiotemporal Traffic Flow Prediction with KNN and LSTM*, Journal of Advanced Transportation, Vol. 2019.
- [30] A. Boukerche, J. Wang, (2020), *Machine Learning-based traffic prediction models for Intelligent Transportation Systems*, Computer Networks, Vol 181.
- [31] H. Menouar, F. Filali and M. Lenardi, (October 2006), *A survey and qualitative analysis of mac protocols for vehicular ad hoc networks*, IEEE Wireless Communications, Vol. 13, no. 5, pages. 30-35,
- [32] F. Cunha, L. Villas, A. Boukerche, G. Maia, A. Viana, R.A.F. Mini, A.A.F. Loureiro, (2016) *Data communication in VANETs: Protocols, applications and challenges*, Ad Hoc Networks, Vol. 44, pages 90-103
- [33] F. Borgonovo, A. Capone, M. Cesana, L. Fratta, (2002), *RR-ALOHA, a reliable R-ALOHA broadcast channel for ad-hoc inter-vehicle communication networks*, Proceedings of Med-Hoc-Net 2002, Baia Chia, Italy
- [34] F. Borgonovo, A. Capone, M. Cesana, L. Fratta, (2004), *ADHOC MAC: New MAC Architecture for Ad Hoc Networks Providing Efficient and Reliable Point-to-Point and Broadcast Services*, Wireless Networks Vol. 10, pages 359–366
- [35] L. Zhang, Z. Liu, R. Zou, J. Guo, L. Yu, (2014), *A Scalable CSMA and Self-Organizing TDMA MAC for IEEE 802.11 p/1609.x in VANETs*, Wireless Personal Communications, Vol. 74, pages 1197–1212
- [36] J. Zhang, Q. Zhang, W. Jia, (2009) *VC-MAC: A Cooperative MAC Protocol in Vehicular Networks*, IEEE Transactions on Vehicular Technology, Vol. 58, No. 3, pages. 1561-1571

- [37] Iris, (24 mayo 2021), *TP-Link USA Outdoor Solution Guide*, TP-Link, <https://www.tp-link.com/us/blog/137/tp-link-usa-outdoor-solution-guide/>
- [38] Wilson Amplifiers, (22 junio 2021), *GUIDE TO SMALL CELLS: FEMTOCELL, PICOCELL, AND MICROCELLS (2021)*, Wilson Amplifiers Premium Wilson Electronics Partner, <https://www.wilsonamplifiers.com/blog/guide-to-small-cells-femtocell-picocell-and-microcells/>
- [39] S. Adireddy, L. Tong, (2002), *Optimal Transmission Probabilities for Slotted ALOHA in Fading Channels*, Proc. CISS'02.
- [40] L. Codeca, R. Frank, T. Engel, (2015), *Luxembourg SUMO Traffic (LuST) Scenario: 24 hours of mobility for vehicular networking research*, 2015 IEEE Vehicular Networking Conference (VNC), pages. 1-8.
- [41] J. Kleinberg, E. Tardos, (2005), *Algorithm Design* Addison-Wesley Longman Publishing Co., Inc, USA.
- [42] E. Romo-Montiel, M. E. Rivero-Ángeles, H. Molina-Lozano, G. Rubino, R. Menchaca-Méndez, R. Menchaca-Méndez, (2018), *Optimal Transmission Thresholds for QoS Guarantees in WSNs*, Proceedings of the 14th ACM International Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet'18), Association for Computing Machinery, New York, NY, USA, pages 108–115.