



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

TELECOM ESCUELA  
TÉCNICA VLC SUPERIOR  
DE UPV INGENIEROS DE  
TELECOMUNICACIÓN

# ACCESO GRANT-FREE PARA TRÁFICO mMTC HETEROGÉNEO EN REDES 5G NR CON PRIORIDADES Y ASIGNACIÓN DINÁMICA DE SLOTS

**Autor: Miguel del Amo Padilla**

**Tutor: Jorge Martínez Bauset**

Trabajo Fin de Grado presentado en la Escuela Técnica Superior de Ingeniería de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Graduado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Curso 2020-21

Valencia, 2 de Julio de 2021



## **Resumen**

Nos encontramos actualmente en un escenario en el que cada vez queremos conectar un mayor número de dispositivos, sea cual sea su uso, a Internet. Es por ello porque ha ganado mucha importancia la nueva generación 5G y el mMTC (massive Machine-Type Communications).

Este trabajo se centra en analizar un sistema en el que un punto de acceso gestione la forma de conceder permiso a los múltiples equipos de usuario (dispositivos o nodos) para transmitir a través de un canal compartido; y se divide la prioridad del tráfico a transmitir en alta y baja prioridad, de manera que el tráfico con mayores exigencias tendrá mayor disponibilidad trabajando en saturación. La simulación trata de acercarse lo máximo posible a un escenario real en el que llegan un gran número de nodos que intentan acceder al canal y obtenemos unos resultados estadísticos que demuestran la validez del modelo.

## **Resum**

Ens trobem actualment en un escenari en el qual cada vegada volem connectar un major nombre de dispositius, siga com siga el seu ús, a Internet. És per això perquè ha guanyat molta importància la nova generació 5G i el mMTC (massive Machine-Type Communications).

Aquest treball se centra en simular un sistema en el qual un punt d'accés gestione la manera de concedir permís als múltiples equips d'usuari (dispositius o nodes) per a transmetre a través d'un canal compartit; i es divideix la prioritat del trànsit a transmetre en alta i baixa prioritat, de manera que el trànsit amb majors exigències tindrà major disponibilitat treballant en saturació. La simulació tracta d'acostar-se el màxim possible a un escenari real en el qual arriben un gran nombre de nodes que intenten accedir al canal i obtenim uns resultats estadístics que demostren la validesa del model.

## **Abstract**

In these days, the international telecommunication stage is going through a lot of changes, the number of devices that are connected and requires services situated in the Internet is becoming higher. Because of that increasing, the new generation 5G and mMTC (massive Machine-Type Communications) will gain importance in the next years.

This project is focused on the study of a system where the Access Point grant the access to the multiple UEs (User Equipments) to the system. The UEs obtain this grant to transmit through a common channel. The User Equipments are divided into two types of devices according to its access priority; these are HPT (High Priority Traffic) and LPT (Low Priority Traffic). The main difference is that LPT devices will use the portion of the channel bandwidth not used by the HPT devices. We are trying to approach the simulation to the real facts in order to test and check the model performance.

# Índice general

1. **Introducción**
  - 1.1. **Motivación**
  - 1.2. **Estructura de la memoria**
  - 1.3. **Metodología de trabajo**
2. **Antecedentes, objetivos y contribuciones**
3. **Evolución de las comunicaciones hacia el IoT**
  - 3.1. **Las comunicaciones M2M y el IoT**
    - 3.1.1. **Historia**
    - 3.1.2. **Tecnologías**
    - 3.1.3. **Aplicaciones**
  - 3.2. **La importancia del 5G en el IoT**
    - 3.2.1. **Áreas de aplicación**
    - 3.2.2. **Desarrollo**
    - 3.2.3. **Tecnologías**
    - 3.2.4. **Rendimiento**
    - 3.2.5. **Seguridad**
    - 3.2.6. **Preocupaciones por la salud**
  - 3.3. **Arquitectura en tres niveles: Tiers**
  - 3.4. **mMTC: Massive Machine Type Communications**
4. **Desarrollo del modelo**
  - 4.1. **Estudios anteriores y aportaciones de nuestro modelo**
  - 4.2. **Presentación del escenario**
    - 4.2.1. **Acceso al medio**
    - 4.2.2. **Multiplexación del canal**
    - 4.2.3. **Estructura de la trama**
  - 4.3. **Régimen de llegadas**
    - 4.3.1. **Tráfico de alta prioridad (HPT)**
    - 4.3.2. **Tráfico de baja prioridad (LPT)**
  - 4.4. **Modelo de asignación dinámica de slots y transmisión de paquetes**
  - 4.5. **Preparación y diseño del modelo de simulación**
    - 4.5.1. **Simulación con tráfico de Bernoulli**
    - 4.5.2. **Simulación con tráfico Beta (3,4) para LPT**
5. **Análisis de resultados y estadísticas**
  - 5.1. **Simulación de un tráfico de Bernoulli de alta y baja prioridad**
    - 5.1.1. **Análisis del tráfico de HPT**
    - 5.1.2. **Análisis del tráfico de LPT**
    - 5.1.3. **Conclusiones**
  - 5.2. **Simulación de un tráfico de Bernoulli de alta prioridad y Beta (3,4) de baja prioridad**
    - 5.2.1. **Análisis del tráfico de HPT**
    - 5.2.2. **Análisis del tráfico de LPT**
    - 5.2.3. **Conclusiones**

- 6. Conclusiones y líneas futuras**
- 7. Referencias bibliográficas**

# Índice de figuras

1. Figura 1. M2M.
2. Figura 2. Tarjeta SIM M2M.
3. Figura 3. System On Chip.
4. Figura 4. MQTT.
5. Figura 5. Sistema de gestión de tráfico DGT
6. Figura 6. App salud.
7. Figura 7. Células 5G vs 4G.
8. Figura 8. Objetivos 5G.
9. Figura 9. Espectro 5G.
10. Figura 10. 5G mmWave.
11. Figura 11. Crecimiento 5G.
12. Figura 12. Arquitectura de 3 niveles.
13. Figura 13. LPWAN vs 5G.
14. Figura 14. Grant-Free vs Grant-Based.
15. Figura 15. HTC y mMTC.
16. Figura 16. ALOHA vs ALOHA Ranurado.
17. Figura 17. CSMA/CD.
18. Figura 18. CSMA-No-Persistente.
19. Figura 19. CSMA-1-Persistente.
20. Figura 20. CSMA-p-Persistente.
21. Figura 21. ALOHA vs ALOHA Ranurado vs CSMA.
22. Figura 22. Successive Interference Cancellation.
23. Figura 23. FDMA, TDMA y CDMA.
24. Figura 24. Trama 5G para  $\beta = 3$ .
25. Figura 25. Distribución Beta (3,4)
26. Figura 26. Cálculo de la asignación de slots para la siguiente subtrama.
27. Figura 27. Cálculo de la probabilidad de permiso para la siguiente subtrama.
28. Figura 28. Llegadas con distribución Beta (3,4).
29. Figura 29. Validaciones internas del código.
30. Figura 30. Recogida de resultados para tráfico de baja prioridad.
31. Figura 31. Caudal HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un tráfico  $M_1 \cdot a_1 = 1$ .
32. Figura 32. Retardo HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un tráfico  $M_1 \cdot a_1 = 1$ .
33. Figura 33. Probabilidad de pérdidas HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un tráfico  $M_1 \cdot a_1 = 1$ .
34. Figura 34. Caudal LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un tráfico  $M_1 \cdot a_1 = 1$  y distintos valores de tráfico LPT  $M_2 \cdot a_2$ .
35. Figura 35. Retardo LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un tráfico  $M_1 \cdot a_1 = 1$  y distintos valores de tráfico LPT  $M_2 \cdot a_2$ .
36. Figura 36. Probabilidad de pérdidas LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un tráfico  $M_1 \cdot a_1 = 1$  y distintos valores de tráfico LPT  $M_2 \cdot a_2$ .

37. **Figura 37. Caudal y Probabilidad de Pérdidas HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min}=1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$ .**
38. **Figura 38. Probabilidad de pérdidas LPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$ .**
39. **Figura 39. KPI<sub>1</sub>: Probabilidad de colisión LPT con Beta(3,4).**
40. **Figura 40. KPI<sub>2</sub>: Probabilidad de éxito LPT con Beta(3,4).**
41. **Figura 41. KPI<sub>3</sub>: Número de transmisiones exitosas LPT por subtrama con Beta(3,4).**
42. **Figura 42. KPI<sub>4</sub>: Retardo medio LPT en subtramas con Beta(3,4).**
43. **Figura 43. KPI<sub>5</sub>: Evolución del número de dispositivos LPT en el sistema para  $M_2 = 10000, 20000, 30000$  con Beta(3,4).**

# Índice de tablas

1. **Tabla I. Distribución temporal del proyecto en diagrama.**

# 1. Introducción

## 1.1. Motivación

El mundo está situado en un momento en el que la sociedad necesita cada vez estar más conectada, buscando la máxima actualidad e instantaneidad en los datos sobre nuestro entorno global. Para cumplir estas necesidades llegan nuevas tecnologías con numerosos avances tecnológicos para satisfacer estos requerimientos que no paran de crecer.

Así, a medida que los desarrollos tecnológicos avanzan, observamos que la cantidad de elementos cotidianos conectados a la red son innumerables, como sucede con la *Internet de las Cosas (IoT)*. Ante este escenario, nos encontramos que cualquier dispositivo puede conectarse a Internet, desde una bombilla hasta unas zapatillas; y podemos extenderlo al contexto de una ciudad, un pabellón polideportivo o cualquier lugar de nuestro día a día. La cantidad de transmisores de datos aumentan y esto hace que se deban desarrollar nuevas tecnologías y metodologías para poder conectar esta gran cantidad de dispositivos a la red.

Se necesitan por tanto nuevas técnicas e infraestructuras que soporten este incremento de información transmitida a través de la red. Las tecnologías anteriores al 5G, ya comenzaban a contemplar este escenario. El 4G LTE (Long Term Evolution) ya contempla esta problemática de la Internet of Things (IoT); tratar de gestionar de la manera más eficiente posible los recursos disponibles para garantizar a los dispositivos la transmisión. Pero la solución que aparenta ser más eficiente para el IoT es el uso de 5G por numerosos motivos, como la cantidad de dispositivos y otros que veremos en este trabajo. Los investigadores han situado el foco del estudio sobre el 5G en tres principales campos: una red móvil con mejores prestaciones (eMBB), la IoT (mMTC) y un tipo de comunicaciones con altas prestaciones y garantías (URLLC). En este trabajo nos centraremos en el estudio de mMTC.

La Internet de las Cosas presenta dispositivos que se van a conectar a la red, pero su principal característica no será la transmisión de grandes cantidades de datos sino todo lo contrario, transmiten pequeñas cantidades de datos normalmente en ráfagas. Dicho esto, los estudios se van a focalizar en intentar converger hacia un modelo de gestión del tráfico por parte del punto de acceso que logre atender a todos los dispositivos y optimizar cada uno de los parámetros de la conexión.

Entendiendo la complejidad de estos estudios, en este trabajo se va a tratar de realizar un modelo de gestión para tráfico con dos prioridades y probar su validez mediante un modelado basado en simulación por eventos discretos que obtengan resultados descriptivos.





## 2. Antecedentes, objetivos y contribuciones

Para poder resumir de forma concreta los objetivos del proyecto, comenzamos explicando cuál es el punto de partida del mismo. Y es, que para la realización del proyecto hemos partido de un artículo publicado para el organismo IEEE <sup>[1]</sup> en el que se defiende un modelo para gestión del tráfico mMTC (una de las tres ramas troncales del 5G), basado en el acceso Grant-Free y la asignación dinámica de slots, que explicando a lo largo del trabajo. Es artículo presenta un escenario concreto de trabajo en el que coexisten dispositivos de dos prioridades distintas, con régimen de llegadas específico con una probabilidad de llegadas definida.

Una vez presentado esto, este trabajo tiene como objetivo principal el de la descripción, estudio, simulación y análisis de un esquema de gestión y asignación del canal de tráfico para el ecosistema de mMTC (massive Machine Type Communications) dentro de 5G, que podemos dividir en los siguientes objetivos más concretos:

- Realizar de una manera básica, entendiendo la complejidad del modelo, un programa con un lenguaje de programación que consiga simularlo de manera real.
- Conseguir unos resultados que sean suficientemente realistas y precisos, y que permitan entender los fenómenos de tráfico que aparecen en un escenario de acceso masivo, así como a proponer soluciones a los problemas que se identifiquen
- Demostrar la validez del modelo de gestión de tráfico estudiado, así como mostrar los registros y estadísticas recabados para afianzarlo.
- Encontrar posibles puntos débiles, para el futuro estudio de nuevas líneas de trabajo para la continuación del estudio y la mejora del modelo teórico.

En el trabajo, las aportaciones realizadas consisten en la construcción de una serie de mejoras que suponen una generalización del modelo mediante la modelación del sistema de forma que se mantiene el escenario propuesto, junto con las suposiciones previas y a partir de esto realizamos las modificaciones. Proponemos un nuevo régimen de llegadas diferente que permite evaluar el sistema en condiciones muy diferentes a las del artículo, ya que buscamos la evaluación en régimen transitorio del sistema. Además, el modelo de simulación construido es adaptable a cambios en el escenario, como por ejemplo a cualquier tipo de distribución de llegadas o a la gestión de las colas si es necesario.



## 3. Evolución de las comunicaciones hacia el IoT

En este capítulo se van a explicar los conceptos teóricos fundamentales para poder ubicar el trabajo en contexto, principalmente de qué trata esto de la Internet de las Cosas IoT, qué relación tiene con la comunicación entre máquinas M2M, las aportaciones del 5G NR (“New Radio”), así como la importancia del mMTC y los distintos protocolos de comunicación para su uso.

### 3.1. Las comunicaciones M2M y el IoT

El IoT puede definirse de manera sencilla como la interconexión de los objetos de todo tipo de nuestro entorno, que intercambiar algún tipo de estímulo e interactúan entre ellos a través de Internet. Podríamos llegar a conectar desde sensores de presión o movimiento hasta la ropa que llevamos, es por ello que decimos que la definición es simple, pero a la vez abarca un amplísimo abanico de tareas que abordar para alcanzar el objetivo: que las máquinas se comuniquen sin interacción humana o con la menor posible.

Las comunicaciones Machine to Machine (M2M) o Máquina a Máquina, están referidas a aquellas comunicaciones que se producen directamente entre dos o más dispositivos a través de cualquier canal. Este tipo de comunicación permite que sensores o medidores desarrollen un software con el que comunican la información que registran. Esta información obtenida tras múltiples interacciones entre máquinas, es finalmente centralizada en un concentrador central que lo analiza, como podemos ver representado en la Figura 1.

Aunque hablamos de algo novedoso y aparentemente un proyecto de futuro, es algo que ya se encuentra implantado, ya que se calcula que aproximadamente hay más de 30 mil millones de dispositivos conectados a Internet.



Figura 1. M2M.

### 3.1.1. Historia

Podemos remontar el inicio de este tipo de comunicaciones en 1991 cuando ya se empieza a tratar temas que tienen que ver con IoT, por ejemplo, en la publicación de un artículo en la revista del IEEE que definía la computación ubicua como el envío de paquetes con poca cantidad de datos a un conjunto de dispositivos con el objetivo de automatizar cualquier tipo de proceso. Poco después se crearon las primeras aplicaciones industriales M2M que desarrolló la empresa Siemens para que sus máquinas se comunicaran a través de redes inalámbricas. El módulo usado se llamó M1 y fue utilizado para telemática de vehículos y aplicaciones de rastreo. Y otras empresas como Microsoft comenzaron a invertir en este tipo de dispositivos M2M. Pero no fue hasta 1999 cuando se usó por primera vez el término de Internet de las Cosas para cubrir nuevas comunicaciones entre dispositivos.

A partir del 2000 los dispositivos comienzan a comunicarse entre ellos incluyendo tecnologías más sofisticadas, como GPS, tarjetas SIM especializadas para la transmisión de datos máquina a máquina (Figura 2) y otras tecnologías que iban surgiendo de manera muy rápida. Y en 2004 se fundaron operadores de redes destinados al espacio M2M que incorporaba nuevas funciones en la comunicación como el direccionamiento IP fijo.



**Figura 2. Tarjeta SIM M2M.**

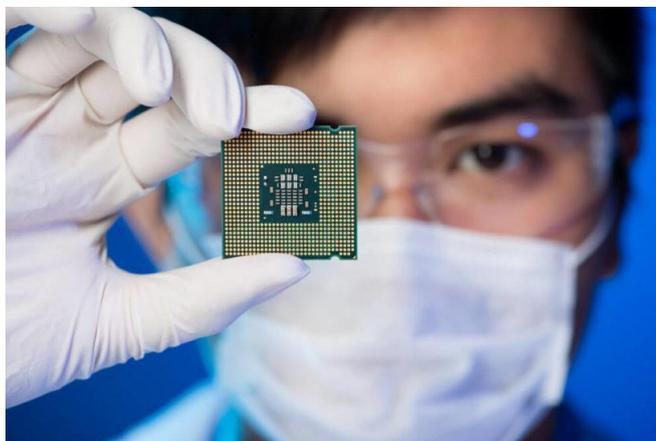
En 2006 se funda la M2Mi, que desarrolla la inteligencia automatizada en los dispositivos para poder comunicarse entre ellos incluyendo en los diseños sensores, servidores, robots y sistemas de redes. Además, poco después se introdujo la gestión en tiempo real de los servicios GSM y CDMA para las comunicaciones entre máquinas, poniendo el foco en mejorar la eficiencia y ahorrar el coste de la red. Con la llegada del 4G/LTE llegaron nuevas aplicaciones destinadas a IoT y se requería una transmisión segura de los datos usando servicios de VPN para M2M, llamado APN. Con la implementación de todas estas mejoras, en 2008 ya existían casi 50 millones de conexiones entre máquinas.

A partir de 2010 el sector se empieza a centrar en el ámbito de la electrónica de consumo y se inicia un crecimiento enorme, miles de compañías comienzan a estudiar protocolos para la comunicación, nuevos diseños de dispositivos, redes mucho más eficientes, protocolos para la seguridad de los datos, etc. Es por ello que afirmamos que, aunque IoT tenga aspecto de algo totalmente nuevo, se encuentra hoy en día en un punto donde gran parte de la puesta en escena ya está llevada a cabo.

### 3.1.2. Tecnologías

Las tecnologías que forman parte del IoT se encuentran en continuo proceso de desarrollo, investigación y mejora.

Podemos comenzar hablando sobre aspectos energéticos, un punto a tener muy en cuenta en cada aparato que se despliegue, que va también de la mano del tamaño del dispositivo que va a conectarse a la red. Destacan los System on Chip que podemos ver en la Figura 3, que son circuitos integrados con múltiples componentes que forman sistema completo y se encargan de todas las funciones de procesamiento. Se ha conseguido desarrollar SoC de bajo consumo logrando un rendimiento más que aceptable para la mayoría de objetos del IoT que sólo van a necesitar transmitir datos de forma esporádica.



**Figura 3. System On Chip.**

Por otro lado, nos encontramos con las barreras de los protocolos de comunicación entre dispositivos. Y el problema se encuentra en que cada fabricante diseña sus protocolos propietarios para sus dispositivos, esto provoca incompatibilidades entre dispositivos. El protocolo abierto MQTT que observamos en la Figura 4 fue creado por IBM y se presenta como un mecanismo estándar que facilita la comunicación entre dispositivos de diferentes compañías.

Y otra parte importante es la tecnología de radio que proporciona acceso a los dispositivos de IoT. La comunicación mediante WiFi supone un alto consumo energético y bajo alcance, por lo que se opta por las redes móviles LTE (4G) y el principal avalista de que IoT cumpla sus objetivos, el 5G.

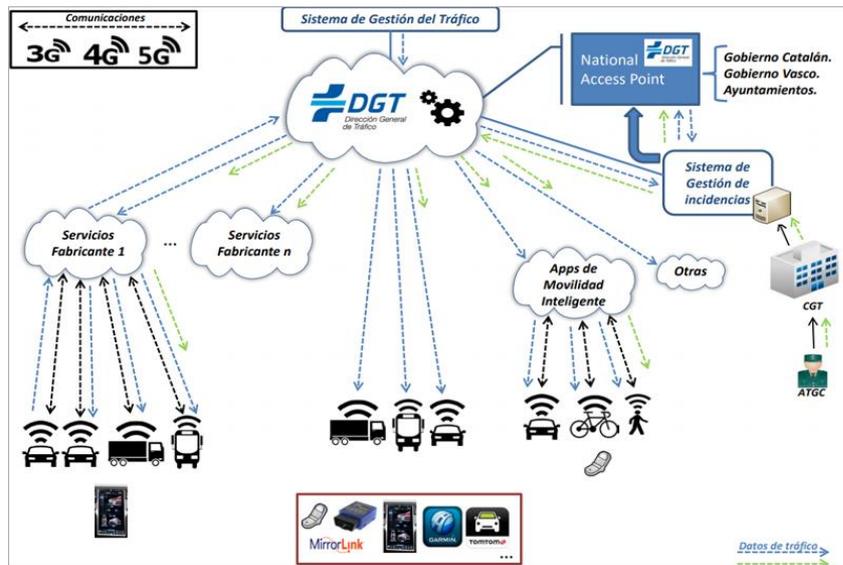


**Figura 4. MQTT.**

### **3.1.3. Aplicaciones**

Principalmente podemos realizar la clasificación de las aplicaciones entre: aplicaciones de consumidor común, aplicaciones con fines comerciales o empresariales, aplicaciones de construcción o agricultura y aplicaciones de bienestar, salud y educación.

- **Aplicaciones de consumidor común.**  
Basadas en optimizar los productos que son comprados por los usuarios, como podría ser el control de la calefacción, electrodomésticos o sistemas inteligentes de transporte para ayudar al conductor de un vehículo que incluyen aplicaciones que permiten monitorear la carga del vehículo, su temperatura, etc. Además, el mundo del entretenimiento y la tecnología textil es una gran apuesta por parte de muchas compañías.
- **Aplicaciones comerciales o empresariales.**  
Aplicaciones como el monitoreo de contadores o medidores inteligentes que previenen el fraude y en control de la información, por ejemplo, el Sistema de Gestión de Tráfico de la DGT de la Figura 5. Por otra parte del mercado, existen medios que evalúan los hábitos de los compradores, mostrándole la publicidad idónea para que tenga la máxima influencia en él; algo similar ocurre con las vallas publicitarias inteligentes. Aunque pueda suponer grandes inversiones, la recaudación supera sobradamente el gasto en la mayoría de los casos.



**Figura 5. Sistema de gestión de tráfico DGT.**

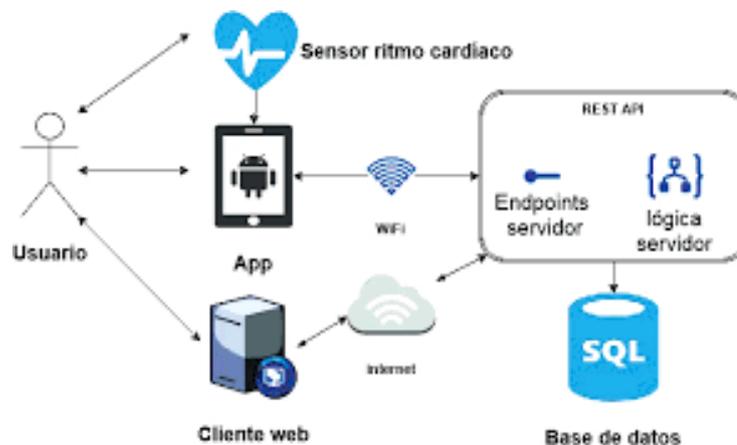
- **Aplicaciones de construcción o agricultura.**

Se basan en garantizar de manera más fiable la seguridad de estructuras y minimizar el riesgo de generar problemas. Estas infraestructuras pueden ser desde navales hasta puentes o placas fotovoltaicas. Además, el IoT proporcionará un seguimiento mucho más sencillo y barato de estas infraestructuras.

Por otra parte, el IoT permitirá controlar y automatizar todos los procesos agrícolas de una manera similar, mediante sensores con unos indicadores que permitan medir la humedad tanto del ambiente como del suelo, consumo de combustible e incluso las infecciones que sufre una plantación.

- **Aplicaciones de bienestar, salud y educación.**

En estos ámbitos la Internet de las Cosas tendrá una influencia muy significativa, ya que suponen una gran mejora el nivel de vida en dos aspectos fundamentales del ser humano. Por una parte, los dispositivos IoT de la medicina sirven para múltiples usos: sensores para monitorear la salud de los ancianos, monitores cardíacos o de presión sanguínea, incluso habitaciones de hospital inteligentes que controlen todos los indicadores de los pacientes o prótesis que se alimentan eléctricamente del propio movimiento del cuerpo humano. En la Figura 6 podemos observar el esquema de una aplicación en la que el usuario puede consultar de forma continua cualquier parámetro controlado por un sensor que mida valores de su organismo.

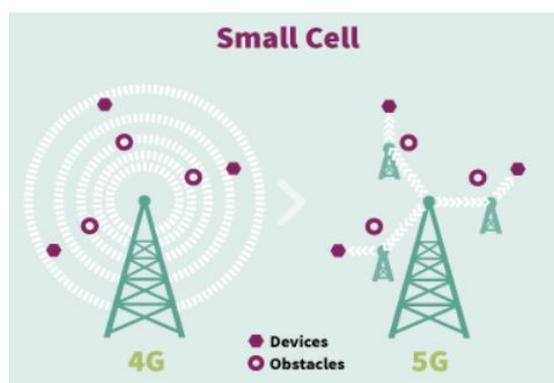


**Figura 6. App salud.**

En la educación cabría destacar dos líneas de actuación. En primer lugar, las innovaciones respecto a la forma en la que enseñar o hacer que los alumnos aprendan de manera adaptada a cada uno. Y en segundo lugar encontramos los dispositivos que directamente realizan tareas como exámenes o la impartición de clases mediante realidad virtual.

### 3.2. La importancia del 5G en el IoT

5G es el estándar tecnológico de quinta generación para redes celulares de banda ancha que empezó a usarse en 2019 como claro sucesor de la tecnología LTE/4G de cuarta generación, de la que disponen los dispositivos actuales. Como tecnología de redes celulares, está formado por distintas celdas geográficas de pequeño tamaño con una antena central que proporciona acceso a todos los aparatos inalámbricos del área, véase la Figura 7. Los avances más destacados de esta nueva tecnología será el aumento de la velocidad de transmisión hasta 10 Gbps y del número de dispositivos por kilómetro cuadrado hasta 1 millón, lo que hará posible nuevas y mejores aplicaciones de IoT y M2M.



**Figura 7. Células 5G vs 4G.**

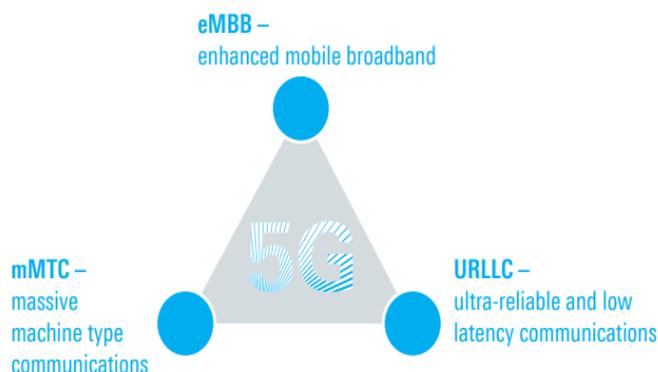
En las pruebas que se han realizado durante el año 2020 y principios de 2021 se han logrado alcanzar velocidades de hasta 1Gbps en banda de frecuencia desde 25 a 39 GHz (ondas milimétricas) y podría ser superior en los próximos años ya que se aumentarán frecuencias más

altas. Por otra parte, con frecuencias inferiores, de entre 2.5 y 3.7 GHz, se han conseguido velocidades de más de 500 Mbps y una distancia alcanzada mucho mayor que con altas frecuencias, hasta varios kilómetros con una antena. Las desventajas de la banda de alta frecuencia son obviamente el menor alcance y la dificultad para atravesar obstáculos a cambio de una velocidad mucho mayor. Adicionalmente, 5G ha heredado uno de los últimos avances de LTE, MIMO (Multiple-Input Multiple-Output) y que consiste en el incremento del número de pequeñas antenas en cada celda y así poder gestionar de forma más cómoda la transmisión en distintos rangos de frecuencia.

Para terminar, cabe mencionar que, si bien el 5G dará paso al despegue final del IoT, es cierto que los ritmos de desarrollo no son paralelos <sup>[5]</sup>. La Internet de las Cosas ha avanzado sin necesidad del despliegue del 5G, y esto se debe a lentitud que supone la construcción de nuevas infraestructuras necesarias para esta nueva tecnología. Aunque IoT desarrolle nuevos proyectos, es incuestionable que no va a mostrar su total potencial hasta la consolidación del 5G.

### 3.2.1. Áreas de aplicación

Los principales campos de aplicación del 5G siguen la línea de las mejoras, y la UIT (Unión Internacional de Telecomunicaciones) ha definido 3 áreas (Figura 8): tecnología móvil de banda ancha eMBB (enhanced Mobile BroadBand), comunicaciones de baja latencia ultra fiables URLLC (Ultra-Reliable and Low Latency Communications) y comunicaciones masivas tipo máquina mMTC (massive Machine Type Communications).



**Figura 8. Objetivos 5G.**

- **eMBB.** Las redes de banda ancha móvil mejorada representan, con la llegada del 5G, un mejor rendimiento y mayores velocidades en zonas con altas cargas de tráfico como grandes establecimientos de la metrópoli.
- **URLLC.** Las comunicaciones de baja latencia están destinadas a la transferencia de datos con altas garantías.
- **mMTC.** La tecnología 5G conectará en un futuro cercano la gran mayoría de los 50 mil millones de dispositivos IoT que hay en el mundo, dispositivos móviles, drones, automóviles, domótica... Nos centraremos más adelante en este campo del 5G para desarrollar el modelo.

### 3.2.2. Desarrollo

En los primeros años de evolución, el 5G ha comenzado usándose principalmente en las redes de operadores móviles. Sin embargo, está ya entrando en el ámbito privado con aplicaciones de IoT industrial y empresarial, así como en comunicaciones más críticas.

Como ya se ha comentado, en el ámbito de las redes de operadores ya se han realizado avances muy significativos; casi 300 operadores en todo el mundo ya están probando redes 5G en diferentes países como Corea del Sur, el primero en lanzarlo a gran escala a través de equipos principalmente de la empresa surcoreana Samsung.

Todavía no existen gran número de compañías con dispositivos 5G en el mercado, pero cada vez van surgiendo más proveedores de teléfonos móviles, hotspots, equipos interiores y exteriores, módulos, adaptadores, dispositivos USB, etc. En el campo del IoT, ya existen conjuntos de chips modem 5G comerciales, procesadores y se esperan muchos más.

### 3.2.3. Tecnologías

Como ya se ha mencionado anteriormente, 5G utiliza massive MIMO (Multiple-Input Multiple-Output), que aumentan la densidad de capacidad con numerosas antenas por célula y también el rendimiento de la misma. Además, con 5G se desarrolla paralelamente MIMO multiusuario (MU-MIMO) que implica que una sola antena de un dispositivo se comunice con múltiples antenas de diferentes tipos; a diferencia de lo que ocurre cuando un dispositivo dispone de varias antenas, una para cada tipo de comunicación.

De la misma forma, se han incluido también nuevas frecuencias de radio presentadas a continuación y en la Figura 9:

- **Rango de frecuencia 1.** Frecuencia menor a 6 GHz con un ancho de banda máximo del canal de 100 MHz. La banda más utilizada para 5G dentro de este rango es de 3.3 a 4.2 GHz, debido a la gran ocupación del canal en esta zona del espectro.
- **Rango de frecuencia 2.** Frecuencias mayores a 24 GHz con un ancho de banda máximo de 400 MHz y con agregación de dos canales disponible. Se usan bandas en 26 GHz o incluso en 39 GHz obteniendo una velocidad de transferencia más alta. Estas altas frecuencias provocan una decaída en la distancia máxima alcanzada obligando a colocar estaciones base cada pocos cientos de metros y además de la dificultad de penetrar objetos sólidos como paredes, automóviles, árboles o altos edificios.



**Figura 9. Espectro 5G.**

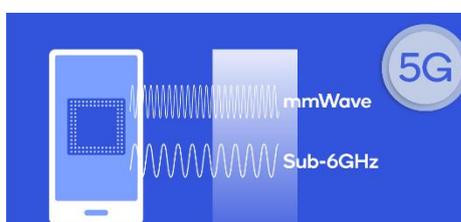
Una vez distribuido el espectro del canal y elegidas las bandas para 5G, ahora el dilema surge a la hora de cómo realizar la asignación del canal para los distintos dispositivos que quieren transmitir a través de él y a veces de forma simultánea, obligando a investigar, desarrollar e implementar diferentes pruebas y modelos que resulten en el mínimo retardo posible y también el mínimo consumo energético, para una utilización óptima del canal y de la energía.

Aquí surgen diferentes líneas de trabajo, investigaciones o ideas de diferentes tipos que realmente aún continúan en desarrollo debido a la contemporaneidad del problema y a que el modelo más óptimo va a depender del tipo de dispositivos donde se vaya a implantar, la forma que tenga de abastecerse su consumo eléctrico y la tasa que necesiten. Todos estos modelos son desarrollados enfocando el 5G hacia la comunicación masiva de tipo máquina mMTC y entraremos más en detalle en el siguiente apartado.

### 3.2.4. Rendimiento

Podemos dividir el rendimiento del 5G en diferentes características:

- **Velocidad.** La tecnología de anterior generación, 4G LTE consigue una velocidad de hasta 300 Mbps. Estas velocidades tendrán un equivalente en 5G, en la banda por debajo de 6 GHz (banda media). Pero la gran mejora del 5G se encuentra en el 5G mmWave o de onda milimétrica, con velocidades de varios Gbps, se han conseguido de aproximadamente 2Gbps en pruebas y se espera acercarse a la decena de Gbps.



**Figura 10. 5G mmWave.**

- **Latencia.** La latencia de la propagación de las ondas a través del aire oscila entre 8 y 12 milisegundos y habría que sumarle a esto la latencia de los servidores, que se encuentra

entre 10 y 20 milisegundos. En pruebas de laboratorio se ha llegado a conseguir una latencia de servidor de 4 ms, pero será complicado implantarlo fuera de él.

- **Tasa de error.** 5G utiliza una codificación del canal que mantiene baja la tasa de error de bit o Bit Error Rate (BER) a cambio de sacrificar ancho de banda.
- **Distancia.** Esta distancia dependerá de la banda de frecuencia: aproximadamente 200 metros para onda milimétrica, hasta 100 km con banda media y banda baja.

### 3.2.5. Seguridad

Existe una gran preocupación por la seguridad en 5G, ya que muchos investigadores ponen en duda que la tecnología 5G desarrollada hasta el momento esté lo suficientemente probada como para transmitir enormes cantidades de datos. Y si enfocamos el objetivo hacia IoT, multiplicamos el problema porque además de generar y transmitir cada vez más datos, a su vez también nos encontramos con un incremento en los posibles puntos de ataque de tipo DDoS (Denegación de Servicio Distribuida), cryptojacking (secuestrar recursos de un dispositivo o conjunto de dispositivos para destinarlos a la minería de criptomonedas), etc. Hay que tener en cuenta que se esperan 21.5 mil millones de dispositivos conectados a 5G en los próximos 3 años. Además, si a todo esto le sumamos que se teme que se dé uso de los datos para el espionaje entre países, se convierte en un problema mundial de gran envergadura.

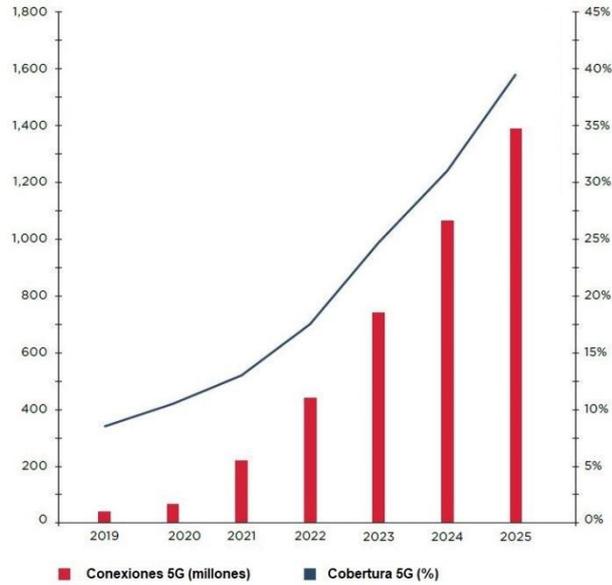
Por último, existen una serie de teorías conspiratorias que relacionan el 5G con el SARS-CoV-2 (Covid-19), que ha dado lugar a ataques revolucionarios contra centros de telecomunicaciones en diversos países.

### 3.2.6. Preocupaciones por la salud

Para que la tecnología de 5G se desarrollara, debía existir y existe un consenso científico global que afirma que es segura. Aunque es un aspecto demostrado que los límites de seguridad actuales para la exposición a la radiofrecuencia siguen siendo más que aceptables para proteger la salud pública, existen miles de propagandas de desinformación.

Se ha emitido programación, se ha publicado en medios y en internet información que afirma que provoca cáncer cerebral, infertilidad, autismo, Alzheimer o tumores cardiacos; y todo ello sin apoyo científico alguno.

Estas “desinformaciones” han provocado durante el desarrollo del 5G que varios países tuvieran que paralizar el proceso por el miedo que causan en la población.

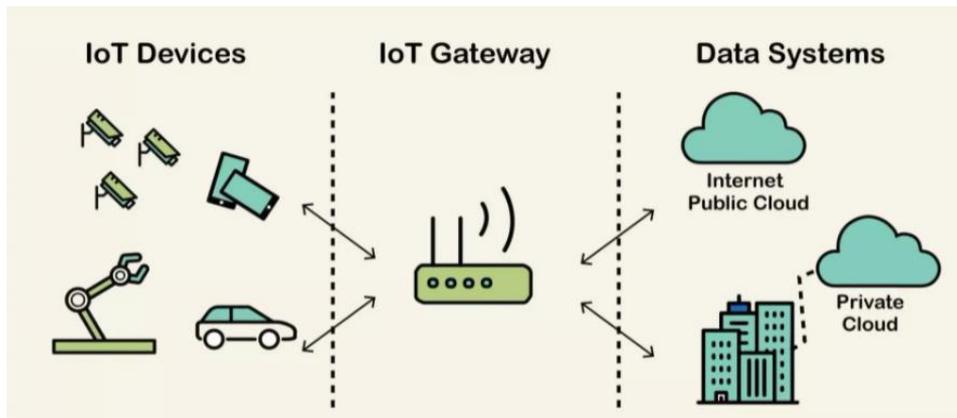


**Figura 11. Crecimiento 5G.**

### 3.3. La arquitectura de 3 niveles: Tiers

El esquema del flujo de datos de las comunicaciones IoT en 5G estaría formado principalmente por tres estructuras diferenciadas en la Figura 12:

- Dispositivos. Representan los transmisores de información recopilada a través de sensores o de cualquier otra forma.
- Gateway o punto de acceso <sup>[10]</sup>. Representa al controlador que recibe la información de los dispositivos mediante el canal inalámbrico y se conecta a la red de Internet.
- Internet.



**Figura 12. Arquitectura de 3 niveles.**

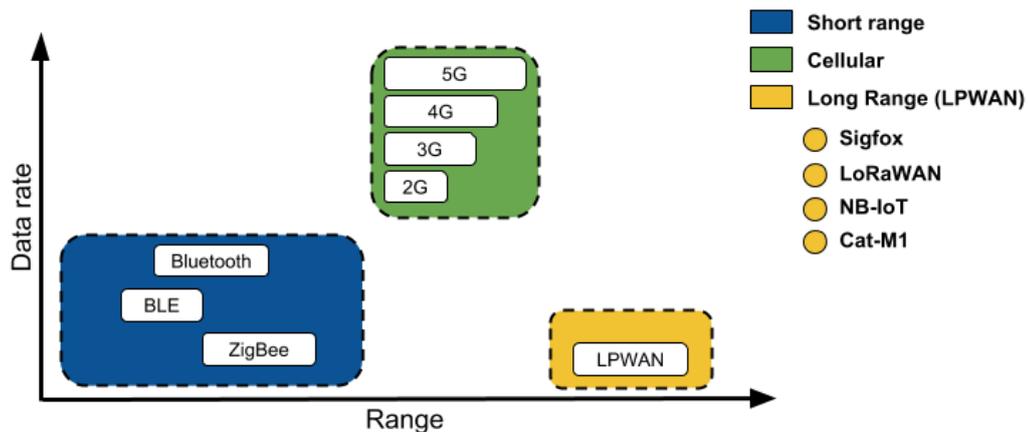
Como podemos observar en la Figura 12, el punto clave de esta arquitectura es el Gateway IoT. Éste realiza las siguientes funciones:

- Establece un puente entre los sensores de los dispositivos IoT y la red de Internet.
- Recoge y traduce los datos del protocolo de los sensores procesándolos y filtrándolos para su posterior transmisión.
- También sirve como caché de datos para guardar información determinada y ahorrar tráfico de datos tanto con internet como con los dispositivos.
- Ofrece servicios de seguridad adicional.
- Dispone de inteligencia adicional para algunos dispositivos IoT.
- Gestiona el canal de transmisión: concede acceso a los dispositivos puedan transmitir en el medio inalámbrico mediante algoritmos predefinidos que permiten gestionar la congestión del medio de transmisión, maximizando el caudal a través del mismo (reduciendo el número de retransmisiones y el retardo de los paquetes). También gestiona la señalización con el resto de Internet.

### **3.4. mMTC: Massive Machine Type Communications**

Entendemos el massive Machine Type Communications como la gran área de aplicación de la nueva tecnología de quinta generación 5G. Este tipo de comunicaciones es muy distinto a lo que se denomina HTC (Human Type Communications), comunicaciones de tipo humano, se produce un cambio de paradigma que revoluciona las comunicaciones y da lugar a nuevos modelos de comunicación y desafíos que abordar. Este nuevo tipo de comunicaciones, las mMTC no requieren una alta velocidad de descarga, al contrario de lo que sucede con las comunicaciones HTC que priorizan la alta velocidad. El motivo es, principalmente, el tipo de información que se transmite en las mMTC son paquetes con una pequeña carga útil, ya que suelen ser mediciones de sensores y señales de monitorización. Y por esta razón, en las comunicaciones masivas de tipo máquina se debe priorizar en reducir al máximo o incluso eliminar las señales de control, puesto que su tamaño puede llegar a ser tanto como el de las señales de datos.

La tecnología mMTC trajo consigo un debate entre distintas tecnologías celulares para proporcionar acceso a los dispositivos situados en lugares aislados. Son las redes LPWAN <sup>[8]</sup>, redes de área amplia y bajo consumo; todas ellas tienen en común la transmisión de bajos ratios de información, cada una con diferentes frecuencias de transmisión y estándares. De esta manera, nos encontramos con NB-IoT <sup>[7]</sup> (Narrow Band-IoT) que es una tecnología de ancho de banda estrecho para mMTC, la tecnología LoRaWAN, una solución sin licencia para mMTC e IoT y Sigfox, compañía que desarrolla una tecnología simple para conexiones de muy bajo ancho de banda. La apuesta del 3GPP (3rd Generation Partnership Project) es NB-IoT, que consiste en una tecnología que satisface las necesidades de dispositivos de una velocidad de datos baja, normalmente con baterías como fuente de alimentación. Su principal objetivo es que se consiga un estándar alrededor de ella para los dispositivos IoT y se ha desarrollado para que dispositivos simples se conecten a una red de operador mediante el espectro con licencia, no a redes IP.

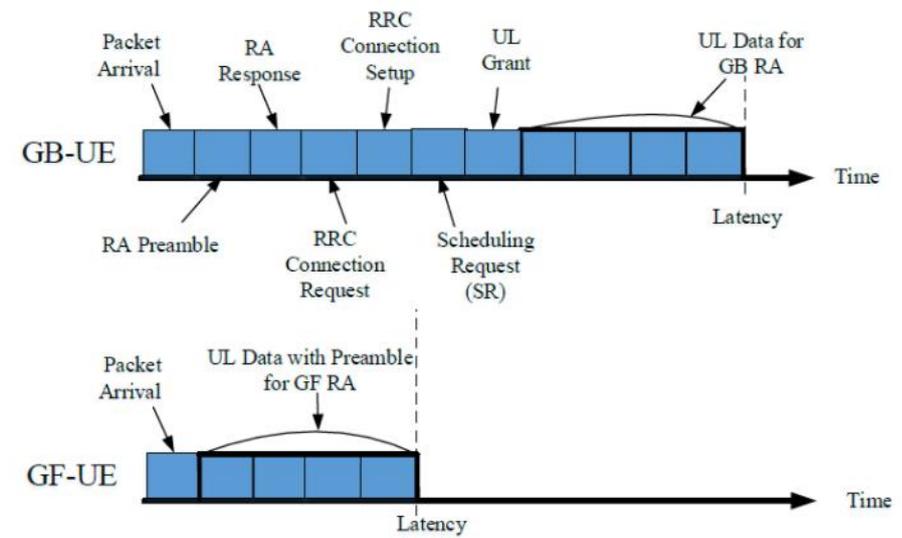


**Figura 13. LPWAN vs 5G.**

Pero ahora vamos a centrarnos en el caso en el que existen miles de dispositivos tratando de transmitir, que es aquí donde se hace clave el uso del 5G. Y más en concreto vamos a centrarnos en la forma de transmitir los datos a través del canal inalámbrico. El 3GPP ha reconocido que además del desarrollo del 5G NR, es necesario mejorar sus estándares para admitir aplicaciones de IoT de manera más eficiente. Se ha comenzado a investigar nuevos mecanismos adaptados a IoT, como **NOMA** <sup>[9]</sup> (Non-Orthogonal Multiple Access) y el **Acceso Grant-Free** (acceso sin concesión). Éste último permite que el nodo o dispositivo UE (User Equipment) transmita datos en forma “arrive-and-go”, el dispositivo llega (se conecta), transmite, y se va (se desconecta), dentro de un intervalo definido. La diferencia con los modelos que usan el mecanismo Grant-Based Access (acceso basado en concesión) es que con el nuevo mecanismo los dispositivos no necesitan esperar a que la estación base les comunique un permiso para transmitir.

El acceso Grant-Free es mucho más óptimo para las comunicaciones mMTC, ya que presenta:

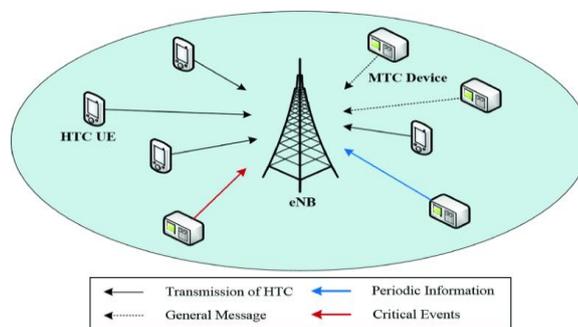
- Una menor latencia o retardo de transmisión. Ver Figura 14.
- La sobrecarga por señalización se ve muy reducida o eliminada como es el caso de nuestro modelo de acceso. Podemos observar en la Figura 14 como se reduce enormemente la cantidad de datos transmitidos en forma de control o señalización al eliminar el preámbulo usado para la contienda en el modelo Grant-Based.
- Mayor duración de la batería de los dispositivos, gracias a: la disminución del tráfico de señalización y el tiempo de encendido de los dispositivos por el mecanismo arrive-and-go.



**Figura 14. Grant-free vs Grant-based.**

Del mismo modo, el acceso múltiple no ortogonal (NOMA) es un mecanismo que permite la superposición de distintos dispositivos transmitiendo en el espectro, utilizando multiplexación en potencia o en código. La combinación de NOMA y Grant-Free da lugar a escenarios mMTC con un consumo de energía reducido y acceso rápido a la red por parte de los equipos, entre otras muchas ventajas. Sin embargo, la estandarización de los accesos Grant-Free y NOMA están aún en una etapa inicial.

En esta línea de progreso, el objetivo a alcanzar es el de tratar de unificar las distintas investigaciones realizadas en este campo para conseguir un modelo conjunto de transmisión y acceso para cuando haya miles de máquinas o dispositivos IoT compitiendo por la concesión del acceso. Para ello se están llevando a cabo estrategias que combinan técnicas basadas en modelos anteriores junto con estas estrategias de acceso innovadoras que unen por un lado el acceso aleatorio y por otro la programación adaptada a mMTC. En el modelo que hemos desarrollado, se combinan técnicas de acceso Grant-Free y además incluye acceso con prioridades para dos tipos distintos de tráfico. Asimismo, todo ello tiene que estar preparado para coexistir con los tipos de comunicaciones HTC ya existentes como podemos ver en la Figura 15 de a continuación.



**Figura 15. HTC y mMTC.**



## 4. Desarrollo del modelo

### 4.1. Estudios anteriores y aportaciones de nuestro modelo

En el escenario actual descrito alrededor de la Internet de las Cosas, 5G y la comunicación masiva tipo máquina, el número de equipos de usuario UEs (User Equipment) se encuentra en un aumento exponencial y, por tanto, el objetivo con 5G es admitir una densidad de usuarios mucho mayor y con un retardo mucho menor que con LTE. Una de las claves en el sentido de la búsqueda de este objetivo es encontrar la forma óptima en la que los nodos (UEs) se conectan al punto de acceso o nodo B de nueva generación, gNB (Next Generation Node B), al RACH (Random Access Channel) a través de un mecanismo de acceso aleatorio RA (Random Access).

El modelo que se estudia en profundidad con este proyecto es el de **un acceso sin concesión con asignación dinámica de slots para dos tipos de tráfico con prioridades y patrones de tráfico diferentes mMTC en redes 5G NR**. Este modelo es presentado de forma analítica en <sup>[1]</sup>, que estudia el comportamiento del sistema en régimen permanente, es decir, el tráfico de llegada a los dispositivos se mantendrá constante (se aplica una distribución de llegadas que sea constante) durante toda la duración de la simulación. De esta forma se analiza el funcionamiento del sistema con unos parámetros concretos determinados, pero no se consigue observar el comportamiento de un régimen transitorio, como por ejemplo la llegada de una ráfaga de paquetes.

En el **presente proyecto** analizamos tanto el régimen permanente como el transitorio, esto nos permite obtener unos resultados más genéricos sobre el funcionamiento del sistema, ya que observamos cómo responde el sistema ante diferentes estímulos de tráfico, así como demostrar los objetivos fijados que validen el uso del mismo. Además, este modelo permite estudiar escenarios con perfiles de tráfico más generales que el de Bernoulli, así, modelamos una nueva distribución de tráfico de baja prioridad. Vamos a presentar a lo largo de este capítulo todas las características del modelo, los análisis realizados, los resultados esperados, así como las modificaciones realizadas en forma de mejoras para estudiar a continuación el rendimiento de las mismas.

- **Acceso Grant-Free.** Existen algunos modelos que incluyen una subtrama de contienda en la que los dispositivos comunican que quieren transmitir y el punto de acceso les asigna un slot en la subtrama de datos. En nuestro caso, la contienda se realiza directamente mediante la transmisión del paquete y esto tiene sentido ya que la cantidad de datos que se van a transmitir en un paquete es muy pequeña y la carga de señalización sería relativamente alta. El dispositivo (o nodo) que quiere transmitir datos, no va a dedicar recursos para solicitar acceso, simplemente va a transmitir en un slot elegido, si lo hace con éxito ya ha terminado su transmisión y si colisiona con la transmisión de otro nodo en el mismo slot, tendrá que volver a intentarlo más tarde.

- **Asignación dinámica de slots**, es decir, en cada subtrama, será variable el número de slots que se asignan a cada tipo de tráfico o prioridad ofreciendo flexibilidad en la gestión de la congestión. El modelo de asignación se explicará más adelante.
- **Dos tipos de tráfico**. Existirán nodos con un tráfico de alta prioridad HPT (High Priority Traffic) y de baja prioridad LPT (Low Priority Traffic). El régimen de llegadas de paquetes, para ser transmitidos al punto de acceso, podrá ser diferente para cada tipo de tráfico.

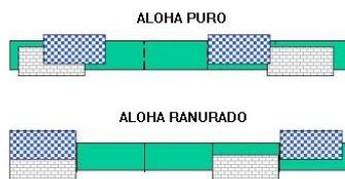
## 4.2. Presentación del escenario

### 4.2.1. Acceso al medio

Existen distintos mecanismos de acceso aleatorio al medio:

- **ALOHA**. Protocolo de acceso al medio en el que cuando la estación tiene un paquete para transmitir, lo envía. Estos paquetes contienen una identificación del destino y un código de detección de errores. La estación, una vez transmite, se queda a la espera de recibir confirmación (ACK) durante un tiempo determinado; si lo recibe la considera transmitida y si no lo recibe la vuelve a enviar excepto si lleva  $n$  intentos de retransmisión sin recibir ACK. En ALOHA, la utilización máxima del canal es del 18%.
- **ALOHA ranurado FSA (Frame Slotted Aloha)**. Es una mejora de ALOHA en la que los dispositivos están sincronizados en el tiempo, ahora existen ranuras (slots) temporales para todos los dispositivos, se puede ver en la Figura 16. En este caso, sólo se puede transmitir un paquete al comienzo del slot, si un paquete se genera a mitad de un slot, se espera al siguiente para enviarlo. Ahora, la utilización máxima del canal es del doble, 36% aproximadamente.

El sistema estudiado en el proyecto utiliza un acceso de tipo **ALOHA ranurado**.



**Figura 16. ALOHA vs ALOHA Ranurado**

- **CSMA (Carrier Sense Multiple Access)**. El acceso múltiple con detección de portadora es un protocolo de acceso al medio basado en ALOHA, pero ahora la estación “escucha” el medio antes de transmitir, y si lo encuentra vacío transmite. Esto es básicamente la detección de portadora, detectar la presencia de una señal portadora de otro nodo antes de transmitir. Existen distintas variantes de CSMA:
  - **CSMA/CD (Collision Detection)**: cada estación detecta cuándo hay una colisión mediante una señal de especial de colisión.

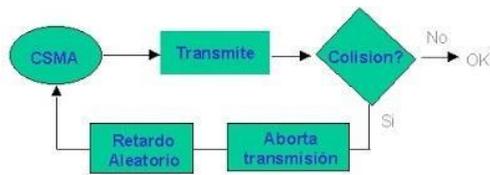


Figura 17. CSMA/CD.

- **CSMA/CA** (Collision Avoidance): incluye prevención de colisiones haciendo que cada equipo anuncie su intención de transmitir antes de hacerlo.
- **CSMA-No-Persistente**: si el dispositivo encuentra el canal libre, transmite, y sino espera un tiempo aleatorio para volver a escuchar.

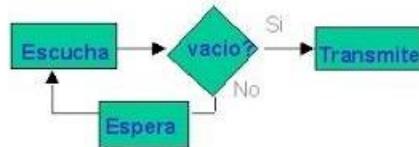


Figura 18. CSMA-No-Persistente.

- **CSMA-1-Persistente**: cada estación transmite en el momento en el que ve el canal libre.



Figura 19. CSMA-1-Persistente.

- **CSMA-p-Persistente**: una vez encuentra el canal libre, el equipo decidirá si transmite o no a través de un algoritmo que dará orden de transmitir con probabilidad  $p$ , o de no transmitir con probabilidad  $1-p$ .



Figura 20. CSMA-p-Persistente.

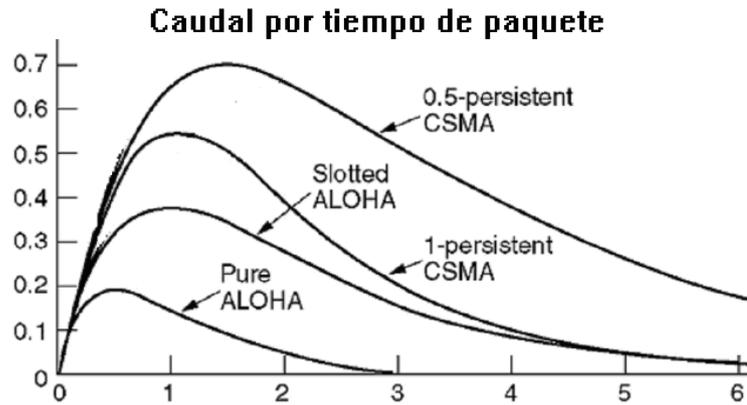


Figura 21. ALOHA vs ALOHA Ranurado vs CSMA.

Además, actualmente ha surgido otra rama de investigación que estudia un tipo de modelo que soluciona las colisiones en el canal llamado **SIC (Successive Interference Cancellation)** [2]. La Cancelación Sucesiva de Interferencias consiste básicamente en la posibilidad de reconstruir un paquete de datos que ha sufrido una colisión con otro paquete de otro dispositivo distinto al intentar transmitir a través del canal. Para poder decodificar la información a pesar de la colisión, este tipo de modelos se basa en la potencia con la que se ha transmitido la señal, el dispositivo que esté más cercano al punto de acceso, alcanzará éste con una mayor señal y por lo tanto será posible bajo determinadas condiciones su decodificación y una vez decodificada, se resta a la señal total recibida y así resultaría la señal más débil. Lo podemos ver de una forma muy descriptiva en la Figura 22 Su principal inconveniente es el consumo energético ya que, con los modelos desarrollados hasta ahora, necesita una mayor cantidad de energía.

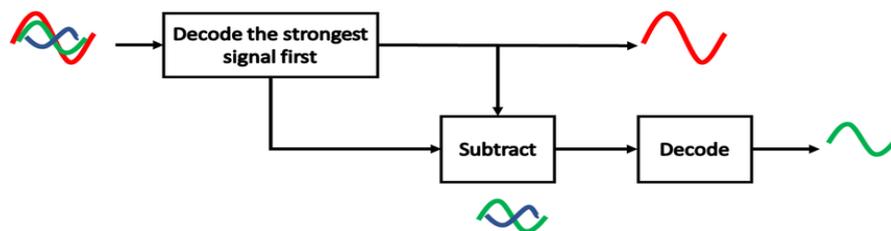


Figura 22. Successive Interference Cancellation.

El esquema modelado en este trabajo utiliza los principios de **ALOHA ranurado (Slotted Aloha)** con algunas modificaciones, debidas a que el modelo se orienta a la estructura de la trama de 5G-NR.

#### 4.2.2. Multiplexación del canal

Las técnicas de multiplexación del canal se presentan a continuación podemos verlo mucho mejor de manera gráfica en la Figura 23:

- **CDMA** (Code Division Multiple Access). Es un método para el control de acceso al medio que usa una tecnología de espectro expandido y un esquema de codificación mediante el que a cada estación se le asigna un código único.
- **FDMA** (Frequency Division Multiple Access). Es una técnica de multiplexación en la que el espectro se divide en canales con distinta frecuencia y asignando a cada usuario un canal de manera que no interfieren entre sí.
- **TDMA** (Time Division Multiple Access). Técnica que consiste en la multiplexación por división de tiempo.

En las tramas de 5G se realiza una multiplexación FDM ortogonal (OFDM), en la que pueden utilizarse hasta 14 símbolos en un único slot.

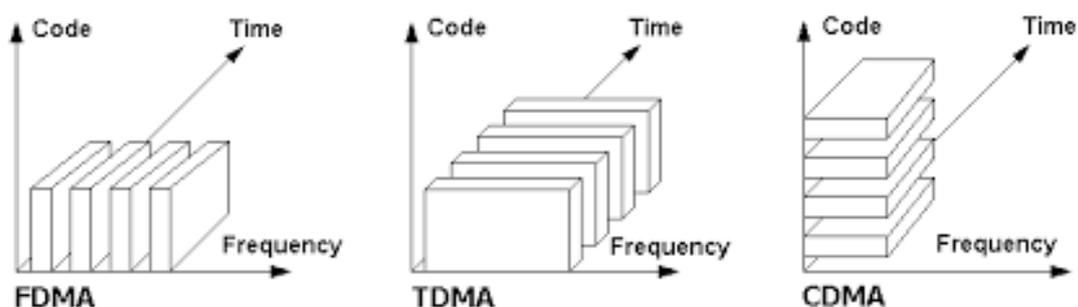


Figura 23. FDMA, TDMA y CDMA.

#### 4.2.3. Estructura de la trama

La estructura de las tramas 5G tiene una anchura de portadora de  $2^\beta \cdot 15$  kHz, para valores de  $\beta \in \{0, 4\}$ . Para cada valor de  $\beta$ , el número de slots por trama será  $2^\beta$ , lo que representa valores desde 1, 2, 4, 8 o 16 slots por subtrama. Esto representa una duración de slot de 1ms ( $\beta = 0$ ) y  $62.5\mu\text{s}$  ( $\beta = 4$ ). La duración de una subtrama es de 1ms, gráficamente más claro en la Figura 24.

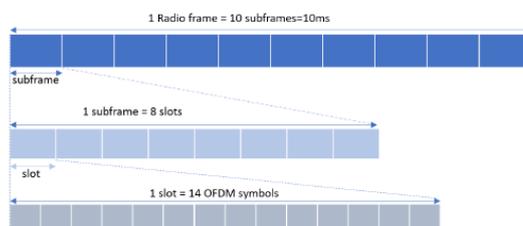


Figura 24. Trama 5G para  $\beta = 3$ .

### 4.3. Régimen de llegadas

El prototipo del trabajo, como ya hemos comentado, presenta dos tipos de prioridades con dos tipos de tráfico distintos.

#### 4.3.1. Tráfico de alta prioridad (HPT)

Los nodos de alta prioridad recibirán un tráfico (**HPT**) de **Bernoulli**, que será independiente en cada nodo.

Una distribución de Bernoulli es un modelo teórico que representa una variable aleatoria discreta en la que sólo pueden ocurrir dos eventos: éxito o fracaso. Aplicándolo al modelo, significa que la probabilidad de que llegue un paquete a uno de los nodos durante una subtrama es  $a_1$ , llamada probabilidad de activación de los nodos de alta prioridad, mientras que la probabilidad de que no llegue ningún paquete durante una subtrama será  $1 - a_1$ .

#### 4.3.2. Tráfico de baja prioridad (LPT)

En los nodos de baja prioridad vamos a trabajar con dos tipos distintos de tráfico, por un lado, van a ser objeto de un tráfico con llegadas que siguen una distribución de Bernoulli, y por otro lado un tráfico de distribución Beta(3,4). Estos dos escenarios son totalmente distintos y tratan de evaluar el sistema en un estado estacionario, con Bernoulli, y en un estado transitorio, con Beta(3,4)

En el primer caso estudiaremos el sistema con los nodos de baja prioridad recibiendo un tráfico de Bernoulli, tal que la probabilidad de que llegue un paquete a un nodo de este tipo durante una subtrama será  $a_2$ , o también probabilidad de activación de los nodos de baja prioridad, mientras que la probabilidad de que no llegue ningún paquete durante una subtrama será  $1 - a_2$ .

El modelado del comportamiento de los nodos y sus colas con este tipo de tráfico tiene como objetivo el estudio de los parámetros que presenta el sistema en régimen permanente, que es alcanzado debido a que la probabilidad llegada de tráfico a los nodos con una distribución de Bernoulli no presenta variaciones en función del tiempo, es decir, será un tipo de tráfico constante durante todo el tiempo de simulación.

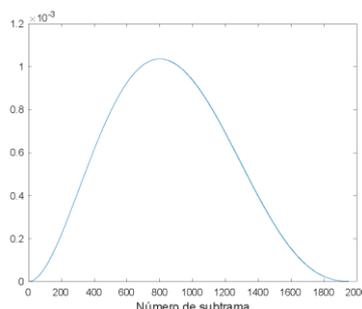
Sin embargo, el tipo de tráfico que más se aproxima a la realidad, según el 3GPP, sería una distribución que no alcanza un régimen permanente. Esto es, la probabilidad de llegadas de tráfico será variable en función del tiempo.

Por este motivo hemos propuesto el siguiente tipo de distribución de tráfico para los nodos de baja prioridad, una distribución Beta(3,4) que se observa en la Figura 25. La distribución Beta(3,4) es la distribución propuesta por el 3GPP para modelar el tráfico para mMTC de tipo transitorio, porque el análisis de sus resultados se asemeja de forma más próxima al tráfico real de mMTC con la tecnología 5G.

Esta distribución tiene una probabilidad de activación que presenta la siguiente forma:

$$p_A(t) = \frac{60 \cdot i^2 (Tdist - i)^3}{Tdist^6}$$

Donde Tdist es el periodo de distribución y es la duración en subtramas durante las que sucederán todas las llegadas, que son instantes de tiempo predefinido en los que el dispositivo tiene la oportunidad de acceder al canal y comunicar con el punto de acceso. Y el parámetro  $i$  irá incrementando con el número de subtrama. Podemos observar, con el paso de aproximadamente el periodo de distribución, Tdist, la probabilidad de activación será 0.



**Figura 25. Distribución Beta(3,4).**

Este tipo de distribución de la Figura 25, a diferencia de la de Bernoulli, tiene más de dos posibles valores, es decir, el número de llegadas que pueden ocurrir durante una subtrama puede tener distintos valores ahora. El número de llegadas que ocurra en cada subtrama va a depender, por un lado, obviamente, de la probabilidad de activación calculada en el instante concreto y, por otro lado, del número total de dispositivos o nodos del sistema. Será la multiplicación de estos dos factores, redondeándolo al entero más próximo, el número de llegadas al sistema en una subtrama.

#### **4.4. Mecanismos de gestión de tráfico: asignación de dinámica de slots y probabilidad de acceso.**

La asignación dinámica de slots en cada subtrama ofrece una gran flexibilidad a la hora de gestionar la congestión del tráfico de ambas prioridades, pero principalmente del tráfico de alta prioridad. Con el algoritmo de nuestro modelo, el número de slots dedicados a cada tipo de tráfico en cada subtrama irá variando.

Para conseguir que este dinamismo sea óptimo, el algoritmo de <sup>[1]</sup> se basa en la observación del tráfico de alta prioridad para poder ofrecerle a éste el máximo caudal posible con el mínimo retardo y la mínima probabilidad de pérdidas posible. Como segundo factor también se realiza una observación del tráfico de baja prioridad para evitar que éste sufra una alta congestión por asignar todo para alta prioridad. Para ello, se marcarán tanto el mínimo como el máximo número de slots que podrán ser asignados al tráfico HPT. Pero el número de slots que se asignan a ambas prioridades también vendrá determinado por el valor de  $\beta$ , es decir, del número de slots por subtrama. Este máximo número de slots asignados a HPT serán todos los slots de la subtrama menos uno o dos slots; y el mínimo serán dos o tres slots, aunque el tráfico requiera menos cantidad de recursos en determinado momento, según las estimaciones.

El algoritmo se basará en las observaciones realizadas en la subtrama anterior para estimar cuál será el número de dispositivos tanto de alta como de baja prioridad que van a querer transmitir un paquete hacia el gNB. A partir de esta estimación, se calculará una probabilidad que será probabilidad de permiso R tanto para HPT como para LPT y será enviada a los distintos dispositivos en forma de mensaje de broadcast (difusión). Con esta probabilidad de permiso, R, los dispositivos que quieran transmitir lo harán y con probabilidad 1-R, esperarán a la siguiente subtrama. Como el algoritmo se despliega en cada subtrama, la probabilidad de permiso cambiará para cada subtrama y por lo tanto cada nodo transmitirá con una probabilidad distinta a la anterior. Para calcular esta probabilidad se tienen en cuenta dos factores:

- En primer lugar, se van a contabilizar para los slots asignados a un tipo de tráfico concreto, el número de slots que vacíos (ningún dispositivo ha elegido este slot para transmitir), el número de slots con éxito (sólo un dispositivo ha elegido transmitir en este slot) y el número de colisiones (slots en los que más de un dispositivo quiere transmitir).
- En segundo lugar, el punto de acceso gNB estima el número de nuevas llegadas, sumado al número de dispositivos que habían retrasado su transmisión por la probabilidad de permiso, más el número de dispositivos que habían colisionado. Todos estos dispositivos se considerarán nodos activos para la siguiente subtrama, ya que tratarán de transmitir en ella si se lo permite la probabilidad de permiso.

Para describir el algoritmo, primero vamos a comentar la nomenclatura utilizada:

- V es el número de slots que tiene una subtrama,  $m_x$  será el número de slots asignados a cada tipo de tráfico,  $x = 1$  para alta prioridad y  $x = 2$  para baja prioridad. Entonces  $m_1 + m_2 = V$ .
- El número de slots estimados para una subtrama será  $w_x$ .  
 $h_x, s_x, c_x$  es el número de slots vacíos, con éxito y con colisión respectivamente para el tipo de tráfico x. Entonces  $h_x + s_x + c_x = m_x$
- $u_{1,max}$  es el número máximo de slots asignados a tráfico HPT, y  $u_{1,min}$  el mínimo.
- Además, el subíndice t indicará que nos referimos a la subtrama actual y t+1 indicará que nos referimos a la siguiente subtrama

Ahora vamos a indicar, por pasos, cómo realizar los cálculos:

- 1- Observamos para la subtrama actual, el número de vacíos, éxitos y colisiones. Con esto y con el número de slots que se habían estimado para dicha subtrama:

$$w_{x,t+1} = \max\left(w_{x,t} + \frac{c_{x,t}}{e - 2} - (h_{x,t} + x_{x,t}), 0\right)$$

La expresión refleja en un término un incremento en el número de slots a medida que hay más colisiones detectadas multiplicado por un factor, y en otro término una disminución en el número de slots estimados debido a los slots vacíos más los éxitos. Además, observamos que  $w_{x,t+1}$  nunca es negativo.

A continuación se estiman las nuevas llegadas para cada tipo de tráfico, que en el estado estacionario podemos decir que se alcanza un equilibrio y que  $\lambda_{x,t} = s_{x,t} \cdot \lambda_{x,t} = s_{x,t}$

- 2- Calculamos el número de slots que se van a asignar para tráfico de HPT en la siguiente subtrama:

$$m_{1,t+1} = \max(u_{1,min}, \min(\lceil w_{1,t+1} \rceil, u_{1,max}))$$

Es decir, el número de slots asignados a alta prioridad será el número de slots estimados (entero superior) siempre que esté por encima del mínimo y por debajo del máximo, en cuyo caso se tomaría el valor extremo. Una vez calculado el número de slots para HPT, sencillamente el resto se asignará a baja prioridad.

$$m_{2,t+1} = V - m_{1,t+1}$$

- 3- Por último, calculamos la probabilidad de permiso que el gNB enviará en mensaje de broadcast a todos los dispositivos, que será  $R_{x,t+1} = \frac{m_{x,t+1}}{w_{x,t+1}}$ , siempre que esté en el rango entre 0 y 1, si es menor que 0 será 0 y si es mayor que 1 será 1.

Una vez tenemos la probabilidad de permiso y el número de slots asignados a cada tipo de tráfico para la subtrama siguiente, ya tenemos el algoritmo recursivo diseñado.

#### 4.5. Preparación y diseño del modelo de simulación

Para validar el modelo que he estudiado en este trabajo, he realizado un programa de simulación programado en C++. Para desarrollar este modelo se ha usado la herramienta de programación y simulación Microsoft Visual C++ 2010, que es una versión de Visual Studio específica para el lenguaje de programación C++. Es un entorno de desarrollo muy completo y profesional. Contiene diversas funcionalidades, para distintos tipos de aplicaciones y para facilitar al máximo la programación, la prueba y la depuración de las aplicaciones que se desarrollan.

La construcción desde cero de un programa de simulación que se ha basado en el diseño de un modelo de gestión de colas para dos tipos de tráfico con llegadas de distinto tipo de distribución. Para llevarlo a cabo hemos implementado las herramientas y algoritmos necesarios que hacen que el modelaje del sistema sea un fiel reflejo de lo que se ocurre realmente. Asimismo, también hemos logrado obtener de manera adicional indicadores que evalúan la capacidad del sistema, KPIs (Key Performace Indicators).

El programa diseñado se encarga de:

- Leer los parámetros de un fichero de texto que recogerá los parámetros que pueden variarse para distintos tipos de simulaciones, como pueden ser número de slots por trama y la tasa de llegadas.
- Simular las colas de los dispositivos de forma independiente para los dispositivos de alta prioridad y para los de baja prioridad. Además, también gestiona las distribuciones de llegada de paquetes a los nodos y la forma de añadirlos a una cola FIFO.
- Realizar el cálculo de la asignación dinámica de slots para cada subtrama, calculando la probabilidad de permiso y el número de slots asignados para cada tipo de tráfico.

En la Figura 26 tenemos una captura de una parte del código, que en concreto se encarga de realizar los cálculos que determinan cuál es el número estimado de dispositivos que van a transmitir en la siguiente subtrama y así calcular en número de slots que se va a asignar a cada prioridad. A continuación, la siguiente captura, Figura 27 muestra el

cálculo de la probabilidad de permiso mediante la cual los nodos decidirán si transmitir o no hacerlo.

```

w1 = w1 + cf * (real)collision_num1 - (real)(holes_num1 + sacc_num1);
w2 = w2 + cf * (real)collision_num2 - (real)(holes_num2 + sacc_num2);

if(w1 < 0)
{ w1 = (real)0 + sacc_num1; }
else
{ w1 = w1 + (real)sacc_num1; }

if(w2 < 0)
{ w2 = (real)0 + sacc_num2; }
else
{ w2 = w2 + (real)sacc_num2; }

// Calculamos el número de slots asignados a LPT y a HPT para la siguiente trama*/
// Aplicamos: m1 = max(u1_min, min(w1,u1_max)
w1_c = (int)ceil(w1);
if(w1_c>u1_max){
    m1 = u1_max;
} else if(w1_c>u1_min){
    m1 = w1_c;
} else {
    m1 = u1_min;
}

//El número de slots asignados a LPT es el resto
m2 = V - m1;

```

**Figura 26. Cálculo de la asignación de slots para la siguiente subtrama.**

```

//Actualizamos la nueva Permission Probability para HPT y para LPT, R y R2 para la siguiente trama
R = ((real)m1)/w1;
if (R > 1.0)
{ R = 1.0;}

R2 = (real)m2/w2;
if (R2 >1.0 )
{ R2 = 1.0; }

//Ir sumando número de slots asignados a cada tipo de tráfico
NS_HPT = NS_HPT + m1 ;
NS_LPT = NS_LPT + m2 ;
// Conviene hacerlo no acumulado, sino en cada trama llamar a una función

```

**Figura 27. Cálculo de la probabilidad de permiso para la siguiente subtrama.**

- Emular por un lado un proceso de llegadas para el régimen de tráfico de alta prioridad y, por otro lado, para el tráfico de baja prioridad. Se hará de manera, por lo tanto, que se empleen funciones y estructuras de datos totalmente independientes. Se han descrito anteriormente distintos escenarios para ambos tipos de prioridades, por lo que la programación de cada uno requiere de la distinción de ambos tipos de manera independiente pero integrados en un mismo espacio temporal (una subtrama de 5G). En la Figura 28 se observa cómo se ha procedido a simular la distribución Beta(3,4) para las llegadas de baja prioridad.

```

pA = 60*pow(Tdist,2)*pow((2000-Tdist),3)/pow(2000,6) ; // Probabilidad de llegadas de LPT
if(pA > 1) { pA = 1; }
else if (pA < 0) { pA = 0; }
else {
arr_n_LPT = round(pA*NS_LPT); // Número de llegadas LPT en la trama actual
//N_ACT = N_ACT - arr_n_LPT ;
insert_LPT(arr_n_LPT) ;
Tdist++;

```

**Figura 28. Llegadas con distribución Beta(3,4).**

- Autocomprobaciones para asegurar que el proceso modelado funciona no sufre bugs o errores internos. En la Figura 29 podemos observar como el número de slots de una trama asignados a cada prioridad, debe de repartirse entre slots vacíos, slots con colisión o slots con éxito.

```

if((holes_num1+collision_num1+sacc_num1) != m1){
printf("holes_num1 + collision_num1 + sacc_num1 != m1 >> ERROR.\n");salir(MYERROR);
}
if((holes_num2+collision_num2+sacc_num2) != m2){
printf("holes_num2 + collision_num2 + sacc_num2 != m2 >> ERROR.\n");salir(MYERROR);
}

```

**Figura 29. Validaciones internas del código.**

- Recogida de resultados de parámetros que midan la validez del modelo como el retardo, probabilidad de pérdidas, etc. y, asimismo, cálculo de otros estadísticos a partir de los resultados obtenidos, como se muestra en la Figura 30, en la que están calculados algunos de los estadísticos necesarios para obtener el desempeño del modelo. Aquí incluimos los indicadores KPI.

```

/*Recopilar estadísticas LPT*/
Thpk_LPT = (real) total_sacc_LPT / (real) NS_LPT;
Thpk_sf_LPT = (real) total_sacc_LPT / (real) LPT_sf;

LPT_D95 = rep_delay_d(LPT_D, LPT_DELAY_DB, LPT_Delay_Dtb) ;
lost_pack_LPT; // total paquetes perdidos
PP_LPT = (real)lost_pack_LPT/(real)TOT_ARR_LPT; // probabilidad de pérdidas
NS_LPT; // número total de slots asignados a LPT
Avg[4] = (real)NS_LPT/(real)TSim; // número medio de slots LPT por trama

KPI_1 = (real)total_collision_LPT/((real)NS_LPT*50) ; // Collision probability (PKI 1)
KPI_2 = (real)total_sacc_LPT/(real)NS_LPT ; // Access success probability (PKI 2)
KPI_3 = (real)total_sacc_LPT/LPT_sf; // Mean value of successfull transmissions
KPI_4; // Delay
KPI_5; // Máximum number of transmissions per slot

```

**Figura 30. Recogida de resultados para tráfico de baja prioridad.**

- Generación de ficheros de resultados y estadísticos para poder ser analizados y también para poder ser importados en otros entornos como ficheros para la realización de gráficos en Matlab más fácilmente interpretables.
- Por último, importación de ficheros de texto con los resultados en Matlab para la representación gráfica de los mismos y facilitar así su análisis y comprensión.

Los KPI que hemos tratado anteriormente son los dictados por el 3GPP para el RACH (Random Access Channel) en <sup>[15]</sup> y son los siguientes:

- 1- Probabilidad de colisión,  $P_C$ , que se representa como la fracción de slots en los que se produce una colisión, respecto del número total de slots asignados al tipo de tráfico en estudio.
- 2- Probabilidad de éxito,  $P_S$ , definida como la fracción de nodos que completan su transmisión con éxito.
- 3- Estadísticas del número de transmisiones exitosas de cada nodo, por subtrama.
- 4- Estadísticas del retardo. Están referidas al tiempo transcurrido entre el instante de llegada del paquete y el instante en el que se transmite con éxito. Para ello, por ejemplo, se obtiene la media del retardo, la función de distribución acumulada, CDF (Cumulative Distribution Function) o algunos percentiles significativos.
- 5- Máximo número de nodos que transmiten en una misma ranura o slot.

#### **4.5.1. Simulación con tráfico de Bernoulli**

El programa de simulación del tráfico está basado, por un lado, en la generación de llegadas de paquetes a cada nodo o dispositivo que forma el conjunto de nodos de un tipo (alta o baja prioridad), que será del orden de algunas decenas de nodos. Este proceso se realiza de forma aleatoria, en función de la probabilidad de activación que se haya fijado. Si un nodo recibe un paquete, tenemos en cuenta su instante de llegada para calcular posteriormente el retardo del paquete hasta que se transmite. El paquete será añadido a la “cola”, que tendrá un tamaño máximo de un paquete. Si un dispositivo recibe un paquete y tiene su cola llena, este paquete se perderá y será anotado como paquete perdido para posteriormente calcular la probabilidad de pérdidas del sistema para ambos tipos de tráfico.

Por otro lado, si el nodo dispone de un paquete para transmitir, lo hará con una probabilidad que será adaptativa de acuerdo al modelo que hemos explicado en el apartado anterior. El dispositivo que transmite, por tanto, un paquete elige uno de los slots disponibles en la subtrama para su tipo de tráfico. Cada nodo que vaya a transmitir hará lo propio, por lo tanto, habrá slots en los que quieran transmitir más de un dispositivo, habrá colisión y ninguno tendrá éxito. Finalmente, si tiene éxito, el paquete será extraído de la cola del dispositivo y se anotará cuál ha sido el tiempo que ha pasado en la cola (retardo) para obtener estadísticas del retardo medio de los paquetes que transmiten con éxito.

#### **4.5.2. Simulación con tráfico Beta(3,4) para LPT**

El diseño de una distribución de tráfico distinta supone un cambio en la gestión de los dispositivos y las colas. Como punto principal, vamos a destacar que en este caso los dispositivos compiten por un permiso de transmisión y una vez consiguen el acceso, transmiten todo su tráfico al punto de acceso a través de otro canal dedicado. Por lo tanto, para esta distribución vamos a tener un número mucho mayor de nodos sólo van a competir por intentar transmitir, hasta que consigan tener éxito, en ese momento se considerará que dejan el sistema. Es decir, un nodo o dispositivo llega al sistema en un instante determinado porque quiere transmitir, intenta conseguir el permiso una serie de veces (este número tendrá un límite el cuál si es alcanzado, el nodo abandonará el sistema y se contabilizará como una pérdida) y una vez ha transmitido, abandona el sistema.

Además, esta distribución presenta un proceso de llegadas transitorio en el cuál llegarán una gran cantidad de dispositivos al sistema de forma simultánea, teniendo un pico de llegadas por encima de la capacidad del sistema, pero disminuirá de forma que se podrá gestionar la transmisión de casi todos ellos, con un retardo aceptable teniendo en cuenta que serán dispositivos con información declarada de baja prioridad.

Para este tipo de distribución, en la simulación tendremos en cuenta el instante de llegada, así como el número de retransmisiones realizadas para intentar transmitir, ya que alcanzado el máximo se declarará como perdido y dejará de intentarlo (en el modelo tomamos como máximo 10 retransmisiones, ya que un número mayor puede causar un retardo inasumible para el servicio). Con ello recogeremos parámetros con los que obtener resultados que nos permitan analizar el desempeño del modelo. Este tipo de distribución se gestiona de manera independiente de la distribución de Bernoulli en la simulación en la que conviven ambas. Tan sólo interaccionan para la asignación dinámica de slots, en las que el tráfico de baja prioridad se queda únicamente con los slots que no necesita el tráfico de alta prioridad.



## 5. Análisis de resultados y estadísticas

En este capítulo vamos a reflejar los resultados que hemos obtenido mediante el uso del programa de simulación para el modelo del proyecto, implementando todas las diferentes variantes explicadas en el capítulo anterior.

### 5.1. Simulación de un tráfico de Bernoulli de alta y baja prioridad

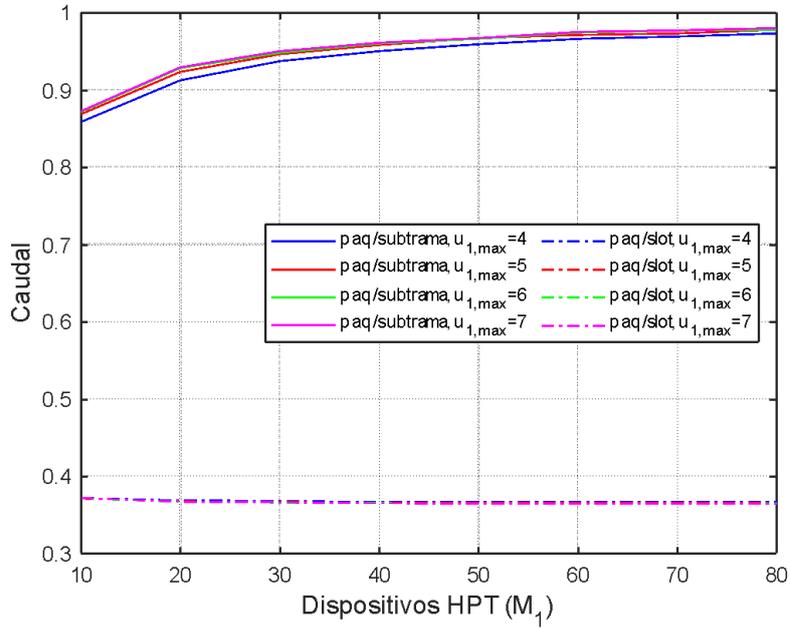
En primer lugar, vamos a expresar los resultados obtenidos mediante la ejecución de la simulación del modelo empleando un tráfico de Bernoulli, con una determinada probabilidad de activación de los nodos ( $a_1$ ), para HPT; y el mismo tipo de tráfico con una probabilidad de activación ( $a_2$ ) para LPT. De la misma forma, el número de equipos de cada tipo será distinto  $N_{HPT}$  y  $N_{LPT}$ .

Estos resultados nos van a permitir realizar una comparativa con los resultados obtenidos a través del modelo analítico de <sup>[1]</sup>. Así, hemos podido realizar una validación exacta de la propia simulación, y también una evaluación del desempeño del modelo presentado para a continuación poder compararlo.

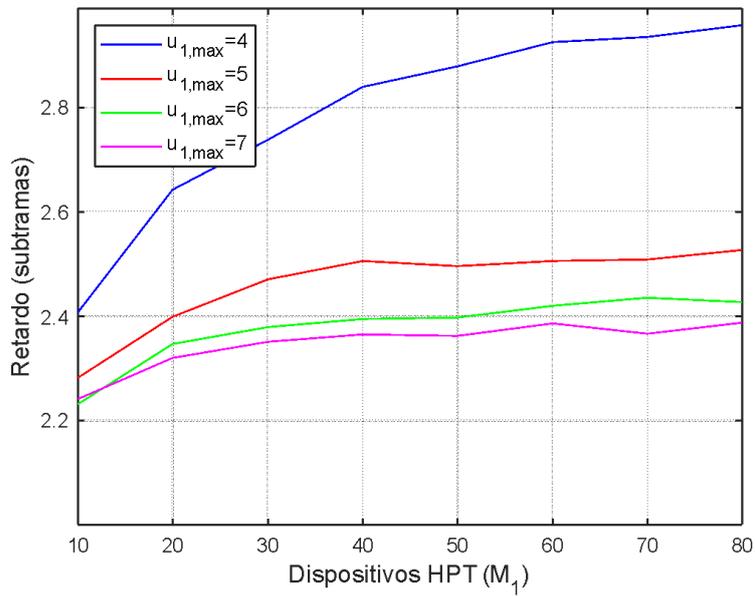
#### 5.1.1. Análisis del tráfico de HPT

En esta primera simulación hemos tomado distintos valores para el tráfico de alta prioridad, para evaluar su caudal, retardo en el acceso y probabilidad de pérdidas. Para ello se asignarán distintos valores a la probabilidad de activación, así como al número total de nodos o dispositivos del sistema. El tráfico de baja prioridad se mantiene constante en un valor que es indiferente, ya que el tráfico de alta prioridad no debe verse afectado por el menos prioritario.

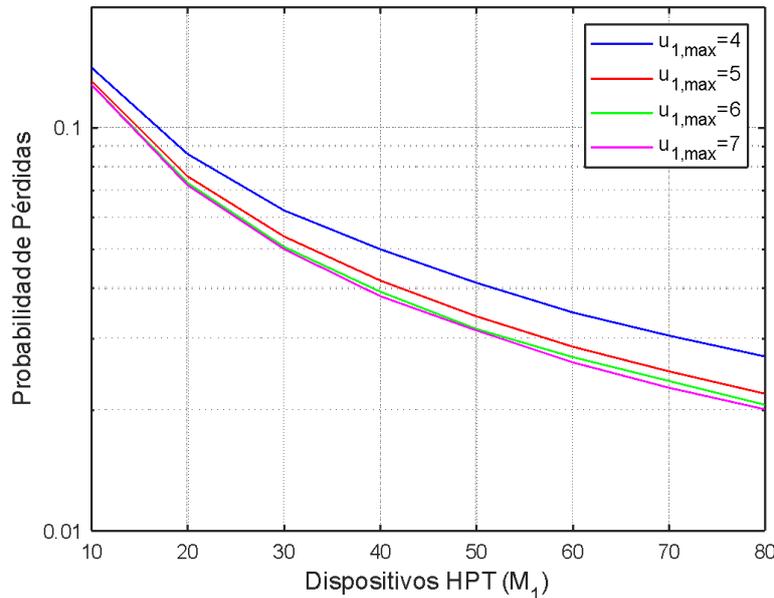
- El **caudal cursado** (Throughput), descrito como caudal a partir de ahora, se ha calculado como la relación de paquetes transmitidos exitosamente sobre el número total de slots asignados a este tipo de tráfico, en este caso HPT. Se expresa como caudal por slot y caudal por subtrama.
- El **retardo** se calculará aplicando una media móvil <sup>[13]</sup>. Esta media sólo representa el retardo de los paquetes que son transmitidos, no se tiene en cuenta los paquetes que se pierden.
- La **probabilidad de pérdidas** se calcula con la diferencia entre paquetes que han llegado menos los paquetes que se han transmitido con éxito, sobre los paquetes que han llegado.



**Figura 31. Caudal HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$ .**



**Figura 32. Retardo HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$ .**



**Figura 33. Probabilidad de pérdidas HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$ .**

Podemos ver claramente en la Figura 31 como un mayor número de dispositivos HPT  $M_1$ , el caudal por subtrama aumenta, pero el caudal por slot no lo hace tan claramente. Esto se debe a que este caudal por slot va a depender del número de slots que se asignan a HPT en cada subtrama, pero por el contrario el caudal por subtrama no. El número slots asignados a alta prioridad aumenta cuando aumenta  $M_1$  debido a que aumentará el número de dispositivos que tienen un paquete para ser transmitido (hay más dispositivos con un paquete esperando para transmitir, ya sea porque ha colisionado anteriormente o porque se ha generado nuevo con probabilidad  $a_1$ ) y esto producirá que se asignen un mayor número de slots a alta prioridad y un mayor caudal cursado. El caudal por slot es algo superior al máximo caudal obtenido en ALOHA-Ranurado,  $1/e \approx 0.368$ , que sería obtenido con un número infinito de población en lugar de con 10,20...80 dispositivos.

En la Figura 32 observamos el retardo en subtramas que sufren los paquetes de los dispositivos HPT para acceder al servicio, donde se puede observar como el aumento de población implica que se necesite un mayor número de subtramas para conformar el tráfico y además el retardo será superior cuando  $u_{1,max}$  sea menor, debido a que habrá una pequeña disminución en el número de slots dedicados a HPT. En la Figura 33, con la probabilidad de pérdidas pasa exactamente lo mismo, al haber un mayor número de paquetes esperando en la cola para ser transmitidos de nodos HPT, la probabilidad de que exista una colisión es mayor y por lo tanto la probabilidad de pérdidas también.

### 5.1.2. Análisis del tráfico de LPT

Para la segunda simulación mantenemos un tráfico ofrecido de alta prioridad constante y se varía el tráfico que se ofrece a los dispositivos con menor preferencia. Esto nos permitirá observar, en este caso, el desempeño de los nodos de baja prioridad en términos de caudal por slot y por subtrama, retardo en el acceso y probabilidad de pérdidas.

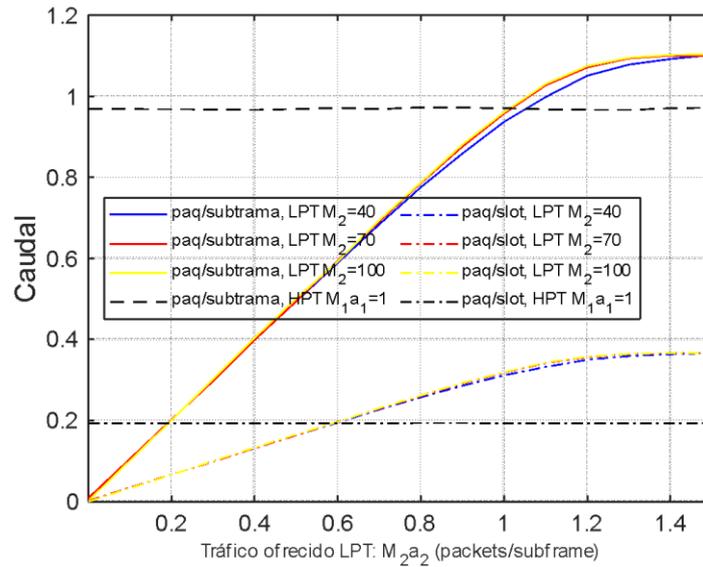


Figura 34. Caudal LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un caudal ofrecido  $M_1 \cdot a_1 = 1$  y distintos valores de caudal ofrecido LPT  $M_2 \cdot a_2$ .

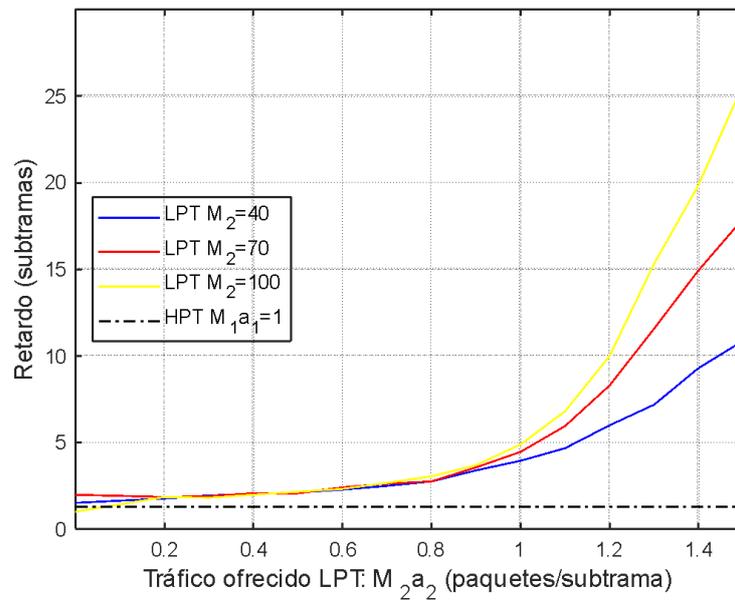
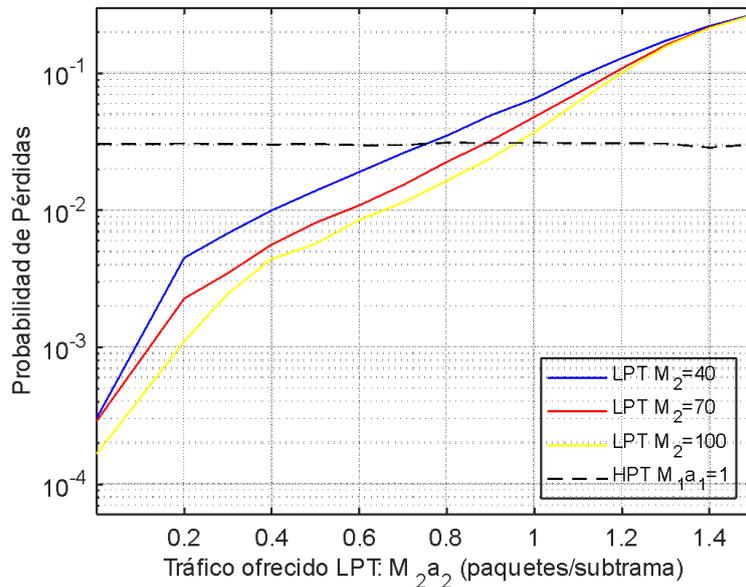


Figura 35. Retardo LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un caudal ofrecido  $M_1 \cdot a_1 = 1$  y distintos valores de caudal ofrecido LPT  $M_2 \cdot a_2$ .



**Figura 36. Probabilidad de pérdidas LPT con  $u_{1,max} = 7$ ,  $u_{1,min} = 5$ , un caudal ofrecido  $M_1 \cdot a_1 = 1$  y distintos valores de caudal ofrecido LPT  $M_2 \cdot a_2$ .**

En este caso nos encontramos con un caudal por subtrama y por slot en la Figura 34 que aumenta de manera lineal hasta que se alcanza el límite del caudal. Este límite se encuentra determinado por el número medio de slots asignados a baja prioridad, sobre el número  $e$ , esto se debe a que nos encontramos ante un modelo de colisiones basado en los principios de ALOHA. El número medio de slots asignados a LPT obtenido en esta simulación es 2.8, por lo que el límite se encuentra aproximadamente en 1.1 paquetes por subtrama, sea cual sea el tráfico ofrecido. Con el caudal por slot ocurre algo similar.

En la Figura 35, a medida que tráfico ofrecido crece, el número de colisiones aumentará y el retardo aumentará también. Este mismo razonamiento nos hace deducir por qué aumenta la probabilidad de pérdidas en la con el tráfico ofrecido ( $M_2 \cdot a_2$ ) en la Figura 36.

### 5.1.3. Conclusiones

Con estas primeras simulaciones, además de haber ayudado a comprender de manera gráfica el desempeño y funcionamiento del sistema, conseguimos el objetivo de demostrar la validez de la simulación, ya que los resultados obtenidos son los esperados para los distintos parámetros asignados al sistema. En <sup>[1]</sup> observamos unos resultados similares a los obtenidos con el programa de simulación utilizado para el presente proyecto.

## **5.2. Simulación de un tráfico de Bernoulli de alta prioridad y Beta(3,4) de baja prioridad**

Una de las principales novedades que incorpora el presente trabajo respecto a <sup>[1]</sup> se encuentra en la introducción del nuevo patrón de tráfico, ya descrito anteriormente en el apartado 4.2.3 . para nodos de baja prioridad y manteniendo un tráfico constante de Bernoulli para los dispositivos con mayor prioridad. De esta forma, presentaremos el desempeño del modelo tratado, pero en un nuevo escenario, con condiciones totalmente diferentes, teniendo en cuenta las enormes diferencias que existen entre una distribución de Bernoulli y una distribución Beta(3,4).

Vamos a presentar los resultados análogamente a como los hemos presentado para el apartado anterior, distintos tipos de indicadores que nos marcarán el desempeño en cada escenario, para distintas situaciones de la realidad. Estos indicadores serán, entre otros, los KPIs que se describieron en el apartado 4.5.

Para poder hacer un análisis en profundidad sobre el comportamiento del sistema, vamos a aplicar distintas cargas de tráfico de alta prioridad, para ver cómo responde el modelo.

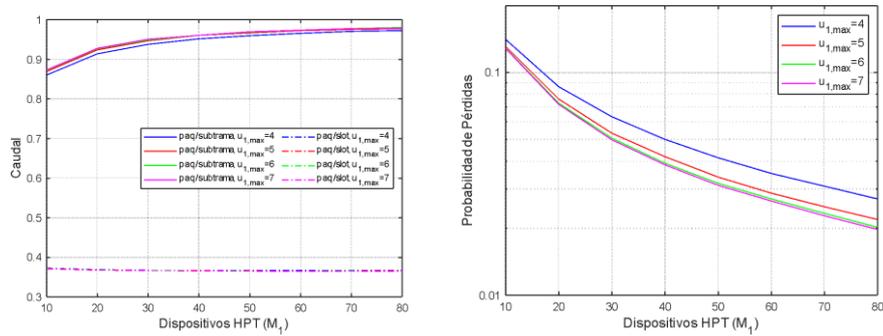
El tráfico de baja prioridad será un tráfico Beta(3,4) y que va a presentar las siguientes características:

- 30000 dispositivos de usuario (UEs). Este es el número total de dispositivos que “llegarán” al sistema. Cada dispositivo será registrado como una llegada.
- Tdist, el tiempo durante el que se producen las llegadas es de 2000 subtramas temporales.

Se realiza una simulación con unos parámetros del tráfico de alta prioridad iguales a los de la simulación del apartado 5.1.1. Con esto podemos estudiar dos aspectos del modelo. Por una parte, demostramos que el tráfico de alta prioridad no va a depender en ninguno de los casos del caudal ofrecido a los nodos de baja prioridad, es decir, el comportamiento de los parámetros del tráfico HPT será exactamente el mismo que en las simulaciones del apartado mencionado, como podemos ver en la Figura 37. Por otra parte, vamos a realizar un análisis del tráfico de baja prioridad, teniendo en cuenta que el tráfico Beta(3,4) que se le está ofreciendo a los distintos nodos es muy diferente a un tráfico constante como el de Bernoulli.

### **5.2.1. Análisis del tráfico de HPT**

Tras analizar los resultados obtenidos con los distintos valores de tráfico de HPT (distintos valores de  $M_1$ , de  $a_1$  y de  $u_{1,max}$ ) de forma similar al apartado anterior, comprobamos que, efectivamente, el sistema que regula el tráfico de alta prioridad y la asignación de recursos al mismo funciona de la misma forma sea cual sea el tráfico ofrecido a los nodos de baja prioridad.

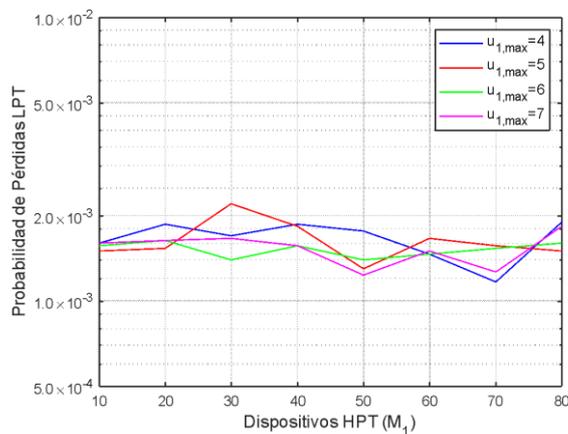


**Figura 37. Caudal y Probabilidad de Pérdidas HPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$ , un tráfico  $M_1 \cdot a_1 = 1$ .**

### 5.2.2. Análisis del tráfico de LPT

A continuación, vamos a presentar los resultados obtenidos tras la obtención y procesado de los datos del tráfico de baja prioridad, analizando con todo detalle diferentes parámetros relevantes que nos informarán sobre la idoneidad del uso de este modelo para un tráfico propuesto por la importante organización 3GPP.

Comenzaremos representando la probabilidad de pérdidas (Figura 38) para distintos parámetros de los nodos de alta prioridad, que será el que determine los recursos que restan para el tráfico de baja prioridad.



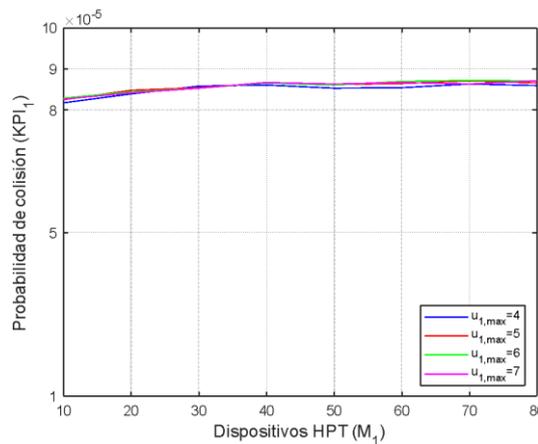
**Figura 38. Probabilidad de pérdidas LPT con distintos valores de  $u_{1,max}$ ,  $u_{1,min} = 1$  y un caudal ofrecido  $M_1 \cdot a_1 = 1$  con Beta(3,4).**

Además, podemos señalar otro parámetro obtenido en la simulación, que el número medio de slots por subtrama que se asignan al tráfico de baja prioridad, que se encuentra en 5.45 slots, de los 8 slots totales en una subtrama; el resto se asignan al tráfico de alta prioridad. Es decir, aunque el tráfico de alta prioridad sea de  $M_1 \cdot a_1 = 1$ , se reserva un número de slots superior, debido a que

el modelo nunca ajusta al máximo el número de slots para alta prioridad, sino que concede recursos de sobra para que nunca sufra congestión.

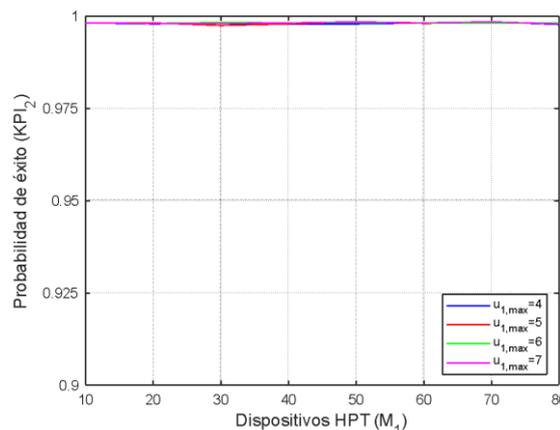
Prosiguiendo con los resultados de baja prioridad, ahora mostramos los indicadores clave, **KPIs**, que hemos explicado anteriormente, y son:

- La Probabilidad de colisión que vemos en la Figura 39. La fracción de slots en los que se produce una colisión respecto del total de slots que son asignados al tráfico de baja prioridad. Se observa que esta probabilidad es muy baja, es decir, se consigue que el número de colisiones sea casi insignificante.



**Figura 39. KPI<sub>1</sub>: Probabilidad de colisión LPT con Beta(3,4).**

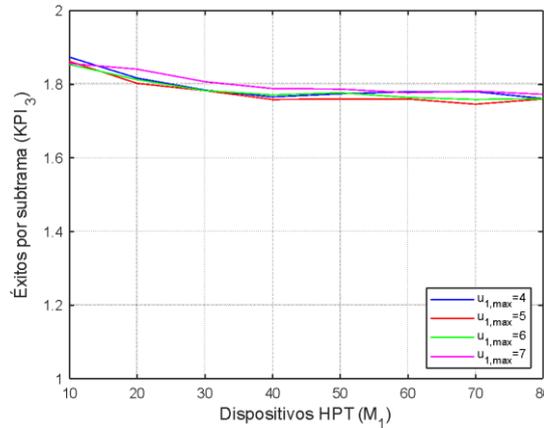
- La probabilidad de éxito, es decir, el número de nodos que logran transmitir al punto de acceso, respecto al número de nodos totales (30000). En la Figura 40 vemos que se aproxima mucho a 1, está por encima de 0.99, casi todos los que lo intentan, consiguen transmitir con éxito. Cabe remarcar que el modelo diseñado permite un número de máximo de 10 retransmisiones hasta que el nodo desiste en su intento de transmitir y se considera una pérdida más. Esto supone un gran punto a favor del modelo.



**Figura 40. KPI<sub>2</sub>: Probabilidad de éxito LPT con Beta(3,4).**

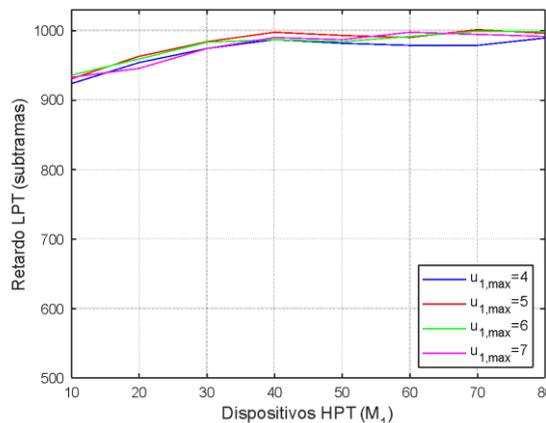
- Número medio de éxitos en una subtrama (8 slots o ranuras). A medida que aumenta el tamaño de la población de alta prioridad, el número de slots dedicados a esta aumenta y

por esta razón el caudal cursado de baja prioridad disminuye. Nos encontramos con un caudal cercano a 2 paquetes por subtrama, teniendo en cuenta las tramas desde que se inicia el tráfico LPT hasta que no queda ningún nodo por transmitir. Este indicador también muestra un buen funcionamiento del sistema.



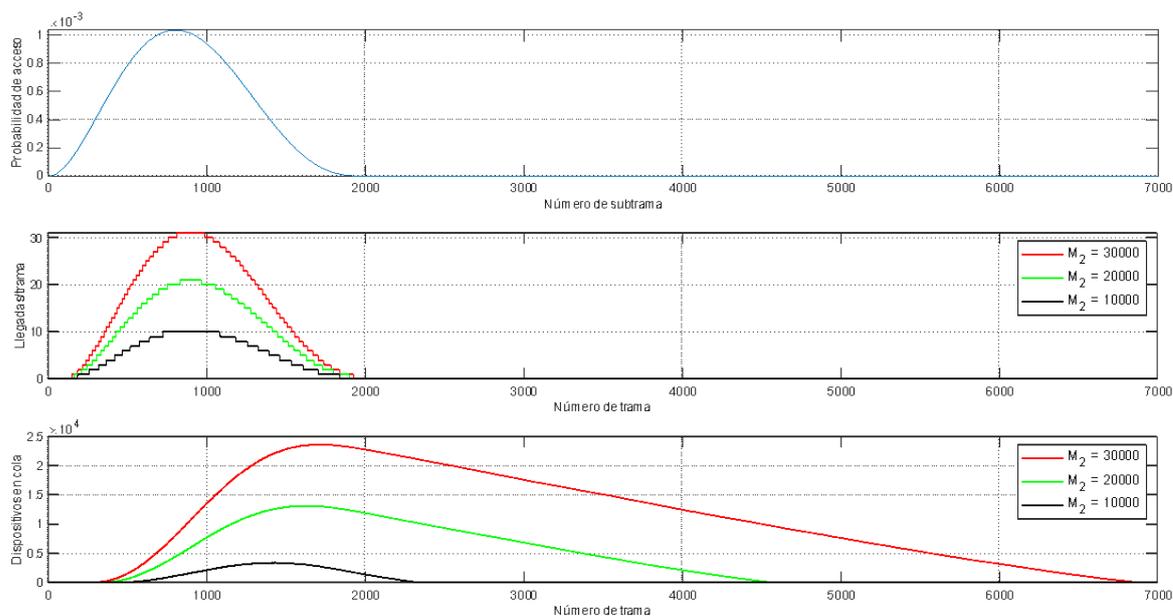
**Figura 41. KPI<sub>3</sub>: Número de transmisiones exitosas LPT por subtrama con Beta(3,4).**

- El retardo medio experimentado por los paquetes de esta prioridad es relativamente alto (Figura 42) comparado con lo experimentado para los nodos de alta prioridad, pero aun así, debemos tener en cuenta que una subtrama tiene una duración de 1ms y nos encontramos, por lo tanto, con un retardo que no alcanza 1s en media. Además del retardo medio un indicador del retardo que proporciona información importante son los percentiles. Uno de los percentiles más importante es el percentil 95, que indica el retardo por debajo del cuál se pueden agrupar el 95% de las transmisiones, que presenta de media 1873.42 subtramas (el 95% de las transmisiones LPT presentan un retardo menor de 1873.42 subtramas). Estos datos sobre el retardo en el acceso son asumibles para un tipo de tráfico de datos de baja prioridad, que deberá ser información con una menor sensibilidad al retardo, como por ejemplo la transmisión de una señal modo alarma que envía un dispositivo para indicar que uno de sus sensores de luz ha bajado por debajo del umbral y se debe encender una bombilla.



**Figura 42. KPI<sub>4</sub>: Retardo medio LPT en subtramas con Beta(3,4).**

- El último indicador es el número máximo de llegadas en la misma subtrama. Éste nos proporciona información sobre la distribución elegida Beta(3,4), más que del tratamiento que hace el sistema. Pero sí que nos dará información sobre la saturación a la que sometemos al sistema en determinado momento. En la Figura 43, podemos observar tres gráficas: la probabilidad de activación o de acceso según la distribución Beta(3,4), el número de llegadas en cada subtrama y el número de dispositivos en el sistema. Hemos representado diferentes curvas, para tres valores de dispositivos LPT que van a intentar acceder al sistema,  $M_2$ .



**Figura 43. KPI<sub>5</sub>: Evolución del número de dispositivos LPT en el sistema para  $M_2 = 10000$ , 20000, 30000 con Beta(3,4).**

En la segunda gráfica se observan las llegadas al sistema. Estas llegadas se producen en forma de oleada durante un corto periodo de tiempo, alcanzando para  $M_2 = 30000$  un máximo de 31 llegadas en la misma trama, para  $M_2 = 20000$  un máximo de 21 llegadas y para  $M_2 = 30000$  un máximo de 10 llegadas. Si tenemos en cuenta que el número de slots en una subtrama es de 8 y que el número medio de slots asignados a alta prioridad es de 2.768, entonces tenemos que el número medio de slots para baja prioridad será de 5.232. Esto produce que el número de dispositivos en el sistema aumente de forma inevitable (tercera gráfica), ya llegan más de los que se pueden servir, hasta que el propio sistema comienza a servir más nodos de los que llegan y la curva toma un sentido descendente. Para  $M_2 = 30000$ , el número de dispositivos en el sistema alcanza un máximo de 23585 dispositivos en el sistema, un número elevadísimo de nodos que pueden transmitir en un determinado momento y un máximo de 3356 para  $M_2 = 10000$ . Nótese que, aunque el número de dispositivos es muy grande, el sistema lo gestiona obteniendo unos resultados expresados en las figuras anteriores, un 99% de los dispositivos transmiten con éxito, un retardo medio de 1000 tramas o una probabilidad de colisión inferior a  $10^{-4}$ . Esto es debido al algoritmo que sigue el modelo en el cuál adapta la probabilidad de permiso a

los nodos que hay en el sistema y a la saturación que está sufriendo el mismo (colisiones) para que el sistema no colapse y se produzcan pérdidas masivas.

### **5.2.3. Conclusiones**

En estas últimas simulaciones expuestas justificamos el buen funcionamiento del sistema con un tráfico distinto al tráfico analizado en <sup>[1]</sup>. Además de ello, la distribución propuesta es una distribución que somete al sistema a una gran carga de trabajo durante un tiempo concreto provocando un régimen transitorio en el cual se puede examinar la adaptabilidad del sistema.



## 6. Conclusiones y líneas futuras

El principal objetivo del trabajo era el de poder llevar a cabo un estudio de un modelo de tráfico para 5G mMTC en el que un elevado número de dispositivos de dos prioridades conectados al mismo punto de acceden al servicio y consiguen unos objetivos de retardo, caudal y pérdidas. Además, poner a prueba al sistema estimulándolo con distintos tipos de tráfico.

En primer lugar, en el proyecto se ha logrado estudiar cuáles son las necesidades tecnológicas que exige la nueva forma de conexión que se está desarrollando y se prevé un enorme crecimiento en los próximos años. La cantidad de dispositivos conectados a Internet con exigencias de baja transferencia de datos implica inversiones tecnológicas tanto en infraestructura como en nuevos modelos de gestión del tráfico.

Y, en segundo lugar, mediante el programa de simulación se ha conseguido probar el desempeño del modelo, demostrar su buen hacer analizando cada uno de los parámetros principales.

Teniendo en cuenta que el modelo de referencia es un modelo de gestión de tráfico de un artículo presentado ante un organismo de estandarización internacional como el IEEE, se ha conseguido realizar aportaciones al mismo.

Podemos señalar como principal línea futura el desarrollo de un modelo en el que la adaptabilidad de la asignación de slots, que ya se realiza de forma dinámica, para que se tenga en cuenta la congestión (colisiones, éxitos y vacíos) de varias subtramas anteriores, por ejemplo, 10 tramas. Esto proporcionaría al sistema una mayor adaptabilidad y también un mayor margen al punto de acceso para comunicar a los dispositivos los nuevos parámetros en cada subtrama (probabilidad de permiso).

Por último, también cabe destacar que el modelo de simulación está diseñado para la inserción de colas en los dispositivos, de manera que pueda estudiarse este nuevo comportamiento en un futuro trabajo. Esta nueva línea de trabajo nos deriva obligatoriamente a calcular el tamaño óptimo de las cosas en función del caudal de tráfico que vaya a recibir el dispositivo, de manera que se deberán obtener diferentes resultados y serán comparados con el modelo actual en el que no existen colas.



## 7. Referencias bibliográficas

- [1] IEEE Transactions of Communications Vol. 69 (2021, Mayo). Priority Enabled Grant-Free Access With Dynamic Slot Allocation for Heterogeneous mMTC Traffic in 5G NR Networks. <https://ieeexplore.ieee.org/document/9335255>
- [2] Instituto ITACA, UPV (2019, Enero 23). Performance evaluation of framed slotted ALOHA with reservation packets and successive interference cancellation for M2M networks. <http://personales.upv.es/~jmartine/Publications/CN2019.pdf>
- [3] SDx Central (2019, Junio 16). What is the Internet of Things. <https://www.sdxcentral.com/5g/iot/definitions/the-internet-of-things-explained/>
- [4] GSMA (2017, Septiembre). 5G, Internet de las cosas (IoT) y Accesorios tecnológicos. [https://www.gsma.com/latinamerica/wp-content/uploads/2019/04/gsma\\_2017\\_5g\\_iot\\_wearable\\_web\\_ES\\_definitivo.pdf](https://www.gsma.com/latinamerica/wp-content/uploads/2019/04/gsma_2017_5g_iot_wearable_web_ES_definitivo.pdf)
- [5] La Vanguardia (2020, Julio 11). IoT y 5G, dos ritmos de implantación. <https://www.lavanguardia.com/tecnologia/20200711/482206308003/iot-5g-dos-ritmos-implantacion.html>
- [6] SDx Central (2019, Diciembre 17) 5G and IoT Standards. <https://www.sdxcentral.com/5g/iot/definitions/5g-iot-standards/>
- [7] IEEE Internet of Things Journal, Vol. 6 (2019 Junio). "Exploring the Performance Boundaries of NB-IoT". <https://ieeexplore.ieee.org/document/8666729>
- [8] Universitat Oberta de Catalunya (2019, Junio). Estudio en detalle de NB-IoT. Comparación con otras tecnologías LPWAN considerando diferentes patrones de tráfico. <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/94946/6/jsaizmiTFM0619memoria.pdf>
- [9] IEEE Communications Standards Magazine (2019, Diciembre). Towards the standarization of Grant-Free Operation and the associated NOMA strategies in 3GPP. <https://ieeexplore.ieee.org/document/9031550>
- [10] Lanner (2019, Septiembre 25). ¿Qué es un Gateway IoT? <https://www.lanner-america.com/es/blog-es/que-es-un-gateway-iot/>
- [11] IEEE Communications Society (2020, Noviembre 15). Massive Machine-Type Communications for IoT. <https://www.comsoc.org/publications/magazines/ieee-wireless-communications/cfp/massive-machine-type-communications-iot>
- [12] IEEE Communications Surveys and Tutorials Vol. 22 (2020). Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey. <https://ieeexplore.ieee.org/document/9097306>

[13] Investopedia (2021, Abril). Moving Average (Media Móvil). <https://www.investopedia.com/terms/m/movingaverage.asp#:~:text=In%20statistics%2C%20a%20moving%20average,commonly%20used%20in%20technical%20analysis>.

[14] IEEE Global Communications Conference (2015, Diciembre). Performance Analysis of RACH Procedure with Beta Traffic-Activated Machine-Type-Communication. <https://ieeexplore.ieee.org/document/7417095>

[15] 3GPP(2011, Septiembre 16). Study on RAN improvements for Machine-Type Communications.