



Programa de Doctorado en Tecnologías Industriales y de
Telecomunicación

**Mecanismos de *scheduling* para la
integración de redes 5G y TSN en entornos
de la Industria 4.0**

Autora: Ana Larrañaga Zumeta

Directora: Dra. María del Carmen Lucas-Estañ

Codirector: Dr. Javier Manuel Gozávez Sempere

Codirector: Dr. Imanol Martínez Coca

Universidad Miguel Hernández de Elche



La presente Tesis Doctoral, titulada “**Mecanismos de *scheduling* para la integración de redes 5G y TSN en entornos de la Industria 4.0**”, se presenta bajo la modalidad de tesis por compendio de las siguientes publicaciones de acuerdo con la Normativa de Estudios de Doctorado de la Universidad Miguel Hernández de Elche:

Artículo publicado en revista de impacto, indexada según JCR Science Edition:

- *An open-source implementation and validation of 5G NR configured grant for URLLC in ns-3 5G LENA: A scheduling case study in industry 4.0 scenarios*
Ana Larrañaga, M. Carmen Lucas-Estañ, Sandra Lagén, Zoraze Ali, Imanol Martínez, Javier Gozávez

Journal of Network and Computer Applications

Volume 215, 2023, article 103638

ISSN: 1084-8045, Editorial: Elsevier

<https://doi.org/10.1016/j.jnca.2023.103638>

Factor de impacto JCR-SCI: 8.7, Cuartil Q1

Esta tesis incluye artículos en congresos internacionales con revisión por pares sobre artículo completo y que cumple los criterios de calidad establecidos en la normativa relativa a los criterios específicos de evaluación para la obtención de SEXENIOS de investigación para el subcampo “6.2 Ingenierías de la Comunicación, Computación y Electrónica” publicados en la Resolución de 21 de noviembre de 2022, de la Comisión Nacional Evaluadora de la Actividad Investigadora (Disposición 21729 del BOE núm. 305 de 2022). Esta normativa establece como aportaciones valorables las publicaciones en congresos que aparecen en el ranking GII-GRIN-SCIE (GGs) o similares. Junto a la referencia de cada publicación, se presenta la posición de la conferencia en el ranking GGS y en otros rankings reconocidos en el área de conocimiento como CORE o Qualis:

- *Configured Grant Scheduling for the Support of TSN Traffic in 5G and Beyond Industrial Networks*

M.C. Lucas-Estañ, A. Larrañaga, J. Gozávez, I. Martínez

2023 98th IEEE Vehicular Technology Conference (VTC2023-Fall)

Hong Kong, 2023

Editorial: IEEE (**Aceptado para publicar**)

GGs rating: Clase A (Very good event)

Categoría en otros rankings: A1 según *Qualis* y B según *CORE*



- *Analysis of 5G-TSN Integration to Support Industry 4.0*
A. Larrañaga, M. C. Lucas-Estañ, I. Martínez, I. Val and J. Gozávez
2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)
Vienna, Austria, 2020
pp. 1111-1114
ISBN: 978-1-7281-8957-4. Editorial: IEEE
doi: 10.1109/ETFA46521.2020.9212141
GGG rating: Clase B (Good quality events)
Categoría en otros rankings: B1 según *Qualis*
Ha sido citado 53 veces (Google Scholar)

- *5G Configured Grant Scheduling for 5G-TSN Integration for the Support of Industry 4.0*
A. Larrañaga, M. C. Lucas-Estañ, I. Martínez and J. Gozávez
2023 18th IEEE Wireless On-Demand Network Systems and Services Conference (WONS)
Madonna di Campiglio, Italy, 2023
pp. 72-79
ISBN: 978-3-903176-56-0, Editorial: IEEE
doi: 10.23919/WONS57325.2023.10062219
GGG rating: Clase C (Work in Progress)
Categoría en otros rankings: B4 según *Qualis*



Además de las publicaciones aportadas como indicio de calidad de la tesis, parte del trabajo realizado en esta tesis también fue presentado en la siguiente publicación de conferencia:

- *5G NR Configured Grant in ns-3 Network Simulator for Ultra-Reliable Low Latency Communications*
Ana Larrañaga, M. Carmen Lucas-Estañ, Imanol Martínez, Javier Gozávez
Procedia Computer Science, Special issue: The 5th International Conference on Emerging Data and Industry 4.0 (EDI40)
Volume 201, 2022
Pages 495-502
ISSN: 1877-0509, Editorial: Elsevier
<https://doi.org/10.1016/j.procs.2022.03.064>



La Dra. M^a del Carmen Lucas-Estañ, directora, el Dr. Javier Manuel Gozávez Sempere, codirector, y el Dr. Imanol Martínez Coca, codirector industrial, de la tesis doctoral titulada “Mecanismos de *scheduling* para la integración de redes 5G y TSN en entornos de la Industria 4.0”

INFORMAN:

Que Ana Larrañaga Zumeta ha realizado bajo nuestra supervisión el trabajo titulado “Mecanismos de *scheduling* para la integración de redes 5G y TSN en entornos de la Industria 4.0.” conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmamos para los efectos oportunos, el 26 de septiembre de 2023.

Directora de la tesis

Dra. M^a del Carmen Lucas-Estañ

Codirector de la tesis

Dr. Javier Manuel Gozávez Sempere

Codirector de la tesis

Dr. Imanol Martínez Coca



El Dr. Óscar Reinoso García, Coordinador del Programa de Doctorado en Tecnologías Industriales y de Telecomunicación,

INFORMA:

Que Ana Larrañaga Zumeta ha realizado bajo la supervisión de nuestro Programa de Doctorado el trabajo titulado “Mecanismos de *scheduling* para la integración de redes 5G y TSN en entornos de la Industria 4.0” conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, el 26 de septiembre de 2023.

Dr. Óscar Reinoso García,
Coordinador del Programa de Doctorado en Tecnologías Industriales y de
Telecomunicación

Financiación

Esta tesis ha sido financiada por el proyecto Basque Industry5G (B-INDUSTRY5G) y el proyecto 5G for Basque RIS3 (5G4BRIS3) del programa ELKARTEK del Gobierno Vasco, y parcialmente por MCIN/AEI/10.13039/501100011033 a través del proyecto PID2020-115576RB-100, y por el proyecto Zero-SWARM del programa Horizon Europe de la Unión Europea (Agreement 101057083).

Agradecimientos

Hoy, un día de verano de 2023, mientras doy por concluida mi enésima revisión de la tesis, me he detenido para dar las gracias a todas las personas que han hecho posible que yo este llegando a la meta final de esta tesis.

En primer lugar, me gustaría agradecer a mi directora M^a del Carmen Lucas-Estañ. Estos cuatro años me has guiado y me has enseñado a buscar mi propio camino. Gracias por todo el esfuerzo que has hecho para sacar esto adelante, por todas esas horas, reuniones, emails, ideas, correcciones, consejos y enseñanzas. Todas tus propuestas y correcciones han contribuido positivamente en la calidad de esta tesis. También me gustaría agradecer a mis codirectores, Javier e Imanol, aun teniendo una presencia más discreta durante la tesis, su apoyo en momentos clave ha sido fundamental. También me gustaría hacer una pequeña mención a Aitor, mi tutor en Ikerlan estos últimos meses. Su ayuda ha sido esencial en esta última etapa final de la tesis, gracias por leer todo el documento y hacer aportaciones valiosas. Querría extender estos agradecimientos a Ikerlan, por darme la oportunidad de poder realizar esta tesis, hacer una estancia y asistir a conferencias.

Quiero dar las gracias al equipo de CTTC, especialmente a Sandra y a Zoraze. A esta estancia de 3 meses en Castelldefels (Catalunya) llegué en el momento más bajo de la tesis, nervios, prisa, un poco perdida y me impulsasteis hacia la meta, un chute de motivación. Moltes gràcies. Special thanks to Zoraze, no one has given me as many GitLab tips as you did.

Ikerlan además de darme la oportunidad de realizar esta tesis, me ha traído a mi vida personitas muy especiales. Alaine y Fátima. Mis excompañeras del pisito, amigas del trabajo, ahora mi familia, qué hubiese sido yo y esta tesis sin vosotras. Hemos celebrado los logros y sufrido esos días amargos donde la tesis parecía un laberinto sin salida, juntas. Me habéis escuchado y corregido múltiples presentaciones que he tenido que hacer a lo largo de estos cuatro años. Después de un día agotador, os habéis sentado conmigo en la cocina de nuestra casa para analizar el código que tantos quebraderos de cabeza me daba. Hemos compartido momentos de paseos, escalada, natación, cocina, desayunos con *Egun on*-es adorables, conversaciones profundas y conversaciones chorras, muchas risas y también algún que otro lloro. Sin duda estos cuatro años habéis sido alegría, apoyo, y todo lo que se puede buscar en unas buenas amigas. Muchas gracias, amigas, no podíais no aparecer en este apartado, porque sí, un cachito de esta tesis también es vuestro.

En el top 1 de la lista de personas que me han acompañado y ayudado durante esta tesis, no hay duda de que te encuentras tú, Charly. Aunque muchos no lo sepan, has contribuido en esta tesis de una manera activa. Me has ayudado a buscar aplicaciones o incluso poner en marcha servidores para agilizar todo el proceso de la obtención de resultados, has estado

a las 23:59 de la noche conmigo subiendo un artículo que daba error de formato mientras yo me tiraba de los pelos, has leído los diferentes artículos que había escrito, y escuchado también las diferentes ponencias que iba a hacer. Pero, sobre todo, a lo largo de esta tesis has sido mi refugio, el que ha tenido que escuchar todas mis quejas y buenas noticias contadas de mil maneras diferentes diciendo lo mismo mientras ponías cara de que no te estabas aburriendo con el monotema, me has animado, siempre has confiado en que podría sacar este trabajo adelante, has entendido lo que esto suponía para mí, y te has puesto en mi lugar. Por todo esto, y mucho más, Eskerrik Asko maitia.

En estos agradecimientos no puede faltar mi familia, Ama, Aita, Maria eta Inaxio. Eskerrik asko ama eta aita, por escucharme todas esas chapadas que os he metido a la hora de comer o mientras tomábamos un vinito por las calles de Zarautz un viernes, por mimarme cuando he estado estresada, y por tener paciencia en los días donde no me soportaba ni a mí misma. Eskerrik asko Maria eta Inaxio por enseñarme muchas cosas de la vida que de una manera indirecta he terminado aplicando en la tesis y en mi vida laboral, como ser valiente y saber sacar las castañas del fuego. Mila esker familia telerín, porque hasta el logro más pequeño lo hemos celebrado juntos.

Quiero terminar dándole las gracias a mi Abuelo, *Aitona Benantzio*, que ya no estás, pero sé que estarías más feliz y más orgulloso que nadie. Eskerrik asko aitona, por confiar en mí, por motivarme a estudiar una ingeniería, porque la cara de orgullo y felicidad que tenías el día de mi graduación aún la recuerdo. Aitona, tesi honen lorpena ere zuregatik da, mila esker.

Índice de Contenido

Índice de Contenido	i
Listado de abreviaturas.....	iii
Listado de figuras.....	vii
Listado de tablas.....	ix
Resumen.....	xi
Abstract	xiii
1. Introducción	1
1.1. Objetivos	4
1.2. Estructura y resumen de la tesis	5
2. Estado del arte de la integración de 5G en redes TSN	11
2.1. Comunicaciones Industriales: Casos de uso 5G-TSN y requisitos.....	11
2.2. Red TSN.....	16
2.3. Redes móviles 5G.....	20
2.4. Integración de redes 5G y TSN	29
2.5. Estado de la literatura sobre esquemas de <i>scheduling</i> en redes 5G-TSN integradas	30
3. Análisis de la integración de 5G- TSN según el modelo del <i>bridge</i> virtual.....	37
3.1. Red 5G como <i>bridge</i> TSN virtual	38
3.2. Latencia en el <i>bridge</i> 5G virtual.....	40
3.3. Conclusiones	44
4. Diseño de esquemas de <i>scheduling Configured Grant</i> para redes 5G integradas con redes TSN	47
4.1. <i>Scheduling</i> en redes 5G integradas con redes TSN	48
4.2. <i>Optimum Flexible configured grAnt Scheduling for TSN traffic (O-FAST)</i>	51
4.3. <i>Heuristic Flexible configured grAnt Scheduling for TSN traffic (H-FAST)</i>	58
4.4. Escenario de simulación	62
4.5. Evaluación del rendimiento	64
4.6. Conclusiones	71
5. <i>Scheduling</i> para redes 5G integradas en redes TSN compatibles con el estándar 3GPP.....	75
5.1. Vector de activación de las configuraciones CG.....	77
5.2. <i>Heuristic 3GPP-Compliant Flexible configured grAnt Scheduling for TSN traffic (C-FAST)</i>	81
5.3. Evaluación del rendimiento.....	86
5.4. Conclusiones	90
6. Implementación de <i>Configured Grant</i> en 5G-LENA.....	93
6.1. Simulador de red 5G-LENA.....	95

6.2.	Implementación de <i>Configured Grant</i>	98
6.3.	Escenarios evaluados	105
6.4.	Validación.....	106
6.5.	Evaluación	109
6.6.	Conclusiones y resultados adicionales.....	114
7.	Conclusiones y líneas futuras de investigación	117
7.1.	Contribuciones científicas y conclusiones	118
7.2.	Líneas futuras de investigación	120
	Bibliografía.....	123
	Anexo A. Publicaciones	123
	Anexo B. Cálculo de la latencia máxima y el tiempo de llegada al <i>bridge</i> 5G virtual	173

Listado de abreviaturas

3GPP	Third Generation Partnership Project
5G	Fifth Generation
5G-ACIA	5G Alliance for Connected Industries and Automation
5QI	5G QoS Identifier
AF	Application Function
AMF	Access and Mobility Function
AUSF	Authentication Server Function
BD	Bridge Delay
BE	Best Effort
BLER	Block Error Rate
BW	Bandwidth
BWP	Bandwidth Part
C2C	Controller-to-controller
C2D	Controller-to-device
C-FAST	Heuristic 3GPP-Compliant Flexible configured grAnt Scheduling for TSN traffic
CG	Configured Grant
CN	Core Network
CNC	Centralized Network Configuration
CP	Control Plane
CQI	Channel Quality Indicator
CRC	Cyclic Redundancy Check
CUC	Centralized User Configuration
CG	Configured Grant
D2Cmp	Device-to-compute
DCI	Downlink Control Information
DL	Downlink
dLCP	Dynamic Logical Channel Prioritization
DS-TT	Device-Side TSN translator
ED	End Device
eMBB	Enhanced Mobile Broadband

EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplexing
gNB	Next Generation Node B
GCL	Gate Control List
GBR	Guaranteed Bit Rate
GFBR	Guaranteed Flow Bit Rate
gPTP	Generic Precision Time Protocol
H-FAST	Heuristic Flexible configured grAnt Scheduling for TSN traffic
HP	Hyper Period
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IIC	Industrial Internet Consortium
IIoT	Industrial Internet of Things
ISA	International Society of Automation
L2C	Line controller-to-controller
LOS	Line of Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MCG	Multiple Configured Grant
MCS	Modulation and Coding Scheme
MDBV	Maximum Data Burst Volume
MFBR	Maximum Flow Bit Rate
mMTC	Massive Machine Type Communications
NF	Network Function
NFV	Network Function Virtualization
NIST	National Institute of Standards and Technology
NOMA	Non-Orthogonal Multiple Access
NPN	Non-Public Network
NR	New Radio
NRF	Network Repository Function
NSA	Non-Standalone
NW-TT	Network-Side TSN Translator
OAI	Open Air Interface

O-FAST	Optimum Flexible configured grAnt Scheduling for TSN traffic
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PCC	Policy and Charging Control
PCF	Policy Control Function
PDB	Packet Delay Budget
PER	Packet Error Rate
PLC	Programmable Logic Controller
PDU	Packet Data Unit
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RB	Resource Block
RRC	Radio Resource Control
SA	Standalone
SCS	Sub Carrier Spacing
SDN	Software Defined Network
SDR	Software Defined Radio
SFN	System Frame Number
SMF	Session Management Function
SNR	Signal to Noise Ratio
SPS	Semi Persistent Scheduling
SINR	Signal to Interference and Noise Ratio
SR	Scheduling Request
TAS	Time Aware Scheduler
TB	Transport Block
TDD	Time Division Duplex
TDMA	Time Division Multiplexing Access
TSC	Time Sensitive Communications
TSCAC	Time Sensitive Communications Assistance Container
TSCAI	Time Sensitive Communications Assistance Information
TSN	Time Sensitive Networking
TTI	Transmission Time Interval

UDM	Unified Data Management
UE	User Equipment
UL	Uplink
UP	User Plane
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communications

Listado de figuras

Figura 1. Integración de las redes 5G y TSN para aplicaciones de la Industria 4.0.....	3
Figura 2. Casos de uso de la Industria 4.0.	12
Figura 3. Modelo de configuración de red TSN centralizada.	18
Figura 4. El puerto de salida 1 del bridge TSN que admite transmisiones deterministas a través del scheduler TAS o TSN.....	19
Figura 5. Scheduling de TSN para el puerto de salida 1.....	19
Figura 6. Resource grid para numerología 0 y numerología 1.....	22
Figura 7. Configured Grant del tipo 1.....	24
Figura 8. Ejemplo de preasignación periódica a través de la configuración CG.	26
Figura 9. Arquitectura de 5G SA.	27
Figura 10. Arquitectura y modelo de integración de las redes 5G y TSN.	30
Figura 11. Superposición de recursos radio utilizando diferentes configuraciones CG. ..	35
Figura 12. Los pasos que se realizan para configurar una red 5G-TSN integrada.....	38
Figura 13. Ejemplo de una red 5G-TSN integrada.	40
Figura 14. El camino extremo a extremo de los flujos TSN a) F1 y b) F2.	41
Figura 15. Ejemplo del resource grid y los parámetros de los schedulers.	50
Figura 16. Cálculo del HP para un ejemplo donde 3 flujos TSN transmiten paquetes con diferentes periodicidades.	52
Figura 17. Proceso de asignación de recursos radio utilizando el esquema de scheduling O-FAST.....	53
Figura 18. Bloques creados dentro del HP para un ejemplo donde 3 flujos TSN transmiten paquetes con diferentes periodicidades.....	54
Figura 19. Asignación de recursos radio para los paquetes pkt1,1, pkt2,1y pkt3,1.	60
Figura 20. Plano de usuario del escenario evaluado.	62
Figura 21. Porcentaje de esquemas de scheduling satisfactorias en los que se garantizan los requisitos de latencia a todos los paquetes TSN transmitidos (P_{satisf}) para H-FAST, O-FAST y el esquema de referencia en función del número de flujos TSN evaluados.	67
Figura 22. Porcentaje de colisiones ($P_{pktConflict}$) con el esquema de referencia.	68
Figura 23. Latencia media experimentada con H-FAST, O-FAST y el esquema de referencia en función del número de flujos TSN evaluados.	69
Figura 24. Desviación típica media de la latencia cuando se utilizan los schedulers propuestos en escenarios en los que H-FAST encuentra solución.....	70
Figura 25. Tiempo computacional promedio en segundos para asignar recursos en función de los tres <i>schedulers</i>	71

Figura 26. Ejemplo explicativo del uso del vector de activación de las configuraciones CG.	78
Figura 27. Relación del paquete N con el indicador j	80
Figura 28. Ejemplo de solución de <i>scheduling</i> al aplicar H-FAST y C-FAST.	82
Figura 29. Latencia media experimentada con C-FAST, H-FAST y el esquema de referencia en función del número de flujos TSN evaluados.....	88
Figura 30. Desviación típica media de la latencia cuando se utilizan los schedulers H-FAST y C-FAST.	89
Figura 31. Tiempo computacional promedio en segundos para asignar recursos en función de los tres planificadores.	90
Figura 32. Representación lógica de la clase RAN de 5G-LENA.	95
Figura 33. Ejemplo de asignación de recursos radio de los distintos modos de acceso múltiple.	97
Figura 34. Máquina de estados de CG implementado en el UE MAC.....	99
Figura 35. Proceso de asignación recursos radio para transmisiones en UL (función DoScheduleUIData) en 5G-LENA.....	100
Figura 36. Asignación de recursos de radio con las políticas diseñadas.	103
Figura 37. Máxima latencia UL en función del ancho de banda (paquetes de 10 bytes).109	
Figura 38. Latencia UL experimentada en función del ancho de banda (tamaño de paquete de 10 bytes).	111
Figura 39. Porcentaje de recursos radio utilizados para la transmisión de paquetes con respecto a 5GL-TDMA (tamaño de paquete de 10 bytes).....	111
Figura 40. La latencia experimentada en UL como función del tamaño del paquete (BW = 20 MHz).	112
Figura 41. Rendimiento en función del MCS (BW = 20 MHz, tamaño de paquete = 10 bytes).	113
Figura 42. Latencia experimentada con los distintos esquemas en función del formato del slot (BW = 20 MHz y tamaño de paquete de 10 bytes).	114

Listado de tablas

Tabla 1. Tipos de tráfico en la automatización industrial y requisitos de servicios [3]. ...	14
Tabla 2. Requisitos de comunicación para una celda de trabajo en una fábrica con comunicaciones inalámbricas definidas por NIST [26].	16
Tabla 3. Artículos de esquemas de scheduling para redes 5G-TSN.	33
Tabla 4. Artículos de esquemas de scheduling semi-estáticos.....	35
Tabla 5. Parámetros de 5G QoS para los flujos TSN F1 y F2.	44
Tabla 6. Parámetros.	50
Tabla 7. Número de CGs y % de UEs para los esquemas de scheduling H-FAST y C-FAST.....	87
Tabla 8. Latencia experimentada con el mecanismo dinámico y con CG	107
Tabla 9. Expresiones finales de la máxima latencia en UL con el scheduler correspondiente.....	108

Resumen

La Industria 4.0 busca la digitalización y la automatización de las fábricas convencionales para poder ofrecer sistemas de fabricación más personalizados, eficientes y sin defectos. Este nuevo paradigma requiere de redes de comunicaciones flexibles y con capacidad de adaptación y reconfiguración, capaces de adaptarse de manera eficiente a las cambiantes demandas de producción. Además, las aplicaciones industriales emergentes requerirán de comunicaciones con baja latencia y muy alta fiabilidad, y con un alto grado de determinismo. *Time Sensitive Networking* (TSN) se está convirtiendo en la tecnología estándar basada en Ethernet para dar soporte a servicios con requisitos deterministas y ultra fiables. A pesar de las ventajas que ofrece, TSN presenta limitaciones en cuanto a su flexibilidad, capacidad de adaptación y compatibilidad con nodos móviles, requisitos indispensables para dar soporte a las industrias del futuro. La red móvil de 5ª generación (5G) ha sido diseñada para proporcionar comunicaciones de muy baja latencia y alta fiabilidad. Sin embargo, 5G no puede garantizar el determinismo que proporciona una red cableada. Es por ello, que la integración de las redes 5G y TSN será clave para lograr el nivel de determinismo y flexibilidad demandado por los entornos de la Industria 4.0.

La investigación sobre las redes 5G-TSN integradas se encuentra en etapas tempranas y aún queda trabajo por hacer para integrar adecuadamente ambas redes de manera que satisfagan plenamente los requisitos de la Industria 4.0. Un aspecto crítico para lograr la integración de ambas redes es la coordinación de sus *schedulers*. En este contexto, en esta tesis se han estudiado y propuesto esquemas de *scheduling* para garantizar la coordinación entre las redes 5G y TSN. El objetivo de la tesis es avanzar hacia la integración eficiente y efectiva de ambas redes 5G y TSN.

Para ello, en esta tesis se ha realizado en primer lugar un análisis sobre la topología de la red integrada 5G-TSN, así como de los requisitos de los distintos flujos de datos a transmitir. Utilizando esta información, la red TSN decide la ruta y el *scheduling* para cada flujo TSN. El esquema de *scheduling* de 5G debe coordinar su decisión con el *scheduling* de TSN para garantizar los requisitos de latencia y determinismo *end-to-end*. Para ello, en esta tesis se han propuesto nuevos esquemas de *scheduling* para garantizar la transmisión de tráfico TSN determinista y periódico en redes 5G. Los esquemas de *scheduling* propuestos están basados en el mecanismo *Configured Grant* (CG) definido en *5G New Radio* (NR) que asigna recursos radio periódicos a los usuarios (UE o *User Equipment*). Los esquemas propuestos consideran la gestión de múltiples flujos TSN con diferentes periodicidades. El objetivo de estos esquemas de *scheduling* es garantizar los requisitos de latencia y determinismo para todos los flujos TSN en la red 5G y minimizar la latencia

experimentada por dichos flujos TSN. En esta tesis se han desarrollado dos esquemas de *scheduling* para la red 5G basados en optimización y en un algoritmo heurístico, respectivamente. Los resultados concluyen que los esquemas propuestos mejoran significativamente la cantidad de flujos TSN que pueden ser atendidos en la red 5G-TSN integrada en comparación con esquemas de *scheduling* comúnmente utilizados. La propuesta basada en un algoritmo heurístico es capaz de alcanzar un rendimiento similar a la propuesta que utiliza técnicas de optimización, al mismo tiempo que disminuye considerablemente el coste computacional.

Los esquemas de *scheduling* propuestos buscan la mejor solución posible, lo que puede dar lugar a que el número de configuraciones CG que pueden ser asignados a un UE supere el máximo permitido por el estándar de *Third Generation Partnership Project* (3GPP). Para solucionar este problema, esta tesis define y propone el uso de un vector de activación que indica la configuración CG que está activa para la transmisión de cada paquete de un flujo TSN. Este vector se introduce en el mecanismo que está definido en los estándares de 3GPP para identificar los recursos radio asignados para la transmisión de cada paquete. Es importante señalar que las configuraciones CG asignadas a un flujo TSN pueden ser activadas y desactivadas por el *Next Generation Node B* (gNB) durante una sesión. Por lo tanto, el uso de este vector evita el intercambio de mensajes entre el UE y el gNB para habilitar y deshabilitar las diferentes configuraciones CG asignadas a un flujo TSN. Este vector se utiliza para coordinar el acceso de los UEs a los recursos radio y evitar los conflictos de *scheduling* (la asignación de los mismos recursos radio para transmitir diferentes paquetes) que pueden producirse con los esquemas de *scheduling* comúnmente utilizados. En esta tesis se ha propuesto un nuevo esquema de *scheduling* que considera el uso de este vector de activación. Al igual que las propuestas anteriores, este nuevo esquema pretende garantizar los requisitos de latencia de extremo a extremo para todos los flujos TSN en la red 5G-TSN integrada. Los resultados muestran que este *scheduler* garantiza la transmisión del mismo número de flujos que el *scheduler* heurístico propuesto anteriormente, pero cumpliendo con la cantidad de configuraciones CG que se pueden crear según los estándares de 3GPP.

Finalmente, como parte de esta tesis, también se ha llevado a cabo la preparación del simulador ns-3 5G-LENA para comunicaciones de baja latencia. En particular, se ha implementado el mecanismo de CG en 5G-LENA. Para modelar con precisión la flexibilidad de 5G NR, también se ha mejorado la implementación del modo de acceso *Orthogonal Frequency Division Multiple Access* (OFDMA) en 5G-LENA, y se han implementado diferentes políticas de *scheduling* para escenarios de la Industria 4.0. Los desarrollos anteriores se han validado mediante la comparación de las latencias obtenidas en el simulador con las latencias obtenidas en distintos estudios analíticos.

Abstract

Industry 4.0 seeks to digitize and automate conventional factories in order to provide more personalized, efficient, and zero-defect manufacturing systems. This new paradigm requires flexible, adaptive, and reconfigurable communications networks, capable to efficiently adapt to the changing production demands. In addition, emerging industrial applications will require communications with low latency and high reliability, and with a high degree of determinism. Time Sensitive Networking (TSN) is becoming the standard Ethernet-based technology to support services with deterministic and ultra-reliable requirements. Despite its advantages, TSN has limitations in terms of flexibility, adaptability, and support for mobile nodes, which are essential for future industries. The 5th generation mobile network (5G) has been designed to provide ultra-reliable and low-latency communications. However, 5G cannot guarantee the determinism that a wired network offers. Hence, the integration of 5G and TSN networks will be key to achieve the determinism and flexibility demanded by Industry 4.0.

Research on integrated 5G-TSN networks is in early stages, and there is still work to be done to achieve the integration of 5G and TSN networks to fully meet the requirements of Industry 4.0. One critical aspect to achieve the integration of both networks is the coordination of their schedulers. In this context, this thesis focuses on the study and proposal of novel scheduling schemes to guarantee the coordination of 5G and TSN networks. The aim of the thesis is to advance towards the efficient and effective integration of 5G and TSN networks.

For this purpose, this thesis has first analyzed the topology of the integrated 5G-TSN network, as well as the requirements of the different data flows to be transmitted. Using this information, the TSN network decides the path and the scheduling for each TSN flow. The 5G scheduling scheme must coordinate its decision with TSN scheduling to ensure end-to-end latency and determinism requirements. To achieve this, this thesis proposes novel scheduling schemes for the support of periodic deterministic TSN traffic in 5G networks. The proposed scheduling schemes are based on Configured Grant (CG) scheduling defined in 5G New Radio (NR) that pre-allocates periodic radio resources to each User Equipment (UE). The proposed schemes consider the management of multiple TSN flows with different periodicities. The objective of these scheduling schemes is to guarantee the latency and determinism requirements for all TSN flows in the 5G network and to minimize the latency experienced by these TSN flows. This thesis proposes two scheduling schemes based on optimization and heuristic approaches, respectively. The results conclude that the proposed schemes outperform the state-of-the-art in the capacity to support multiple TSN flows with different periodicities. The proposal based on a

heuristic algorithm is able to achieve similar performance as the one based on the use of optimization techniques while considerably decreasing the computational cost.

The proposed scheduling schemes aim to find the best possible solution, which can result in the number of configured grants assigned to a TSN flow exceeding the maximum allowed by the Third Generation Partnership Project (3GPP) standard. To address this, this thesis defines and proposes the use of an activation vector that indicates the configured grant that is active for the transmission of each packet of a TSN flow. This vector is introduced in the mechanism defined in the 3GPP standards to identify the radio resources allocated for the transmission of each packet. It is important to note that the configured grants assigned to a TSN flow can be enabled and disabled by the Next Generation Node B (gNB) during a session. The use of this activation vector then avoids the exchange of messages between the UE and the gNB to enable and disable the different configured grants allocated to a TSN flow. This vector is used to coordinate the access of UEs to the radio resources and avoid scheduling conflicts (the allocation of the same radio resources to transmit different packets) that can happen with state-of-the-art scheduling schemes. This thesis then proposes a new scheduling scheme that considers the use of this activation vector. Similar to the previous proposals, this scheme aims to guarantee the end-to-end latency requirements for all the TSN flows in the integrated 5G-TSN network. The results show that this scheduler guarantees the transmission of the same number of flows as the previously proposed heuristic scheduler while complying with the number of configured grants that can be established according to 3GPP standards.

Finally, as part of this thesis, the preparation of the ns-3 5G-LENA simulator for low-latency communications has also been carried out. In particular, the CG mechanism has been implemented in 5G-LENA. To accurately model the flexibility of 5G NR, the implementation of Orthogonal Frequency Division Multiple Access (OFDMA) in 5G-LENA has also been improved, and different scheduling policies for Industry 4.0 scenarios have been implemented. The previous developments have been validated by comparing the latencies obtained in the simulator with the latencies obtained in different analytical studies.

1. Introducción

Actualmente, las industrias están presenciando una transformación digital hacia lo que se conoce como Industria 4.0. El objetivo de la Industria 4.0 es interconectar y digitalizar la industria tradicional para crear fábricas inteligentes y adaptables que optimicen el uso de los recursos y ofrezcan nuevos servicios más eficientes y con cero defectos [1]. Bajo este nuevo paradigma, las fábricas del futuro serán capaces de monitorizar y supervisar toda su línea de producción, tomando decisiones de manera autónoma. Para ello se implementarán tecnologías avanzadas como la realidad aumentada, robots móviles cooperativos o drones [2]. Uno de los pilares fundamentales de esta digitalización industrial es el Internet Industrial de las Cosas (*Industrial Internet of Things* o IIoT) que permitirá tanto la monitorización precisa del entorno industrial como la ejecución de procesos en tiempo real. El IIoT conectará máquinas, robots móviles, sensores y usuarios, facilitando la recopilación, análisis y distribución instantánea de datos. Ciertos casos de uso de la Industria 4.0 requieren la transmisión de comandos de monitorización y actuación con una frecuencia de envío de datos extremadamente alta entre los dispositivos finales [3]. Además, la transmisión de estos datos debe realizarse de manera determinista, es decir, estos paquetes deben ser entregados al dispositivo final dentro de un tiempo especificado y con una alta fiabilidad. En una comunicación determinista, si el paquete no se entrega dentro del tiempo especificado, el sistema lo considera perdido. En estos casos, perder paquetes de información podría afectar los procesos en una cadena de producción, por lo que es esencial contar con comunicaciones de alta fiabilidad y precisión. *Time Sensitive Networking* (TSN) se está convirtiendo en la tecnología cableada estándar para entornos de la Industria 4.0 debido a su alta fiabilidad y determinismo. TSN ha sido diseñada para proporcionar una infraestructura de red común para dar soporte a un amplio abanico de servicios industriales, incluyendo aplicaciones que requieren comunicaciones deterministas, fiables y de bajas latencias.

Uno de los conceptos que introduce el paradigma Industria 4.0 es el diseño de fábricas modulares capaces de adaptar rápidamente sus cadenas de producción a las necesidades cambiantes de la demanda y realizando una producción cada vez más personalizada a las necesidades de los clientes [4]. Además, las fábricas del futuro deberán dar soporte al intercambio de información en el marco de numerosas aplicaciones en las que intervendrán un número cada vez mayor de nodos móviles (robots, maquinaria y trabajadores). Para dar soporte a estos entornos industriales móviles y con capacidad de adaptación y reconfiguración, es necesario el uso de redes de comunicaciones móviles e inalámbricas que minimicen el uso de cables para aumentar la flexibilidad del sistema y permitan la reconfiguración de la red adaptándose a las necesidades cambiantes del entorno industrial. Estas comunicaciones inalámbricas deben ser fiables y garantizar la transmisión de datos de manera determinista. En este sentido, la tecnología de 5ª Generación (5G) de redes móviles es una excelente opción para las comunicaciones industriales debido a las características que ofrece. Por un lado, 5G brinda la posibilidad de implementar redes privadas en entornos industriales, lo que permite un mayor control, seguridad y fiabilidad de las comunicaciones inalámbricas. Por otro lado, 5G ofrece una mayor capacidad, baja latencia y conectividad masiva, lo que hace a esta tecnología idónea para satisfacer los exigentes requisitos de comunicación de las aplicaciones industriales. En cambio, 5G no es capaz de ofrecer el determinismo demandado por las aplicaciones industriales más exigentes. Para alcanzar el nivel de flexibilidad requerido por las fábricas del futuro a la vez que se garantiza el nivel de determinismo en las comunicaciones demandado por las aplicaciones industriales, es necesaria la integración eficiente y efectiva de las redes 5G y TSN, de manera que se exploten de forma conjunta los beneficios de cada una de estas tecnologías.

En las últimas versiones de los estándares del *Third Generation Partnership Project* (3GPP), se aborda la integración de 5G con TSN. El 3GPP propone un modelo de integración en el que la red 5G se integra en la red TSN como un *bridge* TSN virtual, con el objetivo de garantizar una operación fluida entre la red 5G y TSN [5] (Figura 1). Un *bridge* TSN es un *switch* Ethernet que recibe y transmite flujos TSN según el *scheduling* de TSN. Los flujos TSN se transmiten entre dos dispositivos finales a través de uno o más *bridges* TSN o un *bridge* 5G virtual. La gestión de la red TSN se realiza de manera centralizada por el CNC (*Centralized Network Configuration*). El CNC está encargado de establecer el *scheduling* de los distintos flujos TSN en los distintos *bridges* que forman la red, es decir, a través del *scheduling* TAS (*Time Aware Shaper*) decide los instantes de tiempo de transmisión de cada flujo TSN en los distintos *bridges* de la red. El *scheduling* se realiza con el objetivo de cumplir con los requisitos extremo a extremo de los flujos TSN y conociendo las capacidades de los distintos *bridges* (TSN y 5G virtual). Cada *bridge* por tanto se configura para cumplir con el *scheduling* planificado y cumplir con los tiempos de

llegada y salida del *bridge* especificados por el *scheduling* de TSN. Aunque la arquitectura de integración de TSN y 5G está bien definida, 3GPP no define como coordinar los *schedulers* de ambas redes para conseguir un servicio determinista extremo a extremo. Es importante recordar que 5G es capaz de proporcionar comunicaciones con muy baja latencia, pero no es capaz de ofrecer el determinismo proporcionado por TSN. En este contexto, operar de manera efectiva en un sistema que integra las redes 5G y TSN sigue siendo un desafío [6].

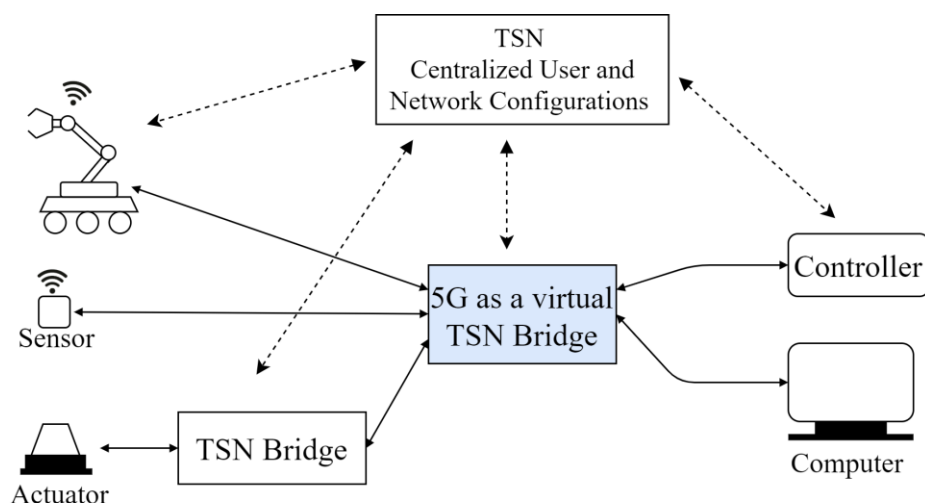


Figura 1. Integración de las redes 5G y TSN para aplicaciones de la Industria 4.0.

El diseño del esquema de *scheduling* en 5G debe tener en cuenta las características del tráfico que generan las aplicaciones industriales y que se va a enviar a través de la red 5G-TSN integrada. El tipo de tráfico más común generado por las aplicaciones industriales es tráfico periódico y determinista [7]. Los dispositivos finales generan paquetes de datos de manera periódica que son transmitidos a través de la red integrada 5G-TSN hacia otros dispositivos finales, esto es lo que se conoce como flujo de datos TSN. Además, los datos deben ser recibidos cumpliendo con una latencia determinada (entre unos límites de tiempo). 5G introduce el uso de mecanismos de *scheduling* semi-estáticos referenciados como *Configured Grant* (CG) en enlace ascendente o *uplink* (UL) y *Semi Persistent Scheduling* (SPS) en enlace descendente o *downlink* (DL). Estos mecanismos asignan de forma previa recursos radio periódicos a los usuarios (UEs) y estaciones base (gNB o *Next generation Node B*) para que puedan transmitir en los recursos preasignados tan pronto como se generen los datos eliminando la necesidad de tener que solicitar recursos radio para la transmisión de cada paquete. Los esquemas de *scheduling* semi-estáticos son una herramienta muy útil para dar soporte a tráfico periódico que permiten disminuir la latencia de transmisión y la señalización en la red. Sin embargo, dadas las características y requisitos demandados por el tráfico TSN, es necesario diseñar mecanismos de *scheduling*

que permitan garantizar en cualquier situación los requisitos de determinismo demandados. Por ejemplo, si existen flujos de datos TSN con distintas periodicidades la asignación de recursos radio periódicos puede dar lugar a conflictos entre distintos paquetes. Es decir, los recursos asignados a diferentes flujos TSN pueden superponerse en frecuencia y tiempo después de varios periodos, desembocando así en un problema crítico de fiabilidad en aplicaciones industriales.

1.1. Objetivos

El objetivo de esta tesis es avanzar hacia la integración eficiente y efectiva de las redes 5G y TSN para casos de uso industriales. Para ello, esta tesis se centra en la coordinación de los esquemas de *scheduling* de las redes 5G y TSN para garantizar los requisitos extremo a extremo demandados por las aplicaciones industriales en términos de determinismo y latencia. En concreto, en esta tesis se estudian y se diseñan nuevos mecanismos de *scheduling* para redes 5G integradas en redes TSN que realizan la gestión de los distintos flujos a transmitir de manera coordinada con el *scheduling* realizado por la red TSN. La tesis se centra en la gestión del tráfico enviado a través del enlace ascendente o UL por la red 5G, considerando que conlleva mayor dificultad gestionar y coordinar el acceso de los UEs al canal radio. A continuación, se enumeran y se detallan los objetivos que aborda esta tesis.

Objetivo 1: Conocer el estado actual de estandarización de la integración de las redes 5G y TSN realizando el estudio de los estándares y documentos de trabajo del 3GPP, así como de distintas asociaciones como 5G-ACIA (*5G Alliance for Connected Industries and Automation*). Además, se enmarca en este objetivo también conocer a fondo el estado de desarrollo de las tecnologías 5G y TSN estudiando los estándares del 3GPP donde se define 5G e IEEE (*Institute of Electrical and Electronics Engineers*) 802.1Q donde se define TSN para proponer soluciones que cumplan con los estándares de las distintas tecnologías. Este primer objetivo también abarca conocer el estado del arte sobre mecanismos de coordinación o esquemas de *scheduling* coordinados para las redes 5G y TSN para la identificación de retos que deben ser todavía abordados.

Objetivo 2: Analizar la integración 5G-TSN y cuantificar los requisitos de latencia que la integración de redes TSN y 5G en un entorno de Industria 4.0 establece sobre el diseño de un *bridge* 5G virtual. Para ello, se considerarán aplicaciones industriales priorizadas por la asociación industrial 5G-ACIA.

Objetivo 3: Diseñar y evaluar esquemas de *scheduling* para redes 5G para dar soporte a tráfico TSN teniendo en cuenta las características de este tipo de tráfico. En particular, se estudiarán redes 5G integradas en redes TSN como un *bridge* virtual. Los esquemas de

scheduling propuestos para redes 5G deberán ser capaces de coordinar sus decisiones con el *scheduling* de la red TSN de manera que se garanticen los requisitos de latencia y determinismo de las comunicaciones extremo a extremo. Para alcanzar este objetivo, se hará uso del conocimiento adquirido en la consecución del objetivo 1 así como el conocimiento y resultados obtenidos sobre los requisitos de latencia que deben garantizarse en el *bridge* 5G virtual.

Objetivo 4: Garantizar que las propuestas de coordinación de las redes 5G y TSN y los esquemas de *scheduling* propuestos cumplan con los estándares de IEEE 802.1Q y 3GPP. Además, se podrá identificar y proponer modificaciones o mejoras al estándar para mejorar la gestión, rendimiento y eficiencia de las redes 5G.

Objetivo 5: Analizar la disponibilidad de herramientas de simulación para evaluar las propuestas realizadas en esta tesis. En el caso de que no las hubiera, se desarrollarán las herramientas de simulación necesarias para la evaluación de las propuestas realizadas. En el desarrollo de estas herramientas, se analizará la necesidad de implementación de los distintos procesos, funcionalidades y características de las redes 5G y TSN que influyan en la evaluación a realizar, buscando un compromiso entre complejidad y coste de implementación y precisión en la emulación de las redes y sus procesos. Siempre que sea posible, las herramientas de simulación desarrolladas se dejarán en abierto para la comunidad investigadora.

1.2. Estructura y resumen de la tesis

1.2.1. Organización

Esta memoria de tesis doctoral se organiza en capítulos que resumen las contribuciones científicas realizadas por la autora de esta tesis, abordando todos los objetivos mencionados previamente. Cada capítulo incluye una introducción y los resultados más importantes obtenidos junto con una discusión detallada de los mismos. En la introducción de cada capítulo se hace referencia a las publicaciones que amplían y detallan los resultados presentados en cada capítulo. Al final de cada capítulo se incluye un apartado de resumen y conclusiones que recapitulan los resultados y aspectos importantes abordados en ese capítulo, remitiendo a las publicaciones originales para obtener más detalles.

El capítulo 2 introduce los requisitos de las comunicaciones industriales y los fundamentos teóricos de los estándares de TSN, de 5G y de la integración de ambas redes. Además, en el capítulo 2 también se analizan los trabajos que se han realizado en el estado del arte acerca de la integración de 5G y TSN y se expone la problemática que se aborda en esta tesis. Los capítulos 3 a 6 desarrollan las contribuciones científicas de esta tesis y

presentan los resultados obtenidos, incluyendo su motivación y discusión. Las conclusiones de la tesis junto con las posibles líneas de investigación futuras se presentan en el capítulo 7. Por último, el Anexo A contiene una copia de los artículos publicados incluidos en esta tesis por compendio, mientras que el Anexo B describe de manera más específica cómo se han calculado los parámetros de los escenarios evaluados para los capítulos 4 y 5.

1.2.2. Materiales y métodos

Con el fin de alcanzar los objetivos planteados en esta tesis, se ha realizado un estudio que abarca varios aspectos. En primer lugar, se ha hecho un análisis de las aplicaciones y los requisitos que se deben cumplir en el contexto de la Industria 4.0. Este análisis ha permitido comprender las necesidades específicas de las aplicaciones industriales y su impacto en las comunicaciones. Posteriormente, se ha estudiado el funcionamiento de las comunicaciones tanto en la red 5G como en la red TSN; este estudio se ha centrado principalmente en los mecanismos de gestión de los recursos. Se han examinado las características y los modos de funcionamiento de cada una de estas redes de manera individual. Además, se ha prestado especial atención a los avances y desarrollos relacionados con la integración de 5G y TSN. Se han revisado los estándares y propuestas existentes para entender cómo estas dos tecnologías pueden coordinarse para poder soportar nuevas aplicaciones industriales. En el caso de TSN, se ha estudiado el conjunto de estándares del IEEE 802.1 que definen el funcionamiento de la red TSN. Los estándares principales para esta tesis están definidos en IEEE 802.1Q [8]. Por otro lado, el conjunto de estándares de 3GPP 5G define el funcionamiento de la red móvil en todos sus aspectos, mientras que la integración de ambas redes está definida principalmente en 3GPP TS 23.501 [5]. El contenido más relevante de estos estándares está resumido en el capítulo 2.

Con el fin de comprender los retos que la integración 5G-TSN supone, se propone en primer lugar un estudio analítico de los requisitos de latencia que las aplicaciones de la Industria 4.0 imponen a la red integrada 5G-TSN. En particular se cuantifican los requisitos de latencia que debe satisfacer la red 5G en el marco de esta integración. Para ello, se ha considerado el modelo de integración definido por el 3GPP en el que la red 5G funciona como un *bridge* TSN virtual. Este análisis tiene en cuenta la latencia introducida por los distintos procesos que tienen lugar en esta integración y el intercambio de información que se realiza entre las redes 5G y TSN. En este estudio se tienen en cuenta las indicaciones de la asociación industrial 5G-ACIA para establecer el escenario y marco de estudio. Este trabajo se ha recogido en el capítulo 3.

En el marco de esta tesis doctoral, se adopta el uso de simulaciones para llevar a cabo la evaluación y optimización de los nuevos mecanismos de *scheduling* en redes integradas 5G-TSN presentados en los capítulos 4 y 5. Se opta por esta metodología científica frente

a, por ejemplo, desarrollos analíticos dada la necesidad de implementar todos los procesos que afectan a la integración (5G-TSN), y la dificultad de modelarlos de forma analítica sin realizar suposiciones que simplifiquen su modelado, pero que puedan poner en riesgo la validez científica que se puede derivar de los modelos. En concreto, se ha hecho uso de Matlab, en el que se han simulado las características del sistema y de la red que afectan al rendimiento de las propuestas de *scheduling*, y que permiten por tanto realizar una evaluación precisa del rendimiento, validez y beneficio de los esquemas de *scheduling* propuestos. Además, se ha utilizado la herramienta *Parallel Computing, Optimization y Symbolic Math* de Matlab para resolver problemas de optimización. Las simulaciones se han llevado a cabo en un servidor con una CPU Intel(R) Core (TM) i7-5930K @ 3.50GHz y RAM de 16GB.

También se ha trabajado con el simulador ns-3 5G-LENA [9], implementando nuevas funcionalidades introducidas por 5G necesarias para la aplicación de los mecanismos de *scheduling* desarrollados en esta tesis para dar soporte al tráfico industrial. Este trabajo se ha llevado a cabo en el capítulo 6, con el objetivo de poder utilizar los esquemas propuestos en esta tesis en futuros trabajos e investigaciones realizadas en Ikerlan en las que se hará uso de esta plataforma para aplicaciones cercanas a la industria.

1.2.3. Resultados y discusión

Para una integración eficaz y eficiente de las redes 5G y TSN, la coordinación de los *schedulers* de ambas redes es fundamental. En esta red integrada, el *scheduling* de los paquetes de los flujos TSN en la red 5G debe realizarse de manera que se cumplan los tiempos de llegada y salida previamente planificados por la red TSN. Como primer resultado de la tesis, en el capítulo 3 se ha realizado un análisis de las latencias requeridas en el *bridge* 5G virtual para soportar tráfico TSN. En este estudio se ha considerado una aplicación industrial presentada por 5G-ACIA en [10] que representa el reto que supone la integración 5G-TSN para dar soporte a aplicaciones de la Industria 4.0. En este ejemplo, una red 5G-TSN proporciona conectividad a una aplicación industrial que controla una planta logística que trabaja en lazo cerrado. A partir de este análisis, se han derivado expresiones para calcular el requisito de la latencia máxima que pueden soportar los paquetes de cada flujo TSN dentro de la red 5G. Además, se ha determinado el nivel de calidad de servicio (*Quality of Service* o QoS) que 5G debe garantizar para los flujos del ejemplo de 5G-ACIA. También se ha realizado una descripción detallada de la integración de ambas redes, destacando la importancia que tiene el *scheduling* en 5G para que la integración de 5G y TSN sea correcta. Estos primeros resultados se recogen en el artículo [11].

Como segundo resultado, en el capítulo 4 se han diseñado y evaluado dos esquemas de *scheduling* basados en CG para redes 5G integradas con redes TSN. El objetivo de ambos esquemas propuestos es alcanzar una solución de *scheduling* en 5G que tome decisiones de manera coordinada con las decisiones previamente tomadas por el *scheduling* de la red TSN. El *scheduling* de TSN determina el tiempo de llegada y salida de cada paquete en cada *bridge* TSN y 5G virtual. Para establecer el requisito de latencia de cada paquete dentro de la red 5G, se han utilizado las expresiones derivadas en el estudio analítico presentado en el capítulo 3. En la literatura, los esquemas de *scheduling* basados en CG suelen asignar recursos radio de manera periódica para cada flujo TSN, en función de la periodicidad de los paquetes (por ejemplo, [12]). Sin embargo, esta estrategia puede ocasionar conflictos entre paquetes de diferentes flujos TSN, ya que los recursos radio asignados a flujos con distintas periodicidades pueden solapar en tiempo y frecuencia en algún momento. Para resolver esta problemática, ambos esquemas de *scheduling* de 5G propuestos en el capítulo 4 buscan satisfacer la demanda de recursos radio y los requisitos de latencia de todos los paquetes de todos los flujos TSN generando múltiples configuraciones CG para cada flujo TSN. Una vez alcanzado este objetivo, los esquemas de *scheduling* propuestos intentan minimizar la latencia experimentada por los distintos flujos TSN. El primer *scheduling* propuesto aplica técnicas de programación lineal entera binaria para encontrar la solución óptima al problema que minimiza la latencia que experimentan los flujos TSN. El segundo *scheduling* propuesto está basado en un algoritmo heurístico que busca alcanzar la mejor solución posible, pero manteniendo un coste computacional bajo. Los resultados obtenidos en este capítulo demuestran que ambos esquemas mejoran significativamente la cantidad de flujos TSN que pueden ser atendidos en la red 5G-TSN integrada en comparación con el esquema de referencia. Los resultados del *scheduling* heurístico y basado en optimización están recogidos en los artículos de conferencia [13] y [14], respectivamente.

Es importante destacar que las propuestas presentadas en el capítulo 4 están enfocadas a buscar la mejor solución y, por tanto, no consideran la limitación impuesta por el 3GPP, que establece un máximo de 12 configuraciones CG por UE [15]. Esto implica que tanto los *schedulers* propuestos en el capítulo 4 como la fórmula que utiliza el UE para identificar a partir de una configuración CG los recursos radio a utilizar en la transmisión de cada paquete tal y como está estandarizado en [16], pueden ser mejorados para cumplir con esta restricción. En este contexto, en el capítulo 5 se ha ampliado la investigación del capítulo 4, considerando la limitación del número de configuraciones CG que se deben crear para admitir todos los flujos TSN. En este capítulo se ha desarrollado y evaluado un esquema de *scheduling* basado en CG para redes 5G integradas con redes TSN, partiendo del esquema de *scheduling* heurístico diseñado en el capítulo 4. Este nuevo esquema de *scheduling* tiene los mismos objetivos que los esquemas de *scheduling* del capítulo 4, pero,

a su vez, intenta minimizar la cantidad de configuraciones CG que necesita crear. Para cumplir con esta limitación, se ha propuesto un vector de activación y desactivación de las configuraciones CG asignadas a un UE, tal que el UE va a utilizar este vector de activación para identificar qué configuración CG está activa en cada momento sin necesidad de intercambiar mensajes entre el UE y el gNB. En consecuencia, el UE podrá identificar los recursos radio a utilizar en la transmisión de cada paquete sin sufrir colisiones entre los recursos asignados a otros UEs y, en consecuencia, se disminuirá la cantidad de configuraciones CG necesarias. Por lo tanto, el uso del vector de activación permite coordinar el acceso de los UEs a los recursos radio asignados en las configuraciones CG que pueden solaparse en determinados momentos, dando solución a posibles conflictos de una manera eficiente y sin aumentar la señalización entre el UE y el gNB para activar y desactivar las diferentes configuraciones CG. Para ello, en este capítulo se propone la incorporación de este vector a la fórmula definida en el estándar de 3GPP [16]. Esta fórmula la utilizan los UEs para identificar los recursos radio asignados para la transmisión de los distintos paquetes. Los resultados obtenidos en este capítulo demuestran que el nuevo *scheduler* propuesto puede soportar el mismo número de flujos que el *scheduler* heurístico propuesto en el capítulo 4 pero cumpliendo con el requisito impuesto por 3GPP. Los resultados de este capítulo forman parte de una publicación aún en preparación.

Como último resultado, en el capítulo 6, se ha implementado el mecanismo de *scheduling Configured Grant* y se ha modificado la implementación de OFDMA (*Orthogonal Frequency Division Multiple Access*) en el entorno de simulación ns-3 5G-LENA para poder modelar con precisión la alta flexibilidad ofrecida por 5G NR. El objetivo de este capítulo ha sido preparar el simulador ns-3 5G-LENA para permitir evaluar en trabajos futuros el funcionamiento de estos *schedulers* en conjunto con trabajos de investigación que se quieran llevar a cabo en Ikerlan. Los resultados de latencia obtenidos utilizando CG en 5G-LENA coinciden con los valores de latencia obtenidos en estudios analíticos en la literatura (3GPP TR 37.910 [17]), lo cual valida la implementación de CG en 5G-LENA. Además, se han implementado dos políticas de *scheduling* para demostrar la flexibilidad y capacidad de 5G NR para soportar servicios críticos. Los resultados muestran que el uso de CG con las políticas de *scheduling* propuestas y el modo de acceso múltiple OFDMA reduce la latencia máxima experimentada y esto se logra utilizando menos recursos radio por UE. La implementación de CG y OFDMA en 5G-LENA podrá ser utilizada tanto por Ikerlan como por la comunidad científica para realizar simulaciones a nivel de sistema de red de 5G NR para soportar servicios críticos que demandan bajas latencias. El código de CG implementado en 5G-LENA está disponible en abierto [18] y utiliza la versión de ns-3 disponible en [19]. Con estas mejoras, el simulador de 5G-LENA ahora está preparado para continuar con la línea de investigación abordada en esta tesis. Estos resultados están recogidos en los artículos [20] y [21].

2. Estado del arte de la integración de 5G en redes TSN

Este capítulo presenta el estado actual y los requisitos de las comunicaciones inalámbricas en la industria, así como las principales características de las tecnologías consideradas en esta tesis (TSN y 5G). Esta descripción es fundamental para la realización de la tesis, ya que presenta el conocimiento previo necesario para realizar las propuestas de *scheduling* 5G-TSN desarrolladas en esta tesis. En concreto, el apartado 2.1 presenta los principales casos de uso de la Industria 4.0, tipos de tráfico generado y requisitos de las comunicaciones inalámbricas en entornos industriales. A continuación, se presentan las principales características y el funcionamiento de las tecnologías de comunicaciones TSN y 5G (apartados 2.2 y 2.3, respectivamente). El apartado 2.4 presenta el marco de integración 5G-TSN que se ha definido en los estándares del 3GPP. Finalmente, el apartado 2.5 presenta el estado del arte sobre mecanismos de *scheduling* en redes integradas 5G-TSN propuestos en la literatura.

2.1. Comunicaciones Industriales: Casos de uso 5G-TSN y requisitos

Diversas asociaciones, como 5G-ACIA, han estudiado los requerimientos de comunicación de la Industria 4.0. En el documento [3], han recopilado casos de uso y requisitos industriales definidos por 3GPP [22], *International Electrotechnical Commission* (IEC)/IEEE [23], IEEE [8] e *Industrial Internet Consortium* (IIC) [24]. Los casos de uso principales que se quieren abordar con esta integración y que se incluyen en la Figura 2 son las comunicaciones de controlador a controlador (C2C o *Controller to*

Controller), comunicaciones de controlador a dispositivos (*C2D* o *Controller to Device*) y comunicaciones de dispositivos a ordenadores (*D2Cmp* o *Device to Compute*).

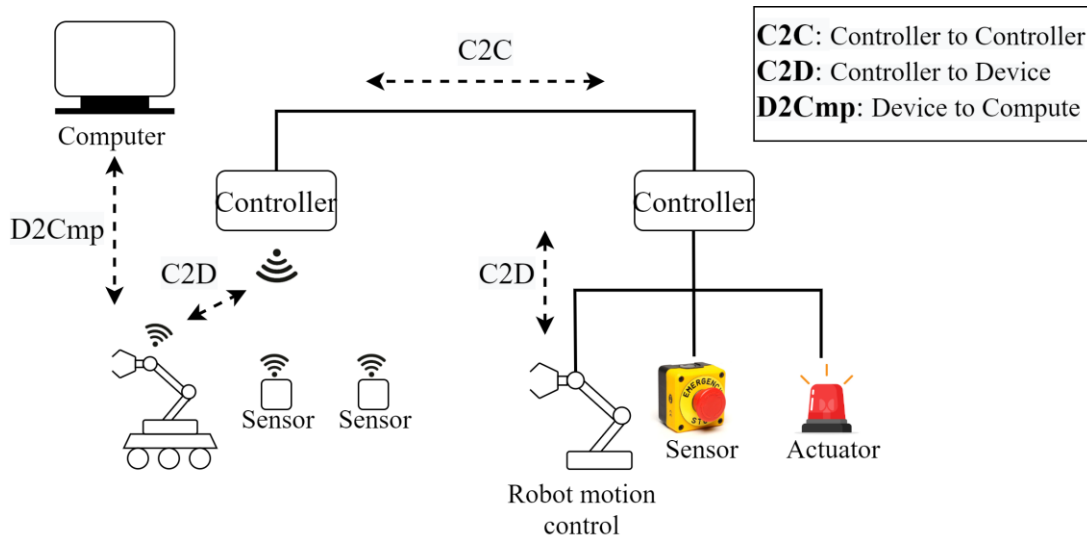


Figura 2. Casos de uso de la Industria 4.0.

La comunicación C2C es necesaria cuando dispositivos con controladores separados interactúan para realizar una tarea conjuntamente [25]. En este contexto, existen dos tipos de comunicaciones C2C: local y remota. Por un lado, se requiere comunicación local cuando los controladores están cerca uno del otro dentro de un mismo entorno, como por ejemplo, cuando se manipula un mismo componente conjuntamente a través de diferentes dispositivos móviles dentro de una fábrica [25]. Por otro lado, la comunicación C2C remota se utiliza para dispositivos que interactúan de manera autónoma con su controlador local y solo necesitan comunicación remota ocasionalmente. Esto se puede ilustrar con un ejemplo de controlador de la línea de producción a controlador (*L2C* o *Line Controller to Controller*), en el que el control remoto de líneas de ensamblaje de PCBs o *Printed Circuit Board*, que habitualmente funcionan de manera autónoma, son controladas de forma remota para implementar cambios en el producto [25].

La comunicación C2D se realiza entre un controlador, que podría ser un *Programmable Logic Controller* (PLC), y sensores y actuadores. Estas comunicaciones se pueden dividir en dos clases: control distribuido y control centralizado. La comunicación y la toma de decisiones en un control distribuido se realiza entre el controlador y sensores/actuadores que están distribuidos en diferentes puntos de la fábrica. En caso del control centralizado, el controlador se coloca en una ubicación centralizada, desde donde realiza el control sobre los actuadores que están distribuidos en la fábrica. En [3], se describe un ejemplo de aplicación del control de un brazo robótico que recoge y coloca objetos en las cajas correspondientes. Este control se basa en la información obtenida a través del análisis de

imágenes en 3D capturadas por una cámara. Si el análisis de imágenes en 3D se realiza de forma centralizada en una sala de control, se utilizaría una comunicación C2D centralizada. En cambio, si los dispositivos de control distribuidos tuviesen capacidad suficiente para analizar las imágenes 3D, se realizaría una comunicación C2D distribuida.

En cuanto a las comunicaciones D2Cmp, no son críticas y se utilizan, por ejemplo, para las aplicaciones con procesos automatizados, donde se monitorizan, se recopila información y se analizan los datos. Con esta comunicación no se controlan los dispositivos. Un ejemplo de comunicaciones D2Cmp podrían ser los campos de petróleo que están instalados en diferentes puntos geográficos y se monitorizan a través de los sensores que informan sobre el estado operativo de los pozos y dispositivos [25].

Los casos de uso definen diferentes requisitos de comunicación en base al tipo de tráfico que se transmite. Los tipos de tráfico que se tienen en cuenta en [3] son: tráfico isócrono¹, tráfico cíclico (síncrono y asíncrono), eventos de control, control de red, configuración y diagnóstico, audio/voz, video y tráfico *Best Effort* (BE). Estos tipos de tráfico se caracterizan según con qué frecuencia se transmiten (periódicamente o esporádicamente), el periodo típico que deben de satisfacer (*Typical Period*) y con qué garantía se debe de satisfacer este periodo (*Data Delivery Guarantee*), la tolerancia al *jitter*, la tolerancia a las pérdidas, el tamaño de paquete y el nivel de criticidad para las aplicaciones industriales. Sus requisitos se resumen en la Tabla 1.

El tráfico periódico, se transmite de manera regular siguiendo un patrón cíclico, mientras que el tráfico esporádico se transmite en cualquier momento, sin seguir un patrón cíclico, como por ejemplo, una alarma de incendios. Los paquetes deben llegar al receptor dentro del requisito establecido por *Data Delivery Guarantee*, en caso contrario el paquete se considera perdido o descartado por la aplicación. Existen tres tipos de garantías para la entrega de los paquetes en el receptor: *deadline*, latencia y *throughput* (están ordenados desde el más restrictivo hasta el menos restrictivo). Para satisfacer la garantía de *deadline* los paquetes deben llegar al receptor en un tiempo específico, es decir, no es tolerante al *jitter*. El *deadline* describe la latencia superior, y normalmente es igual a un periodo de transmisión o *Typical Period* en la Tabla 1. Los paquetes satisfacen el requisito o garantía de latencia cuando los paquetes son entregados al receptor dentro de un intervalo establecido. Esta garantía es menos restrictiva que la de *deadline* ya que permite cierta

¹ El tráfico isócrono es un tipo de tráfico de datos en el cual los paquetes son transmitidos en intervalos de tiempo fijos y consistentes. La sincronización y determinismo son cruciales para este tipo de tráfico [3].

Tabla 1. Tipos de tráfico en la automatización industrial y requisitos de servicios [3].

Traffic types	Periodic/ Sporadic	Typical period	Data delivery guarantee	Tolerance to jitter	Tolerance to loss	Typical data size (Bytes)	Criticality
Isochronous	P	100 μ s ~ 2ms	Deadline	0	None	Fixed: 30 ~ 100	High
Cyclic- Synchronous	P	500 μ s ~ 1ms	Latency bound (t)	$\leq t$	None	Fixed: 50 ~ 1000	High
Cyclic- Asynchronous	P	2 ms ~ 20 ms	Latency bound (t)	$\leq t$	1 ~ 4 Frames	Fixed: 50 ~ 1000	High
Events: control	S	10 ms ~ 50 ms	Latency bound (t)	n.a.	Yes	Variable: 100 ~ 200	High
Events: alarm & operator commands	S	2 s	Latency bound (t)	n.a.	Yes	Variable: 100 ~ 1500	Medium
Network control	P	50 ms ~ 1 s	Throughput	Yes	Yes	Variable: 50 ~ 500	High
Configuration & diagnostics	S	n.a.	Throughput	n.a.	Yes	Variable: 500 ~ 1500	Medium
Video	P	Frame Rate	Throughput	n.a.	Yes	Variable: 1000 ~ 1500	Low
Audio/Voice	P	Sample Rate	Throughput	n.a.	Yes	Variable: 1000 ~ 1500	Low
Best effort	S	n.a.	None	n.a.	Yes	Variable: 30 ~ 1500	Low

tolerancia al *jitter*. Por último, cuando los paquetes deben de garantizar la entrega de los paquetes cumpliendo un *throughput* específico, es suficiente que los paquetes lleguen al receptor dentro del rendimiento establecido, es decir, no requiere una garantía estricta de tiempo de entrega. El tráfico determinista engloba los tráficos que se caracterizan por garantizar la entrega de los paquetes dentro del *deadline* o latencia especificada. El tráfico periódico determinista es la clase de tráfico más común en la industria [7]. El tráfico no determinista se encuentra comúnmente en aplicaciones que requieren actualizaciones de software o descargas de archivos [7], es decir, no tiene que garantizar la entrega de los

paquetes dentro de un tiempo especificado (por ejemplo, el tráfico BE). En cuanto a la fiabilidad requerida por los diferentes tipos de tráfico, el tráfico isócrono y cíclico síncrono no admiten la pérdida de paquetes en la comunicación, mientras que los tráficos basados en eventos, de configuración, voz/audio, vídeo, y BE tienen una mayor tolerancia a las pérdidas de los paquetes en la red. Por lo tanto, considerando tanto los requisitos de tiempo como de fiabilidad, los tráficos más críticos que se consideran en la industria son el tráfico isócrono, el cíclico, de eventos y el que controla la red. En este contexto, los casos de uso C2D, L2C y C2C utilizan principalmente los tipos de tráfico isócrono, cíclico, de eventos y configuración y diagnóstico. Mientras que en comunicaciones D2Cmp se utilizan los tipos de tráfico menos restrictivos, como por ejemplo: el tráfico de configuración, audio/voz, video y *Best Effort*. Un análisis detallado de todos los casos de uso y sus requisitos se puede encontrar en los documentos de 5G-ACIA [3] y [25].

Existen otras fuentes que también analizan los requisitos de las aplicaciones de la Industria 4.0. Por ejemplo, *National Institute of Standards and Technology* (NIST) clasifica los casos de uso de la Industria 4.0 en cinco clases [26] basándose en los requisitos proporcionados por diferentes organismos estandarizados como *International Society of Automation* (ISA) y *European Telecommunications Standards Institute* (ETSI). La Tabla 2 muestra los requisitos de latencia mínima/típica y fiabilidad, así como el número típico/máximo de nodos que hay en una celda para cada clase. La latencia representa el tiempo que un paquete necesita para ser transmitido y recibido correctamente en la capa de aplicación de una estación final. La fiabilidad de las comunicaciones está sujeta a la probabilidad de error preestablecida, es decir, para que una comunicación sea fiable la cantidad de transmisiones fallidas permitidas debe ser menor que la probabilidad de error. Una transmisión puede ser errónea por diferentes motivos: si se pierde el paquete transmitido, si llega el paquete al receptor más tarde que la latencia máxima determinada, o si se recibe el paquete con errores. En [26], la clase 0 incluye servicios o aplicaciones con los requisitos de tráfico más críticos en los que se envían mensajes de seguridad transmitidos para prevenir daños en equipos o personas. La clase 0 requiere latencias entre 0.5 y 4 ms. Las clases 1, 2 y 3 incluyen casos de uso dedicados al control. La clase 1 se refiere al control regulatorio y su objetivo es mantener las condiciones necesarias, como por ejemplo de velocidad, dentro del proceso controlado. Algunas de las aplicaciones típicas de la clase 1 son cortadores láser, control de acoplamiento/interbloqueo de robots y el control preciso de robots basado en la posición. La clase 2 se refiere al control supervisado y es responsable de ajustar las condiciones de operación para obtener un proceso eficiente. Una aplicación típica de la clase 2 es una supervisión basada en PLC, donde los PLCs envían comandos a los actuadores para realizar tareas. En la clase 3, el control se realiza manualmente, con un ser humano, en lugar de utilizar retroalimentación. Estas tres clases también presentan estrictos requisitos de latencia mínima que varían entre 0.25 ms a 4 ms. Finalmente, la clase

4 consiste en dispositivos de detección y monitorización, y representa datos de tráfico que tienen requisitos de latencia menos estrictos, demandando entre 4 y 50 ms. Como se muestra en la Tabla 2, todas las clases demandan una fiabilidad elevada de 10^{-7} .

Estos casos de uso consideran aplicaciones que involucran dispositivos móviles y objetos rotatorios, al mismo tiempo que exigen el cumplimiento de requisitos estrictos. Por lo tanto, este apartado destaca la necesidad de integrar 5G en una red TSN debido a su flexibilidad, robustez y baja latencia.

Tabla 2. Requisitos de comunicación para una celda de trabajo en una fábrica con comunicaciones inalámbricas definidas por NIST [26].

Requisitos	Clase 0	Clase 1	Clase 2	Clase 3	Clase 4
Latencia mínima/típica (ms)	0.5/4	0.25/4	4/20	0.5/4	4/50
Fiabilidad	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}
# (típico/máx. de nodos)	8/16	10/30	10/30	1/4	100/300

2.2. Red TSN

TSN es una extensión de la capa de enlace Ethernet desarrollada para admitir transmisiones deterministas con requisitos de tiempo real y alta fiabilidad. TSN es un conjunto de estándares abiertos de capa 2 especificados por IEEE 802.1. Las principales características de TSN son la gestión de recursos, la planificación determinista del tráfico, sincronización temporal de los dispositivos y alta fiabilidad. Los mecanismos de gestión de recursos están definidos en IEEE 802.1Qcc [8], donde se definen modelos de configuración TSN y arquitecturas de red totalmente centralizadas, totalmente distribuidas o mixtas. Para lograr transmisiones deterministas, TSN define el estándar IEEE 802.1Qbv, que se encarga de planificar la transmisión de los distintos flujos TSN a través de los distintos dispositivos que conforman la red. Además, TSN define distintos mecanismos de priorización de tráfico en los estándares IEEE 802.3br y IEEE 802.1Qbu. La sincronización temporal de los dispositivos es algo crítico en redes TSN y se realiza a través de *Generic Precision Time Protocol* o gPTP (IEEE 802.1AS) [27], que es necesario para otros estándares de TSN como 802.1Qbv. TSN garantiza alta fiabilidad en las comunicaciones mediante distintos mecanismos como la transmisión de múltiples copias del mismo paquete a través de diferentes rutas (IEEE 802.1CB), y mecanismos que proporcionan protección de la red contra violaciones de ancho de banda, fallos en el funcionamiento y comportamientos maliciosos (IEEE 802.1Qci). A continuación, se proporciona más

información sobre los estándares IEEE 802.1Qcc y IEEE 802.1Qbv por su importancia en la integración de 5G-TSN y, por tanto, su relevancia en el contexto de esta tesis.

2.2.1. Modelos de configuración de la red TSN

Una red TSN está compuesta por estaciones finales (ED o *End Devices*), que son la fuente y el destino de los flujos TSN, y *bridges* que son *switches* de Ethernet especiales capaces de transmitir y recibir tramas TSN de acuerdo con el *scheduling* definido por la red TSN (Figura 3). Los flujos TSN consisten en una secuencia de paquetes que se transmiten periódicamente entre diferentes estaciones finales y deben cumplir requisitos estrictos de latencia de extremo a extremo.

El estándar TSN IEEE 802.1Qcc [8] define tres modelos de configuración de red: un modelo completamente distribuido, uno completamente centralizado y un modelo mixto con configuración de red centralizado y configuración de usuarios distribuido. En el modelo completamente distribuido, solo se tienen en cuenta los EDs y los *bridges* de TSN que se utilizan para transmitir los flujos TSN entre los diferentes EDs. Los EDs comunican directamente los requisitos de los flujos TSN a los *bridges*. Cada *bridge* propaga la configuración de usuario y de red a los *bridges* vecinos. Los *bridges* son responsables de gestionar los recursos localmente. En un modelo de red centralizado y usuario distribuido, las estaciones finales comunican directamente los requisitos del flujo TSN a los *bridges*, y estos reenvían esta información a CNC. Los *bridges* también transmiten su capacidad e información de topología activa al CNC. El CNC tiene una visión completa de la red TSN, por lo que proporciona información de configuración (como por ejemplo el *scheduling* de los flujos TSN) a cada *bridge* que hay en el camino entre los EDs para cumplir con los requisitos de todos los flujos TSN. En el modelo totalmente centralizado (Figura 3), los *bridges* TSN comunican su capacidad de red e información de topología activa al nodo CNC. Los EDs envían sus requisitos de flujo TSN al nodo *Centralized User Configuration* (CUC), que reenvía esta información al CNC (el CUC puede adaptar estos requisitos de flujo de los EDs antes de enviarlos al CNC); el CNC tiene entonces una vista completa de la red TSN y puede determinar la configuración de cada *bridge* para cumplir con los requisitos de los flujos TSN. El CNC puede enviar información de configuración específica al CUC, que este a su vez, puede enviar información de configuración a las estaciones finales.

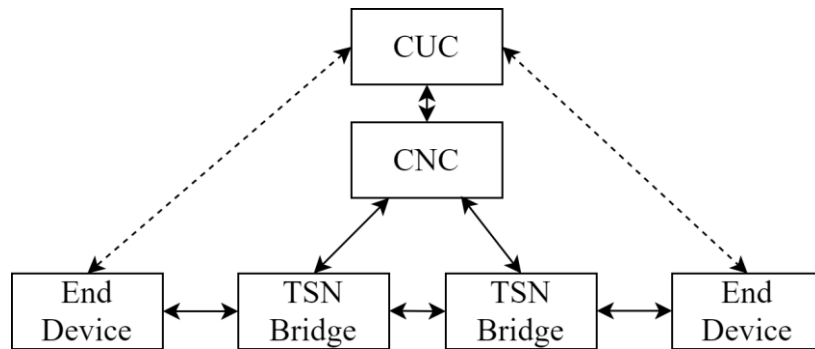


Figura 3. Modelo de configuración de red TSN centralizada.

2.2.2. Scheduling en TSN

TSN define un *scheduler* llamado *Time Aware Shaper* o TAS en IEEE 802.1Qbv [8] para garantizar la transmisión satisfactoria del tráfico determinista en un tiempo acotado. La Figura 4 muestra un *bridge* TSN que puede transmitir tráfico TSN en los instantes de tiempo indicados por el *scheduler* TAS (esta figura está basada en el artículo [28]). El *bridge* TSN de la Figura 4 tiene dos puertos de entrada y tres puertos de salida. Cuando los datos llegan a un *bridge* TSN, se reenvían al puerto de salida correspondiente. En el puerto de salida, los datos se clasifican en diferentes *buffers* basándose en la prioridad del tráfico. Cada *buffer* tiene una prioridad diferente, desde el menos prioritario representado con 0 (tráfico BE), hasta el más prioritario representado con 7. Cada uno de estos *buffers* está controlado por una puerta. Estas puertas tienen dos estados: abierto y cerrado. Los paquetes contenidos en el *buffer* se envían al siguiente *bridge* o estación final cuando la puerta del *buffer* está en estado abierto, en caso contrario, el *buffer* no puede transmitir paquetes.

El *scheduling* TAS establece en cada puerto de salida de cada *bridge* el intervalo de tiempo en el que cada clase de tráfico tiene acceso al canal en base a la información recibida desde el CUC y los *bridges*. Tras crear el *scheduling* TAS o TSN, el CNC se encarga de transmitir la información del *scheduling* TAS a cada *bridge*. De este modo, los *bridges* se configuran con la información del *scheduling* TAS y así garantizan que los paquetes prioritarios cumplan con los requisitos de tiempo demandados. La Figura 5 muestra un ejemplo del *scheduling* TAS, donde se puede observar en que instantes de tiempo cada *buffer* del puerto 1 tiene acceso al canal y puede transmitir sus paquetes. El *scheduling* TAS de este ejemplo tiene dos *slots* que comienzan en el instante de tiempo T1 y T2, respectivamente. Estos *slots* de tiempo se repiten con un periodo común o *Commun Period*.

Para configurar los *bridges* con los parámetros calculados a través del *scheduling* TAS o TSN, se utiliza el mecanismo *Gate Control List* (GCL). La GCL contiene una secuencia de estados. En cada estado, GCL determina qué *buffers* se abren o se cierran. En la Figura 4, GCL está en el primer estado (periodo de tiempo que comienza en el instante T_1 y termina en el instante T_2), por lo tanto, tanto la clase 7 como la clase 0 tienen las puertas abiertas para transmitir sus paquetes. Si hay más de un *buffer* abierto al mismo tiempo, el envío de los flujos TSN depende de la prioridad del *buffer*. Por lo tanto, los paquetes de la clase 7 tienen prioridad de transmisión frente a los paquetes del *buffer* de la clase 0.

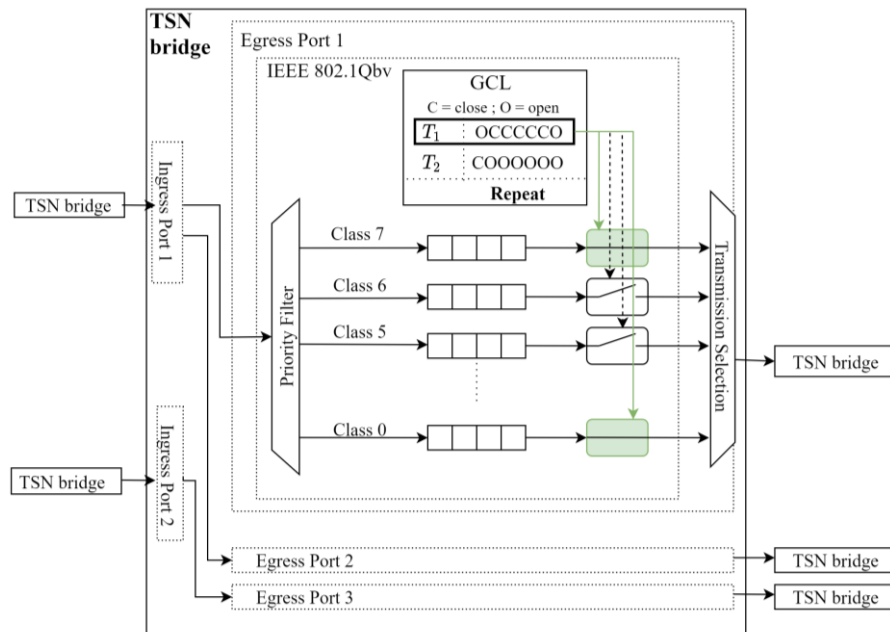


Figura 4. El puerto de salida 1 del *bridge* TSN que admite transmisiones deterministas a través del *scheduler* TAS o TSN.

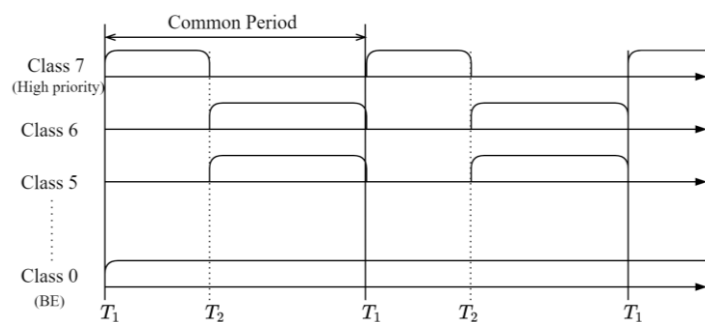


Figura 5. *Scheduling* de TSN para el puerto de salida 1.

Para admitir este tipo de *scheduling*, todos los nodos, tanto los usuarios finales como los *bridges*, deben estar sincronizados utilizando un reloj maestro y varios relojes esclavos, como se define en el estándar IEE 802.1AS.

2.3. Redes móviles 5G

La red móvil 5G ha sido diseñada para satisfacer las necesidades de diversos sectores, como la industria, el transporte, el entretenimiento y la salud. El 5G proporciona tasas de datos más altas (*Enhanced Mobile Broadband* o eMBB), da soporte a una gran cantidad de dispositivos (*Massive Machine Type Communications* o mMTC) y es capaz de garantizar altos niveles de fiabilidad y bajas latencias (*Ultra Reliable Low Latency Communications* o URLLC).

En cuanto a su arquitectura, la red 5G está formada por la red de acceso radio o *Radio Access Network* (RAN) y por la red del núcleo o *Core Network* (CN). La RAN incluye el UE y la estación base 5G, denominada *next generation Node B* o gNB. La CN se encarga del enrutamiento y la señalización de los paquetes entre la RAN y la red de datos externa, como por ejemplo una red TSN o internet. La red 5G puede funcionar en dos modos: *Non Standalone* (NSA) y *Stand Alone* (SA). El modo 5G NSA, utiliza la RAN de 5G junto con la CN de *Long Term Evolution* (LTE), mientras que el modo 5G SA, utiliza tanto la RAN como el CN de 5G.

En este contexto, 5G define una nueva tecnología de acceso radio denominada 5G *New Radio* o 5G NR mucho más flexible que permite adaptarse a distintos requisitos de latencia, ancho de banda y fiabilidad. A su vez, la red 5G CN ofrece una mayor flexibilidad que LTE CN, ya que, en vez de utilizar dispositivos físicos dedicados para implementar las distintas funciones y mecanismos del CN, 5G CN está basado en las tecnologías *Software Defined Network* (SDN) y *Network Function Virtualization* (NFV) que permiten virtualizar estas funciones denominadas funciones de red o *Network Functions* (NF) que se pueden implementar en dispositivos de computación estándares. Las NFs proporcionan una gran flexibilidad de implementación de la red y ofrecen la capacidad de acercar al usuario final las funciones de control del CN. Esta mejora es útil cuando se contemplan casos de uso que requieren bajas latencias. Esta tesis se basa en casos de uso de comunicaciones URLLC en el contexto de la Industria 4.0. Estos casos de uso requieren un sistema que admita el modo de funcionamiento 5G SA.

Para garantizar la seguridad y accesibilidad de los datos en un escenario industrial, las redes privadas son fundamentales. A diferencia de una red que ofrece servicios de red móvil al público en general, una red privada 5G puede proporcionar servicios de red 5G exclusivamente a una organización. Estas redes, también conocidas como redes no públicas

(NPN o *Non-Public Network*), pueden ser implementadas en ubicaciones específicas, como por ejemplo, una fábrica [29].

Los siguientes apartados explican con más detalle diferentes aspectos de la red 5G por la importancia que tienen en la integración de 5G-TSN: las características de 5G NR (apartado 2.3.1), el *scheduling Configured Grant* para transmisiones de paquetes periódicos en UL (apartado 2.3.2), la arquitectura de 5G CN junto con la gestión de los flujos QoS en 5G (apartado 2.3.3) y las redes privadas de 5G (apartado 2.3.4).

2.3.1. 5G NR (*New Radio*)

La nueva interfaz radio definida en 5G, 5G NR, utiliza el esquema de acceso múltiple OFDMA (*Orthogonal Frequency Division Multiple Access*) que permite a los dispositivos compartir recursos radio en los dominios de frecuencia y tiempo. La asignación de recursos radio se organiza en RBs (*Resource Block*) en el dominio de la frecuencia y en *slots* en el dominio del tiempo [30]. El *resource grid* (Figura 6) en el dominio del tiempo se divide en tramas de 10 ms, que a su vez se subdividen en subtramas de 1 ms. En el dominio de la frecuencia, el *resource grid* está limitado por el ancho de banda. Un recurso radio está compuesto por un RB, que contiene 12 subportadoras consecutivas en el dominio frecuencial y un único símbolo OFDM (*Orthogonal Frequency Division Multiplexing*) en el dominio del tiempo. Mientras que el espaciado de subportadoras (*Sub-Carrier Spacing* o SCS) está fijado en 15kHz en LTE, 5G define múltiples numerologías que permiten el uso de diferentes SCS. Una numerología define el valor de SCS y la duración de un *slot*. La numerología² μ puede tomar valores desde 0 a 4, siendo SCS igual a 15, 30, 60, 120 y 240 kHz, respectivamente [31]. La duración de un *slot* está dada por $1/2^\mu$ ms, lo que resulta en una duración de un *slot* de 1 ms para $\mu=0$ teniendo 1 *slot* por subtrama y 0.00625ms para $\mu=4$ teniendo 16 *slots* por subtrama (Figura 6). Un *slot* consta de 14 símbolos OFDM en el dominio del tiempo cuando se utiliza el Prefijo Cíclico Normal (CP). El CP normal se puede utilizar con todas las numerologías. También se puede utilizar un CP extendido con $\mu=2$. En este caso, un *slot* consta de 12 símbolos OFDM. Las numerologías 0, 1 y 2 se pueden utilizar en el rango de frecuencias más bajas (FR1 410 MHz – 7.125 GHz) y las numerologías 2, 3 y 4 se pueden utilizar en el rango de frecuencias más altas (FR2 24.25 GHz – 52.6 GHz). 5G NR permite que las transmisiones comiencen en cualquier símbolo OFDM dentro de un *slot* y que se utilice solo el número de símbolos necesarios para la transmisión. Esto da lugar a transmisiones de *mini-slots* cuando las transmisiones utilizan solo parte de los símbolos de un *slot* o transmisiones de *slots* completos cuando se utilizan todos los símbolos. Las transmisiones de *mini-slot* pueden utilizar entre 1 y 13 símbolos

² En los últimos *releases* de 3GPP se definen numerologías desde 0 a 6 que corresponden al uso de SCS desde 15 kHz a 960 kHz [30].

OFDM en UL y entre 2 y 13 símbolos OFDM en DL [32]. La introducción de múltiples numerologías con un SCS más alto y duraciones de *slot* más cortos, y el uso de transmisiones de *mini-slot* permiten reducir considerablemente la latencia en comparación con LTE que solo consideraba transmisiones de *slots* completos y SCS de 15 kHz.

5G puede utilizar los modos de *Frequency Division Duplex* (FDD) o *Time Division Duplex* (TDD). Por un lado, en TDD, cada *slot* dentro de una trama se puede configurar para transmisiones en UL o en DL. A su vez, un único *slot* se puede dividir en segmentos de símbolos consecutivos que se pueden utilizar para DL o UL. Por otro lado, FDD organiza las transmisiones en UL y en DL en frecuencias separadas. FDD puede reducir la latencia de la comunicación ya que se pueden realizar transmisiones en UL y en DL simultáneamente.

5G NR utiliza las modulaciones *Quadrature Phase Shift Keying* (QPSK) y *Quadrature Amplitude Modulation* (QAM) de 16, 64 o 256 para los canales de datos. 5G NR define tres tablas de esquemas de modulación y codificación (*Modulation and Coding Scheme* o MCS) que proporcionan un compromiso diferente entre la eficiencia espectral y la protección contra errores. Las tablas de MCS 1 y 2 en [33] garantizan una tasa de error de bloque (*Block Error Rate* o BLER) del 10% y la tabla MCS 3 en [33] garantiza un BLER de 10^{-5} cuando se utiliza de acuerdo con las tablas de CQI (*Channel Quality Indicator*) [33].

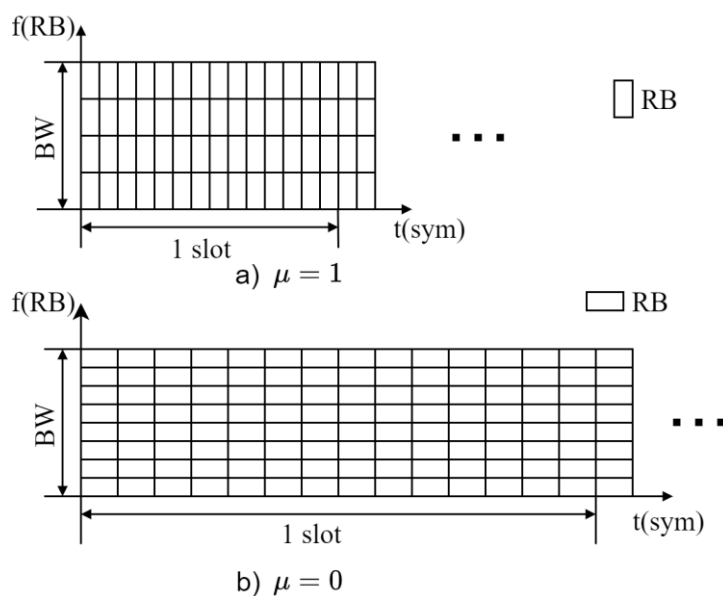


Figura 6. *Resource grid* para numerología 0 y numerología 1.

2.3.2. *Configured Grant* en 5G

Para asignar recursos radio a los UEs y a los gNB, 5G NR utiliza los *schedulers* tanto dinámicos como semi-estáticos. Con el *scheduling* dinámico, el gNB asigna recursos radio cada vez que hay un paquete a transmitir por el canal radio. Cuando se genera un paquete para transmitir en DL, el gNB envía un mensaje de control al UE que indica los recursos radio que se han asignado para la transmisión del paquete en DL. En caso de que sea en UL, el UE envía una solicitud de *scheduling* de recursos (*Scheduling Request* o SR) al gNB. El gNB responde con el mensaje *grant* al UE con la información sobre los recursos radio asignados para transmitir el paquete. El UE trasmite el paquete de datos en los recursos radio que se han indicado en el *grant*, junto con la información del estado del *buffer*. Esta información indica si el UE tiene más datos para transmitir. Si este es el caso, el gNB asignará recursos radio para una nueva transmisión e informará al UE. Este proceso se repite mientras que el UE tenga datos pendientes para transmitir.

El *scheduling* dinámico hace un uso eficiente de los recursos radio. Sin embargo, el intercambio de mensajes de control entre el UE y el gNB antes de transmitir los paquetes aumenta la latencia de transmisión. Esto es más crítico para las transmisiones en UL ya que la cantidad de mensajes intercambiados entre el UE y el gNB es mayor. Por lo tanto, un *scheduler* dinámico puede no ser el más adecuado para las aplicaciones industriales que requieran bajas latencias. Como solución, 5G NR proporciona el *scheduling* semi-estático para la transmisión de paquetes: SPS (*Semi Persistent Scheduling*) en DL y CG (*Configured Grant*) en UL. CG y SPS asignan recursos radio a los UEs antes de que se generen los paquetes de datos, y esta asignación se repite con una periodicidad dada. Cuando se genera un paquete de datos, este puede transmitirse en los recursos preasignados. CG y SPS evitan el intercambio de señalización entre el UE y el gNB para solicitar/informar sobre los recursos radio asignados, lo que reduce la latencia de la transmisión. Dentro de una celda de servicio, para cada *Bandwidth Parts* (BWP³), el UE puede tener hasta 12 CGs configurados para transmisiones en UL y hasta 8 configuraciones SPS para transmisiones en DL [15]. Cuando se configuran más de un CG o SPS, la red decide qué configuración CG/SPS está activa en cada momento y puede activar todas las configuraciones simultáneamente.

³ Un BWP es un conjunto contiguo de RBs con una numerología específica en una portadora determinada. Se trata de una forma flexible de asignar los recursos para dar soporte a diferentes capacidades de ancho de banda para los UEs [86].

La creación de las configuraciones CG se realiza durante la configuración de la conexión, donde el gNB estima los recursos radio que necesita el UE en función de las características del tráfico que va a transmitir (por ejemplo, el tamaño y la periodicidad del paquete de datos). Si el gNB dispone de suficientes recursos para abastecer al UE, asigna los recursos radio e informa al UE sobre los recursos radio preasignados y su periodicidad utilizando la señalización de *Radio Resource Control* (RRC). Si el gNB no dispone de suficientes recursos radio para satisfacer la solicitud de un UE, no se establece una conexión. En [15] se definen dos tipos de CGs: tipo1, que activa la configuración CG desde el momento en que se configura mediante la señalización RRC, y tipo 2, donde la configuración CG se activa o desactiva utilizando mensajes de control en DL (*Downlink Control Information* o DCI). La Figura 7 muestra la creación de un CG de tipo 1. Debemos tener en cuenta que si las características del tráfico generado por un UE cambian (por ejemplo, su tamaño o periodicidad), el CG debe modificarse, y esto se puede hacer a través de los procedimientos de señalización RRC definidos en 3GPP [34].

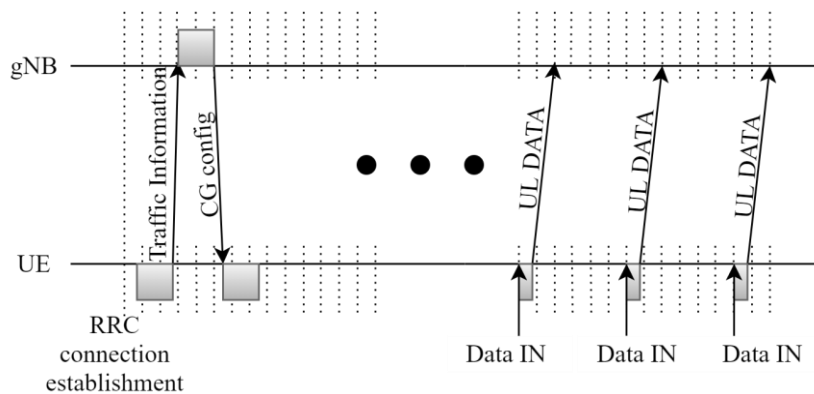


Figura 7. Configured Grant del tipo 1.

El UE determina los recursos radio a utilizar en la transmisión de cada paquete utilizando la expresión (1) dada en 3GPP TS 38.321 en el apartado 5.8.2 [16]. Esta expresión considera una asignación de recursos radio secuencialmente y determina los recursos radio para el N-ésimo paquete ($N \geq 0$).

$$\begin{aligned}
 &[(SFN \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot}) \\
 &+ (\text{slot number in the frame} \times \text{numberOfSymbolsPerSlot}) + \text{symbol number in the slot}] = \\
 &(\text{timeReferenceSFN} \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot} \quad (1) \\
 &+ \text{timeDomainOffset} \times \text{numberOfSymbolsPerSlot} + S + N \times \text{periodicity}) \\
 &\text{modulo} (1024 \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot})
 \end{aligned}$$

En (1), $numberOfSlotsPerFrame$ y $numberOfSymbolsPerSlot$ representan el número de *slots* que hay en una trama y el número de símbolos que hay en un *slot*, respectivamente. El valor de $numberOfSlotsPerFrame$ depende de la numerología, y puede tener desde 1 hasta 16 *slots* dentro de una subtrama en función de si se ha configurado con la numerología 0 o 4, respectivamente. Por otro lado, $timeReferenceSFN$, $timeDomainOffset$ y S identifican los recursos radio asignados para la transmisión del primer paquete del flujo de datos. $timeReferenceSFN$, $timeDomainOffset$ y S representan la trama, el número del primer *slot* asignado dentro de la trama y el primer símbolo asignado dentro de dicho *slot*, respectivamente. La periodicidad con la que se pueden asignar los recursos radio a un UE está definida por el 3GPP en [34] y corresponde al campo *periodicity* en la definición de CG, pudiendo tomar cualquiera de estos valores: 2, 7, $n \cdot 14$ con $n = \{1, 2, 4, 5, 8, 10, 16, 20, 32, 40, 64, 80, 128, 160, \dots\}$ donde el valor n depende del SCS seleccionado. Además, 3GPP ha definido una nueva variable que se llama *periodicityExt* que considera todas las periodicidades múltiplos de un *slot* de 14 símbolos OFDM. Si este parámetro está definido en la configuración CG, se ignora el valor de *periodicity* y se utiliza el valor de *periodicityExt* en su lugar. El periodo de tiempo más largo para la sincronización temporal o *timing synchronization* sin que se reinicie a 0 es igual a 1023 SFN (*System Frame Number*), que corresponde a 10240 ms (siendo $1024 \text{ SFN} \times 10\text{ms}$). Por lo tanto, los parámetros definidos en (1) se configuran con referencia a este valor máximo. Finalmente, los parámetros SFN, número de *slot* en la trama ('slot number in the frame') y el número de símbolo en el *slot* ('symbol number in the slot') en (1) identifican el primer símbolo que se ha asignado para la transmisión del paquete N.

La Figura 8 muestra un ejemplo de preasignación periódica de recursos radio y define los valores que tomarían $timeReferenceSFN$, $timeDomainOffset$ y S para este ejemplo. Se considera que los *slots* son completos (de 14 símbolos OFDM) y se utiliza una trama configurada con numerología 1, es decir, cada subtrama contiene 2 *slots*. En este ejemplo, el paquete de referencia o el primer paquete asignado con la configuración CG definirá $(timeReferenceSFN, timeDomainOffset + N \times periodicity, S) = (\text{trama } 1, \text{ slot } 6, \text{ símbolo } 7)$, mientras que el segundo paquete ($N=1$) tendrá los recursos asignados para $(timeReferenceSFN, timeDomainOffset + N \times periodicity, S) = (\text{trama } 1, \text{ slot } 11, \text{ símbolo } 7)$.

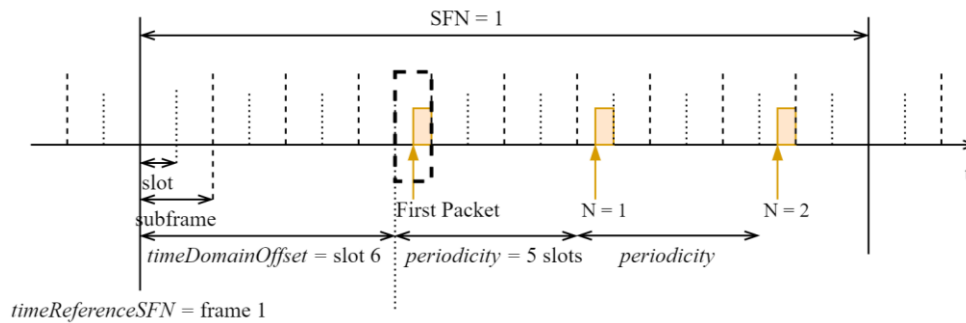


Figura 8. Ejemplo de preasignación periódica a través de la configuración CG.

2.3.3. 5G CN y gestión de QoS

La red 5G CN es altamente flexible y consta de una serie de funciones de red que son independientes para el plano de usuario (*User Plane* o UP) y el plano de control (*Control Plane* o CP). El UP transporta el tráfico de usuario de la red y en el CP se envía el tráfico de señalización y control [5]. La Figura 9 muestra las distintas NF que forman la red 5G CN.

Estas funciones de red se gestionan a través de la NRF (*Network Repository Function*), que actúa como intermediario entre las diferentes funciones de red. Cuando un UE quiere acceder a una red de datos externa a través de la red 5G, primero debe autenticarse y registrarse. Para lograr esto, el UE se comunica con la función AMF (*Access and Mobility Function*) que controla el acceso del UE. El AMF, a su vez, se comunica con la función AUSF (*Authentication Server Function*) que autentifica al UE y almacena las claves de autenticación. Además, el AMF se comunica con la base de datos UDM (*Unified Data Management*) que contiene los datos del usuario.

Una vez autenticado, se debe autorizar los servicios. La función UPF (*User Plane Function*) se encarga de establecer la ruta en el plano de usuario entre el gNB y la red de datos externa. Su objetivo es transmitir los datos desde la red 5G hacia la red de datos externa y viceversa, garantizando el cumplimiento de los requisitos de calidad de servicio (*Quality of Service* o QoS). En el plano de control, la SMF (*Session Management Function*) interactúa con el plano de usuario para establecer o modificar sesiones de PDU (*Packet Data Unit*) que definen los servicios de conectividad entre el UE y la red de datos externa. La SMF es responsable de crear las reglas de QoS que los servicios deben cumplir. Para ello, la SMF interactúa con la PCF (*Policy Control Function*) que proporciona reglas de políticas (reglas PCC o *Policy and Charging Control rules*). La función AF (*Application Function*) se utiliza para intercambiar datos de control con la red externa, como, por

ejemplo, los requisitos que la red 5G debe cumplir para admitir el servicio ofrecido por la red externa. El AF interactúa con el PCF para el control de políticas y puede influir en la ruta del tráfico en el UP.

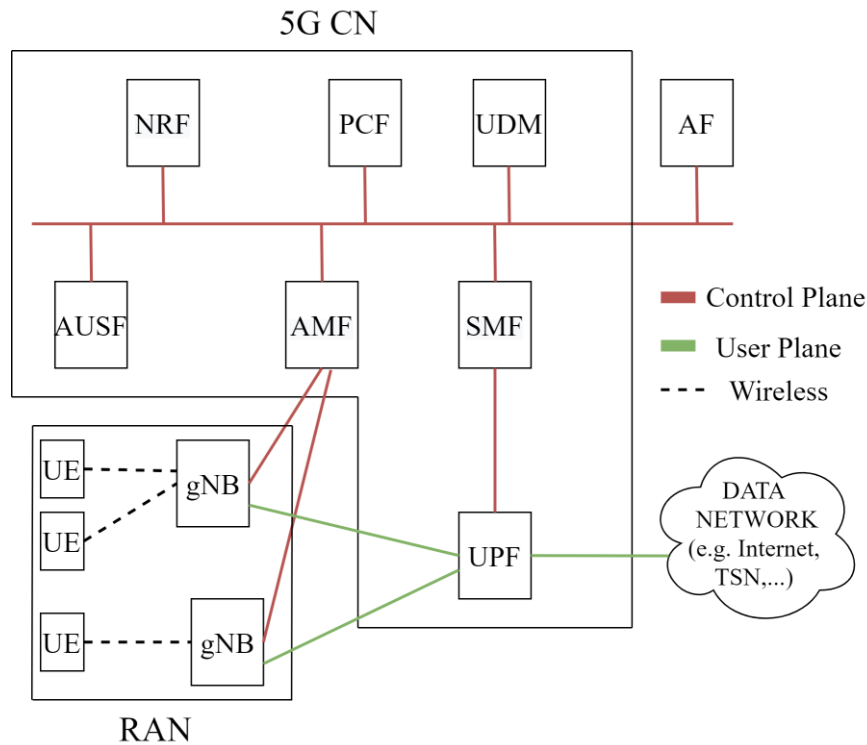


Figura 9. Arquitectura de 5G SA.

Como se introdujo en el párrafo anterior, los paquetes de datos se transmiten de extremo a extremo entre la red de datos externa y el UE a través de sesiones PDU. Dentro de una misma sesión PDU, es posible tener varios servicios que se diferencian entre sí mediante flujos QoS [5]. Un flujo de QoS es la granularidad más fina de diferenciación dentro de una sesión PDU. Un flujo de QoS en 5G puede ser GBR (*Guaranteed Bit Rate*) o no GBR. Además, dentro de los flujos de QoS GBR se diferencia entre los que requieren bajas latencias (*Delay Critical GBR*) y los que no. Cada flujo de QoS está asociado a un perfil de QoS o *QoS profile*. El perfil de QoS está definido por distintos parámetros como, *Guaranteed and Maximum Flow Bit Rate* (GFBR y MFBR, respectivamente), *maximum packet loss rate* y un identificador de QoS de 5G o *5G QoS Identifier* (5QI). El 5QI es un escalador que se utiliza como referencia para un conjunto de características de QoS: prioridad, *Packet Delay Budget* (PDB), *Packet Error Rate* (PER) y *Maximum Data Burst Volume* (MDBV). El PDB es el tiempo máximo que un paquete puede experimentar entre el UE y la UPF que se conecta a la red de datos. El PDB se divide en PDB RAN y PDB CN. El MDBV es la cantidad máxima de datos que la RAN puede servir dentro de un periodo de PDB RAN. Existe una serie de valores de 5QI estandarizados disponibles en [5]. Por

ejemplo, el 5QI igual a 82, con nivel de prioridad igual a 19, se define para aplicaciones de automatización discreta y establece un PDB de 10 ms. Los 5QI también pueden definirse dinámicamente o se pueden preconfigurar. Finalmente, estos valores son utilizados por el gNB para asignar recursos radio a cada UE en base a los requisitos de QoS.

2.3.4. Red 5G Privada

Las redes privadas o NPN también son clave para los casos de uso industriales con requisitos URLLC. Las NPN se dividen en dos categorías. Por un lado, están las NPN desplegadas como redes aisladas y autónomas, y por otro lado están las NPN desplegadas en conjunto con una red pública [29]. La primera categoría implica una configuración única, mientras que la segunda categoría se divide en tres configuraciones según el grado de interacción entre la infraestructura privada con la red pública, tal y como se detallará más adelante.

En una red privada totalmente aislada, la red está totalmente separada de la red pública. La empresa que opera con la NPN tiene la responsabilidad exclusiva de su funcionamiento y todos los aspectos relacionados con el servicio. La única forma de comunicación entre la red privada y la pública es a través de un cortafuegos.

Cuando la red privada funciona en conjunto con una red pública, ciertos casos de uso pueden ser completamente soportados por la red pública, mientras que otros requieren una NPN dedicada. Se consideran tres escenarios: cuando solamente se comparte la red de acceso, cuando se comparte tanto la red de acceso como el plano de control y cuando el tráfico privado y público están alojados en la red pública.

En el primer caso, cuando la red de acceso está compartida por ambas redes, tanto la NPN como la red pública utilizan la misma infraestructura de acceso. Los datos de la red privada se mantienen dentro de un perímetro específico, como una fábrica, mientras que el tráfico de la red pública se transfiere a la red pública. Además, es posible configurar también gNBs adicionales que solo sean accesibles para los usuarios de la NPN. En el segundo caso, tanto la red de acceso como el plano de control están compartidos por ambas redes. En este caso, la red pública es responsable de las tareas de control de red. Sin embargo, los flujos de tráfico de la NPN permanecen dentro de la fábrica, mientras que el tráfico de la red pública se transfiere a la red pública. En el último caso, tanto el tráfico de la red pública como el tráfico de la red privada se encuentran fuera de la fábrica, pero se tratan como si fueran partes de redes completamente diferentes. Esto se logra mediante la virtualización de las funciones de red.

2.4. Integración de redes 5G y TSN

La integración de 5G y TSN proporcionará el nivel de determinismo y flexibilidad demandado por la Industria 4.0. El 3GPP define el marco que hay que seguir para llevar a cabo esta integración [5][35][36]. El 3GPP propone un modelo de integración en el que la red 5G se integra en la red TSN como un *bridge* virtual (Figura 10). En cada puerto de entrada/salida del *bridge* 5G virtual se implementan unos traductores TSN o *TSN Translator* (TT), que funcionan como puntos de interconexión entre 5G y TSN. En los puertos de entrada/salida que corresponden a los UEs se implementa el *Device-Side TSN Translator* (DS-TT), mientras que en los puertos de entrada/salida correspondientes al 5G CN se implementa el *Network-Side TSN Translator* (NW-TT) [37]. Los TT permiten que 5G informe a la red TSN sobre sus capacidades y estado actual. Los TT reciben, interpretan y aplican comandos de configuración de TSN provenientes del CNC. Además, los TT proporcionan funcionalidades de puerto de entrada y salida del *bridge* TSN, lo que implica que retienen y reenvían los flujos TSN de acuerdo con el *scheduler* TAS dado por la red TSN. A este mecanismo se le llama *hold and forward buffering* y permite que los TT en 5G puedan almacenar los paquetes para enviarlos en los instantes de tiempo necesarios de manera que se minimice la variación de la latencia o *jitter*.

Se puede ver también en la Figura 10 que la red integrada 5G-TSN considera un modelo de gestión centralizado. En este modelo centralizado, el CNC recopila información sobre las capacidades y el estado actual de todos los *bridges* que conforman la red, incluyendo el *bridge* virtual 5G [35]. En concreto, cada *bridge* envía al CNC los siguientes datos: ID del *bridge*, puertos Ethernet y sus direcciones MAC, información sobre los usuarios y *bridges* vecinos, latencia mínima/máxima que puede garantizar (*Bridge Delay* o BD), el tiempo de propagación y las clases de tráfico soportadas. En caso de 5G, el TSN AF recibe la información necesaria sobre el *bridge* 5G virtual y la envía al CNC para registrar o actualizar la información sobre un nuevo *bridge*. Con toda la información disponible, el CNC decide la ruta y se calcula el *scheduling* TAS para todos los flujos TSN. Finalmente, el CNC configura los *bridges* TSN y 5G virtual de acuerdo con la decisión del *scheduling* y así garantiza los requisitos de extremo a extremo de todos los flujos TSN. Por cada UE y UPF hay una lista GCL que controla los instantes en el que se transmite el tráfico desde el *bridge* 5G virtual al siguiente *bridge* TSN o dispositivo final. La lista GCL en el UE controla el tráfico que se transmite en DL (desde el gNB al UE), mientras que la lista GCL en el UPF controla el tráfico que se transmite en UL (desde el UE al gNB).

Los requisitos QoS de TSN se deben mapear a los QoS de la red 5G. Este mapeo se realiza a través de la función PCF. La función PCF asigna un identificador 5QI a cada flujo TSN basándose en la información recibida desde el CNC y a su vez desde el TSN AF,

como se describe en el apartado 5.28.4 en [5]. Los flujos TSN en la red 5G se clasifican como *Delay Critical* GBR. El anexo I de [5] propone diferentes formas de calcular el MDBV y MFBR en base a la señalización recibida desde CNC. En cambio, los estándares 3GPP no especifican directamente cómo se mapean algunos de los requisitos de QoS de TSN a los requisitos de QoS de 5G, pero existen propuestas que abordan esta cuestión, como se menciona en las referencias [38] y [39].

El TSN AF también determina el *Time Sensitive Communications Assistance Container* (TSCAC) que describe las siguientes características del tráfico de TSN: el tiempo de llegada del paquete con referencia al puerto de entrada, periodicidad, dirección de flujo (enlace ascendente o descendente), *survival time*⁴ y el dominio de gPTP del flujo TSN. El PCF recibe la información de TSCAC y lo reenvía al SMF. A continuación, el SMF utiliza el TSCAC para derivar la información de asistencia TSC (*Time Sensitive Communications Assistance Information* o TSCAI) correspondiente a ese flujo de QoS y envía tanto el TSCAI como el 5QI al gNB. El TSCAI especifica la periodicidad, el momento en que el paquete llega a 5G, dirección del flujo y tiempo de supervivencia o *survival time* en relación con el reloj de 5G. Al igual que el 5QI, en el anexo I de [5] también hay propuestas para calcular el tiempo de llegada y la periodicidad de los flujos según los datos recibidos por el CNC. El TSCAI y el 5QI pueden ser utilizados por el gNB para decidir el *scheduling* de recursos radio en 5G para el tráfico TSN, con el objetivo de cumplir con los requisitos de extremo a extremo de los flujos TSN.

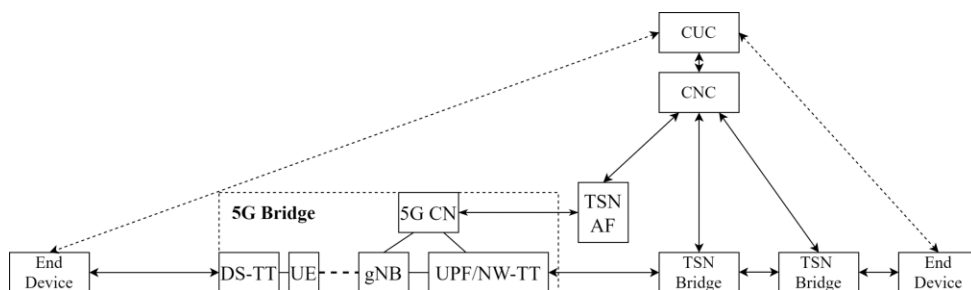


Figura 10. Arquitectura y modelo de integración de las redes 5G y TSN.

2.5. Estado de la literatura sobre esquemas de *scheduling* en redes 5G-TSN integradas

La coordinación de los *schedulers* en 5G y TSN es clave para lograr una integración efectiva y eficiente y cumplir con los requisitos de las aplicaciones de la Industria 4.0. En la literatura, se pueden identificar dos enfoques principales para abordar este problema. En

⁴ El tiempo de supervivencia o *survival time* es el tiempo en el que una aplicación puede mantenerse en funcionamiento sin recibir ningún paquete, según se describe en [5].

primer lugar, existen algunos trabajos que proponen abordar el problema de *scheduling* en la red TSN de manera conjunta con el *scheduling* de los paquetes en la red 5G. Para ello, plantean un problema de optimización que considera las restricciones que ambas redes deben cumplir [6][40][41]. David Ginthör et al. en [6] proponen un problema de optimización para resolver el *scheduling* conjunto de la red TSN y la red 5G. El problema de optimización tiene como objetivo minimizar la cantidad de recursos desaprovechados o no utilizados en ambas redes TSN y 5G. En [40], David Ginthör et al. buscan identificar una estrategia de *scheduling* óptima que mejore la capacidad de asignación de recursos radio de manera dinámica en la red 5G, teniendo en cuenta la incertidumbre del canal inalámbrico. Xiaolong Wang et al. en [41] formulan un problema de optimización para el *scheduling* conjunto de 5G-TSN para maximizar el número de flujos TSN que puedan garantizar los requisitos de latencia extremo a extremo. En [42], Sushmit Bhattacharjee et al. presentan un orquestador TSN orientado a *network slicing* que se comunica tanto con el CNC como con el controlador SDN para optimizar las operaciones de *slice-aware/unaware*. Este trabajo se centra en la interacción entre TSN y 5G para la gestión de los *network slices*, pero no especifica en detalle como interactúan ambos *schedulers*.

Sin embargo, abordar de forma conjunta el *scheduling* en ambas redes conlleva un alto coste computacional. Por este motivo, otros trabajos proponen ejecutar el *scheduling* de la red TSN y de la red 5G de manera independiente pero coordinada, es decir, teniendo en cuenta información compartida entre ambas redes. En este caso, en primer lugar, se realiza el *scheduling* de la red TSN teniendo en cuenta las capacidades de la red 5G. Posteriormente, la red 5G realizará el *scheduling* de los paquetes a transmitir por el canal radio teniendo en cuenta la planificación realizada por el CNC de la red TSN. Por lo tanto, existen trabajos que se enfocan en el desarrollo del *scheduler* de 5G para gestionar adecuadamente los flujos TSN. Estos trabajos se centran en el diseño de esquemas, tanto dinámicos como semi estáticos, que permiten asignar de manera eficiente los recursos radio y garantizar la calidad de servicio requerido por los flujos TSN [43]–[48].

Jingwei Yang et al. en [43] proponen maximizar la relación entre el uso de recursos radio en 5G para usuarios que no tengan requisitos de latencia frente a la latencia experimentada por los flujos TSN. Para ello, en primer lugar, asignan recursos radio a los flujos TSN, desde el más prioritario hasta el menos prioritario, con el objetivo de minimizar la latencia. Para los flujos TSN, el *scheduler* propuesto considera el uso de la modulación más restrictiva para aumentar la fiabilidad. Posteriormente, asignan los recursos radio restantes a los flujos eMBB para maximizar el rendimiento general. La asignación del MCS a utilizar en la transmisión de los flujos eMBB se realiza de forma dinámica, en base a la SNR media de todos los recursos radio asignados para el flujo correspondiente. David Ginthör et al. en [44] consideran una red 5G en la que se aplica *RAN slicing*. En este trabajo

se utiliza un *slice* para tráfico prioritario o flujos TSN, y otro *slice* para tráfico no prioritario (tráfico eMBB). Este segundo *slice* lo dividen en recursos radio que se pueden o no interrumpir. En el diseño del *slice* para el tráfico eMBB se establecen recursos radio que solo pueden ser utilizados por tráfico eMBB y otros en los que la transmisión del tráfico eMBB puede ser interrumpida y utilizada por la *slice* TSN. En concreto, buscan la proporción óptima entre los recursos radio reservados exclusivamente para transmisiones eMBB y los que pueden ser interrumpidos, con el objetivo de cumplir con los requisitos QoS de los servicios, maximizar el rendimiento general de la red y lograr un cierto grado de equidad entre los *slices* y usuarios. Una vez definidas los *slices*, proponen un *scheduler* dinámico basado en el estado actual del canal inalámbrico y las demandas de los flujos TSN. En [45], Renato B. Abreu et al. proponen dos esquemas de *scheduling*, tanto dinámico como semi-persistente, para flujos TSC (*Time Sensitive Communications*) transmitidos en el enlace DL con el objetivo de optimizar la asignación de los recursos radio y maximizar el número de usuarios TSC admitidos en la red. En este estudio, se presenta un esquema de SPS que tiene en cuenta las interferencias entre celdas, sobrecarga de canal de control y condiciones variables del canal, a la hora de asignar recursos radio. El esquema SPS propuesto adapta el MCS de manera lenta, considerando las estadísticas de SINR (*Signal to Interference and Noise Ratio*) experimentadas en el UE durante un gran número de asignaciones periódicas. En este trabajo determinan el intervalo adecuado para actualizar el esquema de MCS en caso de SPS, con el objetivo de minimizar la señalización de control necesaria, al tiempo que se maximiza la eficiencia del *scheduler*. En el artículo [46], Yajing Zhang et al. plantean un mecanismo de *scheduling* predictivo para múltiples prioridades basado en SPS para redes 5G-TSN, que considera tanto datos periódicas como aperiódicos. Presenta un mecanismo de inyección de datos TSN adaptativo, desde el *bridge* 5G virtual a la red TSN, para compensar la variación de la latencia causada por las redes 5G. Para reducir la máxima latencia experimentada en la red 5G y para realizar una mejor predicción de los RBs que hay que reservar, se ha propuesto una estrategia de *risk-sensitive learning*. En [39], Yajing Zhang et al. proponen un esquema dinámico, basado en QoS, para realizar la interconexión entre las redes 5G y TSN. Se centran en los problemas de mapeo de VNF. Primero, formulan un problema de optimización considerando requisitos de *time sensitive*. Después, plantean un algoritmo heurístico para el mapeo y *scheduling* de VNF en la red 5G-TSN. También emplean un mecanismo de prioridad, para prevalecer las aplicaciones prioritarias ante las no prioritarias. Yeping Cai et al. en [47] han propuesto un método de mapeo dinámico de QoS. Es decir, se ajusta de manera flexible el mapeo del 5QI en función de las cargas de la red. Además, presenta un *scheduling* semi-persistente adaptativo de recursos radio que se aplica en dos fases: la primera fase se enfoca en la asignación semi-estática de recursos para tráfico TSN, mientras que en la segunda fase se realiza la asignación dinámica de recurso para el tráfico de baja prioridad aplicando el criterio de maximizar el uso de los recursos del sistema. Los recursos asignados a los flujos TSN son

fijos y se repiten con cierta periodicidad. Si un flujo TSN necesita más recursos que los que ya tiene preasignados, se notifica al gNB, el cual adapta los recursos asignados para realizar su transmisión. En el artículo [48], Dennis Krummacker et al., consideran que un mismo UE puede transmitir varios flujos TSN. Cada uno de estos flujos, será transmitido por un canal lógico diferente. En este contexto, proponen un mecanismo de priorización dinámica de estos canales lógicos, llamado dLCP (*dynamic Logical Channel Prioritization*). Cada clase de tráfico definido en el GCL se mapea a un canal lógico diferente. El gNB asigna RBs para cada UE utilizando el mecanismo *Configured Grant* o CG tal y como está definido en 3GPP. Después, aplicando el mecanismo dLCP, en cada *Transmission Time Interval* (TTI), se selecciona que canal lógico transmitirá los datos en los RBs asignados para dicho UE. En este artículo, concluyen que dLCP permite cumplir con los requisitos de TSN utilizando los mecanismos nativos de la RAN, y permite reducir la complejidad de los traductores TSN propuestos por 3GPP. En el artículo [49], MiJeong Yang et al., proponen un *scheduling* en 5G basado en el mecanismo CG para resolver el problema de desajuste que puede haber entre la periodicidad del tráfico TSN y el periodo con el que se pueden repetir las configuraciones CG. Para ello, en este artículo crean varias configuraciones CG que se activan y se desactivan en base a un *threshold* preconfigurado. En [50], Nan Jiang et al. también quieren resolver el desajuste entre la periodicidad del tráfico y las periodicidades permitidas en 5G NR. En este trabajo proponen un *scheduling* basado en SPS que alinea el periodo de asignaciones hasta que se minimice el desajuste temporal entre el periodo del tráfico y el periodo de repetición de las configuraciones SPS. La Tabla 3 clasifica los artículos que se han mencionado.

Tabla 3. Artículos de esquemas de *scheduling* para redes 5G-TSN.

Tema		Artículos
<i>Scheduling</i> 5G-TSN conjunto		[6][40][41]
<i>Scheduling</i> de los flujos TSN en la red 5G	Dinámicos	[43][44][45]
	Semi-estáticos	[45][46][47][48] [49] [50][51][52]

Como se indica en los artículos [51][52], el uso de CG y SPS es la mejor solución para reducir las latencias y soportar tráfico TSN dentro de la red 5G. En los últimos años, se han realizado varios estudios a cerca del uso de CG y SPS para tráfico periódico. SPS se propuso originalmente para mejorar el QoS de los servicios de Voz sobre protocolo de internet (VoIP) [53]. El CG se propuso para reducir la latencia en el *handshake* entre el gNB y el UE en transmisiones UL, proporcionando un acceso rápido a los UEs [12]. En el caso de la integración 5G-TSN, debido a los diferentes requisitos que cada flujo puede tener, es necesario utilizar múltiples *grants* preconfigurados (MCG o *Multiple Configured Grant*). Un UE puede tener múltiples configuraciones CG para admitir varios servicios o

mejorar la fiabilidad de la comunicación. En el grupo de trabajo 1 de 3GPP RAN [54] [55] [56] se proponen *schedulers* que utilizan múltiples CGs o MCG con diferentes *timeOffset* para un mismo UE, lo que le permite seleccionar la configuración más cercana para iniciar una transmisión sin tener que esperar mucho tiempo hasta el próximo periodo y garantizar el número de repeticiones requerido. Sin embargo, este enfoque genera una gran sobrecarga de señalización de control para activar las diferentes configuraciones en los UEs. En [57], Trung-Kien Le et al., utilizan múltiples configuraciones CG para aumentar la fiabilidad de las comunicaciones. En este artículo proponen el uso de k repeticiones en el que los recursos que se utilizan para transmitir las réplicas se comparten únicamente por un grupo limitado de UEs. En [58], Anusha Gunturu et al., han propuesto un método para seleccionar de manera óptima uno de los *grants* o CG configurados que cumpla con los requisitos tanto de latencia como de fiabilidad, utilizando una predicción de enlace basada en la tasa de información de bits recibida. Cuando el UE genera un paquete, calcula el número de símbolos necesarios para transmitir el tamaño de bloque de transporte. El UE crea tres conjuntos: las configuraciones CG que satisfacen el número mínimo de símbolos para transmitir el tamaño del bloque de transporte y la potencia de transmisión, las configuraciones CG que cumplen los criterios de latencia y las configuraciones CG que cumplen el criterio de fiabilidad. Si más de un CG satisface los criterios tanto de fiabilidad como de latencia, se selecciona la configuración CG que proporciona latencia mínima. Si no es así, el mecanismo propuesto en [58] selecciona el CG del conjunto basado en el criterio de latencia o del conjunto basado en el criterio de fiabilidad según los requisitos de la aplicación. En [59] y [60], Yan Liu et al. han desarrollado el *scheduling* de MCG de *Non-Orthogonal Multiple Access* (NOMA) para mitigar los tiempos de transmisiones y los problemas de congestión de la red. Con NOMA, varios UEs pueden transmitir su paquete utilizando el mismo recurso físico. En este trabajo, se ha propuesto un algoritmo para maximizar el número de paquetes transmitidos con éxito. Por lo tanto, se ha encontrado cuales son los conjuntos de CG más adecuados para minimizar el número de paquetes que no se transmiten dentro del límite de latencia determinado.

En todos los trabajos citados anteriormente sucede que, si las periodicidades de los diferentes CGs no son múltiplos entre sí, los recursos asignados a diferentes flujos pueden superponerse en el tiempo después de un cierto número de periodos. La Figura 11 muestra un escenario que asigna una configuración CG para cada flujo TSN. Este escenario contiene dos flujos TSN con periodicidades iguales a 3 y 4 símbolos OFDM, respectivamente. Como se puede ver en la figura, los recursos asignados a ambos flujos TSN se superponen en el símbolo 12. Esa superposición se produce debido a las diferentes periodicidades de los dos flujos. En el trabajo [61] y [62], Ye Feng et al. diseñan un *scheduler* predictivo que intenta evitar conflictos entre las asignaciones de recursos de distintas configuraciones SPS con diferentes periodicidades. El *scheduler* propuesto asigna un número máximo de RBs en

frecuencia (menor que el número de RBs dentro del ancho de banda) para evitar el solapamiento entre diferentes SPS. Sin embargo, esta restricción puede no ser aplicable si el tamaño del paquete a transmitir es grande y requiere una cantidad de recursos mayor para su transmisión. Además, limita el número de configuraciones que se puedan soportar simultáneamente. En el trabajo [63], Yungang Pan et al. proponen optimizar el número de configuraciones CG creadas donde se tienen en cuenta múltiples periodicidades para crear estas configuraciones. Para ello, proponen un *scheduling* heurístico que combina la transmisión de datos adyacentes. En cambio, no definen en detalle como funcionaría este *scheduling*.

La Tabla 4 clasifica los artículos de esquemas semi-estáticos citados en este apartado. La primera fila representa los *schedulers* que utilizan una configuración CG o SPS por UE, y este *grant* se repite periódicamente. La segunda fila representa *schedulers* que utilizan múltiples configuraciones CG por UE. Esta segunda fila se divide en artículos que no abordan y si abordan la problemática que se ha tratado en esta tesis.

Tabla 4. Artículos de esquemas de *scheduling* semi-estáticos.

Tema		Artículos
Un CG (o SPS) por UE (o gNB)		[12] [45][47][48][52] [53]
<i>Multiple Configured Grant (MCG)</i>	No consideran la superposición	[49][51] [54] [55][56] [57][58][60][61]
	Consideran la superposición	[62][63] [64]

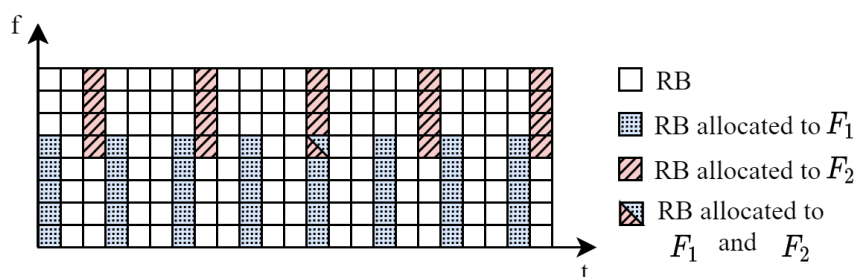


Figura 11. Superposición de recursos radio utilizando diferentes configuraciones CG.

En esta tesis se proponen nuevos *schedulers* para redes 5G integradas en redes TSN, que tienen en cuenta e intentan evitar los conflictos en la asignación de recursos a diferentes configuraciones CG cuando las periodicidades requeridas por distintos flujos no son múltiplo entre sí. Además, los esquemas de *scheduling* definidos cumplirán con el estándar de 3GPP. El objetivo de estos *schedulers* es cumplir con los requisitos de latencia de los

flujos TSN utilizando información proporcionada por la red TSN sobre las características de los flujos y del *scheduler* de TSN.

3. Análisis de la integración de 5G-TSN según el modelo del *bridge* virtual

Para una integración eficaz y eficiente de 5G-TSN, la coordinación entre ambas redes es fundamental. Tal y como se ha presentado en el apartado 2.4, los estándares del 3GPP definen el marco para lograr esta integración. En particular, el 3GPP propone que la red 5G se integre en la red TSN como un *bridge* TSN virtual. En una red integrada 5G-TSN, la ruta de comunicación entre dos dispositivos finales puede establecerse a través de uno o más *bridges* TSN y el *bridge* 5G virtual. Cada *bridge* actúa como un *switch* TSN que recibe y transmite tramas TSN según lo planificado por el CNC de la red. Al comienzo de la comunicación, el CUC recibe la información del tráfico y se lo transmite al CNC. A su vez, todos los *bridges* (TSN y 5G virtual) de la red transmiten sus capacidades al CNC. El CNC tiene información sobre la topología de la red y las capacidades de los *bridges*. En base a esta información, el CNC decide la ruta de comunicación y planifica los tiempos de llegada y salida a cada *bridge* para cada flujo TSN. Por lo tanto, el *scheduling* de los paquetes de los flujos TSN en la red 5G debe hacerse de manera que se cumplan con los tiempos de llegada y salida previamente planificados por el CNC. La Figura 12 muestra el orden que sigue la red 5G-TSN integrada para configurar ambas redes de forma coordinada.

3GPP establece que el *bridge* 5G virtual, junto con la información de las capacidades que le transmite al CNC, debe informar sobre la latencia máxima y mínima que los flujos TSN pueden soportar dentro del *bridge*. Esta información es esencial para poder realizar el *scheduling* de los distintos flujos TSN en la red 5G-TSN integrada. Dada la importancia de

este parámetro, en este capítulo se ha analizado la latencia mínima y máxima que el *bridge* 5G virtual debe garantizar considerando los procedimientos utilizados para establecer una conexión integrada 5G-TSN con tráfico TSN. Para realizar este análisis, se ha considerado un escenario industrial descrito por 5G-ACIA⁵ en [10].

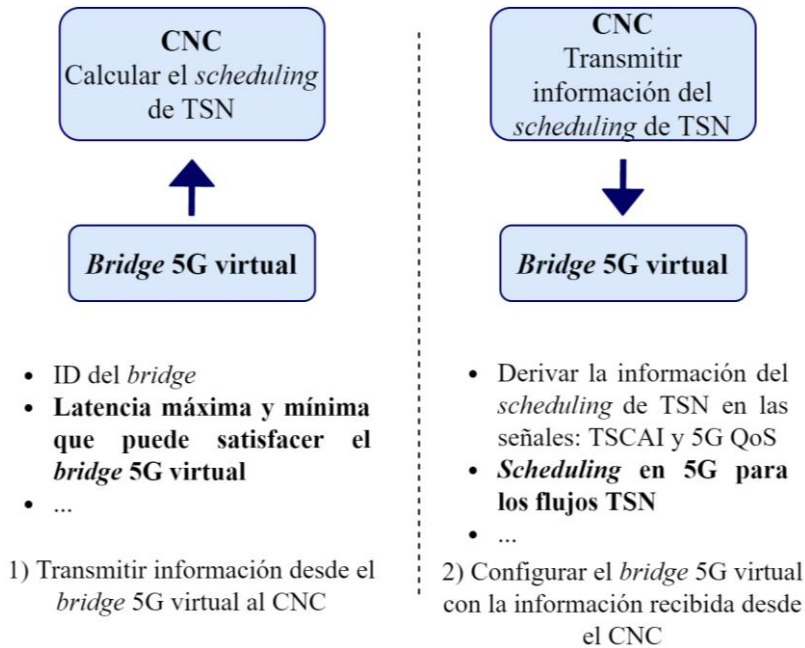


Figura 12. Los pasos que se realizan para configurar una red 5G-TSN integrada.

A continuación, el apartado 3.1 resume los pasos definidos en 3GPP para configurar ambas redes coordinadamente, mientras que el apartado 3.2 describe la cuantificación de la latencia máxima y mínima del *bridge* 5G virtual para el ejemplo de 5G-ACIA. Finalmente, las conclusiones del capítulo se recogen en el apartado 3.3. Este trabajo está publicado en el artículo de conferencia [11].

3.1. Red 5G como *bridge* TSN virtual

Este apartado describe primero la información que el *bridge* 5G virtual envía al CNC (apartado 3.1.1). En base a esta información, el CNC puede calcular el *scheduling* de TSN. Después, se explica la configuración que realiza el *bridge* 5G virtual al recibir la información del *scheduling* de TSN enviada por el CNC (apartado 3.1.2).

⁵ 5G-ACIA es un consorcio constituido por diversas industrias (tales como Audi, Cisco o Bosch) para llevar la tecnología 5G al ámbito industrial.

3.1.1 Información del *bridge* 5G virtual

El estándar de 3GPP TS 23.501 [5] especifica que el *bridge* 5G virtual transmite la siguiente información al CNC: ID del *bridge* 5G virtual, puertos Ethernet y su dirección MAC, información de *bridges* vecinos, la latencia máxima y mínima que puede satisfacer el *bridge* 5G virtual, el tiempo de propagación y las clases de tráfico que soporta. La latencia mínima y máxima que cada *bridge* puede garantizar se calcula para cada puerto de entrada/salida y clase de tráfico que se transmite por dichos puertos. Esta latencia representa el tiempo que un flujo TSN, para una clase de tráfico específica, tarda en ser enviado desde un puerto de entrada a un puerto de salida dentro de un mismo *bridge*. En caso de 5G, esta latencia se calcula como:

$$BD = t_{residence, UE-DSTT} + d_{dependent} + d_{independent} \quad (2)$$

siendo $t_{residence, UE-DSTT}$ el tiempo necesario para que un paquete se envíe desde el puerto de entrada del *bridge* 5G virtual a través del traductor DS-TT y se reciba en el UE y viceversa. Este tiempo se conoce como tiempo de residencia. El tiempo experimentado por un paquete dentro del *bridge* depende de los parámetros $d_{dependent}$ y $d_{independent}$. El tiempo $d_{dependent}$ depende de la longitud del paquete y el *bitrate* del enlace. El tiempo que es independiente a la longitud del paquete ($d_{independent}$), se refiere al tiempo necesario para almacenar y transmitir un paquete a través del *bridge*. El valor de $d_{independent}$ depende de factores como el esquema de *scheduling* en 5G, el tiempo de procesamiento, la estructura de la trama de 5G o las retransmisiones, entre otros.

3.1.2 Configuración del *bridge* 5G virtual

Cuando el *bridge* 5G virtual recibe la información del *scheduling* de TSN, la red 5G debe ser configurada para que pueda satisfacer dicho *scheduling*. Para admitir transmisiones de TSN en 5G, la red 5G mapea la información recibida desde el CNC a los parámetros 5G QoS y TSCAI. Estos parámetros indican los requisitos que los flujos TSN deben cumplir dentro del *bridge* 5G virtual. En base a estos datos, el gNB utilizando un *scheduling* de 5G, asigna recurso radio tanto a los UEs como al propio gNB para transmitir los paquetes de todos los flujos TSN y así lograr una integración coordinada y efectiva entre 5G y TSN.

Como ya se ha citado en el apartado 2.4, 3GPP en el anexo I del estándar 23.501 [5] especifica cómo se extrae información de algunos de estos requisitos. Para TSCAI, 3GPP indica cómo se calculan los parámetros *Burst Arrival Time*, o instante de llegada del

paquete al puerto de entrada y periodicidad para cada flujo TSN. Para 5QI, 3GPP especifica que el tráfico de TSN se define como *Delay Critical GBR*, y describe cómo se calculan los parámetros MDBV y GFBR.

3.2. Latencia en el *bridge* 5G virtual

En este estudio se ha considerado el escenario industrial definido por 5G-ACIA en [10]. En [10], una red 5G-TSN proporciona conectividad a una aplicación que controla una planta logística que trabaja en bucle cerrado. En este caso de uso, un UE de la red 5G está conectado localmente (por cable) a un sensor (S) y un actuador (A) que se comunican a través de la conexión inalámbrica 5G con un PLC remoto. El PLC está conectado a un *bridge* TSN mediante un cable. El sensor envía información al PLC cada 1 ms, y en el siguiente ciclo de control, el PLC envía un comando al actuador. La Figura 13 muestra la red 5G-TSN analizada. El escenario incluye dos ordenadores que intercambian información de la planta, pero las latencias que tienen que cumplir no son estrictas. En el resto del análisis nos enfocamos en el tráfico TSN de mayor prioridad.

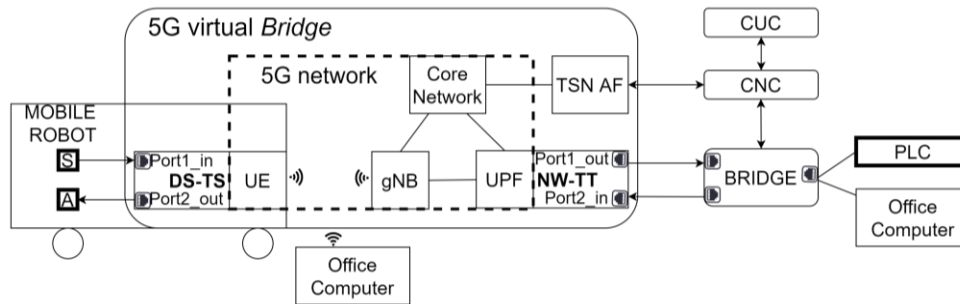


Figura 13. Ejemplo de una red 5G-TSN integrada.

Inicialización: Los usuarios finales (el sensor, el actuador y el PLC) informan sobre los requisitos del tráfico al CUC, que este, a su vez, reenvía esta información al CNC. Se ha considerado la transmisión de dos flujos TSN, F_1 y F_2 . El flujo TSN F_1 se transmite desde el sensor al PLC en UL, mientras que el flujo TSN F_2 se transmite desde el PLC al actuador en DL a través del *bridge* 5G virtual. Los paquetes de tamaño $s_1=64$ Bytes y $s_2=256$ Bytes se generan con una periodicidad de 1 ms y la comunicación entre dos usuarios finales se debe realizar dentro de 1 ms ($l_{e2e}=1\text{ms}$) [10]. Los paquetes s_1 se transmiten desde el sensor (S) al PLC y los paquetes s_2 se transmiten desde el PLC al actuador (A).

Topología de la red 5G-TSN y capacidades del *bridge*: El CNC debe descubrir la topología de la red y las capacidades que tiene cada *bridge* del escenario (TSN y 5G), para luego crear el *scheduling* para todos los flujos TSN. Para ello, cada *bridge* envía al CNC

los siguientes datos: ID del *bridge*, puertos Ethernet y sus direcciones MAC, información sobre los usuarios y *bridges* vecinos, latencia mínima/máxima que puede garantizar (*Bridge Delay* o *BD*), el tiempo de propagación y las clases de tráfico soportadas.

Para calcular la latencia mínima/máxima que los *bridges* (5G y TSN) pueden garantizar, se va a utilizar la fórmula (2) definida en el apartado 3.1.1. La Figura 14 representa el camino que ambos flujos TSN recorren. En este escenario, los paquetes del flujo TSN F_1 son los más pequeños, mientras que los paquetes del flujo TSN F_2 son los más grandes. Por lo tanto, la latencia mínima de los *bridges* (5G y TSN) se va a calcular utilizando el flujo TSN F_1 (BD_{5G}^1 y BD_{TSN}^1 en la Figura 14) y la latencia máxima se va a calcular utilizando el flujo TSN F_2 (BD_{5G}^2 y BD_{TSN}^2). En la Figura 14, l_S , l_{PLC} y l_A son los tiempos de la aplicación (l_{app}), mientras que l_{prop} es el tiempo de propagación entre diferentes *bridges* y entre el *bridge* y el dispositivo final.

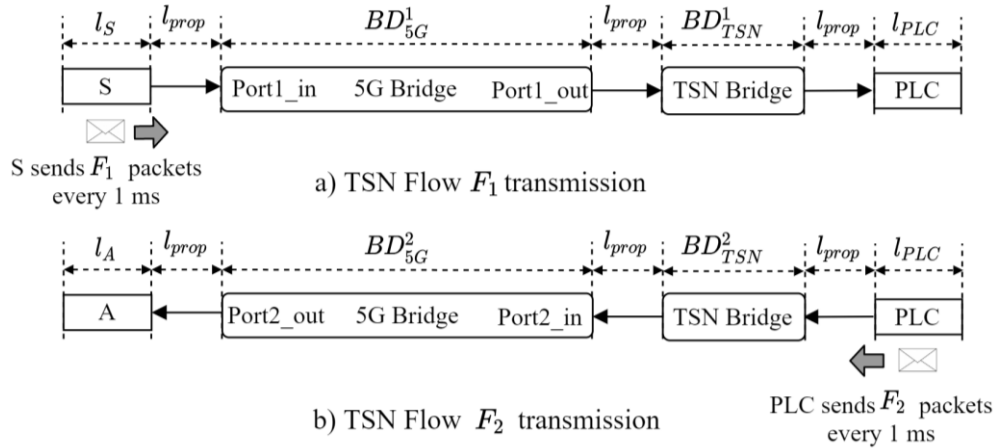


Figura 14. El camino extremo a extremo de los flujos TSN a) F_1 y b) F_2 .

Para calcular la latencia del *bridge* TSN, se asume un enlace con un *bitrate* igual a 100 Mbps, que es una velocidad de bits típica en la industria para latencias entre 25 μ s y 2 ms [64]. Los tiempos dependientes al tamaño del paquete mínimo y máximo son (3) y (4), respectivamente, y representan el tiempo de transmisión desde un puerto de entrada hasta un puerto de salida. Para ello, se tiene en cuenta el tamaño de paquete mínimo que corresponde al paquete que se transmite desde el sensor al PLC (s_1) y el máximo, que es el paquete que se transmite desde el PLC al actuador (s_2).

$$a_{dependent, TSNbridge}^{min} = \frac{64 \text{ bytes} \times 8 \text{ bits}}{100 \text{ Mbps}} = 5.12 \mu\text{s} \quad (3)$$

$$d_{dependent}^{max}{}_{TSNbridge} = \frac{256 \text{ bytes} \times 8 \text{ bits}}{100 \text{ Mbps}} = 20.5 \mu\text{s} \quad (4)$$

En cuanto al tiempo independiente al tamaño del paquete, $d_{independent}$, se asume que una vez el paquete llega al puerto de salida del *bridge* TSN se transmite de inmediato sin añadir ningún retraso; es decir, se considera que la puerta del *buffer* para esta clase de tráfico está abierta. En consecuencia, $d_{independent}$ incluye únicamente los tiempos de procesamiento, que es igual a $1.5\mu\text{s}$ según [65]. En base a (2) y considerando $t_{residence,UE-DSTT}$ igual a 0 en caso del *bridge* TSN, el tiempo mínimo y máximo que un paquete puede experimentar dentro del *bridge* TSN es:

$$BD_{TSN}^l = BD_{TSN}^{min} = 5.12 \mu\text{s} + 1.5 \mu\text{s} = 6.62 \mu\text{s} \quad (5)$$

$$BD_{TSN}^2 = BD_{TSN}^{max} = 20.5 \mu\text{s} + 1.5 \mu\text{s} = 22 \mu\text{s} \quad (6)$$

En el caso del *bridge* 5G virtual se ha considerado el uso de un 5QI preconfigurado en la RAN. Este 5QI está asociado a valores de QoS: PDB, MDBV y GFBR. Estos valores no están estandarizados y se pueden establecer de acuerdo con las características de la red. En esta sección, se estiman los valores de PDB, GFBR y MDBV que deben tener para satisfacer los requisitos de este ejemplo.

La red 5G debe garantizar la transmisión de los flujos F_1 y F_2 dentro del tiempo BD_{5G}^1 y BD_{5G}^2 para poder soportar los requisitos l_{e2e} de latencia extremo a extremo. El PDB es el tiempo máximo que un paquete puede experimentar entre el UE y la UPF que se conecta a la red de TSN. Por lo tanto, el PDB debe ser igual o superior que la latencia que garantiza el *bridge* 5G virtual. Para cumplir la latencia l_{e2e} en esta tesis hemos estimado el PDB como:

$$PDB \leq l_{e2e} - l_{app} - l_{prop} - BD_{TSN} \quad (7)$$

siendo l_{app} el tiempo de procesamiento a nivel de aplicación en los usuarios finales, l_{prop} el tiempo de propagación del paquete desde que sale de un *bridge* hasta que entra en el siguiente *bridge* y BD_{TSN} es la latencia con la que garantiza el *bridge* TSN transmitir un paquete desde su puerto de entrada hasta su puerto de salida.

En base a la ecuación (7), la latencia máxima o PDB que debe soportar la red 5G será igual a $PDB_1 = 743.38 \mu s$ y $PDB_2 = 728 \mu s$ para los flujos TSN F_1 y F_2 , respectivamente. Para ello, se ha asumido un tiempo de procesamiento en la aplicación l_{app} igual al 25% del total de la latencia *end to end* l_{e2e} , y se han tenido en cuenta los cálculos de latencia TSN realizados en las ecuaciones (5) y (6). El tiempo de propagación l_{prop} se ha considerado despreciable debido a las cortas distancias entre dispositivos.

Usando los valores estimados para PDB (PDB_1 y PDB_2) y el tamaño de paquetes (s_1 y s_2) se puede calcular tanto el $GFBR_i$ como $MDBV_i$ siendo $i \in [1,2]$. El *bitrate* mínimo necesario para garantizar las transmisiones de los flujos TSN dentro de la red 5G se calcula a través de la fórmula $GFBR_i > \frac{s_i \times 8 \text{ bits}}{PDB_i}$. El MDBV es la cantidad máxima de datos que se puede transmitir cuando la puerta del *buffer* de una clase determinada está abierta. Por lo tanto, considerando el uso de *Robust Header Compression* (RHC) y la cabecera L2 de 5G de 6 bytes [52], el MDBV mínimo que se puede transmitir por la red 5G tras quitar las cabeceras de TSN y poner las cabeceras de 5G es igual a $MDBV_1 \geq 54$ bytes y $MDBV_2 \geq 246$ bytes para el flujo TSN F_1 y F_2 , respectivamente.

Considerando la ecuación (2) se puede estimar el tiempo del *bridge* 5G virtual (ecuaciones (8) y (9)). Para calcular el tiempo dependiente del tamaño del paquete, $d_{dependent}$, se han tenido en cuenta los parámetros $GFBR_i$ y $MDBV_i$ calculados previamente, con $i \in [1,2]$. Mientras que el tiempo independiente $d_{independent}$ depende del *scheduling* en 5G, la estructura de la trama y las retransmisiones. En este caso, se ha considerado el uso de CG en UL y de SPS en DL para la transmisión del tráfico TSN, con el fin de reducir la latencia de la transmisión. Además, no se han considerado retransmisiones. Por lo tanto, el tiempo que el paquete debe esperar para transmitirse es igual a 0 ($d_{independent} = 0$).

$$BD_{5G}^1 < \frac{MDBV_1 \times 8 \text{ bits}}{GFBR_1} + d_{independent} \approx 627 \mu s \quad (8)$$

$$BD_{5G}^2 < \frac{MDBV_2 \times 8 \text{ bits}}{GFBR_2} + d_{independent} \approx 699.6 \mu s \quad (9)$$

Tras establecer una sesión PDU, TSN AF recibe el valor de tiempo de residencia ($t_{residence, UE-DSTT}$). Por lo tanto, se actualiza el *bridge delay*: $BD_{5G}^1 < (627 \mu s + t_{residence, UE-DSTT})$ y $BD_{5G}^2 < (699.6 \mu s + t_{residence, UE-DSTT})$. Finalmente, los tiempos BD_{5G}^1 y BD_{5G}^2 deben ser menores o iguales a los PDB estimados anteriormente, PDB_1 y PDB_2 , respectivamente.

Configuración de la red 5G-TSN: Cuando el CNC recibe la información de los *bridges*, calcula el *scheduling* de TSN, así como las rutas por las cuales se deben enviar los flujos TSN. En este caso se transmiten dos flujos uno en DL y otro en UL, por lo que estos dos flujos, dentro del mismo *bridge*, se transmiten a través de diferentes puertos de salida. En consecuencia, cada flujo debe cumplir con el GCL del puerto de salida que le corresponde (en la Figura 14, al flujo TSN F_1 le corresponde el puerto de salida Port1_out y al flujo TSN F_2 le corresponde el puerto de salida Port2_out). En base a los parámetros calculados, el flujo TSN F_1 llega al *bridge* 5G virtual en el instante (l_s) y sale antes de ($l_s + 627 \mu s + t_{residence,UE-DSTT}$). Mientras que el flujo TSN F_2 llega al *bridge* 5G virtual en el instante ($l_{PLC}+22 \mu s$) y sale antes de ($l_{PLC}+22 \mu s+699.6 \mu s + t_{residence,UE-DSTT}$).

Una vez realizada el *scheduling* de TSN, el CNC envía información detallada sobre el instante de llegada, máxima latencia que puede soportar y periodicidad de cada flujo TSN a todos los *bridges*, incluyendo al *bridge* 5G virtual. Utilizando esta información, el *bridge* 5G virtual puede determinar el patrón de tráfico y asignar el QoS adecuado. Como se presenta en [51], la asignación entre las prioridades de los flujos TSN y las prioridades 5QI es flexible. 5G establece valores de prioridad estandarizados de 5QI que pueden tomar un valor de hasta 90. Las prioridades 5QI estandarizados para los recursos *Delay Critical* GBR son menores que 24. Además, para las aplicaciones de automatización discreta la prioridad de 5QI es igual a 19. En este contexto, se ha considerado que 5G debe asignar una prioridad igual o inferior a 19 para soportar los flujos TSN. La Tabla 5 presenta los parámetros de QoS que se han estimado para cada flujo TSN. Finalmente, el perfil de QoS y la información TSCAI (periodicidad, dirección del flujo, instante de llegada y *survival time*) se envían al gNB para calcular la asignación de recursos radio que necesita cada flujo TSN dentro de 5G.

Tabla 5. Parámetros de 5G QoS para los flujos TSN F_1 y F_2 .

	Resource type	Priority	MDBV	GFBR	PDB
QoS UL	DelayCritical	<19	64 Bytes	>688.7 kbit/s	$\leq 743.4 \mu s$
QoS DL	DelayCritical	<19	256 Bytes	>2813.1 kbit/s	$\leq 728 \mu s$

3.3. Conclusiones

En este tercer capítulo se ha analizado la integración de 5G-TSN para soportar aplicaciones industriales deterministas y críticas en tiempo. Para ello, se ha cuantificado el tiempo máximo que un flujo TSN puede experimentar dentro del *bridge* 5G virtual para cumplir con los requisitos de la aplicación. Para este análisis se ha escogido una aplicación industrial basada en un ejemplo del documento de 5G-ACIA que representa un control industrial en lazo cerrado [10].

En base a este capítulo, en esta tesis se han podido identificar algunas limitaciones que requieren un estudio más detallado para integrar eficazmente 5G y TSN. A continuación, se listan las limitaciones identificadas:

- Scheduling de los flujos TSN en 5G: Se debe diseñar un *scheduler* en 5G que garantice la transmisión de los flujos TSN dentro de la red 5G cumpliendo con los tiempos establecidos por TSN. Teniendo en cuenta que el tráfico TSN que se ha considerado es periódico, el uso de un *scheduler* SPS en DL o un CG en UL parece ser la opción de *scheduling* más adecuada. El *scheduling* en 5G determinará si el *bridge* 5G virtual puede garantizar el *scheduling* que se ha realizado en TSN. Además, se debe analizar el caso en que la periodicidad del tráfico no coincide con las periodicidades que se pueden admitir en la RAN de 5G.
- Dimensionamiento de la RAN en 5G y definición de nuevos valores estandarizados de 5QI: Los QoS estandarizados en 5G establecen un PDB igual o superior a 5ms [5]. En base a los resultados obtenidos en este capítulo (Tabla 5), estos PDB no son adecuados para este tipo de aplicaciones. Las redes privadas de 5G, como ya se ha explicado en el apartado 2.3.4, permiten implementar un CN local y así reducir la latencia experimentada por los flujos TSN dentro del *bridge* 5G virtual. Por lo tanto, una red 5G privada con un dimensionamiento adecuado de la RAN permitirá cumplir con los valores de PDB inferiores a 1ms [17]. Además, aunque se puedan utilizar valores preconfigurados o dinámicos de 5QI, la definición de nuevos valores estandarizados de 5QI para este tipo de tráfico sería interesante.
- Estimación de la latencia del *bridge* 5G virtual: El estándar 3GPP define los conceptos básicos para estimar esta latencia; sin embargo, no se representa el mapeo exacto entre los parámetros 3GPP y la latencia del *bridge*. La estimación de esta latencia no es exacta ya que puede variar por diferentes aspectos, como, por ejemplo, el *scheduling* de los recursos radio en 5G y el tráfico que se va a transmitir en cada momento. Por esta razón, la transmisión de algunos datos iniciales para medir el PDB en 5G, antes de enviar la información de las capacidades a CNC, puede resultar necesario para poder calcular de una forma más exacta la latencia que los flujos TSN van a experimentar en el *bridge* 5G virtual.

El artículo de conferencia [11] recoge todo los puntos estudiados en este capítulo.

4. Diseño de esquemas de *scheduling Configured Grant* para redes 5G integradas con redes TSN

Un aspecto crítico para conseguir la integración completa de las redes 5G-TSN es la coordinación de los *schedulers* de ambas redes. El *scheduling* de cada flujo de tráfico TSN en ambas redes debe planificarse como si de una única red se tratara, de manera que se satisfagan los requisitos de latencia y determinismo de las aplicaciones industriales. Para ello, el *scheduling* de cada flujo de tráfico en la red TSN debe realizarse teniendo en cuenta las capacidades y estado actual de la red 5G, y viceversa. Tal y como se presentó en el apartado 2.5, en la literatura se está trabajando en dos enfoques diferentes para abordar este problema. Por un lado, se proponen trabajos que calculan a través de problemas de optimización el *scheduling* en la red TSN y en la red 5G de manera conjunta. Sin embargo, estas soluciones pueden conllevar un alto coste computacional. Por otro lado, se propone realizar el *scheduling* de la red TSN y de la red 5G de manera coordinada pero no conjunta, y así minimizar el coste computacional. Es decir, ambos esquemas de *scheduling* se calcularían de manera separada, pero considerando la información compartida entre ambas redes. En este caso, en primer lugar, se realiza el *scheduling* de la red TSN teniendo en cuenta las capacidades de todos los *bridges* (TSN y 5G virtual), y posteriormente, se realiza el *scheduling* del tráfico TSN en el *bridge* 5G virtual. En 5G, el *scheduling* más adecuado para dar soporte al tráfico periódico para transmisiones en UL es el *scheduling Configured Grant* o CG. Los esquemas de *scheduling* basados en CG comúnmente utilizados en la literatura asignan recursos radio de manera previa a cada flujo TSN F_i y esta asignación se repite periódicamente (por ejemplo, [12]). Sin embargo, esta asignación periódica puede resultar en conflictos de asignación de recursos entre distintos flujos TSN cuando la

periodicidad con la que se asignan los recursos es diferente para cada flujo. En la literatura hay algunas propuestas que abordan este problema ([61], [62] y [63]), en cambio, ninguna de estas propuestas soluciona el problema de manera integral (ver apartado 2.5). En este contexto, en este capítulo, se definen y se evalúan dos nuevos esquemas de *scheduling* en 5G, uno basado en políticas de optimización y el otro basado en un algoritmo heurístico, para gestionar el tráfico TSN en la red 5G. Ambos esquemas de *scheduling* propuestos buscan alcanzar la mejor solución posible que evite las colisiones entre diferentes paquetes debido a la asignación de un mismo recurso a flujos TSN con distintas periodicidades. Por lo tanto, estos esquemas de *scheduling* intentan maximizar el número de flujos TSN que se pueden soportar por la red, y a su vez, minimizar la latencia experimentada por dichos flujos.

En los siguientes apartados de este capítulo se presenta, en primer lugar, la definición de distintos parámetros que se van a utilizar a lo largo de este capítulo (el apartado 4.1). Los apartados 4.2 y 4.3 presentan los dos *schedulers* propuestos (basado en optimización y basado en algoritmos heurísticos, respectivamente) para redes 5G integradas en redes TSN. El apartado 4.4 define el escenario que se ha utilizado para realizar las evaluaciones de los *schedulers* propuestos. Las métricas analizadas y los resultados obtenidos están definidos en el apartado 4.5. Los resultados obtenidos demuestran que ambos esquemas propuestos mejoran significativamente la cantidad de flujos TSN que pueden ser atendidos de manera satisfactoria en la red 5G-TSN integrada en comparación con el esquema de *scheduling* comúnmente utilizado en la literatura. Finalmente, en el apartado 4.6 se resumen las conclusiones obtenidas a partir de los estudios presentados en este capítulo. También se indican otros resultados que no se han incluido en este capítulo y que se pueden encontrar en nuestras publicaciones de conferencia [13] y [14] que están adjuntadas en el Anexo A.

4.1. *Scheduling* en redes 5G integradas con redes TSN

En redes 5G-TSN integradas, es el CNC de la red TSN el que decide el *scheduling* de los distintos flujos TSN en cada *bridge* (TSN o 5G virtual) de la red, tal y como se presentó en el apartado 2.4. Es decir, el CNC establece los instantes de llegada y salida de cada flujo TSN a cada *bridge* (TSN o 5G virtual) de la red para cumplir con el requisito de determinismo y latencia *end-to-end* de cada flujo TSN. Para el *bridge* 5G virtual, la latencia máxima que debe experimentar cada flujo TSN en la red 5G (l_{5G_i}) se calcula como la diferencia entre el tiempo de salida y de llegada al *bridge* 5G virtual. Conociendo el tiempo de llegada al *bridge* 5G virtual, la red 5G debe realizar por tanto el *scheduling* de los paquetes de cada flujo TSN garantizando que sean recibidos en el puerto de salida del *bridge* 5G virtual, como muy tarde, en el instante de salida indicado por el CNC.

A través del *bridge* 5G virtual se transmiten N_F flujos TSN entre dispositivos finales conectados a la red TSN. Cada flujo TSN, identificado como F_i con $i \in [1, N_F]$, consiste en paquetes de datos de tamaño s_i que llegan al puerto de entrada de 5G con una periodicidad p_i . El tamaño de los paquetes s_i viene determinado por el MDBV definido en el 5QI y la periodicidad p_i viene indicada en la señal TSCAI. Cabe destacar que s_i y p_i pueden ser diferentes para cada flujo TSN. Cada flujo TSN demanda una cantidad específica de recursos radio d_i , que se calcula en función del tamaño de los paquetes s_i y del MCS que se utilizará para la transmisión de los paquetes. Los paquetes a transmitir en un flujo TSN F_i se representan como $pkt_{i,j}$ donde i indica el flujo TSN F_i y j indica el número del paquete dentro del flujo TSN F_i . Los instantes de llegada y de salida de un paquete $pkt_{i,j}$ vienen dados por $A_{i,j} = A_i + j \cdot p_i$ y $D_{i,j} = A_{i,j} + l_{5G_i}$, respectivamente, siendo A_i el instante de llegada del primer paquete del flujo F_i que viene indicado en el TSCAI, y l_{5G_i} la latencia máxima en la red 5G requerida por el flujo TSN F_i . La latencia máxima en la red 5G para cada flujo TSN viene indicado en el parámetro PDB del 5QI (apartado 2.3). Esta tesis se centra en el estudio del tráfico en sentido ascendente y, por tanto, se considera que todo el tráfico TSN que llega a la red 5G se transmite en UL; la dirección de cada flujo TSN está representada por f_i . Se ha considerado que cada flujo TSN F_i se transmite por un UE diferente al gNB. Por lo tanto, un flujo TSN F_i se define como $F_i = \{A_i, l_{5G_i}, d_i, p_i, f_i\}$. La Tabla 6 recoge los distintos parámetros utilizados en este capítulo⁶.

Se ha considerado una red 5G privada o *Non Public Network* (NPN) que proporciona conectividad a una celda industrial. La red tiene asignado un ancho de banda BW que se divide en R_{BW} RBs (ver la Figura 15 donde se representa el *resource grid*). La red 5G opera en modo TDD⁷. La red 5G utiliza una numerología μ . En cada *slot* de 14 símbolos OFDM se reservan sym_{DLCTRL} y sym_{ULCTRL} símbolos al principio y al final de cada *slot* para transmitir mensajes de control en DL y UL, respectivamente. El resto de los símbolos se reservan para la transmisión de datos en UL⁸. La duración en tiempo de un símbolo OFDM se define como t_{sym} . Para cada paquete a transmitir por el canal radio, el esquema de *scheduling* de la red 5G debe decidir el número d_i^S de símbolos OFDM y d_i^R de RBs que se le debe asignar en función del tamaño del paquete. Además, el esquema de *scheduling* también debe decidir la posición de los recursos asignados en el *resource grid* tanto en frecuencia como en tiempo para cumplir con sus requisitos de latencia. El primer símbolo

⁶ La Tabla 6 contiene otros parámetros que no se han descrito en este apartado pero que van a ir apareciendo a lo largo de todo el capítulo.

⁷ Se ha escogido trabajar en modo TDD ya que en Alemania se asignó la banda 3700-3800 MHz para la industria [87], y esta banda opera en modo TDD [68].

⁸ Esta es una situación ideal en la que todos los recursos radio se reservan para UL, aunque en escenarios reales parte de estos recursos deben reservarse para transmitir datos en DL.

asignado para la transmisión de un paquete $pkt_{i,j}$ vendrá dado por $s_{i,j}$ y puede ser igual a cualquier número natural mayor o igual que 0.

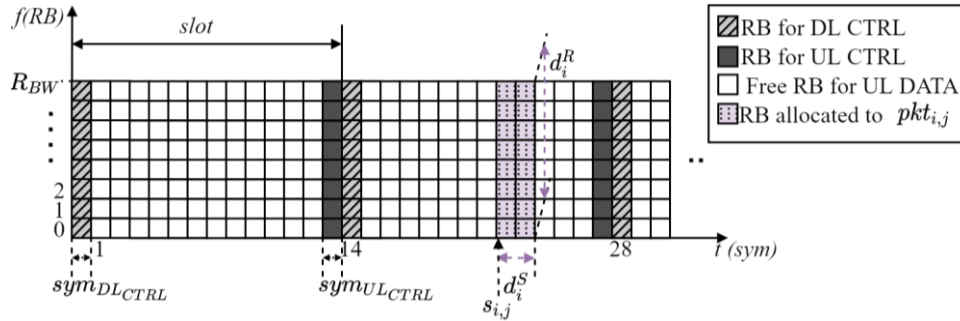


Figura 15. Ejemplo del *resource grid* y los parámetros de los *schedulers*.

Tabla 6. Parámetros.

Parámetro	Definición
F_i	Flujo TSN
$pkt_{i,j}$	Paquete de datos j del flujo TSN F_i
N_F	Número total de flujos TSN transmitidos
l_{5G_i}	Latencia máxima que debe experimentar un paquete del flujo TSN F_i
$A_{i,j}$ y $D_{i,j}$	Instante de llegada y de salida de un paquete $pkt_{i,j}$ al <i>bridge 5G</i> virtual
p_i	Periodicidad con la que se generan los paquetes del flujo TSN F_i
f_i	Sentido del enlace en el que se transmite el flujo TSN F_i (UL o DL)
s_i	Tamaño de los paquetes de datos del flujo TSN F_i
d_i	Número total de recursos radio necesario para la transmisión de un paquete de datos del flujo TSN F_i
d_i^S y d_i^R	Número de símbolos y RBs necesarios para la transmisión de un paquete de datos del flujo TSN F_i
R_{BW}	Número de RBs disponibles en un símbolo OFDM
$slot$	Número de símbolos en un <i>slot</i>
sym_{DLCTRL} y sym_{ULCTRL}	Número de símbolos reservados para la transmisión de mensajes de control en DL y UL en cada <i>slot</i>
$t_{UE,tx}$ y $t_{gNB,rx}$	Tiempos de procesamiento de los paquetes en el UE y en el gNB
t_{sym}	Duración en tiempo de un símbolo OFDM
HP	Hiperperiodo ($HP = \text{mcm}(p_i)$ con $i = [1 \dots N_F]$)
N_{pkt}^i	Número de paquetes que contiene el flujo TSN F_i dentro de HP
B_z	Bloques agrupando los paquetes de los flujos TSN incluidos en el HP cuya transmisión podría solaparse en tiempo y frecuencia
BF_z	Conjunto de paquetes $pkt_{i,j}$ que mantienen la asignación de recursos recibida
S	Número máximo de símbolos dentro del bloque B_z
N_{pkt}	Número de paquetes dentro de un bloque B_z
$N_{pkt}^{i,z}$	Vector de los índices j que identifican a cada paquete $pkt_{i,j}$ del flujo TSN F_i que están dentro del bloque B_z
s_{min}	El símbolo del HP en el que el primer paquete del bloque B_z llega al <i>bridge 5G</i> virtual (puede tomar valores enteros mayor o igual que 0)
$s_{i,j}^{first}$	Primer símbolo del bloque B_z asignado al paquete $pkt_{i,j}$ (puede tomar valores enteros mayor o igual que 1)
$s_{i,j}$	Primer símbolo del HP asignado al paquete $pkt_{i,j}$ ($s_{i,j} = s_{min} + s_{i,j}^{first} - 1$)

4.2. *Optimum Flexible configured grAnt Scheduling for TSN traffic* (O-FAST)

El primer esquema de *scheduling* propuesto está basado en técnicas de optimización lineal y tiene como objetivo principal cumplir con el requisito de latencia máxima que debe experimentar cada flujo TSN en la red 5G (l_{5G_i}). Este esquema de *scheduling* explota la información proporcionada por el CNC de la red TSN sobre los instantes de llegada (A_i), periodicidad (p_i) y latencia máxima (l_{5G_i}) para determinar una asignación de recursos radio flexible que cumpla con los requisitos de latencia de cada flujo TSN. El esquema de *scheduling* está basado en el mecanismo *Configured Grant* o CG y, por tanto, asigna recursos radio periódicamente a cada flujo TSN. Para evitar posibles situaciones de conflicto en las que más de un paquete de diferentes flujos TSN reciban los mismos recursos radio, el esquema de *scheduling* propuesto crea varias configuraciones CG para cada flujo TSN. Es importante destacar que los recursos radio definidos por cada configuración CG se pueden reasignar periódicamente siguiendo la fórmula definida en el apartado 5.8.2 del estándar 3GPP TS 38.321 [16]. El esquema propuesto busca satisfacer la demanda de recursos radio y los requisitos de latencia de todos los flujos TSN. Una vez alcanzado este objetivo, el esquema de *scheduling* propuesto intentará minimizar la latencia experimentada por los distintos flujos TSN. La nueva propuesta de *scheduling* plantea un problema de optimización lineal entera binaria y aplica técnicas de optimización para obtener la solución óptima al problema. Al nuevo esquema de *scheduling* propuesto se le ha denominado *Optimum Flexible configured grAnt Scheduling for TSN traffic* (O-FAST).

Este esquema de *scheduling* considera que se transmite tráfico TSN que genera paquetes periódicos. Aunque la periodicidad p_i de cada flujo TSN puede ser diferente, es posible encontrar un patrón de llegada de paquetes a la red 5G que se repite periódicamente considerando todos los flujos TSN. La duración en tiempo de este patrón de llegada de paquetes se denomina hiperperiodo o *HP* y, por tanto, el patrón se repite con una periodicidad igual a *HP*. Un *HP* puede contener múltiples paquetes de cada flujo TSN, es decir, el flujo F_i tendrá $j \in [1, N_{pkt}^i]$ paquetes dentro del *HP* siendo $N_{pkt}^i = HP/p_i$.

La Figura 16 ilustra el concepto del hiperperiodo o *HP* utilizando un ejemplo con tres flujos TSN que tienen diferentes periodicidades (30, 45 y 90 ms). En este ejemplo es posible observar un patrón de llegada de paquetes que se repite cada 90 ms cuando se considera el conjunto de los tres flujos TSN. O-FAST utiliza el concepto del *HP* para abordar el problema de *scheduling*. Conociendo los instantes de llegada de los paquetes de cada flujo TSN, O-FAST identifica el *HP* y realiza la asignación de recursos radio para la transmisión de los paquetes que hay dentro del *HP*. La solución de *scheduling* obtenida o

lo que es lo mismo, los recursos radio pre-asignados para la transmisión de cada paquete para cada flujo TSN, se repetirá con una periodicidad igual a HP .

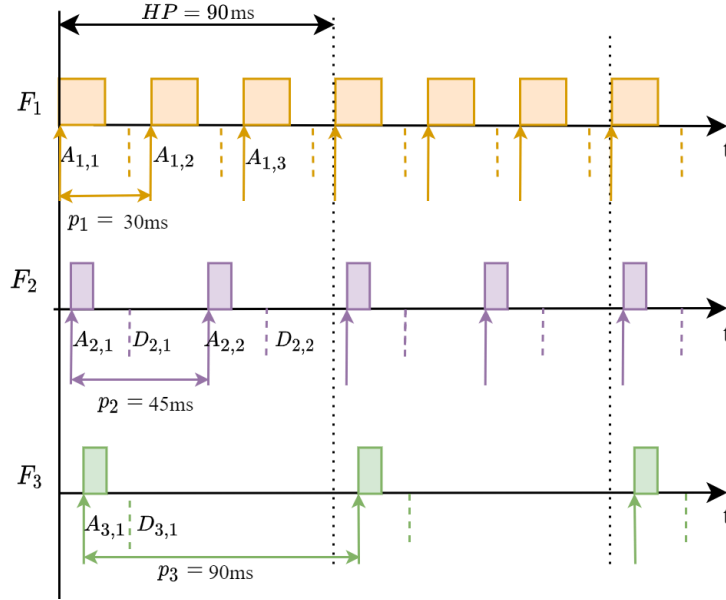


Figura 16. Cálculo del HP para un ejemplo donde 3 flujos TSN transmiten paquetes con diferentes periodicidades.

El funcionamiento de O-FAST se presenta en la Figura 17. En primer lugar, O-FAST calcula el HP como el mínimo común múltiplo de las periodicidades de los paquetes transmitidos en cada flujo TSN, es decir, $HP = \text{mcm}(p_i)$, con $i \in [1, N_F]$. A continuación, O-FAST crea bloques de paquetes agrupando los paquetes de los flujos TSN incluidos en el HP cuya transmisión podría solaparse en tiempo. Las transmisiones de dos paquetes $pkt_{i,j}$ y $pkt_{m,n}$ pueden solapar en tiempo si $A_{i,j} \leq A_{m,n} \leq D_{i,j}$ o $A_{i,j} \leq D_{m,n} \leq D_{i,j}$ ($A_{i,j}$, $A_{m,n}$, $D_{m,n}$ y $D_{i,j}$ son los tiempos de llegada y salida de los paquetes $pkt_{i,j}$ y $pkt_{m,n}$, respectivamente). El proceso de creación de estos bloques se presenta en el algoritmo I.

En el ejemplo de la Figura 18 se crearían dos bloques B_z : el primer bloque B_1 contiene los paquetes $pkt_{1,1}$, $pkt_{2,1}$ y $pkt_{3,1}$ mientras que el segundo bloque B_2 contiene los paquetes $pkt_{1,2}$, $pkt_{1,3}$ y $pkt_{2,2}$. El vector $N_{pkt}^{i,z}$ indica los índices de los paquetes del flujo TSN F_i que hay en un bloque B_z . En el caso de la Figura 18, para el bloque B_1 los índice j de cada flujo son $N_{pkt}^{1,1} = N_{pkt}^{2,1} = N_{pkt}^{3,1} = \{1\}$ mientras que para el bloque B_2 , $N_{pkt}^{1,2} = \{2,3\}$, $N_{pkt}^{2,2} = \{2\}$ y $N_{pkt}^{3,2} = \{\emptyset\}$. El parámetro s_{min} representa el primer símbolo del bloque B_z dentro del HP , y se calcula considerando el instante en el que llega el primer paquete del bloque B_z al *bridge* 5G virtual. En la Figura 18, para B_2 , $s_{min} = \left\lceil \frac{A_{1,2}}{t_{sym}} \right\rceil$. A su vez, cada bloque B_z contiene s símbolos OFDM

desde 1 hasta S . $s = 1$ del bloque B_z corresponde al símbolo s_{min} del HP , y $s = S$ es el número máximo de símbolos que hay dentro del bloque B_z (por ejemplo, en la Figura 18, para B_2 , $S = \frac{\lfloor D_{1,3} \rfloor - \lfloor A_{1,2} \rfloor}{t_{sym}}$). El primer símbolo que se asigna para el paquete $pkt_{i,j}$ dentro del HP es $s_{i,j} = s_{min} + s_{i,j}^{first} - 1$, siendo $s_{i,j}^{first}$ el primer símbolo en $s \in [1, S]$ del bloque B_z que se asigna al paquete $pkt_{i,j}$.

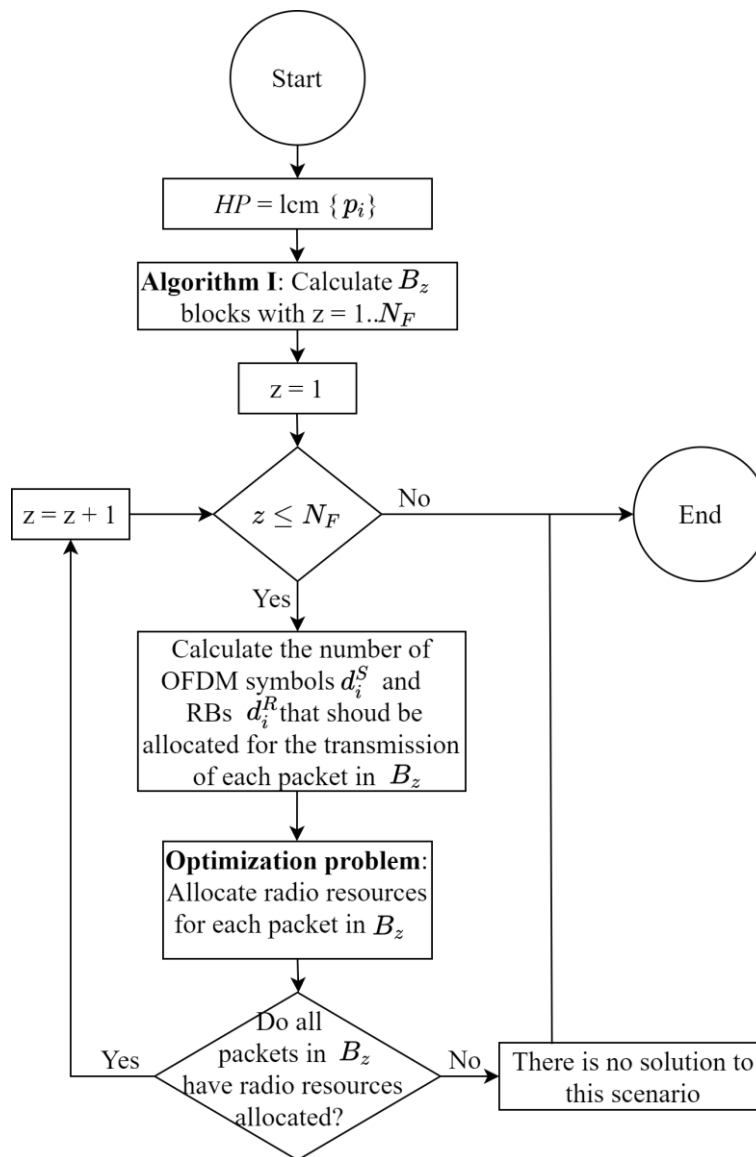


Figura 17. Proceso de asignación de recursos radio utilizando el esquema de *scheduling O-FAST*.

ALGORITHM I: *BLOCK CREATION*

1. Input: $F_i \forall i \in [1, N_F]$
2. $I = \{A_{i,j}\} \forall i \in [1, N_F], \forall j \in [1, HP/p_i]$
3. $z=1, B_z = \emptyset$
4. **While 1**
5. $\{i, j\} = \arg \left\{ \min_{m,n} \{I\} \right\} \forall m \in [1, N_F], \forall n \in [1, HP/p_m]$
6. $I = I - \{A_{i,j}\}$
7. **If** $B_z = \emptyset$
8. $D_{aux} = D_{i,j}, A_{aux} = A_{i,j}$
9. $B_z = \{pkt_{i,j}\}$
10. **Else**
11. **If** $A_{i,j} \leq D_{aux}$
12. $B_z = B_z \cup \{pkt_{i,j}\}$
13. **If** $D_{i,j} > D_{aux}$
14. $D_{aux} = D_{i,j}$
15. **End If**
16. **Else**
17. $z=z+1, B_z = \emptyset$
18. Go to 7
19. **End If**
20. **End If**
21. **If** $I = \emptyset$
22. Go to 24
23. **End If**
24. **End While**

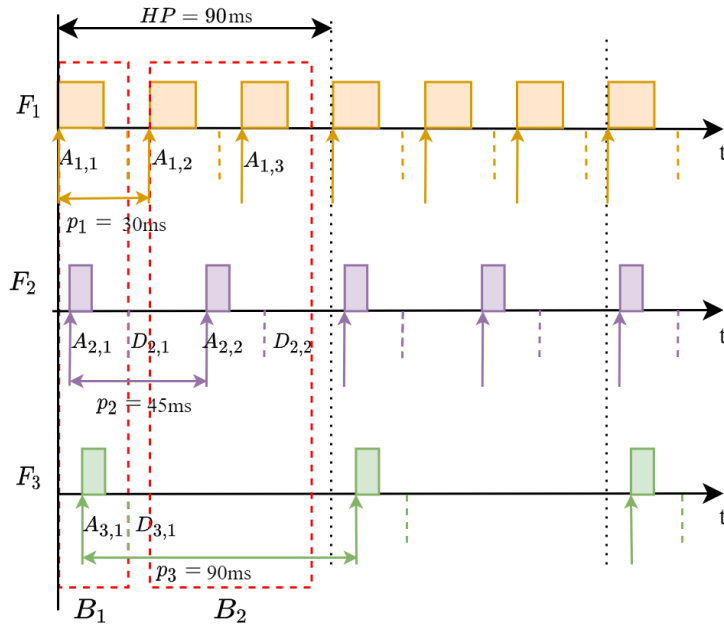


Figura 18. Bloques creados dentro del HP para un ejemplo donde 3 flujos TSN transmiten paquetes con diferentes periodicidades.

Para cada bloque B_z , O-FAST comienza calculando el número d_i^S de símbolos OFDM y el número d_i^R de recursos radio que deben asignarse a cada paquete del flujo F_i para satisfacer su demanda de recursos radio d_i (Figura 17). Para ello, O-FAST minimiza el número de símbolos asignados para la transmisión de cada paquete. De este modo, O-FAST asigna $d_i^R=d_i$ RBs en un símbolo OFDM ($d_i^S=1$) cuando d_i es inferior al número de RBs disponibles en el ancho de banda (R_{BW}), es decir, $d_i \leq R_{BW}$. Si $d_i > R_{BW}$, O-FAST asignará $d_i^R = R_{BW}$ RBs en $d_i^S = \lceil d_i / R_{BW} \rceil$ símbolos OFDM.

Tras calcular la demanda de símbolos OFDM d_i^S y de RBs d_i^R para cada flujo F_i , O-FAST comienza con el proceso de *scheduling* para los paquetes de datos dentro del bloque B_z . O-FAST realiza la asignación de recursos radio a los paquetes de cada bloque B_z por separado, siendo $z \in [1, 2, \dots]$. Para decidir los recursos radio que se van a asignar para llevar a cabo la transmisión de cada paquete incluido en B_z , O-FAST plantea un problema de optimización que se presenta a continuación en los apartados 4.2.1 y 4.2.2. El esquema de *scheduling* O-FAST finaliza si todos los paquetes en el *HP* tienen recursos asignados, o si no es posible encontrar recursos radio para todos los paquetes del *HP*.

4.2.1. Función Objetivo

O-FAST busca minimizar la latencia experimentada por todos los paquetes de los flujos TSN incluidos en B_z . La latencia experimentada por un paquete $pkt_{i,j}$ puede expresarse como:

$$s_{i,j} \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j} \quad (10)$$

donde $s_{i,j} \cdot t_{sym}$ es el instante en el que comienza la transmisión del paquete según los símbolos OFDM asignados, $d_i^S \cdot t_{sym}$ es el tiempo de transmisión del paquete por el canal radio, $t_{gNB,rx}$ es el tiempo de procesamiento del mensaje en el receptor y $A_{i,j}$ es el instante en que el paquete $pkt_{i,j}$ llega al *bridge* 5G virtual. Cuando no se conocen los símbolos que se han asignados al paquete $pkt_{i,j}$, la latencia experimentada podría expresarse como:

$$\sum_{s \in [1, S]} ((s + s_{min} - 1) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j}) \cdot X_s^{i,j} \quad (11)$$

donde se define la variable incógnita $X_s^{i,j} \in \{0, 1\}$, para $s \in [1, S]$, que toma el valor 1 cuando $(s_{min} + s - 1) \cdot t_{sym}$ representa el instante en el que comienza la transmisión del paquete y valdrá

0 de lo contrario. Utilizando (11), O-FAST define la función objetivo mostrada en (12) en la que busca minimizar la suma de la latencia experimentada por todos los paquetes $pkt_{i,j}$ incluidos en B_z , con $i \in [1, N_F]$ y $j \in N_{pkt}^{i,z}$.

$$\min_{X_s^{i,j}} \left(\sum_1^{N_F} \sum_{j \in N_{pkt}^{i,z}} \sum_{s \in [1,S]} ((s+s_{min}-1) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j}) \cdot X_s^{i,j} \right) \quad (12)$$

4.1.1. Restricciones del problema de *scheduling*

En este apartado se presentan las condiciones que debe cumplir la solución de *scheduling* obtenida. Estas condiciones deben expresarse en forma de ecuaciones lineales y restringirán el valor que pueden tomar las variables incógnitas $X_s^{i,j}$.

En primer lugar, O-FAST busca una solución al problema de *scheduling* que cumpla con los requisitos de latencia de cada flujo TSN. Esto implica que cada paquete de datos $pkt_{i,j}$ debe ser entregado y estar disponible en el puerto de salida del *bridge* 5G virtual para su transmisión al siguiente *bridge* antes del instante de salida establecido $D_{i,j}$. El instante en el que se completa la transmisión de un paquete $pkt_{i,j}$ por la red 5G se define en (13).

$$\sum_{s \in [1,S]} ((s+s_{min}-1) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx}) \cdot X_s^{i,j} \quad (13)$$

Para cumplir con los requisitos de latencia de cada flujo TSN, O-FAST establece que el tiempo dado en (13) debe ser menor que $D_{i,j}$, tal y como se indica en (14):

$$\sum_{s=1}^S ((s+s_{min}-1) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx}) \cdot X_s^{i,j} \leq D_{i,j} \quad \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \quad (14)$$

Además, O-FAST tiene en cuenta el tiempo de procesamiento de los paquetes en el transmisor $t_{UE,tx}$, que en este caso es el UE. Este tiempo de procesamiento es necesario para generar y codificar el paquete que será transmitido por el canal radio. Los recursos radio deben asignarse en símbolos OFDM posteriores a $A_{i,j} + t_{UE,tx}$, es decir, posteriores al instante de llegada del paquete más el tiempo de procesamiento en el transmisor. Por tanto, el primer símbolo OFDM s del bloque B_z que podría asignarse al paquete $pkt_{i,j}$ vendría dado por $\left\lceil \frac{A_{i,j} + t_{UE,tx}}{t_{sym}} \right\rceil + 1$. Así pues, las soluciones que asignen símbolos con un valor $s_{min} + s - 1$ anterior a

$\left\lceil \frac{A_{i,j} + t_{UE,tx}}{t_{sym}} \right\rceil + 1$, no son soluciones factibles, y la variable $X_s^{i,j}$ debe tomar un valor nulo. Por tanto, O-FAST establece la siguiente restricción:

$$X_s^{i,j} = 0, \forall s < \left\lceil \frac{A_{i,j} + t_{UE,tx}}{t_{sym}} \right\rceil - (s_{min} - 1) \quad \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \quad (15)$$

O-FAST establece que todos los recursos radio asignados a un paquete $pkt_{i,j}$ deben pertenecer a un mismo *slot*, tal y como se indica en 3GPP en el apartado 5.1.2.1 del estándar TS 38.214 [33]. Para ello, O-FAST define la restricción dada en (16), donde $sym_{ULCTRL} = 1$ representa el número de símbolos OFDM del *slot* que se reservan para transmisiones de mensajes de control en UL. En (16), $((s_{min} + s - 1) - \left\lfloor \frac{(s_{min} + s - 1)}{14} \right\rfloor \cdot 14) \cdot X_s^{i,j}$ representa el número del primer símbolo OFDM dentro del *slot* asignado para la transmisión del paquete $pkt_{i,j}$. Por tanto, (16) establece que el número de este primer símbolo debe ser menor o igual que el número máximo de símbolos dedicados para la transmisión de datos en UL ($14 - sym_{ULCTRL}$) menos los símbolos que se requieren d_i^S dentro del mismo *slot* para transmitir el paquete.

$$\sum_{s=1}^S ((s + s_{min} - 1) - \left\lfloor \frac{s_{min} + s - 1}{14} \right\rfloor \cdot 14) \cdot X_s^{i,j} \leq 14 - sym_{ULCTRL} - d_i^S \quad \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \quad (16)$$

Como se presentó en el apartado 4.1, se ha establecido que los primeros sym_{DLCTRL} símbolos de cada *slot* se utilizarán para transmitir señales de control en sentido DL. Por lo tanto, los primeros sym_{DLCTRL} símbolos de cada *slot* no pueden asignarse para la transmisión de ningún paquete y, por tanto, $X_s^{i,j} = 0$ para todo s que cumpla la restricción definida en (17).

$$X_s^{i,j} = 0, \forall (\text{mod}(s_{min} + s - 1, 14) \leq (sym_{DLCTRL} - 1)) \quad \forall s \in [1, S], \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \quad (17)$$

Cada paquete $pkt_{i,j}$ debe tener recursos radio asignados para poder transmitirse por el canal. A su vez, cada paquete va a utilizar una única configuración CG, por lo tanto solamente una variable $X_s^{i,j}$ con $s = [1, S]$ va a ser igual a 1. O-FAST define la siguiente restricción:

$$\sum_{s=1}^S X_s^{i,j} = 1 \quad \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \quad (18)$$

O-FAST minimiza el número de símbolos asignados para la transmisión de cada paquete. De este modo, O-FAST asigna $d_i^R = d_i$ RBs en un símbolo OFDM ($d_i^S = 1$) cuando d_i es inferior al número de RBs disponibles en el ancho de banda (R_{BW}), es decir, $d_i \leq R_{BW}$. Debido a que un mismo RB no puede ser compartido entre diferentes paquetes, dos o más paquetes pueden transmitirse en el mismo instante de tiempo solo si la suma de los RBs demandados d_i^R por cada paquete es menor o igual que R_{BW} . Para ello, se ha definido la restricción (19) que determina que el número total de RBs que se utilizan para todos los paquetes del bloque B_z en cada símbolo $s = [1, S]$ y la suma de estos RBs en un mismo símbolo OFDM no puede ser mayor que R_{BW} RBs:

$$\sum_{i \in [1, N_F]} \sum_{j \in N_{pkt}^{i,z}} (X_s^{i,j} \cdot d_i^R) \leq R_{BW} \quad \forall s \in [1, S] \text{ con } d_i \leq R_{BW} \quad (19)$$

Si $d_i > R_{BW}$, O-FAST asignará $d_i^R = R_{BW}$ RBs en $d_i^S = \lceil d_i / R_{BW} \rceil$ símbolos OFDM. Debido a como se crean las configuraciones CG [16], O-FAST establece que cada paquete se transmite en símbolos OFDM consecutivos. O-FAST define la siguiente restricción:

$$X_s^{i,j} \cdot R_{BW} + \sum_{s'=s+1}^{s+d_i^S-1} \sum_{\substack{m \in [1, N_F] \\ m \neq i}} \sum_{n \in N_{pkt}^{i,z}} X_{s'}^{m,n} \leq R_{BW} \quad \forall s \in [1, S], \forall i \in [1, N_F], \forall j \in N_{pkt}^{i,z} \text{ con } d_i > R_{BW} \quad (20)$$

4.3. Heuristic Flexible configured grAnt Scheduling for TSN traffic (H-FAST)

El esquema O-FAST propuesto y presentado en el apartado anterior es una propuesta de *scheduling* basado en técnicas de optimización lineal. La principal ventaja de los algoritmos basados en técnicas de optimización es que pueden alcanzar soluciones que cumplan la demanda de recursos y requisitos de todos los flujos TSN, mientras que un *scheduling* basado en un procedimiento heurístico podría no ser capaz de encontrar una solución que satisfaga la demanda de recursos para todos los paquetes de todos los flujos TSN. Sin embargo, las soluciones basadas en técnicas de optimización suelen presentar un coste computacional muy alto. Por este motivo, en esta tesis se define una nueva propuesta de *scheduling* heurístico. El esquema propuesto tiene el mismo objetivo que O-FAST, es decir, busca satisfacer la demanda de recursos radio y los requisitos de latencia de todos los flujos TSN evitando posibles situaciones de conflicto en las que más de un paquete de diferentes flujos TSN reciban los mismos recursos radio e intenta minimizar la latencia

experimentada por los distintos flujos TSN. Al nuevo esquema de *scheduling* propuesto se le ha denominado *Heuristic Flexible configured grAnt Scheduling for TSN traffic* (H-FAST).

El funcionamiento de H-FAST se presenta en el algoritmo II. H-FAST comienza calculando el *HP* y creando los bloques B_z , con $z \in [1, 2, \dots]$, al igual que O-FAST (algoritmo I). A continuación, H-FAST resuelve el problema de *scheduling* para cada bloque B_z de manera independiente. Primero, H-FAST calcula la demanda de símbolos OFDM d_i^S y de RBs d_i^R para cada flujo F_i (línea 6 del algoritmo II). Tras realizar este cálculo, H-FAST comienza con el proceso de *scheduling* para los paquetes de datos dentro del bloque B_z .

H-FAST sigue un proceso iterativo. En cada iteración el esquema de *scheduling* propuesto intenta encontrar una solución que satisfaga el requisito de latencia (l_{5G_i}) para todos los paquetes. Para ello, el esquema de *scheduling* propuesto comienza a asignar recursos radio en función del requisito de latencia l_{5G_i} ; es decir, asigna recursos radio primero al paquete de datos en B_z que deba cumplir el requisito de latencia l_{5G_i} más restrictivo y termina asignando recursos radio al paquete de datos en B_z que deba cumplir el requisito de latencia menos restrictivo. H-FAST asigna a cada paquete de datos d_i^R RBs en los primeros d_i^S símbolos OFDM consecutivos en los que haya recursos radio disponibles tras el tiempo de procesamiento requerido para generar el paquete radio en el transmisor ($A_{i,j} + t_{UE,tx}$). Además, el paquete debe de recibirse en el traductor NW-TT, tras ser procesado en el gNB, antes del tiempo de salida establecido ($D_{i,j} - t_{gNB,rx}$). La asignación de recursos radio se realiza en líneas 12-26 del algoritmo II. El primer símbolo asignado al paquete $pkt_{i,j}$ se guardará en la variable $s_{i,j}$ (línea 16 del algoritmo II).

La Figura 19 muestra un ejemplo sencillo de asignación de recursos radio para los paquetes $pkt_{1,1}$, $pkt_{2,1}$ y $pkt_{3,1}$. Cada paquete es de un flujo F_i diferente con $i \in [1, 2, 3]$ y se representa con un color diferente (naranja, morado y verde, respectivamente). Las flechas indican el instante en el que cada paquete llega a *bridge* 5G virtual, mientras que las líneas discontinuas indican el instante de tiempo en el que cada paquete debe transmitirse desde el *bridge* 5G virtual al siguiente *bridge* o dispositivo final. La demanda de recursos radio es diferente para cada paquete, por ejemplo, el paquete $pkt_{1,1}$ demanda $d_1^R = 4$ RBs en $d_1^S = 3$ símbolos OFDM consecutivos, mientras que el paquete $pkt_{3,1}$ demanda $d_3^R = 3$ RBs en $d_3^S = 1$ símbolo OFDM. En este ejemplo, el requisito de latencia que deben cumplir los paquetes desde el más restrictivo hasta el menos restrictivo son: $l_{5G_3} < l_{5G_2} < l_{5G_1}$. Por lo tanto, el esquema de *scheduling* H-FAST primero va a asignar recursos al paquete $pkt_{3,1}$, luego al paquete $pkt_{2,1}$ y finalmente al paquete $pkt_{1,1}$. El primer símbolo asignado a cada paquete se

indica con el parámetro $s_{3,l}$, $s_{2,l}$ y $s_{1,l}$, respectivamente. Por esta razón, $s_{1,l} > s_{2,l}$ aunque el instante de llegada del paquete $A_{1,l}$ sea menor que $A_{2,l}$.

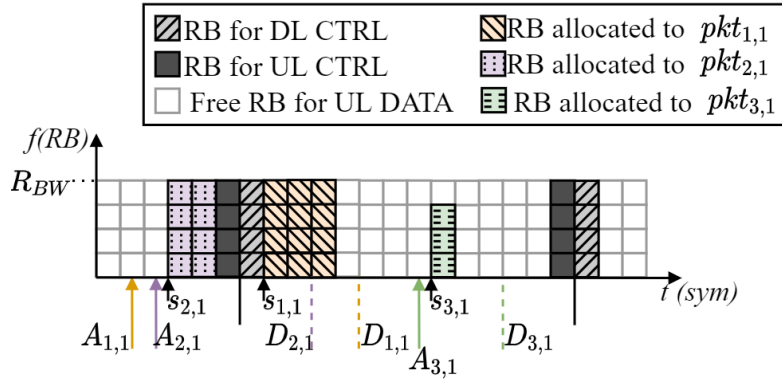


Figura 19. Asignación de recursos radio para los paquetes $pkt_{1,l}$, $pkt_{2,l}$ y $pkt_{3,l}$.

Si un paquete recibe recursos radio justo antes del tiempo límite $D_{i,j} - t_{gNB,rx}$, este paquete se incluye en un subconjunto denominado BF_z (líneas 17-18 del algoritmo II), cuyo propósito se explica más adelante. En el caso de que no sea posible encontrar recursos radio que satisfagan los requisitos de latencia para un paquete $pkt_{i,j}$ (línea 27), se reinicia el proceso de asignación de recursos (líneas 32-36) en busca de una solución diferente que pueda satisfacer la demanda de recursos para todos los paquetes incluidos en B_z . Para ello, se reestablece el conjunto inicial de paquetes incluidos en el bloque B_z (línea 32). Del conjunto inicial se eliminan los paquetes incluidos en BF_z (línea 33) que mantendrán la asignación de recursos recibida hasta el momento. Para el resto de los paquetes que quedan incluidos en B_z , se elimina la asignación de recursos recibida (línea 34) y deben volver a recibir recursos radio que satisfagan su demanda y sus requisitos de latencia l_{5G_i} . Para ello, H-FAST comienza una nueva iteración para buscar una solución para todos los paquetes de B_z . En esta nueva iteración, el primer paquete en recibir recursos será el paquete que no pudo recibir recursos radio en la iteración anterior, el cual también pasa al conjunto BF_z . Posteriormente, los paquetes en B_z se atenderán en función del requisito de latencia l_{5G_i} . El proceso iterativo finalizará cuando se cumpla una de las dos condiciones siguientes: 1) todos los paquetes en el bloque actual han recibido recursos radio (y, por tanto, B_z está vacío, línea 44), o 2) no es posible encontrar una solución factible al problema que cumpla con el requisito de latencia l_{5G_i} para todos los paquetes de datos dentro del bloque actual (línea 30).

ALGORITHM II: *SCHEDULING*

```

1. Input:  $F_i \forall i \in [1, N_F]$ ,  $t_{sym}$ ,  $t_{gNB,rx}$ ,  $t_{UE,tx}$ 
2.  $HP = \text{lcm}\{p_i\}$  with  $i \in [1, N_F]$ 
3. Create packet blocks  $B_z$ , with  $z=1, 2, \dots$  (Algorithm I)
4. Initialize  $BF_z \neq \emptyset$ 
5. For  $z=1$  to  $N_B$ 
6.   Calculates  $d_i^S$  and  $d_i^R$  for all packets in  $B_z$ 
7.    $N_{pkt}$  = number of packets in  $B_z$ 
8.    $pkt_{i,j}$  = packet with lower  $l_{5G}$  in  $B_z$ 
9.    $iter=0$ ,  $BF_z = \emptyset$ 
10.  For  $iter = 0$  to  $\infty$ 
11.    While  $B_z \neq \emptyset$ 
12.       $s_{aux}$  = first OFDM symbol after  $A_{i,j}+t_{UE,tx}$ 
13.      While  $(s_{aux}+d_i^S-1) \cdot t_{sym} \leq D_{i,j}-t_{gNB,rx}$ 
14.        If there are  $d_i^R$  unallocated RBs in symbols  $s_{aux}$  until  $s_{aux}+d_i^S-1$ 
15.           $pkt_{i,j}$  receives  $d_i^R$  RBs in symbols  $s_{aux}$  until  $s_{aux}+d_i^S-1$ 
16.          First allocated symbol:  $s_{i,j} = s_{aux}$ 
17.          If  $(s_{aux}+d_i^S-1) \cdot t_{sym} == D_{i,j}-t_{gNB,rx}$ 
18.             $BF_z = BF_z + \{pkt_{i,j}\}$ 
19.          Else
20.             $N_{pkt} = N_{pkt} - 1$ 
21.          EndIf
22.          Goto line 27
23.        Else
24.           $s_{aux}++$ 
25.        EndIf
26.      EndWhile
27.      If  $(s_{aux}+d_i^S-1) \cdot t_{sym} > D_{i,j}-t_{gNB,rx}$ 
28.         $BF_z = BF_z + \{pkt_{i,j}\}$ 
29.        If number of packets in  $BF_z + B_z == N_{pkt}$ 
30.          Goto line 46
31.        Else
32.          Reinitiate  $B_z$  and  $N_{pkt}$ 
33.           $B_z = B_z - BF_z$ 
34.          Free radio resources allocated to packets in  $B_z$ 
35.          Goto line 10 (new iteration)
36.        EndIf
37.      Else
38.         $B_z = B_z - \{pkt_{i,j}\}$ 
39.        If  $B_z \neq \emptyset$ 
40.           $pkt_{i,j}$  = packet with lower  $l_{5G}$  in  $B_z$ 
41.        EndIf
42.      EndIf
43.    EndWhile
44.    Goto line 5
45.  EndFor
46. End For

```

4.4. Escenario de simulación

Para la evaluación de los esquemas de *scheduling* propuestos para redes 5G integradas con TSN se ha considerado un escenario en el que la red 5G-TSN proporciona conectividad a una planta industrial en la que se implementa una aplicación de control de supervisión de circuito cerrado o *closed-loop supervisory controller* [22]. La Figura 20 muestra una red 5G privada integrada con TSN que da soporte al escenario evaluado. Esta aplicación integra un PLC que recibe datos de monitorización de 30 sensores (S_1, \dots) que generan un total de N_F flujos TSN ($N_F = 30$). Los flujos TSN, identificados como F_i con $i \in [1, \dots, N_F]$, se transmitirán desde los sensores hasta el PLC a través de distintos *bridges* TSN y el *bridge* 5G virtual. En función de los datos recibidos, el PLC enviará una orden a un actuador representado como A en la Figura 20.

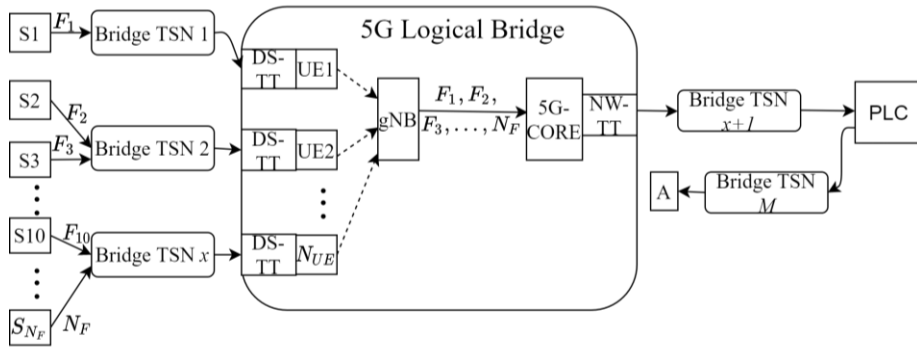


Figura 20. Plano de usuario del escenario evaluado.

Cada flujo TSN F_i , está definido por el tamaño de los paquetes transmitidos (s_i), la periodicidad con la que se crean los paquetes (p_i) y la latencia *end-to-end* demandada entre el sensor y el PLC ($l_{e2e,i}$). Se establece un tamaño de paquete igual a s_i para todos los flujos TSN en una misma simulación; s_i puede tomar cualquier valor entero entre 40 y 250 bytes elegido de manera aleatoria en cada simulación. En base al artículo [26], $l_{e2e,i}$ se establece igual a 4 ms. Cada paquete de datos se debe recibir antes de que se genere el siguiente paquete para el mismo flujo TSN F_i [3], por lo que la periodicidad p_i de generación de paquetes para cada flujo TSN F_i se establece igual o mayor que $l_{e2e,i}$. En concreto, la periodicidad p_i de generación de paquetes para cada flujo TSN F_i se selecciona aleatoriamente entre tres valores posibles p_1, p_2 o p_3 . Las variables p_1, p_2 y p_3 toman valores enteros establecidos aleatoriamente entre 4 y 20 ms.

Como se introdujo en el apartado 2.4, el CNC recopila información sobre la topología de la red, de las capacidades de cada *bridge* que integra la red 5G-TSN y de los requisitos de cada flujo. En base a esta información, el CNC calcula el tiempo de llegada A_i y la

latencia máxima l_{5G_i} que deben experimentar los paquetes de cada flujo TSN dentro del *bridge* 5G virtual para cumplir con los requisitos de latencia *end-to-end* $l_{e2e,i}$. Así pues, el CNC informa a la red 5G sobre los siguientes parámetros (utilizando el TSCAI y el 5QI): la periodicidad p_i , el tamaño de los paquetes s_i , la dirección del flujo f_i que es dirección UL en este caso (desde el UE hasta el gNB), el tiempo de llegada A_i en el puerto de entrada del *bridge* 5G virtual, y el requisito de latencia máxima del flujo TSN dentro del *bridge* 5G virtual l_{5G_i} . En esta tesis, se ha hecho uso de las expresiones derivadas en el estudio presentado en el apartado 3.2 para establecer el tiempo de llegada A_i y el requisito de latencia l_{5G_i} considerando el instante de tiempo en el que se genera cada paquete, el requisito de latencia *end-to-end* $l_{e2e,i}$ y la topología de red. Al considerar una red 5G privada, se asume que el tiempo de tránsito de los paquetes de los flujos TSN a través del *core* de 5G es despreciable [66]. El procedimiento seguido para calcular A_i y l_{5G_i} se presenta en el Anexo B. Como resultado, se ha obtenido que los tiempos de llegada A_i para cada flujo se establece de manera aleatorio entre $75\mu\text{s}$ y $700\mu\text{s}$. La latencia l_{5G_i} se calculará en función del valor del tiempo de llegada y del requisito de latencia *end-to-end* siguiendo la ecuación (33) del Anexo B. En base a esta ecuación, la latencia más restrictiva y menos restrictiva son 2.27 ms y 2.9 ms, respectivamente.

Se considera una red 5G con un ancho de banda de 20 MHz y SCS de 30kHz como se recomienda en [67] para entornos industriales. La celda funciona en modo TDD [68] y para cada *slot* de 14 símbolos OFDM se reserva el primer símbolo OFDM para transmitir mensajes de control en DL y el último símbolo OFDM para mensajes de control en UL, mientras que los 12 símbolos OFDM restantes se utilizan para transmitir datos en UL. Se asume que los UEs están situados en lugares que garantizan condiciones de *Line of Sight* (LOS) con el gNB. Se ha considerado que todos los paquetes se transmiten utilizando el MCS 12 en la tabla 1 de MCSs definida en 3GPP TS 38.214 [33]. Este MCS garantiza un buen compromiso entre robustez y eficiencia espectral. El MCS 12 corresponde al uso de un orden de modulación Q_m igual a 4 y tasa de codificación R igual a 434/1024. El número de RBs necesarios para transmitir cada paquete de datos, puede calcularse en función del MCS utilizado en la transmisión según la expresión presentada en (21). En (21), s_i es el tamaño del paquete a transmitir en bytes, s_{header} es el tamaño en bytes del encabezado IPv4 que se agrega al paquete de datos y s_{CRC} el tamaño en bytes del código de redundancia cíclica que se agrega al final del paquete. tbs_i representa el tamaño de bloque de transporte más pequeño de los valores disponibles dados en [16] que se puede utilizar para transmitir $s_i + s_{header}$ bytes, y $N_{sc, RB}$ representa el número de subportadoras en un bloque de recursos y es igual a 12.

$$d_i = \left\lceil \frac{(tbs_i(s_i + s_{header}) + s_{CRC}) \cdot \delta}{R \cdot Q_m \cdot N_{sc, RB}} \right\rceil \quad (21)$$

4.5. Evaluación del rendimiento

En esta sección se evalúa el rendimiento de los esquemas de *scheduling* propuestos para redes 5G integradas en redes TSN. Para llevar a cabo esta evaluación, se ha considerado el escenario descrito en 0. El rendimiento obtenido con los esquemas de *scheduling* propuestos se ha comparado con un esquema de referencia. El esquema de referencia es un esquema de *scheduling* comúnmente utilizado en la literatura (por ejemplo, [12]) que asigna recursos radio periódicamente a cada flujo TSN F_i con la misma periodicidad p_i con la que se generan los paquetes. Para garantizar una comparación justa, el esquema de referencia considera el mismo criterio y asigna recursos en primer lugar a los flujos TSN con requisitos de latencia 5G más estrictos. Es decir, el esquema de referencia asigna recursos en primer lugar a los flujos TSN con valores l_{5G_i} más bajos. El análisis se ha llevado a cabo utilizando Matlab y se han utilizado las métricas de evaluación descritas a continuación.

4.5.1. Métricas

Las métricas utilizadas para evaluar el rendimiento obtenido con los distintos esquemas de *scheduling* propuestos son:

- Porcentaje de problemas de *scheduling* con solución satisfactoria (P_{satisf}). Los esquemas de *scheduling* propuestos buscan soluciones de *scheduling* que asignen recursos radio a todos los paquetes de todos los flujos TSN F_i de manera que se satisfagan los requisitos de latencia l_{5G_i} para cada flujo TSN F_i . En caso contrario, el problema quedará sin solución. Para el esquema de referencia, es importante recordar que éste se centra en buscar una solución de *scheduling* para el primer paquete de cada flujo F_i (ya que, para los siguientes paquetes de cada flujo TSN F_i , los recursos radio se asignan con periodicidad p_i). En cambio, se considera una solución satisfactoria aquella solución de *scheduling* en la que no hay conflictos en la asignación de recursos para la transmisión de los distintos paquetes de todos los flujos TSN. Es decir, el mismo recurso radio no se ha asignado para la transmisión de más de un paquete de datos.

- Porcentaje de paquetes con conflicto ($PpktConflict$). Representa el porcentaje de paquetes con conflicto, es decir, el porcentaje de veces que dos o más paquetes reciben los mismos recursos radio para su transmisión.
- Latencia media experimentada (\bar{l}). En esta tesis, la latencia UL $l_{i,j}$ se define como el tiempo que transcurre desde que un paquete es recibido en la capa MAC del UE, hasta que es recibido en la capa MAC del gNB. Por lo tanto, la latencia media experimentada por cada flujo TSN F_i (\bar{l}_i) se calcula teniendo en cuenta la latencia experimentada por todos los paquetes del flujo TSN F_i que se transmiten dentro del *HP* (el número de paquetes transmitidos por flujo TSN F_i en el hiperperiodo o *HP* viene dado por N_{pkt}^i). Por lo tanto, la latencia media experimentada por cada flujo F_i es:

$$\bar{l}_i = \frac{\sum_{j=1}^{N_{pkt}^i} l_{i,j}}{N_{pkt}^i} \quad (22)$$

La latencia media experimentada en cada problema de *scheduling* se calcula teniendo en cuenta la latencia media experimentada por todos los flujos TSN F_i que se transmiten dentro del *HP*. Por lo tanto, la latencia media \bar{l} de un esquema de *scheduling* se calcula teniendo en cuenta todos los problemas de *scheduling* abordados que tengan solución (N_{sch}):

$$\bar{l} = \frac{\sum_{k=1}^{N_{sch}} \left(\frac{\sum_i \bar{l}_i}{N_F} \right)}{N_{sch}} \quad (23)$$

- Desviación Típica ($\bar{\sigma}$). La desviación típica se calcula para cada flujo TSN F_i , ya que de este modo se puede saber el *jitter* que va a sufrir cada paquete dentro de un flujo TSN. Si todos los paquetes $pkt_{i,j}$ de un flujo TSN F_i , siendo $j=[1 \dots N_{pkt}^i]$, experimentan la misma latencia, la desviación típica de dicho flujo TSN será 0.

$$\bar{\sigma} = \frac{\sum_{k=1}^{N_{sch}} \left(\frac{\sum_i \sigma(l_{i,j})}{N_F} \right)}{N_{sch}} \text{ con } j = \{1 \dots N_{pkt}^i\} \quad (24)$$

- Coste computacional. El coste computacional se ha medido en función del tiempo de simulación que conlleva la resolución de los problemas de *scheduling* con cada uno de los esquemas evaluados. Los tiempos de computación que se mostrarán en esta tesis han sido obtenidos con una CPU Intel(R) Core(TM) i7-5930K @ 3.50GHz y RAM de 16GB.

4.5.2. Resultados

Debido a las altas demandas y la disponibilidad limitada de recursos radio, puede ser difícil para los esquemas de *scheduling* encontrar soluciones que satisfagan toda la demanda de recursos radio para todos los paquetes de datos a transmitir de los distintos flujos TSN. En este contexto, se ha analizado en primer lugar la capacidad de los esquemas de *scheduling* propuestos y de referencia de encontrar soluciones de *scheduling* que asignen recursos radio a todos los flujos TSN. Estas soluciones tienen que satisfacer los requisitos de latencia y evitar colisiones entre diferentes paquetes debido a las asignaciones simultáneas. La Figura 21 presenta el porcentaje de soluciones satisfactorias según el *scheduling* (representado como P_{satisf}). Como se puede observar en esta figura, los *schedulings* propuestos superan en un 18% al esquema de referencia en cuanto al número de problemas resueltos para todos los paquetes de todos los flujos TSN. Esto es debido a que las asignaciones de recursos realizadas con el esquema de referencia aseguran la no superposición en tiempo y frecuencia de los recursos radio asignados únicamente al primer paquete de cada flujo TSN. Sin embargo, estas asignaciones no pueden garantizar lo mismo para los paquetes subsiguientes. Esto puede resultar en conflictos de asignación de recursos radio a distintos paquetes que se transmiten con una periodicidad diferente. Esto significa que varios flujos TSN pueden recibir los mismos recursos radio después de varios periodos, tal y como se presentó en el apartado 2.5. Estos conflictos de asignación de recursos radio pueden provocar colisiones entre paquetes, por lo tanto, si hay algún conflicto entre diferentes paquetes P_{satisf} es igual a 0. Es importante destacar que el porcentaje de veces que el esquema de referencia encuentra soluciones satisfactorias libre de conflictos es menor que el 26% cuando el número de flujos TSN que demanda recursos es igual o superior a 15. Sin embargo, con los esquemas de *scheduling* propuestos, cuando el número de flujos TSN es igual a 15 ($N_F=15$), ambas propuestas pueden encontrar soluciones en el 100% de los escenarios evaluados. La Figura 21 también muestra que todos los esquemas de *scheduling* evaluados presentan cierta dificultad a la hora de encontrar soluciones a los problemas de *scheduling* cuando el número de flujos TSN aumenta por encima de 20, lo cual es debido al aumento de la demanda de recursos radio.

Los resultados de la Figura 21 muestran que ambas propuestas H-FAST como O-FAST son capaces de obtener soluciones para un número igual o mayor de problemas de

scheduling que el esquema de referencia. El esquema de *scheduling* que más soluciones encuentra es O-FAST, el esquema que aplica técnicas de optimización para resolver el problema. Por ejemplo, cuando N_F es igual a 30, O-FAST alcanza soluciones para un 45% de los problemas de *scheduling* evaluados, mientras que H-FAST lo hace para el 39%. El número de problemas con solución con la técnica de referencia disminuye hasta el 2%. Sin embargo, es importante destacar que H-FAST es capaz de resolver el mismo número de problemas que O-FAST cuando N_F es igual o menor a 25. Cuando el número de flujos transmitidos es menor que 25, los dos esquemas propuestos son capaces de resolver el 100% de los problemas evaluados.

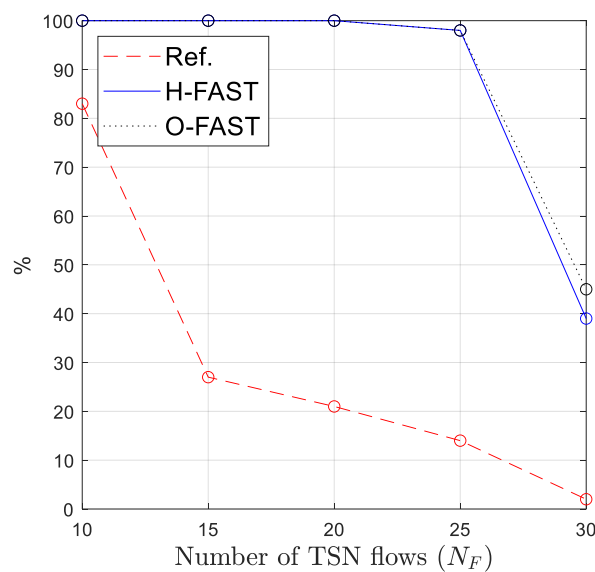


Figura 21. Porcentaje de esquemas de *scheduling* satisfactorias en los que se garantizan los requisitos de latencia a todos los paquetes TSN transmitidos (P_{satisf}) para H-FAST, O-FAST y el esquema de referencia en función del número de flujos TSN evaluados.

La Figura 22 muestra el porcentaje de paquetes que experimentan colisiones en el esquema de referencia debido a las asignaciones simultáneas. Los resultados en la Figura 22 muestran que a partir de 25 flujos TSN, el 13% de los paquetes van a colisionar entre sí por realizar una única configuración CG para cada flujo. Como ya se ha dicho en el apartado 2.1, la fiabilidad de las comunicaciones está sujeta a la probabilidad de error, es decir, para que una comunicación sea fiable la cantidad de transmisiones fallidas permitidas debe ser menor que la probabilidad de error. En la Tabla 2 del apartado 2.1, todas las aplicaciones consideradas tienen que garantizar una fiabilidad de 10^{-7} . Por lo tanto, es importante resaltar que el porcentaje de paquetes con conflicto (13%) es un porcentaje muy elevado teniendo en cuenta los requisitos de fiabilidad de las aplicaciones industriales.

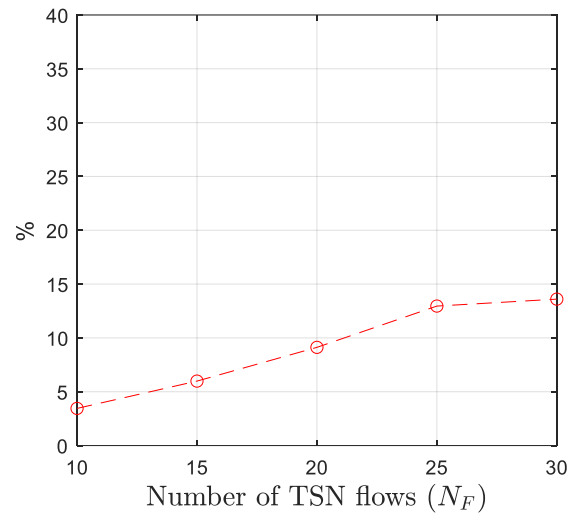
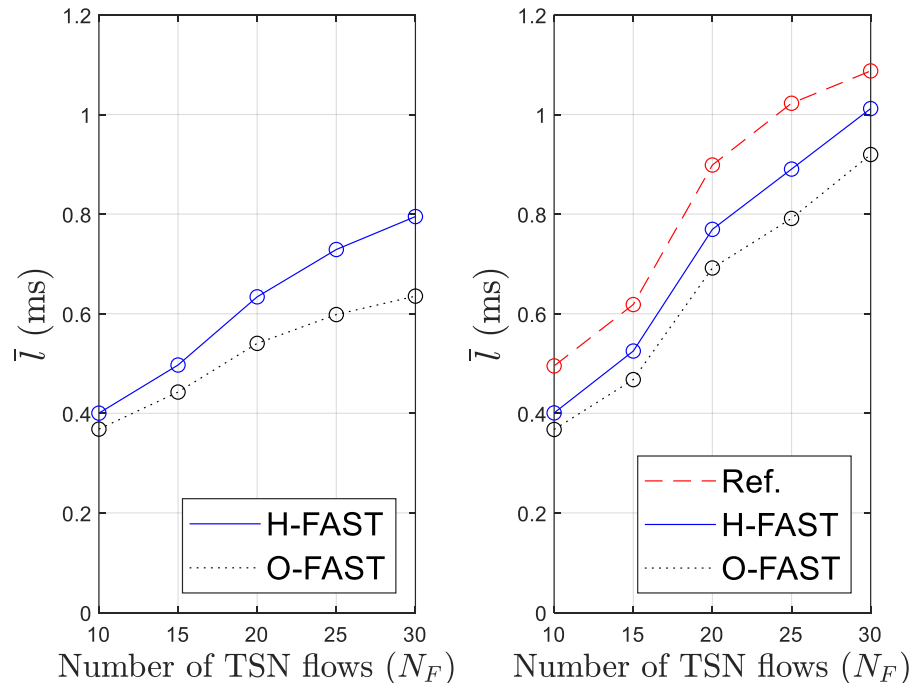


Figura 22. Porcentaje de colisiones ($P_{pktConflict}$) con el esquema de referencia.

En esta sección se analiza el porcentaje de soluciones satisfactorias, es decir, aquellas que cumplen con los requisitos de latencia establecidos para la red 5G. Sin embargo, es interesante analizar la latencia media experimentada, pues cuanto menor sea esta, mayor probabilidad habrá de realizar retransmisiones de paquetes recibidos con errores. Según el Anexo B, el requisito de latencia l_{5G_i} más restrictivo que los flujos TSN deben de soportar dentro del *bridge* 5G virtual es de 2.27 ms. La Figura 23.a representa la latencia media experimentada al utilizar los esquemas de *scheduling* H-FAST y O-FAST en todos los problemas evaluados que H-FAST encuentre solución. Para comparar el rendimiento obtenido con los esquemas propuestos en término de latencia con el esquema de referencia, la Figura 23.b compara la latencia media para los tres esquemas de *scheduling* solamente en aquellos problemas evaluados para los que el esquema de referencia alcanzó soluciones satisfactorias, es decir, cuando el esquema de referencia logra una solución de *scheduling* que garantiza que no hay conflictos en la asignación de recursos para todos los paquetes de todos los flujos TSN. Los resultados de la Figura 23 muestran que la latencia promedio experimentada por los flujos TSN está muy por debajo del requisito de máxima latencia que debe garantizar la red 5G. Los resultados en la Figura 23.a muestran que utilizando O-FAST la latencia media experimentada es menor que con H-FAST. En concreto, O-FAST reduce la latencia media experimentada entre un 8% (cuando $N_F=10$) y un 20% (cuando $N_F=30$) en comparación con H-FAST. Los resultados en Figura 23.b muestran que H-FAST y O-FAST reducen la latencia media experimentada por los paquetes transmitidos en más de un 7% y 15%, respectivamente, para todos los casos analizados en comparación con la

técnica de referencia. Esto se debe a que los esquemas propuestos tienen una mayor flexibilidad en la asignación de recursos radio.



a) Latencia media experimental considerando todos los problemas de *scheduling* evaluados.

b) Latencia media experimental considerando el subconjunto de problemas de *scheduling* evaluados en los que no hay colisiones con la técnica de referencia.

Figura 23. Latencia media experimental con H-FAST, O-FAST y el esquema de referencia en función del número de flujos TSN evaluados.

La Figura 24 muestra la desviación estándar de la latencia experimental en función del número de flujos TSN transmitidos en la red 5G solamente para aquellos problemas en los que H-FAST encuentra una solución. Esta figura muestra solo resultados para H-FAST y O-FAST, ya que la desviación estándar de la latencia experimental al utilizar el esquema de referencia es nula pues se asignan los mismos recursos radio periódicamente a cada paquete del flujo TSN. La mayor flexibilidad a la hora de asignar los recursos radio de los esquemas propuestos permite disminuir la latencia experimental en la red 5G (Figura 23.b) a expensas de un aumento en la desviación estándar de la misma, tal y como se muestra en la Figura 24. A su vez, los paquetes experimentan una desviación típica ligeramente menor cuando se utiliza el esquema de *scheduling* H-FAST en comparación con O-FAST. Esto es debido a que H-FAST asigna los recursos radio siguiendo un mismo orden, es decir, H-FAST siempre empieza a asignar recursos radio a los paquetes que tengan requisitos de latencia más restrictivos. En cambio, O-FAST busca la solución óptima sin seguir este orden.

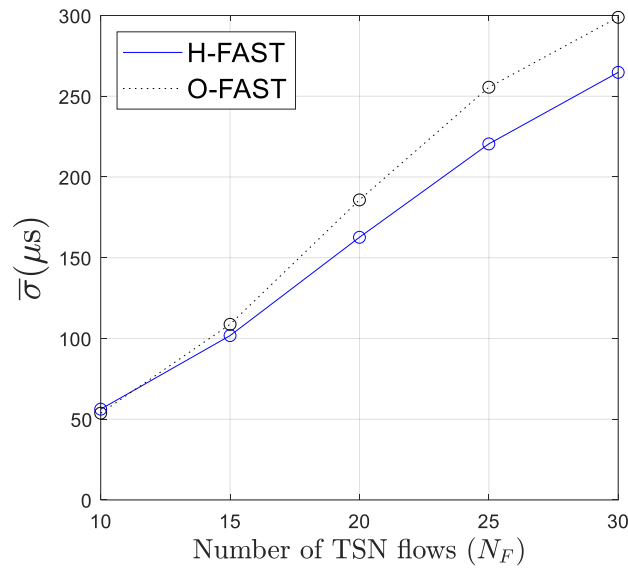


Figura 24. Desviación típica media de la latencia cuando se utilizan los *schedulers* propuestos en escenarios en los que H-FAST encuentra solución.

La Figura 25 muestra la media del tiempo de computación requerido por cada esquema de *scheduling* para resolver los problemas de asignación de recursos. Como se puede observar, el esquema de referencia es el que menor coste computacional presenta, tal y como cabía esperar, mientras que O-FAST es el esquema que mayor coste computacional presenta. Es posible observar que O-FAST tiene un coste computacional mucho mayor. Como muestra la Figura 25, O-FAST requiere una media de 100 s para alcanzar la solución al problema evaluado cuando se transmiten 10 flujos TSN, mientras que el tiempo de computación medio aumenta hasta 1h 30 minutos cuando el número de flujos es de 30. Por otro lado, H-FAST puede alcanzar soluciones al problema de *scheduling* en menos de 61 ms incluso en escenarios con 30 flujos TSN. Aunque como se observó en la Figura 21, O-FAST obtiene un rendimiento ligeramente mejor que H-FAST en términos de problemas con soluciones satisfactorias (solamente en el escenario con mayor número de flujos TSN) y de latencia, H-FAST proporciona un mejor compromiso entre rendimiento y coste computacional.

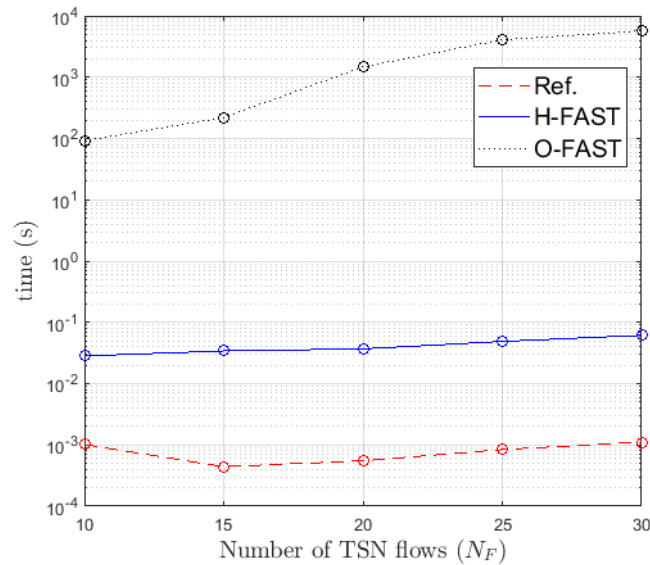


Figura 25. Tiempo computacional promedio en segundos para asignar recursos en función de los tres *schedulers*.

4.6. Conclusiones

En este capítulo se han diseñado y evaluado dos esquemas de *scheduling* basados en CG para redes 5G integradas con redes TSN. En particular, se considera el modelo de integración 5G-TSN, donde la red 5G se integra en la red TSN como un *bridge* virtual. El objetivo de ambos esquemas propuestos es alcanzar una solución de *scheduling* en 5G que tenga en cuenta las decisiones previamente tomadas por el *scheduling* de TSN, con el fin de cumplir de manera eficiente y efectiva con los requisitos de latencia y determinismo de extremo a extremo del tráfico TSN. Para ello, los esquemas de *scheduling* propuestos explotan la información proporcionada por TSN a la red 5G mediante mensajes como TSCAI y el 5QI.

Los esquemas de *scheduling* basados en CG comúnmente utilizados en la literatura asignan recursos radio de forma periódica para cada flujo TSN en función de la periodicidad de los paquetes, siguiendo la fórmula establecida en 3GPP [16]. Sin embargo, esta estrategia puede generar conflictos de asignación de recursos radio entre paquetes de diferentes flujos TSN debido a sus diferentes periodicidades (es decir, los mismos recursos radio pueden ser asignados para la transmisión de diferentes paquetes). Los esquemas de *scheduling* propuestos resuelven esta problemática al configurar varios CGs para cada flujo TSN. De esta manera, se evita la asignación del mismo recurso radio para transmitir

paquetes de diferentes flujos TSN, además de que se sigue respetando la fórmula descrita en 3GPP [16]. En particular, se han propuesto dos esquemas diferentes. Ambos esquemas buscan satisfacer la demanda de recursos radio y los requisitos de latencia de todos los flujos TSN. Una vez alcanzado este objetivo, los esquemas de *scheduling* propuestos intentarán minimizar la latencia experimentada por los distintos flujos TSN. El primero de ellos O-FAST aplica técnicas de programación lineal entera binaria para encontrar la solución óptima a los problemas de *scheduling* según el objetivo propuesto. El segundo esquema de *scheduling* propuesto H-FAST está basado en un algoritmo heurístico que busca alcanzar la mejor solución posible, pero manteniendo un bajo coste computacional.

Los resultados obtenidos en este capítulo demuestran que ambos esquemas O-FAST y H-FAST mejoran significativamente la cantidad de flujos TSN que pueden ser atendidos de manera efectiva en la red 5G-TSN integrada en comparación con el esquema de *scheduling* utilizado como esquema de referencia. O-FAST encuentra soluciones a un mayor número de problemas que H-FAST en el escenario de mayor carga (en términos del número de flujos TSN), aunque la diferencia es tan solo de 6%. Sin embargo, O-FAST presenta un coste computacional mucho mayor que H-FAST, lo cual es una desventaja principalmente en entornos variables y reconfigurables, como son las fábricas del futuro.

En cuanto a la latencia, es crucial enfatizar que lo fundamental es que los paquetes lleguen al puerto de salida dentro del tiempo especificado por el *scheduler* de TSN, para que puedan ser transmitidos al siguiente *bridge* de la red en la ruta del flujo TSN según la planificación realizada por la red TSN. En cambio, una asignación de recursos que permite minimizar la latencia ofrece la ventaja de tener más tiempo disponible en caso de que sea necesario realizar la retransmisión de paquetes erróneos. Los resultados de la implementación de los nuevos mecanismos muestran una reducción en la latencia promedio experimentada en la red 5G, a cambio de tener un aumento en la desviación estándar. Sin embargo, a pesar de que la desviación estándar haya aumentado con los esquemas de *scheduling* propuestos, los traductores TSN que se implementan en los puertos de entrada y salida del *bridge* 5G virtual pueden ayudar a minimizar o eliminar dicha variabilidad en la latencia. Para ello los traductores DS-TT y NW-TT utilizan el mecanismo de almacenamiento y reenvío o *hold and forward buffering* que consiste en asegurar que los paquetes sean transmitidos en el momento preciso, es decir, evitando que la transmisión del paquete al siguiente *bridge* de la red TSN se adelante o se retrase.

En el Anexo A se han incluido nuestros artículo de conferencia [13] y [14]. En caso del artículo [13] se han extendido los resultados y discusiones de este capítulo. En este artículo, se ha evaluado el esquema de *scheduling* heurístico H-FAST y el de referencia en un escenario donde la latencia extremo a extremo es igual a la periodicidad del tráfico. Los

resultados indican que, al aumentar la latencia extremo a extremo, el requisito de la latencia máxima para un flujo TSN dentro de 5G es menos restrictivo y, por lo tanto, el número de soluciones que pueden encontrar los *schedulers* evaluados aumenta. En cambio, la problemática de colisiones entre recursos debido al *scheduler* de referencia sigue siendo la misma.

5. *Scheduling* para redes 5G integradas en redes TSN compatibles con el estándar 3GPP

En una red integrada 5G-TSN, la coordinación de los esquemas de *scheduling* de ambas redes es crucial para lograr una integración exitosa y cumplir con los requisitos de latencia y determinismo de las aplicaciones industriales. En el capítulo anterior, se han propuesto dos esquemas de *scheduling*, O-FAST y H-FAST, para la gestión del tráfico TSN en la red 5G. Los mecanismos de *scheduling* propuestos buscan alcanzar la mejor solución posible (utilizando técnicas de optimización y un esquema heurístico, respectivamente) que maximice el número de flujos TSN que se pueden soportar por la red, y a su vez, minimice la latencia experimentada por dichos flujos. En la búsqueda de esta mejor solución, no se ha tenido en cuenta la limitación impuesta por el estándar 3GPP en cuanto al número máximo de configuraciones CG que pueden asignarse a un UE con el objetivo de analizar el máximo rendimiento alcanzable. En este capítulo, se define un nuevo esquema de *scheduling* que evoluciona el esquema H-FAST para hacerlo compatible con el estándar actual de 3GPP. El esquema propuesto mantiene el objetivo de H-FAST, pero tiene en cuenta el número máximo de configuraciones CG que pueden asignarse a un UE. El gNB puede asignar hasta 12 configuraciones CG en un *bandwidth part* o BWP [15] (ver apartado 2.3).

Tal y como se ha presentado a lo largo de esta tesis, el tipo de *scheduling* más adecuado para dar soporte al tráfico TSN (tráfico periódico y con requisitos deterministas) es el *scheduling* semi-estático, es decir, *Configured Grant* o CG para transmisiones en UL. El mecanismo de CG preasigna recursos radio a cada flujo TSN F_i y esta asignación se repite

periódicamente. Esta asignación periódica puede resultar en conflictos de asignación de recursos entre distintos flujos TSN cuando la periodicidad con la que se asignan los recursos es diferente para cada flujo (ver apartado 2.5). Para abordar este problema, los esquemas de *scheduling* propuestos en el capítulo 4 identifican un periodo común a todos los flujos TSN presentes en la red denominado hiperperiodo o *HP* en base al cual el patrón de llegada de paquetes para todos los flujos se repite. Una vez conocido el *HP*, los esquemas propuestos asignan recursos radio para la transmisión de cada paquete dentro del *HP*, repitiendo esta asignación de forma periódica con una periodicidad igual a *HP*. Esto resulta en la creación de tantas configuraciones CG como paquetes contenga cada flujo TSN F_i dentro del *HP*. Cuando el número de flujos TSN en la red aumenta, aumenta la posibilidad de que incremente el número de flujos con periodicidades diferentes. Esto hace que el *HP* tome un valor mayor, y en consecuencia, el número de paquetes para cada flujo TSN en el *HP* aumente, y con ello aumenta el número de configuraciones CG asignadas a cada flujo TSN.

En este contexto, 3GPP limita el número de configuraciones CG asignadas a un UE a un máximo de 12 configuraciones en un BWP [15]. Por lo tanto, los esquemas de *scheduling* propuestos en el capítulo 4 no cumplen con la limitación impuesta por el 3GPP cuando el número de paquetes dentro de *HP* es mayor que 12. Para disminuir el número de configuraciones CG, hay que permitir configuraciones CG que asignen recursos con una periodicidad menor al *HP*. Sin embargo, como ya se ha dicho, esto puede resultar en conflictos entre la asignación de recursos a distintos flujos TSN. Estos conflictos pueden evitarse si las configuraciones se activan y desactivan para cada UE, solamente en los momentos en los que vayan a utilizarse. Cuando se crea más de una configuración CG para un mismo UE, el gNB decide cuál o cuáles de los CG configurados están activos en cada momento (un UE puede tener varias configuraciones CG activas al mismo tiempo) y así se logra una mejor coordinación entre los UE. Por un lado, el gNB puede activar todas las configuraciones CG a la vez y cada paquete del UE se transmitirá en la configuración CG más cercana. Sin embargo, esto puede generar conflictos entre diferentes configuraciones CG o conllevar un uso ineficiente de los recursos radio y las configuraciones CG. Por otro lado, el gNB puede activar y desactivar las configuraciones CG según sea necesario. Este proceso por parte del gNB conlleva señalización adicional en la red que además puede aumentar la latencia de la comunicación. Para evitar esta señalización y asignar los recursos radio de manera eficiente evitando las posibles colisiones, en esta tesis se propone el uso de un patrón o vector de activación de las configuraciones CG asignadas a un UE. Este vector de activación es calculado por el gNB y es enviado al UE junto con las configuraciones CG. Este vector de activación debe ser utilizado por el UE para identificar qué configuración CG está activa en cada momento e identificar los recursos radio a utilizar en la transmisión de cada paquete. Para ello, esta tesis propone una modificación de la

fórmula definida en el estándar del 3GPP [16] la cual es utilizada por el UE para identificar la configuración activa y los recursos radio a utilizar en la transmisión de cada paquete. El uso del vector de activación permite coordinar el acceso a los recursos radio de UEs con configuraciones CG que se pueden solapar en determinados momentos, dando solución a posibles conflictos de una manera eficiente.

En los siguientes apartados de este capítulo se presenta en primer lugar la definición del vector de activación y la modificación propuesta de la fórmula definida en el estándar 3GPP TS 38.321 [16] para la identificación de los recursos radio asignados en las configuraciones CG (apartado 5.1). En el apartado 5.2 se presenta una nueva propuesta de *scheduling* que hace uso del vector de activación para encontrar la mejor solución posible que minimice las configuraciones CG asignadas a cada UE y se cumpla con la restricción del número de configuraciones CG que pueden asignarse a un UE. En el apartado 5.3 se analiza el rendimiento alcanzado con la nueva propuesta. Los resultados obtenidos demuestran que el nuevo *scheduler* puede soportar el mismo número de flujos que el *scheduler* H-FAST, pero cumpliendo con los requisitos de 3GPP. Finalmente, en el apartado 5.4 se resumen las conclusiones obtenidas del trabajo presentado en este capítulo.

5.1. Vector de activación de las configuraciones CG

En esta tesis se propone la definición del vector de activación asociado a una configuración CG que indica si dicha configuración está activa para la transmisión de cada paquete de un flujo de datos, en este caso, flujo de datos TSN. En este capítulo se va a expresar una configuración CG asignada a un flujo TSN F_i como $CG_{c_i}^i = \{s_{c_i}^i, p_{c_i}^i, V_{c_i}^i\}$, donde $s_{c_i}^i$ indica el primer símbolo asignado al primer paquete del flujo TSN F_i , $p_{c_i}^i$ indica la periodicidad con la que se repite la reserva de recursos, y $V_{c_i}^i$ es el vector de activación/desactivación. El parámetro c_i es el cardinal que identifica la configuración CG asignada al flujo TSN F_i y puede tomar un valor desde 1 hasta el número de configuraciones asignadas a dicho flujo (que estará limitado a 12). El primer símbolo asignado $s_{c_i}^i$ al primer paquete se calcula como: $s_{c_i}^i = s_{aux} - p_{c_i}^i \cdot (j - 1)$, siendo s_{aux} el primer símbolo del HP que se asigna al paquete $pkt_{i,j}$ con el que se crea la configuración $CG_{c_i}^i$. El parámetro $p_{c_i}^i$ es igual a la periodicidad p_i con la que los paquetes del flujo TSN F_i llegan al *bridge* 5G virtual⁹. El vector $V_{c_i}^i = \{a_{c_i,j}^i\}$ tendrá tantos elementos como número de paquetes tenga el flujo TSN F_i dentro del HP. Cada elemento del vector $V_{c_i}^i$ puede tomar un valor igual a 0 o a 1. $a_{c_i,j}^i = 1$ significa que el paquete $pkt_{i,j}$ se transmitirá teniendo en cuenta la configuración $CG_{c_i}^i$, y en

⁹ Esta igualdad es posible ya que en esta tesis se han considerado valores p_i que coinciden con las periodicidades admitidas en la RAN de 5G.

caso contrario, $a_{cij}^l=0$. La Figura 26.a muestra un ejemplo de asignación de recursos radio a tres flujos TSN. Los instantes de llegada de los flujos TSN están representados con una flecha, y cada flujo, así como las asignaciones de recursos radio a cada flujo, se representan con un color diferente (azul, rojo, verde). Las Figura 26.b, 26.c y 26.d ilustran las configuraciones CG que se han creado para los flujos TSN F_1 , F_2 y F_3 , respectivamente, y los vectores de activación asociados. Por ejemplo, la Figura 26.b muestra que el flujo TSN F_1 ha recibido dos configuraciones: $CG_1^1 = \{1, 3, [10000]\}$ y $CG_2^1 = \{0, 3, [01111]\}$. Como se puede observar, solo el primer paquete del flujo F_1 se transmite utilizando CG_1^1 , mientras que el 2º, 3º, 4º y 5º paquetes se transmiten utilizando CG_2^1 . En el caso de la configuración CG_2^1 , s_2^1 se calcula como $s_2^1 = 3 - 3 \cdot 1 = 0$, siendo $s_{aux} = 3$, $p_1 = 3$ y $(j-1) = 2-1=1$.

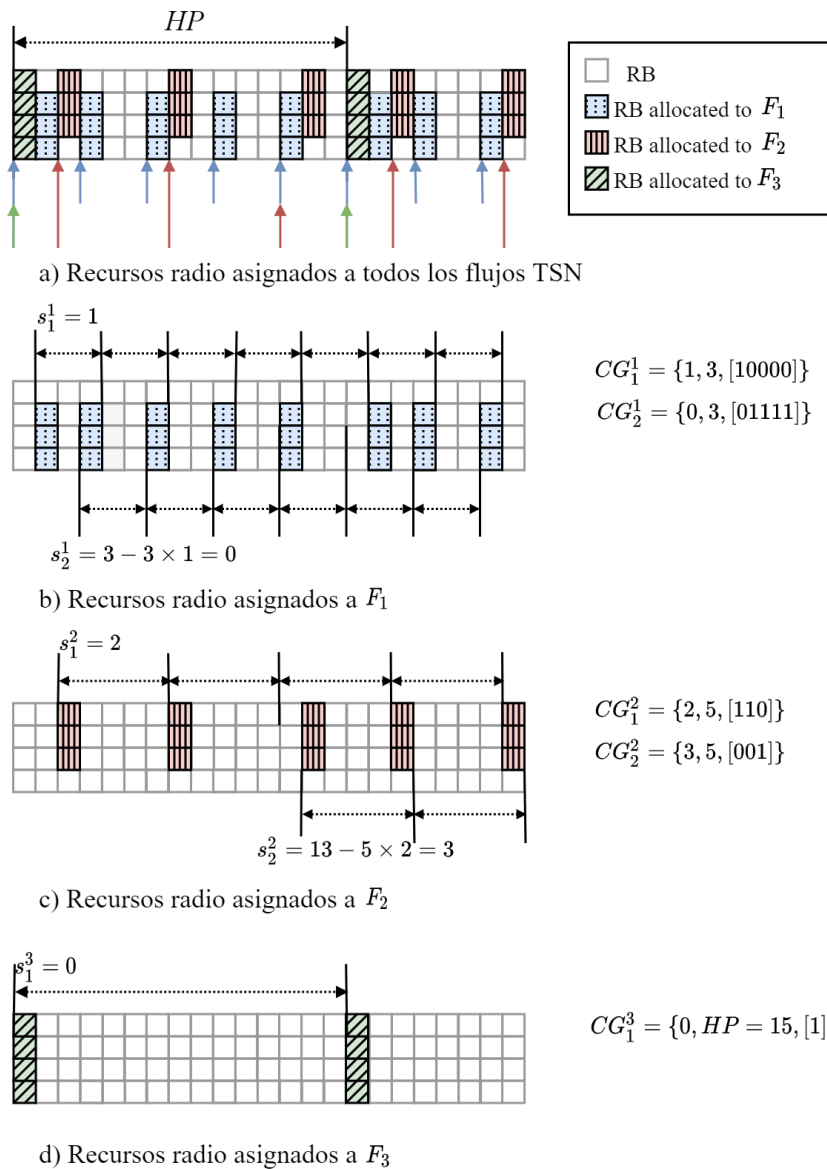


Figura 26. Ejemplo explicativo del uso del vector de activación de las configuraciones CG.

El vector de activación es enviado junto al resto de parámetros de configuración (primer símbolo OFDM asignado $s_{c_i}^i$ y periodicidad $p_{c_i}^i$) al UE. El UE utilizará toda esta información para determinar los recursos radio asignados para la transmisión de cada paquete. El estándar 3GPP TS 38.321 [16] define la expresión (25) para determinar en qué *slots* se encuentran los recursos asignados con la configuración CG para transmitir el N -ésimo paquete ($N \geq 0$).

$$\begin{aligned} & [(SFN \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot}) \\ & + (\text{slot number in the frame} \times \text{numberOfSymbolsPerSlot}) + \text{symbol number in the slot}] = \\ & (\text{timeReferenceSFN} \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot} \quad (25) \\ & + \text{timeDomainOffset} \times \text{numberOfSymbolsPerSlot} + S + N \times \text{periodicity}) \\ & \text{modulo}(1024 \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot}) \end{aligned}$$

En (25), *numberOfSlotsPerFrame* y *numberOfSymbolsPerSlot* representan el número de *slots* que hay en una trama y el número de símbolos que hay en un *slot*, respectivamente. El *numberOfSlotsPerFrame* es igual a 1 *slot* cuando la numerología es igual a 0 e igual a 16 *slots* para la numerología 4, mientras que *numberOfSymbolsPerSlot* depende del número de símbolos del *slot*. Los parámetros *timeReferenceSFN*, *timeDomainOffset* y *S* identifican los recursos radio asignados para la transmisión del primer paquete del flujo de datos (la trama, el número del primer *slot* asignado dentro de la trama y el primer símbolo asignado dentro de dicho *slot*). Esta asignación se repite con periodicidad *periodicity* o *periodicityExt* (este último admite todas las periodicidades múltiples de un *slot* de 14 símbolos OFDM). Con estos parámetros, se puede calcular en qué recursos radio se transmite el paquete N del flujo TSN F_i . Estos recursos radio están definidos por los parámetros: número de trama (SFN), número de *slot* en la trama ('slot number in the frame') y el número de símbolo en el *slot* ('symbol number in the slot').

Para una configuración $CG_{c_i}^i$ asignada a un flujo TSN F_i , $((\text{timeReferenceSFN} \times \text{numberOfSymbolsPerSlot} + \text{timeDomainOffset}) \times \text{numberOfSlotsPerFrame} + S)$ en (25) es igual a $s_{c_i}^i$, y *periodicity* es igual a $p_{c_i}^i$. Según la notación y simbología matemática utilizada en esta tesis, los paquetes de cada flujo se identifican como $pkt_{i,j}$, donde i identifica al flujo TSN al que pertenece y j es el índice que identifica la posición del paquete dentro del *HP*, es decir, j puede tomar valores desde 1 hasta N_{pkt}^i , donde $N_{pkt}^i = HP/p_i$ es el número total de paquetes del flujo TSN dentro del *HP*. Por tanto, es posible establecer la relación

entre j y N ($N \geq 0$) como: $j = N + 1 - N_{pkt}^i \times \left\lfloor \frac{N}{N_{pkt}^i} \right\rfloor$. La relación entre N y j se ilustra en la Figura 27 con un ejemplo en el que $HP = 90$ ms, $p_i = 10$ ms y $N_{pkt}^i = 3$.

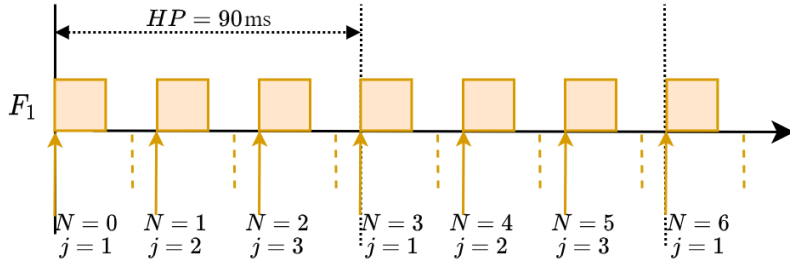


Figura 27. Relación del paquete N con el indicador j .

En esta tesis se propone modificar la expresión en (25) para incorporar el vector de activación y así programar desde el instante en el que se configura el CG los momentos en los que las distintas configuraciones CG estarán activas o inactivas. El vector de activación permite realizar una mejor y más eficiente coordinación del acceso a los recursos radio de distintos UEs además de minimizar la señalización entre el UE y el gNB para llevar a cabo la activación/desactivación de las configuraciones CG de cada UE. En concreto, en esta tesis se propone el uso de la siguiente expresión:

$$\begin{aligned}
 & [(SFN \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot}) \\
 & + (\text{slot number in the frame} \times \text{numberOfSymbolsPerSlot}) + \text{symbol number in the slot}] = \\
 & ((\text{timeReferenceSFN} \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot} \\
 & + \text{timeDomainOffset} \times \text{numberOfSymbolsPerSlot} + S + N \times \text{periodicity}) \\
 & \text{modulo}(1024 \times \text{numberOfSlotsPerFrame} \times \text{numberOfSymbolsPerSlot})) \times a_{c_i,j}^i
 \end{aligned} \tag{26}$$

La expresión (26) incorpora a la expresión (25) la componente $a_{c_i,j}^i$ del vector $V_{c_i}^i$. $a_{c_i,j}^i$ corresponde al paquete $pkt_{i,j}$ del flujo F_i dentro de un HP tal que $j = N + 1 - N_{pkt}^i \times \left\lfloor \frac{N}{N_{pkt}^i} \right\rfloor$ (es decir, corresponde al paquete N). $a_{c_i,j}^i$ multiplica a toda la expresión al lado derecho del igual, habilitando ($a_{c_i,j}^i = 1$) o deshabilitando ($a_{c_i,j}^i = 0$) el uso de dicha configuración CG para la transmisión del paquete N . Cuando un UE tenga que transmitir un paquete del flujo TSN F_i , recorrerá las distintas configuraciones $CG_{c_i}^i$, con $c_i = 1, 2, \dots$, que tenga asignadas para

determinar qué configuración CG y, por tanto, qué recursos radio debe utilizar para la transmisión del paquete.

5.2. Heuristic 3GPP-Compliant Flexible configured grAnt Scheduling for TSN traffic (C-FAST)

En este apartado se propone un nuevo esquema de *scheduling* que mantiene el mismo objetivo que H-FAST y O-FAST en busca de la solución que permita soportar todos los flujos TSN evitando las posibles colisiones, pero teniendo en cuenta el número máximo de configuraciones CG que pueden ser asignados a un UE siendo por tanto este esquema compatible con el estándar. El esquema propuesto en este capítulo asigna, siempre que sea posible, recursos radio periódicos para los paquetes de un mismo flujo TSN dentro del hiperperiodo o *HP*, de manera que se pueda atender al máximo número de paquetes posible con una misma configuración CG y así minimizar el número de configuraciones CG necesarias para un único flujo TSN. Es decir, en lugar de crear una nueva configuración CG para cada paquete dentro del *HP* con periodicidad igual a *HP*, el nuevo esquema de *scheduling* intenta reutilizar configuraciones CG ya creadas (con periodicidad menor a *HP*) para la transmisión de paquetes del mismo flujo TSN F_i dentro del *HP* siempre que sea posible. Esta estrategia es factible si se utiliza el vector de activación/desactivación que se ha propuesto en el apartado 5.1, ya que evitará los posibles conflictos entre los recursos radio asignados a distintos flujos TSN. De esta manera, es posible reducir la cantidad total de configuraciones CG necesarias para soportar los flujos TSN, respetando el límite establecido por el 3GPP en [15]. Al nuevo esquema de *scheduling* propuesto se le ha denominado *Heuristic 3GPP-Compliant Flexible configured grAnt Scheduling for TSN traffic (C-FAST)*.

C-FAST está basado en un algoritmo heurístico y, por tanto, es una evolución de H-FAST. En la Figura 28 se muestra un ejemplo para ilustrar el modo de operación de C-FAST y resaltar la principal diferencia con respecto a H-FAST. En este ejemplo, el primer paquete en recibir recursos es el primer paquete del flujo TSN F_2 por tener requisitos de latencia más estrictos. Después, se le asignan recursos radio al primer paquete del flujo TSN F_1 y seguidamente al segundo paquete del flujo F_2 . En el caso de H-FAST, como se puede observar en la Figura 28.a, se crean dos configuraciones CG con periodicidad *HP* para el flujo F_1 . La primera configuración CG indica los recursos radio que se van a utilizar para transmitir el primer paquete y recibe recursos en el 2º símbolo OFDM (siendo $s_{aux}=1$). La segunda configuración CG se utiliza para indicar los recursos radio en los que se va a transmitir el segundo paquete. Este segundo paquete tiene asignados los recursos radio en el 4º símbolo OFDM (con $s_{aux}=3$). Con esta asignación de recursos, H-FAST consigue minimizar la latencia media experimentada por los paquetes. A diferencia de H-FAST, C-

FAST busca reducir el número de configuraciones CG. Por lo tanto, C-FAST crea una única configuración CG para ambos paquetes del flujo TSN F_i dentro del HP . Los recursos radio asignados mediante esta configuración CG se repiten con una periodicidad igual a p_i , que es la periodicidad con la que el tráfico TSN F_i llega al *bridge* 5G virtual, tal y como se muestra en la Figura 28.b.

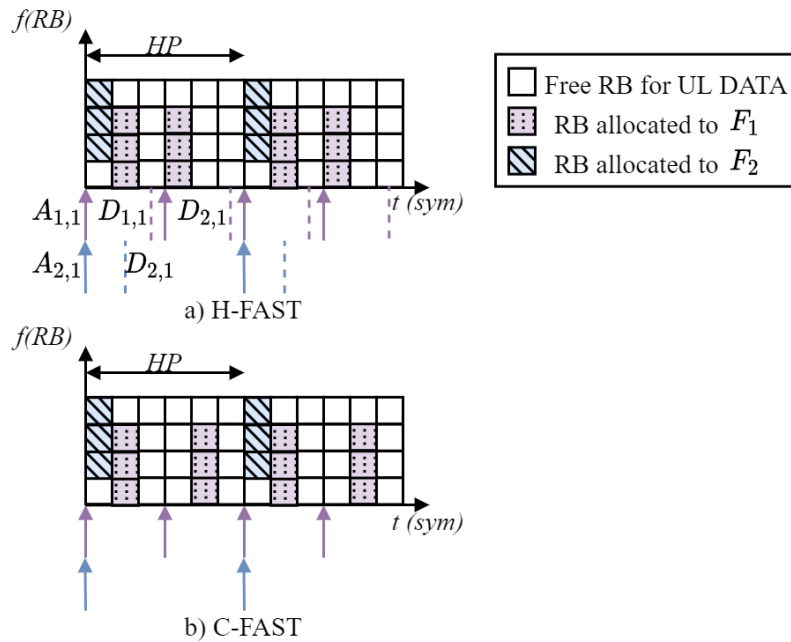


Figura 28. Ejemplo de solución de *scheduling* al aplicar H-FAST y C-FAST.

El funcionamiento de C-FAST se describe en el algoritmo III. En primer lugar, al igual que H-FAST, C-FAST calcula el HP como el mínimo común múltiplo de las periodicidades de los paquetes transmitidos en cada flujo TSN F_i , es decir, $HP = \text{mcm}(p_i)$. A continuación, los paquetes de datos de los distintos flujos TSN que están incluidos en el HP se agrupan en diferentes bloques siguiendo el algoritmo I presentado en el capítulo 4 (línea 3 del algoritmo III). Es importante recordar que cada bloque agrupa los paquetes de distintos flujos TSN cuya transmisión podría solapar en tiempo según su tiempo de llegada y salida a la red 5G. Al igual que H-FAST, C-FAST realiza la asignación de recursos radio a los paquetes contenidos en cada bloque B_z por separado, siendo $z \in [1, 2, \dots, N_B]$.

Para cada bloque B_z , C-FAST comienza calculando el número d_i^S de símbolos OFDM y el número d_i^R de recursos radio que debe asignar a cada paquete del flujo F_i para satisfacer su demanda de recursos radio d_i . El parámetro d_i indica el número de recursos radio que requiere cada paquete en base a la cantidad de datos a transmitir y el MCS a utilizar en la transmisión. El valor de d_i se calcula utilizando la expresión (21) definida en el capítulo 4.

Los valores de d_i^S y d_i^R se calculan igual que en H-FAST. Es decir, C-FAST minimiza el número de símbolos asignados para la transmisión de cada paquete. Por lo tanto, asigna $d_i^R = d_i$ RBs en un símbolo OFDM ($d_i^S = 1$) cuando d_i es inferior al número de RBs disponibles en el ancho de banda (R_{BW}), i.e., cuando $d_i \leq R_{BW}$. Mientras que, si $d_i > R_{BW}$, C-FAST asignará $d_i^R = R_{BW}$ RBs en $d_i^S = \lceil d_i / R_{BW} \rceil$ símbolos OFDM.

Tras calcular d_i^S y d_i^R para cada flujo F_i , C-FAST comienza con la asignación de recursos radio a cada paquete de datos dentro del bloque B_z . Éste es un proceso iterativo (líneas 8-41 del algoritmo III). En cada iteración, el esquema de *scheduling* C-FAST intenta encontrar una solución de *scheduling* en la que todos los paquetes del bloque B_z reciben recursos radio para su transmisión cumpliendo con sus requisitos de latencia l_{5G_i} . En el momento en el que no se encuentren recursos radio que cumplan con los requisitos de latencia de un paquete, se iniciará una nueva iteración buscando una solución alternativa. El proceso iterativo finaliza cuando se alcanza una solución que cumple con los requisitos de todos los paquetes en B_z , o cuando no es posible encontrar una solución factible al problema.

En la primera iteración, C-FAST recorre los flujos TSN según sus requisitos de latencia l_{5G_i} de menor a mayor (línea 10). Para cada flujo TSN F_i , C-FAST asigna recursos radio al primer paquete del flujo. A partir de esta asignación de recursos, C-FAST crea una configuración CG_I^i , reservando los mismos recursos radio de forma periódica con periodicidad igual a $p_{c_i}^i$ (línea 7 del Algoritmo IV). A continuación, se comprueba si el resto de los paquetes del mismo flujo F_i pueden transmitirse utilizando los recursos radio reservados por esta configuración (líneas 22-37 del Algoritmo IV) y rellenando el vector de activación en concordancia (línea 26 del Algoritmo IV). Los paquetes pkt_{ij} del flujo TSN F_i que se puedan transmitir con los recursos radio indicados en la configuración CG_I^i se quitan del conjunto o bloque B_z (línea 35 del Algoritmo IV). Si uno o varios paquetes pkt_{ij} del flujo TSN F_i no se pueden transmitir con la configuración CG_I^i , estos se introducen de nuevo en el conjunto B_z (línea 38 del Algoritmo IV).

Tras recorrer todos los N_F flujos TSN, C-FAST comprueba si el conjunto B_z contiene paquetes que necesiten una nueva configuración CG para que se puedan transmitir por el canal (líneas 19-27 del Algoritmo III). Si este es el caso, C-FAST comienza a asignar recursos radio a cada paquete en función del requisito de latencia l_{5G_i} , es decir, empezando con el paquete con el requisito de latencia más estricto (línea 20 del Algoritmo III). Para este paquete, C-FAST crea una nueva configuración $CG_{c_i}^i$ con c_i^{++} (Algoritmo IV). A continuación, se comprueba si el resto de los paquetes del mismo flujo F_i que están en el

conjunto B_z se pueden transmitir en los recursos radio indicados en esta nueva configuración $CG_{C_i}^i$. En caso afirmativo, se actualiza el vector de activación, y en caso contrario, este paquete se añade otra vez al conjunto B_z para que C-FAST le asigne una nueva configuración. Este proceso se repite hasta que todos los paquetes de todos los flujos TSN incluidos en B_z tengan una configuración CG asignada.

Finalmente, al igual que H-FAST, si un paquete $pkt_{i,j}$ no puede recibir recursos radio (líneas 19-21 del algoritmo IV), C-FAST realiza el proceso de reasignación. Es decir, el *scheduling* propuesto crea una nueva iteración para buscar una nueva solución para todos los paquetes del *HP*. Por este motivo, cada vez que se asigna una configuración $CG_{C_i}^i$ a un paquete, si los recursos radio recibidos están justo antes del tiempo límite $D_{i,j}-t_{gNB,rx}$, este paquete se incluye en el subconjunto BF_z (líneas 8-9 y líneas 27-28 en el Algoritmo IV). Por lo tanto, en el caso de que no sea posible encontrar recursos radio que satisfagan los requisitos de latencia para un paquete $pkt_{i,j}$ (líneas 19-21 del algoritmo IV), C-FAST reestablece el conjunto inicial B_z y de este conjunto inicial, al igual que en H-FAST, elimina los paquetes incluidos en BF_z que mantendrán la asignación de recursos recibida hasta el momento (líneas 33-34 del Algoritmo III). Es decir, C-FAST borra todas las configuraciones y vectores asociados para todos los paquetes que quedan incluidos en B_z , tal que, simplemente deja las configuraciones que mantendrían los paquetes que están en BF_z (líneas 35-36 del Algoritmo III). En consecuencia, los paquetes incluidos en B_z deben volver a recibir recursos radio que satisfagan su demanda y sus requisitos de latencia l_{5G_i} . Para ello, H-FAST comienza una nueva iteración para buscar una solución para todos los paquetes de B_z . En esta nueva iteración, el primer paquete en recibir recursos será el paquete que no pudo recibir recursos radio en la iteración anterior, el cual también pasa al conjunto BF_z . Posteriormente, los paquetes en B_z se atenderán en función del requisito de latencia l_{5G_i} .

ALGORITHM III: SCHEDULING

```

1. Input:  $F_i \forall i \in [1, N_F]$ ,  $t_{sym}$ ,  $t_{gNB,rx}$ ,  $t_{UE,tx}$ 
2.  $HP = \text{lcm}\{p_i\}$  with  $i \in [1, N_F]$ 
3. Create packet blocks  $B_z$ , with  $z=1, 2, \dots$  (Algorithm II)
4. For  $z=1$  to  $N_B$ 
5.   Calculates  $d_i^S$  and  $d_i^R$  for all packets in  $B_z$ 
6.    $N_{pkt}$  = number of packets in  $B_z$ 
7.   Initialize  $iter=0$ ,  $BF_z = \emptyset$ ,  $m = 0$ ,  $c_i=1 \forall i$ 
8.   While  $iter < \infty$ 
9.     While  $m < N_F$ 
10.       $F_i$  = TSN flow with lower  $l_{5G}$ 
11.      First packet  $pkt_{i,1}$  of TSN Flow  $F_i$ 
12.      [ $B_z$ ,  $CG_{c_i}^i$ ,  $N_{pkt}$ ,  $BF_z$ ] = Creates CG (Algorithm IV)
13.      If  $CG_{c_i}^i == \text{NULL}$ 
14.         $iter = iter + 1$ 
15.        Goto line 28
16.      EndIf
17.       $m++$ 
18.    EndWhile
19.    While  $B_z \neq \emptyset$ 
20.       $pkt_{i,j}$  = packet with lower  $l_{5G}$  in  $B_z$ 
21.       $c_i++$  for TSN flow  $F_i$ 
22.      [ $B_z$ ,  $CG_{c_i}^i$ ,  $N_{pkt}$ ,  $BF_z$ ] = Creates CG (Algorithm IV)
23.      If  $CG_{c_i}^i == \text{NULL}$ 
24.         $iter = iter + 1$ 
25.        Goto line 28
26.      EndIf
27.    EndWhile
28.    If  $iter > 0$  &  $CG_{c_i}^i == \text{NULL}$ 
29.       $BF_z = BF_z + \{pkt_{i,j}\}$ 
30.      If number of packets in  $BF_z + B_z == N_{pkt}$ 
31.        Goto line 42
32.      Else
33.        Reinitiate  $B_z$  and  $N_{pkt}$ 
34.         $B_z = B_z - BF_z$ 
35.        Free radio resources allocated to packets in  $B_z$ 
36.        Update  $CG_{c_i}^i$  for all the TSN flows and erase all the  $CG_{c_i}^i$  where  $V_c^i = \emptyset$ 
37.        Goto line 21
38.      Else
39.        Goto line 4
40.    EndIf
41.  EndWhile
42. End For

```

ALGORITHM IV: NEW CG

```

1. Input:  $F_i, pkt_{i,j}, c_i, t_{sym}, t_{gNB,rx}, t_{UE,tx}, BF_z, N_{pkt}$ 
2.  $B_{aux} = \emptyset$ 
3.  $s_{aux}$  = first OFDM symbol after  $A_{i,j} + t_{UE,tx}$ 
4. While  $(s_{aux} + d_i^S - 1) \cdot t_{sym} \leq D_{i,j} - t_{gNB,rx}$ 
5.   If there are  $d_i^R$  unallocated RBs in symbols  $s_{aux}$  until  $s_{aux} + d_i^S - 1$ 
6.      $pkt_{i,j}$  receives  $d_i^R$  RBs in symbols  $s_{aux}$  until  $s_{aux} + d_i^S - 1$ 
7.     Create  $CG_{c_i}^i = \{s_{c_i}^i, p_{c_i}^i, V_c^i\}$  with  $s_c^i = s_{aux} - p_i \cdot (j-1)$ ,  $p_c^i = p_i$ ,  $a_{c,j}^i = 1$ , and  $a_{c,m}^i = 0 \forall m \neq j$ 
8.     If  $(s_{aux} + d_i^S - 1) \cdot t_{sym} == D_{i,j} - t_{gNB,rx}$ 
9.        $BF_z = BF_z + \{pkt_{i,j}\}$ 
10.    Else
11.       $N_{pkt} = N_{pkt} - 1$ 
12.    EndIf
13.     $B_z = B_z - \{pkt_{i,j}\}$ 
14.    Goto line 19
15.  Else
16.     $s_{aux}++$ 
17.  EndIf
18. EndWhile
19. If  $(s_{aux} + d_i^S - 1) \cdot t_{sym} > D_{i,j} - t_{gNB,rx}$ 
20.   New Iteration (Goto line 39)
21. EndIf
22.  $pkt_{i,j}$  = packet from  $F_i$  with lower  $l_{5G}$  in  $B_z$ 
23. While  $pkt_{i,j} \neq \text{NULL}$ 
24.    $s_{aux} = s_{c_i}^i + (j-1) \cdot p_{c_i}^i$ 
25.   If there are  $d_i^R$  unallocated RBs in symbols  $s_{aux}$  until  $s_{aux} + d_i^S - 1$ 
26.     Update  $V_c^i$  with  $a_{c,j}^i = 1$ 
27.     If  $(s_{aux} + d_i^S - 1) \cdot t_{sym} == D_{i,j} - t_{gNB,rx}$ 
28.        $BF_z = BF_z + \{pkt_{i,j}\}$ 
29.     Else
30.        $N_{pkt} = N_{pkt} - 1$ 
31.     EndIf
32.   Else
33.      $B_{aux} = B_{aux} + \{pkt_{i,j}\}$ 
34.   End If
35.    $B_z = B_z - \{pkt_{i,j}\}$ 
36.    $pkt_{i,j}$  = packet from  $F_i$  with lower  $l_{5G}$  in  $B_z$ 
37. EndWhile
38.  $B_z = B_z + B_{aux}$ 
39. Output:  $B_z, \text{New } CG_{c_i}^i, BF_z, N_{pkt}$ 

```

5.3. Evaluación del rendimiento

En este apartado se compara el rendimiento obtenido con el *scheduler* C-FAST propuesto en este capítulo frente al *scheduler* H-FAST que se ha propuesto en el capítulo 4. La evaluación realizada ha demostrado que los esquemas de *scheduling* C-FAST y H-

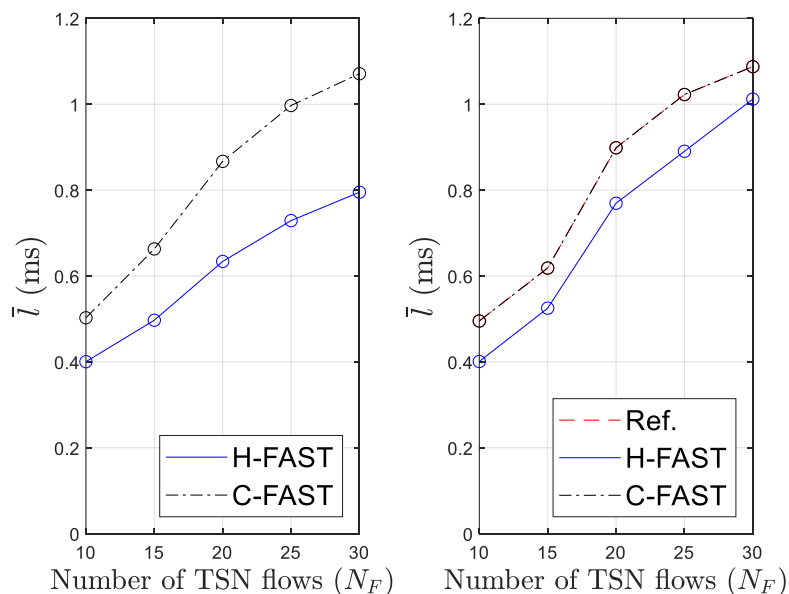
FAST son capaces de resolver el mismo número de problemas de *scheduling* cumpliendo con los requisitos de latencia de todos los paquetes de datos a transmitir de los distintos flujos TSN. Un aspecto crítico que marca la implementabilidad de estas propuestas es si las soluciones de *scheduling* alcanzadas con H-FAST y C-FAST cumplen con el estándar de 3GPP, en concreto, si cumplen con la restricción del número máximo de configuraciones CG por UE que establece el 3GPP (límite que se establece en 12). En este contexto, la Tabla 7 muestra, por un lado, el número máximo y medio de configuraciones CG que se crean por flujo TSN utilizando los diferentes esquemas de *scheduling*, y, por otro lado, el porcentaje de UEs que no cumple con el límite establecido por 3GPP. Los resultados en la Tabla 7 muestran que C-FAST cumple con el límite superior de configuraciones CG impuesto por el 3GPP en todos los escenarios evaluados. Sin embargo, aunque el número de configuraciones CG promedio que crea H-FAST por UE cumple con el límite establecido por el 3GPP en todos los escenarios evaluados, los resultados muestran que hasta el 23% de los UEs reciben un número de configuraciones CG por encima del límite cuando se transmiten $N_F=25$ flujos TSN. Es decir, alrededor de 6 de los 25 UEs que están transmitiendo flujos TSN, reciben más de 12 configuraciones CG. Además, la Tabla 7 muestra que el número máximo de configuraciones CG asignadas a un UE aumenta hasta 72, estando muy por encima del límite (12 CGs). Estos resultados han mostrado que H-FAST no garantiza el cumplimiento del límite establecido por 3GPP en el número de configuraciones CG asignadas a un UE, mientras C-FAST lo cumple en todos los casos.

Tabla 7. Número de CGs y % de UEs para los esquemas de *scheduling* H-FAST y C-FAST.

	Esquema de <i>scheduling</i>	Flujos TSN				
		10	15	20	25	30
Número máximo de CGs asignadas a un UE	H-FAST	72	72	72	72	72
	C-FAST	3	4	5	6	7
Número medio de CGs por UE	H-FAST	7	9.7	9.6	10.61	9.84
	C-FAST	1	1.18	1.28	1.45	1.52
% de UEs con más de 12 CGs	H-FAST	15.45	21.41	19.79	23.29	16.84
	C-FAST	0	0	0	0	0

A continuación, se realiza el análisis de la latencia experimentada. Según el Anexo B, el requisito de latencia l_{5G_i} más restrictivo que los flujos TSN deben soportar dentro del *bridge* 5G virtual es de 2.27 ms. La Figura 29.a representa la latencia media experimentada

al utilizar los esquemas H-FAST y C-FAST. La Figura 29.b muestra una comparación de la latencia media entre los dos *schedulers* propuestos y el esquema de referencia; en la Figura 29.b solo se consideran los problemas de *scheduling* para los que el esquema de referencia alcanza una solución sin conflicto para hacer una comparación justa entre las distintas propuestas. Los resultados en la Figura 29.a muestran que la latencia media experimentada con C-FAST aumenta ligeramente con respecto a la latencia experimentada con H-FAST, debido a que C-FAST utiliza menos configuraciones CG que H-FAST para evitar los conflictos que se dan con el esquema de referencia. Sin embargo, es importante destacar que el requisito de latencia en caso de C-FAST sigue cumpliéndose para todos los paquetes lo cual es lo que exige el servicio. Además, la latencia media experimentada por los paquetes con C-FAST es 52% menor que el requisito de latencia máxima más restrictivo. Como ya se comentó, tener este margen entre la latencia experimentada y el requisito de latencia máximo es importante ante la posible necesidad de realizar retransmisiones de paquetes para aumentar la fiabilidad de las comunicaciones. Los resultados en la Figura 29.b muestran que el *scheduler* C-FAST y el esquema de referencia obtienen la misma latencia. Este resultado era previsible ya que, si el esquema de referencia tiene una solución satisfactoria, entonces C-FAST utilizará únicamente una configuración CG para cada flujo TSN, logrando así la misma latencia que con el esquema de referencia.



a) Latencia media experimentada considerando todos los problemas de *scheduling* evaluados.

b) Latencia media experimentada considerando el subconjunto de problemas de *scheduling* evaluados en los que no hay colisiones con la técnica de referencia.

Figura 29. Latencia media experimentada con C-FAST, H-FAST y el esquema de referencia en función del número de flujos TSN evaluados.

La Figura 30 muestra la desviación estándar de la latencia experimentada en función del número de flujos TSN en la red 5G para H-FAST y C-FAST. El objetivo de C-FAST es reducir el número de configuraciones CG creadas para cumplir con el límite establecido en el estándar de 3GPP. En consecuencia, la desviación estándar de la latencia experimentada por los paquetes de cada flujo TSN disminuye a expensas de un aumento en la latencia de la misma (Figura 29). Esto es debido a que C-FAST prioriza minimizar el número de configuraciones CG a minimizar la latencia mientras se cumpla el requisito demandado. C-FAST consigue disminuir más de un 33% la desviación estándar media para todos los casos.

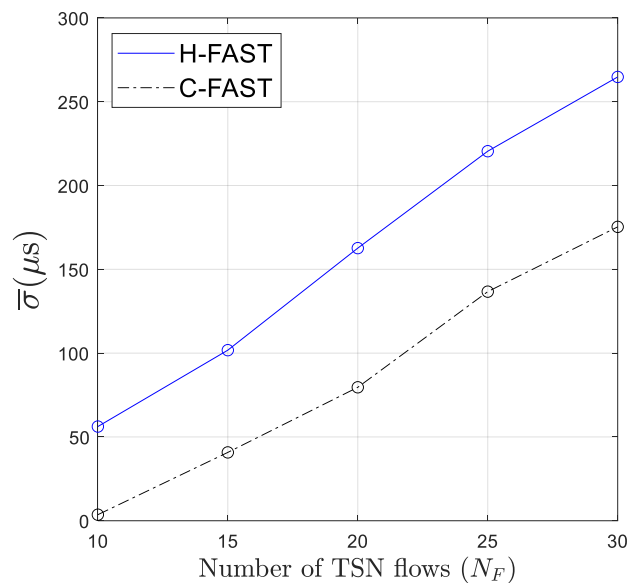


Figura 30. Desviación típica media de la latencia cuando se utilizan los *schedulers* H-FAST y C-FAST.

La Figura 31 muestra la media del tiempo de computación necesario para cada *scheduling* para resolver los problemas de asignación de recursos. En la Figura 31 se observa que, aunque C-FAST presenta un coste computacional ligeramente mayor que H-FAST, este tiempo sigue siendo bajo. Por ejemplo, en el escenario en el que se transmiten 10 flujos TSN, C-FAST alcanza la solución al problema de *scheduling* en tan solo 55 ms. A medida que aumenta el número de flujos, el tiempo de computación medio también aumenta. En el caso de C-FAST, cuando se transmiten 30 flujos TSN, el tiempo de computación medio llega a 181 ms. Al igual que se menciona en el capítulo 4, el esquema de referencia es el que menor coste computacional presenta, pero su rendimiento es inferior en términos de problemas con soluciones satisfactorias. Aunque C-FAST requiere un tiempo de computación ligeramente mayor que H-FAST, éste sigue siendo bajo, y además tienen el mismo rendimiento a la hora de buscar solución a los problemas de *scheduling*. Asimismo, C-FAST requiere un menor número de configuraciones CG, y es el único que

cumple con los requisitos del 3GPP en cuanto al número máximo de configuraciones CG creadas para cada UE.

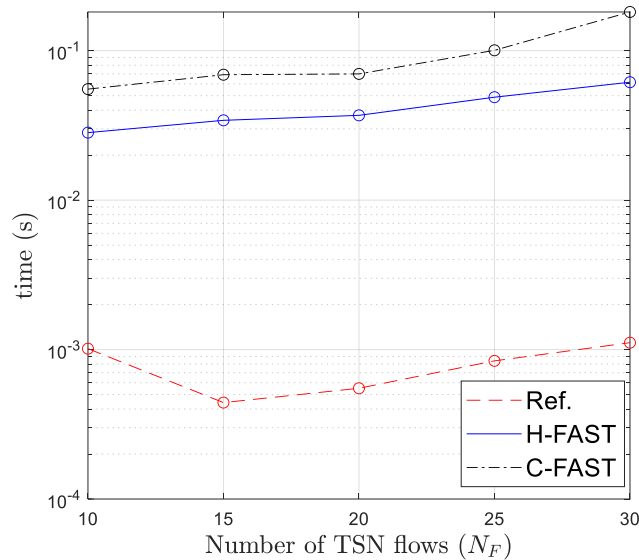


Figura 31. Tiempo computacional promedio en segundos para asignar recursos en función de los tres planificadores.

5.4. Conclusiones

Los esquemas de *scheduling* propuestos en el capítulo anterior, O-FAST y H-FAST, buscan alcanzar la mejor solución posible, sin tener en cuenta la limitación impuesta en el estándar de 5G en cuanto al número de configuraciones CG que pueden ser asignados a un UE en un BWP (12 configuraciones CG por UE) [15]. Para abordar esta problemática, en este capítulo se ha diseñado y se ha evaluado un nuevo esquema de *scheduling* llamado C-FAST, que evoluciona el esquema H-FAST para hacerlo compatible con el estándar actual de 3GPP y, por tanto, pueda ser implementable. Para ello, C-FAST se basa en la activación y desactivación de configuraciones CG asignadas a los UEs. El UE activa las configuraciones únicamente en los momentos en los que lo vaya a utilizar y así evitar conflictos entre las asignaciones de diferentes UEs. Estas configuraciones CG se definen con una periodicidad menor a *HP*, por lo que una misma configuración CG se puede utilizar para transmitir varios paquetes del mismo flujo TSN dentro del *HP*, consiguiendo así reducir el número de configuraciones CG necesarias. Para evitar el incremento de la señalización de control entre el gNB y el UE para activar y desactivar las diferentes configuraciones CG y los posibles retardos asociados, en este capítulo se ha propuesto el uso de un vector de activación. Se propone incluir este vector en la fórmula definida en el

estándar del 3GPP [16] que el UE utiliza para identificar la configuración activa y los recursos radio en los que va a transmitir el paquete.

Los resultados obtenidos en este capítulo demuestran que C-FAST puede satisfacer el mismo número de flujos que H-FAST, pero cumpliendo con el estándar de 3GPP, mientras que H-FAST supera el límite impuesto. Mediante esta modificación, el número máximo de configuraciones CG que se crean con el *scheduler* C-FAST para un mismo número de flujos TSN se reduce significativamente desde 72 hasta 7 en comparación con H-FAST. Además, los resultados muestran que cuando se transmiten $N_F=25$ flujos TSN, el 23% de los UEs recibe más de 12 configuraciones CG, por lo tanto, 6 de los 25 UEs que están transmitiendo flujos TSN, no cumplen con el requisito que establece el estándar de 3GPP. Es decir, los resultados muestran que el uso de este vector en combinación con C-FAST permite que varios UEs que transmiten flujos TSN de distintas periodicidades puedan acceder a los recursos radio de manera coordinada, evitando conflictos entre sí y manteniendo el mismo rendimiento que H-FAST, pero reduciendo el número de configuraciones CG necesario para ello, y pudiendo ser implementable según el requisito de 3GPP.

También se debe destacar la importancia de la latencia en redes TSN. La prioridad radica en asegurarse de que los paquetes sean entregados al puerto de salida del *bridge* 5G virtual dentro del tiempo establecido por el *scheduler* de TSN, para que puedan ser transmitidos al siguiente *bridge* de la red en la ruta del flujo TSN según el *scheduling* realizado por la red TSN. C-FAST continúa cumpliendo la latencia media con un margen del 52% en comparación con la latencia máxima más restrictiva. Además, como era de esperar, C-FAST logra reducir la desviación típica en más de un 33% en comparación con H-FAST. Finalmente, a pesar de que C-FAST requiere un tiempo de computación ligeramente mayor que H-FAST, este tiempo computacional sigue siendo bajo, y, además, es el único *scheduling* propuesto que es implementable para todos los casos evaluados en base al requisito de número de configuraciones CG establecido por el 3GPP.

6. Implementación de *Configured Grant* en 5G-LENA

En esta tesis se han propuesto esquemas de *scheduling* para redes 5G NR integradas con redes TSN para dar soporte a entornos críticos industriales. Las propuestas realizadas se han evaluado en un entorno de simulación desarrollado en Matlab que simula las características del sistema y de la red necesarias para realizar una correcta evaluación del rendimiento, validez y beneficio de los esquemas de *scheduling* propuestos frente a esquemas disponibles en la literatura. Tras demostrar el rendimiento y la ganancia obtenida con las propuestas realizadas en términos de número de flujos TSN soportados en la red 5G-TSN, el siguiente paso en este estudio sería preparar un simulador de redes más avanzado para poder utilizar los esquemas propuestos en esta tesis en futuros trabajos e investigaciones realizadas en Ikerlan en las que se hará uso de una plataforma que simule los distintos protocolos de la red 5G.

Son varios los simuladores de 5G NR utilizados en la comunidad investigadora. Entre los simuladores de código abierto, los más utilizados son OMNeT++ Simu5G [69] y ns-3 5G-LENA [9] por ser estos los que implementan de manera más completa las funcionalidades de 5G NR. Aunque la implementación de los *scheduling* propuestos en esta tesis en simuladores de sistemas avanzados queda fuera de los objetivos de la misma, se ha llevado a cabo la preparación del simulador de 5G NR para que en un futuro se pueda realizar dicha implementación. Se ha optado por el uso de 5G-LENA frente a Simu5G porque era el único simulador de 5G NR *end-to-end* disponible cuando se comenzó este estudio. Sin embargo, 5G-LENA no implementaba el *scheduler Configured Grant* o CG, solo implementaba el *Dynamic scheduling*.

La implementación de CG en simuladores de sistemas avanzados como 5G-LENA es un desafío complejo. Además de la dificultad del código, requiere un conocimiento

profundo de los estándares de 5G NR para entender la implementación de los distintos procesos y funciones. Teniendo en cuenta el conocimiento adquirido durante el desarrollo de esta tesis sobre los estándares de 5G NR y el funcionamiento de los distintos tipos de *scheduling*, se decidió implementar el *scheduling* CG en 5G-LENA como parte de la tesis.

Otro resultado clave de este objetivo ha sido la implementación de la técnica de acceso múltiple OFDMA en 5G-LENA. Tal y como se explica en este capítulo, 5G-LENA implementa distintas técnicas de acceso múltiple. La primera de ellas es TDMA (*Time Division Multiplexing Access*) que permite el acceso de varios UEs en distintos símbolos OFDM, es decir, cada UE recibe todos los RBs disponibles en un símbolo OFDM; en esta tesis, a la implementación de TDMA en 5G-LENA se le ha llamado 5GL-TDMA. 5G-LENA implementa una segunda técnica de acceso múltiple que denomina OFDMA, aunque ésta no implementa la flexibilidad real en el uso de los recursos radio que OFDMA ofrece en un sistema 5G real. Para diferenciar la técnica de acceso múltiple OFDMA 5G NR de la versión implementada en 5G-LENA, en esta tesis, a esta última técnica de acceso múltiple se le llama como 5GL-OFDMA. Mientras que OFDMA 5G NR permite asignar a cada UE un número flexible de símbolos OFDM y RBs, 5GL-OFDMA asigna al mismo UE por lo menos un RB en todos los símbolos OFDM de un mismo *slot*. 5GL-OFDMA limita por tanto la flexibilidad ofrecida por 5G NR y es necesario evolucionar el simulador 5G-LENA para poder emular la verdadera flexibilidad que ofrece 5G NR en el uso de los recursos radio, lo cual es crítico para dar soporte a servicios que requieren bajas latencias.

Este capítulo describe la implementación de la técnica de acceso múltiple OFDMA y el *scheduling* CG realizada en 5G-LENA. En primer lugar, el apartado 6.1 presenta el simulador de red 5G-LENA. La implementación de CG y OFDMA en 5G-LENA se describe en el apartado 6.2. Además, en el apartado 6.2 se presentan dos políticas de *scheduling* definidas para explotar la flexibilidad ofrecida por OFDMA. En el apartado 6.3 se indica el escenario en el que se va a evaluar el *scheduling* CG y el modo de acceso OFDMA, mientras que en el apartado 6.4 se realiza la validación de dichas implementaciones. Para ello, se ha comparado las latencias obtenidas a través del simulador con resultados de latencias obtenidas por un lado a través de estudios analíticos disponibles en la literatura y por otro lado a través de expresiones analíticas formuladas en esta tesis. En el apartado 6.5 se evalúa el uso de las políticas de *scheduling* propuestas que explotan la flexibilidad de OFDMA. Los resultados obtenidos mediante simulación demuestran que la nueva implementación de OFDMA ofrece mejoras significativas en términos de latencia y eficiencia a la hora de asignar los recursos radio en comparación con 5GL-OFDMA. Finalmente, en el apartado 6.6 se resumen las conclusiones obtenidas del trabajo presentado en este capítulo. En el apartado 6.6. también se proporciona una lista de otros resultados y detalles que no se han incluido en este capítulo y que se pueden encontrar en nuestras publicaciones de conferencia [20] y artículo de revista [21], las cuales están adjuntas en el

Anexo A. El código de CG junto con el modo de acceso OFDMA implementados en 5G-LENA están disponible en [18] y utiliza la versión de ns-3 disponible en [19].

6.1. Simulador de red 5G-LENA

5G-LENA es un simulador a nivel de sistema de 5G NR de código abierto. 5G-LENA se ha desarrollado como un módulo de ns-3 [70], un entorno de simulación de comunicaciones de red de código abierto en C++. 5G-LENA es la evolución de LENA, que simula la red de 4ª generación LTE. 5G-LENA se ha centrado en la implementación de la tecnología de acceso radio 5G NR y puede realizar simulaciones *end-to-end* utilizando la red central o *core network* de LTE (EPC o *Evolved Packet Core*). 5G-LENA implementa la mayoría de las funcionalidades principales de las capas física (PHY) y MAC de NR de acuerdo con las especificaciones del 3GPP [15]. 5G-LENA está calibrado tanto en escenarios interiores [9], como en escenarios exteriores referenciados en 3GPP [71].

5G-LENA implementa dos clases principales en C++ para simular las funcionalidades del gNB y el UE: las clases NrGnb y NrUe (ver la Figura 32). Para el gNB y el UE, 5G-LENA modela las distintas capas de la pila de protocolos: PHY, MAC, *Radio Link Control* (RLC), *Packet Data Convergence Protocol* (PDCP) y RRC. Las capas superiores RLC, PDCP y RRC se basan en la implementación de LTE LENA, mientras que las capas PHY y MAC implementan las principales características de 5G NR. Entre otras cosas, 5G-LENA simula la estructura de trama flexible definida en 5G NR [15]. Además, considera el uso de distintas numerologías, en concreto, las numerologías 0, 1, 2, 3 y 4, tal y como se establece en TS 38.300 [15], y los esquemas de modulación y de codificación o MCSs hasta 256-QAM definidos en la tabla 1 y 2 de [33]. El simulador permite la configuración flexible de los modos de duplexación TDD y FDD. Además, permite configurar distintos símbolos OFDM dentro de un *slot* para transmisiones en UL y en DL. 5G-LENA implementa

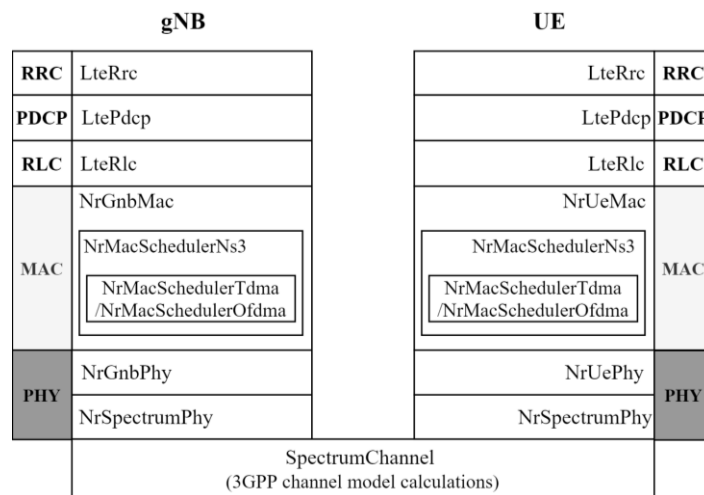


Figura 32. Representación lógica de la clase RAN de 5G-LENA.

diferentes modelos de canal y de propagación definidos en TR 38.901 [72]. 5G-LENA proporciona una implementación realista de los tiempos de *scheduling* k_1 y k_2 definidos para 3GPP 5G NR. k_1 corresponde al tiempo entre la recepción de datos en el UE y la transmisión correspondiente de HARQ-ACK al gNB, mientras que k_2 es el tiempo entre la recepción del UL *grant* en el UE, y la correspondiente transmisión de datos al gNB [73].

6.1.1. Acceso Múltiple

5G-LENA también ha evolucionado la capa MAC para admitir 5G NR. 5G-LENA implementa tres esquemas de acceso múltiple: 5GL-TDMA, 5GL-OFMDA y 5GL-OFDMA con TTI variable [74]. La Figura 33 muestra cómo se asignan los recursos radio a varios UEs cuando se utilizan diferentes esquemas de acceso múltiple. Esta figura representa el *resource grid* de tiempo y frecuencia para un *slot* de 14 símbolos OFDM en el dominio del tiempo y R_{BW} (el número de RBs disponibles en un ancho de banda) RBs en el dominio de la frecuencia. Los cuadrados en la Figura 33 representan un símbolo OFDM en el dominio del tiempo y un RB en el dominio de la frecuencia. El primer y último símbolo están reservados para los canales de control DL y UL, respectivamente. En este ejemplo, cada UE necesita 4 RBs para transmitir su paquete. Los cuadrados coloreados son los RBs que se han asignado a los UEs para la transmisión de paquetes.

La Figura 33.a muestra un ejemplo del esquema de acceso múltiple 5GL-TDMA. Con 5GL-TDMA, los UEs acceden a los recursos radio en diferentes símbolos OFDM, y todos los RBs de un mismo símbolo OFDM deben asignarse al mismo UE. Es decir, 5GL-TDMA asigna R_{BW} RBs de un mismo símbolo OFDM a un único UE, aunque dicho UE solo requiera 4 RBs para transmitir su paquete. 5G-LENA también implementa el esquema de acceso múltiple 5GL-OFDMA (ver Figura 33.b), el cual, es una versión restringida de 5G NR OFDMA. 5GL-OFDMA asigna un RB de cada símbolo OFDM que hay en un *slot* a un mismo UE. 5G-LENA dispone de una versión modificada de 5GL-OFDMA conocida como 5GL-OFDMA con TTI variable (Figura 33.c). Este esquema permite dividir los símbolos OFDM dentro de un *slot* en dos o más segmentos con un número diferente o igual de símbolos OFDM. Los recursos radio dentro de cada uno de estos segmentos pueden ser accedidos por los UEs que están asignados al mismo haz o *beam* de la antena. En cada uno de estos segmentos, se aplica el esquema de acceso múltiple 5GL-OFDMA. Ninguno de estos esquemas de acceso múltiple modela con precisión la flexibilidad ofrecida por OFDMA en 5G NR, ya que OFDMA permite asignar diferente número de RBs y símbolos OFDM a los UEs (ver Figura 33.d).

Los esquemas 5GL-TDMA y 5GL-OFDMA están limitados a asignar un número de recursos radio que sea múltiplo de R_{BW} o de S_{slot} (el número de símbolos OFDM reservados para la transmisión de datos en UL dentro de un *slot*), respectivamente. En el caso de 5GL-OFDMA con TTI variable, el número de recursos radio asignados a los UEs es múltiplo del número de símbolos OFDM dentro de un segmento. Sin embargo, OFDMA en 5G NR permite ajustar de manera más precisa el número de recursos asignados a la demanda de los UEs, en consecuencia, realiza un uso más eficiente de los recursos radio (ver Figura 33.d).

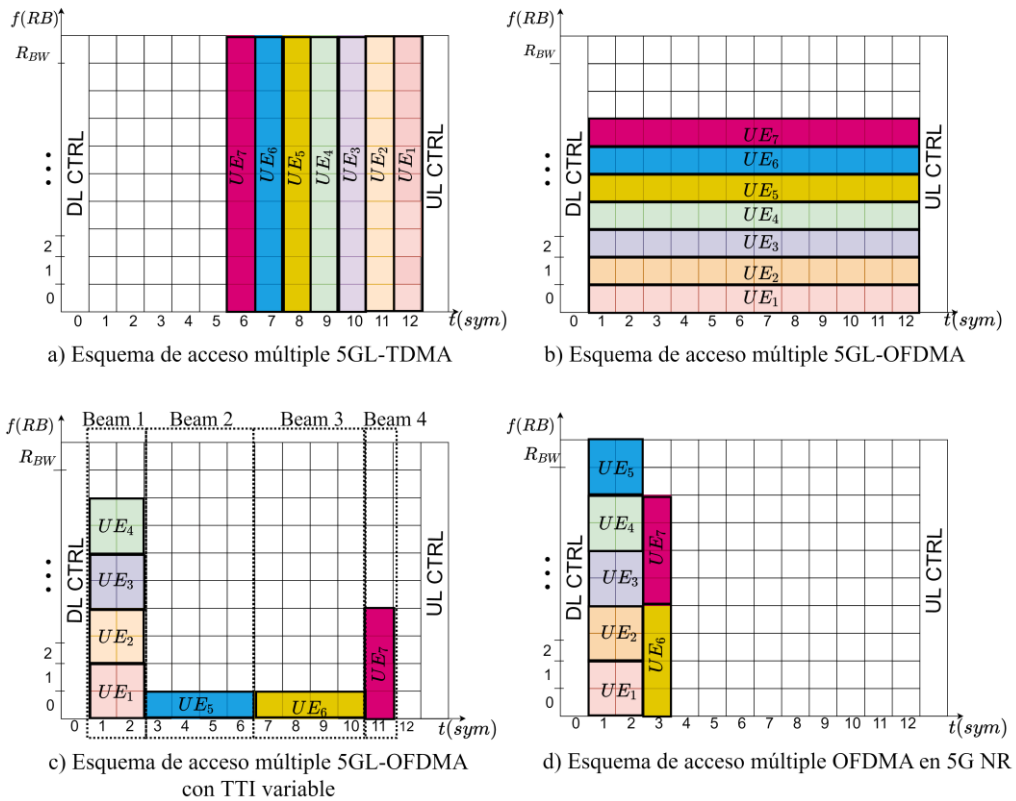


Figura 33. Ejemplo de asignación de recursos radio de los distintos modos de acceso múltiple.

6.1.2. Scheduling

5G-LENA implementa el *scheduling* dinámico tanto en DL como en UL, en cambio, 5G-LENA no disponía de la implementación del *scheduling* semi-estático. Las políticas de *scheduling* implementadas en 5G-LENA son: *round-robin*, *proporcional fair* y *max rate*. Es importante destacar que un UE recibe r RBs en s símbolos OFDM consecutivos, es decir, un UE recibe $r \times s$ recursos radio¹⁰. Debido a las limitaciones del modelo de interferencia implementado en 5G-LENA, si dos UEs reciben RBs en el mismo símbolo

¹⁰ En los capítulos 4 y 5, r RBs para un UE se define como d_i^R RBs para un flujo TSN F_i y s símbolos OFDM para un UE se define como d_i^S símbolos OFDM para un flujo TSN F_i .

OFDM, los dos UEs deben recibir RBs en el mismo número s de símbolos OFDM. Finalmente, 5G-LENA asigna recursos radio para transmisiones UL desde el último símbolo OFDM hasta el primer símbolo OFDM que se haya establecido para transmisiones UL, dentro de un *slot*.

6.2. Implementación de *Configured Grant*

Como se ha presentado en los apartados anteriores, 5G-LENA no implementa el *scheduling* semi-estático, lo que limita su aplicabilidad en estudios de casos de uso de servicios críticos. En este contexto, en el marco de esta tesis se ha evolucionado el simulador 5G-LENA para simular transmisiones en UL utilizando el *scheduler* o mecanismo de *Configured Grant* o CG. En concreto, se ha implementado CG de Tipo 1 que crea la configuración CG o el UL *grant* durante el establecimiento de la sesión. El CG tipo 1 elimina posibles retrasos que podrían introducirse durante la activación y desactivación de las configuraciones CG, por lo que es más adecuado para servicios con requisitos de latencia estrictos en los que se centra esta tesis.

Se han añadido nuevas funcionalidades en el UE y en el gNB para llevar a cabo la implementación del *scheduler* CG. Por un lado, se ha implementado una máquina de estados en el UE para gestionar el proceso de petición de recursos radio desde el UE al gNB, así como la configuración de dichos recursos en el UE. Por otro lado, se ha añadido una nueva función en el gNB para gestionar las peticiones de las configuraciones CG y asignar los recursos radio a cada UE de forma eficiente (ver apartado 6.2.1). También se ha introducido una nueva estructura de mensaje para almacenar la información relevante como los recursos radio preasignados y la periodicidad con la que estos recursos se repiten. El rendimiento de la latencia que se puede alcanzar con el CG depende en gran medida del esquema de acceso múltiple ya que establece cómo se pueden compartir los recursos radio entre los UEs; es decir, determina la flexibilidad del *scheduler* para asignar recursos radio a los UEs (ver apartado 6.2.2). Por este motivo, también se ha implementado la técnica de acceso múltiple OFDMA utilizada en 5G NR en 5G-LENA. Para validar la implementación de CG y analizar la capacidad de 5G NR para admitir servicios industriales críticos, se han implementado dos nuevas políticas de *scheduling* que asignan los recursos radio de manera eficiente y, en consecuencia, minimizan la latencia experimentada por los UEs (apartado 6.2.3).

6.2.1. *Configured Grant*

El *scheduler* CG se ha implementado en 5G-LENA utilizando una máquina de estados que se presenta en la Figura 34. Esta máquina de estados se ha implementado en el UE para gestionar las transmisiones en UL. Al comienzo de la simulación, los UEs se encuentran en el estado *INACTIVE_CG*. Un UE cambia al estado *TO_SEND_TrafficInfo* cuando tiene preparado el mensaje que enviará al gNB con información sobre el tamaño y la periodicidad de los paquetes de datos a transmitir. Cuando el UE envía el mensaje al gNB, el estado cambia a *TO_RECEIVE_CG*. El UE permanece en este estado hasta recibir la configuración CG enviada por el gNB. En ese momento, el UE almacena la configuración CG y cambia al estado *ACTIVE_CG*. Cuando se genera un paquete de datos, el UE comprueba si se encuentra en estado *ACTIVE_CG*. Si este es el caso, significa que el UE ya dispone de una configuración CG por lo que el estado del UE cambia a *SCH_CG_DATA*. En este estado, el UE prepara el paquete de datos para que pueda transmitirlo en los recursos radio indicados en la configuración CG. Tras preparar el paquete, el UE vuelve a cambiar al estado *ACTIVE_CG*. Después, el paquete se transmite al gNB en los símbolos OFDM y en los RBs preasignados a través de la configuración CG.

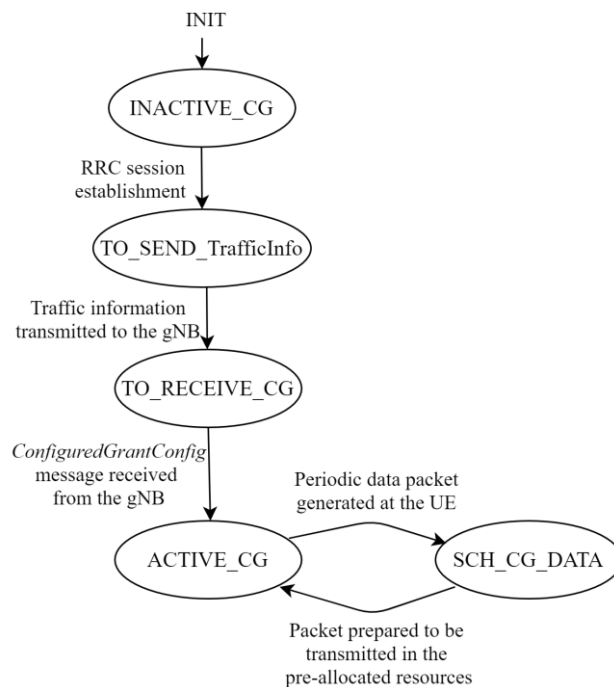


Figura 34. Máquina de estados de CG implementado en el UE MAC.

El gNB es el encargado de decidir los recursos radio que debe asignar a cada UE. Tras recibir el mensaje enviado por un UE solicitando un *configured UL grant* o una configuración CG, el gNB decide los recursos radio que debe asignar al UE en base a una

política de *scheduling*. Este proceso se realiza dentro de la clase `NrMacSchedulerNs3` en la función `DoScheduleUIData` definida en el `gNB` y que se presenta en la Figura 35. La función `AssignULRBG` decide el número de símbolos OFDM y RBs que se deben asignar a cada UE y la función `CreateUICGConfig` genera el mensaje `ConfiguredGrantConfig` con la información sobre los recursos radio preasignados. El mensaje `ConfiguredGrantConfig` es una nueva estructura que almacena la periodicidad de la configuración CG, el tamaño del paquete, el MCS y los recursos que se han asignado al UE correspondiente.

Es importante destacar que la implementación del CG en 5G-LENA también ha implicado la ampliación de varias funciones en la clase `NrPhy` para permitir el *scheduling* periódico. Por ejemplo, la función `StartSlot` de la clase UE PHY se ha extendido para almacenar el mensaje de `ConfiguredGrantConfig` en dicha capa de manera periódica.

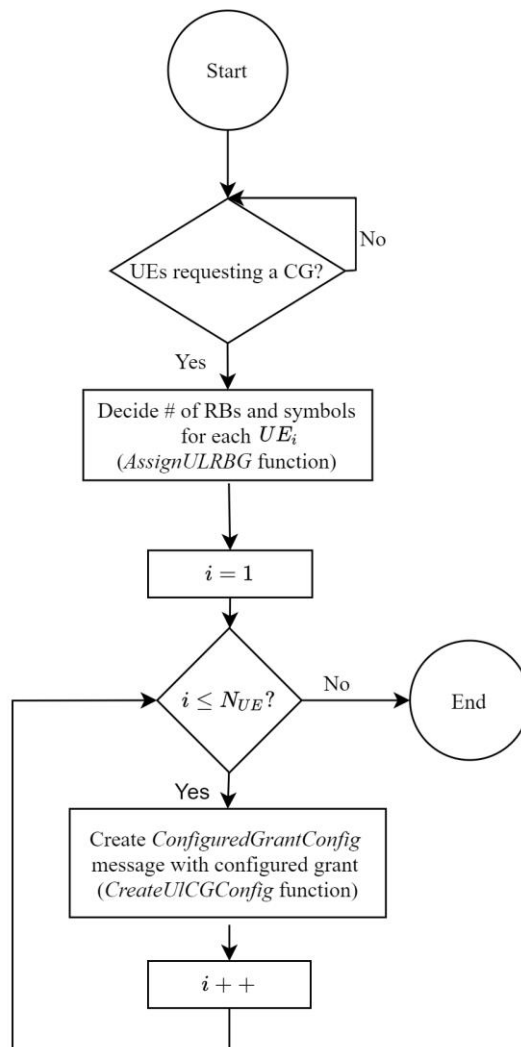


Figura 35. Proceso de asignación recursos radio para transmisiones en UL (función `DoScheduleUIData`) en 5G-LENA.

6.2.2. Acceso múltiple flexible

El rendimiento de latencia alcanzable con el *scheduler* CG depende en gran medida de la técnica de acceso múltiple utilizada. La técnica de acceso múltiple determina cómo los UEs acceden a los recursos radio y define la flexibilidad que tiene el *scheduler* para asignar recursos a los UEs. Por lo tanto, es necesario modelar con precisión el esquema de acceso múltiple OFDMA utilizado en 5G NR para analizar la capacidad real de 5G NR al utilizar CG. Como se ha presentado en este capítulo, ninguno de los esquemas de acceso múltiple implementados en 5G-LENA simulan las verdaderas capacidades en términos de flexibilidad de OFDMA en 5G NR (Figura 33), lo cual puede conllevar a un uso ineficiente de los recursos radio y aumentar las latencias. En consecuencia, puede ser perjudicial para servicios críticos.

En este contexto, en esta tesis se ha extendido la capa MAC de 5G-LENA para implementar OFDMA en 5G NR. El nuevo OFDMA implementado en 5G-LENA permite asignar r RBs en s símbolos OFDM consecutivos a un UE, donde r y s son números enteros en el intervalo $[1, R_{BW}]$ y $[1, S_{slot}]$, respectivamente. R_{BW} representa el número de RBs disponibles y S_{slot} es el número de símbolos OFDM en un *slot* reservados para transmisiones en UL. OFDMA se ha implementado en la clase *NrMacSchedulerOfdma* (ver Figura 32). Más detalles sobre la implementación de OFDMA en 5G-LENA está disponible en el artículo [21] (incluido en el Anexo A).

6.2.3. Scheduling

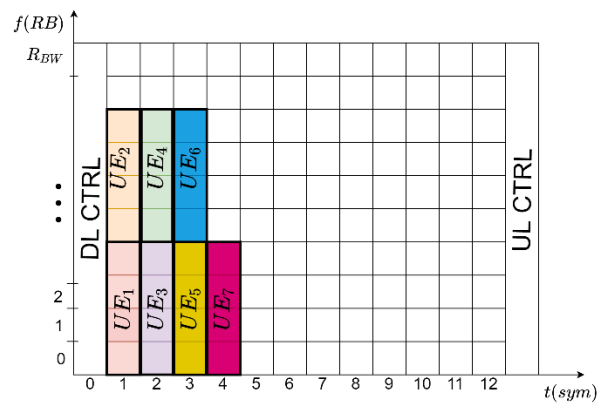
El *scheduling* decide los recursos radio (número de símbolos OFDM y RBs) a asignar a cada UE, y esta tarea se implementa en la función *AssignULRBG* de 5G-LENA (ver Figura 35). Es importante destacar que se ha modificado el orden que sigue 5G-LENA para asignar recursos radio para transmisiones en UL. Como se describe en el apartado 6.1.2, en la implementación original de 5G-LENA, los recursos radio para las transmisiones en UL se asignan desde el último al primer símbolo OFDM dentro de un *slot*. Sin embargo, este orden no está establecido en 5G NR. Por lo tanto, se ha eliminado esta condición. Ahora los recursos se pueden asignar desde el primer hasta el último símbolo OFDM dentro de un *slot*, lo cual es fundamental para garantizar una latencia muy baja.

Además, se han implementado dos nuevas políticas de *scheduling* en 5G-LENA. Estas nuevas políticas se pueden utilizar con el mecanismo de CG. Cada *scheduler* decide el número r_i de RBs y el número s_i de símbolos OFDM que se le asigna a cada UE_i con i desde 1 hasta N_{UE} (N_{UE} representa el número máximo de UEs que han solicitado recursos).

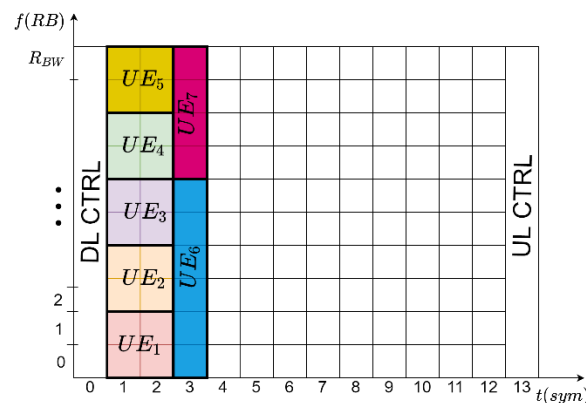
La primera política de *scheduling* a la que se le ha llamada Sym-OFDMA (Figura 36.a), minimiza el número de símbolos OFDM asignados a cada UE. Su funcionamiento se describe en el Algoritmo V. Esta técnica tiene como objetivo minimizar el número de símbolos asignados a cada UE; este fue el mismo criterio aplicado en los esquemas de los capítulos 4 y 5 para calcular la demanda de RBs y símbolos OFDM. Sym-OFDMA sirve los UEs siguiendo el orden de llegada, es decir, desde el UE_1 hasta el UE $UE_{N_{UE}}$ (línea 3 del Algoritmo V). Sym-OFDMA asigna recursos radio desde el primer hasta el último símbolo dentro de un *slot* (las variables n_{slot} , n_s y n_{RB} representan el *slot*, el símbolo dentro del *slot* y el RB, respectivamente, que se está asignando actualmente a un UE). Cada UE_i demanda d_i recursos radio, que se calcula en función del tamaño del paquete a transmitir y el MCS que utiliza en la transmisión del paquete. Si el número d_i de recursos radio demandados por un UE_i es igual o menor que los recursos que hay en un símbolo OFDM (R_{BW}), entonces, el *scheduler* asigna $r_i = d_i$ RBs en $s_i=1$ símbolo OFDM (ver líneas 4-5 del Algoritmo V). Si el UE_i demanda más de R_{BW} recursos radio (es decir, $d_i > R_{BW}$), UE_i recibe $r_i = R_{BW}$ RBs en $s_i = \lceil d_i / R_{BW} \rceil$ símbolos OFDM consecutivos (línea 7 del Algoritmo V). $s_{ini,i}$ y $r_{ini,i}$ y $slot_i$ representan el primer RB, símbolo OFDM y *slot*, respectivamente, asignados a UE_i (ver línea 15). UE_{i+1} recibirá los RBs en el mismo símbolo OFDM que UE_i si el número de RBs no asignados es suficiente para satisfacer la demanda de UE_{i+1} (ver líneas 15-16). De lo contrario, UE_{i+1} recibirá los RBs en el siguiente símbolo OFDM (líneas 9-11). Sym-OFDMA también considera que un UE solo puede recibir RBs dentro de un *slot* (esta restricción está establecida por el estándar de 3GPP [33]). En caso de que no haya suficientes RBs y símbolos no asignados para satisfacer d_i RBs en el *slot* actual, se asignarán recursos para UE_i en el siguiente *slot* (líneas 12-14). La Figura 36.a muestra un ejemplo de asignación de recursos radio realizada por Sym-OFDMA en un escenario donde 7 UEs demandan 4 RBs cada uno para transmitir su paquete. La Figura 36.a representa los recursos radio en un *slot* de 14 símbolos OFDM, donde el primer y último símbolo están reservados para señales de control y el ancho de banda contiene $R_{BW} = 10$ RBs.

La segunda política de *scheduling* a la que se le ha llamada RB-OFDMA, asigna recursos radio a los UEs intentando minimizar el número de RBs que quedan sin asignar. El funcionamiento de RB-OFDMA se presenta en el Algoritmo VI. Con esta política, cada UE debe recibir al menos r_{min} recursos radio en un símbolo OFDM. RB-OFDMA, al igual que Sym-OFDMA, asigna recursos radio a los UEs siguiendo el orden de llegada (*scheduling* basado en *First Come First Served*). La asignación de recursos radio se realiza desde el primero hasta el último símbolo OFDM dentro de un *slot*. RB-OFDMA distribuye los R_{BW} RBs en un símbolo OFDM entre el número máximo de UEs, considerando que $r_i \geq r_{min}$ para cada UE_i , es decir, $n_{UE} = \lfloor R_{BW} / r_{min} \rfloor$ (línea 3 del Algoritmo VI). Los UEs se dividen en conjuntos de n_{UE} UEs (línea 5); el último conjunto de UEs puede tener menos

UEs que n_{UE} . Los UEs de cada conjunto φ compartirán los RBs en los mismos símbolos OFDM. Si el número de UEs en un conjunto φ es igual a n_{UE} , cada UE $_i$ de φ recibirá $r_i = r_{min}$ RBs (línea 7 y 8). Si el número de UEs en un conjunto φ es menor que n_{UE} , los primeros $((n_{UE}/n) - \lfloor n_{UE}/n \rfloor) \cdot n$ UEs en φ recibirán $r_i = \lfloor n_{UE}/n \rfloor \cdot r_{min}$ RBs (línea 10). El resto de los UEs en φ recibirán $r_i = \lfloor n_{UE}/n \rfloor \cdot r_{min}$ RBs (línea 11). Basándose en el número de RBs asignados a cada UE, RB-OFDMA calcula el número d_i^S de símbolos OFDM que cada UE $_i$ necesita para satisfacer su demanda d_i (línea 14). Una vez que se conoce d_i^S , todos los UEs en un grupo φ reciben el número de RBs necesarios para satisfacer al UE que requiere un mayor número de símbolos OFDM (línea 16 y 17). RB-OFDMA también considera que un UE solo puede recibir RBs dentro de un *slot*. En caso de que no haya suficientes RBs y símbolos OFDM no asignados en el *slot* actual para cumplir con r_i RBs y s_i símbolos OFDM que necesita UE $_i$, el gNB asignará recursos para dicho UE en el siguiente *slot* (línea 18-20). La Figura 36.b muestra un ejemplo de la asignación de recursos radio realizada con RB-OFDMA. El ejemplo considera 7 UEs que demandan 4 RBs para transmitir sus datos, siendo $r_{min} = 2$ RBs. En este ejemplo, los N_{UE} UEs se dividen en dos conjuntos. El primer conjunto incluye 5 UEs (igual a $\lfloor R_{BW}/r_{min} \rfloor$) que reciben $r_i = 2$ RBs y $s_i = 2$ símbolos OFDM. El segundo conjunto incluye 2 UEs que reciben $r_i = 4$ RBs y $s_i = 6$ RBs para UE $_7$ y UE $_6$, respectivamente, y un símbolo OFDM ($s_i = 1$).



a) Sym-OFDMA



b) RB-OFDMA

Figura 36. Asignación de recursos de radio con las políticas diseñadas.

ALGORITHM V: SYM-OFDMA

1. Input: $d_i \forall i \in [1, N_{UE}]$
 2. $n_{slot}=1, n_s=1, n_{RB}=1$
 3. **For** $i=1$ to N_{UE}
 4. **If** $d_i \leq R_{BW}$
 5. $r_i=d_i, s_i=1$
 6. **Else**
 7. $r_i=R_{BW}, s_i=\lceil d_i/R_{BW} \rceil$
 8. **End If**
 9. **If** $R_{BW} - n_{RB} + 1 < r_i$
 10. $n_{RB}=1, n_s=n_s + s_i$
 11. **End If**
 12. **If** $n_s + s_i - 1 > S_{slot}$
 13. $n_{slot}=n_{slot}+1, n_s=1, n_{RB}=1$
 14. **End If**
 15. $r_{ini,i}=n_{RB}, s_{ini,i}=n_s, slot_i=n_{slot}$
 16. $n_{RB}=n_{RB} + r_i$
 17. **End For**
-
-

ALGORITHM VI: RB-OFDMA

1. Input: $d_i \forall i \in [1, N_{UE}], r_{min}$
 2. $n_{slot}=1, n_s=1, n_{RB}=1, n_{UEwithRB}=0$
 3. Define $n_{UE}=\lceil R_{BW}/r_{min} \rceil$
 4. **While** there are UEs without resources
 5. Create set ϕ with UE $_i \forall i$ such that
 $i=[n_{UEwithRB}+1, \min(n_{UEwithRB}+n_{UE}, N_{UE})]$
 6. Define n =number of UEs in ϕ
 7. **If** $n=n_{UE}$
 8. $r_i=r_{min} \forall UE_i$ in ϕ
 9. **Else**
 10. $r_i=\lceil \frac{n_{UE}}{n} \rceil \cdot r_{min}$ for the first $(\frac{n_{UE}}{n} - \lfloor \frac{n_{UE}}{n} \rfloor) \cdot n$ UEs in ϕ
 11. $r_i=\lfloor \frac{n_{UE}}{n} \rfloor \cdot r_{min}$ for the last $n - (\frac{n_{UE}}{n} - \lfloor \frac{n_{UE}}{n} \rfloor) \cdot n$ UEs in ϕ
 12. **End If**
 13. **For** all UE $_i$ in ϕ
 14. $d_i^S=d_i/r_i, r_{ini,i}=n_{RB}, n_{RB}=n_{RB} + r_i$
 15. **End For**
 16. $d^S=\max\{d_i^S\} \forall UE_i$ in ϕ
 17. $s_i=d^S \forall UE_i$ in ϕ
 18. **If** $n_s=n_s + d^S - 1 > S_{slot}$
 19. $n_{slot}=n_{slot}+1, n_s=1$
 20. **End if**
 21. $s_{ini,i}=n_s, slot_i=n_{slot} \forall UE_i$ in ϕ
 22. $n_s=n_s + d^S, n_{RB}=1$
 23. $n_{UEwithRB}=n_{UEwithRB} + n_{UE}$
 24. **End While**
-
-

El valor de r_{min} se puede ajustar para optimizar la eficiencia del uso de los recursos radio. En esta tesis, r_{min} ha sido calculada para minimizar el número de recursos radio que quedan sin asignar a los UEs. Si varios valores satisfacen esta condición, se selecciona el valor que proporciona la diferencia más baja entre los recursos radio asignados y la

demanda de los usuarios. El procedimiento para el cálculo de r_{min} se presenta en el Algoritmo III.

ALGORITHM VII: r_{min} in *RB-OFDMA*

1. $aux=R_{BW}$
 2. Create $X=\{x\}$ with $x \in \mathbb{N}$, $1 < x < R_{BW}$, and $\text{mod}(aux/x)=0$.
 3. **If** X is empty
 4. $aux=aux-1$
 5. Go to 3
 6. **End If**
 7. Calculate number $d_i^S(x)$ of OFDM symbols needed to meet the UE demands d_i when UEs receives x RBs, $\forall i \in [1, N_{UE}]$ and $\forall x \in X$.
 8. Set $r_{min} = x \in X$ that satisfies $\min_{i,x} (x \cdot d_i^S(x) - d_i)$
-

6.3. Escenarios evaluados

Se ha considerado un escenario de evaluación en el que la red 5G NR da conectividad a una aplicación de control de lazo cerrado dentro de una celda industrial típica que cubre un área de $10 \times 10 \text{m}^2$ [26]. 5G trabaja en la banda de 3.7-3.8GHz [75], con un BW igual a 10, 20 o 40 MHz, y opera en modo TDD [68]. Se han considerado *slots* de 14 símbolos OFDM y se han evaluado 3 configuraciones distintas, siendo los últimos 13, 9 y 5 símbolos OFDM para transmisiones en UL y el resto de los símbolos OFDM para transmisiones en DL (configuraciones de *slots*: 1D13U, 5D9U y 9D5U). El primer y último símbolo dentro de un *slot* se han reservado para la transmisión de los canales de control en DL y UL, respectivamente. Se utiliza un SCS de 30 kHz, como se recomienda en [32] para entornos industriales. Hay $N_{UE} = 15$ sensores distribuidos aleatoriamente dentro de la celda y están en línea de visión con el gNB. El canal de radio se caracteriza por *fast-fading* y *shadowing*, y sigue un modelo de canal definido en [76]. Los sensores generan paquetes de datos periódicos simultáneamente que se transmiten a un sistema de monitorización central. Los paquetes de datos están constituidos por 10 o 25 bytes, y una periodicidad de 10 ms. Se agrega un encabezado IPv4 con un valor de 22 bytes a este tamaño de paquete. Los sensores transmiten los paquetes de datos en UL hacia el sistema de monitorización central. Todos los sensores utilizan MCS 12 en la tabla 1 del estándar [33] para la transmisión de tráfico periódico; MCS 12 proporciona un buen equilibrio entre robustez y eficiencia de transmisión en el escenario considerado. Los tiempos de procesamiento en el UE y el gNB se calculan según lo indicado en [17] y [33]. Se ha considerado el uso de un código de verificación de redundancia cíclica (CRC).

Se ha comparado el rendimiento de CG utilizando tanto las nuevas políticas de *scheduling* presentadas en el apartado 6.2.3, así como, los esquemas 5GL-TDMA y 5GL-OFDMA ya implementados en 5G-LENA. Para realizar una comparación justa, se aplica el *scheduler First Come First Served* con 5GL-TDMA y 5GL-OFDMA. En este contexto, el *scheduler* atiende a los UEs siguiendo el orden de llegada. Cada UE recibe el número de RBs y símbolos OFDM requeridos para satisfacer su demanda. Cuando se utiliza 5GL-TDMA, todos los UEs recibirán $r_i = R_{BW}$ RBs y $s_i = \lceil r_i/R_{BW} \rceil$ símbolos OFDM consecutivos. Cuando se utiliza 5GL-OFDMA, todos los UEs recibirán $r_i = \lceil d_i/S_{slot} \rceil$ RBs en $s_i = S_{slot}$ símbolos OFDM.

6.4. Validación

En esta sección, se ha llevado a cabo la validación de la implementación de CG en 5G-LENA. La validación se ha basado en la comparación de la latencia obtenida en el simulador con la latencia obtenida en otras fuentes de la literatura y mediante un modelo analítico. En esta tesis, la latencia experimentada en UL representa el tiempo transcurrido desde que se crea un paquete en la capa RLC del UE hasta que se recibe en la capa RLC del gNB. La validación se ha realizado en dos pasos:

- 1) En primer lugar, se realiza la comparación de los resultados de latencia obtenidos mediante simulación de CG en 5G-LENA, con los valores analíticos previamente reportados en 3GPP TR 37.910 [17]. En este caso, se considera que en la red 5G solo hay un UE solicitando recursos, lo que permite analizar la latencia mínima que el UE puede experimentar.
- 2) En segundo lugar, se han desarrollado expresiones analíticas para contrastar las máximas latencias experimentadas en UL obtenidas mediante simulaciones con las obtenidas utilizando el modelo analítico. Esta validación se ha realizado considerando un escenario en el que varios UEs demandan recursos al gNB, el cual emplea las diversas políticas y técnicas de acceso múltiple implementadas en 5G-LENA para asignar los recursos radio.

6.4.1. Validación en entornos con poca carga de usuarios

En primer lugar se ha realizado la comparación de los resultados de latencia obtenidos mediante simulación con CG en 5G-LENA y los resultados de latencia presentados en 3GPP TR 37.910 [17] (ver Tabla 8). En este caso, se considera una red 5G con un BW igual a 20 MHz, en el que un UE transmite paquetes vacíos utilizando 2 símbolos OFDM. Las tramas están configuradas con numerologías (μ) 0, 1 y 2, que corresponden a una duración de símbolo de 71.4 μ s, 25.6 μ s y 17.9 μ s, respectivamente. Se puede observar en la tabla que

los resultados de latencia obtenidos a través de simulaciones de CG en 5G-LENA son muy similares a los resultados analíticos presentados en 3GPP TR 37.910 [17], lo que valida el CG implementado en 5G-LENA. En la tabla también se presentan la latencia que se puede obtener utilizando el *scheduling* dinámico. Como se puede observar en el caso de la numerología 0, la latencia obtenida con CG se reduce en un 93.3% en comparación con la latencia obtenida con el *scheduling* dinámico (columnas 1 y 2 de la Tabla 8).

Tabla 8. Latencia experimentada con el mecanismo dinámico y con CG (resultados simulados y analíticos).

μ	Dynamic scheduling	Configured Grant	
		5G-LENA	TR 37.910 [17]
0	7.31 ms	0.49 ms	0.52 ms
1	3.70 ms	0.30 ms	0.30 ms
2	1.90 ms	0.25 ms	0.24 s

6.4.2. Validación en entornos con carga de usuarios moderada y alta

En este apartado, se realiza una comparación entre los resultados de latencia máxima en UL obtenidos mediante simulación y aquellos obtenidos mediante un modelo analítico. Para ello, se han derivado expresiones analíticas de la latencia máxima en UL experimentada por los UEs cuando utilizan el *scheduler* CG con las distintas políticas de *scheduling* y técnicas de acceso múltiples. Se considera que todos los UEs (UE_i con i desde 1 hasta N_{UE}) transmiten paquetes de 10 bytes. Considerando el modelo derivado en [77], la latencia de un paquete transmitido en UL utilizando una configuración CG (L_{UL}) se calcula como:

$$L_{UL} = t_{UE,tx} + t_{fa} + t_w + t_t + t_{gNB,rx} \quad (27)$$

La expresión en (27) tiene en cuenta los tiempos de procesamiento en el UE y en el gNB que representan el tiempo necesario para generar el paquete en el transmisor y decodificar los datos en el receptor, ($t_{UE,tx}$ y $t_{gNB,rx}$, respectivamente). También considera el tiempo de alineación con las tramas, es decir, el intervalo de tiempo que hay desde la creación de un paquete hasta la próxima oportunidad de transmitir dicho paquete por el canal (t_{fa}) en función de la configuración de la trama utilizada. El modelo de latencia también considera el tiempo de espera que hay entre el tiempo de alineación t_{fa} y los recursos radio que se han asignado para transmitir el paquete (t_w) y el tiempo necesario para la transmisión del paquete por el canal (t_t).

El tiempo de alineación t_{fa} depende de la configuración de la trama y de los *slots*. El tiempo de espera t_w depende del modo de acceso múltiple y de la política de *scheduling*. El tiempo de transmisión t_{tt} es igual a la duración en tiempo de los símbolos (s_i) OFDM utilizados para la transmisión del paquete y se puede calcular como $s_i \cdot t_{sym}$ siendo t_{sym} la duración de un símbolo OFDM. t_{sym} depende de la numerología utilizada para la transmisión del paquete y s_i depende de la política y el modo de acceso múltiple.

En la Tabla 9 se muestran las expresiones de la latencia máxima experimentada en UL con cada política de *scheduling* y técnica de acceso múltiple implementados en 5G-LENA. n_{slot} representa el número de *slots* que son necesarios para asignar recursos a todos los N_{UE} UEs y n_{sym} corresponde al número de símbolos OFDM que se necesita para servir a todos los UEs. En caso de RB-OFDMA (31), el número máximo de UEs que pueden compartir los RBs en un mismo símbolo OFDM viene dado por $n_{UE,1}$. Todos los UEs de este grupo recibirán $r_i^{RB-OFDMA(1)}$ RBs en $s_i^{RB-OFDMA(1)}$ símbolos OFDM consecutivos, mientras que el último grupo (referido como $n_{UE,2}$) tendrá asignados $r_i^{RB-OFDMA(2)}$ RBs en $s_i^{RB-OFDMA(2)}$ símbolos OFDM consecutivos. Estas expresiones se han derivado como parte del trabajo realizado en esta tesis y el proceso seguido para obtener estas expresiones puede encontrarse en el artículo [21].

Tabla 9. Expresiones finales de la máxima latencia en UL con el *scheduler* correspondiente.

5GL-TDMA	$L_{UL} = t_{UE,tx} + (2 \cdot n_{slot} - 1 + N_{UE} \cdot \left\lceil \frac{d_i}{R_{BW}} \right\rceil) \cdot t_{sym} + t_{gNB,rx} \quad (28)$
5GL-OFDMA	$L_{UL} = t_{UE,tx} + (14 \cdot n_{slot} - 1) \cdot t_{sym} + t_{gNB,rx} \quad (29)$
Sym-OFDMA	$L_{UL} = t_{UE,tx} + (2 \cdot n_{slot} - 1 + n_{sym}) \cdot t_{sym} + t_{gNB,rx} \quad (30)$
RB-OFDMA	$L_{UL} = t_{UE,tx} + t_{gNB,rx} + \left(2 \cdot n_{slot} - 1 + \left\lceil \frac{N_{UE}}{n_{UE,1}} \right\rceil \cdot s_i^{RB-OFDMA(1)} + s_i^{RB-OFDMA(2)} \right) \cdot t_{sym} \quad (31)$

La Figura 37 compara la latencia máxima experimentada en transmisiones UL obtenida analíticamente y mediante simulaciones al utilizar CG con 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA y RB-OFDMA para diferentes valores de ancho de banda cuando el tamaño del paquete es de 10 bytes y se utiliza el formato de *slot* 1D13U. Se ha simulado un escenario con 15 UEs, con numerología 1 y un MCS igual a 12. Los resultados analíticos se muestran como puntos discretos y los simulados como líneas. La figura muestra que los resultados simulados y analíticos coinciden para todas las técnicas de acceso múltiple y políticas evaluadas. Estos resultados validan la implementación de CG en 5G-LENA. Además, los resultados muestran que la máxima latencia obtenida con Sym-OFDMA y RB-OFDMA es igual o menor que la máxima latencia experimentada con 5GL-TDMA y 5GL-OFDMA para los distintos anchos de banda evaluados. Esto es gracias a la flexibilidad ofrecida por OFDMA que permite hacer un uso de los recursos radio más eficiente y eficaz. En consecuencia, las nuevas políticas implementadas consiguen menores latencias de transmisión.

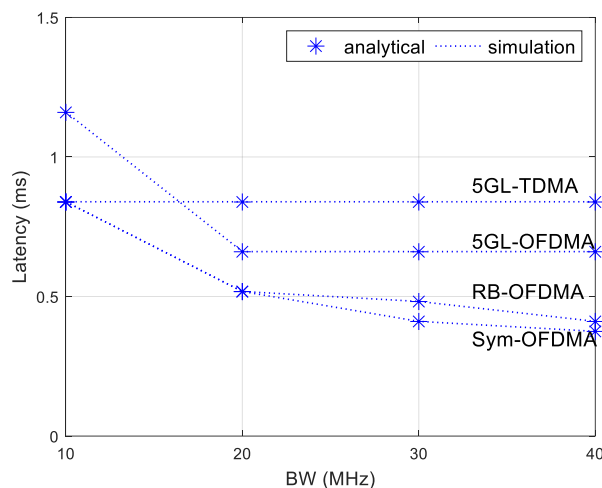


Figura 37. Máxima latencia UL en función del ancho de banda (paquetes de 10 bytes).

6.5. Evaluación

Una vez validada la implementación del CG en 5G-LENA comparándola con modelos analíticos, en este apartado, se evalúa la latencia que se puede lograr utilizando CG con las distintas técnicas de acceso múltiple y políticas de *scheduling* implementadas en 5G-LENA: 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA y RB-OFDMA. Se considera el escenario descrito en la sección 6.3 y se analiza el impacto sobre la latencia de distintos parámetros, tales como el ancho de banda de la celda, el tamaño del paquete, el MCS utilizado y el impacto de la configuración de la trama y del *slot*.

6.5.1. Impacto del ancho de banda de la celda

La Figura 38 muestra un diagrama de caja o *box plot* de la latencia experimentada en UL por los UEs. La línea roja dentro del *box plot* representa el promedio de la latencia experimentada y el borde inferior y superior del cuadro son los percentiles 10 y 90, respectivamente. Las cruces representan los valores mínimos y máximos. La Figura 39 muestra el número de recursos radio utilizados con las diferentes políticas y modos de acceso múltiples, respecto al número de recursos radio que se asignan a través del *scheduler* 5GL-TDMA. Los resultados muestran que Sym-OFDMA y RB-OFDMA logran una latencia máxima y promedio igual o menor que 5GL-TDMA y 5GL-OFDMA para todos los valores de ancho de banda evaluados (Figura 38). Esto es debido a un uso más eficiente de los recursos radio (Figura 39).

Cuando el ancho de banda es de 10 MHz, Sym-OFDMA y RB-OFDMA logran resultados de latencia similares a los de 5GL-TDMA. En este caso, Sym-OFDMA asigna al mismo UE todos los RBs de un símbolo OFDM, al igual que 5GL-TDMA (Figura 39). En cambio, RB-OFDMA, asigna a cada UE RBs en dos símbolos consecutivos, tal que los RBs en un mismo símbolo OFDM son compartidos por dos UEs. Como resultado, el número de recursos utilizado y la latencia máxima experimentada por RB-OFDMA es la mismas que con 5GL-TDMA y Sym-OFDMA.

Cuando el ancho de banda aumenta (20 o 40 MHz), aunque 5GL-OFDMA reduce la latencia máxima experimentada en comparación con 5GL-TDMA, Sym-OFDMA y RB-OFDMA siguen obteniendo menores latencias que 5GL-OFDMA y 5GL-TDMA (Figura 38). En estos casos, Sym-OFDMA permite asignar a distintos UEs RBs en el mismo símbolo OFDM. Por tanto, este *scheduler* requiere menos símbolos OFDM para servir a todos los UEs en comparación con 5GL-TDMA. RB-OFDMA también reduce la latencia máxima experimentada cuando el ancho de banda aumenta de 20 a 40 MHz. En cambio, la latencia mínima y promedio incrementan cuando $BW = 40\text{MHz}$ respecto a la latencia obtenida cuando $BW = 20\text{MHz}$, debido a que todos los UEs asignan el mismo número de símbolos OFDM. Es decir, el primer set de UEs utiliza un mayor número de símbolos OFDM consecutivo con 40MHz que con 20MHz. En caso de 5GL-OFDMA, aun necesitando un menor número de símbolos OFDM para abastecer la demanda (Figura 39), está restringido a asignar a cada UE por lo menos un RB en cada símbolo OFDM del *slot*, por lo que su latencia mínima es igual que el tiempo que se tarda en transmitir un paquete utilizando un *slot* completo. Al poder aprovechar las capacidades de OFDMA en 5G NR, las dos nuevas políticas, tanto Sym-OFDMA como RB-OFDMA, permiten asignar los recursos radio de manera más eficiente y reducir la latencia experimentada.

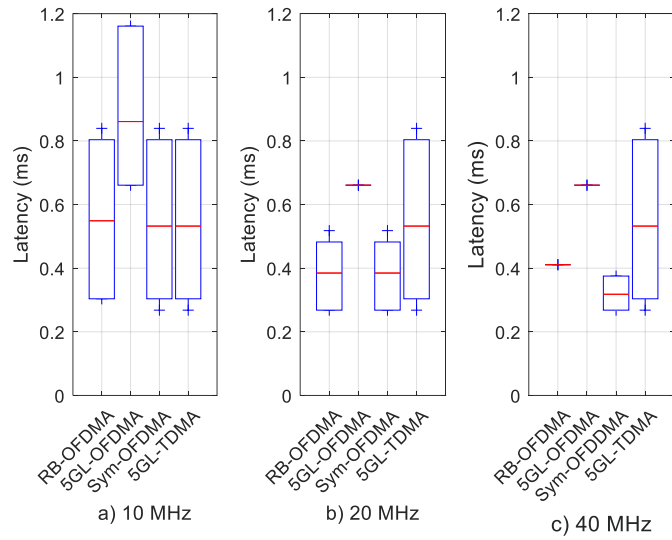


Figura 38. Latencia UL experimentada en función del ancho de banda (tamaño de paquete de 10 bytes).

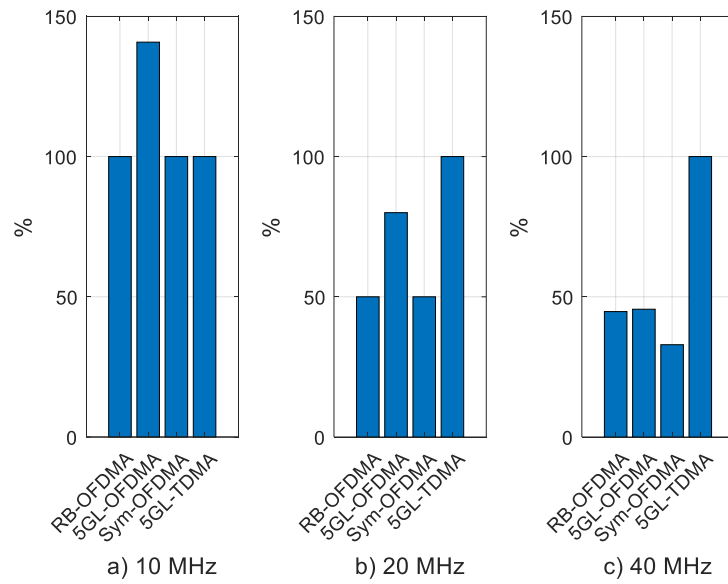


Figura 39. Porcentaje de recursos radio utilizados para la transmisión de paquetes con respecto a 5GL-TDMA (tamaño de paquete de 10 bytes).

6.5.2. Impacto del tamaño de paquete

La Figura 40 muestra *box plots* de la latencia que los paquetes experimentan cuando se transmiten en UL desde el UE al gNB en una red 5G con ancho de banda de 20 MHz¹¹ que utiliza el *scheduler* CG con 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA y RB-OFDMA. En este caso, los UEs transmiten paquetes de 10 y 25 bytes, respectivamente y utilizan el formato de *slot* 1D13U. La Figura 40 indica que la máxima latencia experimentada por los UEs disminuye con el tamaño del paquete para Sym-OFDMA y RB-OFDMA, mientras que se mantiene constante para 5GL-TDMA y 5GL-OFDMA. Esto es debido a que 5GL-TDMA y 5GL-OFDMA asignan más recursos radio que lo que realmente necesitan los UEs para transmitir sus paquetes. Mientras que Sym-OFDMA y RB-OFDMA asignan a cada UE una cantidad de recursos radio más ajustado a lo que los UEs demandan, aprovechando así al máximo la flexibilidad que ofrece OFDMA en 5G NR (ver Figura 40.a). En cambio, cuando el tamaño de paquetes es igual a 25 bytes (ver Figura 40.b), el número de RBs que hay en un símbolo OFDM no es suficiente para satisfacer la demanda de más de un UE, es decir, $R_{BW} < 2 \cdot r_i$. En consecuencia, Sym-OFDMA asigna todos los RBs de un símbolo OFDM a un único UE, tal que los paquetes que se transmiten en los recursos asignados a través de Sym-OFDMA o 5GL-TDMA van a experimentar la misma latencia.

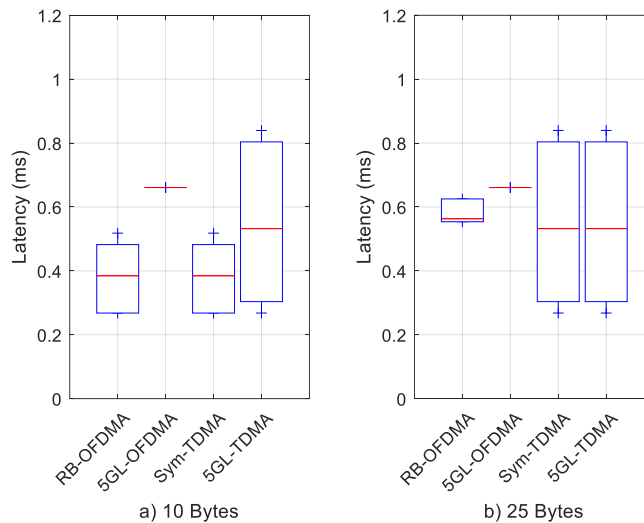


Figura 40. La latencia experimentada en UL como función del tamaño del paquete (BW = 20 MHz).

¹¹ Se ha escogido este ancho de banda para el resto de las simulaciones porque es el que se recomienda para entornos industriales [67].

6.5.3. Impacto de MCS

La Figura 41 muestra la latencia UL máxima experimentada por los UE y el porcentaje de recursos radio asignados utilizando CG con las diferentes políticas y modos de acceso múltiple, cuando se utiliza MCS igual a 12, 20 y 28. El BW considerado es de 20 MHz y se transmiten paquetes de 10 bytes. El porcentaje de recursos radio asignados se calcula con respecto al número de recursos radio utilizados por 5GL-TDMA. La Figura 41.a muestra que Sym-OFDMA y RB-OFDMA reducen la latencia máxima experimentada por los UEs para todos los MCSs evaluados en comparación con 5GL-OFDMA y 5GL-TDMA. Esto se debe al uso más eficiente de los recursos radio (ver Figura 41.b).

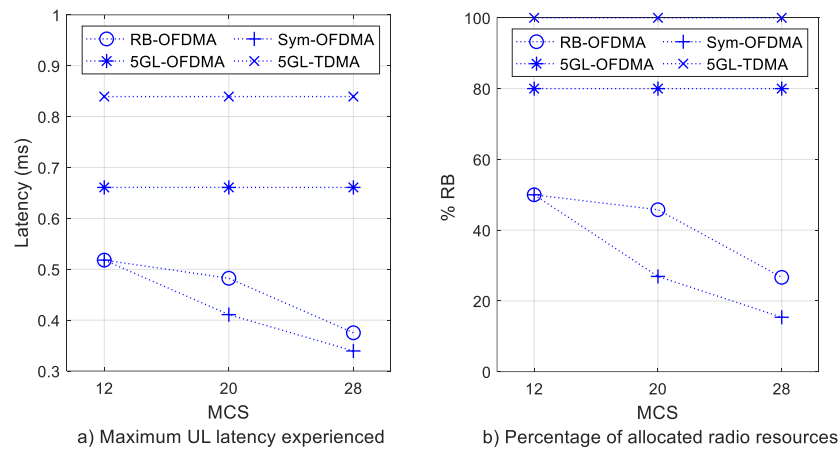


Figura 41. Rendimiento en función del MCS (BW = 20 MHz, tamaño de paquete = 10 bytes).

6.5.4. Impacto de formato de trama y *slot*

Por último, se ha evaluado el uso de 3 configuraciones diferentes de tramas y *slots* TDD (ver Figura 42). En particular, se ha considerado que los primeros 1, 5 y 9 símbolos OFDM dentro de un *slot* se utilizan para transmisiones en DL y los últimos 13, 9 y 5 símbolos OFDM para transmisiones en UL. Las tres configuraciones se denominan, 1D13U, 5D9U y 9D5U. La Figura 42 muestra el *box plot* de la latencia UL experimentada por los UEs cuando se utilizan diferentes configuraciones de trama, *slot*, modos de acceso y *schedulers*. Los resultados muestran que el rendimiento experimentado por el tráfico UL no se ve afectado por la cantidad de tráfico generado en DL, sino más bien por el número de recursos

radio reservados para la transmisión en UL. Este número de recursos está determinado por la configuración de la trama y el *slot*.

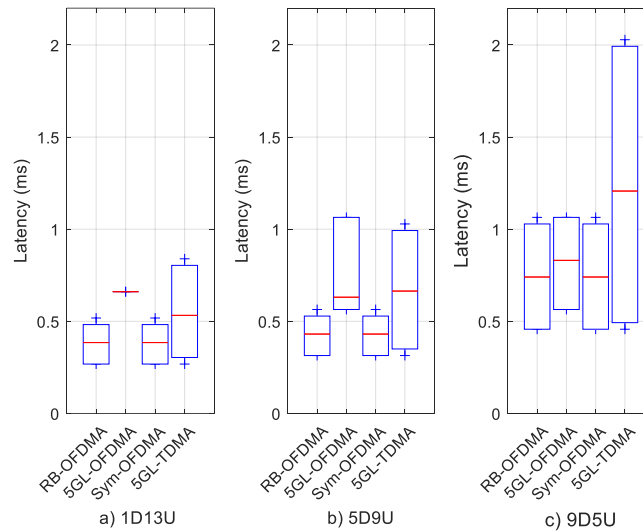


Figura 42. Latencia experimentada con los distintos esquemas en función del formato del *slot* (BW = 20 MHz y tamaño de paquete de 10 bytes).

6.6. Conclusiones y resultados adicionales

En este capítulo se ha presentado una descripción detallada de la primera implementación de CG en el simulador de código abierto de 5G NR 5G-LENA. El mecanismo CG preasigna recursos radio a los UEs y evita el intercambio de señalización entre el UE y el gNB para solicitar/informar sobre los recursos radio asignados, lo que reduce la latencia de la transmisión. Además, se ha modificado la implementación de OFDMA de 5G-LENA, para poder modelar con precisión la alta flexibilidad ofrecida por 5G NR a la hora de asignar los recursos radio. Los resultados de latencia obtenidos utilizando CG en 5G-LENA coinciden con los valores de latencia obtenidos en estudios analíticos en la literatura, lo cual valida la implementación de CG en 5G-LENA.

También se han implementado dos políticas de *scheduling* basadas en CG en un escenario industrial para demostrar la flexibilidad y capacidad de 5G NR para soportar servicios críticos. Estas políticas de *scheduling* explotan la flexibilidad ofrecida por OFDMA para garantizar bajas latencias y un uso eficiente de los recursos radio. Los resultados muestran que el uso de CG con las políticas de *scheduling* propuestas y OFDMA reduce la latencia máxima experimentada por los usuarios en un 48.6% en comparación

con los esquemas de acceso múltiple previamente implementados en 5G-LENA. Esto se logra utilizando un 80.7% menos de recursos radio.

La implementación de CG y OFDMA en 5G-LENA podrá ser utilizada, tanto en investigaciones realizadas por Ikerlan, como también por la comunidad científica para realizar simulaciones a nivel de sistema de red de 5G NR para soportar servicios críticos que demandan bajas latencias. El código de CG implementado en 5G-LENA está disponible en [18] y utiliza la versión de ns-3 disponible en [19]. El simulador de 5G-LENA ahora está preparado para continuar con la línea de investigación de esta tesis.

A continuación, listamos los detalles y resultados adicionales relativos a este capítulo que pueden encontrarse en nuestras publicaciones de conferencia [20] y artículo de revista [21], incluidos en el Anexo A.

- Evaluación de la latencia experimentada en función de la numerología y en función del número de UEs, cuando se utilizan los *schedulings* CG y dinámico, con el modo de acceso 5GL-TDMA y el planificador de *Round Robin*, empezando a asignar desde el último hasta el primer símbolo del *slot* que están reservados para transmisiones en UL. Este análisis se ha incluido en [20].
- Una discusión más detallada de las gráficas obtenidas en la sección de evaluación (apartado 6.5) se encuentra en el artículo [21].
- La definición de los pasos para obtener las expresiones matemáticas de la latencia para el análisis de la implementación (apartado 6.4.2) se encuentra en el artículo [21].

7. Conclusiones y líneas futuras de investigación

La integración de las redes 5G-TSN es fundamental para respaldar los servicios que la Industria 4.0 ofrecerá, en los que la movilidad de los dispositivos y el determinismo de las comunicaciones serán aspectos clave. Actualmente, 3GPP está definiendo los pasos necesarios para lograr una integración coordinada, eficiente y efectiva, en la cual 5G actúa como un *bridge* virtual dentro de la red TSN. Sin embargo, todavía hay varios aspectos que se deben trabajar para que esta integración sea implementable en la industria. En este contexto, un aspecto crítico para cumplir con los requisitos de latencia de extremo a extremo de las aplicaciones industriales es la coordinación de los esquemas de *scheduling* de ambas redes. El *scheduling* de los paquetes de los flujos TSN en la red 5G debe realizarse de manera que se garanticen los tiempos de llegada y salida previamente calculados por la red TSN. El trabajo realizado en esta tesis se ha centrado en la gestión de recursos radio para soportar flujos TSN dentro del *bridge* 5G virtual. El análisis inicial ha facilitado la comprensión de la integración de las redes 5G y TSN. Además, en este primer trabajo se ha realizado un estudio analítico para poder cuantificar la latencia máxima que cada flujo TSN debe garantizar dentro de un *bridge* 5G virtual. También se han propuesto y evaluado tres esquemas de *scheduling* en 5G basados en el mecanismo *Configured Grant* o CG. Estos *schedulers* tienen en cuenta los requisitos necesarios para garantizar la transmisión de los flujos TSN, evitando los conflictos que se pueden dar al asignar recursos a través de configuraciones CG a flujos TSN con distintas periodicidades. Asimismo, se ha propuesto incorporar un vector de activación en la ecuación estandarizada de CG en 3GPP TS 38.321 [16] que determina en que recursos radio debe transmitir su paquete cada UE. A través de este vector de activación, el UE podrá acceder a los recursos radio de forma coordinada, evitando solapamientos entre recursos radio asignados a otros UEs. Esta modificación evita

la señalización entre los UEs y el gNB para activar y desactivar las configuraciones CG correspondientes. Por último, la preparación del simulador ns-3 5G-LENA para comunicaciones de bajas latencias ha sido la última contribución de esta tesis, permitiendo a la comunidad científica continuar con la investigación de diversas tecnologías que requieran comunicaciones de bajas latencias. A continuación, en el apartado 7.1 se exponen con más detalle las principales conclusiones y contribuciones científicas obtenidas del trabajo de investigación realizado en el marco de esta tesis doctoral y en el apartado 7.2 se proponen líneas futuras de investigación.

7.1. Contribuciones científicas y conclusiones

La primera aportación de esta tesis es el resultado del estudio analítico sobre las latencias requeridas al *bridge* 5G virtual para soportar tráfico TSN. Como resultado de este análisis, se han obtenido una serie de expresiones analíticas para calcular el requisito de latencia máxima que pueden soportar los paquetes de cada flujo TSN dentro del *bridge* 5G virtual. Estas expresiones han sido clave para los estudios posteriores.

Para poder coordinar ambas redes eficientemente se han diseñado y evaluado dos esquemas de *scheduling* basados en CG para redes 5G, a los que se les han llamado O-FAST y H-FAST (basado en optimización y heurístico, respectivamente). El principal objetivo de ambos esquemas es evitar conflictos de asignación de recursos radio entre paquetes de flujos TSN que puedan tener diferentes requisitos de periodicidad. Estos esquemas buscan una solución de *scheduling* en 5G que cumpla con los tiempos establecidos por el *scheduling* de TSN para todos los paquetes de distintos flujos TSN. Una vez alcanzado este objetivo, los esquemas de *scheduling* propuestos intentan minimizar la latencia experimentada por los distintos flujos TSN. Para ello, las nuevas propuestas crean varias configuraciones CG para cada flujo TSN. Los resultados muestran que ambos esquemas mejoran significativamente la cantidad de flujos TSN que pueden ser atendidos en la red 5G-TSN integrada en comparación con esquemas de *scheduling* comúnmente utilizados (por ejemplo, [12]). Esto es debido a que los esquemas de referencia no consideran la transmisión de flujos con distintas periodicidades y, por tanto, dichos flujos, en determinados momentos, terminan recibiendo los mismos recursos radio para transmitir los paquetes. Cuando la carga de flujos TSN en el escenario es moderada, O-FAST y H-FAST encuentran recursos radio para todos los flujos en un 100% de los escenarios evaluados, sin embargo, el esquema de referencia solo es capaz de satisfacer los requisitos de los flujos para un porcentaje menor al 26% de los escenarios evaluados. Esto es debido a las colisiones de los paquetes que tiene lugar por los conflictos en la asignación de recursos simultáneos cuando se utiliza el esquema de referencia. De los 74% de los casos que no han encontrado una solución viable para todos los paquetes, en cada escenario

evaluado de media el 6% de los paquetes transmitidos sufren colisión. Según la literatura, todas las aplicaciones industriales definidas en [26] deben garantizar una probabilidad de error igual o inferior a 10^{-7} para que la comunicación sea fiable. Por lo tanto, el porcentaje de error que se obtienen al utilizar el esquema de referencia es mucho más alto que lo permitido. A su vez, O-FAST encuentra un mayor número de soluciones que H-FAST únicamente en el escenario de mayor carga. Sin embargo, la diferencia es tan solo de 6%, mientras que el coste computacional de O-FAST es mucho mayor que el de H-FAST. En resumen, el esquema de *scheduling* H-FAST obtiene un rendimiento muy similar a O-FAST que está basado en optimización y, por tanto, alcanza la solución óptima. Además, H-FAST reduce considerablemente el tiempo de computación, haciendo viable su implementación.

Las propuestas presentadas en el capítulo 4 al querer buscar la mejor solución, no consideran la restricción impuesta por 3GPP, la cual establece un límite máximo de 12 configuraciones CG por UE dentro de un BWP [15]. Como solución, en el capítulo 5 se ha diseñado un nuevo *scheduler* al que se le ha llamado C-FAST, que tiene los mismos objetivos que los *schedulers* definidos en el capítulo 4, pero a su vez, limita el número de configuraciones CG creadas para hacerlo compatible con el estándar actual de 3GPP. Para cumplir con esta limitación, se ha propuesto un vector de activación y desactivación de las configuraciones CG asignadas a un UE. A través de este vector, el UE puede identificar qué configuración CG está activa en cada momento sin necesidad de intercambiar mensajes con el gNB. En consecuencia, este vector disminuye la señalización para activar las configuraciones CG a la vez que los UEs acceden de manera coordinada a los recursos radio evitando conflictos entre sí. En este capítulo se ha propuesto incluir el vector a la fórmula que utilizan los UEs para identificar los recursos radio. Esta fórmula está definida en el estándar de 3GPP [16]. Los resultados obtenidos en el capítulo 5 demuestran que C-FAST puede soportar el mismo número de flujos que el *scheduler* H-FAST, pero cumpliendo con los requisitos de 3GPP ya que el máximo número de configuraciones CG que se crea utilizando C-FAST es menor que 12 en todos los escenarios evaluados, mientras que con H-FAST puede ascender hasta 72 configuraciones CG para un UE. De hecho, en caso de H-FAST, más que el 16% de los UEs obtiene más de 12 configuraciones CG en todos los escenarios evaluados. Es decir, los resultados muestran que el uso de este vector en combinación con C-FAST evita conflictos entre distintos UEs a la hora de acceder a los recursos radio, y al mismo tiempo, mantiene un rendimiento similar al de H-FAST creando una cantidad menor de configuraciones CG, lo que hace posible su implementación conforme al requerimiento de 3GPP. Además, con esta nueva propuesta se disminuye la desviación típica de la latencia experimentada en más de un 33% en comparación con H-FAST y sigue cumpliendo con la latencia media experimentada con un margen del 52% con respecto a la latencia máxima más restrictiva. Finalmente, a pesar de que C-FAST

requiere un coste computacional ligeramente mayor que H-FAST, este coste computacional sigue siendo bajo.

En esta tesis también se ha preparado el simulador ns-3 5G-LENA para posteriormente poder utilizar los esquemas de *scheduling* desarrollados en esta tesis en futuros trabajos que se quieren realizar en Ikerlan en los que se hará uso de la plataforma 5G-LENA. Para ello, se ha implementado el mecanismo *Configured Grant* y se han realizado modificaciones en la implementación de OFDMA en 5G-LENA. Estas modificaciones permiten modelar con mayor precisión la alta flexibilidad ofrecida por 5G NR, lo que facilita la simulación de comunicaciones industriales de bajas latencias en un entorno más realista. Los resultados de latencia obtenidos mediante el uso de CG en 5G-LENA concuerdan con los valores de latencia analíticos obtenidos en el documento de 3GPP TR 37.910 [17]. Por consiguiente, estos resultados validan la implementación de CG en 5G-LENA. Además, los resultados muestran que al utilizar la numerología 0, se reduce en un 93.3% la latencia obtenida con CG en comparación con la latencia obtenida utilizando el *scheduling* dinámico. También se han implementado dos políticas de *scheduling* basadas en CG que demuestran la flexibilidad y la capacidad que tiene 5G NR para soportar servicios críticos industriales. Los resultados muestran que al implementar las políticas de *scheduling* propuestas junto con el modo de acceso OFDMA adaptado, se logra reducir la latencia máxima experimentada por los UEs en un 48.6% en comparación con los esquemas de acceso múltiple previamente implementados en 5G-LENA. El código que se ha implementado en 5G-LENA se encuentra disponible en abierto en [18] y utiliza la versión de ns-3 disponible en [19]. Por lo tanto, esta contribución permite tanto a la comunidad científica como a Ikerlan utilizar el simulador de 5G-LENA y dar continuidad a la línea de investigación abordada en esta tesis.

7.2. Líneas futuras de investigación

Esta tesis doctoral ha demostrado que es posible dar servicios de Industria 4.0 a través de una red integrada 5G-TSN mediante el diseño de esquemas de *scheduling* capaces de gestionar la coexistencia de diferentes flujos TSN. A pesar de los avances realizados en esta tesis, es posible seguir mejorando la integración de las redes 5G y TSN mediante mecanismos de *scheduling*. Un primer paso inmediato sería un esquema de *scheduling* óptimo que minimice tanto la latencia como el número de configuraciones CG creadas. De hecho, sería posible considerar los mismos requisitos que O-FAST, pero la función objetivo se modificaría para incluir la minimización del número de configuraciones CG creadas. De esta manera, se buscaría encontrar una solución que no solo reduzca la latencia, sino también escoja de manera óptima las configuraciones CG creadas y así poder minimizar el *jitter*. Siguiendo la misma línea, otro de los aspectos importantes a estudiar es la posibilidad

de utilizar soluciones de problemas de *scheduling* que no encuentren una asignación para todos los paquetes de todos los flujos TSN, pero que sigan cumpliendo con el tiempo de supervivencia indicado en 3GPP. El tiempo de supervivencia indica el tiempo en el que una aplicación puede mantenerse en funcionamiento sin recibir un paquete [5]. Por lo tanto, con este análisis se podría aumentar el número de soluciones encontradas, ya que si un esquema de *scheduling* no encontrase una asignación de recursos válida para todos los paquetes, pero el número de paquetes que tengan recursos radio asignados fuese suficiente tal que la aplicación pudiese funcionar cumpliendo el tiempo de supervivencia, en ese caso, la solución encontrada por el *scheduling* se daría por buena o satisfactoria.

Además, se debe trabajar en los casos en los que la periodicidad de los flujos TSN no coinciden con las periodicidades admitidas en la RAN 5G. Aunque ya existen algunos estudios que analizan este aspecto tales como [49], [50], [78] y [79], es importante destacar que dichos estudios se centran en soluciones para escenarios en los que se transmite un único flujo TSN. En este contexto, es necesario ampliar la investigación de los esquemas de *scheduling* para abordar esta problemática en escenarios que consideran la transmisión de varios flujos TSN.

También sería interesante trabajar en la problemática descrita en esta tesis para los *schedulings* SPS para transmisiones periódicas en DL. Esta tesis se ha centrado en comunicaciones UL, y para que la integración sea efectiva, es necesario analizar las comunicaciones en ambos sentidos. Con este estudio, se analizaría el funcionamiento de ambos *schedulers* en 5G para UL y DL, simultáneamente.

Otro aspecto importante a estudiar es la selección de los mejores parámetros para la RAN 5G. Mediante este análisis, se buscaría determinar la configuración que favorezca la integración más efectiva de 5G-TSN. En esta tesis, se ha considerado una red 5G con características específicas para la Industria 4.0 siguiendo el documento [67]. Sin embargo, sería interesante analizar el efecto que implica seleccionar diferentes parámetros en la RAN y determinar si al elegir otras características en la RAN, los *schedulers* propuestos podrían soportar más flujos TSN. Es decir, este análisis permitiría evaluar si cambiando la configuración de la RAN es posible aumentar el número de soluciones satisfactorias proporcionadas por cada *scheduler*.

En futuras investigaciones, también se debería de tener en cuenta la retroalimentación entre el *bridge* 5G virtual y el CNC al calcular el *scheduling* de TSN. Actualmente, el estándar de 3GPP especifica que la red 5G transmite la información de la latencia máxima estimada, entre otros datos, del *bridge* 5G virtual al CNC para calcular el *scheduling* de TSN. Sin embargo, es posible que la capacidad real del *bridge* 5G virtual para soportar la

latencia máxima requerida por los flujos TSN no cumpla con el *scheduling* definido por TSN. En los trabajos [6] y [40] se calcula el *scheduling* de TSN conjuntamente a través de un problema de optimización. Sin embargo, realizar el *scheduling* de forma conjunta en ambas redes conlleva un alto coste computacional. Esta tesis se ha enfocado en la coordinación de ambos *schedulers*. El análisis realizado se ha centrado únicamente en una dirección, es decir, se ha considerado el *scheduling* de TSN como dato de entrada y se ha verificado posteriormente si la red 5G puede soportar ese *scheduling* dentro del *bridge* 5G virtual. Como extensión del trabajo, sería interesante analizar el funcionamiento de la retroalimentación. Es decir, una vez que se reciba la información del *scheduling* de 5G, y se verifique si la red 5G puede soportar todos los flujos TSN, en caso de que no sea posible, se debería transmitir la información pertinente al CNC para volver a calcular el *scheduling* de TSN. Por lo tanto, el siguiente paso sería proponer una forma de reajustar el *scheduling* de TSN para poder admitir los flujos que no han recibido recursos radio en la red 5G.

Por otro lado, en esta tesis se han evaluado los esquemas de *scheduling* propuestos en un entorno de simulación desarrollado en Matlab que simula las características del sistema y de la red. El siguiente paso en este estudio sería evaluar el funcionamiento de estos *schedulers* en 5G-LENA junto con otras investigaciones que se quieran llevar a cabo en Ikerlan. Para ello, se partiría de la implementación de *Configured Grant* que se ha realizado en ns-3 5G-LENA durante esta tesis y se tendrían que implementar los *schedulers* propuestos en este entorno de simulación.

Finalmente, las propuestas presentadas en esta tesis podrían ser implementadas y validadas experimentalmente para demostrar su funcionamiento y validez técnica en entornos reales. Por un lado, se podría disponer de dispositivos que proporcionan una red TSN, como por ejemplo [80] o [81], y por otro lado dispositivos *Software Defined Radio* (SDR) en combinación con la plataforma de software de código abierto *Open Air Interface* (OAI) que incluye la pila de protocolos del 3GPP. Esta combinación permitiría emular el funcionamiento en tiempo real de la red integrada de 5G-TSN. Además, se podría emplear un emulador de canal, como el que está descrito en [82], para evaluar estos esquemas de *schedulings* utilizando canales industriales. Este emulador evitaría la necesidad de realizar pruebas de campo directamente en entornos industriales.

Bibliografía

- [1] 5G-ACIA, *5G for Connected Industries and Automation*, 2019.
- [2] 5G-ACIA, *5G for Industrial Internet of Things (IIoT): Capabilities, Features, and Potential*, 2021.
- [3] 5G-ACIA, *Integration of 5G with Time-Sensitive Networking for Industrial Communications*, 2021.
- [4] Platform Industrie 4.0, *Industrie 4.0 Plug-and-Produce for Adaptable Factories: Example Use Case Definition, Models, and Implementation*, white paper, 2017.
- [5] 3GPP TSG RAN, “System Architecture for the 5G System (5GS)”, Rel. 17, TS 23.501, V17.3.0, 2023.
- [6] D. Ginhör, R. Guillaume, J. Von Hoyningen-Huene, M. Schüngel, and H. D. Schotten, “End-to-end Optimized Joint Scheduling of Converged Wireless and Wired Time-Sensitive Networks”, in *Proc. 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 222–229, 2020.
- [7] J. Garcia-Morales, M. C. Lucas-Estañ, and J. Gozalvez, “Latency-Sensitive 5G RAN Slicing for Industry 4.0”, *IEEE Access*, vol. 7, pp. 143139–143159, 2019.
- [8] IEEE, *IEEE 802.1Q: Standard for Local and metropolitan area networks -- Bridges and Bridged Networks*, 2018.
- [9] N. Patriciello, S. Lagen, B. Bojovic, and L. Giupponi, “An E2E Simulator for 5G NR Networks”, *Simulation Modelling Practice and Theory*, vol. 96, article 101933, 2019.
- [10] 5G-ACIA, *A 5G Traffic Model for Industrial Use Cases*, 2019.
- [11] A. Larrañaga, M. C. Lucas-Estañ, I. Martínez, I. Val, and J. Gozalvez, “Analysis of 5G-TSN Integration to Support Industry 4.0”, in *Proc. 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1111–1114, 2020.
- [12] P. Schulz *et al.*, “Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture”, *IEEE Communications Magazine*, vol. 55, issue: 2, pp. 70–78, 2017.
- [13] A. Larranaga, M. C. Lucas-Estan, I. Martínez, and J. Gozalvez, “5G Configured Grant Scheduling for 5G-TSN Integration for the Support of Industry 4.0,” in *Proc. 18th Wireless On-demand Network Systems and Services Conference (WONS)*, pp. 72–79, 2023.
- [14] M. C. Lucas-Estañ, A. Larrañaga, J. Gozalvez, and I. Martínez, “Configured Grant Scheduling for the Support of TSN Traffic in 5G and Beyond Industrial Networks”, accepted for publication in *Proc. 98th IEEE Vehicular Technology Conference (VTC Fall)*, 2023.

- [15] 3GPP TSG RAN, “NR and NG-RAN Overall Description”, Rel. 17, TS 38.300, V17.5.0, 2023.
- [16] 3GPP TSG RAN, “Medium Access Control (MAC) protocol specification”, Rel. 17, TS 38.321, V17.2.0, 2022.
- [17] 3GPP TSG RAN, “Study on self evaluation towards IMT-2020 submission”, Rel. 16, TR 37.910, V16.1.0, 2019.
- [18] The open repository that contains the Configured Grant implementation in 5G-LENA, available at: <https://gitlab.com/ns-3-dev-nr-configuredgrant/nr/-/tree/5g-lena-cg-v2.1.y>
- [19] The open repository that contains the ns-3 version that has to be used with the 5G-LENA CG module, available at: <https://gitlab.com/ns-3-dev-nr-configuredgrant/ns-3-dev/-/tree/ns-3.36-cg>
- [20] A. Larrañaga, M. C. Lucas-Estañ, I. Martinez, and J. Gozalvez, “5G NR Configured Grant in ns-3 Network Simulator for Ultra- Reliable Low Latency Communications”, *Procedia Computer Science*, vol. 201, pp. 495–502, 2022.
- [21] A. Larrañaga, M. C. Lucas-Estañ, S. Lagén, Z. Ali, I. Martinez, and J. Gozalvez, “An open-source implementation and validation of 5G NR configured grant for URLLC in ns-3 5G LENA: A scheduling case study in industry 4.0 scenarios”, *Journal of Network Computer Applications*, vol. 215, article 103638, 2023.
- [22] 3GPP TSG RAN, “Service requirements for cyber-physical control applications in vertical domains”, Rel. 18, TS 22.104, V18.2.0, 2021.
- [23] IEC/IEEE 60802, *Time-Sensitive Networking Profile for Industrial Automation*, draft v1.2, 2020.
- [24] Industrial Internet Consortium (ICC), *Time Sensitive Networks for Flexible Manufacturing Testbed - Characterization and Mapping of Converged Traffic Types*, v1.0, 2019.
- [25] 5G-ACIA, *Key 5G Use Cases and Requirements*, 2020.
- [26] K. Montgomery, R. Candell, Y. Liu, and M. Hany, “Wireless User Requirements for the Factory Workcell”, *NIST Advanced Manufacturing Series*, 300-8, 2020.
- [27] IEEE, *IEEE 802.1AS: Standard for Local and metropolitan area networks--Time and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks*, 2011.
- [28] Silviu S. Craciunas, R. Serna, O. Martin, and C. W. Steiner, “Scheduling real-time communication in IEEE 802.1Qbv time sensitive networks”, in *Proc. 24th IEEE International Conference on Real-Time Networks and Systems*, pp. 183–192, 2016.
- [29] 5G-ACIA, *5G Non-Public Networks for Industrial Scenarios*, 2019.
- [30] 3GPP TSG RAN, “Physical channels and modulation”, Rel. 17, TS 38.211, V17.5.0, 2023.
- [31] 3GPP TSG RAN, “Physical channels and modulation”, Rel. 16, TS 38.211,

- V16.7.0, 2021.
- [32] 3GPP TSG RAN, “Study on New Radio (NR) access technology”, Rel. 17, TR 38.912, V17.0.0, 2022.
 - [33] 3GPP TSG RAN, “Physical layer procedures for data”, Rel. 16, TS 38.214, V16.7.0, 2021.
 - [34] 3GPP TSG RAN, “Radio Resource Control (RRC) protocol specification”, Rel. 16, TS 38.331, V16.6.0, 2021.
 - [35] 3GPP TSG RAN, “Procedures for the 5G System (5GS)”, Rel. 17, TS 23.502, V17.3.0, 2021.
 - [36] 3GPP TSG RAN, “Policy and charging control framework for the 5G System (5GS)”, Rel. 17, TS 23.503, V17.3.0, 2021.
 - [37] 3GPP TSG CN, “Device-Side Time Sensitive Networking (TSN) Translator (DS-TT) to Network-Side TSN Translator (NW-TT) protocol aspects”, Rel. 17, TS 24.535, V17.2.0, 2021.
 - [38] Z. Satka, M. Ashjaei, H. Fotouhi, M. Daneshtalab, and M. Sj, “QoS-MAN : A Novel QoS Mapping Algorithm for TSN-5G Flows”, in *Proc. 28th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, pp. 220–227, 2022.
 - [39] Y. Zhang, Q. Xu, M. Li, C. Chen, and X. Guan, “QoS-Aware Mapping and Scheduling for Virtual Network Functions in Industrial 5G-TSN Network”, in *Proc. 2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2021.
 - [40] D. Ginthor, R. Guillaume, M. Schungel, and H. D. Schotten, “Robust End-to-End Schedules for Wireless Time-Sensitive Networks under Correlated Large-scale Fading”, in *Proc. 17th IEEE International Conference on Factory Communication Systems (WFCS)*, pp. 115–122, 2021.
 - [41] X. Wang, H. Yao, T. Mai, S. Guo, and Y. Liu, “Reinforcement Learning-Based Particle Swarm Optimization for End-to-End Traffic Scheduling in TSN-5G Networks”, *IEEE/ACM Transactions on Networking*, pp. 1–15, 2023.
 - [42] S. Bhattacharjee *et al.*, “Network Slicing for TSN-Based Transport Networks”, *IEEE Access*, vol. 9, pp. 62788–62809, 2021.
 - [43] J. Yang and G. Yu, “Traffic Scheduling for 5G-TSN Integrated Systems”, in *Proc. International Symposium on Wireless Communications Systems (ISWCS)*, pp. 1–6, 2022.
 - [44] D. Ginhör, R. Guillaume, M. Schüngel, and H. D. Schotten, “5G RAN slicing for deterministic traffic”, in *Proc. 2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 0–5, 2021.
 - [45] R. B. Abreu, G. Pocovi, T. H. Jacobsen, M. Centenaro, K. I. Pedersen, and T. E. Kolding, “Scheduling Enhancements and Performance Evaluation of Downlink 5G Time-Sensitive Communications”, *IEEE Access*, vol. 8, pp. 128106–128115, 2020.
 - [46] Y. Zhang, Q. Xu, X. Guan, C. Chen, and M. Li, “Wireless/wired integrated

- transmission for industrial cyber-physical systems: risk-sensitive co-design of 5G and TSN protocols”, *Science China Information Sciences*, vol. 65, article 110204, 2022.
- [47] Y. Cai, X. Zhang, S. Hu, and X. Wei, “Dynamic QoS Mapping and Adaptive Semi-Persistent Scheduling in 5G-TSN Integrated Networks”, *China Communications Magazine*, vol. 20, no. 4, pp. 340-355, 2023.
- [48] D. Krummacker, B. Veith, C. Fischer, and H. D. Schotten, “Analysis of 5G Channel Access for Collaboration with TSN Concluding at a 5G Scheduling Mechanism,” *Network*, vol. 2, no. 3, pp. 440–455, 2022.
- [49] M. J. Yang, S. Y. Lim, S. M. Oh, and J. Shin, “An Uplink Transmission Scheme for TSN Service in 5G Industrial IoT”, in *Proc. International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 902–904, 2020.
- [50] N. Jiang, A. Aijaz, and Y. Jin, “Recursive Periodicity Shifting for Semi-Persistent Scheduling of Time-Sensitive Communication in 5G”, in *Proc. 2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 01-06, 2021.
- [51] Nokia, *5G plug-and-produce. How the 3GPP 5G System facilitates Industrial Ethernet deployments to fuel Industry 4.0 applications*, white paper, 2020.
- [52] M. Khoshnevisan, “5G Industrial Networks With CoMP for URLLC and Time Sensitive Network Architecture”, *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 947–959, 2019.
- [53] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, “Principle and performance of semi-persistent scheduling for VoIP in LTE system”, in *Proc. 2007 International Conference on Wireless Communications, Networking and Mobile Computing (WiMob)*, pp. 2861–2864, 2007.
- [54] 3GPP TSG RAN WG1 #97, “Enhanced UL configured grant transmissions for URLLC”, R1-1906151, Reno, USA, 2019.
- [55] 3GPP TSG RAN WG1, “On enhanced Configured Grant PUSCH for eURLLC”, R1-1812506, Spokane, WA, USA, 2018.
- [56] 3GPP TSG RAN WG1, “Enhanced UL transmission with configured grant for URLLC”, R1-1809165, Gothenburg, Sweden, 2018.
- [57] T. K. Le, U. Salim, and F. Kaltenberger, “Enhancing URLLC Uplink Configured-grant Transmissions”, in *Proc. 93rd IEEE Vehicular Technology Conference (VTC Spring)*, pp.1-5, 2021.
- [58] A. Gunturu, V. S. Tijoriwala, and A. K. R. Chavva, “Optimal Configured Grant Selection Method for NR Rel-16 Uplink URLLC”, in *Proc. 2020 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2020.
- [59] Y. Liu, Y. Deng, M. ElKashlan, A. Nallanathan, and G. K. Karagiannidis, “Optimization of Grant-Free NOMA with Multiple Configured-Grants for mURLLC”, *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1222–1236, 2022.

- [60] Y. Liu, Y. Deng, M. ElKashlan, A. Nallanathan, and G. K. Karagiannidis, “Multiple Configured-Grants Optimization in Grant-Free NOMA for mURLLC Service”, in *Proc. 2022 IEEE International Conference on Communications (ICC)*, pp. 4516–4521, 2022.
- [61] Ye Feng, Lihua Ruan, et al., “Intelligent Radio Resource Allocation for Human-Robot Collaboration”, *IEEE Open Journal of the Communications Society*, vol. 3, pp. 144-158, 2022.
- [62] Y. Feng and E. Wong, “A Predictive Semi-Persistent Scheduling Scheme for Low-Latency Applications in LTE and NR Networks”, in *Proc. 2019 IEEE International Conference on Communications (ICC)*, pp. 1-6, 2019.
- [63] Y. Pan, R. Mahfouzi, S. Samii, P. Eles, and Z. Peng, “Resource Optimization with 5G Configured Grant Scheduling for Real-Time Applications”, in *Proc. Design, Automation and Test Conference and Exhibition (DATE)*, pp. 1–2, 2023.
- [64] A. Nasrallah *et al.*, “Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G Ull research”, *IEEE Communications Surveys and Tutorials*, vol. 21, issue 1, pp. 88–145, 2019.
- [65] D. Pannell, “Choosing the Right TSN Tools to Meet a Bounded Latency”, IEEE SA Ethernet & IP @ Automotive Technology Day, 2019.
- [66] B. Coll-Perales *et al.*, “End-to-End V2X Latency Modeling and Analysis in 5G Networks”, *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 5094–5109, 2023.
- [67] 3GPP TSG RAN, “Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)”, Rel. 16, TR 38.824, V16.0.0, 2019.
- [68] 3GPP TSG RAN, “Base Station (BS) radio transmission and reception”, Rel. 17, TS 38.104, 2021.
- [69] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Viridis, “SiMu5G—An OMNeT++ library for end-to-end performance evaluation of 5G networks,” *IEEE Access*, vol. 8, pp. 181176–181191, 2020.
- [70] Ns-3 website is available at: <https://www.nsnam.org/>
- [71] K. Koutlia, B. Bojovic, Z. Ali, and S. Lagén, “Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios”, *Simulation Modelling Practice and Theory*, vol. 119, article 102580, 2022.
- [72] T. Zugno, M. Polese, N. Patriciello, B. Bojović, S. Lagen, and M. Zorzi, “Implementation of a Spatial Channel Model for ns-3”, in *Proc. of the 2020 Workshop on ns-3 (WNS3)*, pp. 49–56, 2020.
- [73] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, “The impact of NR Scheduling Timings on End-to-End Delay for Uplink Traffic”, in *Proc. 2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2019.
- [74] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, “An improved MAC layer for the 5G NR ns-3 module”, in *Proc. of the 2019 Workshop on ns-3 (WNS3)*, pp.

- 41–48, 2019.
- [75] European 5G Observatory (2020), *5G private licences spectrum in Europe*. Available at: <https://5gobservatory.eu/5g-private-licences-spectrum-in-europe/>
 - [76] 3GPP TSG RAN, “Study on channel model for frequencies from 0.5 to 100 GHz”, Rel. 17, TR 38.901, v17.0.0, 2022.
 - [77] M. C. Lucas-Estañ *et al.*, “An Analytical Latency Model and Evaluation of the Capacity of 5G NR to Support V2X Services using V2N2V Communications”, *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 2293-2306, 2021.
 - [78] 3GPP TSG RAN WG2, “Support for arbitrary TSC message periodicities”, R2-1913453, Chongqing, China 2019.
 - [79] 3GPP TSG RAN WG2, “Solutions to mitigate the periodicity misalignment between TSC traffic and CG/SPS”, R2-1907233, Reno, USA, 2019.
 - [80] SoC-e.com: System on Chip & FPGA IP Core Development company website, available at: <https://soc-e.com/>
 - [81] TTTech company website, available at: <https://www.tttech.com/>
 - [82] I. Val, F. Casado, P. M. Rodriguez, and A. Arriola, “FPGA-based wideband channel emulator for evaluation of Wireless Sensor Networks in industrial environments”, in *Proc. 19th IEEE Emerging Technology and Factory Automation (ETFA)*, pp.1-7, 2014.
 - [83] Z. Satka *et al.*, “Developing a Translation Technique for Converged TSN-5G Communication”, in *Proc. 18th IEEE International Conference on Factory Communication Systems (WFCS)*, pp. 1-8, 2022.
 - [84] Example of a sensor datasheet: MA734 | 8-Bit to 12.5-Bit, 3 μ s Low-Latency Contactless Angle Sensor, available at: <https://www.monolithicpower.com/en/ma734.html>
 - [85] Rockwell Automation, Dataheet of 1756 ControlLogix I/O Modules Specifications Technical Data, available at: https://literature.rockwellautomation.com/idc/groups/literature/documents/td/1756-td002_-en-e.pdf
 - [86] X. Jiang, M. Luvisotto, Z. Pang, and C. Fischione, “Latency Performance of 5G New Radio for Critical Industrial Control Systems” in *Proc. 24th IEEE International Conference on Emerging Technologies Factory Automation (ETFA)*, pp. 1135–1142, 2019.
 - [87] F. P. Mats Norin, Rowan Högma, Mats Buchmayer, Gösta Lemne, *5G spectrum for local industrial networks*, white paper, 2022.

Anexo A. Publicaciones

Esta tesis doctoral se presenta bajo la modalidad de tesis por compendio de trabajos previamente publicados. En este anexo incluimos una copia de las publicaciones especificadas a continuación. El orden en el que están incluidos es el mismo en el que se listan a continuación.

Artículo publicado en revista de impacto, indexada según JCR Science Edition:

- *An open-source implementation and validation of 5G NR configured grant for URLLC in ns-3 5G LENA: A scheduling case study in industry 4.0 scenario*
Ana Larrañaga, M. Carmen Lucas-Estañ, Sandra Lagén, Zoraze Ali, Imanol Martínez, Javier Gozávez
Journal of Network and Computer Applications
Volume 215, 2023, article 103638
ISSN: 1084-8045, Editorial: Elsevier
<https://doi.org/10.1016/j.jnca.2023.103638>

Artículos en congresos internacionales con revisión por pares y comité científico de prestigio, publicados en editoriales de prestigio internacionales:

- *Configured Grant Scheduling for the Support of TSN Traffic in 5G and Beyond Industrial Networks*
M.C. Lucas-Estañ, A. Larrañaga, J. Gozávez, I. Martínez
2023 98th IEEE Vehicular Technology Conference (VTC2023-Fall)
Hong Kong, 2023
Editorial: IEEE (**Aceptado para publicar**)
- *Analysis of 5G-TSN Integration to Support Industry 4.0*
A. Larrañaga, M. C. Lucas-Estañ, I. Martínez, I. Val, J. Gozávez
2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)
Vienna, Austria, 2020
pp. 1111-1114
ISBN: 978-1-7281-8957-4. Editorial: IEEE
doi: 10.1109/ETFA46521.2020.9212141

- *5G Configured Grant Scheduling for 5G-TSN Integration for the Support of Industry 4.0*
A. Larrañaga, M. C. Lucas-Estañ, I. Martínez, J. Gozávez
2023 18th IEEE Wireless On-Demand Network Systems and Services Conference (WONS)
Madonna di Campiglio, Italy, 2023
pp. 72-79
ISBN: 978-3-903176-56-0, Editorial: IEEE
doi: 10.23919/WONS57325.2023.10062219

- *5G NR Configured Grant in ns-3 Network Simulator for Ultra-Reliable Low Latency Communications*
Ana Larrañaga, M. Carmen Lucas-Estañ, Imanol Martínez, Javier Gozávez
Procedia Computer Science
Volume 201, 2022
Pages 495-502
ISSN: 1877-0509, Editorial: Elsevier
<https://doi.org/10.1016/j.procs.2022.03.064>

An open-source implementation and validation of 5G NR Configured Grant for URLLC in ns-3 5G LENA: a scheduling case study in Industry 4.0 scenarios

Ana Larrañaga^a, M. Carmen Lucas-Estañ^b,
Sandra Lagén^c, Zoraze Ali^c, Imanol Martinez^a, Javier Gozalvez^b

^a*HW and Communication Systems Area, Ikerlan Technology Research Centre, Mondragón 20500, Spain*

^b*UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain*

^c*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Barcelona, Spain*

Abstract Factories are undergoing a digital transformation towards more cost-efficient, zero-defect manufacturing. The digitalized factories require communication networks capable of satisfying their strict latency and reliability demands. 5G and beyond networks are being designed to efficiently support services demanding Ultra-Reliable and Low Latency Communications (URLLC). At the MAC level, the use of dynamic scheduling for uplink transmissions entails a non-negligible latency introduced by the signaling messages exchanged to request and inform about the radio resources allocated for each packet transmission. To reduce the transmission latency, 5G defines Configured Grant (CG) for UL transmissions that pre-allocates radio resources to the User Equipments (UEs) and eliminates the need for requesting resources for each transmission. In this context, the availability of 5G NR simulation tools that accurately implement all 5G NR functionalities, and in particular, the technological enablers introduced in 5G NR to support URLLC, is key to analyze the capability of 5G and beyond networks to support time-critical services and research on new solutions. The availability and access to such tools are limited, and to the best of the authors' knowledge, there are currently no open-source 5G NR simulators that implement configured grant in 5G NR. To overcome this issue, this work presents the first implementation of configured grant in an open-source 5G NR simulator. In particular, configured grant has been implemented in the ns-3 5G-LENA system-level simulator, and it is publicly available. To accurately model the flexibility of 5G NR, we have also improved the implementation of Orthogonal Frequency Division Multiple Access (OFDMA) access mode in 5G-LENA according to 5G NR. To validate the implementation of CG and analyze the capability of 5G NR to support time-critical services, we analyze the latency performance that can be achieved using CG with different scheduling policies in Industry 4.0 scenarios. The results show that the latency values achieved with CG in 5G-LENA match with those reported by previous analytical studies. In addition, this study shows the importance of efficiently using radio resources to reduce the latency experienced and meet the requirements of critical services.

Keywords: 5G, Configured Grant, Grant-free scheduler, scheduling, software simulation, URLLC, ns-3, 5G-LENA

1. Introduction

Factories are undergoing a digital transformation towards more cost-efficient, zero-defect manufacturing environments capable of flexibly adapting to changes in production and demand [1]. Industrial applications such as digital twins, motion control, or control-to-control communications demand strict latency and cycle times requirements ranging from milliseconds to microseconds [2] [3]. They require communication networks capable of satisfying their communication requirements. 5G networks are designed to support Ultra-Reliable and Low-Latency Communications (URLLC). In fact, 5G and its future evolution are considered critical enablers for Industry 4.0 paradigm. 5G introduces several mechanisms at the PHY and MAC layers to reduce latency. For example, 5G includes flexible numerologies with different slot durations (from 1 ms to 0.0625 ms) and mini-slot transmissions. In addition to dynamic scheduling, 5G introduces semi-static scheduling to

reduce communication latency. With dynamic scheduling, the base station or gNB informs the UEs about the radio resources used for each packet transmitted in Downlink (DL) before the transmission of each packet. When a UE has a packet to transmit in Uplink (UL), the UE requests resources from the gNB, and the gNB replies with a grant and the information about the allocated resources. This process introduces a non-negligible transmission delay, which is higher in UL due to the higher number of messages exchanged between the UE and the gNB. Semi-static scheduling (referred to as Configured Grant or CG for UL transmissions and Semi-Persistent Scheduling or SPS for DL transmissions) pre-allocates radio resources periodically to UEs [4], and UEs can transmit in the pre-allocated resources as soon as they have data to transmit. Semi-static scheduling eliminates the latency introduced by requesting radio resources for each transmission.

Availability and access to accurate simulation tools are critical to analyze the capability of 5G NR to support critical

industrial services and research new solutions. However, to the best of the authors' knowledge, there are currently no open-source 5G NR simulators that implement configured grant in 5G NR, which is critical for URLLC services. To overcome this issue, this work presents the first implementation of configured grant in an open-source 5G NR simulator. In particular, configured grant has been implemented in 5G-LENA [5]. 5G-LENA is an open-source discrete-event network simulator of the 5G NR based on ns-3 [6]. 5G-LENA implements most of the main 5G NR functionalities, such as the NR frame structures, the numerologies, or Bandwidth Parts (BWPs). However, 5G-LENA does not implement configured grant scheduling. The code of the configured grant implementation for 5G-LENA is publicly available in [7], and it uses the ns-3 version available in [8].

The latency performance achievable with CG strongly depends on the multiple access scheme since it establishes how radio resources can be shared by the UEs and determines the flexibility of the scheduler to allocate radio resources to the UEs. It is then necessary to accurately model the OFDMA multiple access scheme used in 5G NR in order to analyze the actual capability of 5G NR using CG. 5G-LENA implements several multiple access schemes, but none of them models the flexibility of OFDMA that allows simultaneous radio resource allocations in the time and frequency domain. In this context, we have also implemented OFDMA in 5G-LENA according to 5G NR.

To validate the implementation of CG and analyze the capability of 5G NR to support critical industrial services, we consider a scheduling case study in an Industry 4.0 scenario. In particular, this work analyzes the latency performance that can be achieved using CG with different scheduling policies. The results show that the latency values achieved with CG in 5G-LENA match with those reported by previous analytical studies. In addition, the results highlight the importance of making efficient use of radio resources to reduce the latency experienced and meet the requirements of critical services.

The rest of the paper is organized as follows. 5G New Radio is presented in Section II, and Section III describes the 5G-LENA network simulator. Section IV presents the implementation of CG in 5G-LENA. Section IV also presents the modifications done in 5G-LENA to accurately model OFDMA and the scheduling policies used to evaluate the performance of CG. Section V describes the evaluation scenario and the schemes used as a reference. In Section VI, we derive analytical expressions of the maximum latency experienced with CG and the different scheduling policies. We compare in Section VI the latency achieved analytically and by simulation to validate the implementation of CG. Section VII presents the performance results. Section VIII concludes the paper.

2. 5G New Radio

5G uses OFDM (Orthogonal Frequency Division Multiplexing). Radio resources are organized in Resource Blocks (RBs) in the frequency domain and in slots in the time domain (a slot consists of 14 or 12 OFDM symbols when the normal or extended cyclic prefix is used, respectively). A radio resource is composed of an RB, which contains 12 consecutive subcarriers in the frequency domain and a single OFDM symbol in the time domain. While the Subcarrier Spacing (SCS) is fixed at 15 kHz in Long Term Evolution (LTE), 5G defines multiple numerologies μ that allow the use of different SCSs. A numerology μ is given by the use of an SCS and slot duration. Numerologies μ from 0 to 4 use an SCS equal to 15, 30, 60, 120, and 240 kHz SCS, respectively [9]. The slot duration is given by $1/2^\mu$ ms, which results in slot durations from 1 ms for $\mu=0$ to 0.0625 ms for $\mu=4$ (in LTE, the 14 symbols-slot duration is always equal to 1 ms). Slots are organized in frames of 10 ms. A slot consists of 14 OFDM symbols in the time domain when the normal Cyclic Prefix (CP) is used (the normal CP can be used with all the numerologies). An extended CP can also be used with $\mu=2$. In this case, a slot consists of 12 OFDM symbols. Numerologies 0, 1, and 2 can be used in the lower frequency range (410 MHz-7.125 GHz), and numerologies 2, 3, and 4 can be used in the higher frequency range (24.25 GHz-52.6 GHz). 5G NR allows transmissions to start at any OFDM symbol within a slot and to use only the number of symbols needed for the transmission. This results in mini-slot transmissions when transmissions only use part of the symbols of a slot (mini-slot transmissions can use between 1 and 13 OFDM symbols in UL and between 2 and 13 OFDM symbols in DL [10]), or full-slot transmissions when all the symbols are used (LTE only considers full-slot transmissions). The introduction of multiple numerologies with higher SCS and lower slot durations, and the use of mini-slot transmissions allow to considerably reduce the latency compared to LTE.

5G can use Frequency-Division Duplex (FDD) or Time-Division Duplex (TDD) modes. On the one hand, TDD provides high flexibility as each slot in a frame can be configured for UL or DL transmissions. Furthermore, a single slot can be split into segments of consecutive symbols that can be used for UL or DL [11]. On the other hand, FDD organizes UL and DL transmissions on separate frequencies. FDD can reduce communication latency since resources are always available simultaneously for UL and DL transmissions.

5G NR uses Low Density Parity Check coding with Quadrature Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (QAM), 64 QAM, or 256 QAM for data channels. 5G NR defines three Modulation and Coding Scheme (MCS) tables that provide a different trade-off between spectrum efficiency and protection against error. MCS tables 1 and 2 in [12] guarantee a Block Error Rate (BLER) of 10%, and MCS table 3 in [12] guarantees a BLER

of 10^{-5} when used according to the Channel Quality Indicator (CQI).

5G NR uses OFDMA multiple access scheme that allows UEs to share radio resources in the frequency and time domains. With OFDMA, an RB can be used by different UEs in different OFDM symbols, and different RBs in the same OFDM symbol can be assigned to different UEs. As a result, UEs can receive any number of OFDM symbols and RBs. Transmissions in 5G NR can be performed using dynamic or semi-static scheduling in both UL and DL [13]. With dynamic scheduling, the gNB allocates resources dynamically for each packet transmission. When a packet is generated in DL for a UE, the gNB sends a control message to inform the UE about the allocated radio resources. When a UE needs to transmit a packet in UL, the UE sends a scheduling request (SR) to the gNB. The gNB then replies with a grant message to the UE with the information about the radio resources to use for the packet transmission. The UE transmits the packet together with the buffer status report (BSR) to inform the gNB if it has more data to transmit. If this is the case, the gNB will allocate radio resources for a new transmission and will inform the UE. This process is repeated while the UE has pending data to transmit. Dynamic scheduling makes efficient use of radio resources. However, the signaling exchange between the UE and the gNB before the packet transmission can increase the latency of the transmission. This is more critical for UL transmissions since the amount of signaling between the UE and the gNB is higher. This cannot be appropriate for latency-critical industrial applications. Semi-static scheduling can also be used for packet transmission in 5G NR: SPS in DL and CG in UL. CG and SPS pre-assign radio resources periodically to the UEs before the data packets are generated. When a data packet is generated, the packet can be transmitted in the pre-allocated resources. CG and SPS avoid the signaling exchange between the UE and the gNB to request/inform about the allocated radio resources, reducing the transmission latency. The gNB estimates the radio resources necessary to support the UE based on the characteristics of the traffic it will transmit (e.g., data packet size and periodicity). If the gNB has sufficient resources available to support the UE, it allocates the radio resources and informs the UEs about the pre-allocated radio resources and their periodicity using Radio Resource Control (RRC) signaling during the connection setup. A connection is not established if the gNB does not have enough radio resources to satisfy the request of a UE. Two types of CG are defined in [13]: type 1, which activates the configured uplink grant from the moment it is configured via RRC signaling, and type 2, where the grant is activated or deactivated using DL control messages. CG type 1 is illustrated in Figure 1. We should note that if the characteristics of the traffic generated by the UE change (e.g., its size or periodicity), the configured UL grant should be modified, and this can be done through RRC

signaling procedures defined in 3rd Generation Partnership Project (3GPP) [14].

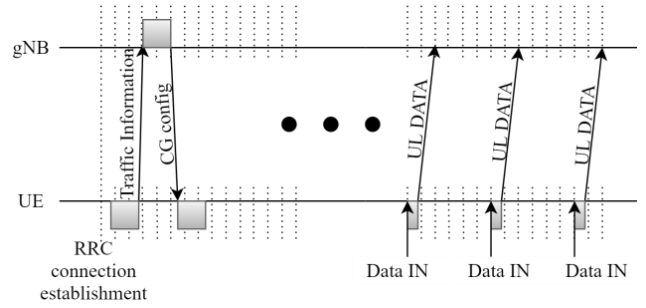


Figure 1. Configured grant type 1.

3. 5G-LENA network simulator

3.1. 5G-LENA overview

5G-LENA is an open-source discrete-event network simulator of the 5G NR [5] implemented over ns-3. ns-3 is an open-source C++ simulation environment for networking research that offers a solid simulation core that supports research on both IP and non-IP-based networks. 5G-LENA is the evolution of LENA that was initially developed to implement the Radio Access Network (RAN) and the core network of LTE. The efforts on the development of 5G-LENA focused first on the RAN. 5G-LENA implements the fundamental PHY-MAC NR features in line with the NR specifications [13], and it has been calibrated in indoor hotspots [5] and outdoor 3GPP reference scenarios [15].

5G-LENA implements two main C++ classes to simulate the functionalities of the gNB (NrGnb class) and the UE (NrUe class) (see Figure 2). For the gNB and the UE, 5G-LENA models the different layers of the protocol stack: PHY, MAC, Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), and RRC. The upper layers RLC, PDCP, and RRC currently rely on LTE LENA implementation. The PHY and MAC layers implement the main NR features. For

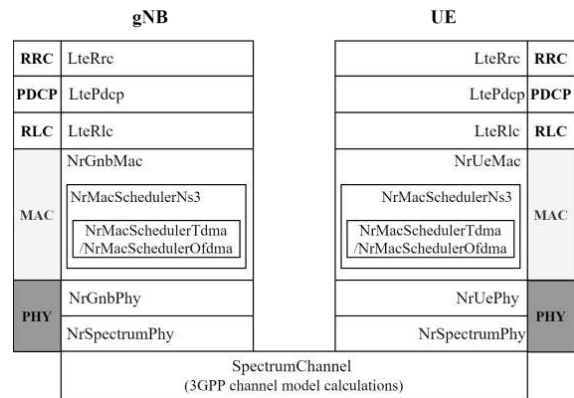


Figure 2. Logical representation of the 5G-LENA RAN class.

example, the PHY layer implements the flexible frame structure defined in [13] for 5G NR. Both TDD and FDD duplexing modes can be configured. In TDD, slots can be flexibly configured for DL or UL transmissions as established by 5G 3GPP standards. 5G-LENA also allows flexible configuration of the number of OFDM symbols within a slot to be used for UL and DL transmissions. The 5G-LENA PHY layer considers the use of the new numerologies defined in [13] for 5G NR, both normal and extended CP, and the different modulation and coding schemes defined in [12]. 5G-LENA provides a realistic implementation of K1 and K2 scheduling timings defined for 3GPP 5G NR. The K1 scheduling timing corresponds to the time between DL data reception at the UE and the corresponding HARQ-ACK feedback transmission to the gNB, while K2 is the delay between UL grant reception at the UE and the corresponding UL data transmission to the gNB. 5G-LENA allows the configuration of such parameters to simulate UEs of different capabilities and processing times [16]. Two beamforming

methods have been added, long-term covariance matrix and beam-search, using either ideal beamforming or realistic beamforming based on uplink Sounding Reference Signals [17]. Recently, dual-polarized Multiple-Input Multiple-Output (MIMO) for the downlink has been implemented [18]. 5G-LENA also supports different channel and propagation models according to the 3GPP spatial channel model defined in TR 38.901 [19].

3.2. Multiple access and scheduling in 5G-LENA

5G-LENA has also evolved the MAC to support 5G NR. 5G LENA implements three multiple access schemes referred to as 5GL-TDMA, 5GL-OFDMA, and 5GL-OFDMA with variable TTI [20]¹. Figure 3 shows how radio resources are allocated to several UEs when different multiple access schemes are used. It depicts the time-frequency resource grid for a slot of 14 OFDM symbols in time and R_{BW} number of RBs in the frequency domain. Squares in Figure 3 represent

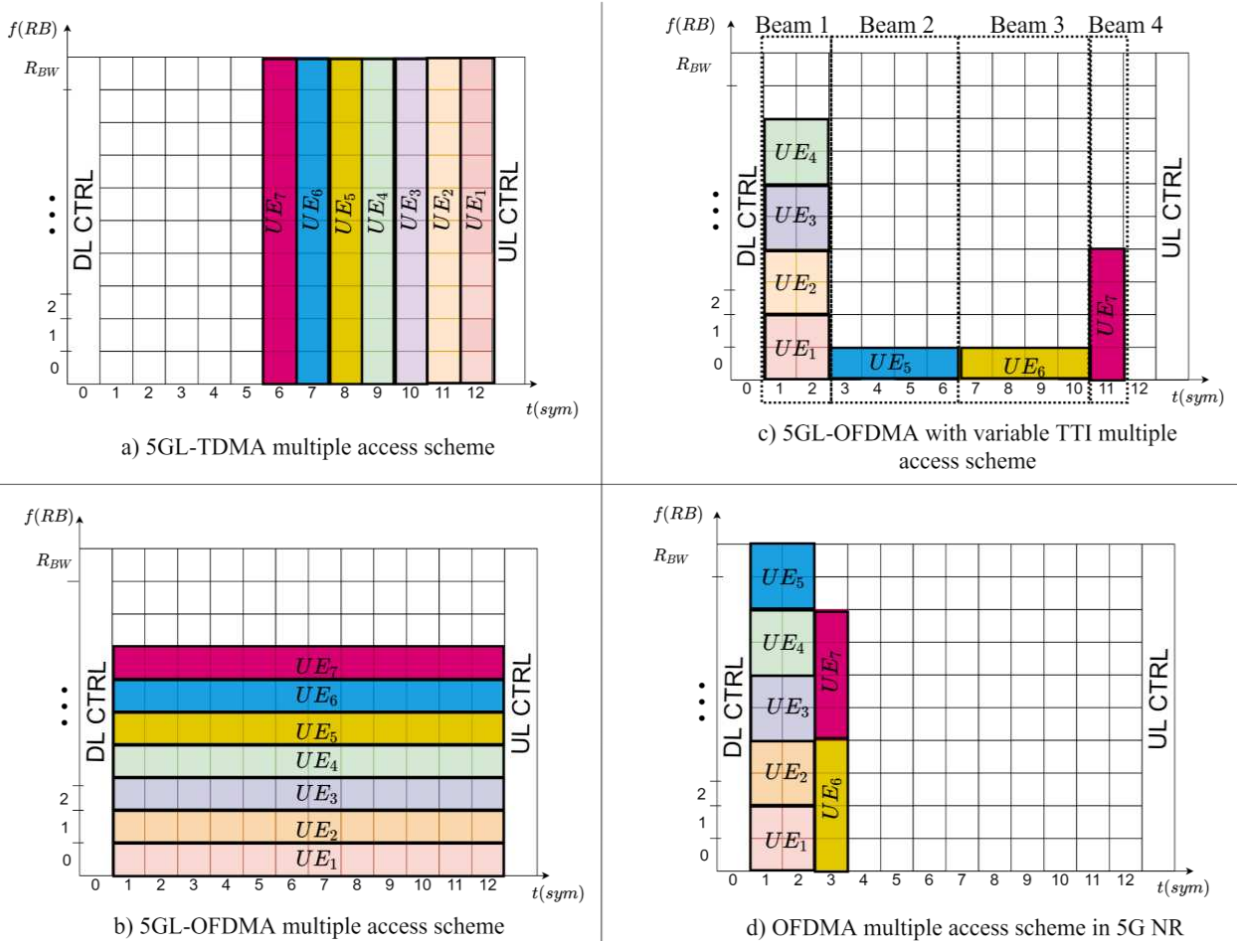


Figure 3. Example of radio resource allocation with the different multiple access modes.

¹ Within 5G-LENA, the implemented multiple access schemes are referred to as TDMA, OFDMA and OFDMA with variable TTI. We added the '5GL-' prefix to differentiate between the particular implementation of the multiple access schemes in 5G-LENA and the TDMA and OFDMA multiple access schemes defined in 5G NR (as they may have some differences).

an OFDM symbol in the time domain and one RB in the frequency domain. The first and last symbols are reserved for DL and UL control channels. In this example, each UE needs 4 RBs to transmit its packet. Colored squares represent the RBs assigned to UEs for packet transmission. With 5GL-TDMA, UEs access radio resources at different OFDM symbols, and all RBs in a symbol must be assigned to the same UE (see Figure 3.a). Although UEs only require 4 radio resources to transmit their packets, they receive all the RBs (R_{BW}) in an OFDM symbol when 5GL-TDMA is used. 5GL-OFDMA is a constrained version of OFDMA defined in 5G NR. 5GL-OFDMA allocates an RB in all the OFDM symbols within a slot to the same UE. This is shown in Figure 3.b. 5G-LENA implements a modified version of 5GL-OFDMA referred to as 5GL-OFDMA with variable TTI. It allows dividing the OFDM symbols within a slot into two or more segments with a different or equal number of OFDM symbols. The radio resources within each of these segments can be accessed by the UEs assigned to the same antenna beam. In each of these segments, 5GL-OFDMA is applied (see Figure 3.c).

None of these multiple access schemes accurately models the flexibility offered by OFDMA in 5G NR. OFDMA allows allocating different numbers of RBs and OFDM symbols to UEs. An example is shown in Figure 3.d. The 5GL-TDMA and 5GL-OFDMA schemes are constrained to allocate a number of radio resources that is multiple of R_{BW} (the number of available RBs in a particular bandwidth) or S_{slot} (that represents the number of OFDM symbols reserved for UL data transmission within a slot), respectively. In the case of 5GL-OFDMA with variable TTI, the number of radio resources allocated to UEs is multiple of the number of OFDM symbols within a segment. OFDMA in 5G NR allows adjusting the number of assigned resources more accurately to the UEs' demand. This enables a more efficient use of radio resources.

5G-LENA implements dynamic scheduling in DL and UL; semi-static scheduling is not yet implemented in 5G-LENA. Different scheduling policies are currently implemented, such as round-robin, proportional fair, etc., but other policies can also be included. It is important to note that a UE receives a number r of RBs in s consecutive OFDM symbols, i.e., a UE receives $r \times s$ radio resources. Due to interference model limitations in 5G-LENA, if two UEs receive RBs in the same OFDM symbol, the two UEs must receive RBs in the same number s of OFDM symbols [5]. Finally, 5G-LENA allocates radio resources for UL transmissions from the last symbol to the first within a slot.

4. Implementation of Configured Grant in 5G-LENA

4.1. Configured Grant

As presented in previous sections, 5G-LENA does not implement semi-persistent scheduling, which limits its applicability for the study of critical use cases. In this

context, 5G-LENA has been extended in this work to simulate UL transmissions using configured grant. We have implemented configured grant Type 1 that configures and stores the UL grant at the session establishment. Configured grant Type 1 avoids potential delays that might be introduced in the activation/deactivation of the configured UL grant. Therefore, it is more suitable for services with stringent latency requirements.

Configured grant has been implemented in 5G-LENA using a state machine that is presented in Figure 4. The state machine is implemented at the UE to manage UL transmissions. INACTIVE_CG is the initial state and is activated for each UE when the simulation starts. The UE informs the gNB about characteristics of the data traffic it intends to transmit, in particular, the size and periodicity of the data packets. To this end, the UE prepares a message with the information about the characteristics of the data traffic and the UE state changes to TO_SEND_TrafficInfo. The UE transmits the prepared message to the gNB in the next slot, and changes to TO_RECEIVE_CG state. When the gNB receives information, it checks the necessary radio resources to support the UE transmissions, and if they are available, it allocates them following the selected scheduling policy. This process is performed within the NrMacSchedulerNs3 class in the DoScheduleUIData function (Figure 5). This function executes the radio resource allocation process for configured UL grant transmissions. Figure 5 illustrates the operation of the DoScheduleUIData function. This function first checks if there are UEs waiting to receive a configured UL grant. If this is the case, the gNB decides the number of OFDM symbols and RBs that should be allocated to each UE. This decision is implemented in AssignULRBG function. Once the decision for radio resource allocation has been made,

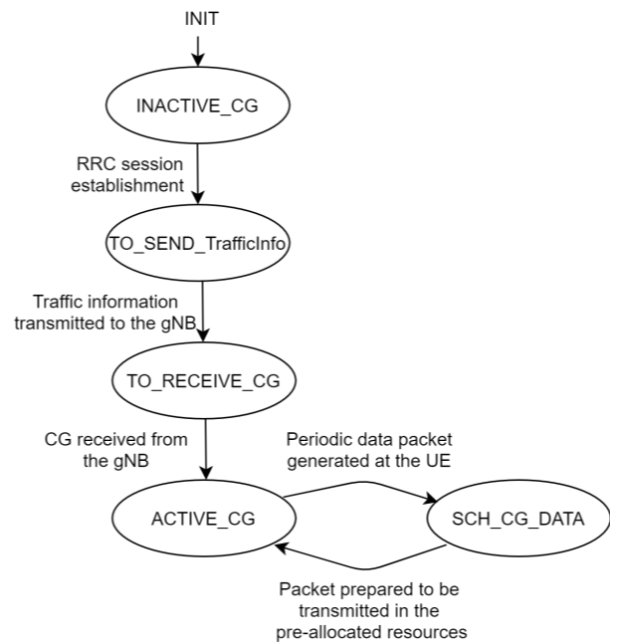


Figure 4. Configured grant state machine.

CreateUICGConfig function is called to generate the *ConfiguredGrantConfig* message [14] with the configured UL grant and information about the pre-allocated radio resources. This message is a new structure that stores the traffic periodicity, the packet size, the MCS, and the resources that have been allocated to the corresponding UE. The gNB transmits the *ConfiguredGrantConfig* message to the UE. When the UE receives the message, it stores the configured UL grant and changes to ACTIVE_CG state. The UE checks if it is in ACTIVE_CG state (that means it has a configured UL grant) each time a new data packet is generated. If this is the case, the UE changes to SCH_CG_DATA. The UE prepares the data packet to be transmitted and changes to ACTIVE_CG state. The packet is then transmitted to the gNB on the pre-allocated resources. The gNB waits to receive data packets from the UE on the pre-allocated resources. It is important to note that the implementation of configured grant in 5G-LENA has also entailed the extension of several functions at the PHY layer class to allow periodic scheduling (for example, StartSlot function of the UE PHY class has been extended to store the *ConfiguredGrantConfig* message in the UE PHY layer in a periodic way).

4.2. Flexible multiple access and radio resource allocation

Flexibility in radio resource allocation is constrained by the multiple access scheme used in the system. OFDMA in 5G NR allows the scheduling of UEs in the time and frequency domain. However, none of the multiple access schemes implemented in 5G-LENA implements the real capabilities of OFDMA in 5G NR. The 5GL-TDMA, 5GL-OFDMA, and 5GL-OFDMA with variable TTI multiple access schemes assume different constraints for scheduling radio resources in frequency or time among different UEs (see Section III). The constraints introduced by the multiple access schemes implemented in 5G-LENA may lead to inefficient use of radio resources. UEs may receive a higher number of radio resources than they actually need. This is especially the case when traffic is characterized by small packet sizes. This can result in higher latencies that can be detrimental to time-critical services.

We have extended the 5G-LENA MAC layer to implement a more accurate version of OFDMA in 5G NR. The objective is to provide an open-source software tool that accurately simulates the actual features and capabilities of 5G NR to support latency-critical services. The new OFDMA implemented in 5G-LENA allows allocating r RBs in s consecutive OFDM symbols to a UE, with r and s being any integer number in $[1, R_{BW}]$ and $[1, S_{slot}]$, respectively (s is always equal to S_{slot} with 5GL-OFDMA, and r is equal to R_{BW} with 5GL-TDMA).

OFDMA has been implemented in the NrMacSchedulerOfdma class (see Figure 2). This class contains AssignULRBG and CreateUICGConfig functions. DoScheduleUIData function in the MAC layer of the gNB in 5G-LENA (Figure 5) is responsible for calling these functions. As mentioned in the previous subsection, AssignULRBG function decides the number of OFDM symbols and RBs that should be allocated to each UE. The number of OFDM symbols and RBs allocated to each UE depends on several factors. First, it depends on the amount of data to be transmitted by the UE and the MCS to use in the packet transmission. The decision also depends on the multiple access scheme used. In this context, the AssignULRBG function included in 5G-LENA has been modified to allow the allocation of radio resources among different UEs simultaneously in frequency and time according to OFDMA in 5G NR. The final radio resource allocation depends on the scheduling policy implemented.

Finally, we have also modified the order followed in 5G-LENA to allocate radio resources for UL transmissions within a slot. As presented in section III, radio resources are allocated from the last to the first symbol within a slot in 5G-LENA. This order is not established in 5G NR. Therefore, we have eliminated this condition. Resources can now be allocated from the first to the last symbol within a slot, which is also key to guarantee very low latency.

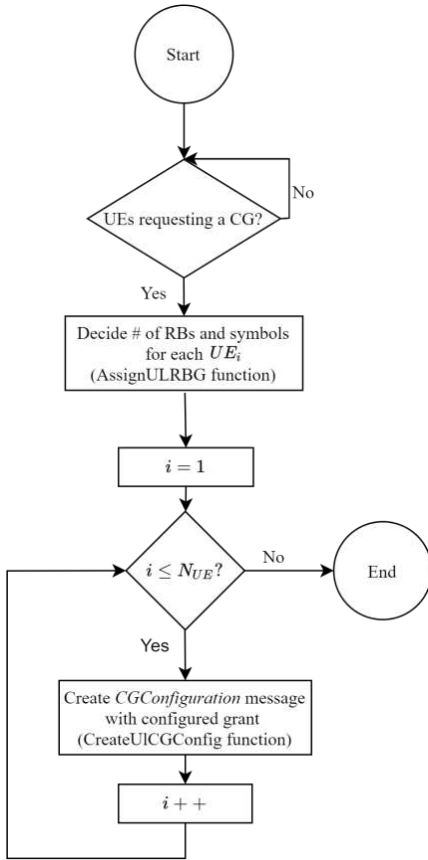


Figure 5. Radio resource allocation process for UL transmissions (DoScheduleUIData function) in 5G-LENA.

4.3. Scheduling policies

Two new scheduling policies have been implemented in 5G-LENA to be applied with CG. These scheduling policies exploit the flexibility of OFDMA in 5G NR. The designed scheduling policies decide the number r_i of RBs and the number s_i of OFDM symbols allocated to each UE $_i$, with $i \in \mathbb{N}$ and $i \in [1, N_{UE}]$. Both scheduling schemes aim to use radio resources efficiently and minimize the experienced latency. To this end, they try to allocate the lowest number of radio resources that satisfies the radio resource demand of each UE. The defined scheduling policies aim to show the importance of accurately emulating the capabilities and flexibility of 5G NR, and they do not search for optimal solutions. This flexibility can be exploited to increase performance and more efficiently support latency-critical services.

The first scheduling policy minimizes the number of OFDM symbols allocated to each UE, referred to as Sym-OFDMA scheduler. The operation of Sym-OFDMA is presented in Algorithm I. Sym-OFDMA serves UEs following a first-come, first-served basis, i.e., from UE $_1$ to UE $_{N_{UE}}$ (line 3 in Algorithm I). Sym-OFDMA allocates radio resources from the first to the last symbol within a slot (variables n_{slot} , n_s , and n_{RB} represent the slot, symbol within the slot, and RB, respectively, that is currently being allocated to a UE). Each UE $_i$ demands d_i radio resources that is calculated as a function of the size of the packet to transmit and the MCS to use in the packet transmission. If the number d_i of radio resources demanded by a UE $_i$ is equal to or lower than R_{BW} (the number of RBs in a symbol), UE $_i$ receives d_i consecutive RBs in an OFDM symbol, i.e., $r_i = d_i$ and $s_i = 1$ (see lines 4-5 in Algorithm I). If the UE $_i$ demands more than R_{BW} radio resources ($d_i > R_{BW}$), UE $_i$ receives R_{BW} RBs in $\lceil d_i / R_{BW} \rceil$ consecutive OFDM symbols (line 7 in Algorithm I). $r_{ini,i}$, $s_{ini,i}$ and $slot_i$ represent the first RB, OFDM symbol, and slot, respectively, allocated to UE $_i$ (see line 15). UE $_{i+1}$ will receive RBs in the same OFDM symbol as UE $_i$ if the number of unallocated RBs are enough to satisfy the demand of UE $_{i+1}$ (see lines 15-16). Otherwise, UE $_{i+1}$ will receive RBs in the next OFDM symbol (lines 9-11). Sym-OFDMA also considers that a UE can only receive RBs within a slot. In case there are not enough unallocated RBs and symbols to meet d_i in the current slot, the UE $_i$ will be served in the next slot (lines 12-14). Figure 6.a shows an example of the radio resource allocation done by Sym-OFDMA in a scenario where 7 UEs demand 4 radio resources each to transmit using configured grant. Figure 6 represents the radio resources in one slot of 14 OFDM symbols, where the first and last symbols are reserved for control signals, and bandwidth is divided in $R_{BW}=10$ RBs.

ALGORITHM I: SYM-OFDMA

1. **Input:** $d_i \forall i \in [1, N_{UE}]$
2. $n_{slot}=1, n_s=1, n_{RB}=1$
3. **For** $i=1$ to N_{UE}
4. **If** $d_i \leq R_{BW}$
5. $r_i=d_i, s_i=1$
6. **Else**
7. $r_i=R_{BW}, s_i=\lceil d_i / R_{BW} \rceil$
8. **End If**
9. **If** $R_{BW} - n_{RB} + 1 < r_i$
10. $n_{RB}=1, n_s = n_s + s_i$
11. **End If**
12. **If** $n_s + s_i - 1 > S_{slot}$
13. $n_{slot} = n_{slot} + 1, n_s = 1, n_{RB} = 1$
14. **End If**
15. $r_{ini,i} = n_{RB}, s_{ini,i} = n_s, slot_i = n_{slot}$
16. $n_{RB} = n_{RB} + r_i$
17. **End For**

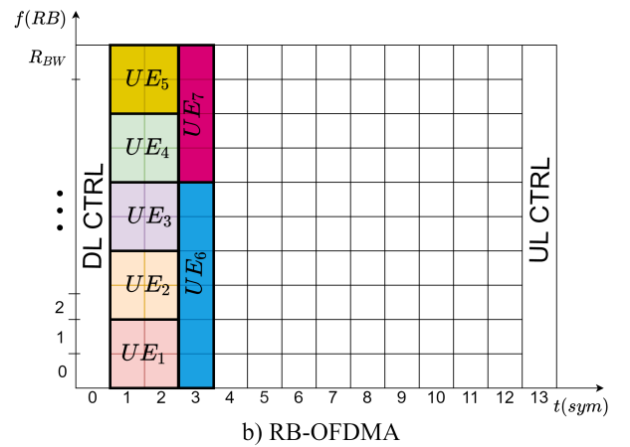
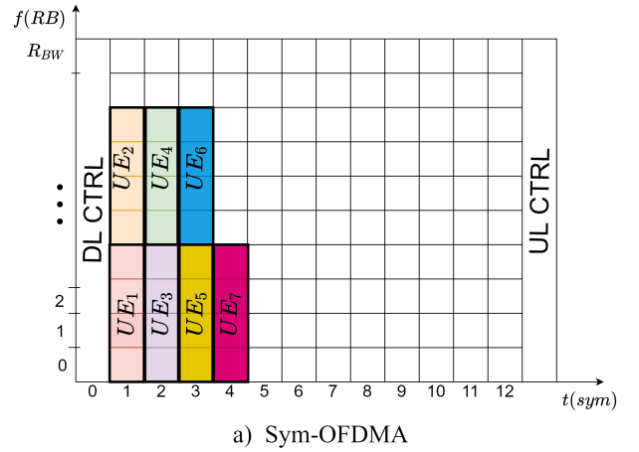


Figure 6. Radio resource allocation with the designed scheduling policies.

The second scheduling policy calculates the number of radio resources (r_i) and OFDM symbols (s_i) that should be allocated to each UE $_i$ to minimize the number of RBs that are not allocated to UEs. It also establishes that each UE $_i$ has to receive at least r_{min} RBs (r_{min} can take any integer number² between 1 and R_{BW}). This second scheduling policy is referred to as RB-OFDMA. The operation of RB-OFDMA is presented in Algorithm II. RB-OFDMA also serves UEs following a first-come, first-served order, and allocates radio resources from the first to the last symbol within a slot. RB-OFDMA distributes the R_{BW} RBs in a symbol among the maximum number of UEs (n_{UE}) considering that $r_i \geq r_{min}$ for each UE $_i$, i.e., $n_{UE} = \lfloor R_{BW}/r_{min} \rfloor$ (line 3 of Algorithm II). The UEs are divided in sets of n_{UE} UEs (line 5); the last set of UEs can have less than n_{UE} . The UEs in each set ϕ will share the RBs in the same OFDM symbols. If the number of UEs in a set ϕ is equal to n_{UE} , each UE $_i$ in ϕ will receive $r_i = r_{min}$ RBs (lines 7 and 8). If the number of UEs in a set ϕ is lower than n_{UE} , the first $((n_{UE}/n) - \lfloor n_{UE}/n \rfloor) \cdot n$ UEs in ϕ will receive $r_i = \lfloor n_{UE}/n \rfloor \cdot r_{min}$ RBs (line 10). The rest of UEs in ϕ will receive $r_i = \lfloor n_{UE}/n \rfloor \cdot r_{min}$ RBs (line 11). Based on the number of RBs allocated to each UE, RB-OFDMA calculates the number d_i^S of OFDM symbols that each UE $_i$ needs to meet its demand d_i

ALGORITHM II: RB-OFDMA

1. Input: $d_i \forall i \in [1, N_{UE}]$, r_{min}
 2. $n_{slot} = 1$, $n_s = 1$, $n_{RB} = 1$, $n_{UEwithRB} = 0$
 3. Define $n_{UE} = \lfloor R_{BW}/r_{min} \rfloor$
 4. **While** there are UEs without resources
 5. Create set ϕ with UE $_i \forall i$ such that
 $i = [n_{UEwithRB} + 1, \min(n_{UEwithRB} + n_{UE}, N_{UE})]$
 6. Define $n =$ number of UEs in ϕ
 7. **If** $n = n_{UE}$
 8. $r_i = r_{min} \forall$ UE $_i$ in ϕ
 9. **Else**
 10. $r_i = \lfloor \frac{n_{UE}}{n} \rfloor \cdot r_{min}$ for the first $(\frac{n_{UE}}{n} - \lfloor \frac{n_{UE}}{n} \rfloor) \cdot n$ UEs in ϕ
 11. $r_i = \lfloor \frac{n_{UE}}{n} \rfloor \cdot r_{min}$ for the last $n - (\frac{n_{UE}}{n} - \lfloor \frac{n_{UE}}{n} \rfloor) \cdot n$ UEs in ϕ
 12. **End If**
 13. **For** all UE $_i$ in ϕ
 14. $d_i^S = d_i / r_i$, $r_{ini,i} = n_{RB}$, $n_{RB} = n_{RB} + r_i$
 15. **End For**
 16. $d^S = \max \{d_i^S\} \forall$ UE $_i$ in ϕ
 17. $s_i = d^S \forall$ UE $_i$ in ϕ
 18. **If** $n_s = n_s + d^S - 1 > S_{slot}$
 19. $n_{slot} = n_{slot} + 1$, $n_s = 1$
 20. **End if**
 21. $s_{ini,i} = n_s$, $s_{lot,i} = n_{slot} \forall$ UE $_i$ in ϕ
 22. $n_s = n_s + d^S$, $n_{RB} = 1$
 23. $n_{UEwithRB} = n_{UEwithRB} + n_{UE}$
 24. **End While**
-

(line 14). Once d_i^S is known, all UEs in a group ϕ receive the number of RBs necessary to satisfy the UE that requires a greater number of OFDM symbols (lines 16-17). RB-OFDMA also considers that a UE can only receive RBs within a slot. In case there are not enough unallocated RBs and symbols to meet r_i and s_i in the current slot, the UE $_i$ will be served in the next slot (lines 18-20). Figure 6.b shows an example of the radio resource allocation done with RB-OFDMA. The example considers that 7 UEs demand 4 radio resources to transmit their data using configured grant and $r_{min} = 2$. In this example, the N_{UE} UEs are divided into two sets of UEs. The first set includes 5 UEs (equal to $\lfloor R_{BW}/r_{min} \rfloor$) that receive $r_i = 2$ RBs and $s_i = 2$ OFDM symbols. The second set includes 2 UEs that receive $r_i = 4$ RBs and $r_i = 6$ RBs for UE $_7$ and UE $_6$, respectively, and $s_i = 1$ OFDM symbol.

The value of r_{min} can be tuned to optimize radio resource efficiency. In this work, we calculate r_{min} with the aim of minimizing the number of RBs not allocated to any UE. To this end, we search for the value of r_{min} that minimizes the number of not allocated RBs in a symbol (lines 1-6 of Algorithm III). If several values satisfy this condition, we select the value that provides the lowest difference between the number of allocated radio resources to UEs ($r_i \cdot s_i = x \cdot d_i^S(x)$) and their demands (line 8).

ALGORITHM III: r_{min} in RB-OFDMA

1. $aux = R_{BW}$
 2. Create $X = \{x\}$ with $x \in \mathbb{N}$, $1 < x < R_{BW}$, and $\text{mod}(aux/x) = 0$.
 3. **If** X is empty
 4. $aux = aux - 1$
 5. Go to 3
 6. **End If**
 7. Calculate number $d_i^S(x)$ of OFDM symbols needed to meet the UE demands d_i when UEs receive x RBs, $\forall i \in [1, N_{UE}]$ and $\forall x \in X$.
 8. Set $r_{min} = x \in X$ that satisfies $\min_{i,x} (x \cdot d_i^S(x) - d_i)$
-

5. Evaluation Scenario & Reference Schemes

We consider an evaluation scenario where a single 5G NR cell covers a typical work cell of 10 x 10 m² where a closed-loop control application is implemented. The cell is assigned a bandwidth of BW in the 3.7–3.8 GHz band³, with BW equal to 10, 20, or 40 MHz, and operates in TDD mode [21]. We consider slots with 14 OFDM symbols and evaluate 3 slot configurations for which the last 13, 9, and 5 OFDM symbols are used for UL and the rest of OFDM symbols for DL. We refer to the different configurations as 1D13U, 5D9U, and 9D5U, respectively. The first and last symbols within a slot are reserved for the transmission of the control channels in DL and UL, respectively. We consider the use of 30 kHz SCS as

² r_{min} can be configured based on traffic characteristic to optimize system performance.

³ The 3.7–3.8 GHz band is considered in some European countries for non-public network deployments [28].

recommended in [22] for industrial environments (we also evaluate the use of 15 and 60 kHz when specifically indicated). There are $N_{UE}=15$ sensors randomly distributed in the work cell, and we consider that they are in the Line of Sight with the gNB. The radio channel is characterized by fast-fading and shadowing, and they follow a spatial channel model defined in [23]. Following [24], sensors generate periodic data packets that are transmitted to a central monitoring system. All sensors participating in the closed-loop control application generate data packets simultaneously. Data packets are characterized by a size p_i that is set equal to 10 or 25 bytes and a periodicity equal to 10 ms [24]. Sensors transmit the data packets in UL to the gNB towards the central monitoring system. The IPv4 header with a value of 22 bytes is added to this packet size, $p_{headers}$. We also consider the use of cyclic redundancy check (CRC) code. All sensors use MCS 12 in MCS table 1 [12] for the periodic traffic transmissions; MCS 12 provides a good trade-off between robustness and transmission rate in the considered scenario. Furthermore, we consider the use of one MIMO transmission layer (referred to as v). Processing times in the UE and gNB are calculated as indicated in [25] and [12]. The values for the different evaluation parameters are summarized in Table 1⁴.

Table 1. Evaluation parameters.

Parameters	Value
Simulation Duration	10 seconds
Work Cell dimensions	10 x 10 m ²
Number of UEs (N_{UE})	15
Packet Size (p_i)	10 or 25 bytes
Packet Periodicity	10 ms
Frequency Band	3.7 – 3.8 GHz
Bandwidth (BW)	10, 20 or 40MHz
Numerology (μ)	1 (SCS = 30 kHz)
MCS	12
MIMO transmission layers (v).	1

We compare the performance achieved with the designed CG scheduling policies when 5G NR OFDMA is used with the performance obtained using the 5GL-TDMA and 5GL-OFDMA schemes already implemented in 5G-LENA. To make a fair comparison, a First Come First Served scheduler is applied with 5GL-TDMA and 5GL-OFDMA. In this context, the scheduler serves UEs following a first-come, first-served basis, i.e., from UE₁ to UE _{N_{UE}} . Each UE receives the required number of RBs and OFDM symbols to satisfy its demand. When 5GL-TDMA is used, all UEs will receive r_i

$=R_{BW}$, and s_i will be calculated for each UE as $s_i=\lceil d_i/R_{BW} \rceil$ to meet its demand. When 5GL-OFDMA is applied, all UEs will receive $s_i= S_{slot}$ OFDM symbols, and r_i is calculated as $r_i=\lceil d_i/S_{slot} \rceil$.

6. Analytical Validation

In this section, we derive the analytical expressions that model the maximum UL latency experienced by the UEs using CG and the different scheduling policies and multiple access schemes. In this work, the UL latency accounts for the elapsed time from when a packet is created at the RLC layer of a UE until it is received at the RLC layer of the gNB. We compare the results achieved analytically with those achieved by simulation in order to validate the implementation in CG in 5G-LENA.

6.1. Analytical modeling

This subsection derives the analytical expressions of the maximum UL latency (L_{UL}) experienced by the UEs. We consider that all UEs transmit packets of the same size. Following [26], the latency of a packet transmitted in UL with CG is given by the following components. First, we need to account for the processing times at the UE and the gNB that represent the time required to generate the packet at the transmitter and decode the data at the receiver, respectively ($t_{UE,tx}$, and $t_{gNB,rx}$). Other latency components are the frame alignment time (t_{fa}) that accounts for the time interval from the creation of a packet until the next transmission opportunity for the Physical Uplink Shared Channel (PUSCH), the waiting time for the resources allocated for the packet transmission (t_w), and the transmission time of the packet (t_{tt}). L_{UL} is then calculated as:

$$L_{UL}=t_{UE,tx}+t_{fa}+t_w+t_{tt}+t_{gNB,rx} \quad (1)$$

The processing time at the transmitter and the receiver ($t_{UE,tx}$ and $t_{gNB,rx}$) are calculated as indicated in [23]. t_{fa} depends on the frame structure, i.e., on the configuration of slots and OFDM symbols within a slot for the transmission of control and data channels for DL and UL transmissions. The waiting time t_w accounts for the time interval that a UE has to wait after t_{fa} for the assigned resources. t_w depends on the way radio resources are allocated to the UEs. Therefore, t_w depends on the multiple access mode and the scheduling policy. The transmission time t_{tt} is equal to the time duration of the OFDM symbols used for the packet transmission, and it can be calculated as $s_i T_{sym}$, T_{sym} is the duration of an OFDM symbol. T_{sym} depends on the numerology used for the transmission of the packet, and s_i depends on the multiple

⁴ We ran sufficient simulations to achieve statistically valid results: a minimum of 100 runs (each run simulates 10 s of network operation) for each evaluated configuration was executed.

access mode and scheduling scheme. For numerology 1 with SCS 30 kHz, T_{sym} is equal to 35.67 ms. We calculate t_{fa} , t_w , and t_{tt} for the different scheduling policies and multiple access schemes studied in this work.

6.1.1. 5GL-TDMA

5GL-TDMA allocates all RBs in an OFDM symbol to the same UE. As a result, a UE_{*i*} receives $r_i^{5GL-TDMA} = R_{BW}$ RBs and $s_i^{5GL-TDMA} = \lceil d_i / R_{BW} \rceil$ OFDM symbols. R_{BW} depends on the bandwidth and is given in [21].

t_{tt} for a UE_{*i*} can be calculated as $\lceil d_i / R_{BW} \rceil \cdot T_{sym}$, where d_i is the number of radio resources demanded by UE_{*i*}. d_i can be calculated as a function of the packet size (p_i) and the MCS used to transmit the packet as:

$$d_i = \left\lceil \frac{(tbs_i(p_i + p_{headers}) + CRC)}{R \cdot Q_m \cdot v \cdot N_{sc, RB}} \right\rceil \quad (2)$$

In (2), $tbs_i(p_i + p_{headers})$ represents the smallest transport block size from the available values given in [27] that can be used to transmit $p_i + p_{headers}$ bits. Q_m is the modulation order, R is the code rate, and v is the number of MIMO transmission layers. The waiting time for the last UE served (the one that experiences the highest latency) is equal to the transmission time required by the other $N_{UE} - 1$ UEs requesting resources. Therefore, we can calculate $t_w + t_{tt}$ as:

$$t_w + t_{tt} = \sum_{i=1}^{N_{UE}} \left\lceil \frac{d_i}{R_{BW}} \right\rceil \cdot T_{sym} \quad (3)$$

If all UEs transmit packets of the same size, (3) can be expressed as:

$$t_w + t_{tt} = N_{UE} \cdot \left\lceil \frac{d_i}{R_{BW}} \right\rceil \cdot T_{sym} \quad (4)$$

t_{fa} accounts for the delay introduced by the transmission of other channels (control channels in UL and DL -PUCCH or PDCCH- or the transmission of the PDSCH). We consider that the first and last OFDM symbols of a slot are reserved for the transmission of control channels and 12 OFDM symbols are used for data transmission in UL. To calculate t_{fa} , we need to know how many slots are needed to allocate resources for all UEs, which is represented as n_{slot} , and is calculated as:

$$n_{slot} = \left\lceil \frac{\sum_{i=1}^{N_{UE}} s_i^{5GL-TDMA}}{12} \right\rceil = \left\lceil \frac{N_{UE} \cdot s_i^{5GL-TDMA}}{12} \right\rceil \quad (5)$$

The first and last OFDM symbols in each slot are dedicated to control channels transmissions. Then, t_{fa} is calculated as:

$$t_{fa} = (2 \cdot n_{slot} - 1) \cdot T_{sym} \quad (6)$$

Using (4), (5), and (6), the maximum latency experienced by UEs using 5GL-TDMA is calculated as:

$$L_{UL} = t_{UE,tx} + (2 \cdot n_{slot} - 1 + N_{UE}) \cdot \left\lceil \frac{d_i}{R_{BW}} \right\rceil \cdot T_{sym} + t_{gNB,rx} \quad (7)$$

6.1.2. 5GL-OFDMA

5GL-OFDMA allocates the same RB in all OFDM symbols within a slot to the same UE. 5GL-OFDMA then allocates $r_i^{5GL-OFDMA} = \lceil d_i / S_{slot} \rceil$ RBs in $s_i^{5GL-OFDMA} = S_{slot}$ OFDM symbols for each UE_{*i*} (d_i is calculated in (2)). The latency $t_{fa} + t_w + t_{tt}$ for 5GL-OFDMA is calculated as the time needed to transmit n_{slot} slots minus the time duration of the last OFDM symbol. n_{slot} is the number of slots used to transmit the packets of all UEs. $t_{fa} + t_w + t_{tt}$ is then calculated as:

$$t_{fa} + t_w + t_{tt} = n_{slot} \cdot T_{slot} - T_{sym} = (14 \cdot n_{slot} - 1) \cdot T_{sym} \quad (8)$$

In (8), T_{slot} is the time duration of a slot, which is equal to $14 \cdot T_{sym}$. The parameter n_{slot} in (8) is calculated as:

$$n_{slot} = \left\lceil \sum_{i=1}^{N_{UE}} r_i^{5GL-OFDMA} / R_{BW} \right\rceil = \left\lceil \frac{N_{UE} \cdot r_i^{5GL-OFDMA}}{R_{BW}} \right\rceil \quad (9)$$

Using (9) and (10), the maximum latency experienced by UEs using 5GL-OFDMA is calculated as:

$$L_{UL} = t_{UE,tx} + (14 \cdot n_{slot} - 1) \cdot T_{sym} + t_{gNB,rx} \quad (10)$$

6.1.3. Sym-OFDMA

Sym-OFDMA allocates to each UE_{*i*} the number of RBs necessary to satisfy d_i in the minimum number of OFDM symbols possible. We can then distinguish two cases. When $d_i > R_{BW}$, the UE_{*i*} receives $r_i^{Sym-OFDMA} = R_{BW}$ RBs in $s_i^{Sym-OFDMA} = \lceil d_i / R_{BW} \rceil$ consecutive symbols. When $d_i \leq R_{BW}$, Sym-OFDMA allocates $r_i^{Sym-OFDMA} = d_i$ RBs and $s_i^{Sym-OFDMA} = 1$ to UE_{*i*}. In this case, several UEs can share the RBs in the same OFDM symbol. The number of UEs sharing RBs in the same OFDM symbol is given by $\lfloor R_{BW} / d_i \rfloor$ when all UEs demand the same number d_i of RBs. We can then calculate the number of OFDM symbols n_s needed to serve all UEs as:

$$n_s = \begin{cases} \frac{N_{UE}}{\lfloor R_{BW}/d_i \rfloor}, & d_i \leq R_{BW} \\ \left\lceil \frac{d_i}{R_{BW}} \right\rceil \cdot N_{UE}, & d_i > R_{BW} \end{cases} \quad (11)$$

Using (11), the number of slots needed to serve all UEs is equal to $n_{slot} = \lceil n_s/12 \rceil$; each slot has 12 OFDM symbols for UL data transmission. Considering that the first and last OFDM symbols within slots are used for control channels, $t_{fa} + t_w + t_{tt}$ can be calculated as:

$$t_{fa} + t_w + t_{tt} = (2 \cdot n_{slot} - 1 + n_{sym}) \cdot T_{sym} \quad (12)$$

The maximum latency experienced using Sym-OFDMA is calculated as:

$$L_{UL} = t_{UE,tx} + (2 \cdot n_{slot} - 1 + n_{sym}) \cdot T_{sym} + t_{gNB,rx} \quad (13)$$

6.1.4. RB-OFDMA

RB-OFDMA establishes that each UE_{*i*} has to receive at least r_{min} RBs. Considering this constraint, it distributes the R_{BW} RBs in a symbol among the maximum number of UEs. The maximum number of UEs that can share the RBs in an OFDM symbol is given by $n_{UE,1} = \lfloor R_{BW}/r_{min} \rfloor$. In this context, the N_{UE} UEs are divided in $\lfloor N_{UE}/n_{UE,1} \rfloor$ groups of $n_{UE,1}$ UEs that receive $r_i^{RB-OFDMA(1)} = r_{min}$ RBs in $s_i^{RB-OFDMA(1)} = \lceil d_i/r_{min} \rceil$ OFDM symbols, and one last group with $n_{UE,2} = N_{UE} - \lfloor N_{UE}/n_{UE,1} \rfloor \cdot n_{UE,1}$ UEs that receive $r_i^{RB-OFDMA(2)} = \lceil R_{BW}/r_{min} \rceil / n_{UE,2} \cdot r_{min}$ RBs in $s_i^{RB-OFDMA(2)} = \lceil d_i/r_i^{RB-OFDMA(2)} \rceil$ OFDM symbols. We can then calculate the number n_s of OFDM symbols necessary to serve all UEs as:

$$n_s = \lfloor N_{UE}/n_{UE,1} \rfloor \cdot s_i^{RB-OFDMA(1)} + s_i^{RB-OFDMA(2)} \quad (14)$$

Using (14), the number of slots needed to serve all UEs is calculated as $n_{slot} = \lceil n_s/12 \rceil$. Considering that the first and last OFDM symbols within slots are used for control channels, $t_{fa} + t_w + t_{tt}$ can be calculated using (12) as $(2 \cdot n_{slot} - 1 + n_{sym}) \cdot T_{sym}$. The maximum latency experienced using RB-OFDMA is calculated as:

$$L_{UL} = t_{UE,tx} + t_{gNB,rx} + (2 \cdot n_{slot} - 1 + \left\lceil \frac{N_{UE}}{n_{UE,1}} \right\rceil \cdot s_i^{RB-OFDMA(1)} + s_i^{RB-OFDMA(2)}) \cdot T_{sym} \quad (15)$$

6.2. Validation

Figure 7 compares the maximum latency experienced in UL transmissions obtained analytically and through

simulations when using CG with 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA, and RB-OFDMA for different bandwidth values (BW) when packets size is equal to 10 bytes and slot format 1D13U is used. Figure 7 clearly shows that the simulated and analytical results precisely match all the multiple access schemes and scheduling policies evaluated. It is important to note that the maximum latency obtained with Sym-OFDMA and RB-OFDMA is equal to or lower than the maximum latency experienced with 5GL-TDMA and 5GL-OFDMA for all the evaluated BW values. The same maximum latency is achieved for all BW s when 5GL-TDMA is used. This is because 5GL-TDMA assigns all the RBs in an OFDM symbol to the same UE regardless of the number of RBs available or BW and its RB demand. As a result, 5GL-TDMA uses the same number of OFDM symbols to serve the N_{UE} UEs for all the evaluated BW values, and the latency experienced by the last served UE does not change with BW . When 5GL-OFDMA is used, the maximum latency experienced by a UE decreases 43.1% when BW increases from 10 to 20 MHz. This is because two slots are needed to allocate resources to all the UEs when $BW=10$ MHz. When BW increases to 20 MHz, the radio resource demand of all UEs can be satisfied with the RBs in one slot. Since each UE receives RBs in all the OFDM symbols dedicated for data within a slot, the maximum latency experienced by a UE remains constant as BW increases above 20 MHz. Both Sym-OFDMA and RB-OFDMA provide the lowest maximum latency values for all the BW evaluated. In addition, the maximum latency experienced with both scheduling policies decreases when BW increases. This is due to the greater flexibility introduced by OFDMA that allows radio resources to be allocated more efficiently and the experienced latency to be reduced. The performance achieved with both Sym-OFDMA and RB-OFDMA is analyzed in more depth in the following section.

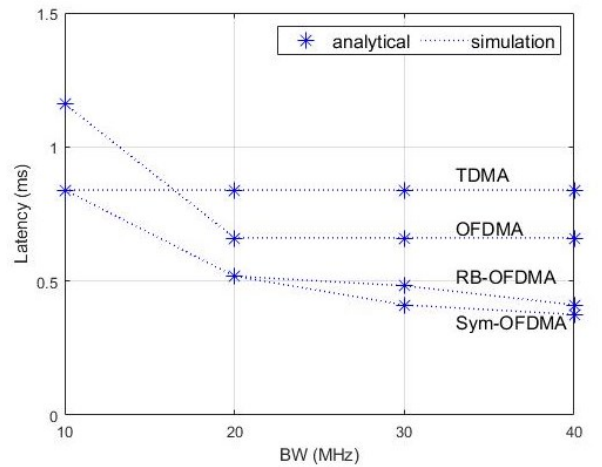


Figure 7. Maximum UL latency as a function of the bandwidth (packet size of 10 bytes).

We also compare the latency results obtained with CG using 5G-LENA with the analytical values reported in [25]

(third column in Table 2). Table 2 also includes the latency results obtained using dynamic scheduling for performance comparison. We consider a scenario with only one UE that transmits empty packets using 2 OFDM symbols. We evaluate the use of numerologies (μ) 0, 1, and 2, which correspond to a subcarrier spacing of 15, 30, and 60 kHz and a symbol duration of 71.4 μ s, 35.6 μ s, and 17.9 μ s, respectively. We should note that the higher the numerology, the lower the symbol duration and, therefore, lower latency. We consider a slot format 1D13U and $BW=20$ MHz. The results in Table 2 show that the latency results obtained with CG implemented in 5G-LENA are in line with the analytical results presented in 3GPP TR 37.910 [25], which validates the implemented CG in 5G-LENA. Moreover, the latency obtained with configured grant has a reduction of 93.3% compared to the latency obtained with the dynamic scheduler in the case of numerology 0.

Table 2. Latency experienced with Dynamic Scheduling and Configured Grant

μ	Dynamic scheduling	Configured Grant	
		5G-LENA	TR 37.910 [25]
0	7.31 ms	0.49 ms	0.52 ms
1	3.70 ms	0.30 ms	0.30 ms
2	1.90 ms	0.25 ms	0.24 ms

7. Performance evaluation

In this section, we analyze the performance achieved with the different multiple access schemes and scheduling policies in the evaluation scenario presented in section V.

7.1. Impact of the cell bandwidth

We analyze the latency that can be achieved with the different multiple access schemes and scheduling policies when different cell bandwidths are considered. Figure 8 shows a boxplot of the UL latency experienced by UEs using CG with 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA, and RB-OFDMA when bandwidth is equal to 10, 20 and 40 MHz and packet size is 10 bytes. In Figure 8, the red line within the box represents the average of the experienced latency, and the edges of the box are the 10th and 90th percentiles. The crosses represent the minimum and maximum values. At least otherwise indicated, we consider a packet size of 10 bytes and slot format 1D13U. Figure 9 depicts the number of radio resources used with the different multiple access modes and scheduling policies calculated with respect to the number of radio resources used by 5GL-TDMA. Figure 8 shows that Sym-OFDMA and RB-OFDMA equal or reduce the maximum and average latency experienced by the UEs for all the evaluated BW values compared with 5GL-OFDMA and 5GL-TDMA. This is due to a more efficient use of radio resources, as shown in Figure 9. When the

bandwidth is low ($BW=10$ MHz), both Sym-OFDMA and RB-OFDMA achieve similar latency results to that achieved with 5GL-TDMA. 5GL-TDMA assigns all the RBs in a symbol to the same UE. In this case, the number of RBs needed to transmit 10-bytes packets is equal to the number of RBs available in the 10 MHz bandwidth. Sym-OFDMA then allocates all the RBs in a symbol to a UE, achieving the same latency performance as 5GL-TDMA. Figure 9.a also shows that both 5GL-TDMA and Sym-OFDMA use the same amount of radio resources. When $BW=10$ MHz, RB-OFDMA allocates RBs in two consecutive symbols to each UE, and the RBs in a symbol are shared by two UEs. As a result, the maximum latency experienced with RB-OFDMA is the same as with 5GL-TDMA and Sym-OFDMA, as shown in Figure 8. Figure 9.a shows that RB-OFDMA also uses the same number of radio resources as 5GL-TDMA and Sym-OFDMA. 5GL-OFDMA provides the largest latency values when the bandwidth is 10 MHz. 5GL-OFDMA allocates RBs in all symbols within a slot to UEs. When bandwidth is equal to 10 MHz, only 8 of the 15 UEs can receive resources in the first slot after the packets are generated, and they experience a latency of 0.65 ms. The rest of the UEs receive resources in the next slot and experience a latency of 1.15 ms. Figure 9.a shows that 5GL-OFDMA uses 40% more radio resources than 5GL-TDMA, Sym-OFDMA and RB-OFDMA.

When the bandwidth increases from 20 to 40 MHz, 5GL-OFDMA reduces the maximum latency experienced compared with 5GL-TDMA. Figure 8 shows that all UEs experience the same latency when BW is equal to 20 and 40 MHz. This happens because all UEs share RBs in all the OFDM symbols of the first slot after the packets are generated. Sym-OFDMA and RB-OFDMA achieve the lowest latency for the UEs. Furthermore, both Sym-OFDMA and RB-OFDMA use a lower number of radio resources to serve the N_{UE} UEs than 5GL-TDMA and 5GL-OFDMA

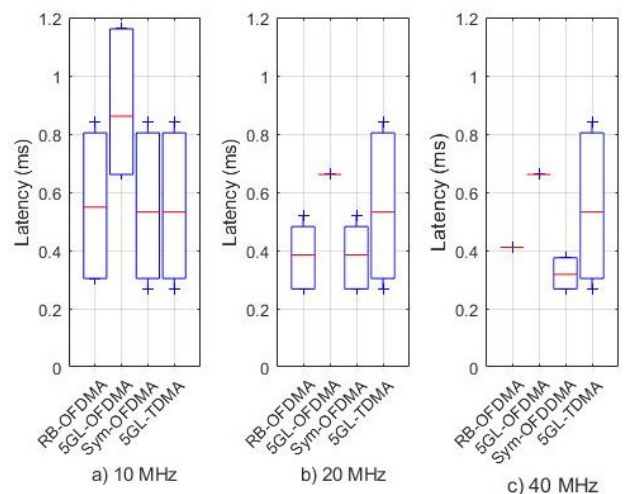


Figure 8. UL latency experienced as a function of the bandwidth (packet size of 10 bytes).

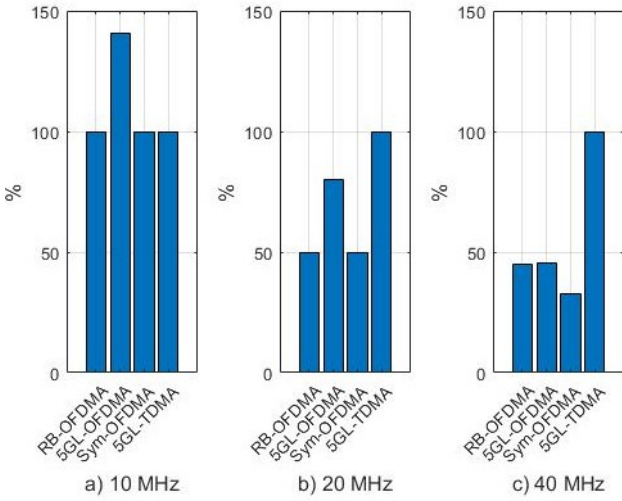


Figure 9. Percentage of radio resources used for the transmission of packets with respect to 5GL-TDMA (packet size of 10 bytes).

(Figure 9). Sym-OFDMA allows several UEs to allocate RBs in the same OFDM symbol. As BW and the number of available RBs increase, more UEs receive RBs in the same OFDM symbol. Sym-OFDMA then requires a lower number of OFDM symbols to serve all UEs compared with 5GL-TDMA. As a result, the latency experienced by the UEs decreases. Figure 8 also shows that the maximum latency experienced with RB-OFDMA also reduces when the bandwidth increases. However, the minimum and average latency experienced increases when bandwidth increases from 20 to 40 MHz. This is due to the different number of OFDM symbols allocated to the UEs when BW is equal to 20 and 40 MHz, respectively. When $BW=20$ MHz, each UE receives RBs in only one OFDM symbol, and the RBs in a symbol are shared by 2 UEs. When bandwidth is equal to 40 MHz, RB-OFDMA allocates RBs in 5 consecutive OFDM symbols to each UE, and the RBs in a symbol are allocated to all the UEs.

7.2. Impact of the packet size

Figure 10 shows the boxplot of the UL latency experienced by the UEs using CG with 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA and RB-OFDMA when packets of 10 and 25 bytes are transmitted, respectively, and considering $BW=20$ MHz, and slot format 1D13U. Figure 10 shows that the latency experienced by the UEs increases with the packet size for Sym-OFDMA and RB-OFDMA, while it remains constant for 5GL-TDMA and 5GL-OFDMA. This is the case because 5GL-TDMA and 5GL-OFDMA allocate more radio resources than demanded by each UE when 10 bytes of data are transmitted per packet. The number of allocated radio resources is enough to satisfy the radio resource demand when the packet size increases to 25 bytes. Therefore, the experienced UL latency remains constant with 5GL-TDMA and 5GL-OFDMA because they allocate the same radio resources to UEs when the packet size is 10 and

25 bytes. On the other hand, Sym-OFDMA and RB-OFDMA allocate to each UE a number of radio resources more adjusted to their demands by taking full advantage of the 5G NR OFDMA flexibility. For that reason, the number of allocated radio resources to each UE increases when packet size increases. This also results in the increase of the UL latency experienced by the UEs. However, it is important to highlight that the maximum UL latency experienced with Sym-OFDMA and RB-OFDMA is always lower than the one experienced with 5GL-TDMA and 5GL-OFDMA, respectively. In the case of Sym-OFDMA, the number of available RBs is not enough to satisfy the demand of more than one UE, and each UE receives RBs in different OFDM symbols. As a result, Sym-OFDMA and 5GL-TDMA achieve the same latency performance. However, Sym-OFDMA only uses 66% of the radio resources used by 5GL-TDMA, and the non-allocated radio resources could be used by other UEs. RB-OFDMA has higher flexibility than Sym-OFDMA since it can allocate any number of RBs and OFDM symbols to each UE; the number of OFDM symbols is always 1 with Sym-OFDMA if the radio resource demand of the UE can be satisfied with the available RBs in an OFDMA symbol. Thanks to this, it achieves a lower maximum UL latency when the packet size is equal to 25 bytes.

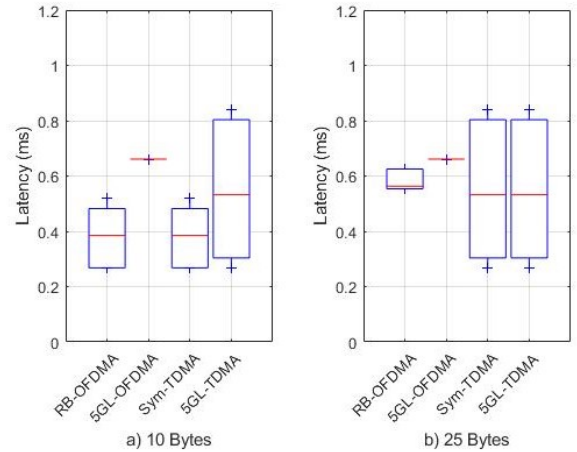


Figure 10. UL latency experienced as a function of the packet size ($BW=20$ MHz).

7.3. Impact of the MCS

Now, we evaluate the impact of the MCS used for the packet transmissions. Figure 11 shows the maximum UL latency experienced by the UEs and the percentage of allocated radio resources using CG with 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA and RB-OFDMA when MCS 12, 20 and 28 are used and $BW=20$ MHz and packets of 10 bytes are transmitted. The percentage of radio resources allocated is calculated with respect to the number of radio resources used by 5GL-TDMA. Figure 11.a shows that Sym-OFDMA and RB-OFDMA reduce the maximum latency experienced by the UEs for all the evaluated MCSs compared with 5GL-OFDMA

and 5GL-TDMA. This is thanks to the most efficient use of radio resources, as shown in Figure 11.b. When MCS increases, a lower number of RBs are demanded by each UE to transmit their packets. Sym-OFDMA and RB-OFDMA adjust the number of allocated resources to the UEs demand exploiting the flexibility offered by OFDMA. However, 5GL-TDMA and 5GL-OFDMA maintain the same radio resource allocation for all MCS values evaluated due to the constraints introduced by the multiple access schemes implemented in 5G-LENA. As a result, Sym-OFDMA and RB-OFDMA reduce the number of allocated radio resources and the maximum experienced UL latency compared with 5GL-TDMA and 5GL-OFDMA. For example, Sym-OFDMA and RB-OFDMA reduce the maximum UL latency by 48.63% and 43.22%, respectively, compared with 5GL-OFDMA when MCS 28 is used. These results are achieved using only 80.7% and 66.7% of the radio resources used by 5GL-OFDMA. Compared with 5GL-TDMA, Sym-OFDMA and RB-OFDMA reduce the maximum UL latency 59.5% and 55.3%, respectively, using only 15.38% and 26.67% of the radio resources.

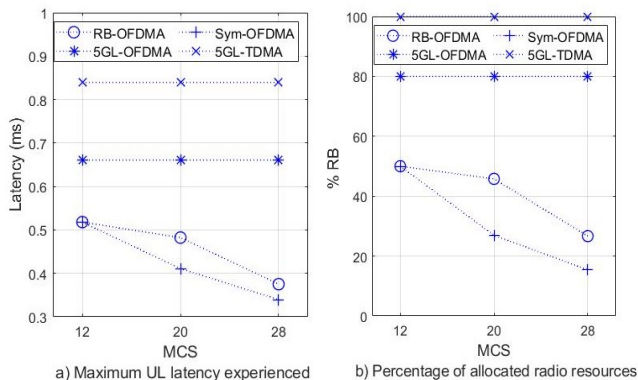


Figure 11. Performance as a function of the MCS ($BW = 20$ MHz, packet size = 10 bytes).

7.4. Impact of the frame and slot format

Finally, we evaluate the performance achieved using CG with 5GL-TDMA, 5GL-OFDMA, Sym-OFDMA and RB-OFDMA when only part of the OFDM symbols within a single slot is reserved for UL traffic. To this end, we evaluate the use of 3 different TDD frame and slot configurations. In particular, we consider that the first 1, 5, and 9 OFDM symbols within a slot of a frame are used for DL transmissions, and the last 13, 9, and 5 OFDM symbols of each slot are used for UL transmissions; the three configurations are referred to as 1D13U, 5D9U, and 9D5U, respectively⁵. Figure 12 shows the boxplot of the UL latency experienced by UEs when different

frame and slot configurations are used, $BW=20$ MHz, and packet size is equal to 10 bytes. Figure 12 shows that, as expected, the latency increases for all multiple access schemes and scheduling policies when the number of OFDM symbols reserved for UL transmissions decreases. Sym-OFDMA and RB-OFDMA always provide the lowest latency values thanks to the higher flexibility offered by OFDMA multiple access scheme. 5GL-TDMA is the one for which the experienced UL latency increases more. This is because 5GL-TDMA requires a larger number of slots to serve all the UEs as the number of OFDM symbols reserved for UL within a slot decreases.

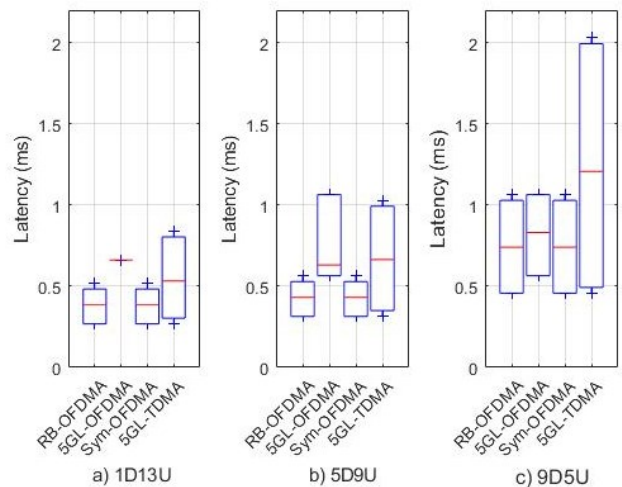


Figure 12. Latency experienced with the different schemes as a function of the slot format ($BW = 20$ MHz, packet size = 10 bytes).

8. Conclusions

This paper has presented a detailed description of the first implementation of configured grant scheduling in an open-source 5G NR simulator (to the best of the authors' knowledge), in particular in 5G-LENA. The code of configured grant is publicly available in [7]. Configured grant pre-allocates radio resources to the UEs and avoids the signaling exchange between the UE and the gNB to request/inform about the allocated radio resources reducing the latency of the transmission. Configured grant is key for the support of time-critical services in 5G networks. This work is a valuable contribution since the availability of simulation tools that accurately model all the functionalities of 5G NR and, in particular, configured grant, is fundamental for the research in 5G and beyond networks supporting time-critical services. In particular, we have presented the new functionalities and modifications included at the MAC and PHY layers of 5G-LENA to integrate configured grant in the

⁵ We consider that several devices generate DL traffic in the system. The results showed that the performance experienced by the UL traffic is not affected by the amount of traffic generated in DL, but by the number of radio resources reserved for the transmission of UL traffic and the frame and slot configuration.

simulator. To accurately model the high flexibility offered by 5G NR in the radio resource allocation process, we have also implemented the 5G NR OFDMA in the simulator that allows radio resources to be shared simultaneously in time and frequency by different UEs. Using OFDMA multiple access scheme, it is possible to transmit using any number of OFDM symbols. The latency results achieved with configured grant in 5G-LENA match with the latency values reported in previous analytical studies, which validates the implementation of configured grant in 5G-LENA.

We have also implemented two scheduling policies that are applied with configured grant and OFDMA to demonstrate the flexibility and capabilities of 5G NR to support time-critical services. The proposed scheduling policies exploit the flexibility offered by OFDMA to guarantee low latencies and efficient use of radio resources. We have considered a case study where a 5G NR cell covers an industrial scenario where a closed-loop control application demands low latency communications. The results have shown that the use of CG with the proposed scheduling policies and OFDMA reduces the maximum latency experienced by UEs by up to 48.63% compared to the latency experienced when the multiple access schemes previously implemented in 5G-LENA are used. This result is achieved using 80.7% fewer radio resources. The results have shown that the use of scheduling policies that make efficient use of radio resources is more critical the smaller the size of the packets to be transmitted.

Acknowledgments

This work has been funded by MCIN/AEI/10.13039/501100011033 through the project PID2020-115576RB-I00, and by European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101057083.

References

- [1] 5G-ACIA, "5G for Connected Industries and Automation," 2019.
- [2] 5G-ACIA, "Key 5G Use Cases and Requirements" 2020.
- [3] S. Zeb, A. Mahmood, S. A. Hassan, M. J. Piran, M. Gidlund, and M. Guizani, "Industrial digital twins at the nexus of NextG wireless networks and computational intelligence: A survey," *J. Netw. Comput. Appl.*, vol. 200, p. 103309, Apr. 2022.
- [4] E. Dahlman and S. Parkvall, "NR - The new 5G radio-access technology," *IEEE 87th Vehicular Technology Conference*, 2018.
- [5] N. Patriciello, S. Lagen, B. Bojovic, and L. Giupponi, "An E2E Simulator for 5G NR Networks," *Simul. Model. Pract. Theory*, vol. 96, Nov. 2019.
- [6] "ns-3 | a discrete-event network simulator for internet systems." Available: <https://www.nsnam.org/>.
- [7] "Open-source implementation of Configured Grant in 5G-LENA." Available: <https://gitlab.com/ns-3-dev-nr-configuredgrant/nr/-/tree/5g-lena-cg-v2.1.y>.
- [8] "Open source implementation ns-3 (updated to use with configured-grant)." Available: <https://gitlab.com/ns-3-dev-nr-configuredgrant/ns-3-dev/-/tree/ns-3.36-cg>.
- [9] 3GPP TSG RAN, "Physical channels and modulation," Release 16, TS 38.211, V16.7.0, 2021.
- [10] 3GPP TR, "Study on New Radio (NR) access technology" Release 17, 38.912, 2022.
- [11] A. A. Zaidi, R. Baldemair, M. Andersson, S. Faxer, V. Moles-cases, and Z. Wang, "The 5G NR Physical Layer Design," *Ericsson Technol. Rev.*, vol. 1, 2017.
- [12] 3GPP TSG RAN, "Physical layer procedures for data" Release 16, TS 38.214, V16.1.0, 2020.
- [13] 3GPP TSG RAN, "NR and NG-RAN Overall Description" Release 16, TS 38.300, V16.7.0, 2021.
- [14] 3GPP TSG RAN, "Radio Resource Control (RRC) protocol specification" Release 16, TS 38.331, V16.6.0, 2021.
- [15] K. Koutlia, B. Bojovic, Z. Ali, and S. Lagén, "Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios," *Simul. Model. Pract. Theory*, vol. 119, p. 102580, Sep. 2022.
- [16] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, "The impact of NR Scheduling Timings on End-to-End Delay for Uplink Traffic," 2019.
- [17] B. Bojović, S. Lagén, and L. Giupponi, "Realistic beamforming design using SRS-based channel estimate for ns-3 5G-LENA module," *ACM Int. Conf. Proceeding Ser.*, pp. 81–87, 2021.
- [18] B. Bojovic, Z. Ali, and S. Lagen, "ns-3 and 5G-LENA Extensions to Support Dual-Polarized MIMO," 2022.
- [19] T. Zugno, M. Polese, N. Patriciello, B. Bojović, S. Lagen, and M. Zorzi, "Implementation of a Spatial Channel Model for ns-3," *ACM Int. Conf. Proceeding Ser.*, no. February, pp. 49–56, 2020.
- [20] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, "An improved MAC layer for the 5G NR ns-3 module," *ACM Int. Conf. Proceeding Ser.*, no. November, pp. 41–48, 2019.
- [21] 3GPP TSG RAN, "Base Station (BS) radio transmission and reception" Release 17, TS 38.104, 2021.
- [22] 3GPP TSG RAN, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)," Release 16, TR 38.824, V16.0.0, 2019.
- [23] 3GPP TSG RAN, "Study on channel model for frequencies from 0.5 to 100 GHz" Release 17, TR 38.901, v17.0.0, 2022.
- [24] K. Montgomery, R. Candell, Y. Liu, and M. Hany, "Wireless User Requirements for the Factory Workcell NIST Advanced Manufacturing Series 300-8 Wireless User Requirements for the Factory Workcell," *Adv. Manuf. Ser. (NIST AMS)*.
- [25] 3GPP TSG RAN, "Study on self evaluation towards IMT-2020 submission" Release 16, TR 37.910, V16.1.0, 2021.
- [26] M. C. Lucas-Estañ *et al.*, "An Analytical Latency Model

and Evaluation of the Capacity of 5G NR to Support V2X Services using V2N2V Communications," 2021.

[27] 3GPP TSG RAN, "Medium Access Control (MAC) protocol specification" Release 16, TS 38.321, V16.6.0, 2021.

[28] European 5G Observatory (2020), "5G private licences spectrum in Europe." Available: <https://5gobservatory.eu/5g-private-licences-spectrum-in-europe/>.

Configured Grant Scheduling for the Support of TSN Traffic in 5G and Beyond Industrial Networks

M.C. Lucas-Estañ¹, A. Larrañaga², J. Gozalvez¹, I. Martínez²

¹UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain

²HW and Communication Systems Area, Ikerlan Technology Research Centre, Mondragón 20500, Spain

{m.lucas, j.gozalvez}@umh.es, {ana.larranaga, imartinez}@ikerlan.es

Abstract— 5G and beyond networks can facilitate the digital transformation of manufacturing and support more flexible and reconfigurable factories with ubiquitous mobile connectivity. This requires integrating 5G networks with industrial networks that increasingly rely on TSN (Time Sensitive Networking) to support deterministic communications with bounded latencies. Deterministic communications are critical for many industrial applications, but 5G does not natively support deterministic communications. To address this limitation, this study proposes the coordination of the 5G and TSN schedulers and presents a novel 5G configured grant scheduling scheme to support TSN traffic. The scheme uses information about the characteristics of the TSN traffic (packet size, periodicity, and arrival time) to coordinate its scheduling decisions with the TSN scheduler. The study demonstrates that the proposed scheme outperforms the state-of-the-art in the capacity to support multiple TSN traffic flows with different periodicities.

Keywords—5G, TSN, industrial networks, 5G-TSN integration, scheduling, deterministic communications, Configured Grant.

I. INTRODUCTION

The digital transformation of manufacturing requires flexible and reconfigurable factories that efficiently integrate (mobile) cooperative robots and cyber-physical systems. Industrial wired networks increasingly rely on TSN (Time Sensitive Networking) to support mixed traffic flows and deterministic communications with bounded latencies. 5G does not natively support deterministic communications but provides the mobility support and reconfigurability required by future factories. In this context, 5G-ACIA (5G Alliance for Connected Industries and Automation) identifies the integration of 5G and beyond with TSN as a fundamental step to support the digital and data-centric transformation of manufacturing [1].

The 3GPP defines the framework for integrating 5G and TSN where the 5G system operates as a logical bridge in a TSN network. However, 3GPP does not specify how the 5G and TSN schedulers should be integrated, and this is critical to satisfy the end-to-end (E2E) latency requirements of industrial TSN traffic over 5G, especially under mixed traffic flows. A first proposal is presented in [2] where authors introduce a scheduler that preempts eMBB (enhanced mobile broadband) traffic when prioritized deterministic traffic must be transmitted. The scheduler seeks minimizing the impact of preemption on the eMBB throughput while guaranteeing the latency requirements of deterministic traffic. The proposal in [3] combines semi-static and dynamic scheduling to support TSN traffic over 5G. The proposal pre-allocates periodic radio resources for the transmission of TSN traffic. If the TSN traffic varies and the pre-allocated resources are insufficient to support it, additional resources are dynamically scheduled to support the additional TSN traffic. A similar approach is followed in [4] to support mixed traffic flows (periodic time-critical, event-triggered time-critical and best-effort). The proposal pre-allocates resources for the periodic time-critical

traffic, reserves resources in each slot for potential event-triggered time-critical traffic, and dynamically schedules best-effort traffic. Best-effort traffic can also be scheduled over unutilized reserved resources with preemption applied if new event-triggered traffic is generated.

The proposals in [3] and [4] advocate for the use of semi-static scheduling (Configured Grant -CG- for uplink traffic or Semi-Persistent Scheduling -SPS- for downlink) to support periodic deterministic TSN traffic. The proposals assign one configured UL (uplink) grant to each TSN traffic flow. Each grant periodically pre-allocates radio resources to each TSN flow with the same periodicity as the TSN traffic. Existing proposals manage each grant or TSN flow independently of each other, and this can generate scheduling conflicts if each TSN flow generates traffic with different periodicities. Fig. 1 illustrates such conflicts that occur when two or more TSN traffic flows are pre-allocated the same resources at a certain time. To address this challenge, this paper proposes a novel 5G configured grant scheduling scheme that can efficiently manage multiple UL TSN flows with different periodicities (it is important to note that periodic deterministic traffic is the most common traffic generated in industrial environments [5]). The proposed scheme uses information about each TSN flow (periodicity, packet size and packet arrival time) to coordinate its scheduling decisions with the TSN network and avoid scheduling conflicts. The scheme avoids conflicts between TSN flows with different periodicities by assigning several configured UL grants to each TSN flow. The scheme identifies a common period (hyperperiod or HP) for all TSN traffic flows, and configures a different UL grant for each packet of a TSN flow in the HP. Each configured UL grant allocates the radio resources that minimize the latency of the corresponding packet, and this radio resource allocation repeats periodically with a periodicity HP. The conducted evaluation shows that the proposed CG scheduling scheme avoids scheduling conflicts, and considerably increases the number of TSN traffic flows that 5G can satisfactorily support, i.e., for which their E2E latency requirement is met.

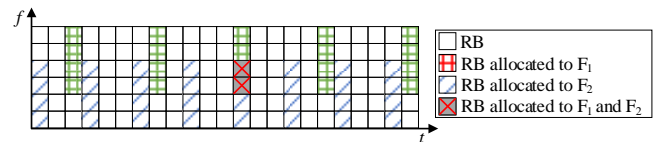


Fig. 1. Scheduling conflict between two TSN flows with different periodicities.

II. 5G-TSN INTEGRATION MODEL

A TSN network is composed of end devices and bridges interconnected using standard Ethernet links. The TSN bridges are Ethernet bridges with special features to guarantee deterministic communications. These features include (among others) a strict time synchronization using the IEEE 802.1AS standard, and a priority-based scheduler that reserves specific transmission intervals for high priority traffic using the IEEE 802.1Qbv standard.

The 3GPP defines in [6] the framework for the integration of 5G and TSN. 3GPP defines that a 5G network should integrate in a TSN network as a logical bridge. The integration model is depicted in Fig. 2, and the 5G System (5GS) is referred to as 5GS Bridge. The 3GPP integration model includes TSN translators (TT) at the interconnection points between the 5G and TSN networks, i.e. at the User Equipment or UE (the device-side TT or DS-TT) and the 5G Core Network (the network-side TT or NW-TT). The TTs act as ingress and egress ports of the logical bridge, and their main functions are to understand, translate and/or execute TSN configuration messages received from the TSN network. The Central Network Configuration (CNC) in a TSN network is in charge of the centralized management of the integrated 5G-TSN network. The CNC is aware of the communication requirements of the end-devices, the status and capabilities of all (TSN and 5GS) bridges in the network, and the delay suffered at each bridge and link of the integrated 5G-TSN network. The CNC uses this information to schedule transmissions following the IEEE 802.1Q standard. In particular, it establishes the communication path and the time instants at which each packet of a TSN flow should arrive and depart from each TSN and 5GS bridge to ensure that the end-to-end requirements of the TSN traffic are met. The CNC sends the scheduling decisions to the TSN and 5GS bridges in the network for their configuration.

It is important to highlight that 3GPP or TSN standards do not specify a scheduling mechanism for the TSN traffic over a 5G network. Such scheduling should be done so that the arrival and departure times at the 5GS bridge established by the CNC are met. To this end, the CNC provides information to the 5G network regarding the packet arrival time, the flow direction, the survival time¹, and the periodicity of the packets² for each TSN flow. The packet arrival time refers to the time when the first packet of the flow arrives at the 5GS bridge (specifically, at the ingress port of the 5GS bridge that is the UE in UL transmissions). In addition, a 5G QoS profile is assigned to each TSN flow. The 5G QoS profile includes the flow's priority level, the maximum data burst volume (MDBV), and its required packet error rate (PER) and packet delay budget (PDB). The PDB indicates the maximum latency for a packet to be transmitted over the 5G network.

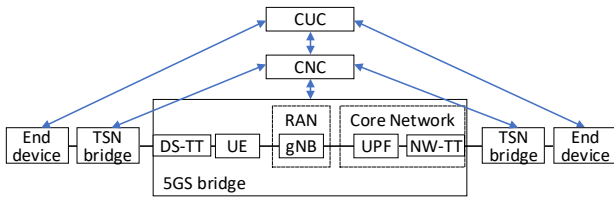


Fig. 2. 5G-TSN integration framework proposed by the 3GPP [6].

III. 5G NR SCHEDULING FOR TSN TRAFFIC

This section presents a novel 5G Configured Grant scheduling scheme capable to support multiple periodic TSN flows in the integrated 5G-TSN framework depicted in Fig. 2. We refer to the proposed scheme as Optimum Flexible configured grAnt Scheduling for TSN traffic (O-FAST). The proposed scheme pre-allocates radio resources periodically to the TSN flows and configures multiple UL grants for each TSN flow if required. The scheme avoids scheduling conflicts

between TSN flows with different periodicities (see Fig. 1). To avoid the conflicts, the proposed scheme exploits the information provided by the CNC to satisfy the latency requirements of all TSN flows. In particular, the scheme utilizes the information about the arrival time of the first packet of each TSN flow at the 5GS bridge, and the size and periodicity of the packets of each TSN flow.

We consider a 5G network that must support N_F TSN traffic flows. Each TSN flow F_i (with $i=1,\dots,N_F$) is characterized by the transmission of periodic packets with size $size_i$ and periodicity p_i . We denote as $pkt_{i,j}$ the packets of a TSN flow F_i , where i and j indicate that it is the j^{th} packet in the TSN flow F_i , and $j=1,2,\dots$. The 5G network knows the arrival time of the first packet $pkt_{i,1}$ of a TSN flow F_i at the 5GS bridge, which is denoted with $A_{i,1}$, and the maximum latency that must be guaranteed in the 5G network for the packets of a TSN flow F_i , referred to as \hat{l}_i^{5G} . Based on the previous information, the arrival time for any packet $pkt_{i,j}$ of a TSN flow F_i can be calculated as $A_{i,j} = A_{i,1} + (j-1)p_i$.

O-FAST first computes the hyperperiod (HP) that is defined as the least common multiple of the periodicities of all the TSN flows: $HP = \text{LCM}\{p_i\}$, $\forall i = 1,\dots,N_F$. O-FAST then schedules the packets of all the TSN flows within HP , and repeats the scheduling in following HP s since all the TSN flows repeat their traffic pattern with periodicity HP . For a TSN flow F_i , O-FAST allocates the radio resources to each packet $pkt_{i,j}$ in a TSN flow F_i that minimize the latency considering the available radio resources. This is not the case with commonly used CG schemes that allocate resources with periodicity p_i for the transmission of all the packets of a TSN flow F_i . O-FAST selects the radio resources reserved to each packet $pkt_{i,j}$ of a TSN flow F_i to avoid scheduling conflicts among different flows (Fig. 1). To this aim, O-FAST divides the scheduling problem into several sub-problems with lower computational cost. It creates different groups G_z of packets from all TSN flows F_i included in an HP . Each group includes the packets $pkt_{i,j}$ (with $i=1,\dots,N_F$ and $j=1,2,\dots$) whose transmission could overlap in time considering their arrival time $A_{i,j}$ at the 5GS bridge and the maximum latency \hat{l}_i^{5G} that must be guaranteed. The transmission of two packets $pkt_{i,j}$ and $pkt_{m,n}$ can overlap in time when $A_{i,j} \leq A_{m,n} \leq A_{i,j} + \hat{l}_i^{5G}$ or $A_{i,j} \leq A_{m,n} + \hat{l}_m^{5G} \leq A_{i,j} + \hat{l}_i^{5G}$. O-FAST then defines an optimization problem for each group G_z to identify the radio resources that must be allocated to each packet $pkt_{i,j}$ (with $i=1,\dots,N_F$ and $j=1,2,\dots$) within G_z to avoid scheduling conflicts.

O-FAST needs to identify the radio resources that must be allocated for the transmission of each packet $pkt_{i,j}$ within G_z to guarantee the maximum latency \hat{l}_i^{5G} established by the TSN scheduler. The number of radio resources needed to transmit a packet $pkt_{i,j}$ of size $size_i$ is calculated as:

$$d_i = \left\lceil \frac{[tbs_i(size_i + header) + CRC] \cdot 8}{R \cdot Q_m \cdot N_{sc, RB}} \right\rceil \quad (1)$$

where $header$ is the length of the IPv4 header (in bits) and CRC is the length of the cyclic redundancy check (CRC) code in bytes. $tbs_i(size_i + header)$ is the smallest transport block size that can transmit a packet of size $size_i + header$. R and Q_m are the coding rate and modulation order, respectively, used to transmit a packet. $N_{sc, RB}$ is the number of subcarriers in a resource block or RB of the 5G NR grid. 5G NR defines a grid

¹ The survival time indicates the time period an application can survive without receiving any data packet.

² This information is included in the Time Sensitive Communication Assistance Information (TSCAI) message [6].

organized in RBs in the frequency domain and OFDM symbols in the time domain. An RB is composed of 12 subcarriers in the frequency domain, and the number of RBs depends on the bandwidth and is represented by R_{BW} . A radio resource consists of an RB in the frequency domain and an OFDM symbol in the time domain. We number the symbols (starting from 0) within an HP , and each symbol has a time duration of t_{sym} that depends on the 5G NR numerology. OFDM symbols are organized in slots that consist of 14 or 12 symbols depending on the 5G NR numerology.

O-FAST establishes that when the number of radio resources d_i needed for the transmission of a packet $pkt_{i,j}$ is lower than R_{BW} , the packet $pkt_{i,j}$ will receive d_i RBs in the same symbol. When d_i is higher than R_{BW} , the packet $pkt_{i,j}$ will receive R_{BW} in $\lceil d_i/R_{BW} \rceil$ consecutive symbols. The number of symbols and RBs needed to satisfy d_i for a packet $pkt_{i,j}$ is represented by d_i^S and d_i^R , respectively. O-FAST allocates to each packet $pkt_{i,j}$ the radio resources that minimize its latency considering the arrival time at the 5GS bridge $A_{i,j}$ and the number of radio resources d_i needed to transmit each packet $pkt_{i,j}$ in a group G_z . We denote with $l_{i,j}$ the latency experienced by the packet $pkt_{i,j}$ in the 5G network. O-FAST then minimizes the sum of the latency $l_{i,j}$ experienced by all the packets $pkt_{i,j}$ in G_z , with $i \in [1, N_F]$ and $j \in [1, N_{pkt}^{i,z}]$, where $N_{pkt}^{i,z}$ represents the number of packets for the TSN flow F_i within G_z :

$$\min \sum_{i=1}^{N_F} \sum_{j=1}^{N_{pkt}^{i,z}} l_{i,j} \quad (2)$$

The latency experienced by a packet in a 5G network can be calculated as the sum of the latency experienced in the radio, transport, and core networks. This study considers a 5G Non-Public Network (NPN) with a core network implemented in the factory premises. In this scenario, the latency experienced in the transport and core network can be considered negligible compared to the latency experienced in the radio network [7]. $l_{i,j}$ can then be estimated as the difference between the time the packet is received in the gNB and the arrival time $A_{i,j}$ of the packet at the 5GS bridge (it is important to note that the ingress port is the UE for UL transmissions). The time at which the packet is received in the gNB depends on: 1) the time at which the transmission of packet $pkt_{i,j}$ starts, 2) the transmission time of the packet $pkt_{i,j}$ through the radio channel that is given by $d_i^S \cdot t_{sym}$, and 3) the processing time $t_{gNB,rx}$ needed to decode the packet at the receiver. If we represent as $s_{i,j}$ the first symbol allocated to a packet $pkt_{i,j}$, the time at which the transmission of packet $pkt_{i,j}$ starts is given by $s_{i,j} \cdot t_{sym}$. Then, $l_{i,j}$ can be expressed as:

$$l_{i,j} = s_{i,j} \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j} \quad (3)$$

$s_{i,j}$ in (3) can be expressed as:

$$s_{i,j} = \sum_{s=0}^S s \cdot X_s^{i,j} \quad (4)$$

with $X_s^{i,j}$ a binary variable equal to 1 when s is the first symbol allocated for the transmission of packet $pkt_{i,j}$ and 0 in other case. S in (4) represents the length of the HP expressed in number of symbols. Using (3) and (4), the objective function of O-FAST defined in (2) can now be expressed as:

$$\min \sum_{i=1}^{N_F} \sum_{j=1}^{N_{pkt}^{i,z}} \left(\sum_{s=0}^S s \cdot X_s^{i,j} \right) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j} \quad (5)$$

with $X_s^{i,j} \in \{0,1\}$, $\forall i \in \{1, \dots, N_F\}$, $\forall j \in \{1, \dots, N_{pkt}^{i,z}\}$, and $\forall s \in \{1, \dots, S\}$.

The solution to the optimization problem must satisfy that the latency $l_{i,j}$ experienced by a packet $pkt_{i,j}$ in the 5GS is equal to or lower than the maximum latency established by the CNC of the TSN network (l_i^{5G}). This constraint to the optimization problem is expressed as:

$$\left(\sum_{s=0}^S s \cdot X_s^{i,j} \right) \cdot t_{sym} + d_i^S \cdot t_{sym} + t_{gNB,rx} - A_{i,j} \leq l_i^{5G}, \quad (6)$$

$$\forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\}.$$

O-FAST takes into account the processing time $t_{UE,tx}$ required by the transmitter (the UE) to generate and encode the packet. The radio resources allocated to the transmission of a packet $pkt_{i,j}$ must then be in symbols after $A_{i,j} + t_{UE,tx}$, and this constraint is expressed as:

$$X_s^{i,j} = 0, \quad \forall s < \left\lfloor \frac{A_{i,j} + t_{UE,tx}}{t_{sym}} \right\rfloor, \quad (7)$$

$$\forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\}$$

5G NR can reserve some symbols within a slot for the transmission of control channels. O-FAST considers that the first $sym_{DL-Ctrl}$ symbols of each slot are reserved for control channels in DL and the last $sym_{UL-Ctrl}$ symbols of each slot are reserved for control channels in UL. To avoid allocating symbols reserved for DL control channels, we define the following constraint for O-FAST:

$$X_s^{i,j} = 0, \quad \forall s \mid 0 < \text{mod}(s, 14) \leq sym_{DL-Ctrl}, \quad (8)$$

$$\forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\}$$

The 3GPP standards establish that all the radio resources allocated for the transmission of a packet must be in the same slot. Assuming a 5G NR numerology with a normal cyclic prefix, each slot consists of 14 symbols. $\lfloor s/14 \rfloor$ is the number of complete slots from the beginning of the HP to the symbol s that represents the first symbol allocated for the transmission of the packet. $s - \lfloor s/14 \rfloor \cdot 14$ indicates the number of the symbol s within the current slot. To ensure that the d_i^S symbols allocated for the transmission of packet $pkt_{i,j}$ are within the same slot and do not include the $sym_{UL-Ctrl}$ symbols reserved for UL control channels, the solution to the optimization problem defined for O-FAST must satisfy:

$$\sum_{s=0}^S (s - \lfloor s/14 \rfloor \cdot 14) \cdot X_s^{i,j} \leq 14 - sym_{UL-Ctrl} - d_i^S, \quad (9)$$

$$\forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\}$$

O-FAST aims to successfully serve all the packets in G_z . As a result, one variable $X_s^{i,j}$ with $s \in \{1, \dots, S\}$ must be equal to one for all packets $pkt_{i,j}$ in G_z , which is expressed as:

$$\sum_{s=0}^S X_s^{i,j} = 1, \quad \forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\} \quad (10)$$

In order to avoid scheduling conflicts, O-FAST must ensure that the same resources are not allocated for the transmission of more than one packet. To this end, O-FAST does not allocate more than R_{BW} RBs in each symbol, which is expressed in (11). If this constraint is fulfilled, it will be possible to allocate different RBs for the transmissions of different packets, and finally avoid conflicts:

$$\sum_{i=1}^{N_F} \sum_{j=1}^{N_{pkt}^{i,z}} X_s^{i,j} \cdot d_i^R \leq R_{BW}, \forall s \in \{1, \dots, S\} \quad (11)$$

If a packet demands more than R_{BW} resources, the d_i resources must be allocated in d_i^S consecutive symbols, which is expressed with the following constraint:

$$X_s^{i,j} \cdot R_{BW} + \sum_{s'=s+1}^{s+d_i^S-1} \sum_{\substack{m \in [1, N_F] \\ m \neq i}} \sum_{n \in [1, N_{pkt}^{i,z}]} X_{s'}^{m,n} \leq R_{BW}, \quad (12)$$

$\forall s \in \{1, \dots, S\}, \forall j \in \{1, \dots, N_{pkt}^{i,z}\}, \forall i \in \{1, \dots, N_F\}$ with $d_i^S > 1$

To summarize, the O-FAST scheduling proposal is defined with the objective function in (5), and the constraints in (6), (7), (8), (9), (10), (11), and (12). The optimization problem results in a Binary Integer Programming (BIP) problem where the unknown variables are binary variables.

IV. EVALUATION SCENARIO

We evaluate the proposed 5G-TSN scheduling scheme in an industrial plant scenario covered by a single-cell 5G private network (NPN) that is integrated with an industrial TSN network following the architecture depicted in Fig. 3. The scenario considers the deployment of a closed-loop supervisory application [8] where a PLC (Programmable Logic Controller) receives monitoring data from N_F sensors (S_1, S_2, \dots, S_{N_F}), and transmits a command to actuator A in Fig. 3. N_F TSN flows from the sensors to the PLC pass through the 5GS bridge. We evaluate the performance of the scheduling scheme with values of N_F ranging between 10 and 30. For each TSN flow F_i , the corresponding sensor generates packets of size $size_i$ with a periodicity p_i . Following [8], $size_i$ and p_i for each F_i are randomly selected between 40 and 250 bytes, and between 4 and 20 ms, respectively. Each packet must be received before the next packet is generated [8]. Consequently, the E2E latency requirement (l_i^{e2e}) for each TSN flow F_i is set equal to p_i . The time at which a TSN packet $pkt_{i,j}$ arrives to the 5GS ($A_{i,j}$) and the maximum latency that 5GS must guarantee (l_i^{5G}) are computed using (13) and (14), respectively, derived following the analysis in [9]:

$$A_{ij} = l_{sensor} + l_{link-ingress_i} + l_{TSN-ingress_i} + t_{DS-TT} \quad (13)$$

$$l_i^{5G} \leq l_i^{e2e} - A_{ij} - l_{link-egress_i} - l_{TSN-egress_i} - t_{NW-TT} - l_{PLC} \quad (14)$$

In (13) and (14), l_{sensor} and l_{PLC} represent the application processing time at the sensor and PLC, respectively, $l_{TSN-ingress_i}$ and $l_{TSN-egress_i}$ represent the latency experienced by the TSN flow F_i in the TSN bridges in the path between the sensor i and the 5GS bridge, and between the 5GS bridge and the PLC, respectively, and $l_{link-ingress_i}$ and $l_{link-egress_i}$ represent the propagation time it takes a packet to travel through the links between the sensor i and the 5GS bridge, and between the 5GS bridge and the PLC, respectively. t_{DS-TT} and t_{NW-TT} are the processing times at the DS-TT and NW-TT.

We simulate a 5G network in Matlab with 20 MHz, a sub-carrier spacing (SCS) equal to 30 kHz, and operating in TDD mode following recommendations in [10] and [11]. The UEs (industrial devices) are deployed under Line of Sight (LoS) conditions with the gNB, and utilize MCS12 (Modulation and Coding Scheme) from Table 1 in [12] for their transmissions. This MCS guarantees a good trade-off between robustness and

transmission rate under LoS conditions as it uses 16QAM (with modulation order $Q_m=4$) and coding rate $R=434/1024$.

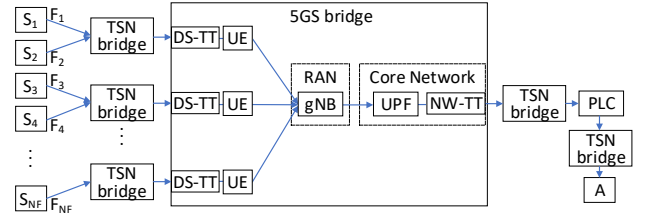


Fig. 3. 5G-TSN integrated network under evaluation.

V. PERFORMANCE RESULTS

We compare the performance of the proposed O-FAST scheduling scheme with a commonly used CG (Configured Grant) scheduling scheme [3][4]. The reference scheme configures a single UL grant for each TSN flow F_i , and periodically assigns radio resources to each flow. The periodicity is set equal to the periodicity p_i of each TSN flow F_i . The scheme assigns each TSN flow the number of radio resources necessary to satisfy its demand d_i . For a fair comparison with O-FAST, the reference scheme serves first those TSN flows with more stringent latency requirements.

Fig. 4 depicts the percentage of scheduling problem for which O-FAST and the reference scheme achieve a solution that meets the latency requirement for all TSN flows as a function of the number N_F of TSN flows. Fig. 4 shows that O-FAST outperforms the reference scheme. The reference scheme significantly degrades its capacity to satisfy the latency requirements of all the TSN flows as the number N_F of TSN flows increases. This is due to the conflicts that can occur when a CG scheduling scheme periodically assigns radio resources to each flow independently of each other, and each flow has different periodicity. In this case, two or more flows may receive the same radio resources at the same time (Fig. 1), and this results in packet collisions as shown in Fig. 5. The figure shows that the percentage of packets that collide with the reference scheme increases with the number N_F of TSN flows. Such collisions are critical for industrial applications that can demand reliability levels as high as 10^{-9} [8]. On the other hand, O-FAST guarantees the absence of packet collisions as its resource allocation avoids conflicts between TSN flows independently of the number of flows and their periodicity. O-FAST can then satisfy the latency requirements of all TSN flows for up to 25 TSN flows. The capacity of O-FAST to satisfy the latency requirements of all TSN flows decreases when N_F increases to 30. This is not due to packet collisions between different TSN flows, but to the fact that there are no feasible solutions that can satisfy the stringent latency requirements for all TSN flows when N_F increases to 30. In any case, O-FAST still significantly outperforms the reference scheme for $N_F=30$.

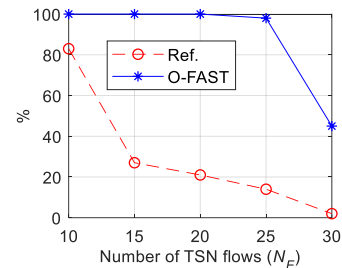


Fig. 4. Percentage of scheduling problems for which the latency requirements for all the TSN flows are satisfied as a function of N_F .

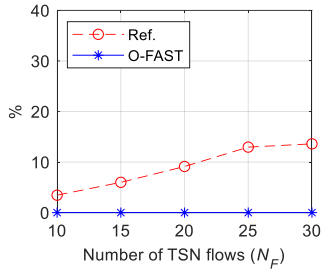


Fig. 5. Percentage of packet collisions as a function of N_F .

Fig. 6 compares the average latency experienced in the 5G by the packets of the different TSN flows with O-FAST and the reference scheme. The figure shows that O-FAST significantly reduces the average latency compared to the reference scheme. This is because the reference scheme allocates the radio resources that minimize the latency for the first packet of each TSN flow. The allocated radio resources then repeat with a periodicity equal to the periodicity p_i of the TSN flow, and do not change even if there are other radio resources available that could decrease the latency for following packets of the TSN flow. This results in that all packets of a TSN flow experience the same latency with the reference scheme, i.e. the jitter is null. On the other hand, O-FAST flexibly allocates to each packet the radio resources that minimize the latency based on the current resource availability. The flexibility of O-FAST comes at the expense of a small jitter (55-300 μ s when N_F varies between 10 and 30 flows). We should note that the jitter can be eliminated at the NW-TT as it can hold the packets and forward them to the next node at the departure time ($A_{i,j} + l_i^{5G}$) established by TSN.

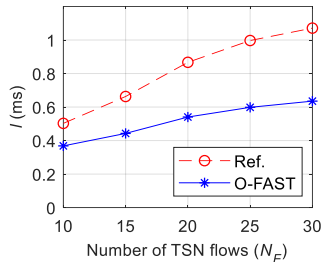


Fig. 6. Average latency (l_i^{5G}) experienced by the packets of the different TSN flows as a function of N_F .

Previous results have shown that O-FAST can avoid scheduling conflicts (and hence packet collisions) and significantly reduce the latency compared to traditional 5G CG scheduling schemes. This results in that O-FAST can effectively support large number of TSN flows with different periodicities. These gains come at the expense of a larger computational cost to decide the resource allocations. The reference scheme can take its scheduling decisions at the millisecond level. On the other hand, O-FAST can require between 1 minute and 1.5 hours to solve the scheduling problem in scenarios with N_F between 10 and 30 flows³. These values can be reduced with sub-optimal scheduling solutions that are left for future work. However, it is important to highlight that O-FAST does not need to be executed in real time since the scheduling solutions can be derived during the offline planning phase of the industrial plant just like it is done for the scheduling of TSN networks. Such planning is done

when the industrial plant is deployed, and the scheduling decisions are maintained as long as the industrial layout is maintained. As a result, O-FAST can compute offline its scheduling decisions, and these decisions are maintained for long periods of time.

VI. CONCLUSIONS

This study has presented a novel 5G configured grant scheduling scheme to support TSN traffic characteristic of industrial networks. The scheme coordinates its scheduling decisions with the TSN network and uses information about each TSN traffic flow to schedule multiple TSN flows with different periodicities. The proposed scheme assigns multiple configured UL grants to each TSN flow and adapts the resource allocation on a hyperperiod basis to avoid scheduling conflicts among TSN flows. The study shows that, compared to existing 5G CG scheduling schemes, the proposed scheme can significantly augment the number of TSN flows that satisfy their E2E latency requirements when transmitted over 5G. The proposed scheme also significantly reduces the average 5G E2E latency at the expense of larger computational times to decide the resource allocations. We should though note that the scheduling decisions are planned offline (non-real-time) when the industrial layout is done, just like it is done with industrial TSN networks.

ACKNOWLEDGMENT

This work has been funded by European Union's Horizon Europe Research and Innovation programme under the Zero-SWARM project (No 101057083), by MCIN/AEI/10.13039/501100011033 through the project PID2020-115576RB-I00, and by Generalitat Valenciana (CIGE/2022/17), and UMH's Vicerrectorado de Investigación grants.

REFERENCES

- [1] 5G-ACIA, *Integration of 5G with Time-Sensitive Networking for Industrial Communications*, white paper, Dec. 2020.
- [2] D. Ginthör, *et al.*, "5G RAN Slicing for Deterministic Traffic", in *proc. 2021 IEEE Wireless Communications and Networking Conference (WCNC)*, Nanjing, China, 2021, pp. 1-6.
- [3] Y. Cai, *et al.*, "Dynamic QoS mapping and adaptive semi-persistent scheduling in 5G-TSN integrated networks", *China Communications*, vol. 20, no. 4, pp. 340-355, April 2023.
- [4] Y. Zhang, *et al.*, "Wireless/wired integrated transmission for industrial cyber-physical systems: risk-sensitive co-design of 5G and TSN protocols", *Sci. China Inf. Sci.*, vol. 65, 110204, 2022.
- [5] 3GPP TR 22.804 V16.3.0, "Study on Communication for Automation in Vertical Domains", Release 16, 2020.
- [6] 3GPP, TS 23.501 V17.3.0, "System Architecture for the 5G System", Release 17, 2021.
- [7] B. Coll-Perales, *et al.*, "End-to-End V2X Latency Modeling and Analysis in 5G Networks", *IEEE Transactions on Vehicular Technology*, vol 72, no. 4, pp. 5094 - 5109, April 2023.
- [8] K. Montgomery, *et al.*, "Wireless User Requirements for the Factory Workcell", *NIST Advanced Manufacturing Series*, 300-8, Nov. 2020.
- [9] A.Larrañaga, *et al.*, "Analysis of 5G-TSN Integration to Support Industry 4.0", *Proc. IEEE ETFA 2020*, Vienna, Austria, Sept. 2020.
- [10] 3GPP, TR 38.824. V16.0.0, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)", Release 16, 2019.
- [11] 3GPP, TS 38.104, V17.10.0, "Base Station (BS) radio transmission and reception", Release 17, 2023.
- [12] 3GPP, TS 38.214, V16.1.0, "Physical layer procedures for data", Release 16, 2020.

³ The scheduling problems are solved using Matlab's toolbox Parallel Computing, Optimization y Symbolic Math. The values reported have been

measured with the toolbox executed in a server with an Intel(R) Core (TM) i7-5930K @ 3.50GHz CPU and 16GB RAM.

Analysis of 5G-TSN Integration to Support Industry 4.0

Ana Larrañaga¹, M. Carmen Lucas-Estañ², Imanol Martinez¹, Iñaki Val¹, Javier Gozalvez²

¹HW and Communication Systems Area, Ikerlan Technology Research Centre, Mondragón 20500, Spain

²UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain

{ana.larranaga, imartinez, IVal}@ikerlan.es, {m.lucas, j.gozalvez}@umh.es

Abstract—Time Sensitive Networking (TSN) is becoming the standard Ethernet-based technology for converged networks of Industry 4.0 due to its capacity to support deterministic latency requirements. However, it cannot provide the required flexibility to support mobile industrial applications required for the factories of the future. This could be enabled through the integration of wireless technologies in factories, and in particular of 5G and Beyond networks since they have been designed to support ultra-reliable and low-latency communications. This has triggered significant interest to integrate 5G and TSN networks, and first frameworks for such integration have been defined. However, the work is at early stages and the solutions to effectively integrate the two networks so that 5G can support TSN QoS levels are yet to be designed. This paper discusses current research and standardization work on 5G-TSN integration, and quantifies for a closed loop control application the 5GS bridge delay. The paper uses an example based on 5G-ACIA [1] to discuss open technical and research challenges to effectively integrate 5G and TSN.

Keywords—TSN; 5G; 5G-TSN; Deterministic; Time-critical; Industry 4.0; QoS support

I. INTRODUCTION

Future smart factories will provide wireless and wired connectivity between the entire production line in order to exchange data and take decisions autonomously. To achieve this paradigm, highly reliable and deterministic communications will be key. In the last decade, industry has introduced wired communications in order to support the strict requirements that it needs. Time-Sensitive Networking (TSN) Ethernet-based technology is being developed to provide deterministic, real-time and ultra-reliable communications. However, TSN cannot guarantee the flexibility required by the factories of the future to support mobility in some industrial applications. Wireless technologies can provide the required flexibility, and the integration of TSN with wireless technologies will be key. 5th Generation (5G) networks are considered one of the best options to support the Industry 4.0 paradigm as it has been designed to support Ultra-Reliable Low Latency Communications (URLLC). This paper discusses the current procedures defined in 3GPP standards for 5G and TSN integration, and illustrates with a closed loop control application example based on 5G-ACIA [1] how to quantify the 5GS bridge delay as a relevant component for an effective 5G-TSN integration. The example is used to identify existing open technical and research questions and limitations that require further study for an effective and efficient 5G-TSN integration.

The rest of the paper is structured as follows. Section II and III explain the most important TSN and 5G-NR (New Radio)

features. Section IV introduces the state of the art of 5G-TSN integration. Section V presents the use case and analyzes the current procedures defined for 5G and TSN integration. Section VI presents main conclusions.

II. TIME SENSITIVE NETWORKING (TSN)

TSN is an extension of traditional Ethernet data-link layer developed to support deterministic transmissions with real time and high reliability requirements. TSN is a set of layer 2 open standards specified by IEEE 802.1.

A TSN network is composed of end stations that are the source and destination of the TSN streams (also referred to as talkers and listeners respectively), and bridges that are special Ethernet switches capable of transmitting and receiving Ethernet frames of a TSN stream following a schedule. TSN standard IEEE 802.1Qcc [2] defines three network management models: a fully distributed model, a centralized network and distributed user model, and a fully centralized one. This paper focuses on the fully centralized model since it is the one that allows the most efficient use of network resources. In this model, the TSN bridges communicate their network capabilities and active topology information to the CNC (Centralized Network Configuration) node. End stations send their TSN stream requirements to the CUC (Centralized User Configuration) node that forwards this information to the CNC (the CUC may adapt these TSN end station stream requirements before forwarding them to the CNC); the CNC has then a complete view of the TSN network and it can determine the configuration of each bridge (e.g. the scheduling) to fulfil the TSN stream requirements. The CNC can send specific TSN configuration information to the CUC, which may send configuration information to the end stations.

TSN defines a Time Aware Shaper (TAS) scheduler in IEEE 802.1Qbv [2] for guaranteeing the transmission of high priority deterministic traffic in a bounded time. To this end, the Ethernet traffic is divided into different classes. There can be up to 8 traffic classes and Best Effort (BE) is the default traffic class. TAS scheduler establishes on every egress port of each bridge the time-interval each traffic class has access to the channel so that the time requirements of high priority traffic are met. TAS scheduler creates a time-protected channel used by TSN traffic class in order to guarantee their deterministic transmission time. When data enters in a TSN bridge, it is forwarded to the corresponding egress port. The data is classified in different queues based on the traffic class. Each queue has different priority, from less priority 0 to high priority 7. The queues are controlled by a Gate Control Lists (GCL) that determines which queues are opened or closed at each time

instant. If there are more than one queue opened at the same time, the forwarding of frames depends on the priority of the queue. In order to support TAS scheduling, all the nodes (talkers, listener and bridges) in the network are synchronized using one master and several slaves' clocks, as defined in the IEEE 802.1AS standard [3].

III. 5G SYSTEM

5G System (5GS) has two main network components: the Radio Access Network (RAN) and the Core Network (CN). The RAN includes the UE (User Equipment) and the gNB (new generation Node B). The 5G CN is highly flexible and consists of a series of independent network functions for the user plane (UP) and the control plane (CP) [4]. UP carries the network user traffic and CP carries signaling and control traffic. The UPF (User Plane Function) handles the UP path between the gNB and the external data network (e.g. a TSN network). In the control plane, the SMF (Session Management Function) interacts with the user plane to establish or modify PDU (Protocol Data Unit) sessions that define the connectivity services between UE and the data network. The PCF (Policy Control Function) provides policy rules to the SMF. The AF (Application Function) interacts with the PCF for policy control and can influence on traffic routing in the UP. To support low latency communications, 5G reduces the processing times at UE and gNB, defines more flexible frames structures, and the use of shorter transmission slots based on different numerologies. These numerologies represent the spacing between sub-carriers and result in slots from 1 to 0.0625 ms. It is possible to transmit only using part of a slot, referred to as minislot. Another key aspect to reduce latency is the definition of grant-free scheduling for uplink transmissions: gNB preallocates resources for a UE. When the UE wants to transmit, it does not need to ask resources to the gNB (skipping the delay associated to this process), and it can transmit data immediately.

5G supports different QoS Flows depending on the service (a QoS Flow is the finest granularity of QoS differentiation in a PDU Session). A QoS Flow in 5GS may either be GBR (Guaranteed Bit Rate) or Non-GBR. GBR QoS Flows may use either GBR or Delay-critical GBR resources. Each QoS Flow is related with a 5G QoS profile that contains, among others, the following QoS parameters: the guaranteed and maximum flow bit rate (GFBR and MFBR respectively), maximum packet loss rate, and a 5G QoS Identifier (5QI). The 5QI is a scalar that is used as a reference to a set of QoS characteristics: priority, packet delay budget (PDB), packet error rate (PER), and maximum data burst volume (MDBV). The PDB is the maximum time that a packet may be delayed between the UE and the UPF that connects to the data network. The PDB is explicitly divided into RAN PDB and CN PDB. The MDBV is the largest amount of data that the RAN is required to serve within a period of RAN PDB. 5QI values can be standardized in [4], preconfigured in the RAN or dynamically defined.

IV. INTEGRATION OF 5G AND TSN

This section presents some of the recent proposals for integration of TSN and 5G that are currently being studied by 3GPP, Avnu Alliance and 5G-ACIA, among others.

A. Architecture for the 5G-TSN Integration

3GPP Release 16 [4] defines the basic functionalities and architecture to support Time Sensitive Communications (TSC) and integrate TSN and 5G. 3GPP Release 16 focused first on TSN networks with fully centralized management. 5GS is integrated within the TSN network as a logical TSN bridge. 3GPP introduces a TSN translator (TT) functionality at the interconnection points between both networks, that is in the device side (DS) and in the network side (NW). The DS-TT and NW-TT act as TSN ingress and egress Ethernet ports in the logical 5GS TSN bridge. Thanks to the TT functionality, the 5GS appears as any other TSN bridge for the TSN network. The DS-TT and NW-TT are connected to UE and UPF respectively, and their functionality is to configure all parameters necessary to coordinate 5G and TSN. In the control plane, the TSN AF is in charge of exchanging control data between 5GS and the TSN CNC in order to configure 5GS as a TSN bridge.

B. 5GS bridge information

5GS reports 5GS bridge information to the TSN network. This information is used by the CNC to adequately configure the TSN network so that it supports the QoS requirements of TSN traffic. The 5GS bridge information includes information about bridge ID, ethernet ports and their MAC address, neighbor discovery information, supported minimum/maximum bridge delay, propagation delay and traffic classes. The 5GS calculates the bridge delay for each port pair and traffic class. The bridge delay represents the time that a frame for a specific traffic class takes to be forwarded from an ingress port to an egress port. The bridge delay (BD) is calculated as follows:

$$BD = t_{residence,UE-DSTT} + d_{dependent} + d_{independent} \quad (1)$$

where $t_{residence,UE-DSTT}$ is the time that a packet needs to be forwarded from DS-TT to UE and vice versa and is referred to as residence time, $d_{dependent}$ is the delay experienced in the bridge that is dependent on the frame length, and $d_{independent}$ is the delay to store and forward a frame through the bridge and that is independent of the frame length; $d_{independent}$ does not depend on the frame length but is a function of the processing time, the 5G frame structure, scheduling, HARQ, etc [5].

C. 5GS bridge configuration

5GS needs information about the QoS requirement of each TSN stream to adequately configure the different management mechanisms at the RAN and CN. QoS requirements of a TSN stream are identified by a stream ID, a priority, a traffic class, a maximum frame size, a maximum latency and source/destination of the flow, among other parameters. In order to support TSN streams in 5G, the TSN QoS requirements need to be mapped to a 5G QoS profile. TSN streams are supported in 5GS as Delay-critical GBR QoS flows. The 5QI selected for the TSN QoS flow should have a MDBV equal to or higher than the TSC Burst Size, and a PDB that is equal or higher than the bridge delay for this traffic class. The QoS mapping between TSN QoS traffic class and 5G QoS profiles is carried out at the TSN AF and the PCF. The TSN AF decides the TSN QoS parameters (i.e. priority and delay) based on the received configuration information from the CNC. The TSN QoS information is sent to the PCF, which is a function of the

5G CN that maps the TSN QoS parameters into a 5QI. 3GPP standards also define the exchange of additional TSC Assistance Information (TSCAI) that may be used in the 5GS for a more efficient scheduling of periodic and deterministic traffic flows in the RAN. TSCAI provides traffic pattern parameters such as burst arrival time with reference to the ingress port, periodicity, and flow direction.

V. BRIDGE DELAY ESTIMATION FOR AN EFFECTIVE 5G-TSN INTEGRATION

Our research focuses on how to effectively integrate 5G and TSN to support latency-critical and deterministic industrial applications. The 5GS bridge and TT functionalities have a strong impact on this integration, so in this first study, we analyze the bridge delay considering the procedures used to establish an integrated 5G-TSN connection with TSN traffic. For the analysis, we consider an industrial scenario based on 5G-ACIA white paper [1] where the 5G-TSN network supports the connectivity of a logistic closed loop control use case. For this use case, a 5G UE is locally connected (wired) to one sensor and one actuator that communicate through the 5G wireless connection with a remote PLC (Programable Logic Controller). The sensor sends information to the PLC every 1 ms. In the next control cycle, the PLC sends a command to the actuator. In general, during the same control cycle the sensor writes data in the PLC and the actuator reads the data from the PLC. However, in this example the actuator reads the data transmitted from the PLC one control cycle later. Figure 1 shows the analyzed 5G-TSN architecture, that considers the deployment of a private 5G network with a dedicated local core network. The PLC is connected to a TSN bridge by means of a cable. The scenario includes two office computers that exchange plant information, but their data is considered best effort traffic. We focus the rest of the analysis on higher priority TSN traffic.

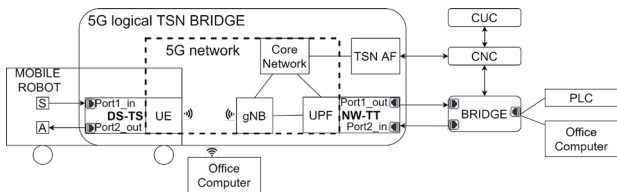


Figure 1. Example of 5G-TSN integrated network architecture.

Initialization: Talkers and listeners send requirements information to the CUC (Table 1) that forwards this information to the CNC. We assume only two streams of TSN traffic class.

TABLE I. TALKER REQUIREMENTS

Talker	Sensor (S)	PLC
Stream ID	Stream 1	Stream 2
Priority	4-7 (TSN traffic, highest priority)	
MaxFrameSize	64 Bytes	256 Bytes
FramePerInterv.	1	
Interval (Periodicity)	1 ms	
Maximum Latency	1 ms	

Physical network topology and bridge capabilities: The CNC has to discover the physical network topology and the TSN capabilities of each bridge. To this end, each bridge sends to the CNC the following data: bridge ID, ethernet ports and their MAC address, neighbor discovery information, supported

minimum/maximum bridge delay, propagation delay and traffic classes. Equation (1) and Table 1 parameters are used in order to calculate the minimum and maximum bridge delay of the TSN and 5GS bridge. Minimum bridge delays are calculated using the stream 1 (from sensor to PLC) which represents the minimum streams size, while maximum bridge delays are calculated using the stream 2 (from PLC to actuator) that represents the maximum stream size.

We consider that, when a TSN stream arrives to the egress port of the TSN bridge, it is transmitted immediately (we consider that the gate is open for this traffic class). In this case, the delay experienced in the bridge includes processing and transmission times. We consider a processing delay in the TSN bridge equal to $1.5 \mu\text{s}$ as presented in [6]. If we assume the use of a link with data bit rate equal to 100 Mbps (typical bit rate in industry for latencies between $25 \mu\text{s}$ -2 ms [7]), the minimum and maximum dependent delay (equal to the transmission time) can be computed as:

$$d_{dependent,TSNbridge}^{min} = \frac{64\text{Bytes} \times 8\text{bits}}{100\text{Mbps}} = 5.12 \mu\text{s} \quad (2)$$

$$d_{dependent,TSNbridge}^{max} = \frac{256\text{Bytes} \times 8\text{bits}}{100\text{Mbps}} = 20.5 \mu\text{s} \quad (3)$$

Finally, the minimum and maximum TSN bridge delays ($BD_{TSNbridge}^{min}, BD_{TSNbridge}^{max}$) are given by:

$$BD_{TSNbridge}^{min} = 5.12 \mu\text{s} + 1.5 \mu\text{s} = 6.62 \mu\text{s} \quad (4)$$

$$BD_{TSNbridge}^{max} = 20.5 \mu\text{s} + 1.5 \mu\text{s} = 22 \mu\text{s} \quad (5)$$

In the 5GS bridge, we consider the use of a preconfigured 5QI in the RAN to calculate the bridge delay; this 5QI is associated to preconfigured values for the QoS parameters, PDB, MDBV and GFBR. These values are not standardized and can be established according to the network capabilities. In this section, we estimate the values for PDB, GFBR and MDBV that should be supported in this example. 5GS should support to transmit TSN stream 1 and stream 2 inside the required maximum latency. The PDB should be equal to or higher than the 5GS bridge delay. The PDB is calculated as:

$$PDB \leq l_{e2e} - l_{app} - l_{links} - BD_{TSNbridge} \quad (6)$$

where l_{e2e} is the required end-to-end latency (1 ms), l_{app} is the processing time at the application level in the end stations (we assume l_{app} equal to the 25% of the end-to-end latency, i.e., $250 \mu\text{s}$), l_{links} represents the propagation time in the wired links (l_{links} is considered equal to 0 based on the short distances between devices), and $BD_{TSNbridge}$ is the TSN bridge delay as calculated in (4) and (5) for stream 1 and 2 respectively. Based on (6), PDB is equal to $743.38 \mu\text{s}$ and $728 \mu\text{s}$ for stream 1 and 2 respectively.

Using the estimated value for PDB, it is possible to calculate the minimum bit rate needs to be guaranteed as:

$$GFBR_{stream1} > \frac{64\text{Bytes} \times 8\text{bits}}{PDB_{stream1}} = 688.7\text{kbit/s} \quad (7)$$

$$GFBR_{stream2} > \frac{256\text{Bytes} \times 8\text{bits}}{PDB_{stream2}} = 2813.1\text{kbit/s} \quad (8)$$

where $GFBR_{stream1}$ and $GFBR_{stream2}$ are the minimum bitrates to transmit stream 1 and stream 2 respectively. We compute the minimum MDBV that 5GS has to support to

transmit stream 1 and stream 2, considering the use of Robust Header Compression (RHC) and 5G L2 header of 6 bytes [8]. RHC compresses the Ethernet header from 18 Bytes to 2 Bytes. In this case, the $MDBV_{stream1}$ is equal to 54 Bytes and $MDBV_{stream2}$ is equal to 246 Bytes. We can then estimate the maximum (stream 2) and minimum (stream 1) 5GS bridge delays as:

$$BD_{5GS}^{min} < \frac{MDBV_{stream1} \times 8}{GFBR_{stream1}} \approx 627 \mu s \quad (9)$$

$$BD_{5GS}^{max} < \frac{MDBV_{stream2} \times 8}{GFBR_{stream2}} \approx 699.6 \mu s \quad (10)$$

After establishing a PDU session, TSN AF receives UE-DS-TT residence time value and it deduces the in/out port pairs. So, the bridge delay for each port pair and traffic class is updated adding the $t_{residence, UE-DSTT}$. Finally, BD_{5GS} for the 5GS bridge is given by $(627 \mu s + t_{residence, UE-DSTT})$ and $(699.6 \mu s + t_{residence, UE-DSTT})$ and they have to be smaller or equal to the estimated PDBs, $743.38 \mu s$ and $728 \mu s$ respectively.

Network configuration: When the CNC receives the bridges information, it calculates the gate control list or GCL of each egress port as well as the paths by which the streams have to be forwarded. In this use case there are two streams with different flow directions. The streams are transmitted through different egress ports within the same bridge, so each egress port has its on GCL. Figure 2 represents the communication path for stream1, as well as the time instant at which stream1 arrives and leaves the 5GS bridge.

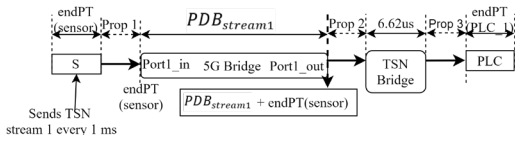


Figure 2. The end-to-end path of stream 1.

The CNC sends to the 5GS bridge information about the arrival time of each data stream and their periodicity. With this information, 5G can determine the traffic pattern information, make the QoS mapping, and configure the schedule. As presented in [9], the mapping between TSN stream priorities and 5QI priorities is flexible. 5G establishes priority values for standardized 5QI up to 90. Priority assigned for standardized 5QI with delay critical GBR resources is lower than 24, and for the 5QI related to discrete automation applications is equal to 19. In this context, we consider a priority equal to or lower than 19. The QoS parameters for each QoS flow are presented in Table 2. The QoS profile and TSCAI information are sent to gNB in order to calculate the scheduling.

TABLE II. QOS PARAMETERS FOR THE DATA STREAMS

	Resource type	Priority	MDBV	GFBR	PDB
QoS UL	DelayCritical	<19	64 Bytes	>688.7 kbit/s	$\leq 743.4 \mu s$
QoS DL	DelayCritical	<19	256 Bytes	>2813.1 kbit/s	$\leq 728 \mu s$

VI. CONCLUSIONS

This WIP paper has discussed and analyzed the integration of 5G and TSN to support latency-critical and deterministic

industrial applications. An important aspect of this integration is the bridge between the technologies. The paper quantifies the 5GS bridge delay for a close loop control industrial application example based on 5G-ACIA white paper [1]. From the work carried out in this paper, it is possible to identify some limitations that need to be further studied to effectively integrate 5G and TSN. For example, all standardized 5QIs in 5G (presented in [4]) establish a PDB equal to or higher than 5 ms. This PDB is not adequate for this kind of applications. Private 5G networks allow to deploy local CN that will allow to reduce considerably the delay experienced by data in the CN. For example, Nokia presented in [9] that a CN PDB lower than $10 \mu s$ can be achieved considering a local optimized CN deployment. An adequate designed and dimensioning of the RAN of the 5GS will allow to meet PDB values lower than 1 ms as shown in [10]. Although pre-configured or dynamic 5QI values can be used, it might be interesting the definition of new standardized 5QI values for this kind of network deployments.

To support this kind of periodic traffic, the use of a semipersistent scheduling in DL or configured scheduling with periodic resources seems to be the most adequate scheduling options. However, it is important to analyze the case when the traffic periodicity does not match with the periodicities that can be supported in 5G RAN. How to address this issue needs to be further studied.

Finally, the 3GPP standard defines the basics concepts to estimate the 5GS bridge delay. However, the exact mapping between the 3GPP parameters and the bridge delay is not presented and is up to the implementation. How to estimate the bridge delay before initiating or establishing a QoS flow is a difficult task in 5GS. To this end, maybe it might be necessary to transmit some initial data to measure the PDB in 5GS before the TSN streams are established.

REFERENCES

- [1] 5G-ACIA, "A 5G Traffic Model for Industrial Use Cases (White Paper)," no. November, 2019.
- [2] IEEE, *IEEE 802.1Q: Standard for Local and Metropolitan Area Networks--Bridges and Bridged Networks*. 2018.
- [3] IEEE, *IEEE 802.1AS: Standard for Local and metropolitan area networks--Time and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks*. 2011.
- [4] 3GPP; Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS) Stage 2 (Release 16) 3GPP TS 23.501 V16.4.0, March, 2020.
- [5] C. Mannweiler *et al.*, "Reliable and deterministic mobile communications for industry 4.0" 24. *ITG-Symposium Mob. Commun. - Technol. Appl.*, pp. 64–69, 2020.
- [6] D. Pannell, "Choosing the Right TSN Tools to meet a Bounded Latency" in *IEEE SA Ethernet & IP @ Automotive Technology Day*, September, 2019.
- [7] A. Nasrallah *et al.*, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 1, pp. 88–145, 2019.
- [8] M. Khoshnevisan, "5G Industrial Networks With CoMP for URLLC and Time Sensitive Network Architecture," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 947–959, 2019.
- [9] Nokia, "5G plug-and-produce. How the 3GPP 5G System facilitates Industrial Ethernet deployments (White Paper)." 2020.
- [10] 3GPP; Technical Specification Group Radio Access Network; Study on self evaluation towards IMT-2020 Submission (Release 16) v16.1.0, September. 2019.

5G Configured Grant Scheduling for 5G-TSN Integration for the Support of Industry 4.0

Ana Larrañaga¹, M. Carmen Lucas-Estañ², Imanol Martinez¹, Javier Gozalvez²

¹HW and Communication Systems Area, Ikerlan Technology Research Centre, Mondragón 20500, Spain

²UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain

{ana.larranaga, imartinez}@ikerlan.es, {m.lucas, j.gozalvez}@umh.es

Abstract— Factories are evolving towards digitalized data-based ecosystems under the paradigm of the Industry 4.0 where new industrial services allow the implementation of more robust, resilient and customized manufacturing systems. Such services (e.g., digital twins, extended reality or cooperative robots) will require highly reliable and deterministic communication networks capable of supporting stringent latency and reliability requirements. 5G networks and their future evolution have the necessary capabilities to meet these requirements. However, the use of 5G in industrial environments requires its effective and efficient integration with Time Sensitive Networking (TSN), which is becoming the standard wired technology for Industry 4.0 environments. TSN provides unprecedented deterministic service levels with perfectly bounded latencies. The integration of the industrial 5G and TSN networks will be key to support the flexibility and determinism demanded by the Industry 4.0 paradigm. A critical aspect to achieve this integration is the coordination of the schedulers of both networks. TSN has information about the capabilities of the 5G-TSN integrated network, and it is in charge of deciding the path and scheduling for each TSN traffic flow. The scheduling in 5G must be done according to the scheduling decisions and information provided by TSN to guarantee the end-to-end latency requirements of TSN traffic. In this context, this paper proposes a novel Configured Grant (CG) scheduling scheme for 5G integrated into a TSN network that aims to meet the latency requirements of the different TSN flows. The proposed scheme exploits the information provided by TSN about the characteristics of the TSN traffic to coordinate its decision with the scheduling of TSN. This study demonstrates that the proposed scheduling scheme considerably increases the number of TSN flows that can be satisfactorily served in the integrated 5G-TSN network compared with a commonly used CG scheduling scheme.

Keywords—5G; 5G-TSN integration; Deterministic; Low Latency; TSN; Configured Grant; scheduling; Industry 4.0.

I. INTRODUCTION

Factories are evolving towards digitalized data-based ecosystems where new industrial services, such as digital twins, extended reality (XR) or cooperative robots, emerge and allow the implementation of more robust and resilient manufacturing systems. Industrial environment and processes will be accurately monitored in real-time thanks to Industrial IoT (Industrial IoT or IIoT) networks that will connect machines, mobile robots, sensors, and user terminals, allowing data to be collected, analyzed, and distributed instantly [1]. Factories of the future will then require highly reliable communication networks capable of supporting the stringent latency, bandwidth, and reliability requirements of

demanding industrial applications. In addition, the communication networks that will support Industry 4.0 must be able to adapt their operation to the changing connectivity and data transfer needs of industrial production systems. 5th Generation (5G) networks and their future evolution have the necessary capabilities to meet these requirements. In fact, 5G technology is considered a key catalyst for the digital transformation of the industry.

The use of 5G in industrial environments requires its effective and efficient integration with industrial wired networks currently implemented in factories to meet the strict reliability and resilience requirements of industrial applications. Time-Sensitive Networking (TSN) is becoming the standard wired technology for Industry 4.0 environments. TSN is based on Ethernet and provides unprecedented deterministic service levels with perfectly bounded latencies. However, TSN does not have the flexibility and reconfiguration capabilities required by the factories of the future. On the other hand, 5G is highly reconfigurable and flexible but cannot efficiently and scalably support deterministic services. Both networks, therefore, have complementary capabilities and the integration of the industrial 5G and TSN networks is key to support the Industry 4.0 paradigm.

3GPP standards already define the framework for the efficient integration of TSN and 5G networks. The 5G network is integrated into the TSN network as a logical TSN bridge of the TSN network (see Fig. 1). A TSN bridge is an Ethernet switch that receives and transmits TSN frames based on a scheduling. In this integrated network, the communication path between two end-devices (for example, a sensor and an actuator) is established through one or more TSN bridges and the 5G logical bridge. In this context, a critical aspect to meet the end-to-end latency requirements of the industrial applications is the coordination of the scheduling of both networks. In the integrated 5G-TSN network, the Central Network Configuration (CNC) has information about the whole network topology and capabilities of the bridges. Based on this information, the CNC decides the communication path and scheduling for each TSN traffic flow. The scheduling decision establishes the arrival and departure time of the packets of each TSN flow in each (TSN and 5G logical) bridge. The scheduling of the packets of the TSN flows in the 5G network has to be done in a way that these arrival and departure times calculated by TSN are satisfied.

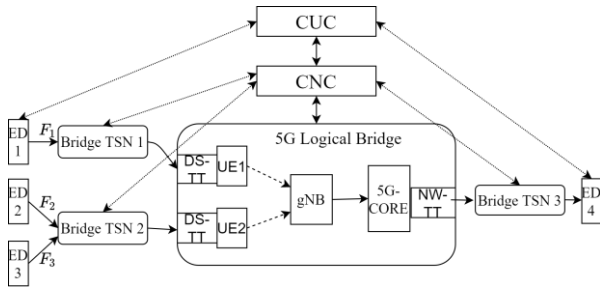


Fig. 1. 5G-TSN integration model.

Some current researcher work have studied scheduling schemes for 5G networks integrated into TSN networks as a logical bridge. For example, [2] proposed a dynamic scheduling scheme for deterministic traffic. The proposed scheme is based on optimization techniques and allocates resources dynamically for each TSN packet to improve radio resource efficiency while guaranteeing the latency requirements of the TSN traffic. With dynamic scheduling, the end users need to request resources and/or receive a grant before transmitting a packet. This exchange of signaling can increase the latency of communication. Configured Grant (CG) in uplink (UL) and Semi-Persistent Scheduling (SPS) in downlink (DL) allocate radio resources periodically for each TSN flow based on the periodicity of the packets. Configured Grant or Semi-Persistent Scheduling are the most proper scheduling (in UL and DL, respectively) for this type of traffic based on its low latency requirements and periodicity. An SPS scheme has been studied in [3] for DL Time Sensitive Communications (TSC) flows. The authors in [3] proposed the adaptation of the modulation and coding scheme (MCS) used for the transmission of each packet based on past channel conditions. [4] proposed a predictive multi-priority scheduling mechanism based on SPS for the 5G-TSN network where periodic and aperiodic data is considered. According to [5], CG or SPS are the most proper scheduling (in UL and DL, respectively) for TSC traffic demanding low latency requirements. The allocation of periodic resources to TSN flows with different periodicity may result in conflicts: the resources allocated to different TSN flows may overlap after several periods. This is illustrated in Fig. 2. The example in Fig. 2 shows two TSN flows, F_1 and F_2 , with periodicity 3 and 5 respectively. Although the radio resources allocated for the transmission of the first packet of each TSN flow are different, some of the resources allocated for the transmission of the third and fifth packets of F_1 and F_2 respectively overlap in frequency and time due to the different packet periodicities of F_1 and F_2 . This is a critical issue that, to the best of the author's knowledge, has not been previously addressed. In this context, this paper progresses the state of the art by proposing a novel scheduling scheme for 5G integrated into TSN networks that addresses this problem. The main objective of the proposed scheduling scheme is to meet the latency requirements of the different TSN flows based on the arrival and departure times calculated by the TSN network. To this end, the proposed scheduling scheme uses the information provided by the TSN network about the characteristics of the

TSN flows (packet sizes, periodicity, etc). The proposed scheduling scheme configures several grants for each TSN flow in order to avoid radio resource allocation conflicts among different TSN flows. This study demonstrates that the proposed scheduling scheme considerably increases the number of TSN flows that can be satisfactorily served in the integrated 5G-TSN network.

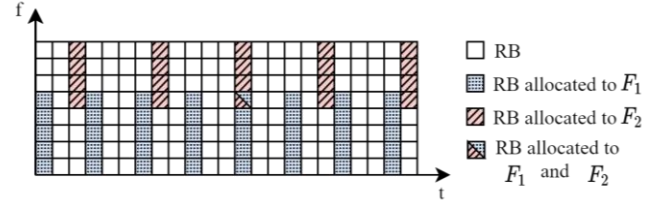


Fig. 2. Example of a radio resource allocation conflict where the same radio resources are allocated to more than one TSN flow when radio resources are allocated considering the packet periodicity.

The paper is organized as follows. Section II presents the current framework for TSN and 5G network integration. Section III presents the proposed 5G scheduler for 5G integrated into a TSN network. Section IV presents the evaluated scenario and Section V presents the performance achieved with the proposed scheme. Section VI concludes the paper.

II. 5G-TSN INTEGRATION

TSN is a set of open standards developed by IEEE 802.1. TSN has been designed to provide communications with a very high level of determinism: TSN guarantees the delivery of the data in a guaranteed time window [6]. This is achieved thanks to the introduction of several technical features, such as strict time synchronization (IEEE 802.1AS) and a traffic scheduler (IEEE 802.1Qbv) that reserves specific time slots for the transmission of high-priority traffic. 5G networks have the capacity to support low latency and highly reliable communications. 5G can complement TSN to provide the level of determinism and flexibility demanded by Industry 4.0. 3GPP standards define the framework for the integration of 5G in TSN networks. A TSN network is integrated by end devices (ED), which are the source and destination of TSN flows, and bridges that are interconnected using standard Ethernet links. In the integration model defined by the 3GPP, 5G appears as a (logical) bridge of the TSN network, as shown in Fig. 1 5G includes TSN translator (TT) functionality at the edges of the network: at the users (UEs) referred to device side TT or DS-TT and at the 5G Core Network referred to network side TT or NS-TT. The TTs are the interconnection points between 5G and TSN. The TTs allow 5G to inform the TSN network about its capabilities and current status, as well as to receive, interpret and apply TSN configuration commands from the TSN network. The TTs provide TSN bridge ingress and egress port operations. For example, TTs are responsible for holding and forwarding the TSN flows to minimize latency variation.

The integrated 5G-TSN network considers a centralized management model. The TSN network incorporates the Centralized User Configuration (CUC) and CNC functions.

The CUC receives information about the communication requirements for each TSN flow from the end devices. The CUC informs the CNC about these requests. The CNC collects information about the capabilities and current status of all the (TSN and 5G logical) bridges in the network [7][8]. For example, bridges inform about the minimum and maximum supported bridge delays, which are the minimum and maximum times respectively that a packet needs to be forwarded from an ingress port to an egress port of the bridge. TSN bridges also inform of the propagation delay or link delay, which is the time a packet takes to travel through the links that interconnect the nodes (end devices and bridges) from the specified port of the station to the neighboring port on a different station. The CNC then decides the path and schedule of all TSN flows using IEEE 802.1Q [9]. The CNC finally configures the (TSN and 5G logical) bridges according to the scheduling decision in order to guarantee the end-to-end requirements of the TSN flows. The CNC applies the time-aware shaper (TAS) defined in IEEE 802.1Qbv that establishes the arrival and departure time of each TSN flow at the ingress and egress ports of each TSN and 5G logical bridge, respectively.

5G translates the information about TSN QoS (Quality of Service) requirements received from the CNC to a 5G QoS Identifier (5QI) for each TSN flow. A 5QI defines a set of 5G QoS characteristics that describe the packet forwarding treatment that a QoS flow receives in the 5G network (between the UE and the last node in the 5G core network). A 5QI includes the priority level of a flow for the scheduling of resources, the packet error rate (PER), the maximum data burst volume (MDBV), and the packet delay budget (PDB), which indicates the maximum latency a packet can experience in the 5G network, among other QoS characteristics. 3GPP standards do not specify how TSN QoS requirements are mapped to 5G QoS requirements. Some proposals are found in [10], [11], and [12]. In addition to the 5QI, 5G determines Time Sensitive Communication Assistance Information (TSCAI). The TSCAI describes the TSN traffic characteristics: burst or packet arrival time with reference to the ingress port, periodicity, flow direction (uplink or downlink), and survival time¹ [8]. The TSCAI and the 5QI may be used by the gNB (new generation Node B) to decide the scheduling of radio resources for the TSN traffic in order to meet the end-to-end requirements of the TSN flows.

III. 5G CG SCHEDULING FOR THE SUPPORT OF TSN TRAFFIC

This work proposes a scheduling scheme for a 5G network integrated into a TSN network to support the stringent communication requirements of industrial applications. The main objective of the proposed scheduling scheme is to guarantee the 5G latency requirements for each TSN flow and ensure that packets are received at the egress port of the 5G logical bridge before the departure time indicated by the TSN network. To this end, the proposed scheduling scheme uses the information provided by the TSN network (5QI and TSCAI) to decide the radio resource allocation solution that better satisfies the latency requirements of all TSN flows. A radio resource is composed of one Resource Block (RB) of 12

subcarriers in the frequency domain and one Orthogonal Frequency Division Multiplexing (OFDM) symbol in the time domain. The scheduling proposal is based on CG, and it allocates radio resources periodically for each TSN flow. The proposed scheduling scheme configures several grants for each TSN flow, if needed, to ensure that radio resources are not assigned to more than one TSN flow simultaneously.

Let us consider that N_F TSN flows must be transmitted through the 5G virtual bridge toward the end devices. For each TSN flow F_i (with $i=1, \dots, N_F$), packets of size s_i arrive with a periodicity p_i ; s_i is given by the MDBV of the 5QI and p_i in the TSCAI. s_i and p_i can be different for each TSN flow. We consider that all the TSN traffic that arrives at the 5G network must be transmitted in UL (flow direction is represented by f_i). Each flow demands d_i radio resources calculated as a function of s_i and the MCS to use for the packet transmission. Packets in a TSN flow F_i are referred to as $pkt_{i,j}$ where j indicates the number of the packet in the flow. The arrival and departure time instants of a packet $pkt_{i,j}$ are given by $A_{i,j}=A_i+j \cdot p_i$ and $D_{i,j}=A_{i,j}+l_{5G_i}$, respectively, where A_i is the burst arrival time indicated in the TSCAI and l_{5G_i} is the latency requirement for packets in the TSN flow F_i indicated by the PDB of the 5QI. A TSN flow F_i is then defined as $F_i=\{A_i, l_{5G_i}, d_i, p_i, f_i\}$. The proposed scheduling scheme exploits this information to decide the radio resources that must be allocated to each TSN flow.

The TSN flows supported by the 5G network transmit periodic packets. Although the periodicity p_i of each TSN flow can be different, it is possible to find a pattern of packet arrival considering all TSN flows that repeat periodically. The periodicity of this packet arrival pattern is referred to as hyperperiod or HP . Fig. 3 shows an example with three TSN flows with different periodicities (30, 45 and 90 ms). It is

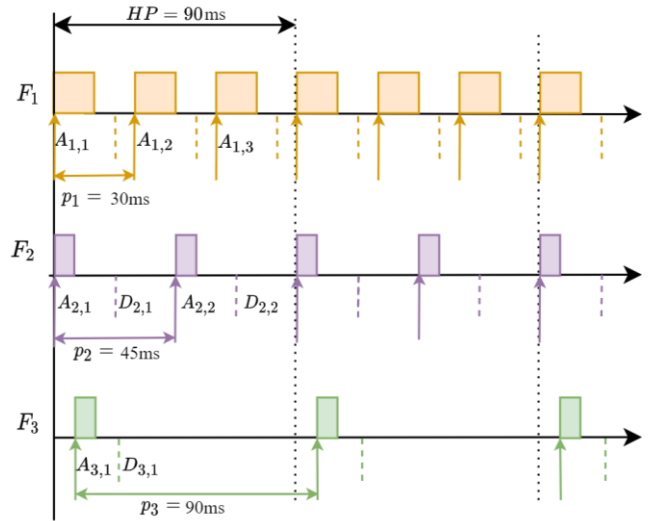


Fig. 3. Hyperperiod for an example case with 3 TSN flows with different periodicities.

¹ Survival time is the time period an application can survive without receiving any burst or packet as described in [24].

possible to see that there is a packet arrival pattern that repeats every 90 ms when considering the set of all TSN flows. The scheduling scheme identifies HP and allocates radio resources for each packet within HP . The allocated radio resources for each TSN flow are pre-assigned periodically with a periodicity HP .

The operation of the proposed scheduling scheme is presented in Algorithm I and Algorithm II. The scheduling scheme first calculates HP . HP is calculated as the least common multiple (lcm) of the periodicity of the packets transmitted in each TSN flow, i.e., $HP = \text{lcm}(p_i)$, with $i \in [1, N_F]$ (line 2 in Algorithm I). The packets of all the TSN flows within the time window given by HP are divided into different groups or blocks. Each block contains the packets whose transmission could overlap in time considering their arrival and departure times. Two packets $pkt_{i,j}$ and $pkt_{m,n}$ overlap if $A_{i,j} \leq A_{m,n} \leq D_{i,j}$ or $A_{i,j} \leq D_{m,n} \leq D_{i,j}$ (see Algorithm II). In the example of Fig. 3, two blocks are created: the first block B_1 contains $pkt_{1,1}$, $pkt_{2,1}$ and $pkt_{3,1}$, while the second block B_2 contains $pkt_{1,2}$, $pkt_{2,2}$ and $pkt_{3,2}$. The proposed scheme addresses the radio resource allocation process separately for each block B_z , with $z=1,2,\dots$. First, the scheduling scheme calculates the number of symbols d_i^S and RBs d_i^R that should be allocated for the transmission of each packet to satisfy d_i . To this end, the scheduling scheme is based on the Sym-OFDMA (Orthogonal Frequency Division Multiple Access) scheduling policy presented in [13]. Sym-OFDMA has been selected because it reduces the transmission latency compared to other scheduling policies. Following Sym-OFDMA, the scheduling scheme tries to minimize the number of symbols assigned for the transmission of each packet. In this context, the scheduling scheme will allocate $d_i^R = d_i$ RBs in one symbol ($d_i^S=1$) when d_i is lower than the number of RBs available in the bandwidth (R_{BW}), i.e. $d_i \leq R_{BW}$. If $d_i > R_{BW}$, the scheduling scheme will allocate $d_i^R = R_{BW}$ RBs in $d_i^S = \lceil d_i / R_{BW} \rceil$ symbols. When d_i^S and d_i^R are known, the scheduling scheme starts an iterative process. At each iteration, the scheduling scheme tries to find a radio resource allocation solution that satisfies the latency and departure times for all packets. The scheduling scheme starts a new iteration if it cannot find adequate radio resources to satisfy the departure time for a packet. The iterative process will finalize when one of the two following

ALGORITHM I: *SCHEDULING*

1. Input: $F_i \forall i \in [1, N_F], t_{sym}$
2. $HP = \text{lcm}\{p_i\}$ with $i \in [1, N_F]$
3. Create packet blocks B_z , with $z=1, 2, \dots$ (Algorithm II)
4. Initialize $BF_z \neq \emptyset$
5. **For** $z=1$ to N_B
6. Calculates d_i^S and d_i^R for all packets in B_z
7. N_{pkt} = number of packets in B_z
8. $pkt_{i,j}$ = packet with lower l_{5G} in B_z
9. $B_z = B_z - \{pkt_{i,j}\}$
10. $iter=0, BF_z = \emptyset$
11. **For** $iter = 0$ to ∞
12. **While** $B_z \neq \emptyset$
13. s = first OFDM symbol after $A_{i,j} + t_{UE,tx}$
14. **While** $s + d_i^S - 1 \leq D_{i,j}$
15. **If** there are d_i^R unallocated RBs in symbols s until $s + d_i^S - 1$
16. $pkt_{i,j}$ receives d_i^R RBs in symbols s until $s + d_i^S - 1$
17. First allocated symbol: $s_{i,j} = s$
18. **If** $s + d_i^S - 1 == D_{i,j}$
19. $BF_z = BF_z + \{pkt_{i,j}\}$
20. **EndIf**
21. **Else**
22. $s++$
23. **EndIf**
24. **EndWhile**
25. **If** $s + d_i^S - 1 > D_{i,j}$
26. $BF_z = BF_z + \{pkt_{i,j}\}$
27. **If** $BF_z + B_z == N_{pkt}$
28. Goto 41
29. **Else**
30. Reinitiate B_z
31. $B_z = B_z - BF_z$
32. Free radio resources allocated to packets in B_z
33. Goto line 11
34. **EndIf**
35. **Else**
36. $pkt_{i,j}$ = packet with lower l_{5G} in B_z
37. $B_z = B_z - \{pkt_{i,j}\}$
38. **EndIf**
39. **EndWhile**
40. **EndFor**
41. **End For**

$N_F]$ (line 2 in Algorithm I). The packets of all the TSN flows within the time window given by HP are divided into different groups or blocks. Each block contains the packets whose transmission could overlap in time considering their arrival and departure times. Two packets $pkt_{i,j}$ and $pkt_{m,n}$ overlap if $A_{i,j} \leq A_{m,n} \leq D_{i,j}$ or $A_{i,j} \leq D_{m,n} \leq D_{i,j}$ (see Algorithm II). In the example of Fig. 3, two blocks are created: the first block B_1 contains $pkt_{1,1}$, $pkt_{2,1}$ and $pkt_{3,1}$, while the second block B_2 contains $pkt_{1,2}$, $pkt_{2,2}$ and $pkt_{3,2}$. The proposed scheme addresses the radio resource allocation process separately for each block B_z , with $z=1,2,\dots$. First, the scheduling scheme calculates the number of symbols d_i^S and RBs d_i^R that should be allocated for the transmission of each packet to satisfy d_i . To this end, the scheduling scheme is based on the Sym-OFDMA (Orthogonal Frequency Division Multiple Access) scheduling policy presented in [13]. Sym-OFDMA has been selected because it reduces the transmission latency compared to other scheduling policies. Following Sym-OFDMA, the scheduling scheme tries to minimize the number of symbols assigned for the transmission of each packet. In this context, the scheduling scheme will allocate $d_i^R = d_i$ RBs in one symbol ($d_i^S=1$) when d_i is lower than the number of RBs available in the bandwidth (R_{BW}), i.e. $d_i \leq R_{BW}$. If $d_i > R_{BW}$, the scheduling scheme will allocate $d_i^R = R_{BW}$ RBs in $d_i^S = \lceil d_i / R_{BW} \rceil$ symbols. When d_i^S and d_i^R are known, the scheduling scheme starts an iterative process. At each iteration, the scheduling scheme tries to find a radio resource allocation solution that satisfies the latency and departure times for all packets. The scheduling scheme starts a new iteration if it cannot find adequate radio resources to satisfy the departure time for a packet. The iterative process will finalize when one of the two following

ALGORITHM II: *BLOCK CREATION*

1. Input: $F_i \forall i \in [1, N_F]$
2. $I = \{A_{i,j} \forall i \in [1, N_F], \forall j \in [0, HP/p_i]\}$
3. $z=1, B_z = \emptyset$
4. **While 1**
5. $\{i, j\} = \arg \left\{ \min_{m,n} \{I\} \right\} \forall m \in [1, N_F], \forall n \in [1, HP/p_m]$
6. $I = I - \{A_{i,j}\}$
7. **If** $B_z = \emptyset$
8. $D_{aux} = D_{i,j}, A_{aux} = A_{i,j}$
9. $B_z = \{pkt_{i,j}\}$
10. **Else**
11. **If** $A_{i,j} \leq D_{aux}$
12. $B_z = B_z \cup \{pkt_{i,j}\}$
13. **If** $D_{i,j} > D_{aux}$
14. $D_{aux} = D_{i,j}$
15. **End If**
16. **Else**
17. $z=z+1, B_z = \emptyset$
18. Go to 7
19. **End If**
20. **End If**
21. **If** $I = \emptyset$
22. Go to 24
23. **End If**
24. **End While**

conditions are met: all packets in B_z have allocated resources, or it is not possible to find a feasible solution to the problem. At each iteration, the proposed scheduling scheme serves packets considering their latency requirements l_{5G_i} : packets with lower l_{5G_i} are served first (line 8 in Algorithm I). The scheduling scheme allocates to each packet $pkt_{i,j}$ d_i^R RBs in the first d_i^S consecutive symbols with at least d_i^R unallocated RBs after the packet is generated ($A_{i,j} + t_{UE,tx}$, where $t_{UE,tx}$ represents the processing time in the transmitter to generate the packet) and before the departure time $D_{i,j}$ (lines 15-21 in Algorithm I). If a packet receives radio resources just before its departure time $D_{i,j}$, the packet will maintain the allocated radio resources although a new iteration is performed. To this end, it is included in a set BF_z . BF_z includes the packets that cannot change their allocated radio resources in potential next iterations (see lines 18-20 in Algorithm I). If it is not possible to find the required RBs and OFDM symbols for a packet, the scheduling scheme will start a new iteration. Before that, B_z is reinitialized but excluding packets in BF_z (lines 25-35 in Algorithm I). The first packet served in the new iteration will be the packet that was not possible to assign radio resources in the previous iteration. This packet is also included in BF_z (line 26 in Algorithm I). The scheduling process finalizes when all packets have allocated resources or when it is not possible to find a feasible solution to the problem (lines 27-29).

IV. EVALUATION SCENARIO

We consider an evaluation scenario where a 5G network is integrated into a TSN network as a logical bridge. The integrated 5G-TSN network provides connectivity to an industrial plant where there is implemented a closed-loop supervisory controller application [14]. This application integrates a Programmable Logic Controller (PLC) which receives monitoring data from sensors (S_1, S_2, \dots, S_{N_F}) that creates a total of N_F TSN flows. Based on the received data, the PLC sends a command to an actuator (A). Fig. 4 shows the evaluated scenario. The PLC, sensors, and actuator are the static end devices that exchange data in the network. The end devices are connected through several bridges, including the 5G logical bridge. The evaluation of the proposed scheduler has been carried out using Matlab.

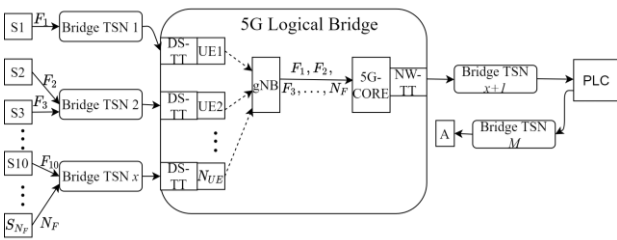


Fig. 4 . Evaluated scenario.

The sensors send the sensed data to the PLC. This results in N_F TSN flows between the sensors and the PLC that are transmitted through the 5G logical bridge (we consider N_F equal to 30). Each TSN flow F_i is defined by the size of the transmitted packets (s_i), the periodicity of the packets (p_i) and

the demanded end-to-end latency (l_{e2e_i}) between the sensor and the PLC. Based on [14], l_{e2e_i} can take values between 4ms and 20ms. We consider that the periodicity p_i is equal to the l_{e2e_i} requirement (this means that each packet needs to be received before the next packet is generated). TSN flows randomly select the periodicity p_i between p_1, p_2 or p_3 , and p_j , where $j \in \{1,2,3\}$, is randomly selected between 4 and 20 ms. The packet size s_i for each TSN flow F_i can take a value between 40 and 250 bytes. The characteristics and requirements that need to be satisfied by the 5G logical bridge of a TSN flow is indicated to the 5G network using the TSCAI and 5QI information. For each TSN flow, the TSN network informs the 5G network about the following information: the periodicity p_i and packet size s_i of the packets, the flow direction f_i , the arrival time A_i at the ingress port of the 5G logical bridge, and the maximum latency requirement of the TSN flow inside the 5G logical bridge l_{5G_i} . In the evaluation scenario, all TSN flows in the 5G network are transmitted in uplink, i.e., from the UEs to the gNB. The arrival time A_i and the 5G latency requirement are calculated by the CNC that has information about the TSN network topology and is in charge of the scheduling of the TSN flows in the TSN network to satisfy the end-to-end latency requirements. The arrival time A_i of the first packet of a TSN flow is equal to the latency that the packet experiences from the moment it is generated in the sensor until it is received in the 5G logical bridge. Then, it depends on the application processing time in the sensor l_{sensor_i} , the number and the processing time of TSN bridges that there are in the path between the sensor that generates the data and the 5G logical bridge $l_{TSN_{b,i}}$ with $b \in [a, \dots, x]$, the propagation time it takes a packet to travel through the links that interconnect the nodes (end devices and bridges) l_{links_i} , and the processing time t_{DS-TT} at the DS-TT translator. The A_i is given by the following expression:

$$A_i = l_{sensor_i} + l_{links_i} + l_{TSN_{a,i}} + \dots + l_{TSN_{x,i}} + t_{DS-TT} \quad (1)$$

$l_{TSN_{b,i}}$ represents the time that packets of a TSN flow i spend in a TSN bridge b . $l_{TSN_{b,i}}$ is null if TSN bridge b is not in the path between the sensor that generates the data and the 5G logical bridge. Otherwise, $l_{TSN_{b,i}}$ considers processing and queuing delays in the TSN bridge (independent delay $t_{indep_{b,i}}$) that vary depending on the traffic load [8], and the transmission delay that depends on the packet size and the bit rate used by the TSN bridge to transmit the data to the next node in the path (dependent delay $t_{dep_{b,i}}$). To calculate the arrival time, we consider that all sensors (30 sensors) connect to a TSN bridge that connects to the 5G logical bridge, and all sensors transmit packets of 250 bytes. Due to short distances, l_{links_i} is considered negligible. We consider DS-TT and NW-TT processing times equal to 0.05 ms and 0, respectively [15]. We assume a data bit rate equal to 100 Mbps, and the minimum independent delay $t_{indep_{b,i}}$ is equal to the TSN bridge processing delay, which is 1.5 μ s as presented in [16]. Then, for the minimum arrival time, we assume that a flow is transmitted through the TSN bridge

directly to the 5G logical bridge as soon as it arrives at the ingress port of the TSN bridge. We have considered that the sensor processing time is $3 \mu\text{s}$ [17], then, the minimum arrival time to 5G logical bridge for a packet of 250 bytes is $3\mu\text{s} + \left(\frac{250 \cdot 8\text{bits}}{100\text{Mbps}} + 1.5\mu\text{s}\right) + 50\mu\text{s} \cong 75\mu\text{s}$. In the case of $700\mu\text{s}$, which represents the maximum arrival time, we consider the last flow to be transmitted from TSN bridge to the 5G logical bridge arrives at $3\mu\text{s} + 30 \cdot \left(\frac{250 \cdot 8\text{bits}}{100\text{Mbps}} + 1.5\mu\text{s}\right) + 50\mu\text{s} \cong 700\mu\text{s}$ to the ingress port of 5G logical bridge. This results in arrival times between $75\mu\text{s}$ and $700 \mu\text{s}$.

We calculate the latency that must be satisfied in the 5G network for each TSN flow F_i as:

$$l_{5G_i} \leq l_{e2e_i} - A_i - l_{links_i} - \sum_{b=x+1}^M l_{TSN_{b,i}} - t_{NW-TT} - l_{PLC_i} \quad (2)$$

In (2), l_{PLC_i} represents the application processing time in the PLC, t_{NW-TT} is the processing time at the NW-TT translator, $l_{TSN_{b,i}}$ represents the time that packets of a TSN flow i spend in a TSN bridge b with $b \in [x+1, \dots, M]$, M is the number of TSN bridges in the TSN flow and l_{links_i} is the propagation time it takes a packet to travel through the links that interconnect the nodes from the 5G logical egress port until the PLC ingress port. To calculate l_{5G_i} , we assume l_{PLC_i} equal to $1007 \mu\text{s}$ as we consider a scan time of analog input data and the execution time of Proportional Integral Derivative (PID) control in the PLC of 1ms [18] and $7\mu\text{s}$ [19], respectively. We have assumed that the scan and processing times are minimal. This data is an approximation since it will vary with the application.

The 5G network deployed in the industrial plant implements a single cell with a bandwidth of 20 MHz. We consider the use of a 30 kHz sub-carrier spacing (SCS) as recommended in [20] for industrial environments. The cell operates in TDD mode [21], and each slot reserves 2 OFDM symbols for the transmission of UL and DL control messages and 12 OFDM symbols for the transmission of UL data. The UEs are placed in locations that guarantee Line of Sight conditions with the gNB. Based on that, we consider that all packets are transmitted using the MCS 12 in MCS table 1 [22], which guarantees a good compromise between robustness and spectral efficiency (modulation order Q_m equal to 4 and coding rate R equal to 434/1024). The number of radio resources to transmit a packet of size s_i can be then calculated as:

$$d_i = \left\lceil \frac{(tbs_i(s_i + s_{header}) + s_{CRC}) \cdot \delta}{R \cdot Q_m \cdot N_{sc, RB}} \right\rceil \quad (3)$$

In (3), s_{header} represents the size in bytes of the IPv4 header added to the data packet, and s_{CRC} is the size in bytes of the cyclic redundancy check (CRC) code added at the end of the packet. $tbs(s_i + s_{header})$ represents the smallest transport block size from the available values given in [23] that can be used

to transmit $s_i + s_{header}$ bytes, and $N_{sc, RB}$ is the number of subcarriers in a resource block ($N_{sc, RB} = 12$).

V. PERFORMANCE EVALUATION

This section evaluates the performance of the proposed scheduling scheme for a 5G network integrated into a TSN network to provide connectivity in industrial scenarios. To this end, we compare its performance to that achieved with a commonly used CG scheduling scheme [13]. The reference scheme allocates radio resources periodically for each TSN flow F_i with a periodicity p_i . Each TSN flow F_i receives the number of radio resources necessary to transmit packets of size s_i according to (3). For a fair comparison with the proposed scheduling scheme, the reference scheme serves TSN flows based on their 5G latency requirements: TSN flows with lower l_{5G_i} values are served first. Since the TSN flows can have different periodicities, several TSN flows could receive the same radio resources after some periods.

Due to the limited availability of radio resources and the high resource demands, it can be difficult for scheduling schemes to find a valid solution that can satisfy all requests. In this context, we first analyze the capabilities of the proposed and reference scheduling schemes to achieve a solution that allocates the requested radio resources for all the TSN flows. Fig. 5 shows the percentage of evaluated scheduling problems for which the proposed and reference schemes satisfy the radio resource request for all the TSN flows. The results in Fig. 5 are shown for different numbers of TSN flows demanding resources in the 5G network (between 10 and 30). Fig. 5 (lines with circles) considers that the end-to-end latency (l_{e2e_i}) demanded by each TSN flow F_i is equal to its packet periodicity p_i , and Fig. 5 (lines with stars) considers that all TSN flows demand a l_{e2e_i} equal to 4 ms. The results in Fig. 5 show that when the latency requirements are more relaxed, both the proposal and the reference scheme obtained solutions that satisfy all the requests. When the latency requirements are more stringent ($l_{e2e_i} = 4 \text{ ms}$ for all the TSN flows) and the number of TSN flows requesting resources increases to 25, the reference scheme fails to find solutions satisfying all the requests in 9%

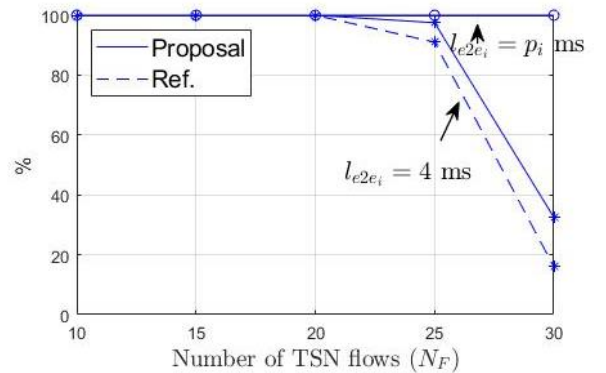


Fig. 5. Percentage of times the proposed and reference scheduling schemes satisfy the radio resource request for all packets of all TSN flows as a function of the number of TSN flows.

of the cases. In this scenario, the proposal achieves solutions satisfying all the requests in the 98% of the cases. When the number of TSN flows increases to 30, the proposal is able to satisfy all the requests in the 33% of the cases, while the reference scheme only in the 17%.

The solutions achieved with the reference scheme ensure that the radio resources allocated to the TSN flows do not overlap in time and frequency simultaneously for the first packet in each TSN flow. However, this is not guaranteed for the following packets in the flows, and several TSN flows can receive the same radio resources after several periods due to the different periodicities of each TSN flow (Fig. 2). This is a radio resource allocation conflict that can ultimately result in packet collisions. Fig. 6 shows the percentage of times that the reference scheme achieves a solution serving all requests, but some radio resources are allocated to more than one TSN flows. The results in Fig. 6 show that the number of times that the solution achieved with the reference scheme results in a radio resource allocation conflict increases with the number of TSN flows demanding resources. It is important to note that the percentage of times that there is a resource allocation conflict with the reference scheme is higher than 74% when the number of TSN flows demanding resources is equal to or higher than 15.

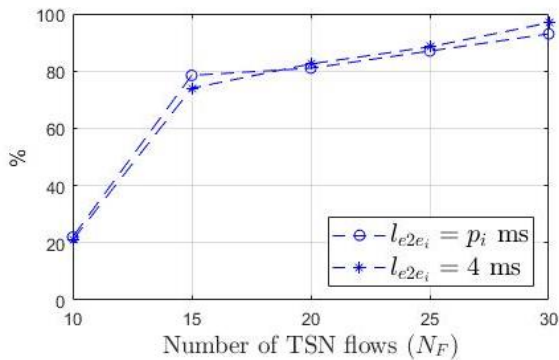


Fig. 6. Percentage of times that the reference scheme achieves a solution serving all requests, but some radio resources are allocated to more than one TSN flows as a function of the number of TSN flows.

Fig. 7 depicts the average of the latency and the average of the standard deviation experienced by the packet of each TSN flow when a different number of TSN flows are transmitted in the 5G network, considering that the end-to-end latency (l_{e2e_i}) demanded by each TSN flow F_i is equal to its packet periodicity p_i . The results show that the proposed scheduling scheme reduces the average latency experienced by the transmitted packets by more than 40% when the number of TSN flows is equal to or higher than 25. This is a result of the higher flexibility of the proposed scheduling scheme. With the reference scheme, a TSN flow receives resources periodically with the same periodicity that packets arrive at the 5G network. Therefore, all packets in a TSN flow experienced the same latency. This is shown in Fig. 7.b,

which shows the average standard deviation experienced by the different TSN flows with the proposed and reference scheme. Fig. 7.b shows that the average standard deviation experienced with the reference scheme is zero. On the other hand, the proposed scheduling scheme allocates radio resources individually for each packet of a TSN flow within a time window of size HP (the resources are allocated periodically with period HP). Thanks to this flexibility, the proposal assigns for each packet of a TSN flow the radio resources that minimize the latency based on the availability of unassigned resources at each moment. This results in a reduction of the latency experienced in the 5G network at the expense of an increase in the standard deviation, as shown in Fig. 7.b. The standard deviation will be minimized by the TSN translators that will hold and forward the packets at the departure time indicated by the TSN network. It is essential to highlight that the important thing is that packets arrive at the egress port of the 5G logical bridge (the TSN translator) at the time required to be transmitted to the next node in the path (a TSN bridge or the end device).

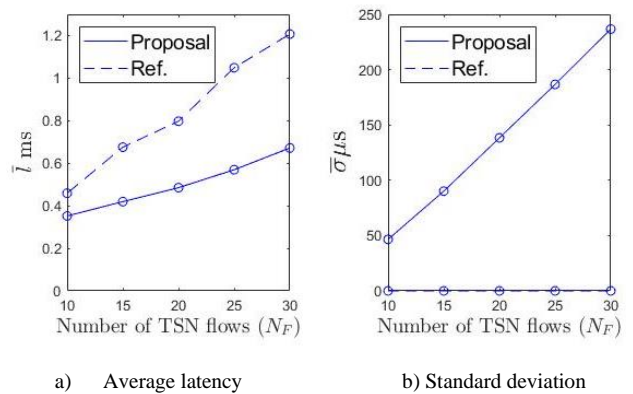


Fig. 7. Average latency and standard deviation of the proposed and reference scheduling schemes as a function of the number of TSN flows when l_{e2e_i} is equal to the periodicity.

VI. CONCLUSIONS

This paper has presented and evaluated a novel Configured Grant scheduling scheme for 5G networks integrated with TSN networks. This paper considers the 5G-TSN integration model, where 5G is integrated into the TSN network as a bridge. The proposed scheduling scheme aims to coordinate its scheduling decision with the scheduling performed in TSN to efficiently and effectively meet the end-to-end latency and determinism requirements of TSN traffic. To this end, the proposed scheduling scheme uses the information provided by TSN about the characteristics of the TSN traffic (periodicity, packet size, etc.) to satisfy the maximum latency that, based on the TSN scheduling decision, must be experienced in the 5G network. Common CG scheduling schemes allocate radio resources periodically for each TSN flow based on the periodicity of the packets. This may result in the use of the same radio resource for the

transmission of some packets by different TSN flows due to the different periodicity of the packets of different TSN flows. The proposed scheduling scheme solves this issue by configuring several grants for each TSN flow in order to avoid radio resource allocation conflicts among different TSN flows. The results achieved in this paper have demonstrated that the proposed scheduling scheme considerably increases the number of TSN flows that can be satisfactorily served in the integrated 5G-TSN network compared to common CG schemes. Furthermore, the average latency experienced in the 5G network is reduced with the proposed scheme.

REFERENCES

- [1] 5G-ACIA, "Key 5G Use Cases and Requirements," 2020.
- [2] D. Ginhör *et al.*, "5G RAN slicing for deterministic traffic", in *IEEE Wireless Communications and Networking Conference (WCNC)*, China, 2021.
- [3] R. B. Abreu *et al.*, "Scheduling Enhancements and Performance Evaluation of Downlink 5G Time-Sensitive Communications," *IEEE Access*, 2020.
- [4] Y. Zhang *et al.*, "Wireless/wired integrated transmission for industrial cyber-physical systems: risk-sensitive co-design of 5G and TSN protocols," *Sci. China Inf. Sci.*
- [5] Nokia, "5G plug-and-produce. How the 3GPP 5G System facilitates Industrial Ethernet deployments (White Paper)." 2020.
- [6] S. S. Craciunas *et al.*, "Scheduling real-time communication in IEEE 802.1Qbv time sensitive networks," *ACM Int. Conf. Proceeding Ser.*, 2020.
- [7] IEEE, *IEEE Std 802.1Qcc: Standard for Local and Metropolitan Area Networks--Bridges and Bridged Networks*, vol. 2018. 2018.
- [8] 3GPP, TSG RAN; System Architecture for the 5G System; Rel. 17, TS 23.501, V17.3.0, 2021.
- [9] IEEE, *IEEE 802.1Q: Standard for Local and metropolitan area networks -- Bridges and Bridged Networks* -. 2018.
- [10] Z. Satka *et al.*, "QoS-MAN: A Novel QoS Mapping Algorithm for TSN-5G Flows", the 28th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'22), Taiwan, 2022.
- [11] Y. Zhang *et al.*, "QoS-Aware Mapping and Scheduling for Virtual Network Functions in Industrial 5G-TSN Network," *2021 IEEE Glob. Commun. Conf. GLOBECOM*, 2021.
- [12] M. J. Yang *et al.*, "An Uplink Transmission Scheme for TSN Service in 5G Industrial IoT," *Int. Conf. ICT Converg.*, 2020.
- [13] A. Larrañaga Zumeta, M. C. Lucas-Estañ, S. Lagén, Z. Ali, I. Martinez, and J. Gozalvez, "An Open-Source Implementation and Validation Of 5g Nr Configured Grant for Urllc in Ns-3 5g Lena: A Scheduling Case Study in Industry 4.0 Scenarios," *SSRN Electron. J.*, Aug. 2022.
- [14] K. Montgomery *et al.*, "Wireless User Requirements for the Factory Workcell", NIST Advanced Manufacturing Series 300-8, Nov. 2020.
- [15] Z. Satka *et al.*, "Developing a Translation Technique for Converged TSN-5G Communication," *IEEE Int. Work. Fact. Commun. Syst. - Proceedings, WFCS*, 2022.
- [16] D. Pannell, "Choosing the Right TSN Tools to meet a Bounded Latency The Need high speeds saves wires in Zonal networks ➤ And Zonal networks bring new requirements that (TSN) solves," in *IEEE SA Ethernet & IP @ Automotive Technology Day*, Detroit, September, 2019.
- [17] "MA734 | 8-Bit to 12.5-Bit, 3µs Low-Latency Contactless Angle Sensor | MPS." [Online]. Available: <https://www.monolithicpower.com/en/ma734.html>. [Accessed: 13-Nov-2022].
- [18] R. Automation, "1756 ControlLogix I/O Modules Specifications Technical Data."
- [19] R. Automation, "Estimated Logix 5000 Controller Instruction Execution Times Reference Manual."
- [20] 3GPP, TSG RAN; Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC); Rel. 16, TR 38.824. V16.0.0, 2019.
- [21] 3GPP, TSG RAN; Base Station (BS) radio transmission and reception; Rel. 17, TS 38.104, 2021.
- [22] 3GPP, TSG RAN; Physical layer procedures for data; Rel. 16, TS 38.214, V16.1.0, 2020.
- [23] 3GPP, TSG RAN; Medium Access Control (MAC) protocol specification; Rel. 16, TS 38.321, V16.6.0, 2021.
- [24] 3GPP, TSG Services and System Aspects; Service requirements for the 5G system; Rel. 18, TS 22.261, 2022.



The 5th International Conference on Emerging Data and Industry 4.0 (EDI40)
March 22-25, 2022, Porto, Portugal

5G NR Configured Grant in ns-3 Network Simulator for Ultra-Reliable Low Latency Communications

Ana Larrañaga^{a,*}, M. Carmen Lucas-Estañ^b, Imanol Martinez^a, Javier Gozalvez^b

^a*HW and Communication Systems Area, Ikerlan Technology Research Centre, Mondragón 20500, Spain*

^b*UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain*

Abstract

5th Generation (5G) and Beyond networks are being designed to support Ultra-Reliable and Low Latency Communications (URLLC). To this end, 5G defines a new radio (NR) interface with a new mechanism at the Physical (PHY) and Medium Access Control (MAC) layers that allow reducing the latency communication. One key mechanism to reduce the latency is the scheduling scheme. Mainly, 5G defines the use of the configured grant (CG) scheduling for uplink (UL) transmissions that eliminates the need to request and assign resources for each packet transmission by pre-allocating resources to the UE. The availability of simulation tools that accurately model the new mechanisms and technologies incorporated in 5G New Radio (NR) is key to research and evaluate new proposals and enhancements to meet the communication requirements of emerging services. In this context, this paper presents the implementation of the configured grant scheduling in the ns-3 network simulator. Remarkably, the configured grant has been implemented within the 5G-LTE-EPC Network simulator (5G-LENA) module that simulates the fundamental PHY-MAC NR features in line with the NR specifications. In addition, this paper validates the configured grant implementation through system-level simulations considering a typical Industry 4.0 scenario characterized by applications demanding URLLC.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Configured Grant, 5G, NR, semi-static scheduling, URLLC, ns-3, 5G-LENA, software simulation

* Corresponding author. Tel.: +34-607-581-466; fax: +34-943-796-944.

E-mail address: ana.larranaga@ikerlan.es

1. Introduction

5G and Beyond networks are being designed to efficiently support very different applications and services with diverse and stringent communication requirements not only in terms of bandwidth and transmission rates but also in terms of reliability and latency. The support of URLLC is key in verticals with increasing levels of automation, such as manufacturing or transportation. 5G incorporates an NR interface that defines new mechanisms and technologies to reduce the contribution of the PHY and MAC mechanisms to the communication latency [1]. At the PHY layer, 5G NR defines a more flexible frame structure, the use of flexible numerology that allows using transmission slots with shorter time-domain (between 1 and 0.0625 ms), and the possibility to use minislots [2]. At the MAC layer, one of the mechanisms that significantly contributes to the communication latency is the scheduling scheme. To reduce this delay, 5G NR introduces semi-static scheduling schemes [3] (referred to as grant-free scheduling), in addition to the dynamic scheduling, that pre-allocate resources to UEs and eliminate the need to request and allocate resources before transmitting each packet.

The availability of simulation tools that accurately model the new mechanisms and technologies incorporated in 5G NR is key to researching and evaluating new proposals and enhancements to meet the communication requirements of emerging services. In this context, it is possible to find in the literature different simulation tools developed to this end. For example, [4] and [5] have developed the necessary software tools to evaluate the basic connectivity of devices to a 5G network. However, these tools are not open to the research community. Open simulation tools are very interesting options since it aggregates the efforts of the research community to develop powerful and accurate simulation environments. OMNeT++ and ns-3 discrete event network simulators are two powerful open-source network simulators that implement modules for the simulation of the 5G NR. In the case of OMNeT++, Simu5G [6] models the different features of 5G NR such as carrier aggregation, different numerologies, and frequency/time division duplexing. On the other hand, ns-3 incorporates a pluggable 5G-LTE-EPC Network Simulator (5G-LENA) module [7] to simulate 5G networks. 5G-LENA implements the 5G NR features at the PHY and MAC layers, such as configurations of the NR frame structures through different numerologies, operations through multiple Bandwidth Parts (BWP), or dynamic scheduling for UL and DL transmissions. Both simulators are very powerful and offer significant advantages for studying and evaluating 5G networks. However, to the best of the authors' knowledge, neither of them implements configured grant, a key mechanism for studying URLLC communications.

We have considered using the 5G-LENA simulator as a tool to support and evaluate our research about the implementation of 5G and Beyond networks to support the communication requirements of critical Industry 4.0 applications. We selected 5G-LENA because it was the unique 5G NR simulator available when starting this study (Simu5G was released after 5G-LENA). To this end, it is necessary to implement the semi-static scheduling in the simulation tool essential to support URLLC applications. In this context, this paper presents the implementation of the semi-static scheduling defined in 5G NR [8] for uplink transmissions (this is referred to as configured grant). We focus on the UL since it is the link where the semi-static scheduling potentially provides higher latency reductions compared with the use of dynamic scheduling; in the uplink, the user equipment (UE) has to request resources to the new generation Node B (gNB) and wait for the allocated resources before transmitting a packet. In addition, we validate this implementation through system-level simulations and compare the performance results with analytical values available in the literature. In addition, we present initial evaluations to show the potential of this software tool to evaluate the potential of configured grant to support Industry 4.0 applications with stringent latency requirements.

The rest of the paper is structured as follows. Section 2 describes the different mechanisms included in 5G NR for radio resource allocation. Section 3 presents the 5G-LENA module. Section 4 presents the implementation of the configured grant scheduling in 5G-LENA module. Section 5 validates the implemented code. Section 6 outlines the main conclusions.

2. Radio resource allocation in 5G New Radio

5G-NR introduces dynamic and semi-static scheduling for supporting services with different communication requirements and traffic patterns [8]. With dynamic scheduling, radio resources are dynamically allocated by the gNB for each packet transmission in downlink (DL) and UL. For each packet to be transmitted in the DL, the gNB informs the corresponding UE about the allocated radio resources where the packet will be transmitted. The control message

with the scheduling information is transmitted. When a UE wants to transmit a packet in the UL using dynamic scheduling, the UE sends a scheduling request (SR) to the gNB. After the scheduling decision, the gNB sends a grant message to the UE that contains the information about the allocated resources. Then, the UE can transmit the packet using the allocated resources. If the UE was not able to transmit all the data in the allocated resources, the UE sends the Buffer Status Report (BSR) together with the data packet to inform the gNB about the amount of data pending to be transmitted. Based on this information, the gNB takes a scheduling decision and sends a grant with the information about the allocated resources to the UE. When the UE receives the grant, it can transmit a new data packet in the allocated resources. This process continues as long as the UE has data to transmit. Fig. 1.a illustrates the dynamic scheduling process for UL transmissions.

The signalling exchange between the UE and the gNB to request/inform about the allocated resources results in a non-negligible delay that can compromise the feasibility of services demanding low latencies, and this latency is more relevant in the UL where two messages need to be exchanged. For example, this signalling exchange results in latency values for UL transmissions between the UE and the gNB in 5G NR of 4.5 and 2.5 ms when using 15 kHz subcarrier spacing and a 14 and 7 symbols-slot respectively [9]. It is important to highlight that these values do not meet the 1 ms-latency requirement established by the 3rd Generation Partnership Project (3GPP) Release 16 in [10] as a general latency requirement for URLLC services¹. To reduce this latency, 5G NR introduces the use of semi-static scheduling, referred to as semipersistent scheduling (SPS), and configured grant (CG) for DL and UL transmissions, respectively [8]. With SPS and CG, radio resources are pre-assigned periodically to the UEs. In this context, when a packet is generated, it can be transmitted immediately in the pre-allocated resources. SPS and CG eliminate the need to exchange signaling messages to request/grant resources for each packet and therefore reduce the transmission latency. With SPS, the periodicity of the pre-allocated DL resources is configured by Radio Resource Control (RRC) signalling when the connection is established, and the allocated DL resources can either be signalled, activated, or deactivated by control messages [8]. In the UL, the 3GPP defines two types of CG: Type 1 and Type 2. With Type 1, the configured uplink grant, including the periodicity, is configured by RRC signaling at the connection establishment. CG Type 2 is similar to SPS. With CG Type 2, the periodicity of the configured UL grant is defined by RRC signaling at the session establishment, while the configured UL grant is either signalled, activated, or deactivated by control messages [8]. Fig. 1.b illustrates the CG scheduling (Type 1) for UL data transmissions.

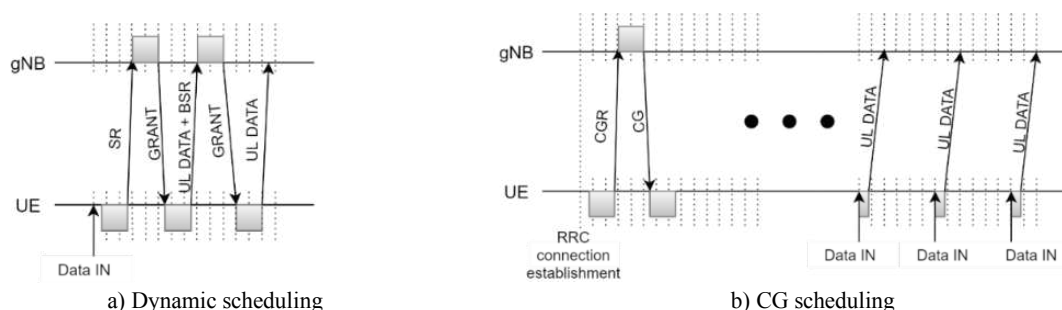


Fig. 1. Signal flow for UL transmissions (shadow rectangles represent the processing times at the gNB and the UE).

3. 5G-LENA Network Simulator

5G-LENA is an open-source discrete-event network simulator of the 5G NR ([11], [7]). It is implemented over ns-3 [12]. ns-3 is an open C++ simulation environment for networking research. The ns-3 simulator offers a solid simulation core that supports research on both IP and non-IP-based networks. LENA [7] was initially developed to implement the Radio Access Network (RAN) and the core network of Long Term Evolution (LTE). 5G-LENA is the evolution of LENA for the simulation of 5G NR. This evolution focused first on the RAN. The 5G-LENA module implements the fundamental PHY-MAC NR features in line with the NR specifications [8].

¹ 3GPP Release 16 establishes in [10] as a general requirement for URLLC services that a packet of 32 bytes must be transmitted with a reliability of $1-10^{-5}$ and a latency deadline of 1 ms.

The 5G NR RAN is modeled in the RAN C++ class. The RAN class implements two different C++ classes to simulate the functionalities of the gNB (NrGnb class) and the UE (NrUe class). In addition to the PHY and MAC layers, where the main NR features are implemented, the upper layers, Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), and RRC of the protocol stack are implemented for the gNB and the UE. The RLC, PDCP, and RRC layers currently rely on LTE LENA implementation. The C++ classes that implement the MAC and PHY layers for the gNB are the NrGnbMac and NrGnbPhy classes, and for the UE are the NrUeMac and NrUePhy classes. The NrGnbMac and NrUeMac classes contain the NrMacSchedulerNs3 class that implements the scheduler for the UL and DL transmissions. The NrMacSchedulerNs3 class uses the functions implemented in the NrMacSchedulerOfdma and NrMacSchedulerTdma classes to allocate resources based on Orthogonal Frequency Division Multiplexing Access (OFDMA) or Time Division Multiplexing Access (TDMA). The NrGnbPhy and NrUePhy classes implement the different numerologies defined in 5G NR (numerologies 0 to 4 with subcarrier spacing equal to 15, 30, 60, 120, and 240 kHz, respectively) [2], and properly model the numerology-dependent slot and OFDM symbol granularity (see [4]). Therefore, the PHY layer of the 5G LENA is able to support the flexible NR frame structure in the time domain and model the transmission and reception of data and control channels through the NrSpectrumPhy class that models the radio channel.

5G-LENA only implements dynamic scheduling for UL and DL transmissions [7]. When a new data packet generated in the UE arrives to the MAC layer from the upper layers, the UE needs to send an SR to the gNB to request radio resources for the transmission of the packet (as presented in Section 0). In this context, the UE MAC layer generates the SR. The UE has a variable denoted as *UE_state* that indicates at which state is the transmission of the SR. *UE_state* is initialized to INACTIVE when the UE is created in the simulation. When the MAC layer generates the SR, it changes the value of *UE_state* from INACTIVE to TO_SEND. At the beginning of each slot, the MAC layer checks if *UE_state* is equal to TO_SEND. In this case, the MAC layer sends the SR to the PHY layer, and *UE_state* is set to ACTIVE. The PHY layer then transmits the SR to the gNB in the radio resources reserved for the Physical UL Control Channel (PUCCH) through the radio channel.

The gNB checks at each slot if control messages from the UEs are received. If an SR is received from a UE, the PHY layer saves the SR information (the Radio Network Temporary Identifier (RNTI) and BWP id) and sends the SR to the gNB MAC layer. The MAC layer takes then a scheduling decision (it decides the radio resources that are assigned to the UE) and generates the grant message to inform the UE about the allocated resources. The grant message is sent to the gNB PHY layer. The PHY layer transmits the grant to the UE in the radio resources reserved for the transmission of the Physical DL Control Channel (PDCCH). When the UE receives the grant, it prepares the allocated resources for the transmission of the data packet. The UE MAC stores the data packet to be transmitted in the allocated resources through the ProcessUIDci function. If all the data is transmitted in the current packet, the *UE_state* is set to INACTIVE. If there is data pending to be transmitted in the UE buffer, the MAC layer sends, together with the data, the BSR to inform the gNB of the amount of data pending to be transmitted at the UE and request additional resources for a new transmission. The UE PHY layer then transmits the data packet to the gNB in the allocated resources (using InsertFutureAllocation function) on the Physical UL Shared Channel (PUSCH). When the gNB receives the data packet, it checks the BSR information and takes a new scheduling decision if there is data pending to be transmitted in the UE. When the scheduling decision is taken, the gNB will be sent a new grant message to the UE. This process will be repeated as long as there is data to transmit in the UE buffer.

4. Configured Grant Implementation

This section presents the implementation of the CG scheduling for UL transmissions carried out by the authors within 5G-LENA. In this work, we have implemented the CG Type 1 that configures the UL grant and the allocated resources when the session is established; we aim to avoid any potential transmission delay that could be introduced due to the activation/deactivation of the configured UL grant with the CG type 2 for services with very stringent latency requirements.

Fig. 2 shows a flow chart of the operation of the CG scheduling at the UE and gNB. When the simulation starts, the UEs are configured with traffic parameters such as packet size and transmission period (traffic parameters are read from the simulation configuration file). UEs that are configured to use CG scheduling have a variable referred to as *UE_CG_status* that identifies at which point of the CG configuration process is the UE. *UE_CG_status* can take the following values: INACTIVE_CG, TO_SEND_CGR, TO_RECEIVE_CG, ACTIVE_CG, and SCH_CG_DATA. *UE_CG_status* is initialized to INACTIVE_CG at the beginning of the simulation (this is shown in (Fig. 2.a)). During

the RRC connection establishment process, the UE sends a CG request (CGR) to the gNB. To this aim, the MAC layer generates the CGR and changes *UE_CG_status* from *INACTIVE_CG* to *TO_SEND_CG*. The MAC layer sends the CGR to the PHY layer of the UE. The PHY layer checks at the start of each slot if *UE_CG_status* is equal to *TO_SEND_CG*. If this is the case, it creates the control message where the CGR is transmitted and sends it to the

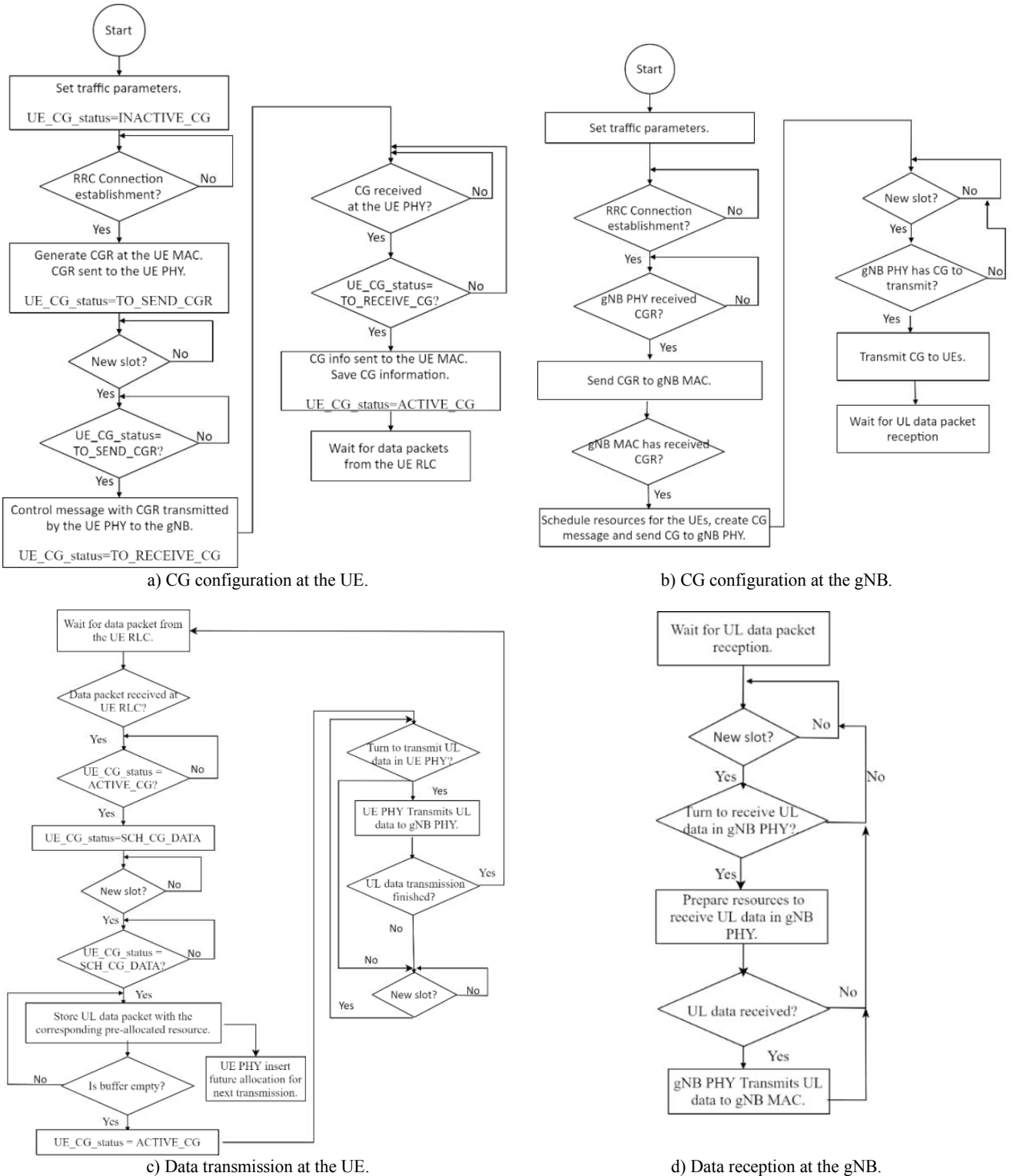


Fig. 2. Operation of the CG scheduling at the UE and gNB.

gNB through the radio channel. UE_CG_status is set to $TO_RECEIVE_CG$. When the gNB receives a CGR at the PHY layer (see Fig. 2.b), the PHY layer sends the CGR message to the MAC layer. The MAC layer then decides the radio resources to be allocated to the UE; the gNB reserves the radio resources needed to transmit the packets with the required periodicity. Once the scheduling decision is taken, the MAC layer prepares the configured grant information (allocated resources and periodicity) to be transmitted to the UE and sends it to the PHY layer. The gNB PHY layer creates the control message where the CG information is transmitted and sends it to the UE through the radio channel. When the UE receives the CG, it first checks whether UE_CG_status is equal to $TO_RECEIVE_CG$ (see Fig. 2.a). If this is the case, the PHY layer sends the CG information to the UE MAC layer, and the UE_CG_status state is changed to $ACTIVE_CG$. The CG information is saved at the PHY and MAC layer to be used to transmit future data packets (to this end, we use function `UIData` and `InsertFutureAllocation` in the PHY layer and `ProcessUIDci` in the MAC layer implemented in 5G LENA).

Once the UE receives the CG, it waits for the generation of new data. Data is generated periodically every T_p ms at the UE RLC layer. When data is received at the UE MAC layer from the RLC layer (see Fig. 2.c), the UE checks if UE_CG_status is equal to $ACTIVE_CG$. If this is the case, the MAC layer sets UE_CG_status to SCH_CG_DATA and stores the UL data packet with the corresponding pre-allocated resource information. Then UE MAC notifies the UE PHY layer that a new CG packet has been generated in the UE MAC. When the buffer is empty, UE_CG_status is changed to $ACTIVE_CG$. The UE PHY layer checks if the UE has pre-allocated resources in the current slot. If not, the UE PHY layer waits for the next slot. If the UE has pre-allocated resources in the current slot, the PHY layer transmits the UL data packet on the pre-allocated resources to the gNB through the radio channel. At the start of each period, the gNB prepares the PHY layer to receive a data packet on the pre-allocated resources (see Fig. 2.d). When the gNB PHY layer receives a data packet, it sends the packet to the MAC layer.

5. Validation

In this section, we validate the implementation of CG in 5G-LENA. To this end, we compare, when possible, the latency results achieved with the implemented CG in 5G-LENA with results available in the literature. In addition, we also compare the performance in terms of latency achieved using dynamic scheduling and CG. For this study, we consider a single-cell scenario with 10 MHz bandwidth at 3.55 GHz covering a typical work cell² of 10×10 m². In the work cell, sensor UEs transmit the sensed data to a central monitoring system (only UL data). All slots in the frame are configured to be used for UL transmissions. The DL CTRL and the UL CTRL messages are transmitted in the first and last symbols of each slot, respectively. We evaluate different numbers N of sensor UEs in the scenario, with N equal to 10, 20, and 30 [13]. We consider that sensors are located at positions that guarantee Line of Sight (LOS) propagation conditions with the gNB. All UEs transmit using a robust Modulation Coding Scheme (MCS) equal to 4 (with a modulation order of 2 and a coding rate of 0.19) to guarantee a low Block Error Rate (BLER). We evaluate different numerologies, particularly, we consider numerologies (μ) 1 and 2 corresponding to a subcarrier spacing (SCS) of 30 and 60 kHz and a slot time duration of 0.5 and 0.25 ms, respectively [2]. Following [13], the UEs transmit packets of 32 bytes with a transmission period T_p equal to 1, 5, and 10 ms. All UEs generate the packets at the same time. Packets are transmitted using 1 OFDM symbol minislot. UEs are scheduled at different symbols, and each UE receives resources in the first available symbol of the first available slot. We analyze the latency at the RLC layer that accounts for the time elapsed between the time when the UE generates the packet at the RLC layer and the time when this packet is received at the RLC layer of the gNB.

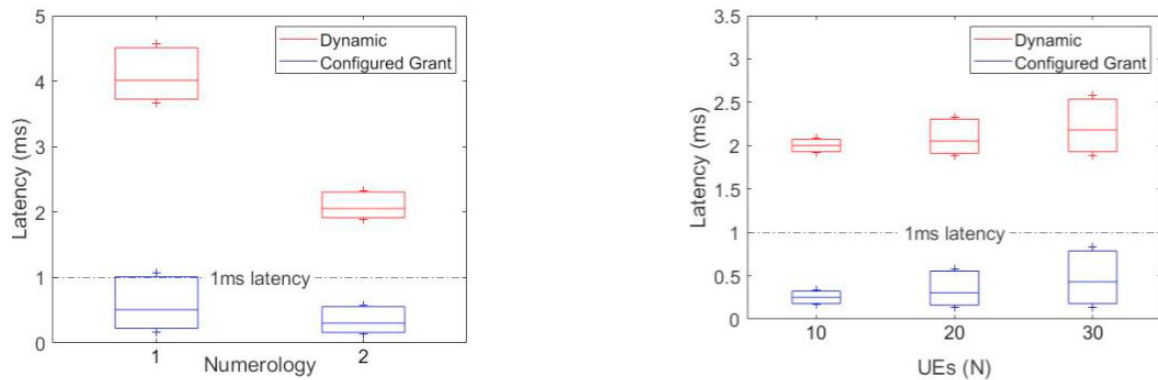
Fig. 3 shows the boxplot of the experienced latency. The top and bottom box edges represent the 10th and 90th percentile of the experienced latency, the center mark represents the median, and the plus symbol represents the maximum and minimum experienced latency values. Fig. 3.a shows the experienced latency as a function of the numerology when $N=20$ UEs and $T_p=10$ ms and CG and dynamic scheduling are used, respectively. Results in Fig. 3.a show that the latency experienced with CG decreases considerably the latency experienced when dynamic scheduling is used. For example, 75% of the transmissions experienced a latency lower than 0.9 and 0.5 ms with CG and μ equal to 1 and 2 respectively, while these values are equal to 4.4 and 2.25 ms with dynamic scheduling. This

² The typical workcell size is 10 m x 10 m [13].

results in a reduction of 79.54 and 77.77% of the 75th percentile of the experienced latency when CG is used with μ equal to 1 and 2 respectively in comparison with the use of dynamic scheduling. The higher latency experienced with dynamic scheduling is due to the exchange of control messages to request and assign radio resources previous to the transmission of each packet. Furthermore, it is important to highlight that the latency results achieved with CG are in line with the analytical results presented in [14]. 3GPP TR 37.910 [14] evaluates the latency achievable in the radio link between the UE and the gNB when there is only one UE in the cell. It reports latency values³ of 0.30 and 0.24 ms when using 2 OFDM symbols-minislot and μ equal to 1 and 2, respectively. The minimum latency achieved with the implemented CG in 5G-LENA is equal to 0.18 and 0.135 ms with μ equal to 1 and 2 respectively, and the latency experienced by the 25% of the transmissions is equal to 0.33 and 0.22 ms respectively (we consider the minimum and 25th percentile of the experienced latency achieved with the implemented CG for comparison with results in [14] since these values correspond to the UE to which resources are first allocated and the evaluation conditions for these UEs are similar to when unloaded scenarios are considered). These results validate our implementation of the CG within 5G-LENA.

Fig. 3.a also shows that the latency experienced with both the dynamic scheduling and CG reduces when numerology increases. It is possible to observe that the latency experienced with $\mu=2$ decreases by approximately 50% compared with the use of $\mu=1$. This is because the time duration of the symbol and the slot decreases by a factor of 2 when numerology increases from 1 to 2.

As presented in [13], closed-loop applications in Industry 4.0 demands a maximum latency of 1 ms. Fig. 3.a shows that the dynamic scheduling can not meet the 1 ms latency requirement in the evaluated conditions. The minimum latency experienced by a UE with dynamic scheduling is equal to 1.89 ms when $\mu=2$ is used. When CG is used, the 1 ms latency requirement can be satisfied using both μ equal to 1 and 2. All transmissions are performed in less than 0.58 ms when $\mu=2$, and 90% of the transmissions are performed in less than 1 ms when $\mu=1$.



a) Latency as a function of the numerology ($N=20$ UEs, $T_p=10$ ms). b) Latency as a function of the number of UEs (N) ($\mu=2$, $T_p=10$ ms).
Fig. 3. Experienced latency with dynamic scheduling and CG.

Fig. 3.b shows the latency experienced by the packets when the number of UEs increases. The results in Fig. 3.b correspond to $\mu=2$ since it is the one that allows achieving the lower latency values with both dynamic scheduling and CG. Fig. 3.b shows that the minimum latency value remains equal to 0.135 ms for all values of N . This is because the minimum value corresponds to the UE that receives resources at the first symbol of the first slot after the packets are generated, and this results in the same latency value for all values of N . However, the mean, 75th percentile, and maximum latency values increase with the increase of UEs. This is because the transmission of the UEs that receive resources in the last position is delayed more, and this increases the latency that these UEs experience.

We have also evaluated the experienced latency for different values of T_p equal to 1, 5, and 10 ms when $N=20$ UEs and μ is equal to 1 and 2. The results show that the same latency values are achieved for given numerology independently of the T_p when CG is used. This is because all packets can be transmitted before the next packet is generated for the UEs. With dynamic scheduling, the latency experienced when $\mu=2$ is equal for all values of T_p .

³ These values correspond to the use of resource mapping Type B and UE capability 2.

However, it is possible to see that the latency considerably increases when $\mu=1$ and $T_p=1$ ms. This is because a new packet is generated before the previous packet has been transmitted for some UEs. In this case, the new packet is queued at the transmitter, and latency considerably increases.

6. Conclusions

This paper has presented the implementation of the configured grant scheduling included in 3GPP Release 16 within the 5G-LENA 5G NR simulator. Configured grant is one of the key mechanisms to reduce communication latency and support URLLC services with stringent latency requirements from critical verticals such as manufacturing. The main features of the implemented code have been briefly described in this paper. In addition, it has also been described the integration of the code within the logical structure of the 5G-LENA simulation tool.

The configured grant implementation has been validated through system-level simulations. Furthermore, we have presented some initial evaluations about the potential to use the configured grant to support Industry 4.0 applications with stringent latency requirements to illustrate the potential of the implemented code.

References

- [1] 3GPP TR 36.881 V14.0.0, “Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Study on latency reduction techniques for LTE (Rel 14).” 2016.
- [2] 3GPP TS 38.211 V16.7.0, “Technical Specification Group Radio Access Network; NR; Physical channels and modulation; (Rel. 16).” 2021.
- [3] 3GPP TS 38.321 V16.6.0, “Technical Specification Group Radio Access Network; NR; Medium Access Control (MAC) protocol specification (Rel 16).” 2021.
- [4] “Wireless Network Simulation (5G, 4G, 3G, 2G, IP, TDM).” [Online]. Available: <https://www.gl.com/telecom-test-solutions/wireless-network-simulation-in-your-lab.html>. [Accessed: 10-Dec-2021].
- [5] “End-to-end 5G Network Simulator - Polaris Networks.” [Online]. Available: <https://www.polarisnetworks.net/5g-network-emulators.html>. [Accessed: 10-Dec-2021].
- [6] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Viridis, “SiMu5G—An OMNeT++ library for end-to-end performance evaluation of 5G networks,” *IEEE Access*, vol. 8, pp. 181176–181191, 2020, doi: 10.1109/ACCESS.2020.3028550.
- [7] N. Patriciello, S. Lagen, B. Bojovic, and L. Giupponi, “An E2E Simulator for 5G NR Networks,” *Simul. Model. Pract. Theory*, vol. 96, Nov. 2019.
- [8] 3GPP TS 38.300 V16.7.0, “Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2 (Rel. 16).” 2021.
- [9] P. Schulz et al., “Network architectures for demanding 5G performance requirements: Tailored toward specific needs of efficiency and flexibility,” *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 33–43, 2019, doi: 10.1109/MVT.2019.2904185.
- [10] 3GPP TR 38.913 V16.0.0, “Technical Specification Group Radio Access Network; Study on Scenarios and Requirements for Next Generation Access Technologies; (Rel 16).” 2020.
- [11] “5G-LENA network simulator: <https://5g-lena.cttc.es/>,” 2021.
- [12] “ns-3 | a discrete-event network simulator for internet systems.” [Online]. Available: <https://www.nsnam.org/>. [Accessed: 09-Dec-2021].
- [13] K. Montgomery, R. Candell, Y. Liu, and M. Hany, “Wireless User Requirements for the Factory Workcell NIST Advanced Manufacturing Series 300-8 Wireless User Requirements for the Factory Workcell,” *Adv. Manuf. Ser. (NIST AMS)*, doi: <https://doi.org/10.6028/NIST.AMS.300-8>.
- [14] 3GPP TR 37.910 V16.1.0, “Technical Specification Group Radio Access Network; Study on self evaluation towards IMT-2020 submission (Rel. 16).” 2021.

Anexo B. Cálculo de la latencia máxima y el tiempo de llegada al *bridge* 5G virtual

En este anexo se describe el procedimiento utilizado para calcular A_i y l_{5G_i} . Estos parámetros son necesarios para la definición de los escenarios para llevar a cabo la evaluación de los esquemas de *scheduling* propuestos tanto en el capítulo 4 como en el capítulo 5. El tiempo de llegada A_i del primer paquete de un flujo TSN al *bridge* 5G virtual se calcula como la suma del instante de tiempo en el que se genera el paquete en el sensor y la latencia que el paquete experimenta desde el momento en que se genera en el sensor hasta que se recibe en el puerto de entrada de 5G. Tal y como se muestra en la Figura 20 del capítulo 4, todos los paquetes de datos enviados por los sensores hacia el PLC se transmiten a través del *bridge* 5G virtual. Entre cada sensor y el *bridge* 5G virtual la información pasa por un *bridge* TSN. En este caso, A_i se calcula mediante la siguiente expresión, en el que se ha considerado que todos los paquetes se generan en los sensores en el instante 0:

$$A_i = l_{sensor_i} + l_{links_i} + \sum_{a=1}^x l_{TSN_{a,i}} + t_{DS-TT} \quad (32)$$

siendo l_{sensor_i} el tiempo de procesamiento de la aplicación en el sensor, x es el número total de *bridges* TSN que hay en el camino entre el sensor que genera los datos y el *bridge* 5G virtual y $l_{TSN_{a,i}}$ representa el tiempo de procesamiento de dichos *bridges* TSN. $l_{TSN_{a,i}}$ se define para $a \in [1, \dots, x]$ siendo $l_{TSN_{a,i}}$ nulo si el *bridge* TSN a no está en el camino entre el sensor que genera los datos y el *bridge* 5G virtual. El tiempo que necesita un paquete para transmitirse a través de los enlaces que interconectan los nodos se representa con l_{links_i} y el tiempo de procesamiento en el DS-TT se define como t_{DS-TT} .

El tiempo $l_{TSN_{a,i}}$ que necesita un flujo para atravesar un *bridge* TSN se calcula siguiendo la fórmula (2) definida en el apartado 3.2. Para ello, consideramos los tiempos de procesamiento y de permanencia en cola en cada *bridge* TSN que varían según la carga de tráfico ($t_{independent}$) y el tiempo de transmisión de cada flujo TSN, el cual depende del tamaño del paquete y la velocidad de transmisión de los datos del *bridge* TSN ($t_{dependent}$). Debido a que las distancias entre los *bridges* TSN son cortas, l_{links_i} se considera despreciable. Los tiempos de procesamiento DS-TT y NW-TT se asumen que son iguales a 0.005ms y 0, respectivamente [83]. Por un lado, en el mejor de los casos, los paquetes de datos son transmitidos por un *bridge* TSN tan pronto como llegan al puerto de salida de

dicho *bridge*. Considerando un tiempo de procesamiento del sensor igual a $3\mu\text{s}$ [84], el tiempo de llegada mínimo al *bridge* 5G virtual para un paquete de 250 bytes es de $3\mu\text{s} + \left(\frac{250 \cdot 8 \text{ bits}}{100 \text{ Mbps}} + 1.5\mu\text{s}\right) + 50\mu\text{s} \cong 75\mu\text{s}$. Por otro lado, el tiempo de llegada máximo para un paquete de 250 bytes al *bridge* 5G virtual se asume que es igual a $3\mu\text{s} + 30 \cdot \left(\frac{250 \cdot 8 \text{ bits}}{100 \text{ Mbps}} + 1.5\mu\text{s}\right) + 50\mu\text{s} \cong 700\mu\text{s}$. Esto es debido a que todos los primeros paquetes de 250 bytes de todos los sensores ($N_F = 30$ sensores) se encolan en el *bridge* TSN, tal que el primer paquete del último flujo llega al *bridge* 5G virtual cuando $A_i = 700\mu\text{s}$. Por lo que, los tiempos de llegada de los paquetes al *bridge* 5G virtual se establecen de forma aleatoria entre $75\mu\text{s}$ y $700\mu\text{s}$.

Para calcular la latencia que debe satisfacer la red 5G para cada flujo TSN F_i se ha utilizado la expresión (7) del apartado 3.2:

$$l_{5G_i} \leq l_{e2e_i} - A_i - l_{links_i} - \sum_{b=x+1}^M l_{TSN_{b,i}} - t_{NW-TT} - l_{PLC_i} \quad (33)$$

siendo l_{PLC_i} el tiempo de procesamiento en el PLC, t_{NW-TT} indica el tiempo de procesamiento en el traductor NW-TT, $l_{TSN_{b,i}}$ representa el tiempo que los paquetes de un flujo TSN i pasan por un *bridge* TSN b siendo $b \in [x+1, \dots, M]$. Los parámetros M y l_{links_i} representan el número de *bridges* TSN y enlaces que debe pasar el flujo TSN i desde que sale del *bridge* virtual 5G hasta que llega al PLC, respectivamente. Para calcular l_{5G_i} , se asume que l_{PLC_i} es igual a $1007\mu\text{s}$ [85] y que solo hay un *bridge* TSN entre el *bridge* 5G virtual y el PLC. Por tanto, l_{5G_i} es igual a la ecuación (34) para cada flujo TSN F_i .

$$l_{5G_i} \leq 4000\mu\text{s} - A_i - 0 - \frac{s_i \cdot 8 \text{ bits}}{100 \text{ Mbps}} - 1.5\mu\text{s} - 0 - 1007\mu\text{s} \quad (34)$$