

ANÁLISIS DEL IMPACTO DEL INTERVALO DE TIEMPO DE TRANSMISIÓN SOBRE LA LATENCIA EN LA RED DE ACCESO RADIO DE SISTEMAS 5G



Jesús Andrés Castro Delgado

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Grupo de Radio e Inalámbricas (GRIAL)
Sistemas de Comunicaciones Móviles e Inalámbricos
Popayán-Cauca
2019

ANÁLISIS DEL IMPACTO DEL INTERVALO DE TIEMPO DE TRANSMISIÓN SOBRE LA LATENCIA EN LA RED DE ACCESO RADIO DE SISTEMAS 5G



Trabajo de grado presentado como requisito para obtener el título de Ingeniero en
Electrónica y Telecomunicaciones

Jesús Andrés Castro Delgado

Director:

Víctor Manuel Quintero Flórez

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Grupo de Radio e Inalámbricas (GRIAL)
Sistemas de Comunicaciones Móviles e Inalámbricos
Popayán-Cauca
2019

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN	1
CAPITULO 1: GENERALIDADES	3
1.1. IMT-2020	3
1.2. TECNOLOGÍA DE ACCESO RADIO 5G	3
1.2.1. Espectro para Sistemas 5G	5
1.2.2. Arquitectura de Protocolos Radio	6
1.2.3. Modulación	8
1.2.4. Forma de Onda	8
1.2.5. Múltiples Antenas	10
1.2.6. Codificación de Canal	11
1.2.7. Esquema de Duplexación	11
1.2.8. Bandas de Operación	11
1.2.9. Estructura de Transmisión en el Dominio del Tiempo	13
1.2.10. Estructura de Transmisión en el Dominio de la Frecuencia	14
1.3. CONCEPTO DE LATENCIA	15
1.4. LATENCIA EN REDES CELULARES	16
1.5. SERVICIOS DE BAJA LATENCIA EN 5G	18
CAPÍTULO 2: DESARROLLO METODOLÓGICO	21
2.1. ANÁLISIS DE REQUERIMIENTOS	21
2.1.1. Descripción del Problema	21
2.1.2. Requerimientos Funcionales	22
2.1.3. Requerimientos no Funcionales	22
2.1.4. Criterios de Selección de la Herramienta de Simulación	23
2.1.5. Herramienta de Simulación ns-3	24
2.2. DISEÑO	26
2.2.1. Componentes de Latencia en el Nivel Físico	26
2.2.2. Modelo de Simulación	28
2.2.3. Escenario de Simulación	28
2.3.1. Conceptos Principales en ns3	31
2.3.2. Generación del Código en ns3	31

2.4. PLAN DE PRUEBAS	33
2.4.1. Diseño de Casos de Estudio	34
2.4.2. Variables de entrada y salida.....	36
2.4.3. Ejecución de las simulaciones	37
CAPITULO 3: ANÁLISIS DE RESULTADOS.....	41
3.1. CASO DE ESTUDIO 1.....	41
3.1.1. Variante 1.....	41
3.1.2. Variante 2.....	46
3.2. CASO DE ESTUDIO 2.....	50
3.3. CASO DE ESTUDIO 3.....	54
3.4. CASO DE ESTUDIO 4.....	55
3.5. COMPARACIÓN DE RESULTADOS	58
CAPÍTULO 4: CONCLUSIONES Y TRABAJOS FUTUROS	63
4.1. INTRODUCCIÓN	63
4.2. CONCLUSIONES.....	63
4.3. TRABAJOS FUTUROS.....	65
REFERENCIAS BIBLIOGRÁFICAS.....	66

LISTA DE FIGURAS

Figura 1.1.	Banda de operación de sistemas 5G..	4
Figura 1.2.	Proceso de estandarización de sistemas 5G..	5
Figura 1.3.	Pila de protocolos radio en el plano de usuario.....	6
Figura 1.4.	OFDM y separación entre subportadoras..	9
Figura 1.5.	Prefijo cíclico en CP-OFDM.	9
Figura 1.6.	Numerología para NR..	10
Figura 1.7.	Enlace de Subida Suplementario.	11
Figura 1.8.	Estructura de trama en NR.....	13
Figura 1.9.	Mini-TTI de 7, 4 y 2 símbolos.	14
Figura 1.10.	Elemento de Recurso en NR.....	15
Figura 1.11.	RTT y latencia E2E..	16
Figura 1.12.	Arquitectura de sistema 5G.....	16
Figura 2.1.	Diagrama de metodología RUP.....	211
Figura 2.2.	Ejemplo de implementación de trama mmWave..	254
Figura 2.3.	Proceso de transmisión en el nivel físico..	276
Figura 2.4.	Modelo de simulación.....	298
Figura 2.5.	Escenario de simulación..	298
Figura 2.6.	Variables de entrada y salida.	375
Figura 3.1.	Latencia en la RAN en el UL para diferentes numerologías.....	42
Figura 3.2.	Latencia en la RAN en el DL para diferentes numerologías.....	42
Figura 3.3.	Tiempo de espera de asignación de TTI.	43
Figura 3.4.	Retardo de cola debido a la carga del sistema..	44
Figura 3.5.	Tiempo de espera de asignación de TTI.	45
Figura 3.6.	Retardo de alineación... ..	45
Figura 3.7.	Retardo de transmisión.	45
Figura 3.8.	Latencia en la RAN en el UL para 14, 7, 4 y 2 símbolos por TTI..	477
Figura 3.9.	Latencia en la RAN en el DL para 14, 7, 4 y 2 símbolos por TTI..	477
Figura 3.10.	Ubicación de la señal DMRS en TTI de 7 y 2 símbolos..	488
Figura 3.11.	Latencia en la RAN en el UL para 400, 900 y 1400 segmentos por segundo.	50
Figura 3.12.	Latencia en la RAN en el DL para 400, 900 y 1400 segmentos por segundo.	50
Figura 3.13.	Proceso de Transmisión en el UL.	52
Figura 3.14.	Envío de mensajes SR según contenido del buffer de transmisión. .	52
Figura 3.15.	Latencia en la RAN en el UL para segmentos de 50 Bytes y 1500 Bytes.	54
Figura 3.16.	Latencia en la RAN en el DL para segmentos de 50 Bytes y 1500 Bytes.	54
Figura 3.17.	Asignación de TTI según el tamaño de segmentos.	55
Figura 3.18.	Latencia en la RAN en el UL para diferentes valores de tiempo	

de procesamiento.	56
Figura 3.19. Latencia en la RAN en el DL para diferentes valores de tiempo de procesamiento.	56
Figura 3.20. Latencia en la RAN y de extremo a extremo para diferentes numerologías.	599
Figura 3.21. Latencia en la RAN para diferentes tamaños de segmentos.	60
Figura 3.22. Latencia en la RAN para diferentes valores de símbolos por TTI. ...	61

LISTA DE TABLAS

Tabla 1.1.	Bandas de operación para FR1.....	12
Tabla 1.2.	Bandas de operación para FR2.....	13
Tabla 1.3.	Duración de símbolo y CP en NR.....	14
Tabla 2.1.	Valores de ejemplo para la trama mmWave.....	25
Tabla 2.2.	Parámetros del escenario de simulación.....	29
Tabla 2.3.	Parámetros del caso de estudio 1.. ..	33
Tabla 2.4.	Parámetros del caso de estudio 2.. ..	34
Tabla 2.5.	Parámetros del caso de estudio 3.. ..	34
Tabla 2.6.	Parámetros del caso de estudio 4.. ..	34
Tabla 2.7.	Desviación estándar de resultados de casos de estudio en el UL.....	37
Tabla 2.8.	Desviación estándar de resultados de casos de estudio en el DL.....	38
Tabla 2.9.	Número de ejecuciones para cada caso de estudio en el UL.....	38
Tabla 2.10.	Número de ejecuciones para cada caso de estudio en el DL.....	39
Tabla 3.1.	Reducción de la latencia en la RAN en el UL para diferentes numerologías.	466
Tabla 3.2.	Reducción de la latencia en la RAN en el DL para diferentes numerologías.	466
Tabla 3.3.	Reducción de la latencia en la RAN en el UL para diferentes valores de símbolos por TTI.	499
Tabla 3.4.	Reducción de la latencia en la RAN en el UL para diferentes valores de símbolos por TTI.	499
Tabla 3.5.	Reducción de la latencia en la RAN en el UL para diferentes niveles de tasa de segmentos.. ..	533
Tabla 3.6.	Reducción de la latencia en la RAN en el DL para diferentes niveles de tasa de segmentos.. ..	533
Tabla 3.7.	Reducción de la latencia en la RAN en el UL para segmentos pequeños y grandes.. ..	555
Tabla 3.8.	Reducción de la latencia en la RAN en el UL para segmentos pequeños y grandes.. ..	555
Tabla 3.9.	Reducción de la latencia en la RAN en el UL para diferentes valores de tiempo de procesamiento.....	577
Tabla 3.10.	Reducción de la latencia en la RAN en el UL para diferentes valores de tiempo de procesamiento.....	577
Tabla 3.11.	Resumen del impacto del TTI en el UL.	588

LISTA DE ACRÓNIMOS

16 QAM	<i>16 Quadrature Amplitude Modulation</i> , Modulación de Amplitud <u>en</u> Cuadratura de 16 estados.
1G	<i>First Generation</i> , Primera Generación.
256 QAM	<i>256 Quadrature Amplitude Modulation</i> , Modulación de Amplitud <u>en</u> Cuadratura de 256 estados.
3GPP	<i>3rd Generation Partnership Project</i> , Proyecto de Cooperación para Sistemas de Tercera Generación.
5G	<i>Fifth Generation</i> , Quinta Generación.
64 QAM	<i>64 Quadrature Amplitude Modulation</i> , Modulación de Amplitud <u>en</u> Cuadratura de 64 estados.
AR	<i>Augmented Reality</i> , Realidad Aumentada.
ARQ	<i>Automatic Repeat Request</i> , Solicitud de Retransmisión Automática.
BPSK	<i>Binary Phase Shift Keying</i> , Modulación por Desplazamiento de Fase Binaria.
BS	<i>Base Station</i> , Estación Base.
BWP	<i>Bandwidth Part</i> , Parte de Ancho de Banda.
CA	<i>Carrier Aggregation</i> , Agregación de Portadoras.
CN	<i>Core Network</i> , Núcleo de Red.
CP	<i>Cyclic Prefix</i> , Prefijo Cíclico.
CP	<i>Control Plane</i> , Plano de Control.
CP-OFDM	<i>Cyclic Prefix Orthogonal Frequency-Division Multiplexing</i> , Multiplexación por División de Frecuencias Ortogonales con Prefijo Cíclico.
D2D	<i>Device to Device</i> , Dispositivo a Dispositivo.
DFTS-OFDM	<i>Discrete Fourier Transform Spread Orthogonal Frequency-Division Multiplexing</i> , Multiplexación por División de Frecuencias Ortogonales con Dispersión por Transformada de Fourier Discreta.
DL	<i>Downlink</i> , Enlace de Bajada.
DMRS	<i>Demodulation Reference Signal</i> , Señal de Referencia para Demodulación.
DS	<i>Delay Spread</i> , Dispersión de Retardo.
E2E	<i>End to End</i> , Extremo a Extremo.
FDD	<i>Frequency Division Duplexing</i> , Duplexación por División de Frecuencia.
FR1	<i>Frequency Range 1</i> , Rango de Frecuencia 1.
FR2	<i>Frequency Range 2</i> , Rango de Frecuencia 2.
gNB	<i>Next Generation Node B</i> , Nodo B de Próxima Generación.
GPL	<i>General Public License</i> , Licencia Pública General.
HARQ	<i>Hybrid Automatic Repeat Request</i> , Solicitud de Retransmisión Automática Híbrida.
IMT-2020	<i>International Mobile Telecommunications for 2020 and beyond</i> , Telecomunicaciones Móviles Internacionales para el 2020 y futuro.

IoT	<i>Internet of Things</i> , Internet de las Cosas.
IP	<i>Internet Protocol</i> , Protocolo de Internet.
ISI	<i>Intersymbol Interference</i> , Interferencia Intersímbolo.
ITU	<i>International Telecommunications Union</i> , Unión Internacional de Telecomunicaciones.
ITU-R	<i>ITU Radio Section</i> , Sección Radiocomunicaciones de la ITU.
LDPC	<i>Low Density Parity Check</i> , Verificación de Paridad de Baja Densidad.
LOS	<i>Line of Sight</i> , Línea de Vista.
LTE	<i>Long Term Evolution</i> , Evolución a Largo Plazo.
MAC	<i>Medium Access Control</i> , Control de Acceso al Medio.
MBB	<i>Mobile Broadband</i> , Banda Ancha Móvil.
MME	<i>Mobility Management Entity</i> , Entidad de Gestión de Movilidad.
NR	<i>New Radio</i> , Nueva Radio.
OFDM	<i>Orthogonal Frequency-Division Multiplexing</i> , Multiplexación por División de Frecuencias Ortogonales.
OSI	<i>Open System Interconnection</i> , Interconexión de Sistemas Abiertos.
PDCP	<i>Packet Data Convergence Protocol</i> , Protocolo de Convergencia de Paquetes de Datos.
PDU	<i>Protocol Data Unit</i> , Unidad de Datos de Protocolo.
PGW	<i>Packet Data Network Gateway</i> , Pasarela de Red de Paquetes de Datos
PHY	<i>Physical Layer</i> , Nivel Físico.
QoS	<i>Quality of Service</i> , Calidad del Servicio.
QPSK	<i>Quadrature Phase Shift Keying</i> , Desplazamiento de Fase en Cuadratura.
R15	<i>Release 15</i> , Especificación 15.
R16	<i>Release 16</i> , Especificación 16.
RAN	<i>Radio Access Network</i> , Red de Acceso Radio.
RLC	<i>Radio Link Control</i> , Control del Enlace Radio.
RTT	<i>Round Trip Time</i> , Tiempo de Ida y Vuelta.
SDAP	<i>Service Data Adaptation Protocol</i> , Protocolo de Adaptación de Servicio de Datos.
SDL	<i>Supplementary Downlink</i> , Enlace de Bajada Suplementario.
SGW	<i>Serving Gateway</i> , Pasarela de Servicio.
SR	<i>Scheduling Request</i> , Solicitud de Recursos.
SUL	<i>Supplementary Uplink</i> , Enlace de Subida Suplementario.
TDD	<i>Time Division Duplexing</i> , Duplexación por División de Tiempo.
TIC	Tecnologías de Información y Comunicación.
TTI	<i>Time Transmission Interval</i> , Intervalo de Tiempo de Transmisión.
UE	<i>User Equipment</i> , Terminal de Usuario.
UG	<i>Uplink Grant</i> , Mensaje de Concesión.
UL	<i>Uplink</i> , Enlace de Subida.
UP	<i>User Plane</i> , Plano de Usuario.

URLLC *Ultra-Reliable Low-Latency Communication*, Comunicación Ultra
Confiable de Baja Latencia.

INTRODUCCIÓN

Las Tecnologías de la Información y Comunicación (TIC) han tenido un desarrollo vertiginoso en las últimas décadas. La capacidad de transmitir y procesar información instantáneamente ha permitido el despliegue de diversas aplicaciones: comercio electrónico, interacciones sociales, aprendizaje virtual, acceso instantáneo a la información, comunicación multimedia, entre otras. Así mismo, muchas industrias se han basado en las TIC para potenciar sus procesos de producción y comercialización [1].

Los sistemas de comunicaciones móviles de Quinta Generación (5G, *Fifth Generation*) ampliarán significativamente las capacidades de las redes móviles actuales. Nuevas tecnologías, funcionalidades, y procedimientos están siendo introducidas para los sistemas 5G, cuyo objetivo es soportar nuevos servicios y aplicaciones para consumidores e industrias que revolucionarán las comunicaciones del futuro [2].

Los sistemas 5G soportarán servicios de Comunicación Ultra Confiable de Baja Latencia (URLLC, *Ultra-Reliable Low-Latency Communication*), que incluyen aplicaciones con requerimientos exigentes en términos de disponibilidad, latencia y confiabilidad como son: Internet táctil, reproducción de video de alta definición, tele-medicina, tele-cirugía, transporte inteligente, conducción autónoma, entre otras [3].

El punto crítico de cualquier sistema de comunicación móvil es la tecnología de acceso radio o inalámbrica, la cual conecta Terminales de Usuario (UE, *User Equipment*) con Estaciones Base (BS, *Base Station*). Se espera que la Red de Acceso Radio (RAN, *Radio Access Network*) 5G provea latencias cercanas a cero. Para lograr esto, la Unión Internacional de Telecomunicaciones (ITU, *International Telecommunications Union*) ha decidido reestructurar la trama de datos de nivel físico para soportar diferentes escenarios de aplicación. Este cambio permitirá adaptar el Intervalo de Tiempo de Transmisión (TTI, *Time Transmission Interval*) según los requerimientos de los servicios, haciendo un mejor uso de los recursos

radio y por consecuencia reduciendo la latencia [4]. El presente trabajo de grado estudia los diferentes parámetros de la estructura de trama de datos que definen el TTI y su impacto sobre la latencia en la RAN.

Este documento se divide en 4 capítulos, cada uno organizado de la siguiente manera: en el primer capítulo, se estudia la tecnología RAN de los sistemas 5G haciendo énfasis en la estructura de trama de nivel físico; el segundo capítulo presenta el desarrollo metodológico para el diseño y obtención del escenario de simulación, casos de estudio y el plan de pruebas a implementar; el tercer capítulo realiza el análisis de los resultados obtenidos y su comparación con trabajos relacionados; por último, en el cuarto capítulo, se presentan las conclusiones y trabajos futuros.

CAPITULO 1: GENERALIDADES

1.1. IMT-2020

En el año 2012, la ITU y más específicamente la Sección de Radiocomunicaciones de la ITU (ITU-R, ITU *Radio Section*) definió el comienzo del desarrollo de los nuevos sistemas de comunicaciones móviles bajo la norma de Telecomunicaciones Móviles Internacionales para el año 2020 y futuro (IMT-2020, *International Mobile Telecommunications for 2020 and beyond*). El marco de trabajo y visión de dicho proyecto están plasmados en la recomendación ITU-R M.2083-0, la cual establece los siguientes requerimientos [5]:

- Ancho de banda de al menos 100 MHz.
- Densidad de dispositivos conectados de al menos 1 millón por Km².
- Velocidad de transmisión de datos pico en el Enlace de Bajada (DL, *Downlink*) de 20 Gb/s.
- Velocidad de transmisión de datos pico en el Enlace de Subida (UL, *Uplink*) de 10 Gb/s.
- Latencia en la RAN de máximo 1 ms para servicios URLLC.

En las siguientes secciones se describen las características de la tecnología de acceso radio 5G definidas hasta el momento que permitirán alcanzar estos objetivos.

1.2. TECNOLOGÍA DE ACCESO RADIO 5G

La tecnología RAN 5G está diseñada para interconectar toda clase de dispositivos que se beneficien de dicha conexión. A partir de esta premisa surgen tres grandes desafíos:

- Crecimiento masivo en el número de dispositivos conectados.
- Crecimiento masivo en el volumen de tráfico.

- Un gran rango de aplicaciones con requerimientos y características diferentes.

Para enfrentar estos retos, la tecnología RAN 5G no solamente requiere nuevas funcionalidades, sino también una mayor cantidad de espectro y nuevas bandas de frecuencia [6]. La Figura 1.1 muestra los rangos de frecuencia en los que operan los sistemas de comunicaciones móviles actuales y los futuros sistemas 5G.

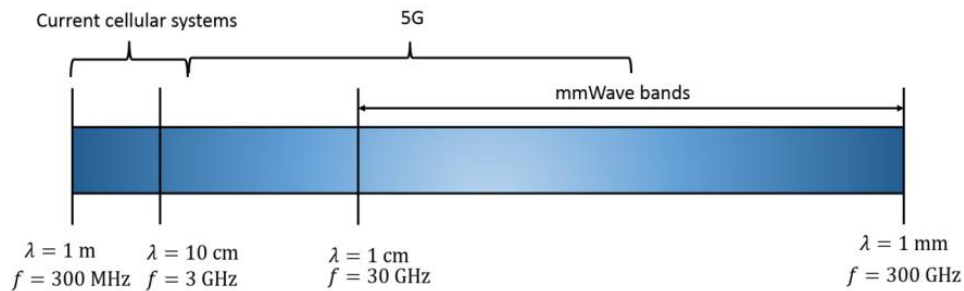


Figura 1.1. Banda de operación de sistemas 5G. Tomada de [7].

Los sistemas celulares actuales operan por debajo de la frecuencia de 6 GHz. Una gran cantidad de espectro está disponible en la banda mmWave (30 GHz - 300 GHz), sin embargo, no hay ningún sistema de comunicaciones móviles comercial operando en dicha banda. El Proyecto de Cooperación para Sistemas de Tercera Generación (3GPP, *3rd Generation Partnership Project*) está desarrollando un estándar global para la tecnología de acceso radio 5G denominado Nueva Radio (NR, *New Radio*), el cual operará desde la frecuencia de 1 GHz hasta 100 GHz. La primera versión del estándar NR fue completada a mediados del 2018 y se definió bajo el nombre de Especificación 15 (R15, *Release 15*). La segunda versión del estándar, definida como Especificación 16 (R16, *Release 16*), se encuentra en desarrollo y será completada a finales del 2019 [8]. El proceso de estandarización de sistemas 5G se muestra en la Figura 1.2.



Figura 1.2. Proceso de estandarización de sistemas 5G. Tomada de [8].

1.2.1. Espectro para Sistemas 5G

Uno de los principales cambios con respecto a los anteriores sistemas de comunicaciones móviles es que 5G usará frecuencias considerablemente altas en la banda mmWave. En la primera versión de NR, 3GPP ha decidido soportar la operación en el rango desde 1 GHz hasta 52.6 GHz. Este cambio es motivado por la disponibilidad de gran cantidad de espectro en dicha banda. Sin embargo, la operación en la banda mmWave conlleva a los siguientes retos [9]:

- Las pérdidas de transmisión crecen considerablemente en esta banda, por lo que es necesario el uso de técnicas multi-antena y de conformación de haces (*beamforming*) para compensar el aumento de la pérdida de propagación.
- El desempeño de los sistemas de radio frecuencia es degradado debido al ruido, y los fenómenos de absorción, refracción, reflexión y dispersión.
- Se requiere considerar la interferencia que pueda ser ocasionada por otros sistemas inalámbricos operando en altas frecuencias, i.e., sistemas satelitales.

1.2.2. Arquitectura de Protocolos Radio

La arquitectura de protocolos radio para NR puede ser dividida en el Plano de Control (CP, *Control Plane*) y el Plano de Usuario (UP, *User Plane*). El UP entrega datos de usuario, mientras que el CP es responsable del establecimiento de las conexiones, la movilidad de usuarios y la seguridad. La pila de protocolos se compone de los siguientes niveles y subniveles: Nivel Físico (PHY, *Physical Layer*), Subnivel de Control de Acceso al Medio (MAC, *Medium Access Control*), Subnivel de Control del Enlace Radio (RLC, *Radio Link Control*), Subnivel de Protocolo de Convergencia de Paquetes de Datos (PDCP, *Packet Data Convergence Protocol*) y Subnivel del Protocolo de Adaptación de Datos de Servicio (SDAP, *Service Data Adaptation Protocol*) [10]. La Figura 1.3 b. muestra la pila de protocolos NR para el plano de usuario.

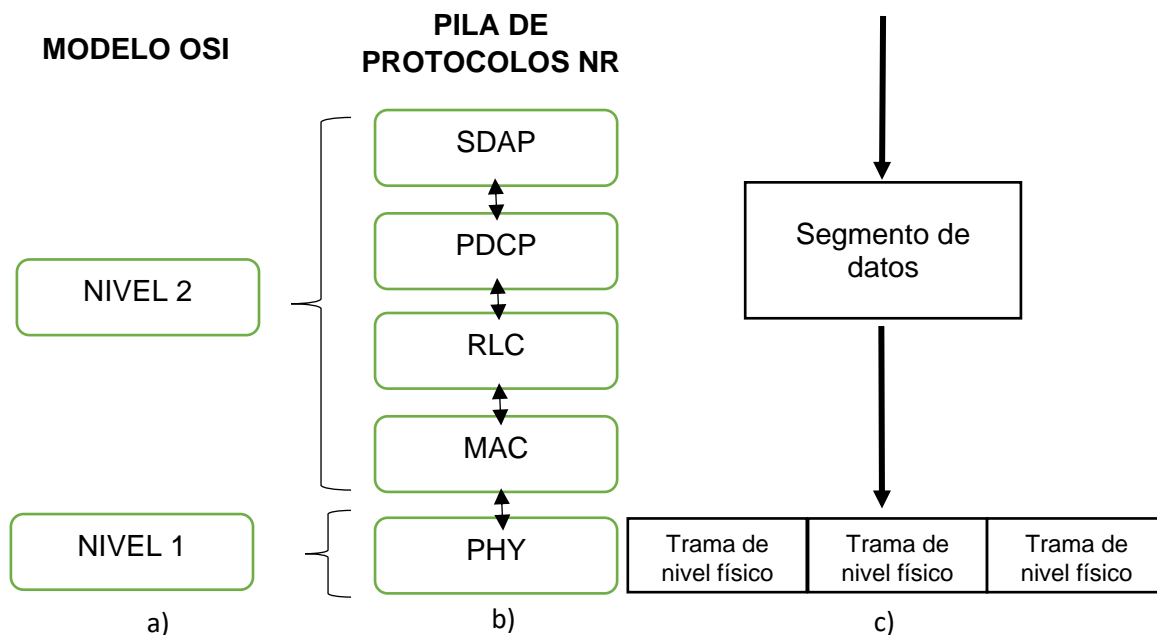


Figura 1.3. Comparación de la pila de protocolos NR con el modelo OSI. a) Modelo OSI. b) Pila de protocolos NR. c) PDU del nivel físico. Tomada de [11].

Teniendo en cuenta el modelo de referencia para Interconexión de Sistemas Abiertos (OSI, *Open System Interconnection*), el nivel PHY corresponde al primer nivel de dicho modelo, mientras que los subniveles MAC, RLC, PDCP y SDAP

corresponden al segundo nivel del mismo, tal como se muestra en la Figura 1.3 a. En este trabajo de grado se denomina trama de nivel físico al PDU de dicho nivel y segmento de datos a la información transportada por niveles superiores que se entrega al nivel físico para su posterior transmisión por el canal radio, tal como se muestra en la Figura 1.3 c. A continuación, se describen las principales funcionalidades de cada nivel y subnivel:

- El subnivel SDAP administra el flujo de paquetes del Protocolo de Internet (IP, *Internet Protocol*) de acuerdo al requerimiento de Calidad del Servicio (QoS, *Quality of Service*).
- El subnivel PDCP es responsable de la compresión y descompresión de encabezados IP, ordenamiento de paquetes, detección de duplicados, cifrado y descifrado. La compresión del encabezado reduce el número de bits a ser transmitidos por la interfaz radio. El cifrado ofrece seguridad y mantiene la integridad de los datos. El ordenamiento y detección de duplicados permite la correcta entrega de las Unidades de Datos del Protocolo (PDU, *Protocol Data Unit*).
- El subnivel RLC realiza la corrección de errores a través del mecanismo de Solicitud de Retransmisión Automática (ARQ, *Automatic Repeat Request*), también desempeña las funciones de segmentación y desegmentación de los paquetes IP y de entrega ordenada de PDUs hacia los niveles superiores.
- El subnivel MAC es responsable de la corrección de errores a través del mecanismo de Solicitud de Retransmisión Automática Híbrida (HARQ, *Hybrid Automatic Repeat Request*) y la planificación (*scheduling*) en el enlace de subida y bajada. El planificador (*scheduler*) asigna los recursos radio en la dimensión tiempo-frecuencia.

- El nivel físico es el encargado de la codificación y decodificación, modulación y demodulación, procesamiento multi-antena y mapeo de señales según los recursos radio disponibles.

Dado que el presente trabajo de grado se enfoca en el nivel físico de sistemas 5G, en las siguientes secciones se describen las principales características de dicho nivel.

1.2.3. Modulación

NR soporta los siguientes esquemas de modulación tanto en el enlace de bajada como de subida: Modulación por Desplazamiento de Fase en Cuadratura (QPSK, *Quadrature Phase Shift Keying*), Modulación de Amplitud en Cuadratura de 16 estados (16 QAM, 16 *Quadrature Amplitude Modulation*), Modulación de Amplitud en Cuadratura de 64 estados (64 QAM, 64 *Quadrature Amplitude Modulation*) y Modulación de Amplitud en Cuadratura de 256 estados (256 QAM, 256 *Quadrature Amplitude Modulation*). El esquema de Modulación por Desplazamiento de Fase Binaria $\pi/2$ ($\pi/2$ BPSK, $\pi/2$ *Binary Phase Shift Keying*) puede ser utilizado en el enlace de subida para transmitir información a una velocidad de transmisión de datos baja [10].

1.2.4. Forma de Onda

NR emplea Multiplexación por División de Frecuencias Ortogonales con Prefijo Cíclico (CP-OFDM, *Cyclic Prefix Orthogonal Frequency-Division Multiplexing*) tanto en el enlace de bajada como de subida. En esta forma de onda, la separación entre subportadoras asegura su ortogonalidad (ver Figura 1.4), evitando la interferencia entre subportadoras y la necesidad de bandas de guarda o filtros pasabanda complejos. Al tener la misma forma de onda en ambos enlaces se simplifica el diseño general del sistema especialmente para aplicaciones de Dispositivo a

Dispositivo (D2D, *Device to Device*). Adicionalmente es posible usar la Multiplexación por División de Frecuencias Ortogonales con Dispersión por Transformada de Fourier Discreta (DFTS-OFDM, *Discrete Fourier Transform Spread Orthogonal Frequency-Division Multiplexing*) para el enlace de subida en escenarios de cobertura limitada [12].

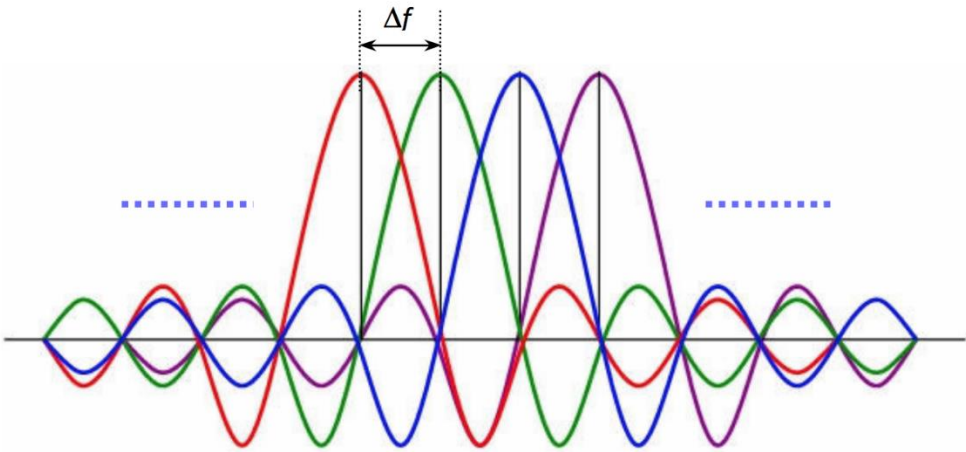


Figura 1.4. OFDM y separación entre subportadoras. Tomada de [12].

La CP-OFDM inserta un Prefijo Cíclico (CP, *Cyclic Prefix*) en el símbolo OFDM para mitigar los efectos de la Interferencia Intersimbólica (ISI, *Intersymbol Interference*) por efectos de multitrayectoria. El CP se define como un intervalo de guarda que consiste en la copia de la última parte del símbolo OFDM. Si la duración del CP es mayor que la Dispersión del Retardo (DS, *Delay Spread*), la forma de onda no sufre de ISI. El CP se muestra en la Figura 1.5 [13].



Figura 1.5. Prefijo cíclico en CP-OFDM. Tomada de [14]

NR define diferentes valores para la duración del símbolo y el espaciado entre subportadoras en OFDM, definiciones conocidas como numerología. En la primera versión de NR, 3GPP define cinco valores para el espaciado entre

subportadoras: 15 KHz, 30 KHz, 60 KHz, 120 KHz y 240 KHz. Esta característica permitirá el soporte de diversos servicios con diferentes requerimientos [10]. En la Figura 1.6 se muestran las diferentes numerologías definidas para NR y el ancho de banda considerando 12 subportadoras.

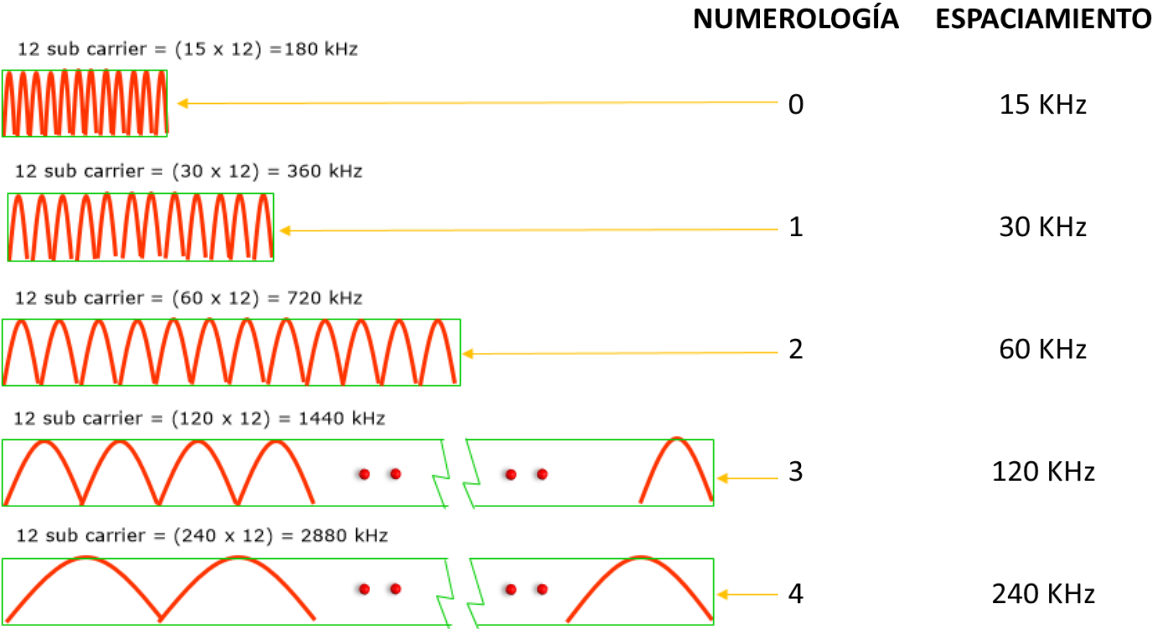


Figura 1.6. Numerología para NR. Tomada de [10].

Para implementar canales con ancho de banda mayor es posible usar Agregación de Portadoras (CA, *Carrier Aggregation*). En NR es posible combinar hasta 16 portadoras.

1.2.5. Múltiples Antenas

Las técnicas multi-antena tienen un papel crucial en NR. Dada las altas frecuencias de operación, NR se ha diseñado en torno al soporte de técnicas multi-antena y conformación de haces (*beamforming*) para proporcionar una cobertura adecuada. 3GPP ha definido procedimientos para implementar estas tecnologías tanto en la BS como en el UE y serán utilizadas no solamente para la transmisión de datos sino también para la transmisión de mensajes de difusión y comunicación inicial [15].

1.2.6. Codificación de Canal

NR emplea Verificación de Paridad de Baja Densidad (LDPC, *Low Density Parity Check*) para transmisión en la Banda Ancha Móvil (MBB, *Mobile Broadband*) y Codificación Polar (*polar codes*) para la transmisión de señales de control [12].

1.2.7. Esquema de Duplexación

NR soporta Duplexación por División de Tiempo (TDD, *Time Division Duplexing*) y Duplexación por División de Frecuencia (FDD, *Frequency Division Duplexing*). Así mismo, NR soporta TDD dinámico donde la asignación de intervalos de tiempo en el enlace de bajada y subida cambia con el tiempo [10, 12].

NR soporta la técnica de Enlace de Subida Suplementario (SUL, *Supplementary Uplink*). Como se muestra en la Figura 1.7, un par convencional de enlaces de subida y bajada pueden asociarse con un enlace adicional de subida operando en frecuencias bajas, con el objetivo de mejorar la cobertura en escenarios de potencia limitada. El Enlace de Bajada Suplementario (SDL, *Supplementary Downlink*) es la contraparte de SUL y asocia un enlace de bajada adicional operando en frecuencias bajas [12].

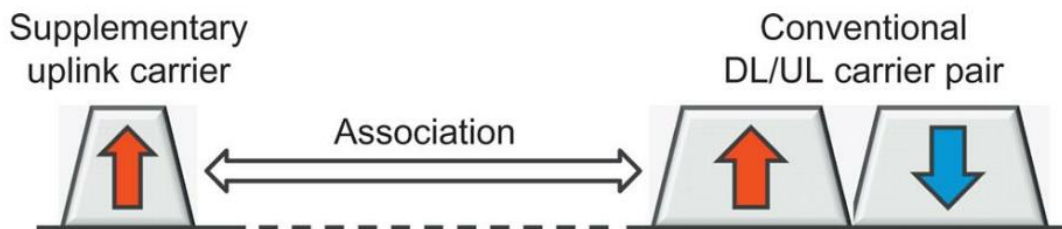


Figura 1.7. Enlace de Subida Suplementario. Tomada de [14]

1.2.8. Bandas de Operación

En la primera versión de NR se han definido dos rangos de operación:

- Rango de Frecuencia 1 (FR1, *Frequency Range 1*) incluye las bandas por debajo de 6 GHz.

- Rango de Frecuencia 2 (FR2, *Frequency Range 2*) incluye las bandas en el rango 24.25 a 52.6 GHz.

3GPP define 26 bandas de operación en FR1 y 3 bandas de operación en FR2 [8]. Las Tablas 1.1 y 1.2 muestran la definición de frecuencias, el nombre de la banda y el modo de duplexación para FR1 y FR2, respectivamente.

Nombre de la banda	Banda de Frecuencia en UL (MHz)	Banda de Frecuencia en DL (MHz)	Modo de Duplexación
n1	1920 - 1980	2110 - 2170	FDD
n2	1850 - 1910	1930 - 1990	FDD
n3	1710 - 1785	1805 - 1880	FDD
n5	824 - 849	869 - 894	FDD
n7	2500 - 2570	2620 - 2690	FDD
n8	880 - 915	925 - 960	FDD
n20	832 - 862 MHz	791 - 821	FDD
n28	703 - 748	758 - 803	FDD
n38	2570 - 2620	2570 - 2620	TDD
n41	2496 - 2690	2496 - 2690	TDD
n50	1432 - 1517	1432 - 1517	TDD
n51	1427 - 1432	1427 - 1432	TDD
n66	1710 - 1780	2110 - 2200	FDD
n70	1695 - 1710	1995 - 2020	FDD
n71	663 - 698	617 - 652	FDD
n74	1427 - 1470	1475 - 1518	FDD
n75	N/A	1432 - 1517	SDL
n76	N/A	1427 - 1432	SDL
n77	3300 - 4200	3300 - 4200	TDD
n78	3300 - 3800	3300 - 3800	TDD
n79	4400 - 5000	4400 - 5000	TDD
n80	1710 - 1785	N/A	SUL
n81	880 - 915	N/A	SUL
n82	832 - 862	N/A	SUL
n83	703 - 748	N/A	SUL
n84	1920 - 1980	N/A	SUL

Tabla 1.1. Bandas de operación para FR1. Tomada de [8].

Nombre de la Banda	Banda de Frecuencia (MHz)	Modo de Duplexación
n257	26500 - 29500	TDD
n258	24250 - 27500	TDD
n260	37000 - 40000	TDD

Tabla 1.2. Bandas de operación para FR2. Tomada de [8]

1.2.9. Estructura de Transmisión en el Dominio del Tiempo

En el dominio del tiempo, la transmisión está organizada en tramas con una duración de 10 ms, cada trama es dividida en 10 subtramas de 1 ms de duración cada una. Una subtrama es dividida a su vez en TTI compuestos por 14 símbolos OFDM. La duración de cada TTI depende de la numerología implementada como se muestra en la Figura 1.8. En NR es posible usar TTI con un número menor de 14 símbolos, dichas estructuras se definen como mini-TTI. En la primera versión de NR se definen mini-TTI con duración de 2, 4 y 7 símbolos OFDM (ver Figura 1.9) [16]. La duración del símbolo para cada numerología se muestra en la Tabla 1.3.

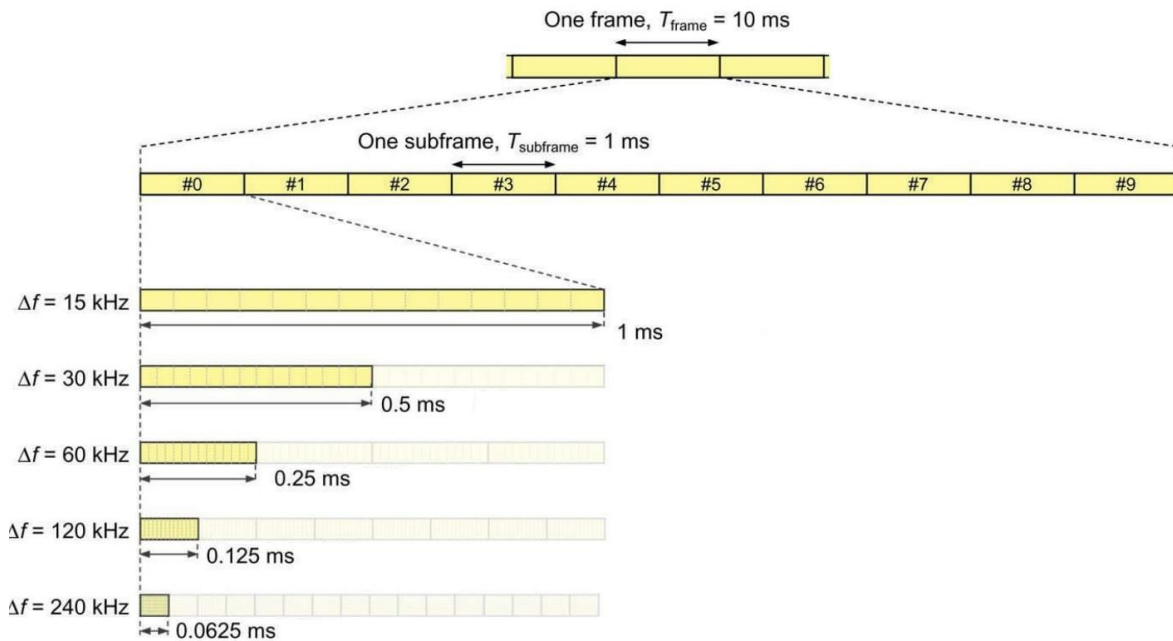


Figura 1.8. Estructura de trama en NR. Tomada de [16].

Numerología	0	1	2	3	4
Espaciamiento entre subportadoras (KHz)	15	30	60	120	240
TTI (ms)	1	0.5	0.25	0.125	0.0625
Duración del símbolo OFDM (μ s)	66.67	33.33	16.67	8.33	4.17

Tabla 1.3. Duración de símbolo en NR. Tomada de [8].

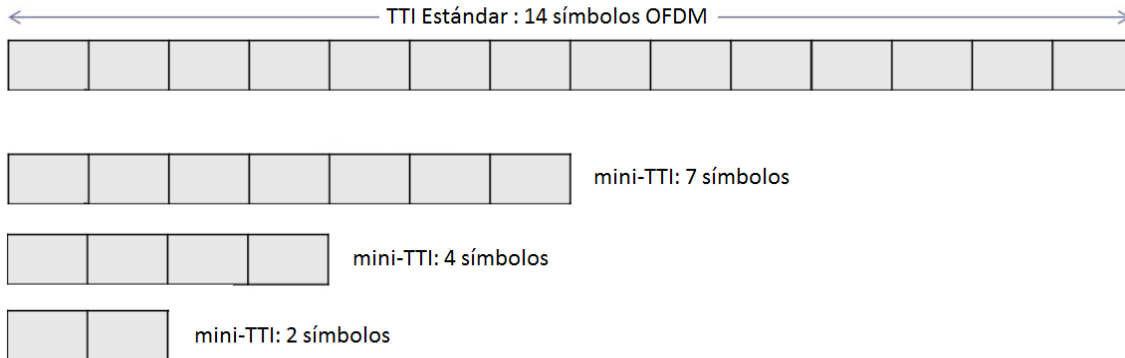


Figura 1.9. Mini-TTI de 7, 4 y 2 símbolos.

1.2.10. Estructura de Transmisión en el Dominio de la Frecuencia

En el dominio de la frecuencia, NR define Bloques de Recursos (RB, *Resource Blocks*) compuestos por 12 subportadoras. Los RB pueden ser agrupados para formar Partes de Ancho de Banda (BWP, *Bandwidth Part*) y asignar recursos de frecuencia dependiendo de los requerimientos de velocidad de transmisión de datos. El rango de frecuencias ocupado por cada RB dependerá de la numerología empleada [8].

Al considerar las dimensiones frecuencia-tiempo es posible obtener la estructura denominada Cuadrícula de Recursos (*Resource Grid*). A diferencia de los Sistemas de Evolución a Largo Plazo (LTE, *Long Term Evolution*), NR define diferentes conjuntos de recursos para la *Resource Grid*, uno por cada numerología. Un elemento de recurso (*Resource Element*) es el recurso compuesto por una

subportadora y la duración de un símbolo OFDM, tal como se muestra en la Figura 1.10.

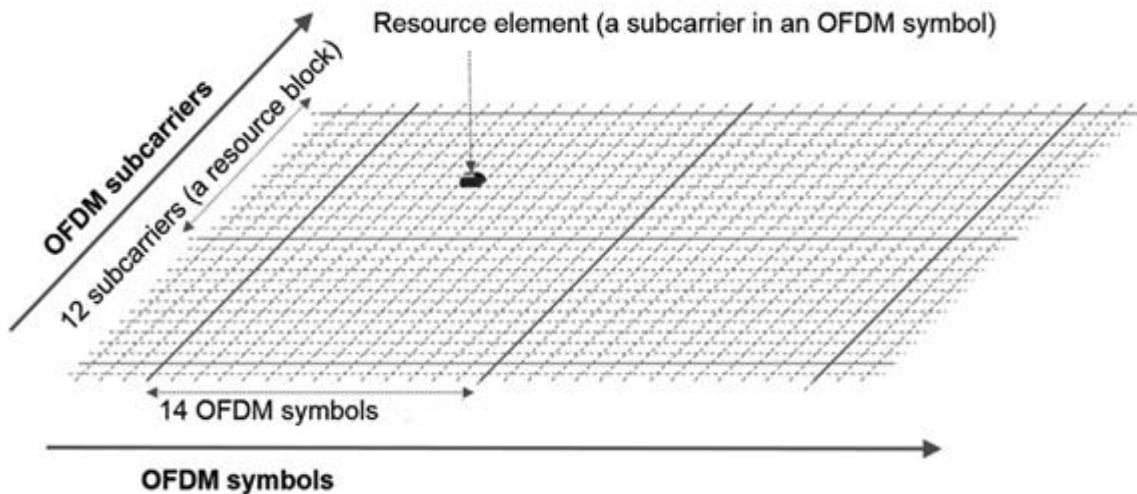


Figura 1.10. Elemento de Recurso en NR. Tomada de [11].

1.3. CONCEPTO DE LATENCIA.

La latencia en redes de telecomunicaciones es definida como el tiempo que toma un segmento de datos en viajar de un nodo a otro. Es causada por diversos retardos de transmisión y de procesamiento de datos que variarán de acuerdo a las características propias de la red.

Existen diferentes definiciones de latencia que dependen de la selección del emisor y receptor en una red. Al considerar el emisor y receptor ubicados en extremos opuestos de la red es posible definir El Tiempo de Ida y Vuelta (RTT, *Round Trip Time*), el cual comprende el tiempo que toma un segmento en ser enviado desde el emisor, en retornar al mismo emisor habiendo pasado por el receptor. La latencia extremo a extremo (E2E, *End to End*) o latencia de una vía, representa el tiempo que toma un segmento en ser transmitido desde el emisor y llegar al receptor, usualmente corresponde a la mitad del RTT [17]. Estas definiciones de latencia se muestran en la Figura 1.11.

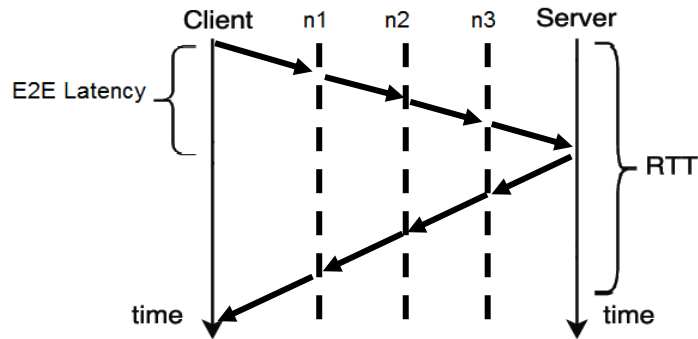


Figura 1.11. RTT y latencia E2E. Tomada de [17].

1.4. LATENCIA EN REDES CELULARES

La latencia en redes celulares puede ser clasificada en dos grupos: latencia en el UP, definida como el tiempo de transmisión entre la BS y el UE; y latencia en el CP, definida como el tiempo de transición entre el estado inactivo y activo de un terminal [18].

La latencia en redes celulares puede originarse en los distintos componentes de la arquitectura del sistema: la RAN, Red de Transporte (*Backhaul*), el Núcleo de Red (CN, *Core Network*) e Internet [19]. Dichos componentes son ilustrados en la Figura 1.12. En NR, las estaciones base se denominan Nodos B de Próxima Generación (gNB, *Next Generation Node B*) que proporcionan conectividad a los UE.

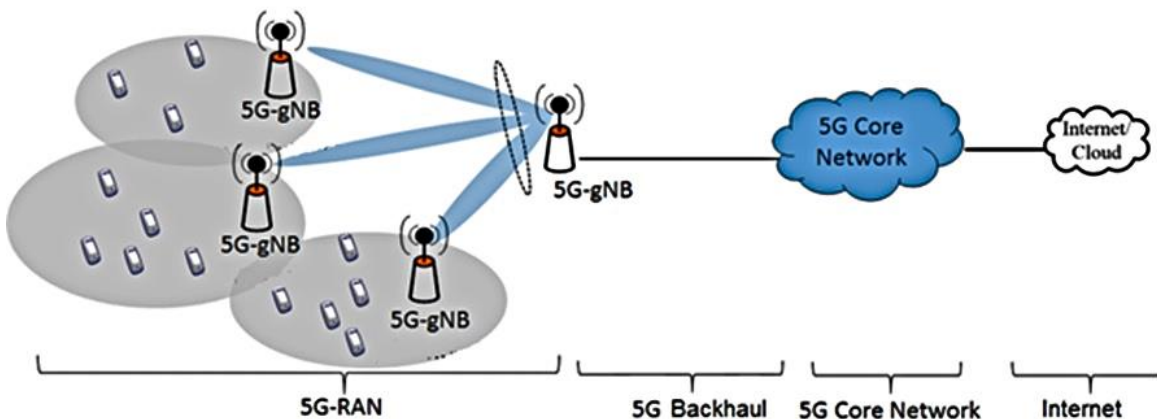


Figura 1.12. Arquitectura de sistema 5G. Tomada de [20].

La latencia extremo a extremo en sistemas 5G puede ser definida como la suma de la latencia presente en los diferentes componentes de la arquitectura:

$$T_{E2E} = T_{Radio} + T_{Backhaul} + T_{core} + T_{Internet} \quad (1.1)$$

donde,

T_{Radio} : representa la latencia en la RAN y consiste en el tiempo de transmisión de un segmento entre la BS y el UE o viceversa. Es originado principalmente por los procesos del nivel físico. Está compuesto por el tiempo de transmisión, tiempo de procesamiento de UE y BS y tiempo de propagación.

$T_{Backhaul}$: es el tiempo de establecimiento de conexiones entre la RAN y el CN. Depende del medio de transmisión que conecta ambas redes, generalmente se usa cable coaxial, fibra óptica o enlace por microondas.

T_{core} : consiste en el tiempo de procesamiento de las diferentes entidades del CN. Los procesos de control, seguridad, asignación de direcciones IP, filtro de paquetes, entre otros, se llevan a cabo en esta red.

$T_{Internet}$: es el retardo de transmisión existente entre el CN e Internet. Depende del ancho de banda de los enlaces entre dichas redes y el protocolo empleado para su comunicación.

Habiendo definido los conceptos y tipos de latencia en redes celulares es importante establecer que el presente trabajo de grado enfoca su estudio en el primer componente de la Ecuación 1.1, es decir la latencia en la RAN. Así mismo, se analizará dicha latencia dependiendo del sentido de comunicación: en el UL o DL.

Para alcanzar el objetivo de 1 ms de latencia en la RAN, el 3GPP ha implementado mejoras en dicha red. Las tecnologías y técnicas seleccionadas para NR en el nivel físico fueron descritas en secciones anteriores. Es posible clasificar estas soluciones como: estructura de trama, técnicas de acceso múltiple, esquemas de

codificación y modulación, sistemas multi-antena, conformación de haces (*beamforming*) y agregación de portadoras.

La flexibilidad de la estructura de trama de datos definida para NR, permite modificar el valor del TTI según los requerimientos de los servicios. Esta característica busca mejorar la latencia en la RAN con respecto a sistemas de comunicaciones móviles anteriores [21]. El presente trabajo de grado analiza el impacto del TTI sobre la latencia en la RAN, más específicamente en el primer nivel del modelo de referencia OSI. Las demás soluciones implementadas por el 3GPP en el nivel físico y en los diferentes componentes de la arquitectura no serán consideradas en el presente trabajo de grado.

1.5. SERVICIOS DE BAJA LATENCIA EN 5G

Los requerimientos de latencia son críticos en algunas aplicaciones como producción industrial automatizada, sistemas de control, robótica, transporte, servicios de salud, entretenimiento, realidad virtual, educación y comunicación. Aunque algunas de estas aplicaciones están siendo implementadas a través de LTE o sistemas móviles 4G, existen requerimientos más exigentes sobre la red móvil en términos de latencia, confiabilidad y seguridad. Algunas de las aplicaciones a implementar en sistemas 5G se describen a continuación [22]:

- **Automatización Industrial:** consiste en el control en tiempo real de maquinaria y sistemas para la producción industrial. El requerimiento de latencia para esta aplicación está entre los 0.25 ms y 10 ms.
- **Sistemas de Transporte Inteligente:** la conducción autónoma y la optimización del tráfico en ciudades requiere de URLLC. La conducción autónoma de vehículos requiere una latencia máxima de 10 ms. Para el control y optimización de tráfico es necesaria una latencia de 10 a 100 ms.
- **Robótica y Tele-presencia:** en el futuro cercano robots controlados de forma remota tendrán aplicaciones en diversos sectores como construcción

y mantenimiento de áreas peligrosas. La respuesta de este tipo de sistemas debe ser del orden de un par de milisegundos.

- **Realidad Virtual:** varias aplicaciones como tele-cirugía requieren altos niveles de sensibilidad y precisión para manipular objetos. Los requerimientos de esta tecnología no pueden ser alcanzados con los sistemas móviles actuales. La latencia para implementar estas aplicaciones debe ser menor a 1 ms.
- **Realidad Aumentada:** la tecnología de Realidad Aumentada (AR, *Augmented Reality*) despliega información relevante teniendo en cuenta la ubicación y estado del usuario en su campo visual. Algunos ejemplos de esta tecnología son: asistencia de conducción, mantenimiento de equipos, guía de recorridos, educación remota y tecnologías de asistencia para policías y bomberos. Sin embargo, la capacidad computacional insuficiente y la latencia presente en los sistemas de comunicaciones móviles actuales no han permitido el despliegue de dichas aplicaciones. En este caso la latencia necesaria es del orden de unos pocos milisegundos.
- **Servicios de Salud:** tele-diagnóstico, tele-cirugía y tele-rehabilitación son algunas aplicaciones en el área de salud que requieren valores mínimos de latencia. Estos servicios requieren una latencia en el rango de 1 a 10 ms.
- **Videojuegos:** latencias entre 30 y 50 ms degradan significativamente la experiencia de juego. En esta aplicación es recomendable contar con latencias menores a 1 ms, el cual consiste en el tiempo de reacción perceptible por el ojo humano.
- **Smart Grid:** consiste en el control dinámico e inteligente de energía eléctrica. Es una de las aplicaciones más sensibles a la latencia debido a la complejidad en el control de sistemas eléctricos y electrónicos. Para su correcta operación la latencia no puede superar 1 ms.
- **Educación y Cultura:** aplicaciones como educación remota o la interpretación remota de instrumentos musicales requieren de una latencia entre 1 a 10 ms.

CAPÍTULO 2: DESARROLLO METODOLÓGICO

La metodología Proceso Racional Unificado (RUP, *Rational Unified Process*) es un proceso de desarrollo incremental en cuatro etapas: análisis de requerimientos, diseño, implementación y pruebas, como se observa en la Figura 2.1 [23]. Esta metodología es utilizada para desarrollar las simulaciones necesarias para el cumplimiento de los objetivos del presente trabajo de grado.

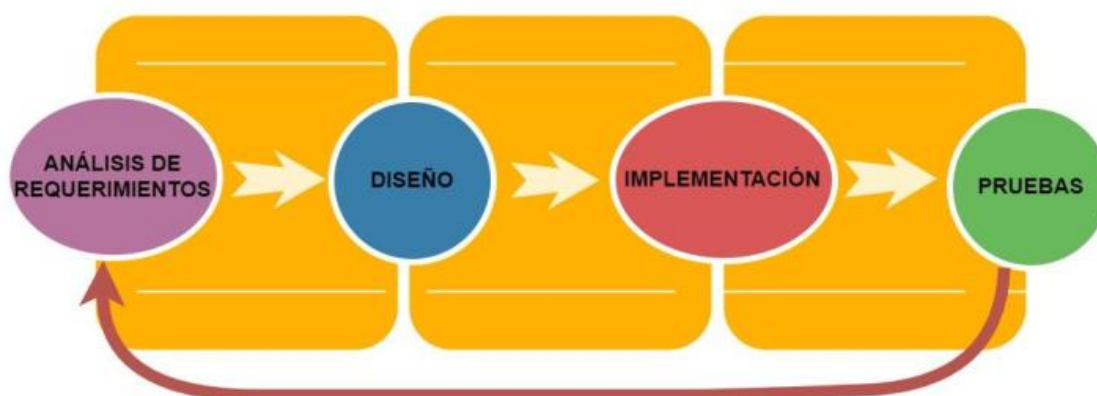


Figura 2.1. Diagrama de metodología RUP. Tomada de [24].

2.1. ANÁLISIS DE REQUERIMIENTOS

En este apartado se describen el contexto y los requerimientos de las simulaciones, los cuales son agrupados en requerimientos funcionales y no funcionales.

2.1.1. Descripción del Problema

Uno de los objetivos de los sistemas 5G es proporcionar una baja latencia para soportar servicios URLLC. Algunos ejemplos de servicios URLLC fueron presentados en el Capítulo 1, en donde se estableció que los requerimientos de latencia para el correcto despliegue de estos servicios se encuentran en el rango de 1 ms a 10 ms.

La R15 define una nueva estructura de trama y numerologías para mejorar la latencia en la RAN. Estas características permiten adecuar el TTI de acuerdo a los requerimientos de los servicios.

Es de importancia establecer que el presente trabajo de grado considera la latencia en el UP, en el nivel físico del modelo de referencia OSI, y en la RAN.

Se diseña el escenario de simulación considerando las componentes de latencia en el nivel físico. Este escenario está caracterizado por parámetros cuya variación impone diferentes condiciones sobre el sistema 5G. Cada parámetro es considerado de forma independiente en casos de estudio en los cuales se varía el TTI para analizar su impacto en la latencia en la RAN en diferentes condiciones.

2.1.2. Requerimientos Funcionales

- Requerimiento 1: configurar parámetros generales del sistema 5G (ancho de banda, frecuencia de operación y densidad de gNBs y UE).
- Requerimiento 2: configurar parámetros asociados al nivel físico de la RAN (estructura de trama, separación de subportadoras y cantidad de símbolos OFDM por TTI).
- Requerimiento 3: configurar parámetros asociados a los servicios a desplegar en el sistema 5G (tamaño y tasa de segmentos de datos).
- Requerimiento 4: obtener la latencia de los segmentos de datos en la RAN.

2.1.3. Requerimientos no Funcionales

- Requerimiento 1: graficar de manera intuitiva los resultados obtenidos en los diferentes casos de estudio para facilitar su análisis.

2.1.4. Criterios de Selección de la Herramienta de Simulación

Los siguientes son los criterios considerados para seleccionar la herramienta de simulación:

- Capacidad de simular redes 5G a nivel de sistema.
- Capacidad de configuración de la trama de nivel físico según la Especificación 15 del 3GPP.
- Capacidad de configuración y despliegue de flujos de datos.
- Software de código abierto con fines académicos.
- Modularidad de la herramienta.

Inicialmente se consideraron las siguientes herramientas de simulación: OMNET++ [25], OPNET [26], Matlab [27] y ns-3, evaluando sus ventajas y desventajas. Las herramientas OMNET++ y OPNET no cuentan con un módulo para desarrollar simulaciones de redes 5G, sin embargo, es posible añadir código fuente adicional para implementar este tipo de redes. El trabajo de codificar un módulo completo en alguna de estas herramientas es una tarea dispendiosa y compleja. Matlab ofrece un módulo específico para sistemas 5G que permite la configuración de los parámetros asociados al nivel físico, sin embargo, dicho módulo permite realizar simulaciones únicamente a nivel de enlace. Finalmente, la herramienta ns-3 cuenta con un módulo de redes 5G completamente funcional que permite la configuración de los parámetros que definen la estructura de trama de nivel físico y la ejecución de simulaciones a nivel de sistema. Esta herramienta de código abierto es desarrollada y mantenida por una comunidad académica extensa por lo que cumple a cabalidad con los criterios de selección definidos anteriormente. A continuación, se presenta una descripción de las funcionalidades más importantes de ns-3, que son utilizadas para el desarrollo del presente trabajo de grado.

2.1.5. Herramienta de Simulación ns-3

ns-3 es un simulador de redes de eventos discretos desarrollado con fines educativos e investigativos. ns-3 es software de código abierto bajo Licencia Pública General (GPL, *General Public License*). Dada la complejidad de las redes de comunicaciones, el simulador realiza cálculos computacionales que de forma teórica serían difíciles de llevar a cabo [28].

Para el presente trabajo de grado se utiliza el módulo de ondas milimétricas (*mmWave module*) desarrollado por la Universidad de Nueva York. Este módulo cuenta con diversos parámetros configurables para simular diferentes condiciones sobre el sistema 5G [29]. Dado que el presente trabajo de grado se enfoca en la estructura de trama de nivel físico, a continuación, se presenta la implementación de la misma en el módulo mmWave.

2.1.5.1. Estructura de Trama Módulo mmWave ns-3

La estructura de trama del módulo mmWave es completamente configurable. La implementación de la trama en el módulo se asemeja a la definida en NR, con la excepción de que los parámetros que conforman la trama pueden ser modificados, teniendo un control total de esta estructura. La trama en el módulo mmWave es subdividida en un número específico de subtramas, a su vez, la subtrama es dividida en TTI. Cada TTI comprende un número específico de símbolos OFDM y podrán ser asignados como información de datos o control tanto en el enlace de subida como de bajada [30].

A continuación, se presenta un ejemplo de implementación de la trama con los valores presentados en la Tabla 2.1. La representación de la trama se muestra en la Figura 2.2.

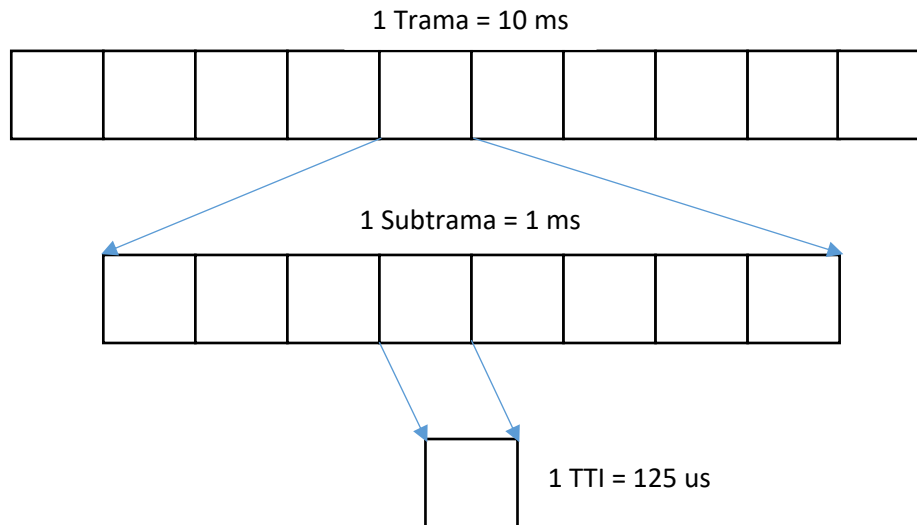


Figura 2.2. Ejemplo de implementación de trama mmWave. Tomada de [30].

Parámetro	Valor
Símbolos por TTI	30
Duración de símbolo	4.16 us
TTI por subtrama	8
Subtramas por trama	10

Tabla 2.1. Valores de ejemplo para la trama mmWave. Tomada de [30].

2.1.5.2. Modelo de Canal implementado en ns-3

El modelo de canal para la banda mmWave implementado en ns-3, es el especificado en el reporte técnico 38.901 del 3GPP [31]. Este modelo es válido para frecuencias entre 6 GHz a 100 GHz y ancho de banda de hasta el 10% de la frecuencia central del sistema.

El reporte técnico 38.901 describe un modelo estadístico espacial en tres dimensiones, basado en campañas de mediciones mmWave realizadas por el 3GPP. El modelo tiene en cuenta la posición, velocidad y el escenario en el cual se encuentran los terminales para determinar las pérdidas de propagación. Los escenarios definidos en este modelo son:

- Microcelda Urbana: escenario urbano con alta densidad de UE. Las BS son ubicadas por debajo de las azoteas de edificios.

- Macrocela Urbana: escenario urbano con baja densidad de UE y mayor cobertura. Las BS son ubicadas por encima de las azoteas de edificios.
- Macrocela Rural: escenario rural de gran cobertura. Válido hasta la frecuencia de 7 GHz.
- Interiores: escenario de oficinas o centros comerciales en donde una misma área o piso es compartido por los UE. La cobertura máxima es de 100 m.

El reporte técnico 38.901 proporciona características opcionales que permiten especificar el canal con más detalle, si las simulaciones así lo requieren. Estas características se describen a continuación:

- Consistencia espacial: definida para escenarios con movilidad alta, como por ejemplo, autopistas.
- Atenuación debida al oxígeno: tiene en cuenta atenuación adicional debida a la absorción atmosférica en determinadas frecuencias.
- Obstrucción o bloqueo: tiene en cuenta la atenuación debido a la presencia de obstáculos entre BS y UE.

2.2. DISEÑO

En esta etapa se diseña el modelo de simulación y el escenario específico de simulación que permitirá evaluar el impacto del TTI en la latencia de la RAN 5G. Para esto, es indispensable analizar las componentes de la latencia en el nivel físico y determinar así los parámetros adecuados para dicha evaluación.

2.2.1. Componentes de Latencia en el Nivel Físico

Inicialmente, los niveles superiores entregan segmentos de datos al nivel físico, el cual los almacena en un *buffer* de transmisión. Seguidamente, el nivel físico espera al siguiente TTI para comenzar la transmisión. La información es puesta entonces en la trama de nivel físico, la cual viaja hasta el nivel físico del receptor. El nivel físico del receptor realiza procesamientos sobre la trama recibida y entregará la información a los niveles superiores. Este proceso se muestra en la Figura 2.3.

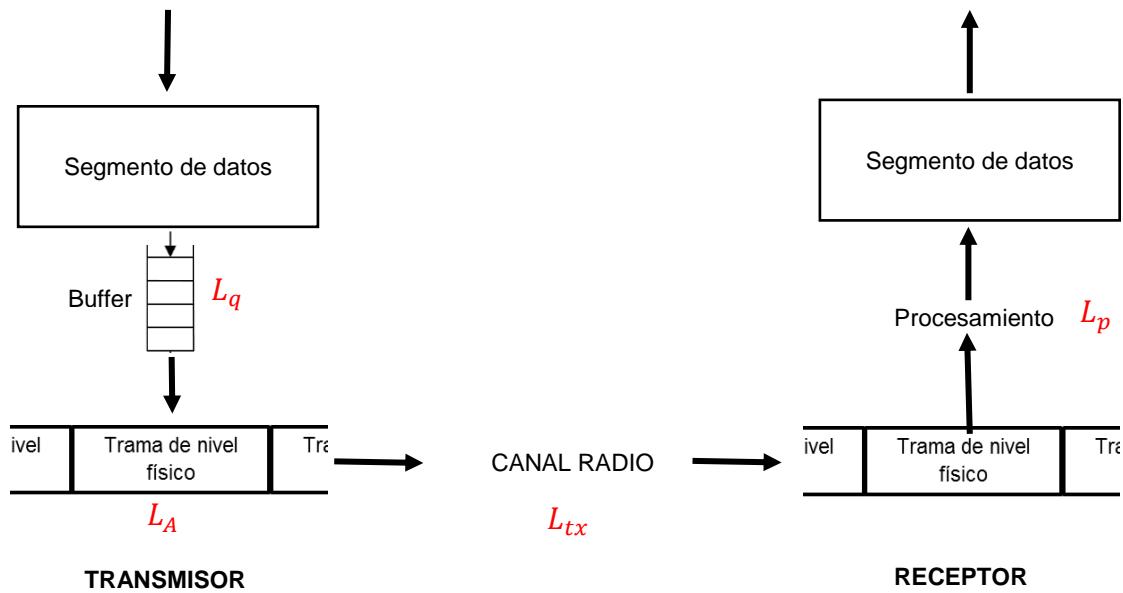


Figura 2.3. Proceso de transmisión en el nivel físico.

La latencia total en el nivel físico puede ser descrita como [32]:

$$L_{PHY} = L_q + L_A + L_{tx} + L_p, \quad (2.1)$$

donde,

L_q : el retardo de cola consiste en el tiempo que tardan los segmentos de datos en llegar al *buffer* de transmisión y ser transmitidos. El nivel de carga del sistema, la tasa de segmentos y el tamaño de los segmentos influyen directamente sobre este tipo de latencia.

L_A : el retardo de alineación de trama comprende el tiempo de espera de los segmentos de datos al siguiente TTI. La reducción del TTI es una solución obvia para disminuir este tipo de retardo.

L_{tx} : el retardo de transmisión consiste en el tiempo que tarda el nivel físico en colocar la trama de nivel físico en el canal radio.

L_p : el retardo de procesamiento en el receptor consiste en el tiempo que tarda el nivel físico del UE o gNB en procesar la trama de nivel físico recibida. Este retardo depende de la capacidad de procesamiento de los equipos radio.

2.2.2. Modelo de Simulación

Considerando los componentes de latencia definidos en la ecuación 2.1, se diseña el modelo de simulación de tal forma que permita modificar, en primer lugar, la carga del sistema. Para esto, se modela la RAN compuesta por un gNB sirviendo a un número determinado de UE, siendo el número de UE el que permita variar la carga del sistema. Los parámetros de tasa y tamaño de segmentos de datos caracterizarán los flujos de datos que simulan la información generada por diferentes servicios. Estos segmentos son transportados por niveles superiores mediante diferentes protocolos. Dado que el gNB no posee la capacidad de implementar dichos niveles, es necesario incluir dos componentes adicionales en el modelo de simulación. Por lo tanto, se incluye un servidor remoto que será utilizado para establecer una comunicación a nivel de transporte con los UE de la RAN y así simular los parámetros de tasa y tamaño de segmentos de datos que caracterizarán los flujos de datos entre el servidor y UE. Dado que dicho nodo no puede conectarse directamente a la RAN, se incluye en el modelo de sistema el CN cuyo único objetivo será proporcionar conectividad entre los UE y el servidor remoto. El modelo de simulación se muestra en la Figura 2.4.

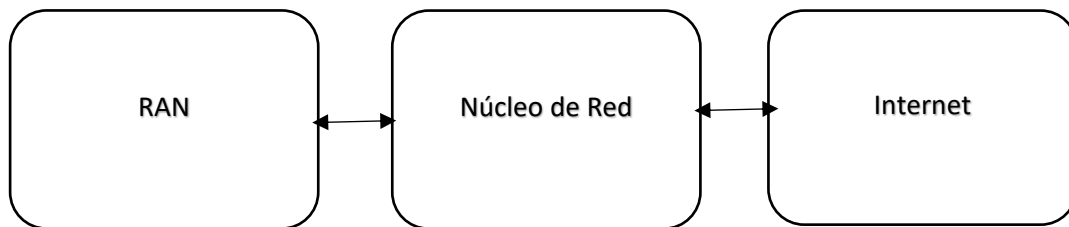


Figura 2.4. Modelo de Simulación.

2.2.3. Escenario de Simulación

El gNB proporciona servicio a UE ubicados en el interior de un edificio, en una misma planta. La planta tiene dimensiones de ancho y alto de 50 m y 50 m, respectivamente. Se inician flujos de segmentos desde los UE hacia el servidor remoto y viceversa.

Se modifica el valor del TTI en la trama de nivel físico implementando numerologías y mini-TTI. Así mismo se varían los parámetros de carga de sistema, tasa de segmentos, tamaño de procesamiento y el retardo de procesamiento para simular diferentes condiciones en el sistema. El escenario de simulación se muestra en la Figura 2.5. Los parámetros del modelo del sistema se muestran en la Tabla 2.2.

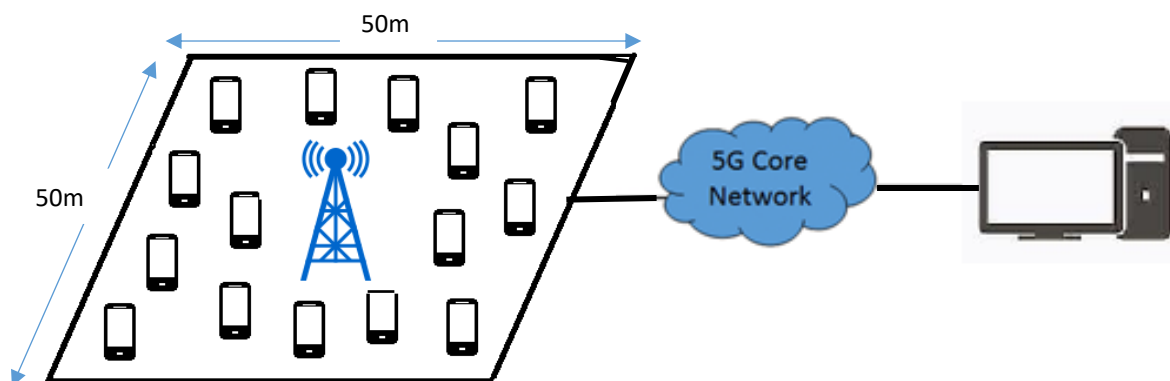


Figura 2.5. Escenario de Simulación.

Parámetro	Valor
Frecuencia central	28 GHz
Ancho de banda	100 MHz
Número de bloques de recursos radio	4
Subportadoras por bloques de recursos radio	12
Esquema de duplexación	TDD
Modulación	64 QAM
Codificación de Canal	LDPC
Potencia de Transmisión	23 dBm
Número de gNB	1
Número de UE	20, 40, 60
Tasa de segmentos	400, 900, 1400 segmentos/s
Tamaño de segmentos	500, 1500 Bytes
Tiempo de procesamiento	0 ms, 0.1 ms, 0.5 ms, 2TTI

Tabla 2.2. Parámetros del escenario de simulación.

Se implementa el modelo de canal descrito en la sección 2.1.5.2. de la siguiente forma:

- Escenario: se implementa un escenario de interiores, que consiste en una planta de un edificio con alta densidad de UE.
- Ubicación de UE y gNB: se ubican los UE aleatoriamente dentro del área. El gNB es ubicado en el centro del área, a una altura de 20 m.
- Velocidad de UE: los UE no poseen velocidad. Su posición es fija durante el 100% del tiempo.
- Obstrucción o bloqueo: no hay obstáculos entre el gNB y los UE. Existe línea de vista (LOS, *Line of Sight*) durante el 100% del tiempo.

De acuerdo con la R15, los tiempos de procesamiento dependen de la numerología y las capacidades de procesamiento de los equipos radio. Se toma como referencia el tiempo de procesamiento para sistemas LTE consistente en 2 ms [33]. Teniendo en cuenta que el TTI en dichos sistemas es de 1 ms, se considera el valor de 2 TTI para representar un tiempo de procesamiento dependiente de la numerología implementada. Así mismo, se consideran diferentes valores fijos que representan las capacidades de procesamiento de los equipos radio. Szczesny et al., en [34] presentan la implementación de un procesador con técnicas de aceleración de hardware que permiten incrementar significativamente la capacidad de procesamiento de protocolos LTE. En dicho artículo se varía el tamaño de los segmentos para evaluar su impacto en el tiempo de procesamiento. Para segmentos de tamaño de 125 bytes, se obtiene un tiempo de procesamiento mínimo y máximo de 0.1 ms y 0.5 ms, respectivamente, dependiendo del protocolo implementado. Estos valores son tomados como referencia, para evaluar el tiempo de procesamiento dependiente de la numerología. El retardo de procesamiento se tiene en cuenta únicamente en recepción.

2.3. IMPLEMENTACIÓN

En esta etapa se implementa el diseño de la sección 2.2.2 en la herramienta software seleccionada. Para esto, es necesario conocer las características de la herramienta y su funcionamiento.

2.3.1. Conceptos Principales en ns3

La herramienta software ns3 se basa en algunos conceptos de redes para facilitar su creación y simulación. Los conceptos más importantes son [35]:

- **Nodo:** representa un terminal, dispositivo o equipo que se conecta a una red.
- **Canal:** es el medio por el cual viaja la información.
- **Dispositivo de Red:** constituye la parte hardware y software que se instala en un nodo específico y que habilita la comunicación de este con otros nodos en la red a través del canal.
- **Aplicación:** constituyen los programas o aplicaciones que se ejecutan en los nodos y que generan segmentos de datos para la comunicación.

A continuación, se definen estos componentes en el contexto específico de un sistema 5G para la generación del código correspondiente.

2.3.2. Generación del Código en ns3

Teniendo en cuenta los componentes definidos en el apartado anterior, es posible implementar el sistema diseñado en la herramienta ns3 de la siguiente forma:

- **Nodos:** para simular la RAN, se crea un nodo que representa a la estación base y un conjunto de nodos que representan a los UE. El CN es representado por defecto en ns3 mediante 2 nodos, los cuales representan la Entidad de Gestión de Movilidad (MME, *Mobility Management Entity*), la

Pasarela de Servicio (SGW, *Serving Gateway*) y la Pasarela de Red de Paquetes de Datos (PGW, *Packet Data Network Gateway*). Finalmente, se crea un nodo adicional para simular el servidor remoto.

- **Canal:** el modelo de canal implementado por defecto en el módulo mmWave de ns3 es el definido en el reporte técnico 38.901 del 3GPP, el cual es el apropiado para sistemas 5G. Este componente se configura de acuerdo a los valores presentados en la Tabla 2.2.
- **Dispositivos de red:** se utilizan los dispositivos de red correspondientes a 5G implementados en el módulo mmWave y los dispositivos de red IP.
- **Aplicación:** se utiliza un generador de tráfico UDP que toma como parámetros la tasa y el tamaño de segmentos.

Una vez definidos los componentes, se procede a generar el código de la siguiente forma:

1. Se crea el objeto que representa al CN mediante el siguiente comando:

```
coreNetwork = CreateObject<MmWavePointToPointEpc> ();
```

2. Se crean dos contenedores para el nodo gNB y los nodos UE. El número de UE es almacenado en la variable ueNum:

```
NodeContainer gNBNodes;  
gNBNodes.Create (1);  
NodeContainer ueNodes;  
ueNodes.Create (ueNum);
```

3. Se instalan los dispositivos de red 5G del módulo mmWave en el gNB y los UE. De esta forma, el gNB y el CN tendrán conectividad:

```
mmWaveHelper->InstallEnbDevice (gNBNodes);  
mmWaveHelper->InstallUeDevice (ueNodes);
```

4. Es necesario especificar a qué gNB cada UE estará conectado, en este caso, se conectan todos los UE al gNB previamente creado:

```
mmWaveHelper->AttachToClosestEnb (ueNodes, gNBNodes);
```

5. En este punto el sistema 5G está configurado correctamente. Se crea el nodo que representa al servidor remoto y el enlace punto a punto entre dicho nodo y el CN:

```
NodeContainer remoteHostContainer;  
server = remoteHostContainer.Create (1);  
p2ph.Install (coreNetwork, server);
```

6. Se instalan los dispositivos de red IP en todos los nodos creados:

```
InternetStackHelper internet;  
internet.Install (remoteHostContainer);  
internet.Install (gNBNodes);  
internet.Install (ueNodes);
```

7. Finalmente, se crea un generador de segmentos UDP teniendo en cuenta los parámetros de tasa y tamaño de segmentos. Dichos valores se almacenan en las variables `interPacketInterval` y `packetSize`, respectivamente.

```
UdpClientHelper dlClient;  
dlClient.SetAttribute ("Interval", interPacketInterval);  
dlClient.SetAttribute ("PacketSize", packetSize);
```

2.4. PLAN DE PRUEBAS

En esta sección se definen los casos de estudio y las variables de entrada y salida que serán tenidas en cuenta para el desarrollo de las simulaciones y la evaluación y análisis de la latencia.

2.4.1. Diseño de Casos de Estudio

Teniendo en cuenta los componentes de la latencia en el nivel físico (ver Ecuación 2.1), es posible determinar que el TTI tiene un impacto directo sobre el retardo de alineación de trama y el retardo de transmisión.

Se toma como referencia el escenario de simulación definido en la Sección 2.2.3 para el diseño de los casos de estudio. Estos casos de estudio tienen en cuenta los parámetros de carga del sistema, tasa y tamaño de segmentos y capacidad de procesamiento. Dado que cada caso de estudio impone diferentes requerimientos, es posible analizar cuál es el valor del TTI adecuado para dichos casos de estudio y su impacto en la latencia. El retardo de procesamiento es analizado en un caso de estudio específico por lo que no es tenido en cuenta en los demás escenarios. Los casos de estudio se definen a continuación:

- **Caso de estudio 1:** este caso de estudio considera la carga del sistema, para esto se varía el número de UE asociados al gNB. Los parámetros de este caso de estudio se presentan en la siguiente tabla:

Número de UE	20, 40, 60
Tasa de Segmentos	400 segmentos/s
Tamaño de Segmentos	50 Bytes
Tiempo de Procesamiento	0 ms

Tabla 2.3. Parámetros del caso de estudio 1.

- **Caso de estudio 2:** en este caso de estudio se varía la tasa de flujo de segmentos. Los parámetros de este caso de estudio se presentan en la siguiente tabla:

Número de UE	20
Tasa de Segmentos	400, 900, 1400 segmentos/s
Tamaño de Segmentos	50 Bytes
Tiempo de Procesamiento	0 ms

Tabla 2.4. Parámetros del caso de estudio 2.

- **Caso de estudio 3:** en este caso de estudio se varía el tamaño de los segmentos. Los parámetros de este caso de estudio se presentan en la siguiente tabla:

Número de UE	20
Tasa de Segmentos	400 segmentos/s
Tamaño de Segmentos	50, 1500 Bytes
Tiempo de Procesamiento	0 ms

Tabla 2.5. Parámetros del caso de estudio 3.

- **Caso de estudio 4:** este caso de estudio considera los tiempos de procesamiento de los equipos radio. Para esto se implementan tiempos de procesamiento fijos y variable que depende del TTI. Los parámetros de este caso de estudio se presentan en la siguiente tabla:

Número de UE	20
Tasa de Segmentos	400 segmentos/s
Tamaño de Segmentos	50 Bytes
Tiempo de Procesamiento	0 ms, 0.1 ms, 0.5 ms, 2TTI

Tabla 2.6. Parámetros del caso de estudio 4.

En cada caso de estudio se varía el TTI, el espaciamiento entre subportadoras y la duración del símbolo de acuerdo a los valores definidos en la Tabla 1.3.

2.4.2. Variables de entrada y salida.

Las variables de entrada y salida se muestran en la Figura 2.6 y se describen a continuación:

Número de UE: esta variable tiene como objetivo variar la carga del sistema. Dado que los recursos radio son limitados, se espera que la variable afecte directamente la latencia en la RAN. Se escogen los valores de 20, 40, 60 UE para simular carga baja, media y alta, respectivamente.

Tasa de segmentos: esta variable define cuán rápido los segmentos de datos son enviados al nivel físico. Se escogen los valores de 400, 900 y 1400 segmentos/s para simular tasa baja, media y alta de segmentos, respectivamente.

Tamaño de segmentos: se escoge esta variable dado que el nivel físico debe segmentar los segmentos y adaptarlos a la trama de nivel físico. El tiempo de procesamiento en dicha tarea se verá reflejado en la latencia. Se escogen los valores de 500 y 1500 B para simular segmentos de datos pequeños y grandes, respectivamente

Tiempo de procesamiento: esta variable simula la capacidad de procesamiento de los equipos radio. Consiste en el tiempo que toma el gNB o UE en procesar la trama de nivel físico recibida. Se escogen los valores fijos de 0 ms, 0.1 ms, 0.5 ms para simular el caso ideal, procesamiento rápido y lento, respectivamente. También se utiliza un valor que depende del TTI, en este caso 2 TTI.

TTI: el TTI toma los valores de las diferentes numerologías presentadas en la Tabla 1.3. De esta forma se cuentan con cinco valores de TTI que son implementados en cada escenario. Así mismo, se utilizan mini-TTI con el objetivo de variar el número de símbolos por TTI.

Latencia en la RAN: consiste en el promedio de los valores de latencia de los segmentos cuando son transmitidos desde el gNB a los UE o viceversa.

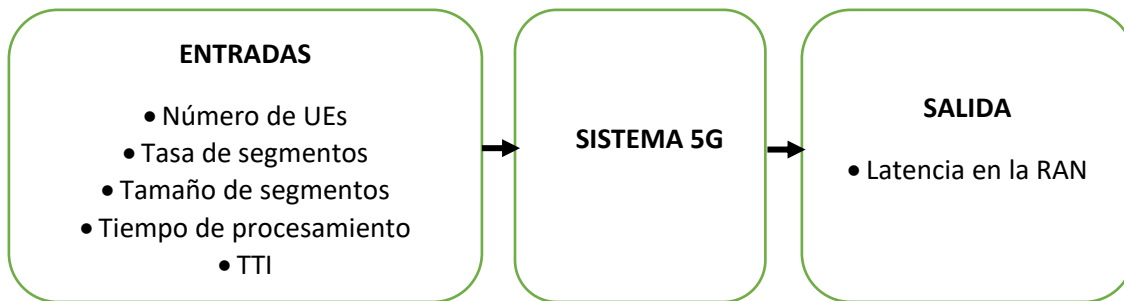


Figura 2.6. Variables de entrada y salida.

2.4.3. Ejecución de las simulaciones

Después de implementar cada caso de estudio se utiliza el método Montecarlo para obtener la tendencia estadística de los resultados. Para esto, se calcula la media y la desviación estándar de los resultados y se establece un margen de error de 10 μ s. Es posible obtener el número de veces que cada caso de estudio debe ser ejecutado de acuerdo al siguiente procedimiento:

1. Se calcula la media de la latencia de los segmentos cuando son transmitidos desde la gNB hacia los UE o viceversa, mediante la Ecuación 2.2:

$$M = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.2)$$

donde:

M : es la media de la latencia de los segmentos en la RAN.

n : es el número de segmentos transmitidos.

x_i : es la latencia del segmento i en la RAN.

2. Se calcula la desviación estándar de la latencia de los segmentos cuando son transmitidos desde la gNB hacia los UE o viceversa, mediante la Ecuación 2.3:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - M)^2 \quad (2.3)$$

donde:

s : es la desviación estándar.

M : es la media de la latencia de los segmentos en la RAN.

n : es el número de segmentos transmitidos.

x_i : es la latencia del segmento i en la RAN.

3. Se calcula el número de simulaciones necesarias para proporcionar un margen de error e mediante la Ecuación 2.4:

$$e = s/\sqrt{N} \quad (2.4)$$

donde:

e : es el margen de error.

s : es la desviación estándar.

N : es el número de simulaciones requeridas.

Dado que en cada caso de estudio se varía un parámetro de acuerdo a dos o más valores, es necesario realizar los cálculos descritos anteriormente para cada valor del parámetro considerado y cada numerología. En las Tablas 2.7 y 2.8 se muestran los valores de la desviación estándar para cada caso de estudio en el UL y DL, respectivamente.

		Numerología				
		1	2	3	4	5
Caso de estudio 1	20 UE	49.4	53.2	52.3	54.6	59.7
	40 UE	77.3	79.5	83.1	82.5	87.8
	60 UE	88.4	89.7	90.6	97.9	98.9
Caso de estudio 2	400 Seg/s	51.7	52.2	56.4	55.6	59.8
	900 Seg/s	58.4	59.5	63.3	67.2	72.2
	1400 Seg/s	61.5	63.1	67.3	72.7	75.4
Caso de estudio 3	50 Bytes	45.3	47.2	41.2	42.4	44.6
	1500 Bytes	46.4	46.7	48.6	49.3	41.8
Caso de estudio 4	0, 0.1,0.5 ms	43.1	45.5	47.1	46.7	45.7
	2 TTI	45.2	47.3	44.6	46.3	48.6

Tabla 2.7. Desviación estándar de resultados de casos de estudio en el UL.

		Numerología				
		1	2	3	4	5
Caso de estudio 1	20 UE	39.7	41.2	46.9	48.3	53.1
	40 UE	76.4	75.4	79.7	70.9	76.9
	60 UE	87.5	86.1	88.4	91.9	94.9
Caso de estudio 2	400 Seg/s	42.7	43.2	45.3	48.3	57.5
	900 Seg/s	43.4	46.4	44.7	52.8	54.4
	1400 Seg/s	45.5	49.1	52.4	56.5	63.2
Caso de estudio 3	50 bytes	42.3	41.4	42.2	44.5	45.7
	1500 bytes	46.4	45.2	47.3	48.2	47.1
Caso de estudio 4	0, 0.1,0.5 ms	44.1	43.2	44.6	45.4	46.8
	2 TTI	45.2	43.1	45.7	44.2	46.2

Tabla 2.8. Desviación estándar de resultados de casos de estudio en el DL.

Seguidamente se calcula el número de ejecuciones de acuerdo a los valores de desviación estándar mostrados en las Tablas 2.7 y 2.8 para proporcionar un margen de error de 10 μ s. En las Tablas 2.9 y 2.10 se muestran el número de ejecuciones para cada caso de estudio en el UL y DL, respectivamente.

		Numerología				
		1	2	3	4	5
Caso de estudio 1	20 UE	24	28	27	29	35
	40 UE	59	63	69	68	77
	60 UE	78	80	82	95	97
Caso de estudio 2	400 Seg/s	26	27	31	30	35
	900 Seg/s	34	35	40	45	52
	1400 Seg/s	37	39	45	52	56
Caso de estudio 3	50 Bytes	20	22	16	17	19
	1500 Bytes	21	21	23	24	17
Caso de estudio 4	0, 0.1,0.5 ms	18	20	22	21	20
	2 TTI	20	22	19	21	23

Tabla 2.9. Número de ejecuciones para cada caso de estudio en el UL.

		Numerología				
		1	2	3	4	5
Caso de estudio 1	20 UE	15	16	21	23	28
	40 UE	58	56	63	50	59
	60 UE	76	74	78	84	90
Caso de estudio 2	400 Seg/s	18	18	20	23	33
	900 Seg/s	18	21	19	27	29
	1400 Seg/s	20	24	27	31	39
Caso de estudio 3	50 Bytes	17	17	17	19	20
	1500 Bytes	21	20	22	23	22
Caso de estudio 4	0, 0.1,0.5 ms	19	18	19	20	21
	2 TTI	20	18	20	19	21

Tabla 2.10. Número de ejecuciones para cada caso de estudio en el DL.

CAPITULO 3: ANÁLISIS DE RESULTADOS

En esta sección se presentan los resultados y el análisis de las pruebas diseñadas en el anterior capítulo.

3.1. CASO DE ESTUDIO 1

En este caso de estudio se varía la carga en el sistema y se consideran diferentes valores del TTI, para analizar su impacto sobre la latencia en la RAN. Existen dos formas en las que el sistema 5G puede aprovechar la flexibilidad de la trama NR para variar el valor del TTI, se estudia cada caso de forma independiente:

1. **Variante 1:** el sistema 5G puede utilizar las numerologías definidas en la Tabla 1.3. Al utilizar una numerología superior, la separación entre subportadoras será mayor y por consiguiente la duración del símbolo menor, como se muestra en la Figura 1.8. De esta forma se reduce el TTI al aumentar la separación entre subportadoras. El TTI está compuesto por 14 símbolos en todos los casos.
2. **Variante 2:** el sistema 5G puede hacer uso de mini-TTI, esto es, emplear una cantidad menor a 14 símbolos por TTI para disminuir su duración, sin necesidad de variar el espaciamiento entre subportadoras.

3.1.1. Variante 1

Para evaluar la primera variante se simula un sistema 5G con diferentes valores de carga. La carga del sistema depende del número de UE asociados al gNB. Se simulan 20, 40, 60 UE que corresponden a carga baja, media y alta, respectivamente. Así mismo, se implementan las numerologías presentadas en la Tabla 1.3 para cada caso. La latencia en la RAN para el UL y DL se muestra en las Figuras 3.1 y 3.2, respectivamente.

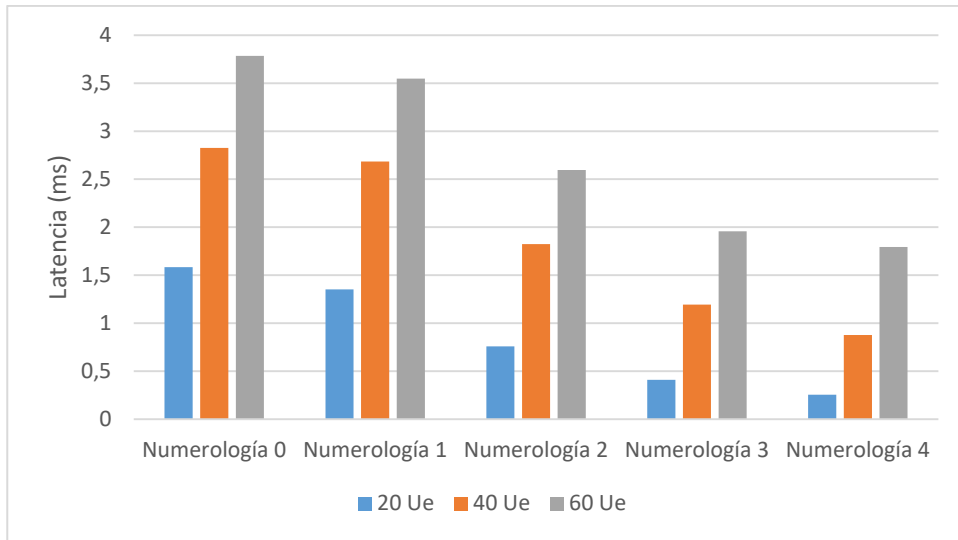


Figura 3.1. Latencia en la RAN en el UL para diferentes numerologías.

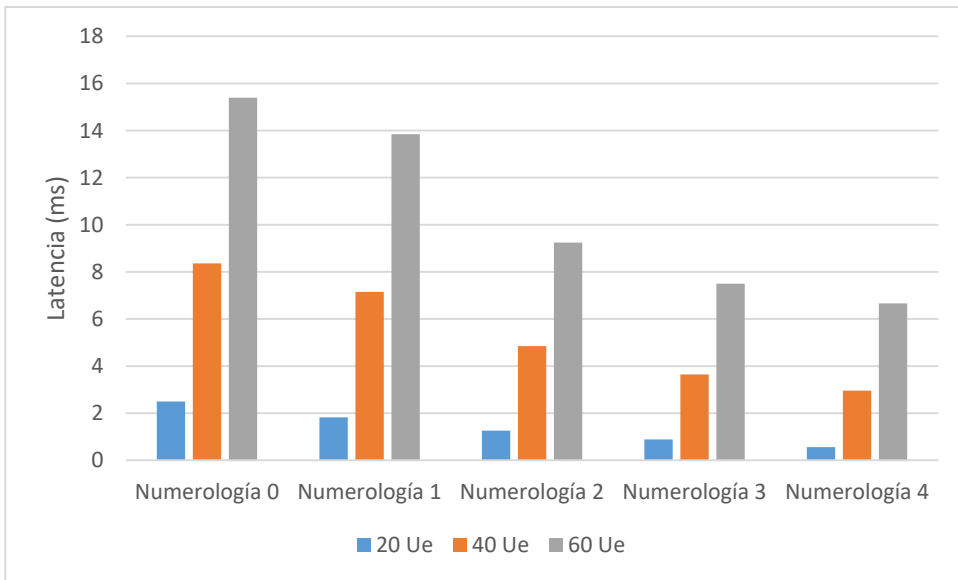


Figura 3.2. Latencia en la RAN en el DL para diferentes numerologías.

Se observa que la latencia en la RAN se incrementa al aumentar la carga del sistema. Es posible explicar esta tendencia analizando el proceso de distribución de TTI por parte del gNB:

- Los segmentos de datos son asignados en TTI para su transmisión. A cada usuario se le asigna un TTI determinado, dependiendo de la carga del sistema. Si existiese un único usuario, se le asignarían TTI consecutivos

teniendo una transmisión continua e ininterrumpida (ver Figura 3.3 a). Si existen N usuarios, cada usuario tendrá que esperar $N-1$ TTI para que se le sea asignado un TTI nuevamente y así continuar con la transmisión de su información (ver Figura 3.3 b). De esta forma, el incremento en la carga aumenta el intervalo de espera de cada usuario y por consiguiente la latencia en la RAN se incrementa.

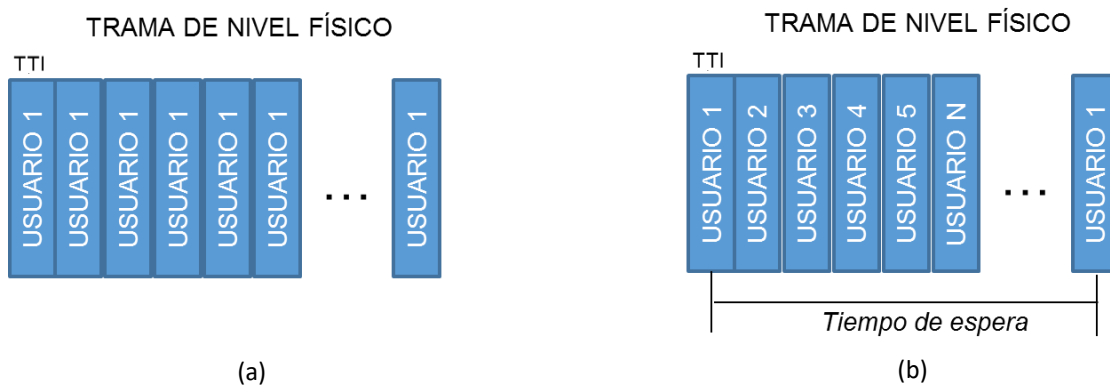


Figura 3.3. Tiempo de espera de asignación de TTI. a) Un usuario b) N usuarios.

- Los segmentos que llegan al *buffer* del gNB deben esperar un tiempo adicional para ser transmitidos hacia los UE debido al proceso anteriormente descrito. Cada segmento puede ser transmitido hasta que todos los segmentos previos a él sean transmitidos, causando que la latencia en la RAN se incremente aún más (ver Figura 3.4). De esta manera, el retardo de cola constituye el componente predominante en la latencia de la RAN para los casos de carga media y alta.

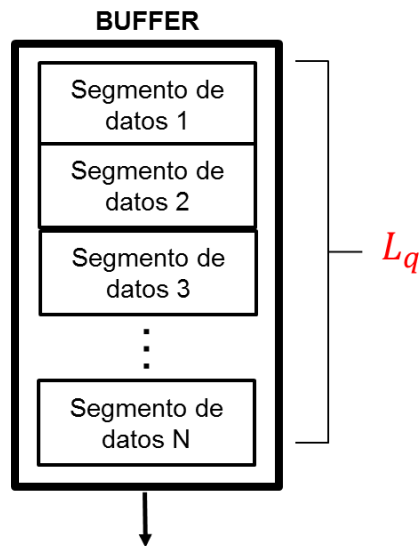


Figura 3.4. Retardo de cola debido a la carga del sistema.

- La latencia en la RAN para el DL es mayor dado que el retardo de cola es superior en el gNB. Así mismo, el gNB debe realizar la planificación de segmentos (*scheduling*) para cada usuario, por lo que al incrementar la carga este proceso tomará más tiempo.

Al implementar numerologías superiores, es decir al disminuir el valor del TTI, se aprecia una reducción significativa en la latencia de la RAN. Teniendo en cuenta la ecuación 2.1, es posible analizar este efecto de la siguiente manera:

- Al disminuir el valor del TTI, el tiempo que cada usuario debe esperar para que se le asigne un TTI disminuye también, reduciendo la latencia en la RAN (ver Figura 3.5).
- El retardo de alineación disminuye proporcionalmente con el valor del TTI. Este retardo es cuantificado como la mitad del valor del TTI, de esta forma cada segmento debe esperar cada vez menos a la siguiente oportunidad de transmisión (ver Figura 3.6).
- El retardo de transmisión es igual al valor del TTI, por lo que disminuye en la misma cantidad (ver Figura 3.7).

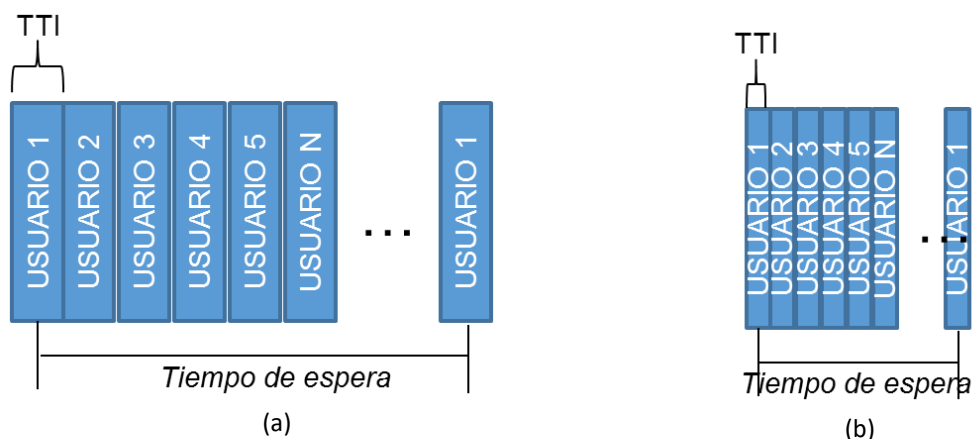


Figura 3.5. Tiempo de espera de asignación de TTI. a) TTI grande b) TTI corto.



Figura 3.6. Retardo de alineación. a) TTI grande b) TTI corto.

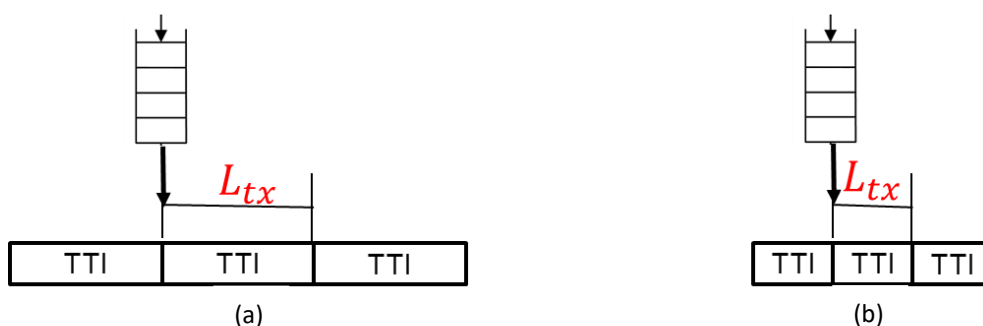


Figura 3.7. Retardo de transmisión. a) TTI grande b) TTI corto.

- Si el sistema tiene una carga baja, el uso de TTI cortos disminuye los retardos presentados en la ecuación 2.1, reduciendo la latencia de la RAN. Sin embargo, al aumentar la carga, el retardo de cola crece en una mayor proporción, superando significativamente la reducción que el TTI obtiene en el retardo de alineación, transmisión y de asignación de TTI. Dado que el

retardo de cola es la componente predominante, las ventajas del uso de TTI cortos desaparecen en el caso de carga alta del sistema.

Seguidamente, se cuantifica la reducción en la latencia en la RAN, que cada numerología proporciona. Para esto, se calcula el porcentaje de reducción en la latencia de la RAN de las numerologías superiores con respecto a la numerología 0. Dicha numerología establece una separación de subportadoras de 15 KHz y una duración de trama de 1 ms, parámetros que son utilizados en los sistemas LTE. Esta información se muestra en las Tablas 3.1 y 3.2 para el UL y DL, respectivamente.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
Carga Baja	14.64 %	52.15 %	74.12 %	83.89 %
Carga Media	4.99 %	35.46 %	57.75 %	68.96 %
Carga Alta	3.25 %	31.44 %	48.28 %	52.59 %

Tabla 3.1. Reducción de la latencia en la RAN en el UL para diferentes numerologías.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
Carga Baja	26.91 %	49.71 %	64.47 %	77.67 %
Carga Media	14.47 %	42.03 %	56.38 %	64.59 %
Carga Alta	10.06 %	39.93 %	51.31 %	56.74 %

Tabla 3.2. Reducción de la latencia en la RAN en el DL para diferentes numerologías.

3.1.2. Variante 2

Las pruebas de la sección 3.1.1 fueron realizadas teniendo en cuenta un número fijo de símbolos por TTI, en este caso 14 símbolos. Para la segunda variante, se varía el número de símbolos por TTI sin modificar el espaciamiento entre subportadoras. Se utiliza la numerología 0, es decir un espaciamiento fijo de 15 KHz. Se consideran los valores de 2, 4 y 7 símbolos por TTI que son comparados con el valor estándar de 14 símbolos por TTI. Se utilizan los valores de 20, 40 y 60 UE para simular carga baja, media y alta, respectivamente. La latencia en la RAN para el UL y DL se muestra en las Figuras 3.8 y 3.9, respectivamente.

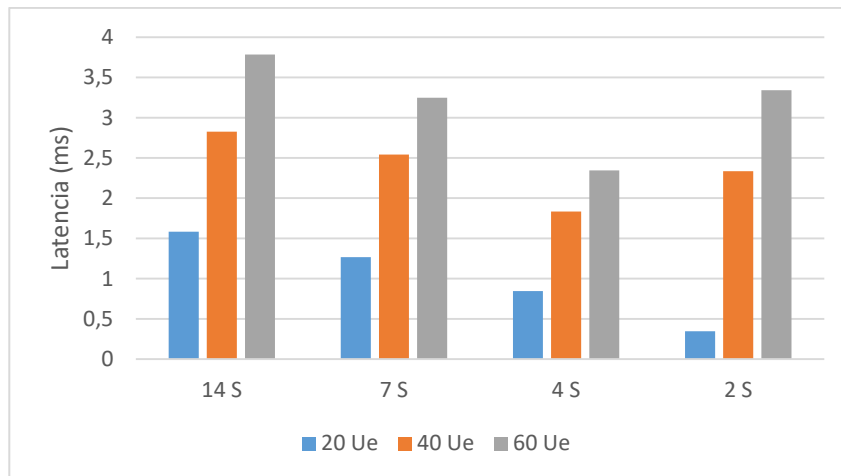


Figura 3.8. Latencia en la RAN en el UL para 14, 7, 4 y 2 símbolos por TTI.

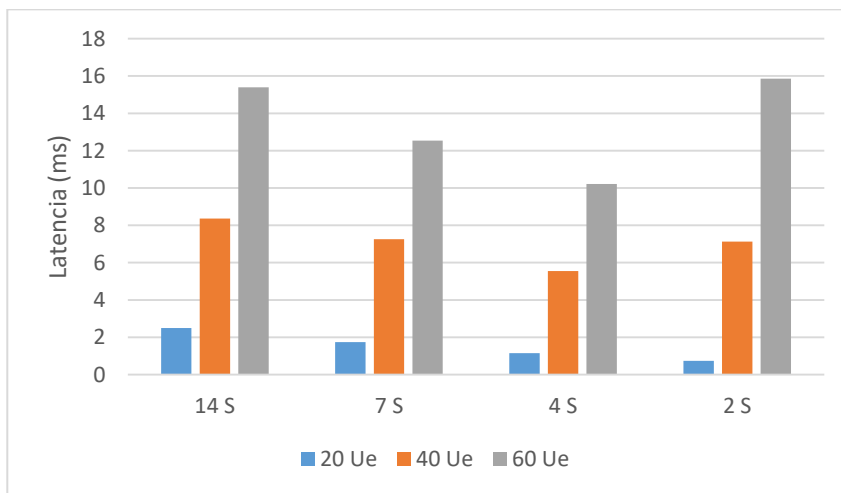


Figura 3.9. Latencia en la RAN en el DL para 14, 7, 4 y 2 símbolos por TTI.

Se observa que la reducción del TTI en este caso también impacta la latencia de la RAN, sin embargo, no posee el mismo efecto que en la variante anterior. De las Figuras 3.8 y 3.9 se puede observar que:

- La carga del sistema afecta negativamente la latencia como en la primera variante. Las razones que explican este efecto fueron descritas anteriormente.

- Considerando la carga baja en el sistema, la reducción del TTI disminuye la latencia en la RAN, siendo 2 símbolos por TTI el valor que proporciona menor latencia.
- Considerando carga media y alta en el sistema, la reducción del TTI no siempre implica disminución de la latencia en la RAN. Es posible observar que los valores de 7 y 4 símbolos por TTI disminuyen la latencia en comparación con el valor de 14 símbolos por TTI, sin embargo, al utilizar el valor de 2 símbolos por TTI la latencia en la RAN se deteriora significativamente. Para estos niveles de carga, el valor de 4 símbolos por TTI proporciona menor latencia.

Esta tendencia puede ser explicada al tener en cuenta el *overhead*¹ introducido por las señales de referencia que el Sistema 5G necesita para operar correctamente. Dichas señales proporcionan información sobre el estado del canal radio y la red, y son transmitidas tanto en el UL como en el DL. La señal de Referencia para Demodulación (DMRS, *Demodulation Reference Signal*), por ejemplo, ocupa dos símbolos del TTI en tres subportadoras de cada bloque de recursos radio. Es posible analizar el *overhead* introducido por esta señal al graficar bloques de recursos radio con diferentes valores de TTI, tal como se muestra en la Figura 3.10.

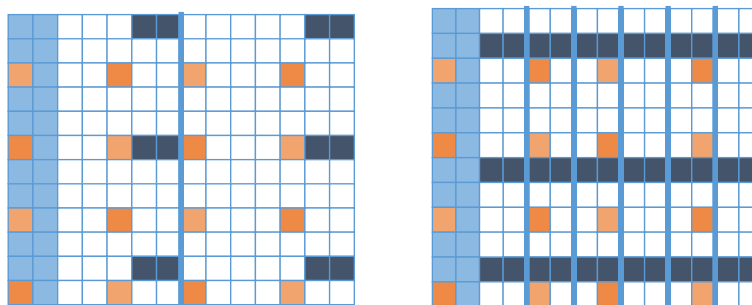


Figura 3.10. Ubicación de la señal DMRS en TTI de 7 y 2 símbolos. Tomada de [36].

¹ *Overhead* hace referencia a la información de señalización y control que es necesaria para llevar a cabo el proceso de comunicación en un sistema de telecomunicaciones.

Para el caso de 7 símbolos por TTI, el *overhead* introducido por la señal DMRS es del 7.14 %. Para el TTI formado por 2 símbolos, el *overhead* corresponde al 25%. Al reducir el número de símbolos por TTI, el *overhead* se incrementa, por consiguiente, los segmentos deben esperar TTI adicionales para su transmisión, incrementando el retardo de cola. Al considerar sistemas con carga media y alta, el retardo de cola influencia la latencia total de la RAN, por lo que al reducir el número de TTI disponibles para la transmisión de segmentos y asignarlos a la información de control, la latencia se deteriora significativamente. En las Tablas 3.3 y 3.4 se muestra el porcentaje de reducción de la latencia de la RAN considerando 7, 4 y 2 símbolos por TTI en el UL y DL, respectivamente.

	7 símbolos	4 símbolos	2 símbolos
Carga Baja	19.94 %	46.54 %	78.16 %
Carga Media	10.02 %	35.10 %	17.37 %
Carga Alta	14.17 %	38.03 %	11.70 %

Tabla 3.3. Reducción de la latencia en la RAN en el UL para diferentes valores de símbolos por TTI.

	7 símbolos	4 símbolos	2 símbolos
Carga Baja	30.10 %	54.06 %	70.57 %
Carga Media	13.23 %	33.66 %	14.78 %
Carga Alta	18.52 %	33.64 %	- 3.00 %

Tabla 3.4. Reducción de la latencia en la RAN en el UL para diferentes valores de símbolos por TTI.

Es posible concluir que la reducción del TTI, ya sea implementando diferentes numerologías o reduciendo el número de símbolos por TTI posee un impacto significativo sobre la latencia de la RAN. La utilización de una u otra alternativa para la reducción de la latencia dependerá de los requisitos de la red. Las numerologías altas exigen un mayor ancho de banda y una frecuencia de operación alta, por lo cual es necesario considerar los efectos de los fenómenos de propagación en ondas de frecuencias altas. Así mismo es necesario considerar la carga del sistema y el posible *overhead* introducido por señales de referencia y/o protocolos de red.

3.2. CASO DE ESTUDIO 2

En este caso de estudio se varía el flujo de segmentos, para esto, se consideran los valores de 400, 900 y 1400 segmentos/s que representan una tasa baja, media y alta de segmentos, respectivamente. Se implementan las numerologías definidas en la Tabla 1.3 para cada caso. La latencia en la RAN para el UL y DL se muestra en las Figuras 3.11 y 3.12, respectivamente.

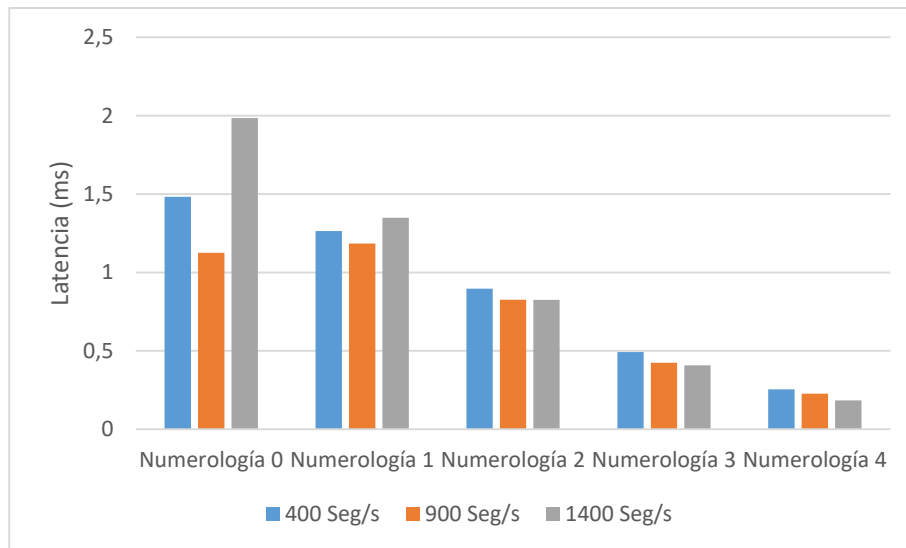


Figura 3.11. Latencia en la RAN en el UL para 400, 900 y 1400 segmentos por segundo.

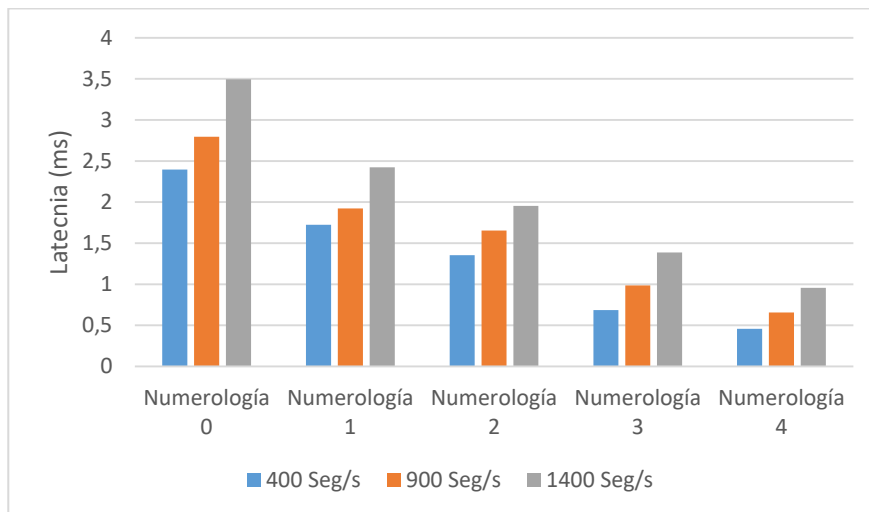


Figura 3.12. Latencia en la RAN en el DL para 400, 900 y 1400 segmentos por segundo.

De las Figuras 3.11 y 3.12 es posible observar que:

- En el DL, el incremento en la tasa de segmentos afecta negativamente la latencia en la RAN. Esta tendencia puede ser explicada debido a que, al aumentar la tasa de segmentos, el *buffer* del gNB se carga más rápidamente, incrementando el retardo de cola. Los segmentos que llegan al *buffer* de transmisión tienen que esperar un tiempo mayor para ser transmitidos. Al reducir el TTI, los segmentos que esperan en el *buffer* de transmisión son transmitidos más rápidamente, reduciendo el retardo de cola y por lo tanto la latencia en la RAN.
- En el UL, el incremento en la tasa de segmentos no posee el mismo efecto que en el DL. Es posible observar, que a diferencia del DL, no es el valor de 400 segmentos/s el que proporciona menor latencia. Este resultado puede ser explicado al analizar el proceso de transmisión de datos en el UL.

En el UL, una vez que el UE posee datos disponibles para su transmisión, envía una Solicitud de Asignación de Recursos (SR, *Scheduling Request*) al gNB. Al recibir dicho mensaje, el gNB responde con un Mensaje de Concesión (UG, *Uplink Grant*) indicando los recursos asignados al UE y la configuración para la transmisión. Dicho proceso se muestra en la Figura 3.13.

El proceso de asignación de recursos en el UL toma más tiempo que en el DL, en el cual el gNB asigna los recursos correspondientes inmediatamente. En el UL, el UE debe esperar a la siguiente oportunidad de transmisión para enviar el SR, dicha transmisión durará por lo menos 1ms. Seguidamente, el gNB debe procesar el SR recibido para luego transmitir el UG en 1ms. El UE finalmente procesa el UG e inicia la transmisión de los datos correspondientes. Este proceso incrementa significativamente la latencia en la RAN, si el UE necesita enviar SRs frecuentemente para acceder a los recursos del sistema.

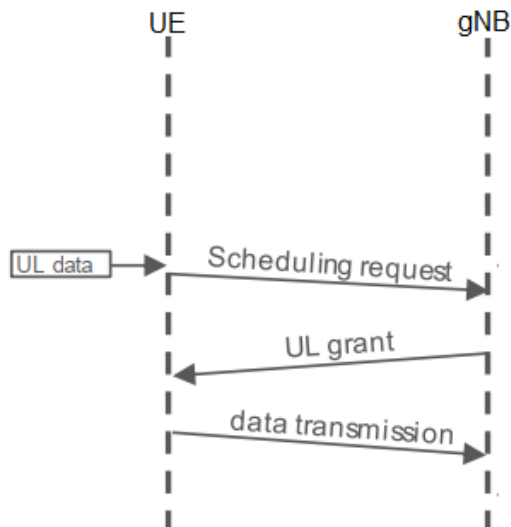


Figura 3.13. Proceso de Transmisión en el UL. Tomada de [37].

El *buffer* de transmisión se llena con los segmentos de datos del UE y este inicia el proceso anteriormente descrito mediante el envío de un SR para la transmisión de dichos datos. Todos los segmentos que están en el *buffer* pueden ser transmitidos usando el mismo SR (ver Figura 3.14 a), pero si el *buffer* está vacío, el UE debe enviar un nuevo SR y empezar el proceso de acceso a recursos nuevamente (ver Figura 3.14 b).

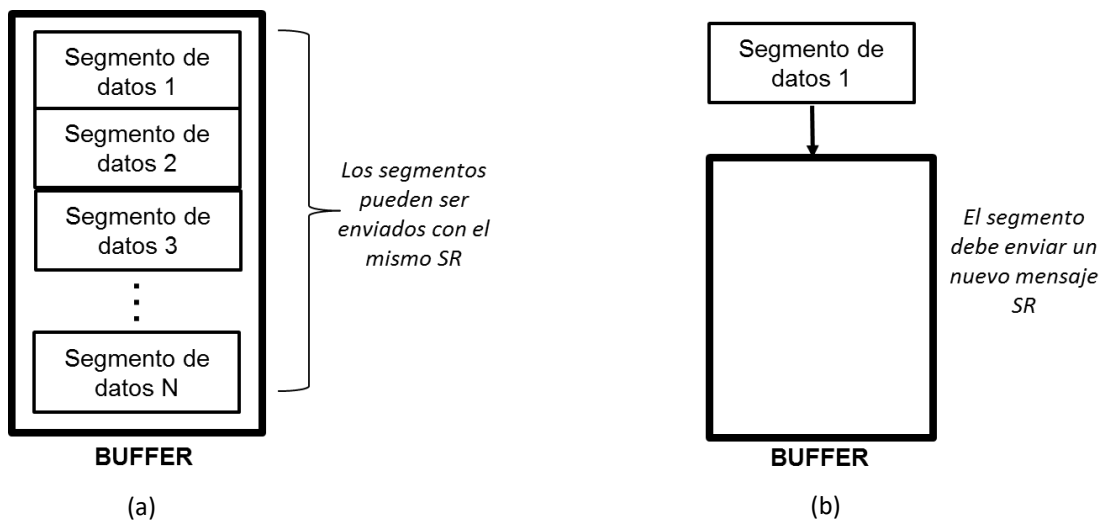


Figura 3.14. Envío de mensajes SR según contenido del buffer de transmisión. a) buffer lleno b) buffer vacío.

Considerando la numerología 0, la tasa media de segmentos (900 segmentos/s) implica el envío de 1 segmento cada milisegundo aproximadamente, valor que disminuye la probabilidad de que cada segmento encuentre un *buffer* vacío, si se compara con la tasa baja de segmentos. De esta forma, para la tasa media de segmentos, el UE envía mensajes SR con menor frecuencia, disminuyendo la latencia en la RAN. Sin embargo, al considerar la tasa alta de segmentos, el *buffer* de transmisión no es capaz de vaciar los segmentos con la misma velocidad con la que llegan, incrementando el retardo de cola y deteriorando la latencia en la RAN.

Para las numerologías superiores (2,3,4), la velocidad de transmisión se incrementa significativamente, esto significa que el *buffer* se vacía con mayor rapidez, incrementando la probabilidad de envíos de mensajes SR. Es posible notar que las tasas baja y media de segmentos proporcionan mayor latencia en este caso, debido al efecto descrito anteriormente. En las Tablas 3.5 y 3.6 se muestra la reducción de latencia en la RAN para las tasas baja, media y alta de segmentos en el UL y DL, respectivamente.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
Tasa Baja	14.71 %	39.54 %	66.79 %	82.80 %
Tasa Media	- 5.23 %	26.55 %	62.36 %	79.85 %
Tasa Alta	32.07 %	58.45 %	79.46 %	90.71 %

Tabla 3.5. Reducción de la latencia en la RAN en el UL para diferentes niveles de tasa de segmentos.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
Tasa Baja	28.03 %	43.43 %	71.33 %	80.91 %
Tasa Media	31.17 %	40.79 %	64.70 %	76.49 %
Tasa Alta	30.65 %	44.07 %	60.33 %	72.61 %

Tabla 3.6. Reducción de la latencia en la RAN en el DL para diferentes niveles de tasa de segmentos.

3.3. CASO DE ESTUDIO 3

En este caso de estudio se varía el tamaño de segmentos, considerando los valores de 50 bytes y 1500 bytes que representan segmentos pequeños y grandes, respectivamente. Se implementan las numerologías definidas en la Tabla 1.3. La latencia en la RAN para el UL y DL se muestra en las Figuras 3.15 y 3.16, respectivamente.

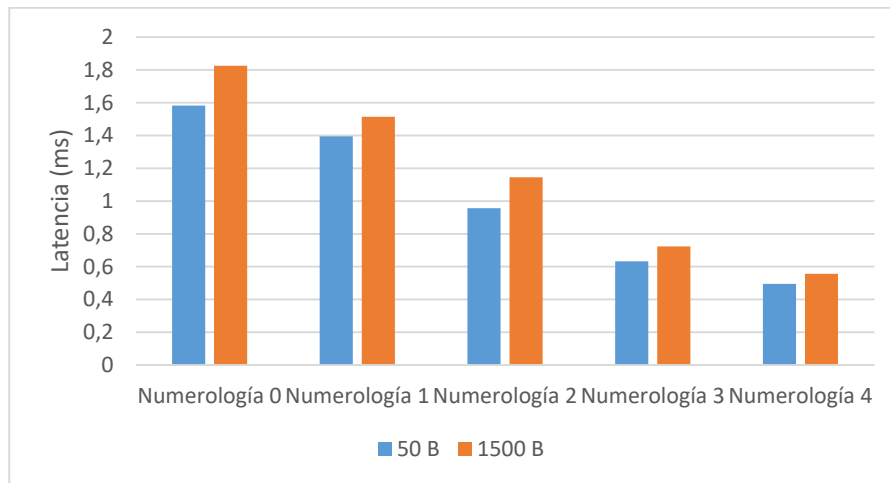


Figura 3.15. Latencia en la RAN en el UL para segmentos de 50 Bytes y 1500 Bytes.

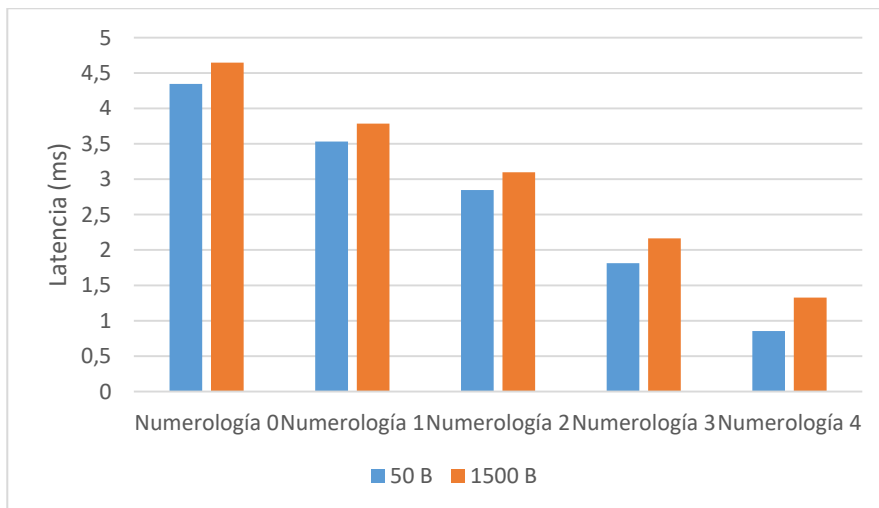


Figura 3.16. Latencia en la RAN en el DL para segmentos de 50 Bytes y 1500 Bytes. Es posible observar que en el UL y DL el tamaño de segmentos afecta negativamente la latencia. Dicha tendencia es explicada debido a que segmentos

más grandes requieren mayor cantidad de procesamiento para su transmisión. Así mismo, se asignan un mayor número de TTI para transmitir segmentos más grandes (ver Figura 3.17), por lo que el *buffer* de transmisión se llenará con más rapidez, aumentando el retardo de cola y la latencia en la RAN.

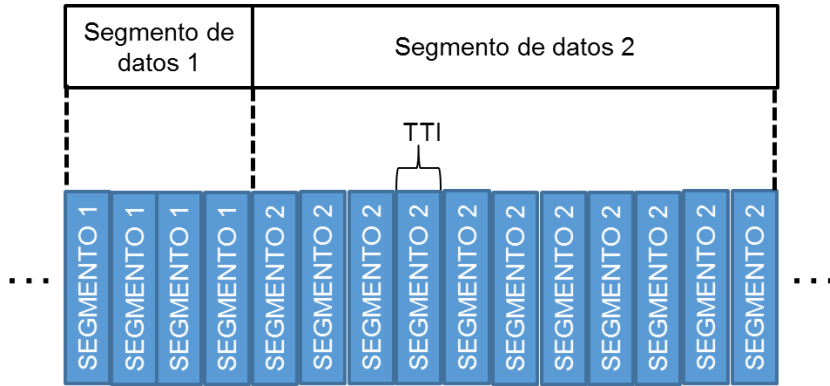


Figura 3.17. Asignación de TTI según el tamaño de segmentos.

La reducción de la latencia en la RAN en el UL y DL para segmentos pequeños y grandes se muestra en las Tablas 3.7 y 3.8, respectivamente.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
50 Bytes	11.89 %	39.57 %	60.04 %	68.73 %
1500 Bytes	17.04 %	37.25 %	60.36 %	69.50 %

Tabla 3.7. Reducción de la latencia en la RAN en el UL para segmentos pequeños y grandes.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
50 Bytes	18.71 %	34.52 %	58.29 %	80.32 %
1500 Bytes	18.53 %	33.34 %	53.42 %	71.43 %

Tabla 3.8. Reducción de la latencia en la RAN en el UL para segmentos pequeños y grandes.

3.4. CASO DE ESTUDIO 4

En este caso de estudio se varía el retardo de procesamiento del UE y el gNB. Para esto, se consideran los valores de 0, 0.1 y 0.5 ms que representan el caso ideal,

procesamiento rápido y lento, respectivamente. También se considera un valor que depende del TTI, en este caso 2 TTI. Así mismo, se implementan las numerologías definidas en la Tabla 1.3. La latencia en la RAN para el UL y DL se muestra en las Figuras 3.18 y 3.19, respectivamente.

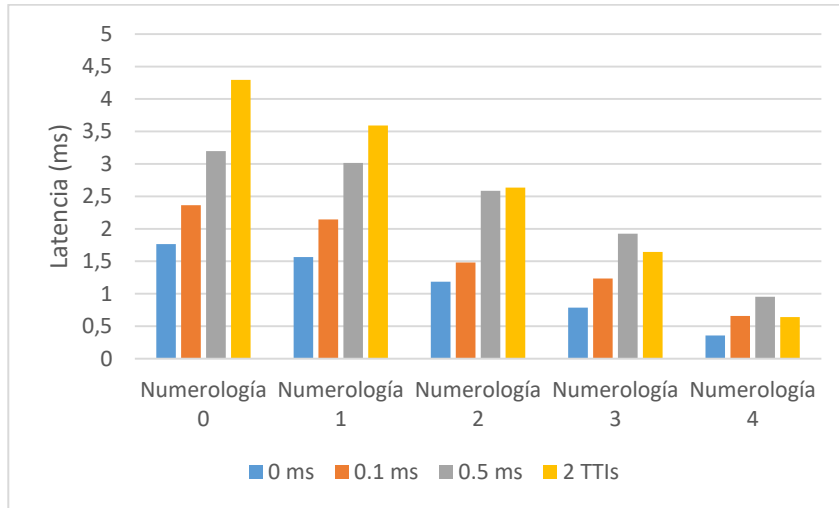


Figura 3.18. Latencia en la RAN en el UL para diferentes valores de tiempo de procesamiento.

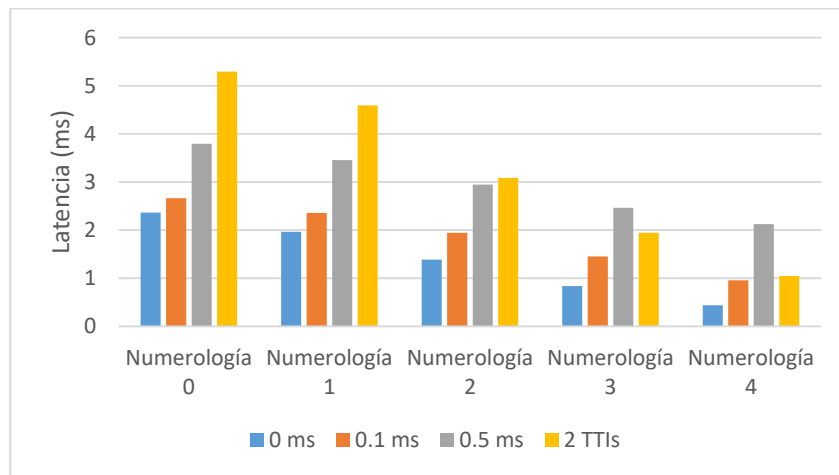


Figura 3.19. Latencia en la RAN en el DL para diferentes valores de tiempo de procesamiento.

De las Figuras 3.18 y 3.19 es posible observar que:

- La capacidad de procesamiento de los equipos radio influye significativamente sobre la latencia de la red. Si se toma como base el caso ideal (0 ms), el retardo de procesamiento puede duplicar la latencia de la

RAN en algunos casos. En la numerología 0, por ejemplo, el retardo de procesamiento de 2 TTI es aproximadamente el doble de la latencia en condiciones ideales.

- Para las numerologías 0, 1 y 2, el valor que proporciona mayor latencia es el de 2 TTI. Esto se debe a que, para estas numerologías, el TTI es relativamente grande, por lo que el valor de 2 TTI supera a los valores fijos considerados.
- Para las numerologías 3 y 4, el TTI se reduce significativamente, por lo que el valor de 2 TTI constituye ahora un tiempo de procesamiento rápido. En este caso dicho valor es menor que el tiempo de procesamiento lento y aproximadamente igual al tiempo de procesamiento rápido.

La reducción de la latencia en la RAN en el UL y DL para diferentes tiempos de procesamiento se muestra en las Tablas 3.9 y 3.10, respectivamente.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
0 ms	11.33 %	32.76 %	55.41 %	79.75 %
0.1 ms	9.30 %	37.30 %	47.78 %	72.20 %
0.5 ms	5.63 %	19.06 %	39.75 %	70.16 %
2 TTI	16.36 %	38.64 %	61.73 %	85.07 %

Tabla 3.9. Reducción de la latencia en la RAN en el UL para diferentes valores de tiempo de procesamiento.

	Numerología 1	Numerología 2	Numerología 3	Numerología 4
0 ms	16.91 %	41.36 %	64.61 %	81.51 %
0.1 ms	11.55 %	26.94 %	45.48 %	64.10 %
0.5 ms	9.04 %	22.41 %	35.13 %	44.34 %
2 TTI	13.27 %	41.69 %	63.28 %	80.25 %

Tabla 3.10. Reducción de la latencia en la RAN en el UL para diferentes valores de tiempo de procesamiento.

Para finalizar este capítulo, se presenta en la Tabla 3.11 un resumen del impacto del TTI en la latencia de la RAN en el UL. Se consideran el proceso de acceso a recursos mostrado en la Figura 3.13 y los valores del TTI definidos en el la Tabla

1.3. La Tabla 3.11 muestra las ventajas de la utilización de TTI cortos en la latencia de la RAN. El retardo de procesamiento es igual a 2 TTI en este caso.

TTI (ms)	1	0.5	0.25	0.125	0.0625
Retardo de alineación (ms)	0.5	0.25	0.125	0.0625	0.03125
Retardo de transmisión del mensaje SR (ms)	1	0.5	0.25	0.125	0.0625
Retardo de procesamiento en el gNB (ms)	2	1	0.5	0.25	0.125
Retardo de transmisión del mensaje UG (ms)	1	0.5	0.25	0.125	0.0625
Retardo de procesamiento en el UE (ms)	2	1	0.5	0.25	0.125
Retardo de alineación (ms)	0.5	0.25	0.125	0.0625	0.03125
Retardo de transmisión de datos (ms)	1	0.5	0.25	0.125	0.0625
Retardo de procesamiento en el gNB (ms)	2	1	0.5	0.25	0.125
Latencia Total (ms)	10.5	5.25	2.625	1.3125	0.65625

Tabla 3.11. Resumen del impacto del TTI en el UL.

La reducción del TTI presenta diversas ventajas para mejorar la latencia de la RAN, sin embargo, es necesario considerar diferentes factores para escoger correctamente este parámetro. Es necesario recalcar que existen otras técnicas y tecnologías que se implementan en los diferentes niveles del modelo OSI y a lo largo de la arquitectura de los sistemas 5G para la reducción de la latencia. La variación del TTI constituye una de estas técnicas que prometen mejorar considerablemente la latencia en los sistemas de comunicaciones móviles futuros y habilitar el despliegue de servicios URLLC.

3.5. COMPARACIÓN DE RESULTADOS

A continuación, se presentan los resultados de trabajos relacionados y su respectivo análisis. El trabajo [38] considera la separación de subportadoras para la reducción del TTI; mientras que el trabajo [39] considera mini-TTI, es decir, diferente número de símbolos por TTI para evaluar su impacto en la latencia.

El artículo [38] presenta una implementación real de un sistema URLLC con el objetivo de analizar la latencia para diferentes configuraciones. En dicho trabajo se consideran las numerologías 0, 1 y 2 definidas en la Tabla 1.3 y se evalúa la latencia en la RAN y de extremo a extremo. La Figura 3.20 muestra los resultados de este trabajo.

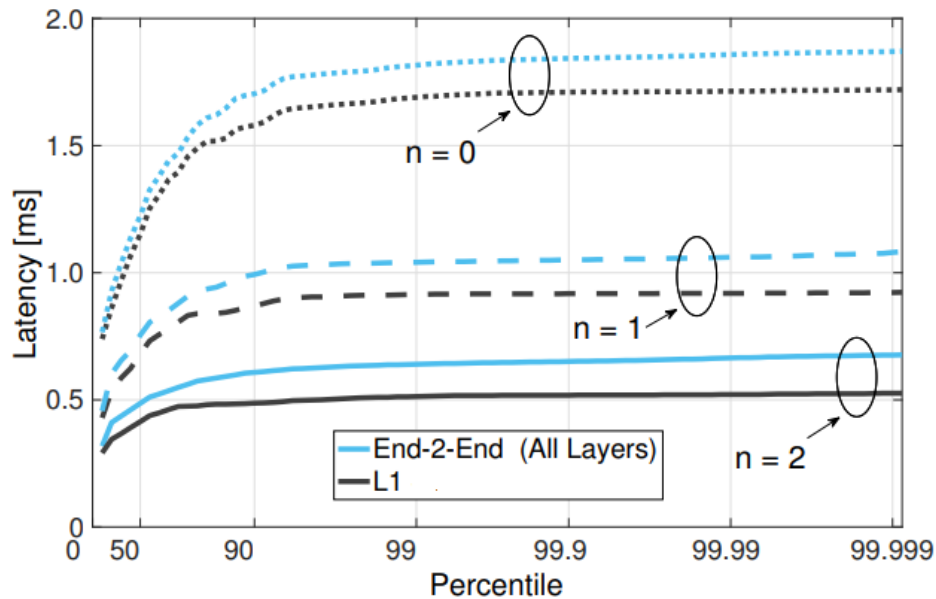


Figura 3.20. Latencia en la RAN y de extremo a extremo para diferentes numerologías. Tomada de [36].

De la Figura 3.20 se observa que la implementación de numerologías superiores disminuye significativamente la latencia tanto en la RAN como de extremo a extremo, lo cual concuerda con los resultados obtenidos en la sección 3.1.1 del presente trabajo de grado. Si se tienen en cuenta los resultados en dicha sección, es posible notar que la latencia obtenida para el caso de carga de sistema baja (ver Figura 3.2) se asemeja a la obtenida en el artículo [38] (ver Figura 3.20). Esto se explica debido a que el artículo [38] solamente considera un receptor para la evaluación de la latencia, o sea carga de sistema baja, caso en el cual todos los recursos radio se asignarán a dicho receptor. En estas condiciones el retardo de cola no influye de forma dramática en la latencia como en los casos de carga de

sistema media y alta analizados en la sección 3.1.1 del presente trabajo de grado, resultando en una latencia relativamente baja.

El artículo [38] también considera diferentes tamaños de segmento y su impacto en la latencia en la RAN, estos resultados se muestran en la Figura 3.21.

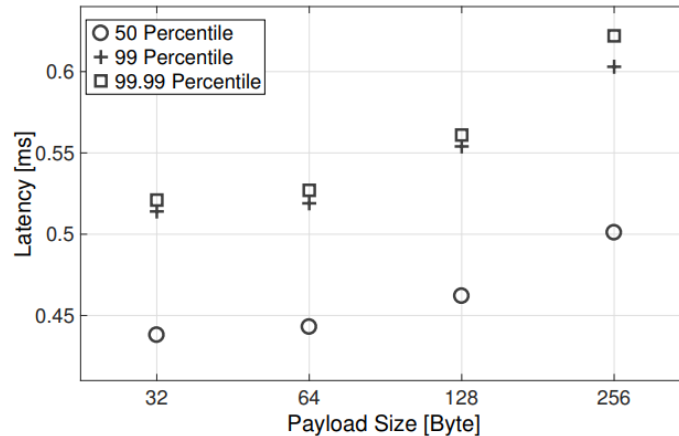


Figura 3.21. Latencia en la RAN para diferentes tamaños de segmentos. Tomada de [36].

De la Figura 3.21 se observa que el tamaño de segmentos influye negativamente en la latencia, lo cual concuerda con los resultados obtenidos en la sección 3.3 del presente trabajo de grado. El tamaño de segmentos impacta en menor medida la latencia de la RAN que la carga del sistema. Esta tendencia fue encontrada en los resultados de la sección 3.3, que muestran un cambio relativamente pequeño en la latencia al incrementar el tamaño del segmento, sin embargo, es necesario considerar el efecto introducido por dicho parámetro en escenarios críticos que requieren niveles mínimos de latencia.

El artículo [39] implementa simulaciones a nivel de sistema para evaluar el impacto del TTI en la latencia. El artículo considera TTI de 14, 7, 4 y 2 símbolos y cargas de sistema alta, media y baja para evaluar su impacto en la latencia. Los resultados de este artículo se muestran en la Figura 3.22.

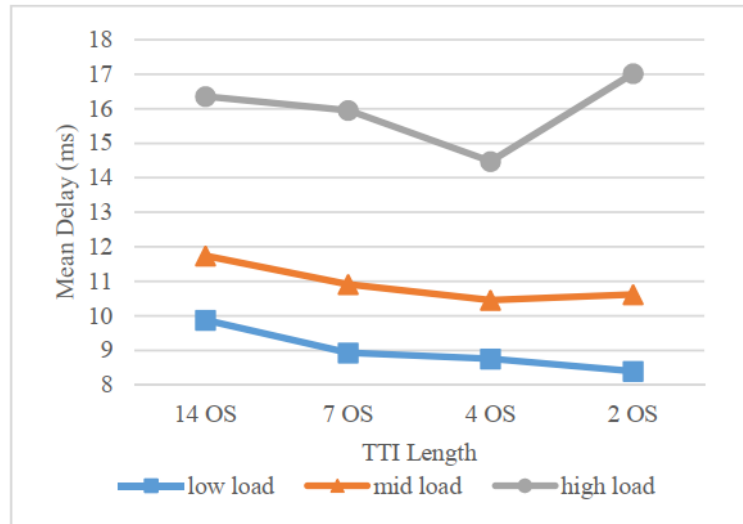


Figura 3.22. Latencia en la RAN para diferentes valores de símbolos por TTI. Tomada de [37].

De la Figura 3.22 se observa que la carga de sistema deteriora significativamente la latencia, lo cual concuerda con los resultados obtenidos en las secciones 3.1.1 y 3.1.2 del presente trabajo de grado. Así mismo, la reducción del TTI mejora la latencia en la mayoría de casos. Sin embargo, al considerar el valor de 2 símbolos por TTI en los niveles de carga medio y alto, la latencia se deteriora, tendencia que también fue encontrada en los resultados de la sección 3.1.2. Como se analizó en dicha sección, este efecto se puede explicar debido al *overhead* introducido por el valor de 2 símbolos por TTI, resultando en una latencia mayor.

CAPÍTULO 4: CONCLUSIONES Y TRABAJOS FUTUROS

4.1. INTRODUCCIÓN

El presente trabajo de grado analizó el impacto del intervalo de tiempo de transmisión (TTI) sobre la latencia en la red de acceso radio (RAN) de sistemas 5G. Para ello, se estudió la estructura de trama de nivel físico de dichos sistemas y se escogieron parámetros apropiados para el diseño de casos de estudio. Dichos casos de estudio fueron implementados en la herramienta software ns3, utilizando el módulo *mmWave* para sistemas 5G. Los resultados de dichas simulaciones se analizaron y compararon con trabajos relacionados. Las conclusiones del presente trabajo de grado y trabajos futuros se presentan en las secciones 4.2 y 4.3, respectivamente.

4.2. CONCLUSIONES

- El TTI impacta significativamente la latencia en la RAN, sin embargo, es necesario considerar los parámetros de carga de sistema, tasa y tamaño de segmentos y capacidad de procesamiento de equipos radio para escoger adecuadamente su valor y aprovechar todas las ventajas que éste ofrece en la reducción de la latencia en la RAN.
- La reducción del TTI debida a la implementación de las numerologías definidas en la R15 posee un impacto positivo sobre la latencia en la RAN al incrementar la carga del sistema. La reducción del TTI disminuye el tiempo de espera de asignación de TTI, el retardo de cola, el retardo de alineación y el retardo de transmisión, las cuales componen la latencia de nivel físico. El porcentaje de reducción depende del nivel de carga del sistema.
- La reducción de la latencia en la RAN, debida a la implementación de numerologías superiores, es más evidente cuando se considera una carga de sistema baja. Al aumentar la carga, el retardo de cola se incrementa considerablemente, deteriorando la latencia en la RAN. En el caso de carga alta o media, el retardo de cola es la componente predominante en la latencia

de nivel físico, por lo que las ventajas de la reducción del TTI sobre la latencia desaparecen.

- La reducción del TTI debido a la implementación de mini-TTI de 7 y 4 símbolos posee un impacto positivo sobre la latencia en la RAN al incrementar la carga del sistema. Sin embargo, en el caso de mini-TTI de 2 símbolos, el *overhead* introducido por las señales de referencia y control se incrementa, ocupando una cantidad considerable de TTI. En esta configuración, los TTI disponibles se reducen, aumentando el retardo de cola y deteriorando la latencia en la RAN.
- La tasa de segmentos de datos influye de forma diferente sobre la latencia de la RAN en el UL y DL. Es necesario considerar el proceso de asignación de recursos radio en el UL, en el cual se hace uso de mensajes SR para la transmisión de datos. Dicho proceso deteriora la latencia en la RAN si se repite frecuentemente. Tasas bajas de segmentos incrementan la probabilidad de que un segmento encuentre el *buffer* de transmisión vacío, obligando al UE a enviar nuevos mensajes SR. Al disminuir el TTI, dicha probabilidad también aumenta, dado a que el retardo de transmisión es disminuido y los segmentos de datos son transmitidos más rápidamente. De esta forma, las numerologías superiores proporcionan menores niveles de latencia en la RAN, si los flujos de datos son transmitidos a una tasa de segmentos alta.
- El tamaño de segmentos influye negativamente en la latencia debido a que se asignan un mayor número de TTI para su transmisión. El impacto de este parámetro no es considerable, sin embargo, debe ser tenido en cuenta para escenarios con requerimientos críticos de latencia.
- El retardo de procesamiento posee un impacto significativo sobre la latencia en la RAN. Este parámetro depende de la capacidad de procesamiento de los equipos radio y el valor del TTI. Al disminuir el TTI, el retardo de procesamiento se reduce considerablemente, lo que constituye una gran ventaja considerando de manera puntual el proceso de asignación de

recursos en el UL, el cual comprende la transmisión de varios mensajes en el UL y DL.

4.3. TRABAJOS FUTUROS

- Analizar el impacto de las demás técnicas y tecnologías definidas en la especificación R15 sobre la latencia en la RAN, tanto independientemente como de forma conjunta.
- Analizar las técnicas y tecnologías definidas en la especificación R15 que buscan mejorar la latencia en el plano de control de sistemas 5G.
- Analizar el impacto de las técnicas y tecnologías definidas en la especificación R15 sobre la confiabilidad en la RAN, tanto independientemente como de forma conjunta.
- Analizar el impacto del TTI y demás técnicas y tecnologías definidas en la especificación R15 sobre el *throughput* en la RAN, tanto independientemente como de forma conjunta.
- Evaluar la viabilidad del despliegue de servicios URLLC en sistemas 5G.
- Analizar el impacto de las técnicas y tecnologías diseñadas para la red de transporte y núcleo de red que buscan mejorar la latencia, la confiabilidad y la capacidad en los sistemas 5G.
- Evaluar el desempeño de sistemas 5G al conectarse con una densidad alta de dispositivos.

REFERENCIAS BIBLIOGRÁFICAS

- [1] N. Isachenko, "El papel de la información y las tecnologías de la información y la comunicación en la sociedad moderna", *Utopía y Praxis Latinoamericana*, vol. 23, pp. 361-367, 2018. Available: 10.5281/zenodo.1512122 [Accessed 30 March 2019].
- [2] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Commun. Surv. Tutor.*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.
- [3] S. Zhang et al., "5G: Towards energy-efficient, low- latency and high-reliable communications networks," in *Proc. IEEE Int. Conf. on Commun. Syst. (ICCS)*, Nov 2014, pp. 197–201.
- [4] M. Shafi *et al.*, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice", *IEEE Journal on Selected Areas in Communications*, vol. 35, n.º 6, pp. 1201-1221, jun. 2017.
- [5] Itu.int. (2018). *M.2083: IMT Vision - "Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond"*. [Online] Available at: <https://www.itu.int/rec/R-REC-M.2083-0-201509-I/es> [Accessed 24 Dec. 2018].
- [6] J. Butler, "5G Spectrum Challenges", *5G Radio Technology Seminar. Exploring Technical Challenges in the Emerging 5G Ecosystem*, 2015. Available: 10.1049/ic.2015.0028 [Accessed 30 March 2019].
- [7] P. Marsch et al., *5G System Design*. Chichester: Wiley, 2018.
- [8] "Release 15 - 3GPP", 2019. [Online]. Available: <https://www.3gpp.org/release-15>. [Accessed: 30- Mar- 2019].

- [9] Y. Niu et al., "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges", *Wireless Networks*, vol. 21, no. 8, pp. 2657-2676, 2015. Available: 10.1007/s11276-015-0942-z.
- [10] B. Rong et al., "Emerging Technologies for 5G Radio Access Network: Architecture, Physical Layer Technologies, and MAC Layer Protocols", *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1-2, 2018. Available: 10.1155/2018/6082161 [Accessed 30 March 2019].
- [11] W. Xiang, *5G Mobile Communications*. Washington: Springer, 2018.
- [12] G. Wunder et al., "New Physical-layer Waveforms for 5G", *Towards 5G*, pp. 303-341, 2016. Available: 10.1002/9781118979846.ch14 [Accessed 30 March 2019].
- [13] J. Du and S. Signell, "Comparison of CP-OFDM and OFDM/OQAM in Doubly Dispersive Channels", *Future Generation Communication and Networking (FGCN 2007)*, 2007. Available: 10.1109/fgcn.2007.89 [Accessed 30 March 2019].
- [14] F. Luo and C. Zhang, *Signal processing for 5G*. Springer, 2018.
- [15] B. Jia et al., "Flexible uplink MU-MIMO scheduling in unlicensed spectrum", *China Communications*, vol. 14, no. 12, pp. 59-71, 2017. Available: 10.1109/cc.2017.8246337 [Accessed 30 March 2019].
- [16] S. Lien et al., "5G New Radio: Waveform, Frame Structure, Multiple Access, and Initial Access", *IEEE Communications Magazine*, vol. 55, no. 6, pp. 64-71, 2017. Available: 10.1109/mcom.2017.1601107 [Accessed 30 March 2019].
- [17] "What is round-trip time? RTT definition | Cloudflare", *Cloudflare*, 2019. [Online]. Available: <https://www.cloudflare.com/learning/cdn/glossary/round-trip-time-rtt/>. [Accessed: 20- Jun- 2019].

- [18] *Techmahindra.com*, 2019. [Online]. Available: <http://www.techmahindra.com/documents/WhitePaper/WhitePaperLatencyAnalysis.pdf>. [Accessed: 24- Dec- 2018].
- [19] N. Maskey et al., "Analysis of latency for cellular networks for smart grid in suburban area", *IEEE PES Innovative Smart Grid Technologies, Europe*, 2014. Available: 10.1109/isgteurope.2014.7028750 [Accessed 30 March 2019].
- [20] A. Prasad et al., "Energy Efficient Coordinated Self-Backhauling for Ultra-Dense 5G Networks", *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017. Available: 10.1109/vtcspring.2017.8108609 [Accessed 4 April 2019].
- [21] I. Parvez et al., "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions", *IEEE Communications Surveys & Tutorials*, 20(4), pp.3098-3130.
- [22] J. Sachs et al., "Adaptive 5G Low-Latency Communication for Tactile Internet Services", *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325-349, 2019. Available: 10.1109/jproc.2018.2864587 [Accessed 30 March 2019].
- [23] L. Vijayasathy and C. Butler, "Choice of Software Development Methodologies: Do Organizational, Project, and Team Characteristics Matter?", *IEEE Software*, vol. 33, no. 5, pp. 86-94, 2016. Available: 10.1109/ms.2015.26 [Accessed 10 May 2019].
- [24] M. Meneses Pino and J. Sanchez Pino, "Análisis de la Capacidad de Transmisión de Datos en una Red Celular con MMWAVE ante Diferentes Estrategias de Compartición de Espectro", *Ingeniería Electrónica y de Telecomunicaciones*, Universidad del Cauca, 2018.
- [25] "OMNeT++ Discrete Event Simulator", *Omnetpp.org*, 2019. [Online]. Available: <https://omnetpp.org/>. [Accessed: 10- May- 2019].

- [26] "OPNET IS NOW PART OF RIVERBED STEELCENTRAL™", *Riverbed*, 2019. [Online]. Available: <https://www.riverbed.com/id/products/steelcentral/opnet.html>. [Accessed: 10- May- 2019].
- [27] "MATLAB - MathWorks", *Mathworks.com*, 2019. [Online]. Available: <https://www.mathworks.com/products/matlab.html>. [Accessed: 10- May- 2019].
- [28] ns-3. (2018). *Documentation*. [online] Available at: <https://www.nsnam.org/documentation/> [Accessed 24 Dec. 2018].
- [29] GitHub. (2018). *nyuwireless-unipd/ns3-mmwave*. [online] Available at: <https://github.com/nyuwireless-unipd/ns3-mmwave> [Accessed 24 Dec. 2018].
- [30] M. Mezzavilla et al., "End-to-End Simulation of 5G mmWave Networks", *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2237-2263, 2018. Available: 10.1109/comst.2018.2828880 [Accessed 30 March 2019].
- [31] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz", Reporte Técnico 38.901, 2018.
- [32] S. Ahmadi, *LTE-Advanced*. "A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies". Elsevier Science, 2013.
- [33] J. Solano Arenas et al., "Ultra-low Latency in Next Generation LTE Radio Access", *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, 2017. [Accessed 23 October 2019].
- [34] D. Szczesny et al., "Performance analysis of LTE protocol processing on an ARM based mobile platform", *2009 International Symposium on System-on-Chip*, 2009. Available: 10.1109/socc.2009.5335678 [Accessed 23 October 2019].
- [35] "ns-3 Tutorial — Tutorial", *Nsnam.org*, 2019. [Online]. Available: <https://www.nsnam.org/docs/tutorial/html/>. [Accessed: 21- Jun- 2019].

[36] "TDoc R1-160934", *Portal.3gpp.org*, 2019. [Online]. Available: <https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=685917>. [Accessed: 02- Jul- 2019].

[37] "5G NR Scheduling Request Procedure | Uplink Grant", *Rfwireless-world.com*, 2019. [Online]. Available: <https://www.rfwireless-world.com/5G/5G-NR-Scheduling-Request-Procedure.html>. [Accessed: 01- Jul- 2019].

[38] T. Wirth et al., "5G New Radio and Ultra Low Latency Applications: A PHY Implementation Perspective", *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016. Available: 10.1109/acssc.2016.7869608 [Accessed 01 July 2019].

[39] Z. Zhang et al., "Performance evaluation of shortened transmission time interval in LTE networks", *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018. Available: 10.1109/wcnc.2018.8377098 [Accessed 01 July 2019].