UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

TESIS DOCTORAL

**CONTRIBUCIÓN AL DISEÑO DE SERVICIOS MULTIMEDIA EN
ARQUITECTURAS DE RED MÓVIL BASADAS EN FUNCIONES DE
RED VIRTUALIZADAS**

CARLOS MARIANO LENTISCO SÁNCHEZ
INGENIERO DE TELECOMUNICACIÓN

2019

Departamento de Ingeniería de Sistemas Telemáticos

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

UNIVERSIDAD POLITÉCNICA DE MADRID

TESIS DOCTORAL

**CONTRIBUCIÓN AL DISEÑO DE SERVICIOS
MULTIMEDIA EN ARQUITECTURAS DE RED MÓVIL
BASADAS EN FUNCIONES DE RED VIRTUALIZADAS**

Autor

Carlos Mariano Lentisco Sánchez

Ingeniero de Telecomunicación

Director

Luis Bellido Triana

Doctor Ingeniero de Telecomunicación

Madrid, Mayo 2019

POLITÉCNICA

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día 9 de mayo de 2019.

Título: Contribución al diseño de servicios multimedia en arquitecturas de red móvil basadas en funciones de red virtualizadas

Autor: Carlos Mariano Lentisco Sánchez

Director: Luis Bellido Triana

TRIBUNAL CALIFICADOR

Presidente: D. Juan Quemada Vives

Catedrático de Universidad
Universidad Politécnica de Madrid

Vocales: D. Alfonso Sánchez-Macian Pérez

Profesor Titular de Universidad
Universidad Antonio de Nebrija

D. Francisco Valera Pintor

Profesor Titular de Universidad
Universidad Carlos III de Madrid

Dña. Jasone Astorga Burgo

Profesora Titular de Universidad
Universidad del País Vasco

Secretario: Dña. Ana Belén García Hernando

Profesora Titular de Universidad
Universidad Politécnica de Madrid

Realizando el acto de defensa y lectura de la Tesis Doctoral en Madrid, el día 9 de mayo de 2019.

CALIFICACIÓN:.....

El Presidente

Los Vocales

El Secretario

Se lo dedico a Papá, a Marijose, a la Nuria y a todos mi amigos

Agradecimientos

Una llamada lo empezó todo, como si de una novela se tratara. Al otro lado estaba Encarna Pastor. Yo no la conocía porque había estudiado Ingeniería de Telecomunicación en la Universidad Carlos III de Madrid. Me interesé por un Máster que imparte el Departamento de Ingeniería de Sistemas Telemáticos, y de entre todos los profesores que figuran en la página web del departamento, la seleccioné a ella por azar. Terminé haciendo esta tesis doctoral y siendo profesor ayudante de este departamento.

No es la primera vez que agradezco a Luis, David y Encarna la confianza que me dieron en aquellos momentos, y que todavía hoy depositan en mí, pero que duda cabe de que merecen especial atención en estos agradecimientos. En especial Luis, que ha dedicado mucho trabajo y esfuerzo en mi formación y en la tutela de esta tesis. Mi admiración y confianza en él me ha permitido estar siempre atento, buscando la excelencia desde la alegría. Gracias a él esta tesis es lo que es y no otra cosa. Gracias por todo Luis.

También querría expresar mi sentido agradecimiento a todos mis compañeros de trabajo y a todo el personal del DIT: profesores, investigadores, personal técnico y de secretaría. Todos juntos hacen de este, el gran departamento que es.

Finalmente, quiero dar las gracias a mi madre y a mi hermano por estar siempre ahí. Gracias por quererme y aceptarme tal y como soy. Gracias por acompañarme en estos años tan llenos de sueños y pesadillas.

Ah! y gracias a la música por conectar cada palabra, cada acento y cada fórmula de esta tesis. Para mí esta tesis es otra más de mis canciones, las dos comienzan igual, con un folio en blanco.

Resumen

La demanda de servicios multimedia en las redes móviles ha experimentado un gran crecimiento en los últimos años. Sin embargo, la retransmisión en directo de eventos sigue planteando diversos desafíos cuando estos servicios se soportan sobre las arquitecturas de red móvil tradicionales.

Los organismos de estandarización han definido protocolos y mecanismos de streaming que dan soporte a los servicios multimedia en las redes móviles, pero el estado del arte muestra la necesidad de proponer soluciones orientadas a reducir el retardo de las retransmisiones en directo y garantizar la continuidad de la reproducción multimedia.

En la tesis doctoral se abordan dos enfoques para diseñar los mecanismos de streaming. El primero, consiste en desplegar las soluciones de mejora en los dispositivos finales de la red, es decir, en los servidores y los dispositivos móviles. El segundo, consiste en adaptar la transmisión del streaming multimedia mediante la actuación de elementos intermedios de la red, como por ejemplo, servidores proxy. Los mecanismos de streaming propuestos usando estos enfoques se han evaluado mediante modelos analíticos y de simulación.

El marco de las nuevas arquitecturas de red móvil, basadas en tecnologías de virtualización, proporciona una flexibilidad que facilita la adopción de los mecanismos propuestos. Las principales ventajas de estas nuevas arquitecturas se obtienen gracias a la combinación de soluciones basadas en la virtualización de las funciones de red y las redes definidas por software, aumentando la programabilidad de las redes y flexibilizando el despliegue de nuevos servicios. En la tesis doctoral se analiza la integración de los servicios de difusión de contenidos multimedia y las mejoras definidas en la tesis para los mecanismos de streaming en las nuevas arquitecturas de red móvil basadas en funciones de red virtualizadas.

Las propuestas de integración realizadas, junto con los resultados obtenidos de la evaluación mediante modelos analíticos y de simulación de los mecanismos de streaming propuestos, muestran cómo es posible reducir la latencia del servicio garantizando la continuidad de la reproducción multimedia en las nuevas arquitecturas de red móvil.

Abstract

The demand for multimedia services in mobile networks has been rapidly increasing over the past years, which has led to a significative Internet traffic growth. Broadcasting live events over current mobile network architectures can reduce bandwidth consumption but poses other challenges in terms of quality of service.

Standardization bodies have defined protocols and streaming mechanisms giving support to multimedia broadcast services in mobile networks. However, the state of the art shows the need to propose solutions aimed at providing a seamless multimedia playback with a reduced service latency. This doctoral thesis addresses this problem by proposing new streaming mechanisms that can improve these key aspects of the service.

Two different approaches are considered when designing the streaming mechanisms proposed in this thesis. The first one is based on implementing the new solutions in end devices such as servers and mobile devices. The second one is based on network assisted streaming using intermediate network nodes, such as proxy servers, to support the multimedia streaming adaptation. The streaming mechanisms proposed following these approaches have been evaluated by using analytical and simulation models.

The upcoming mobile network architectures, based on virtualization technologies, provide the flexibility needed for adopting the proposed streaming mechanisms. The main advantages of these architectures rely on the combination of network function virtualization and software-defined networking, which increases the programmability of networks and facilitates the deployment of new services. In this doctoral thesis, we analyse how both multimedia broadcast services and the new streaming solutions previously proposed can be integrated with the new mobile network architectures based on virtual network functions.

Our proposals for this integration, together with the results obtained from the evaluation of the streaming mechanisms using analytical and simulation models, show how it is possible to reduce service latency while providing a seamless playback in the context of the upcoming mobile network architectures.

Contenido

Capítulo 1: INTRODUCCIÓN Y OBJETIVOS	1
1.1 CONTEXTO Y MOTIVACIÓN	1
1.2 OBJETIVOS	3
1.3 ESTRUCTURA DE LA MEMORIA	4
Capítulo 2: ESTADO DEL ARTE	7
2.1 SERVICIOS DE STREAMING MULTIMEDIA EN REDES MÓVILES	8
2.1.1 Servicios de streaming adaptativo sobre HTTP	8
2.1.2 Algoritmos de adaptación para streaming multimedia basado en DASH	10
2.1.3 Adaptación asistida por la red	11
2.1.4 Servicios de streaming 3GP-DASH sobre eMBMS	13
2.1.5 Problema de investigación: reducción del retardo y continuidad de la reproducción en los servicios de difusión de contenidos	16
2.2 EVALUACIÓN DE LA CALIDAD DE LA EXPERIENCIA	18
2.2.1 Métricas objetivas para evaluar la QoE de los servicios de streaming adaptativo sobre HTTP	19
2.2.2 Problema de investigación: modelo objetivo para evaluar la QoE de los servicios de streaming 3GP-DASH sobre eMBMS	20
2.3 ARQUITECTURA DE LA RED LTE PARA EL SOPORTE DE LOS SERVICIOS DE DIFUSIÓN DE CONTENIDOS	22
2.3.1 Arquitectura de la red móvil LTE	22
2.3.2 Serving Gateway (S-GW)	23
2.3.3 Packet Data Network Gateway (P-GW)	23
2.3.4 Mobility Management Entity (MME)	24
2.3.5 Multimedia Broadcast Multicast Service Gateway (MBMS GW)	25
2.3.6 Broadcast-Multicast Service Centre (BM-SC)	25
2.3.7 Multi-cell/Multicast Coordination Entity (MCE)	26
2.3.8 Definición de los servicios de streaming 3GP-DASH sobre eMBMS	27
2.4 ARQUITECTURAS DE RED MÓVIL BASADAS EN LA VIRTUALIZACIÓN DE LAS FUNCIONES DE RED	29
2.4.1 Introducción a las redes definidas por software y la virtualización de las funciones de red	29
2.4.2 Arquitecturas de red móvil basadas en NFV y SDN	33
2.4.3 Problema de investigación: evolución de la arquitectura de eMBMS combinando NFV y SDN	35

2.5	CONCLUSIONES	36
Capítulo 3:	ANÁLISIS DE LOS SERVICIOS DE STREAMING MULTIMEDIA SOBRE REDES MÓVILES LTE	37
3.1	DISEÑO DE UN MODELO DE SIMULACIÓN Y UN MODELO ANALÍTICO	38
3.2	ANÁLISIS DEL SERVICIO A PARTIR DEL MODELO DE SIMULACIÓN Y EL MODELO ANALÍTICO	41
3.2.1	Relación entre la tasa de codificación AL-FEC y la cobertura del servicio . .	41
3.2.2	Relación entre la tasa de datos y la cobertura del servicio	42
3.2.3	Análisis comparativo de los códigos Raptor y RaptorQ	44
3.3	ESCENARIOS VIRTUALES BASADOS EN SOFTWARE DE CÓDIGO ABIERTO	46
3.3.1	Terminal móvil	47
3.3.2	Servidor de contenidos	47
3.3.3	Conformador de tráfico	47
3.3.4	Creación de escenarios con VNX	48
3.4	USO DE LA PLATAFORMA VIRTUAL PARA ANALIZAR LOS SERVICIOS MULTIMEDIA A PARTIR DE MEDICIONES EXTREMO A EXTREMO	50
3.5	CONCLUSIONES	52
Capítulo 4:	REDUCCIÓN DE LA LATENCIA EN LOS SERVICIOS DE DIFUSIÓN DE CONTENIDOS MULTIMEDIA	53
4.1	ANÁLISIS DE RETARDO	54
4.1.1	Generación de los segmentos multimedia	54
4.1.2	Transmisión de los segmentos a la red del operador	55
4.1.3	Protección de los segmentos multimedia mediante la aplicación de técnicas AL-FEC	55
4.1.4	Transporte de los segmentos multimedia utilizando el protocolo FLUTE . . .	55
4.1.5	Protección frente a variaciones de retardo	56
4.1.6	Recuperación unicast de los segmentos multimedia	57
4.1.7	Retardo inicial de buffering	58
4.1.8	Tiempo límite de reproducción	58
4.2	MÉTODO PARA REDUCIR LA LATENCIA Y EVITAR LAS PARADAS DEL SERVICIO . .	59
4.3	DIMENSIONAMIENTO DEL BUFFER PARA UN DESPLIEGUE ESPECÍFICO DE LA RED MÓVIL	61
4.4	STREAMING MULTIMEDIA DE BAJO RETARDO	63
4.5	EVALUACIÓN DE LA CALIDAD MULTIMEDIA	66
4.6	CONCLUSIONES	68
Capítulo 5:	ANÁLISIS DE LOS SERVICIOS DE DIFUSIÓN DE CONTENIDOS DESDE LA PERSPECTIVA DE LA QOE	70
5.1	HACÍA LA ESTANDARIZACIÓN DE UN NUEVO MODELO PARA MEDIR LA CALIDAD DE LOS SERVICIOS DE STREAMING MULTIMEDIA	71

5.2	ANÁLISIS DEL SERVICIO DE STREAMING 3GP-DASH SOBRE EMBMS UTILIZANDO EL MODELO U-V MOS	72
5.2.1	Relación entre la tasa de datos del servicio y las retransmisiones HTTP	73
5.2.2	Relación entre el retardo inicial de buffering, el bitrate de la representación eMBMS y el tiempo de parada del servicio	75
5.2.3	Definición de los parámetros de servicio a partir de la evaluación de la QoE .	76
5.3	CONCLUSIONES	79
Capítulo 6:	DIFUSIÓN DE CONTENIDOS MULTIMEDIA ASISTIDA POR LA RED	80
6.1	CONTINUIDAD DE LA REPRODUCCIÓN MULTIMEDIA	81
6.2	MPD HÍBRIDO PARA DISTRIBUCIONES UNICAST Y MULTICAST	82
6.2.1	Análisis del impacto de los procesos de handover en la continuidad de los servicios de streaming multimedia	85
6.3	MECANISMO DE RECUPERACIÓN DE ERRORES ADAPTATIVO	87
6.3.1	Adaptación asistida por el BM-SC	88
6.4	PRESTACIONES DEL MECANISMO DE RECUPERACIÓN DE ERRORES ADAPTATIVO .	92
6.4.1	Relación entre la tasa de codificación AL-FEC y las retransmisiones HTTP .	93
6.4.2	Análisis de la continuidad de la reproducción multimedia	94
6.4.3	Rendimiento de los algoritmos de adaptación	95
6.5	CONCLUSIONES	97
Capítulo 7:	ARQUITECTURAS DE RED MÓVIL BASADAS EN NFV Y SDN PARA LA DIFUSIÓN DE CONTENIDOS MULTIMEDIA	99
7.1	DISEÑO DE ARQUITECTURAS DE RED MÓVIL BASADAS EN NFV Y SDN PARA LA DIFUSIÓN DE CONTENIDOS	100
7.1.1	vMBMS: virtualización del plano de control e implementación del plano de datos sobre hardware propietario	101
7.1.2	vMBMS: virtualización del plano de datos y de control	103
7.1.3	vMBMS: VNFs conectadas por una red SDN	105
7.1.4	vMBMS: aplicaciones SDN para implementar el plano de datos	106
7.1.5	Comparación de las diferentes estrategias	109
7.2	DESPLIEGUE DEL MECANISMO DE RECUPERACIÓN DE ERRORES ADAPTATIVO EN ARQUITECTURAS DE RED MÓVIL BASADAS EN VIRTUALIZACIÓN	110
7.2.1	Distribución de segmentos multimedia en arquitecturas de red móvil basadas en NFV y SDN	112
7.2.2	Recuperación adaptativa de errores en arquitecturas de red móvil basadas en NFV y SDN	114
7.3	CONCLUSIONES	115
Capítulo 8:	CONCLUSIONES Y TRABAJOS FUTUROS	116
8.1	CONCLUSIONES GENERALES	116
8.2	TRABAJOS FUTUROS	117

8.3	DIVULGACIÓN DE RESULTADOS	119
8.3.1	Proyectos de investigación subvencionados	119
8.3.2	Publicaciones	120
LISTA DE ACRÓNIMOS		123
REFERENCIAS		123

Índice de figuras

FIGURA 2.1. EJEMPLO DEL FUNCIONAMIENTO DE DASH	9
FIGURA 2.2. SERVICIOS DE STREAMING ADAPTATIVO BASADOS EN SAND	12
FIGURA 2.3. ELEMENTOS DE LA ARQUITECTURA HÍBRIDA DASH/FLUTE	15
FIGURA 2.4. ARQUITECTURA DE REFERENCIA DE LA RED MÓVIL LTE	23
FIGURA 2.5. COMPOSICIÓN INTERNA DEL BM-SC	25
FIGURA 2.6. ARQUITECTURA HÍBRIDA DASH/FLUTE EN LTE	27
FIGURA 2.7. PLATAFORMA DE GESTIÓN Y ORQUESTACIÓN NFV-MANO	30
FIGURA 2.8. ARQUITECTURA DE LAS REDES SDN.	33
FIGURA 3.1. RELACIÓN ENTRE AL-FEC CODE RATE Y LA COBERTURA DEL SERVICIO	41
FIGURA 3.2. RELACIÓN ENTRE TASA DE DATOS Y COBERTURA DEL SERVICIO	42
FIGURA 3.3. IMPACTO DEL ISD Y DE LA DURACIÓN DEL SEGMENTO.	43
FIGURA 3.4. ANÁLISIS COMPARATIVO DE LOS CÓDIGOS RAPTOR Y RAPTORQ	44
FIGURA 3.5. RAPTOR Y RAPTORQ: COMBINACIÓN MCS Y AL-FEC CODE RATE	45
FIGURA 3.6. ESCENARIOS VIRTUALES BASADOS EN VNX	48
FIGURA 3.7. METODOLOGÍA DE EVALUACIÓN DEL SERVICIO.	49
FIGURA 3.8. ESTABILIDAD Y RETARDO: SIN LÓGICA DE ADAPTACIÓN	51
FIGURA 3.9. ESTABILIDAD Y RETARDO: CON LÓGICA DE ADAPTACIÓN	51
FIGURA 3.10. ANÁLISIS DE LA EFICIENCIA	52
FIGURA 4.1. ANÁLISIS DE RETARDO	54
FIGURA 4.2. EJEMPLO DE ANÁLISIS DE RETARDO	60
FIGURA 4.3. BUFFER MÍNIMO EN FUNCIÓN DE $P(f_{RC})$	62
FIGURA 4.4. IMPACTO DEL AL-FEC CR Y T_{sgmnt} EN LA RECUPERACIÓN UNICAST.	64
FIGURA 4.5. RELACIÓN ENTRE LA DURACIÓN DEL SEGMENTO Y EL BUFFER MÍNIMO	65
FIGURA 4.6. EVALUACIÓN DE LA QoE MEDIANTE EL CÁLCULO DE LA Y-PSNR	66
FIGURA 4.7. USUARIOS AFECTADOS POR LA PARADA DEL SERVICIO	67
FIGURA 5.1. DEFINICIÓN DEL SERVICIO DE STREAMING 3GP-DASH SOBRE EMBMS	70
FIGURA 5.2. DIAGRAMA DE BLOQUES PLATAFORMA EXPERIMENTAL U-VMOS	72
FIGURA 5.3. IMPACTO DEL AL-FEC Y T_{sgmnt} EN LOS SEGMENTOS PERDIDOS.	74
FIGURA 5.4. RELACIÓN ENTRE PARADAS DE SERVICIO, BUFFERING INICIAL Y BITRATE	75
FIGURA 5.5. CÁLCULO DE U-VMOS	77
FIGURA 5.6. IMPACTO DEL NIVEL DEL BUFFER EN LA CALIDAD PERCIBIDA.	78
FIGURA 6.1. EJEMPLO DE PARADA DEL SERVICIO	82
FIGURA 6.2. MPD PARA LA IDENTIFICACIÓN DE LA REPRESENTACIÓN EMBMS	83
FIGURA 6.3. ESCENARIOS DE HANDOVER	84
FIGURA 6.4. HANDOVER ENTRE DIFERENTES ÁREAS EMBMS	85

FIGURA 6.5. REPRESENTACIONES DE VIDEO SELECCIONADAS DURANTE UN HANDOVER. . . .	86
FIGURA 6.6. IMPACTO DE UN HANDOVER VERTICAL EN EL BUFFER DEL REPRODUCTOR	87
FIGURA 6.7. MECANISMO DE RECUPERACIÓN DE ERRORES ADAPTATIVO	89
FIGURA 6.8. ALGORITMOS DE ADAPTACIÓN EJECUTADOS EN EL BM-SC	92
FIGURA 6.9. IMPACTO DEL AL-FEC EN EL PORCENTAJE DE SEGMENTOS PERDIDOS	93
FIGURA 6.10. TIEMPO DE PARADA DE SERVICIO	94
FIGURA 6.11. ANÁLISIS DE INESTABILIDAD.	95
FIGURA 6.12. ANÁLISIS DE LA CALIDAD DE VÍDEO PROMEDIO	96
FIGURA 7.1. VMBMS: PLANO DE DATOS EN HARDWARE PROPIETARIO	102
FIGURA 7.2. VMBMS: VIRTUALIZACIÓN DEL PLANO DE DATOS Y DE CONTROL	104
FIGURA 7.3. VMBMS: VNFs CONECTADAS POR UNA RED SDN	105
FIGURA 7.4. VMBMS: SDN PARA IMPLEMENTAR EL PLANO DE DATOS.	107
FIGURA 7.5. COMPOSICIÓN DE LAS VNFs.	109
FIGURA 7.6. PROCEDIMIENTO PARA LA RECUPERACIÓN DE ERRORES	112
FIGURA 7.7. PROCEDIMIENTO DE INICIO DE SESIÓN VMBMS	113
FIGURA 7.8. RECUPERACIÓN ADAPTATIVA DE ERRORES SOBRE VMBMS	114

Índice de tablas

TABLA 2.1. RELACIÓN MOS Y CALIDAD PERCIBIDA	18
TABLA 3.1. MODELO DE SIMULACIÓN: ANÁLISIS DE LOS SERVICIOS DE STREAMING	39
TABLA 4.1. PARÁMETROS DEL SERVICIO: MÉTODO REDUCCIÓN LATENCIA	61
TABLA 5.1. PARÁMETROS DEL SERVICIO: EVALUACIÓN CALIDAD DE LA EXPERIENCIA	73
TABLA 6.1. PARÁMETROS DEL SERVICIO: MECANISMO DE RECUPERACIÓN DE ERRORES	93

Capítulo 1

Introducción y objetivos

1.1. Contexto y Motivación

Varios informes indican que el tráfico de vídeo dominará Internet en los próximos años [CIS] [ERI]. Las últimas predicciones apuntan que en el año 2022 los flujos de vídeo constituirán hasta un 82 % de todo el tráfico que atraviese la red. El crecimiento afectará a los servicios de vídeo bajo demanda y a los servicios de difusión de contenidos generados en directo tanto en las redes fijas como en las redes móviles. Esta tesis doctoral se centra en proponer soluciones que contribuyen al diseño y mejora de los servicios de difusión de contenidos en las redes móviles.

La definición de las redes 5G debe satisfacer los estrictos requisitos de ancho de banda y retardo que demandan los servicios multimedia y otras aplicaciones más novedosas como el Internet táctil [DML⁺17] o las aplicaciones de la industria 4.0 [MCC17]. Con este propósito, los organismos de estandarización han propuesto incorporar en las próximas arquitecturas de red móvil nuevos paradigmas como: la virtualización de las funciones de red o Network Functions Virtualization (NFV) y las redes definidas por software o Software Defined Networks (SDN). Estas tecnologías [Sta15] proporcionan mecanismos de red para desplegar ágilmente nuevos servicios y gestionar los flujos de tráfico de una forma más eficiente.

El servicio evolved Multimedia Broadcast and Multicast Service (eMBMS) [3GP17a], definido por el 3rd Generation Partnership Project (3GPP) para las redes Long Term Evolution (LTE), permite enviar el mismo contenido a múltiples receptores al mismo tiempo utilizando un canal común. Es por ello que su uso es adecuado para realizar retransmisiones en directo. Sin embargo, la tecnología no se ha desplegado comercialmente con éxito (hasta la fecha) porque los operadores de red no han identificado un modelo de negocio claro. Algunos operadores, que ya han realizado despliegues y pruebas sobre pilotos de demostración, reconocen que no existe una “killer app” [ALL] que acelere la penetración de la tecnología en el mercado.

La incorporación de soluciones basadas en SDN/NFV en las arquitecturas de red supone una nueva oportunidad para que los operadores de red desplieguen servicios eMBMS. Como se explica en el Capítulo 7, la combinación de la virtualización de las funciones de red y las redes definidas por software flexibiliza el despliegue de los servicios de difusión de contenidos. Tradicionalmente, estos servicios se soportan sobre equipos basados en hardware propietario,

pero la virtualización de la red hace posible que se desplieguen en servidores genéricos. Por otro lado, el uso de las tecnologías SDN y NFV reduce el CAPEX y el OPEX, lo que puede ayudar a definir un modelo de negocio.

El estado del arte muestra diversas propuestas centradas en mejorar aspectos clave de los servicios de streaming multimedia, como, por ejemplo, el retardo de las retransmisiones en directo o la continuidad de la reproducción. Sin embargo, muchas de estas soluciones son costosas de implementar en las arquitecturas de red móvil clásicas, donde existe un fuerte acoplamiento entre el software y el hardware de las funciones de red. Gracias a la virtualización de las funciones de red es posible reducir el tiempo que transcurre desde que se concibe un servicio o producto hasta que puede lanzarse al mercado. Además, la virtualización de la red permite desplegar nuevos mecanismos de streaming de una forma más flexible, evitando que los fabricantes desarrollen nuevas versiones de sus productos para incorporar la funcionalidad deseada.

El 3GPP define el uso de Dynamic Adaptive Streaming over HTTP (DASH) [ISO14] como mecanismo de streaming para dar soporte a los servicios de difusión de contenidos multimedia basados en eMBMS. DASH fue concebido originalmente como estándar para servicios de streaming adaptativo sobre HTTP pero su uso se ha adaptado a las transmisiones punto-a-multipunto. El 3GPP define así el servicio de streaming 3GP-DASH sobre eMBMS (Capítulo 2), que combina el uso de DASH con técnicas de corrección de errores a nivel de aplicación y un procedimiento de recuperación de errores unicast que se inicia cuando el contenido multimedia no se ha transmitido correctamente sobre eMBMS. La revisión del estado del arte revela la necesidad de proponer, en este contexto, soluciones que permitan reducir la latencia y evitar las paradas del servicio. El retardo y el número de eventos de parada son parámetros que afectan a la calidad de la experiencia, sin embargo, no se han identificado trabajos de investigación que analicen el impacto de estos parámetros en la calidad que perciben los usuarios de los servicios de streaming 3GP-DASH sobre eMBMS.

Por otro lado, el estado del arte también revela la necesidad de definir un marco arquitectural que permita desplegar los mecanismos de streaming propuestos en las redes móviles basadas en las tecnologías de virtualización. Se debe tener en cuenta que, aunque el 3GPP pretende abordar en el futuro la virtualización del servicio eMBMS [5GP], a día de hoy todavía no están claras cuales son las transformaciones que permiten desplegar los servicios basados en eMBMS en un entorno virtualizado. La presente tesis doctoral tiene como objetivo clarificar esta cuestión, considerando soluciones que faciliten la interoperación de las arquitecturas de red clásicas y las futuras arquitecturas de red basadas en funciones de red virtualizadas.

Para validar las propuestas surgidas en el marco de la tesis doctoral se han desarrollado dos plataformas de experimentación. La primera de ellas combina modelos analíticos y de simulación con el objetivo de proporcionar un entorno de experimentación que permita validar las propuestas de una forma flexible. La segunda, basada en el despliegue de escenarios virtuales, se ha desarrollado con el objetivo de realizar experimentos sobre redes móviles reales o pilotos

de demostración. Los resultados que se han obtenido utilizando estas plataformas muestran que las propuestas de la tesis doctoral pueden mejorar las prestaciones de los servicios multimedia en términos de retardo y de continuidad de la reproducción multimedia. La tesis analiza cómo se pueden incorporar estas soluciones en las arquitecturas de red móvil basadas en las funciones de red virtualizadas.

1.2. Objetivos

El objetivo general de la tesis doctoral es formular propuestas que contribuyan al diseño de los servicios multimedia en las arquitecturas de red móvil basadas en tecnologías de virtualización. Para ello, se deberán determinar las modificaciones de los mecanismos de streaming que permiten mejorar aspectos clave de estos servicios, como el retardo de las retransmisiones en directo o la continuidad de la reproducción. Debido a que las arquitecturas de red móvil actuales proporcionan poca flexibilidad para desplegar nuevos servicios, la tesis también se plantea como objetivo determinar las transformaciones que debe sufrir la arquitectura para agilizar el despliegue de los mecanismos propuestos en el marco de la tesis doctoral.

El objetivo general se puede descomponer en los siguientes objetivos específicos:

1. Interpretar los trabajos de investigación que componen el estado del arte para realizar una crítica de la situación actual de los servicios multimedia en las redes móviles. Para ello, se deberán identificar las limitaciones que existen en las arquitecturas de red clásicas y en los mecanismos de streaming para garantizar los estrictos requisitos de retardo de los servicios de difusión de contenidos, así como para garantizar la continuidad de la reproducción multimedia. Debido a que estos factores afectan a la calidad percibida por los usuarios, también se contrastarán diferentes modelos que permitan medir la calidad de la experiencia.
2. Diseñar soluciones de mejora de los mecanismos de streaming que permitan reducir el retardo de las retransmisiones en directo y garantizar la continuidad de la reproducción multimedia. Tras revisar el estado del arte (Capítulo 2) se detecta que este objetivo se puede abordar con dos perspectivas diferentes. El primer enfoque consiste en proponer un método que permita determinar el buffer mínimo que evita las paradas de servicio en la mayor parte de los usuarios. Para ello, se analizará la cadena de retardos que sufre el contenido multimedia cuando se distribuye a través de la red móvil. El segundo enfoque, basado en los servicios de streaming adaptativo asistido por la red, consiste en proponer un mecanismo de recuperación de errores que permita adaptar el flujo multimedia de cada cliente durante la retransmisión unicast del contenido. En este caso, será necesario identificar las modificaciones arquitecturales que permiten desplegar el mecanismo, así como realizar un análisis comparativo entre diferentes algoritmos de adaptación.
3. Emplear un modelo para la evaluación de la calidad de la experiencia que pueda aplicarse

en el marco de la tesis doctoral para determinar el impacto de diferentes parámetros del servicio en la calidad percibida por los usuarios. Tras lo anterior, se desarrollará una metodología que permita definir los parámetros de servicio mediante las medidas obtenidas de los análisis de la calidad de la experiencia.

4. Proponer modificaciones de la arquitectura de red móvil que permitan agilizar el despliegue de los mecanismos de streaming propuestos durante la realización de la tesis doctoral. Para ello, se valorará el uso combinado de soluciones basadas en nuevos paradigmas de red como la virtualización de las funciones de red o las redes definidas por software. Se evaluarán diferentes estrategias de diseño considerando una mayor o menor integración de estas tecnologías en la arquitectura de la red móvil. A continuación, se compararán las diferentes estrategias para seleccionar aquella que proporcione mayor agilidad y menor coste. Finalmente, se analizará cómo se puede desplegar el mecanismo de recuperación de errores unicast propuesto en el objetivo anterior sobre esta nueva arquitectura de red.
5. Validar las diferentes propuestas surgidas durante la realización de la tesis doctoral mediante la realización de experimentos. Para ello se desarrollará una plataforma de experimentación que permita realizar análisis sobre redes móviles reales o sobre pilotos de demostración y que permita evaluar el rendimiento de las soluciones propuestas de una forma flexible. Para analizar los servicios multimedia sobre redes móviles reales se desarrollará una plataforma basada en escenarios virtuales. Para evaluar de una forma flexible las soluciones que se proponen para mejorar los mecanismos de streaming se desarrollará una plataforma que combine modelos de simulación y modelos matemáticos.

1.3. Estructura de la memoria

La memoria de la tesis doctoral se ha estructurado de la siguiente manera:

- Capítulo 2, *Estado del arte*

Este capítulo presenta una revisión del estado del arte sobre diferentes líneas de investigación relacionadas con la tesis doctoral. En primer lugar, el capítulo presenta las bases teóricas que permiten comprender los mecanismos de streaming multimedia, y en particular, cómo se despliegan los servicios de difusión de contenidos multimedia en las redes móviles. Más tarde, se analizan propuestas centradas en la evaluación de la calidad de la experiencia para determinar si existe algún modelo que pueda aplicarse a los servicios de difusión de contenidos. Finalmente, el capítulo aborda la evolución de la arquitectura de la red móvil con el objetivo de entender cómo se pueden desplegar los mecanismos de streaming multimedia en las arquitecturas de red móvil basadas en tecnologías de virtualización.

- Capítulo 3, *Herramientas para el análisis de los servicios de streaming multimedia*

En este capítulo se describen las herramientas que se han utilizado para analizar los servicios de streaming multimedia en las redes móviles y que han permitido validar las diferentes propuestas de la tesis doctoral. La primera parte del capítulo se centra en el análisis de los servicios multimedia mediante la combinación de modelos de simulación y analíticos. Utilizando esta plataforma, ha sido posible analizar el impacto de diferentes parámetros del servicio en la transmisión broadcast de los segmentos multimedia. A continuación, el capítulo presenta una plataforma basada en escenarios virtuales que ha sido desarrollada a partir de componentes software de código abierto. Esta plataforma se ha desplegado con éxito sobre un piloto de demostración de una red móvil LTE, lo que ha permitido obtener medidas extremo a extremo del servicio.

- Capítulo 4, *Método para reducir la latencia y evitar paradas de servicio*

En este capítulo se analiza la cadena de retardos que sufre el contenido multimedia durante su envío sobre la red móvil. El análisis permite determinar cómo se debe sincronizar el servidor de contenidos y los reproductores multimedia para que el contenido multimedia se reciba correctamente durante su transmisión broadcast. A continuación, el capítulo propone un método para reducir la latencia del servicio y evitar las paradas de la reproducción al mismo tiempo. Este método se basa en calcular el buffer mínimo que garantiza la continuidad de la reproducción en la mayor parte de los clientes de streaming.

- Capítulo 5, *Definición de los parámetros de servicio a partir del análisis de la QoE*

Este capítulo presenta una metodología que permite definir diferentes parámetros del servicio a partir del análisis de la calidad de la experiencia. En particular, el capítulo discute como aplicar un modelo de calidad de la experiencia que está en fase de estandarización a un servicio de difusión de contenidos multimedia. Los resultados obtenidos permiten entender el impacto de los parámetros del servicio en la calidad percibida por los usuarios.

- Capítulo 6, *Mecanismo de recuperación de errores adaptativo*

En este capítulo se diseña un mecanismo de recuperación de errores adaptativo que permite reducir la latencia del servicio y garantizar la continuidad de la reproducción multimedia. En este caso, el mecanismo propuesto se basa en los servicios de streaming adaptativo asistido por la red. El capítulo explica cómo se puede ejecutar un algoritmo de adaptación DASH en un elemento intermedio de la red móvil, para adaptar el flujo multimedia de los clientes cuando se establece la recuperación unicast de los segmentos multimedia que se han perdido tras su transmisión broadcast. El capítulo evalúa el rendimiento de dos algoritmos de adaptación diferentes.

- Capítulo 7, *Arquitecturas de red móvil basadas en NFV y SDN para la difusión de contenidos multimedia*

Este capítulo aborda la evolución de la arquitectura de soporte a los servicios de difusión de contenidos multimedia en el marco de las redes móviles basadas en la virtualización de las funciones de red (NFV) y las redes definidas por software (SDN). El capítulo analiza cuatro enfoques diferentes que combinan en menor o mayor medida soluciones de tipo SDN y NFV. Tras realizar un análisis comparativo de los diferentes enfoques, el capítulo explica cómo se puede desplegar el mecanismo de recuperación de errores propuesto en el capítulo anterior sobre este tipo de arquitecturas. Para ilustrarlo, se selecciona la arquitectura que ofrece la mayor agilidad y menor coste.

- Capítulo 8, *Conclusiones y trabajos futuros*

Finalmente, el Capítulo 8 proporciona las conclusiones del trabajo de investigación realizado durante la tesis doctoral. El capítulo analiza el cumplimiento de los objetivos propuestos en esta introducción y proporciona unas indicaciones sobre cuales son las líneas futuras del trabajo de investigación postdoctoral. Finalmente, se proporciona información sobre la divulgación de los resultados obtenidos.

Capítulo 2

Estado del arte

Este capítulo tiene como objetivo interpretar los trabajos de investigación que componen el estado del arte para realizar una crítica de la situación actual de los servicios multimedia en las redes móviles. Existe por tanto una correspondencia directa entre este capítulo y el objetivo específico número 1, aunque también ha servido para definir los objetivos específicos 2, 3 y 4 de la tesis doctoral (ver Capítulo 1).

La estructura del capítulo es la siguiente:

- La Sección 2.1 proporciona la descripción de los mecanismos y protocolos de streaming multimedia que definen los estándares del 3GPP para su uso en las redes móviles. Posteriormente, se justifica la necesidad de proponer nuevas soluciones que permitan reducir la latencia y garantizar la continuidad de la reproducción multimedia en el marco de los servicios de streaming en directo.
- La Sección 2.2 presenta dos tendencias para medir la calidad de la experiencia: el uso de modelos objetivos y subjetivos. Se justifica la necesidad de aplicar un modelo objetivo para analizar un servicio de difusión de contenidos y se analizan los trabajos relacionados para determinar si los modelos existentes pueden aplicarse en el marco de la tesis doctoral.
- Las secciones 2.3 y 2.4 describen la arquitectura de referencia de las redes móviles LTE y muestran cual es su evolución hacia las redes 5G. Para ello, se proporciona una visión general de los paradigmas de red que se están considerando para definir las redes 5G: la virtualización de las funciones de red y las redes definidas por software. Analizando los trabajos publicados, se detecta la necesidad de determinar cuales son las transformaciones que se requieren en las arquitecturas de red móvil actuales para dar soporte a servicios de difusión de contenidos más flexibles.
- Finalmente, la Sección 2.5 presenta las conclusiones generales que se obtienen tras revisar el estado del arte.

2.1. Servicios de streaming multimedia en redes móviles

Esta sección proporciona una descripción del estándar DASH y ofrece una visión general de los algoritmos de adaptación. Más tarde, se aborda una nueva tendencia en streaming multimedia denominada “adaptación asistida por la red”. Tras describir cómo se integra DASH en los servicios de difusión de contenidos, se identifica el problema de investigación.

2.1.1. Servicios de streaming adaptativo sobre HTTP

DASH [ISO14] es un estándar que define un conjunto de mecanismos que permiten adaptar la transmisión del flujo multimedia a las características de los dispositivos y al ancho de banda disponible en la red. El estándar define que, para proporcionar esta capacidad de adaptación, el contenido multimedia debe codificarse con diferentes bitrates y resoluciones. De este modo, se obtienen múltiples versiones alternativas del mismo contenido. De aquí en adelante se utilizará el término representación para hacer referencia a una de estas versiones alternativas.

La adaptación es posible gracias a que el reproductor puede conmutar entre múltiples representaciones disponibles. Para ello, el cliente ejecuta un algoritmo de adaptación que, a partir de estimaciones del ancho de banda disponible y/o del nivel del buffer, selecciona la representación más adecuada para garantizar la continuidad de la reproducción multimedia. El estándar no especifica cómo se debe diseñar la lógica de adaptación, lo que ha motivado la aparición de numerosas propuestas de algoritmos orientados a diferentes tipos de redes, dispositivos y escenarios (subsección 2.1.2).

Mediante la conmutación, la lógica de adaptación trata de evitar que el buffer del cliente se quede sin datos que reproducir. Aunque se ha detectado que la calidad percibida por los usuarios puede verse afectada por la conmutación, también se ha demostrado que las paradas de servicio son, en general, más perjudiciales desde la perspectiva de la calidad de la experiencia [SES⁺15] (Capítulo 5).

El estándar de DASH también indica que, para que sea posible conmutar entre las diferentes representaciones disponibles, es necesario que estas estén fragmentadas en segmentos (ficheros) de la misma duración. Los clientes DASH obtienen los segmentos multimedia desde servidores web estándar, siendo posible replicarlos en redes de distribución de contenidos o Content Delivery Network (CDN)s [YLML10, AGH⁺15]. El uso de las CDN's proporciona una menor latencia en términos de Round Trip Time (RTT) durante el acceso al contenido, así como una mayor disponibilidad.

Para iniciar una sesión de streaming basado en DASH, el cliente debe descargar previamente un fichero de metadatos llamado Media Presentation Description (MPD). El fichero MPD describe las representaciones disponibles, la localización de cada segmento multimedia en el servidor (identificado unívocamente por un URL) y otros parámetros de la sesión, como, por ejemplo, el número mínimo de segmentos que debe almacenarse en el buffer del reproductor antes de que la sesión pueda comenzar. Tras la descarga y análisis del MPD, el reproductor so-

Figura 2.1: Ejemplo del funcionamiento de DASH. La figura muestra que el cliente es capaz de adaptarse a las variaciones de ancho de banda que existen en la red gracias a que puede conmutar entre las diferentes representaciones disponibles.

licita segmentos multimedia pertenecientes a las diferentes representaciones disponibles según dicte el algoritmo de adaptación.

La Figura 2.1 muestra un ejemplo del funcionamiento de DASH. El servidor almacena un contenido multimedia codificado con tres tasas de bits (bitrates) diferentes: calidad alta, media y baja. El ancho de banda disponible en el canal que transporta los segmentos sufre variaciones. Gracias al algoritmo de adaptación, el cliente puede adaptarse a estas variaciones solicitando segmentos que pertenecen a diferentes representaciones multimedia. Como se puede observar en la figura, la representación solicitada en cada momento es aquella cuyo bitrate se ajusta mejor al ancho de banda disponible.

- Servicios de streaming adaptativo basados en HTTP/2

En mayo de 2015 se publicó una nueva versión del protocolo HTTP (HTTP/2) [BPT15]. Esta nueva versión ha sido diseñada para abordar las necesidades de transporte de las páginas web que se desarrollan hoy en día. El protocolo incluye nuevas funciones que no existían en versiones anteriores como, por ejemplo, la compresión de cabeceras, la posibilidad de realizar peticiones de varios recursos sobre la misma conexión TCP o la función *Server Push*, que permite al servidor enviar recursos de forma anticipada al cliente sin que este los haya solicitado previamente, y asumiendo que los va a solicitar en un futuro.

Desde su aparición, han surgido diferentes trabajos que proponen adaptar DASH a las nuevas funcionalidades que ofrece HTTP/2 [YLNS17, HBP⁺18, XZG18, LNN⁺18]. El artículo de Huysegems *et al.* [HHB⁺15] ofrece una buena visión sobre los beneficios que proporciona HTTP/2 al streaming adaptativo. Los autores abordan la aplicación de varias de las funciones de HTTP/2 en este contexto. Por ejemplo, muestran que es posible que el cliente envíe el mensaje de HTTP/2 RST_STREAM [BPT15] para cancelar la descarga en curso de un segmento. Esto es útil en escenarios donde el ancho de banda se ha reducido significativamente y no es posible descargar el segmento multimedia antes de que el buffer del reproductor se vacíe. Otra de sus propuestas es utilizar la capacidad de HTTP/2 para multiplexar el envío de varios recursos de

forma concurrente. Esto permite transmitir varios segmentos en paralelo sobre la misma conexión TCP, lo que puede ser útil si el buffer se ha quedado vacío y hay que cargarlo rápidamente con segmentos de baja calidad. Finalmente, los autores exploran el uso de la función *Server Push*, que permite reducir el retardo de los servicios de streaming en directo gracias a que los segmentos multimedia se envían al cliente tan pronto como se hayan generado en el servidor. El uso de HTTP/1.x puede incrementar el retardo porque requiere que los clientes comprueben periódicamente si los segmentos están disponibles para ser descargados.

Durante la realización de la tesis se ha dirigido un Trabajo Fin de Grado [Ram16] cuyo objetivo era analizar el envío anticipado de vídeo mediante la función *Server Push* de HTTP/2. Los resultados obtenidos vienen a confirmar que es posible reducir el retardo de las retransmisiones en directo, tal y como apuntan Huysegems *et al.*.

2.1.2. Algoritmos de adaptación para streaming multimedia basado en DASH

Algunas propuestas centradas en los servicios de streaming adaptativo sobre HTTP/2 requieren diseñar algoritmos de adaptación que tengan en cuenta las nuevas funcionalidades que ofrece el protocolo, como por ejemplo, la función *Server push* [HPW⁺18]. La lógica de adaptación no ha sido definida en el estándar, por lo que no existen limitaciones para abordar su diseño. El estado del arte muestra que se han propuesto diversos algoritmos de adaptación orientados a diferentes tipos de redes, dispositivos y escenarios [SES⁺15]. La mayoría de estas propuestas están basadas en técnicas pasivas para la estimación del ancho de banda [CB15]. Las técnicas pasivas utilizan información sobre datos que ya han sido descargados por las aplicaciones, es decir, no requieren que se inyecte en la red tráfico adicional. En el caso de DASH, la técnica más común consiste en utilizar el tamaño y tiempo de descarga de los segmentos multimedia. Por ejemplo, una posible estrategia es seleccionar la representación multimedia a partir del throughput instantáneo, que se define como el tamaño del último segmento descargado dividido por el tiempo que ha sido necesario para descargarlo. El problema de utilizar el throughput instantáneo es que el reproductor puede conmutar excesivamente entre las representaciones disponibles, lo que reduce la calidad de la experiencia (Sección 2.2).

Para evitar la excesiva conmutación, los reproductores multimedia implementan un estimador que calcula el ancho de banda promedio considerando los últimos N segmentos descargados. En la literatura se pueden encontrar múltiples ejemplos de este tipo de algoritmos [CCPM13, JSZ14, LZG⁺14].

Durante la realización de esta tesis, se ha colaborado en el diseño y desarrollo de un nuevo algoritmo de adaptación [ABLP18] que se basa en el método matemático Adaptive Forgetting Factor (AFF). Si el ancho de banda varía frecuentemente y de forma irregular o abrupta el estimador tiende a calcular el ancho de banda considerando un número de segmentos N pequeño. Sin embargo, si el ancho de banda varía con poca frecuencia y permanece estable, el estimador tiende a utilizar un número de segmentos N más grande. Los resultados obtenidos durante su evaluación muestran que, en comparación a otras alternativas, el algoritmo reduce las paradas

que se producen durante la reproducción del contenido y el número de conmutaciones.

Otros trabajos [YCH14, LBK15] proponen utilizar técnicas activas para la estimación del ancho de banda. Estas técnicas requieren el envío de paquetes de datos “dummy” que se generan y se transportan sobre la red con el único objetivo de que se pueda realizar la estimación. Es decir, no son datos generados por ningún tipo de aplicación, y por tanto, no tienen ninguna otra utilidad que la propia estimación. El estado del arte muestra que los algoritmos de adaptación basados en técnicas activas obtienen un buen rendimiento si se comparan con los algoritmos basados en técnicas pasivas. Su principal desventaja, sin embargo, es el consumo de ancho de banda, por lo que en escenarios donde existen recursos limitados parece más recomendable utilizar técnicas pasivas.

Los algoritmos de adaptación pueden seleccionar la representación más adecuada utilizando, además de estimaciones de ancho de banda, información del buffer del reproductor. Por ejemplo, si se detecta que el nivel del buffer ha caído por debajo de un umbral mínimo, el algoritmo de adaptación selecciona la representación de menor bitrate para asegurar la continuidad de la reproducción. Un ejemplo, de este tipo de algoritmos es BOLA [SUS16] que combina el cálculo del throughput instantáneo con un método de control del buffer.

2.1.3. Adaptación asistida por la red

El estándar de DASH está dividido en diferentes partes. La subsección 2.1.1 se centra en la segunda edición de la parte 1 del estándar, publicada en 2014 y llamada *Part 1: Media Presentation Description and Segments Formats* [ISO14]. A lo largo del tiempo se han publicado nuevas partes que abordan otras cuestiones relevantes para definir los servicios de streaming basados en DASH, como aspectos de autenticación y cifrado, el uso de HTTP/2, o la adaptación asistida por la red. La adaptación asistida por la red se define en la parte 5 del estándar, publicada en el año 2017 y llamada *Part 5: Server and Network Assisted DASH (SAND)* [ISO17].

SAND se basa en la idea de adaptar la transmisión del flujo multimedia a través de la actuación de diferentes elementos de la red. Los clientes todavía pueden ejecutar un algoritmo de adaptación, pero la selección de la representación multimedia se decide a partir de información que proviene de los servidores y otros elementos intermedios. SAND especifica los mecanismos y mensajes de control que hacen posible la cooperación entre los clientes DASH y estos elementos, que reciben el nombre de DASH-aware Network Elements (DANE).

Así, el algoritmo de adaptación puede utilizar, además de estimaciones del ancho de banda o el nivel del buffer, otras medidas que aportan una información más precisa sobre las condiciones de la red. Por ejemplo, en las redes móviles LTE las estaciones base realizan la planificación de los recursos radio [CPG⁺13] y establecen parámetros del nivel físico que determinan el ancho de banda asignado a un terminal. Como se explica en el Capítulo 6, esta información se puede utilizar para mejorar la capacidad de adaptación de los reproductores multimedia.

Cofano *et al.* [CCZ⁺17] identifican dos posibles enfoques para diseñar mecanismos de adaptación basados en SAND (Figura 2.2). El primero de ellos (Figura 2.2a) consiste en moni-

Figura 2.2: Enfoques para el diseño de servicios de streaming adaptativo basados en SAND

torizar métricas relacionadas con la reproducción del contenido multimedia como, por ejemplo, el nivel del buffer, para tomar decisiones a nivel de red que permitan maximizar la calidad percibida por los usuarios. Por ejemplo, para evitar que el nivel del buffer del reproductor caiga por debajo de cierto umbral, se pueden aplicar políticas de calidad de servicio que permitan asignar más ancho de banda a los flujos multimedia o seleccionar un nodo de la CDN que proporcione una menor latencia. Como se puede observar, el cliente sigue ejecutando el algoritmo de adaptación, pero gracias a la asistencia de los elementos intermedios de la red, es posible reducir el número de conmutaciones o que se seleccionen representaciones de calidades más altas.

El segundo enfoque (Figura 2.2b) se diferencia del anterior en que un elemento de la red (e.g., un proxy) selecciona, en nombre del cliente, la representación multimedia que debe solicitar. Es decir, el algoritmo de adaptación se ejecuta en un DANE y no en el cliente. Como se puede observar en la figura, también es posible actuar sobre la red a partir de la monitorización de las métricas de calidad de la experiencia.

Aunque estas son las dos estrategias más comunes, existen estrategias híbridas que combinan en mayor o menor medida las dos anteriores. También conviene destacar que la actuación sobre la red no es imperativa, aunque se ha demostrado que puede mejorar el rendimiento de los servicios de streaming adaptativo [CCZ⁺17]. La aparición de la virtualización de las funciones de red y las redes definidas por software facilita el despliegue de soluciones de este tipo.

Bhat *et al.* [BRZS17] diseñan una arquitectura que incorpora soluciones de tipo SDN para dar soporte a los servicios de streaming adaptativo basados en SAND. Esta arquitectura considera el enfoque de la Figura 2.2a, donde la responsabilidad de seleccionar la representación multimedia sigue recayendo en el cliente. La solución que proponen permite gestionar la CDN de una forma más eficiente, lo que implica una reducción del número de conmutaciones y la selección de representaciones de vídeo de mayor calidad.

El estado del arte también incluye propuestas basadas en el segundo enfoque (Figura 2.2b).

Por ejemplo, El Essaili *et al.* [ESS⁺15] proponen utilizar un proxy, ubicado cerca de las estaciones base, para seleccionar la representación que se ajusta con más precisión al ancho de banda disponible. En este caso, el cliente no ejecuta el algoritmo de adaptación, sino que se ejecuta en un proxy que utiliza la información de ancho de banda que le proporcionan las estaciones base.

2.1.4. Servicios de streaming 3GP-DASH sobre eMBMS

El uso de HTTP para enviar el mismo contenido a múltiples receptores simultáneamente no es adecuado si el objetivo es gestionar los recursos radio de una forma eficiente. Para resolver este problema, el 3GPP ha definido eMBMS [3GP17a, 3GP18a]. eMBMS es una tecnología definida para las redes móviles LTE que permite transmitir una única copia de los datos (por ejemplo, multimedia o actualizaciones de software) a todos los receptores multicast conectados al servicio. El transporte de los datos sobre eMBMS tiene dos características fundamentales: (1) es unidireccional, y por tanto, requiere el uso de un protocolo diseñado para el transporte unidireccional de ficheros. (2) No es fiable, por lo que es recomendable aplicar técnicas de corrección de errores que mejoren esta fiabilidad.

Aunque DASH se definió inicialmente como una solución de streaming adaptativo sobre HTTP, el 3GPP ha propuesto adaptar su uso a los servicios de difusión de contenidos multimedia. La propuesta, llamada *3GP-DASH streaming in MBMS*, se basa en el uso del protocolo File Delivery over Unidirectional Transport (FLUTE) [PWL⁺12] como solución para enviar los segmentos multimedia, codificados según indica el estándar de DASH, sobre el canal de eMBMS. FLUTE permite encapsular los segmentos como ficheros y enviarlos a los múltiples receptores utilizando el protocolo de transporte User Datagram Protocol (UDP). Aunque FLUTE se puede utilizar para enviar múltiples segmentos generados en tiempo real, el protocolo no fue concebido con ese propósito específico. Esto ha motivado la aparición del protocolo Real-Time Object Delivery over Unidirectional Transport (ROUTE) [WSM⁺16], basado en FLUTE, pero optimizado para los servicios de difusión de contenidos que se construyen en torno a DASH. El protocolo ROUTE ha sido considerado en el estándar de televisión Advanced Television Systems Committee (ATSC) 3.0 [ATS17].

El envío de los segmentos multimedia puede verse afectado por errores de transmisión, por tanto, se recomienda el uso de técnicas que proporcionen una mayor fiabilidad a su envío. El 3GPP propone combinar el uso del protocolo FLUTE con técnicas de corrección de errores a nivel de aplicación: las técnicas Application Layer Forward Error Correction (AL-FEC) [Lub12].

La codificación AL-FEC permite transmitir los datos de la aplicación junto con otros datos de redundancia para proporcionar al receptor la capacidad de corregir los errores introducidos durante la transmisión. Actualmente, Raptor [LSWS07] es la técnica AL-FEC estandarizada por el 3GPP para proporcionar una transferencia fiable del contenido multimedia sobre eMBMS. Posteriormente, surgió el código RaptorQ [LSWS11], cuyo rendimiento se analizará en el Capítulo 3. Los códigos Raptor son códigos bloque, lo que permite fragmentar un flujo de

datos en varios ficheros, denominados *source blocks*, que se codifican de forma independiente. En un servicio de streaming 3GP-DASH sobre eMBMS, cada segmento multimedia es un *source block* que debe ser codificado. Un *source block* se compone de k fragmentos del mismo tamaño llamados *source symbols*. A partir de los *source symbols*, el codificador puede generar r símbolos de redundancia o *repair symbols* (también del mismo tamaño). Para expresar la cantidad de redundancia introducida por el transmisor es común utilizar la métrica *tasa de codificación AL-FEC*, que se define como el cociente entre los k *source symbols* y los $k + r$ símbolos resultantes del proceso de codificación, es decir $k/(k + r)$. El segmento multimedia y los datos de redundancia (es decir, los $k + r$ símbolos) se envían sobre el canal eMBMS a los receptores multicast. El proceso de decodificación concluye con éxito si el número de símbolos recibidos (*source* o *repair*) es ligeramente superior a k .

La capacidad del receptor para corregir errores depende de la tasa de codificación AL-FEC y de la tasa de error del canal eMBMS. Si la tasa de error es elevada, la decodificación AL-FEC del segmento multimedia puede finalizar sin éxito y no es posible entregar el segmento al buffer del reproductor. Con el objetivo de aumentar la fiabilidad de la transmisión, el 3GPP también propone utilizar el modo de transmisión Multicast/Broadcast over Single Frequency Network (MBSFN) [3GP18a]. MBSFN se basa en la cooperación de varias estaciones base, sincronizadas de forma muy precisa en tiempo y frecuencia, para transmitir la misma señal radio. El uso de MBSFN mejora la relación señal a ruido o Signal to Interference plus Noise Ratio (SINR) de los receptores multicast. Es decir, la tasa de error del canal se reduce, lo que incrementa la probabilidad de que un receptor decodifique correctamente el segmento multimedia.

A pesar de que MBSFN y las técnicas AL-FEC pueden aplicarse de forma combinada, su uso no asegura la entrega de los segmentos multimedia. Para garantizar la entrega, el 3GPP propone utilizar HTTP como mecanismo de recuperación de errores unicast, i.e., para recuperar los segmentos que no han sido transmitidos correctamente sobre eMBMS. Esta es la razón para adaptar el uso de DASH a los servicios de difusión de contenidos multimedia. Si el flujo multimedia se codifica como indica el estándar de DASH, es posible combinar las técnicas AL-FEC con el uso de HTTP para recuperar los segmentos perdidos sobre eMBMS. Esto se debe a que la fragmentación del contenido multimedia en segmentos permite utilizar códigos bloque como Raptor.

La generación de múltiples representaciones no tiene cabida en este contexto, puesto que solo se puede enviar una representación sobre el canal unidireccional de eMBMS (de aquí en adelante se utilizará el término *representación eMBMS* para hacer referencia a dicha representación). La representación eMBMS se codifica con un bitrate que se ajusta a la tasa de datos máxima del servicio. En los servicios unicast, el ancho de banda del canal varía dinámicamente en función del esquema de modulación y codificación o Modulation and Coding Scheme (MCS) y de los recursos radio asignados por la estación base [DM11]. En los servicios eMBMS, el MCS y el número de recursos asignados al canal son parámetros constantes. Por tanto, la tasa de datos del servicio no varía durante la sesión eMBMS. En definitiva, DASH

no se utiliza en este contexto como solución de streaming adaptativo porque sobre el canal de eMBMS solo se envía una única representación. DASH se utiliza como estándar para codificar y fragmentar la representación eMBMS y para permitir el uso de HTTP como mecanismo de recuperación de errores.

La Figura 2.3 muestra los componentes de la arquitectura de soporte a los servicios de streaming 3GP-DASH sobre eMBMS. El proveedor de contenidos genera la representación eMBMS y el fichero MPD. Los segmentos multimedia se envían sobre el canal de eMBMS como ficheros gracias al uso del protocolo FLUTE. Para proporcionar más fiabilidad a la entrega los segmentos han sido previamente protegidos mediante la aplicación de las técnicas AL-FEC. El decodificador AL-FEC del cliente intenta reconstruir el segmento a partir de los paquetes FLUTE (símbolos) que ha recibido. Si el segmento se decodifica correctamente, se almacena en una memoria cache local. Si no, se considera que el segmento se ha perdido y que debe ser recuperado utilizando HTTP.

El reproductor DASH solicita los segmentos multimedia a medida que estos son generados por el proveedor de contenidos. La disponibilidad de los segmentos se informa a los clientes mediante el campo *availabilityStartTime* del MPD, que indica el instante de tiempo a partir del cual un segmento está disponible para ser enviado. Cuando el reproductor solicita un segmento, primero comprueba si ya está disponible en su memoria cache. Si el segmento no está en cache, significa que el proceso de decodificación no se ha llevado a cabo con éxito y que es necesario recuperar el segmento por HTTP utilizando un canal unicast. Debido a que solo se envía una representación sobre eMBMS, el reproductor solicita el mismo segmento que fue transmitido por el canal de multidifusión.

Figura 2.3: Arquitectura híbrida DASH/FLUTE. La figura muestra los componentes de la arquitectura que dan soporte a los servicios 3GP-DASH sobre eMBMS. Esto incluye un mecanismo para la recuperación unicast de errores.

2.1.5. Problema de investigación: reducción del retardo y continuidad de la reproducción en los servicios de difusión de contenidos

El rendimiento de los servicios de streaming multimedia en directo puede verse afectado por el retardo extremo a extremo. Debido a que la demanda de estos servicios ha crecido significativamente en los últimos años, múltiples autores han abordado este problema por su relevancia.

El estado del arte muestra que existen diferentes tipos de soluciones para reducir la latencia del servicio dependiendo de la tecnología utilizada para distribuir el contenido. Por ejemplo, Yang *et al.* [YL15] proponen un mecanismo que permite reducir el retardo cuando un usuario de un servicio de IPTV cambia de canal. Los autores demuestran que es posible reducir este retardo, denominado *zapping time*, aumentando el ancho de banda excepcionalmente.

Por otro lado, existen diferentes propuestas que tratan de caracterizar de forma analítica los servicios de streaming en directo [BSA⁺12, TC12, AM14]. Sus autores proponen diferentes métodos para calcular el buffer mínimo que evita las paradas durante la reproducción, pero no consideran DASH como mecanismo para distribuir el contenido multimedia. Con respecto a los servicios de streaming basados en DASH, Lohmar *et al.* [LEF⁺11] analizan la cadena de retardos que sufre un segmento multimedia desde que se genera en el servidor de contenidos hasta que se recibe en el terminal del cliente. Los autores también calculan el buffer mínimo que evita las paradas en la reproducción, pero su método no aplica a los servicios de streaming 3GP-DASH sobre eMBMS porque la entrega de los segmentos multimedia se retarda por el uso de las técnicas AL-FEC y la recuperación de errores unicast. En esta tesis se propone un método para calcular el buffer mínimo que permite garantizar la continuidad de la reproducción en los servicios de streaming 3GP-DASH sobre eMBMS (Capítulo 4).

Por otro lado, Lohmar *et al.* [LEF⁺11] analizan como sincronizar la transmisión del servidor de contenidos con la petición de los segmentos que realiza el cliente. Como se explica en el Capítulo 4, es necesario añadir un retardo adicional para evitar problemas de sincronismo en los servicios de streaming 3GP-DASH sobre eMBMS. Tras analizar la cadena de retardos del servicio, el capítulo explica como definir el fichero MPD correctamente para evitar dichos problemas.

En relación a los servicios de streaming adaptivo sobre HTTP/2, Wei *et al.* [WS14] analizan si es posible reducir el tamaño del segmento multimedia para reducir el retardo. Los autores argumentan que el uso de HTTP puede incrementar excesivamente la sobrecarga de la comunicación (si el tamaño del segmento es menor, el número de transacciones HTTP puede aumentar excesivamente). Sin embargo, el uso de HTTP/2 puede reducir la sobrecarga gracias al uso de la función *Server Push* (Subsección 2.1.1). En los servicios de streaming 3GP-DASH sobre eMBMS la mayor parte de los segmentos se transmiten correctamente sobre eMBMS por lo que el impacto en la sobrecarga de la comunicación se reduce de forma significativa. El Capítulo 4 analiza si es posible reducir el tamaño del segmento en este contexto.

El estado del arte también incluye propuestas que se centran en la distribución de conte-

nidos sobre eMBMS, pero abordan la reducción del retardo en servicios genéricos de entrega de ficheros. Aunque estos ficheros podrían ser de tipo multimedia, las soluciones propuestas están orientadas a servicios donde los usuarios visualizan el contenido una vez que este ya ha sido descargado completamente. Es decir, las soluciones no están diseñadas con el objetivo de satisfacer los estrictos requisitos de retardo de los servicios de streaming 3GP-DASH sobre eMBMS, donde se requiere una entrega fiable y en tiempo real.

Montserrat *et al.* [MCFG12] analizan el impacto de diferentes parámetros de los servicios basados en eMBMS (por ejemplo, el MCS), en el tiempo necesario para transmitir un conjunto de ficheros. Los autores consideran dos mecanismos para la recuperación de los ficheros que no se han transmitido correctamente en el canal multicast. El primero consiste en utilizar conexiones punto a punto una vez que ha finalizado la sesión eMBMS. El segundo, más apropiado cuando se ha perdido una cantidad significativa de ficheros, consiste en establecer una nueva sesión eMBMS más robusta. Ninguno de estos mecanismos de recuperación de errores es el que define el 3GPP para su aplicación en los servicios de streaming 3GP-DASH sobre eMBMS.

Existen diferentes trabajos de investigación que proponen soluciones para proporcionar más robustez a las sesiones eMBMS. Una de estas soluciones consiste en modificar dinámicamente la cantidad de datos de redundancia que añade el codificador AL-FEC a los ficheros que se envían sobre eMBMS [FFBG12, FG14]. Sin embargo, esta propuesta considera que el envío de los ficheros se realiza en carrusel, es decir, una y otra vez hasta que todos los terminales han recibido todos los ficheros. Los clientes envían periódicamente mensajes al servidor para reportar las pérdidas de ficheros que han detectado, y a partir de esta información, el servidor determina la tasa de codificación AL-FEC óptima que permite reducir el tiempo que se requiere para que todos los terminales reciban todos los ficheros.

En los servicios de streaming 3GP-DASH sobre eMBMS, los problemas de fiabilidad que pueden surgir durante la transmisión de los segmentos multimedia sobre eMBMS se corrigen gracias al establecimiento de un procedimiento de recuperación de errores unicast. En este contexto, DASH no se utiliza como solución de streaming adaptativo, sino porque permite combinar el uso de las técnicas AL-FEC con la recuperación unicast de los segmentos multimedia. En general, la lógica de adaptación que implementan los clientes DASH trata de evitar que el buffer del reproductor se quede sin segmentos multimedia a través de la conmutación de las diferentes representaciones que se definen en el MPD. En un servicio de streaming 3GP-DASH sobre eMBMS estándar no se ejecuta ninguna lógica de adaptación, lo que se debe a que solo se transmite una representación multimedia sobre el canal multicast.

El Capítulo 6 propone codificar el contenido que se va a transmitir sobre eMBMS en múltiples calidades. Solo se transmite una representación sobre eMBMS, pero las demás están disponibles para ser solicitadas durante la recuperación unicast de los segmentos perdidos. La hipótesis de partida es que de esta forma es posible reducir el retardo y evitar las paradas en la reproducción. Para ello, es necesario diseñar un mecanismo que permita al cliente seleccionar la representación multimedia más adecuada al ancho de banda disponible en el canal unicast.

MOS	Calidad	Degradación de la calidad
1	Mala	Muy molesta
2	Pobre	Molesta
3	Razonable	Algo molesta
4	Buena	Perceptible pero no molesta
5	Excelente	Imperceptible

Tabla 2.1: Relación entre las puntuaciones de la escala MOS y la calidad percibida

Este mecanismo debe tener en cuenta que el ancho de banda del canal unicast es variable, al contrario de lo que ocurre en eMBMS, donde el ancho de banda permanece constante durante toda la sesión.

El uso de técnicas pasivas para la estimación del ancho de banda (subsección 2.1.2) no parece muy adecuado en este contexto, puesto que la mayor parte de los segmentos se recibe correctamente sobre eMBMS y el cliente no dispone de suficiente información para realizar una estimación precisa. Una solución alternativa es utilizar medidas de tráfico de otras aplicaciones que se ejecutan en el terminal (es decir, aplicar técnicas activas). Sin embargo, esta solución requiere que el reproductor DASH se comuniquen con otras aplicaciones, lo que aumenta su complejidad.

El streaming adaptativo basado en SAND (subsección 2.1.3) puede resolver esta problemática. Sin embargo, no se han identificado trabajos en el estado del arte que evalúen estos mecanismos de streaming en el contexto de los servicios de 3GP-DASH sobre eMBMS. La propuesta que se detalla en el Capítulo 2 de esta tesis doctoral se basa en el enfoque de Essaili *et al.* [ESS⁺15], que propone mejorar la adaptación de los reproductores utilizando un proxy que ejecuta un algoritmo de adaptación a partir de información del ancho de banda asignado a los terminales.

2.2. Evaluación de la calidad de la experiencia

La Unión Internacional de Telecomunicaciones (UIT) define la calidad de la experiencia o Quality of Experience (QoE) [ITU17] como la aceptabilidad de una aplicación o servicio, tal y como se percibe subjetivamente por los usuarios finales. Esto incluye cualquier efecto del sistema extremo a extremo (cliente, terminal, red o infraestructura del servicio). La aceptabilidad puede verse afectada por las expectativas y contexto del usuario.

La calidad de la experiencia puede medirse a través de la realización de encuestas que muestran de forma directa como valoran los usuarios el servicio. Las valoraciones se presentan en forma de puntuaciones con una escala conocida como Mean Opinion Score (MOS). La Tabla 2.1 muestra la escala de puntuaciones y su significado en términos de calidad percibida.

La valoración que realizan los usuarios es subjetiva, por lo que estos modelos son comúnmente llamados modelos subjetivos para la evaluación de la QoE. Esta metodología de evalua-

ción es costosa, entre otras cosas, porque se requiere mucho tiempo para preparar y realizar las encuestas. Otra desventaja es que estos modelos no se pueden utilizar para evaluar la calidad percibida por los usuarios en tiempo real. Por esta razón, ha crecido el interés en modelos que utilizan mediciones de parámetros objetivos del servicio para estimar la calidad percibida, esto es, modelos objetivos para la evaluación de la QoE.

Los modelos objetivos de QoE se pueden clasificar en tres categorías [CWZ15]: Full Reference (FR), Reduced Reference (RR) y No Reference (NR). Los modelos de la categoría FR requieren acceso completo a la señal de vídeo original, antes de que esta haya sido codificada. La calidad percibida se mide cuantificando la diferencia que existe entre el contenido multimedia original y la versión distorsionada que recibe el reproductor. Dos ejemplos de este tipo de modelos son Peak Signal-to-Noise Ratio (PSNR) y Video Quality Metric (VQM). El PSNR calcula la distorsión que existe en cada pixel del fotograma. Para calcular el PSNR se pueden utilizar los componentes de luminancia o de crominancia, pero en general, es suficiente con realizar el cálculo a partir de la componente de luminancia porque la percepción visual es más sensible al brillo. El modelo VQM [PW04] analiza, además de la distorsión de estas componentes, otros factores como, por ejemplo, efectos de desenfoco o percepción no continua del movimiento de la imagen. Por esta razón, se considera que el modelo VQM correla mejor con el sistema visual humano que el PSNR.

Los modelos RR requieren acceso parcial al contenido multimedia original. Para estimar la calidad percibida, primero se extrae el valor de algunos parámetros del contenido original. Después, se obtiene el valor de estos mismos parámetros, pero de la versión distorsionada que recibe el cliente. Finalmente, se realiza una comparación que sirve para estimar la calidad percibida. Este tipo de modelos proporciona una estimación menos precisa que la que se obtiene con los modelos de tipo FR, pero tienen la ventaja de que se transmiten menos datos por la red para realizar la estimación.

Los modelos NR, al contrario que los modelos FR o RR, no requieren el acceso al contenido original. Estos modelos establecen una asociación directa entre la calidad percibida y diferentes parámetros de la red y de la aplicación. Este tipo de modelos es muy apropiado para analizar el impacto de parámetros del servicio que no están relacionados con la degradación de la calidad de la imagen, como por ejemplo, las interrupciones de servicio o el retardo inicial de buffering.

2.2.1. Métricas objetivas para evaluar la QoE de los servicios de streaming adaptativo sobre HTTP

El 3GPP ha definido un conjunto de métricas objetivas que pueden medirse para evaluar la QoE de los servicios de streaming multimedia que se despliegan sobre las redes móviles. Esto incluye servicios de streaming 3GP-DASH sobre eMBMS [3GP17a] y servicios de streaming DASH convencionales que solo utilizan conexiones unicast para el transporte de los segmentos multimedia [3GP17b]. Durante la revisión del estado del arte, se ha detectado que la terminología utilizada no siempre es consistente, es decir, se utilizan diferentes términos para definir la

misma métrica de QoE. A continuación, se indica la nomenclatura que se utilizará en el resto del documento para hacer referencia a las métricas de QoE que se han utilizado para evaluar las diferentes propuestas de la tesis doctoral.

- *Retardo inicial de buffering* [LBP17b] es el tiempo que transcurre desde que el cliente inicia la sesión de streaming, hasta que el buffer del reproductor almacena los segundos de vídeo que indica el parámetro *minBufferTime* del fichero MPD y la reproducción puede dar comienzo.
- *Tiempo de parada de servicio* [SES⁺15] es el tiempo que ha permanecido parada la reproducción como consecuencia de la ausencia de datos en el buffer del reproductor. Esto puede producirse porque los segmentos multimedia se han recibido más tarde del instante en que debían reproducirse. Normalmente, se mide como un ratio, es decir, como el porcentaje de tiempo que ha estado detenida la reproducción. Tanto la frecuencia como la duración de los eventos de parada afectan a la calidad percibida.
- *Inestabilidad* [JSZ14] es el número promedio de conmutaciones que se producen durante la reproducción del contenido. La conmutación entre diferentes reproducciones multimedia previene que el buffer del reproductor se vacíe en un servicio de streaming adaptativo basado en DASH. Este parámetro no aplica, por tanto, a los servicios de streaming 3GP-DASH sobre eMBMS estándar, dado que solo se envía una representación multimedia sobre el canal de eMBMS.
- *Calidad de vídeo promedio* [JSZ14] es el bitrate promedio normalizado con respecto al bitrate de la representación de mayor calidad.
- *Eficiencia* [JSZ14] es la capacidad del algoritmo de adaptación para ajustarse al ancho de banda disponible. Cuanto mayor es la diferencia entre el bitrate de la representación seleccionada y el ancho de banda disponible menor es la eficiencia.

2.2.2. Problema de investigación: modelo objetivo para evaluar la QoE de los servicios de streaming 3GP-DASH sobre eMBMS

El estado del arte muestra que se han propuesto diversos modelos de la categoría NR para evaluar la QoE de los servicios de streaming basados en DASH. Oyman y Singh [OS12] revisan la parte del estándar del 3GPP que se centra en los servicios de streaming unicast [3GP17b] para luego analizar el impacto de las paradas de servicio en la calidad percibida. El modelo que proponen no considera otros parámetros de influencia como la inestabilidad o el retardo inicial de buffering.

Claeys *et al.* [CLFT14] también han propuesto un modelo para evaluar la QoE de un servicio de streaming multimedia basado en DASH. Este modelo estima la calidad percibida a partir del tiempo de parada de servicio, la inestabilidad y la calidad de vídeo promedio. Dado que el

modelo no considera el retardo inicial de buffering su aplicación es limitada en el contexto de los servicios de streaming en directo, caracterizados por unos estrictos requisitos de retardo.

Hoßfeld *et al.* [HSBP13] proponen un modelo que relaciona el retardo inicial de buffering y la duración de los eventos de parada con una puntuación MOS. El modelo no considera otros parámetros objetivos como la calidad de vídeo promedio. Además, las puntuaciones están obtenidas a partir de encuestas realizadas para evaluar los servicios unicast basados en DASH, es decir, no se pregunta a los encuestados sobre servicios de streaming 3GP-DASH sobre eMBMS.

Rodriguez *et al.* [RRA⁺16] han propuesto un modelo que incorpora todos los parámetros que afectan a la QoE. Sin embargo, su evaluación solo considera vídeos codificados a una resolución igual o menor a 480p, cuando hoy en día las resoluciones de vídeo más comunes son 720p o 1080p. Por otro lado, su modelo tampoco ofrece la posibilidad de seleccionar otros parámetros relacionados con la codificación del contenido, como el codec de vídeo usado o el tipo de perfil. El perfil se define como el conjunto de técnicas de codificación y algoritmos de compresión que utiliza el codec para generar el flujo multimedia [WSBL03a].

También se han propuesto modelos para evaluar la QoE de otro tipo de servicios, como servicios de streaming convencionales sobre HTTP, servicios de IPTV o servicios que se basan en el envío de ficheros multimedia sobre eMBMS. Sin embargo, ninguna de estas propuestas se basan en DASH ni tampoco están orientadas a los servicios de difusión de contenidos multimedia en tiempo real considerando los mecanismos de recuperación de errores que define el 3GPP.

En los servicios de IPTV, la QoE se ve seriamente afectada por la pérdida de paquetes, que provoca distorsiones en los fotogramas que recibe el reproductor [THB08]. En un servicio de streaming 3GP-DASH sobre eMBMS, la aplicación de las técnicas AL-FEC protege el envío del contenido multimedia frente a errores. Los segmentos se decodifican correctamente o no en función de los símbolos (paquetes) AL-FEC recibidos. Los segmentos que no se decodifican correctamente se recuperan por HTTP, lo que puede causar la parada del servicio si el buffer del reproductor se queda vacío antes de que haya finalizado el proceso de recuperación. Pero, los fotogramas que componen los segmentos no se ven afectados por la pérdida de paquetes o el retardo extremo a extremo. Por esta razón, los modelos diseñados para evaluar los servicios de IPTV no son adecuados para mediar la calidad de la experiencia de los servicios de streaming 3GP-DASH sobre eMBMS. Por el mismo motivo, tampoco resulta adecuado utilizar ningún modelo perteneciente a la categoría FR como, por ejemplo el PSNR. Estos modelos no consideran el tiempo de parada de servicio, se limitan a realizar comparaciones pixel-a-pixel de los fotogramas de vídeo para detectar distorsiones en la imagen.

La UIT está trabajando en la estandarización de un nuevo modelo objetivo de QoE de tipo NR que pueda aplicarse a cualquier servicio de streaming multimedia con independencia de la infraestructura de red que le de soporte. La propuesta que ha realizado Huawei para este modelo, llamada User, Unified, Ubiquitous-Mean Opinion Score for Video (U-vMOS) [HUA], obtuvo la aprobación de la UIT y es clave para la definición del nuevo estándar. Este modelo

estima la calidad percibida a partir del valor de diferentes parámetros de codificación del vídeo, así como de medidas de las métricas de QoE definidas en la subsección 2.2.1.

Debido a que se ha adoptado un marco muy general para diseñar este modelo de evaluación, es posible aplicarlo en el marco de los servicios de streaming 3GP-DASH sobre eMBMS, tal y como se discute en el Capítulo 5. De esta forma, ha sido posible realizar un análisis que puede utilizarse para definir diferentes parámetros del servicio (red y aplicación) desde la perspectiva de la QoE.

2.3. Arquitectura de la red LTE para el soporte de los servicios de difusión de contenidos

La infraestructura de la red móvil, basada en la conmutación de paquetes desde la tercera generación, Universal Mobile Telecommunications System (UMTS), evolucionó hacia la red móvil LTE para dar respuesta a la creciente demanda de servicios de datos y afrontar las limitaciones de capacidad que se vislumbraban ante dicha demanda [COLM11]. La aparición de los smartphones también incentivó la demanda de los servicios de datos, y en concreto de los servicios multimedia, básicamente, debido a que las características técnicas de los terminales hacían posible que los usuarios pudiesen visualizar imágenes y reproducir vídeos en cualquier momento y en cualquier lugar.

Sin embargo, se ha detectado que las redes móviles LTE no son adecuadas para desplegar servicios novedosos que presentan unos requisitos de latencia y ancho de banda muy estrictos como, por ejemplo, servicios de telemedicina, el Internet táctil o aplicaciones de realidad aumentada/realidad virtual. Esto se debe a que las redes LTE proporcionan poca flexibilidad para desplegar ágilmente nuevos servicios. La funcionalidad de los elementos de la red se asigna de forma estática y, además, existe una fuerte dependencia funcional entre los diferentes elementos [RBB⁺16]. La aparición de nuevos paradigmas de red, como la computación en el borde de la red, las redes definidas por software o la virtualización de las funciones de red aportarán a las redes 5G la capacidad necesaria para programar los elementos de la red de forma que se satisfagan los requisitos de latencia, ancho de banda y fiabilidad que demandan las aplicaciones.

Esta sección proporciona una descripción de la arquitectura de la red LTE, poniendo énfasis en los elementos que dan soporte a los servicios eMBMS. Más adelante, la sección 2.4 discute la evolución de esta arquitectura teniendo en cuenta soluciones que proponen adaptar los servicios eMBMS a los paradigmas de red mencionados anteriormente.

2.3.1. Arquitectura de la red móvil LTE

La Figura 2.4 muestra la arquitectura de referencia de la red móvil LTE. Incluye, todas las entidades funcionales que componen la red, es decir, las entidades que dan soporte a los servicios unicast y las entidades que se han definido específicamente para dar soporte a los servicios eMBMS. Como se puede observar, la red de acceso radio o Evolved Universal Terrestrial Radio Access Network (E-UTRAN) se compone únicamente por estaciones base o evolved Node B

(eNB)s. En LTE, las estaciones base integran toda la funcionalidad de la red de acceso, encargándose tanto de la transmisión de la señal radio, como del control de la interfaz (lo que incluye la gestión de los recursos o la gestión de movilidad).

El núcleo de la red móvil o Evolved Packet Core (EPC) se compone del resto de entidades funcionales. A continuación, se proporciona una breve descripción de las principales entidades funcionales del EPC, haciendo especial énfasis en aspectos que son clave para el soporte de los servicios de streaming 3GP-DASH sobre eMBMS. La subsección finaliza profundizando en los procedimientos que debe invocar un proveedor de contenidos para definir este servicio.

2.3.2. Serving Gateway (S-GW)

La entidad Serving Gateway (S-GW) actúa como pasarela durante el intercambio de tráfico unicast entre la red de acceso radio y el EPC. En concreto, se encarga de encaminar el tráfico de bajada al eNB donde el terminal está conectado y el tráfico de subida al Packet Data Network Gateway (P-GW) que corresponda (ver 2.3.3). En procesos de traspaso o handover [ABG14] (entre eNBs) el S-GW actúa como punto de anclaje, es decir, gestiona la movilidad del usuario para que el tráfico de datos se encamine correctamente independientemente de la estación base a la que se conecte. Además, el S-GW es capaz de almacenar temporalmente los paquetes destinados a un usuario hasta que se completa el establecimiento de la portadora, así como de filtrar datos potencialmente peligrosos para evitar que estos lleguen al EPC [COLM11].

2.3.3. Packet Data Network Gateway (P-GW)

La entidad P-GW actúa como pasarela durante el intercambio de tráfico entre el EPC y una red de conmutación de paquetes externa. Es decir, se encarga de encaminar el tráfico de bajada

Figura 2.4: Arquitectura de referencia de la red móvil LTE. Incluye los elementos funcionales que dan soporte a los servicios multimedia sobre eMBMS.

al S-GW donde el terminal está registrado y el tráfico de subida a la red externa. Actúa como punto de anclaje cuando se produce un traspaso o handover entre la red LTE y una red no 3GPP, como por ejemplo WiMAX. Además, es responsable de asignar la dirección IP del terminal durante el establecimiento de la conexión y de garantizar que se cumplen las políticas de Quality of Service (QoS) definidas en el Policy and Charging Rules Function (PCRF), entidad de LTE que gestiona los servicios portadores de la red [COLM11].

La Figura 2.4 muestra que el punto de referencia SGi conecta el P-GW con el Broadcast Multicast Service Centre (BM-SC) (ver 2.3.6). Como se explicará más tarde, el BM-SC es el punto de entrada para las transmisiones que realiza el proveedor de contenidos sobre eMBMS. A través del punto de referencia SGi es posible retransmitir datos enviados sobre eMBMS sobre conexiones unicast. Esto es útil para recuperar datos que no se han recibido correctamente sobre eMBMS debido a errores de transmisión. En el caso particular de un servicio de streaming 3GP-DASH sobre eMBMS, el cliente puede solicitar al BM-SC los segmentos que no se han transmitido correctamente sobre eMBMS utilizando HTTP. Estos se entregarán a través del P-GW y del S-GW que corresponda.

2.3.4. Mobility Management Entity (MME)

Mobility Management Entity (MME) es la principal entidad de control de las redes LTE. En lo que respecta a eMBMS, se encarga de transmitir mensajes de control a las estaciones base para notificarles el inicio o parada de la sesión eMBMS. También realiza otras funciones, como, por ejemplo:

- Participa en el intercambio de mensajes de señalización que se generan en los diferentes procedimientos que se llevan a cabo entre el EPC y los terminales móviles, por ejemplo, durante la conexión del terminal en la red.
- Gestión de la movilidad, lo que incluye el seguimiento de la posición del terminal o la transmisión del contexto del usuario durante los procesos de handover. El contexto del usuario incluye, entre otros parámetros: datos de la localización del usuario, identificadores y claves de seguridad [ABC⁺10].
- Control de acceso. Utiliza los datos almacenados en la entidad Home Subscriber Server (HSS), una base de datos que almacena los datos de suscripción de los usuarios, en los procesos de autenticación e intercambio de claves que se llevan a cabo para permitir o denegar el acceso de los usuarios a la red.
- Selección de los nodos P-GW y S-GW a partir de la localización del terminal y de información de carga de la red.

2.3.5. Multimedia Broadcast Multicast Service Gateway (MBMS GW)

Multimedia Broadcast Multicast Service Gateway (MBMS GW) es un elemento funcional definido específicamente para dar soporte a los servicios eMBMS. Actúa como pasarela, transmitiendo los datos de eMBMS a los eNBs que pertenecen al área donde se despliega el servicio. Esta transmisión se basa en IP multicast.

Por otro lado, el MBMS GW utiliza el punto de referencia Sm (Figura 2.4) para mandar mensajes de control a la entidad MME de forma que se señalice, hacia la red de acceso radio, el inicio/parada de la sesión eMBMS o la actualización de sus parámetros. Algunos parámetros de la sesión eMBMS son [3GP17c, Rue13]: la duración de la sesión eMBMS, parámetros de QoS o parámetros relacionados con la formación del área eMBMS.

2.3.6. Broadcast-Multicast Service Centre (BM-SC)

El BM-SC es un elemento que también ha sido definido específicamente para dar soporte a los servicios eMBMS. Actúa como punto de entrada de las transmisiones que realiza el proveedor de contenidos sobre eMBMS. Se conecta al MBMS GW a través de los puntos de referencia SGmb y SGi-mb. El punto de referencia SGi-mb transporta el tráfico de datos, mientras que el SG-mb se utiliza para configurar la sesión eMBMS y gestionar los grupos multicast. Es decir, el SG-mb se utiliza para transportar los mensajes de control que indican al MBMS GW el inicio/final de la sesión eMBMS o la actualización de sus parámetros.

El BM-SC proporciona diversas funciones para dar soporte al envío de datos sobre eMBMS (Figura 2.5), y en particular, para dar soporte a los servicios de streaming 3GP-DASH sobre eMBMS. A continuación, se resumen estas funciones [3GP17c]:

- Función de pertenencia: se encarga de proporcionar autorización a los usuarios que solicitan activar el servicio. También puede almacenar datos de los usuarios suscritos a los servicios eMBMS o generar registros para la tarificación del servicio.

Figura 2.5: Composición interna del BM-SC. La figura muestra las diferentes funciones (Func.) que realiza el BM-SC.

- Función de sesión y transmisión: se trata de un conjunto de funciones que se pueden agrupar en dos categorías. La primera, llamada *MBMS Delivery Functions*, incluye funciones que se relacionan directamente con el envío de datos sobre eMBMS. Por ejemplo, incluye funciones que se encargan de la planificación de las transmisiones (y retransmisiones) de datos sobre sesiones eMBMS, o de etiquetar cada sesión para que los clientes puedan distinguirlas entre sí. La segunda categoría, llamada *Associated Delivery Functions*, incluye funciones de propósito general que pueden llevarse a cabo antes, durante o al finalizar la sesión eMBMS.

Los servicios de streaming 3GP-DASH sobre eMBMS utilizan funciones de ambas categorías. Mediante una función de tipo *MBMS Delivery* se realiza el envío de los segmentos multimedia a los receptores multicast utilizando el protocolo FLUTE. Otra función de tipo *MBMS Delivery* se encarga de aplicar las técnicas AL-FEC para dar mas fiabilidad al envío del contenido sobre eMBMS. Por otro lado, y dado que algunos segmentos pueden perderse durante su envío sobre eMBMS, es posible utilizar una función de tipo *Associated Delivery* para establecer su recuperación por HTTP (a través del P-GW y el S-GW).

- Función de proxy: se trata de un agente proxy para la señalización entre el MBMS GW y otras funciones del BM-SC. Por ejemplo, los mensajes de control que genera la función de sesión y transmisión para reservar y liberar los recursos asignados a la sesión eMBMS, se envían a través del proxy, al MBMS GW. El encaminamiento de los mensajes de control es transparente para el MBMS GW.

El proxy también puede actuar como dispositivo intermedio en el plano de datos, es decir, encaminando los datos eMBMS que se envían desde la función de sesión y transmisión del BM-SC al MBMS GW.

Por otro lado, la función también podría ser capaz de generar registros para cobrar al proveedor de contenidos por la transmisión de datos sobre eMBMS.

- Función de anuncio del servicio: proporciona a los clientes la descripción del contenido multimedia que va a transmitirse sobre eMBMS. Esta descripción indica, por ejemplo, el tipo de codec que se ha utilizado para codificar el audio y el video. También proporciona a los clientes otra información relacionada con la sesión eMBMS, como el instante de inicio de la transmisión o una identificación del servicio.
- Función de seguridad: funciones que pueden utilizarse para garantizar la integridad y/o confidencialidad de los datos transmitidos sobre eMBMS. Esta función se encarga de transmitir las claves de seguridad a los terminales que se conectan al servicio.

2.3.7. Multi-cell/Multicast Coordination Entity (MCE)

La entidad Multi-cell/multicast Coordination Entity (MCE) se encarga del control de admisión y de asignar los recursos radio que los eNBs del área MBSFN utilizan para enviar los

datos eMBMS. También se encarga de configurar la transmisión radio, por ejemplo, determinando el esquema de modulación y codificación (MCS) que se utiliza en la sesión eMBMS. El MCE puede integrarse como parte de la funcionalidad del eNB o desplegarse como un sistema autónomo.

2.3.8. Definición de los servicios de streaming 3GP-DASH sobre eMBMS

Tras describir los elementos funcionales que componen las redes LTE es posible explicar cómo se integran los mecanismos y protocolos de streaming que dan soporte a los servicios 3GP-DASH sobre eMBMS en estas redes. La Figura 2.6 muestra dicha integración. El BM-SC incluye dos funciones de tipo *MBMS Delivery*: la primera se encarga de aplicar las técnicas AL-FEC para proteger el envío de los segmentos. La segunda de la encapsulación de los segmentos como ficheros y su transporte sobre el canal del eMBMS mediante el uso de un servidor FLUTE. Por otro lado, el BM-SC incluye una función de tipo *Associated Delivery* que se encarga de la recuperación unicast de los segmentos perdidos sobre eMBMS. Esta recuperación se realiza a través de un servidor HTTP.

Los segmentos generados por el servidor de contenidos se transmiten desde el BM-SC, y a través del MBMS GW, a todos los clientes conectados al servicio. La transmisión de los segmentos sobre el canal eMBMS no está libre de errores. El receptor intenta decodificar los segmentos multimedia a partir de los símbolos AL-FEC recibidos. Si la decodificación se lleva a cabo con éxito, el segmento se almacena en una memoria cache intermedia hasta que el reproductor lo solicita. Si el proceso de decodificación falla, el segmento perdido se recupera por HTTP gracias al servidor web que se encuentra instalado en el BM-SC. La solicitud HTTP viaja en este caso a través del S-GW y el P-GW hasta el BM-SC. La respuesta viaja en el camino inverso.

La Figura 2.6 también muestra que el proveedor de contenidos y el BM-SC se conectan entre sí a través del punto de referencia xMB. Utilizando este punto de referencia, el proveedor de contenidos puede invocar procedimientos del BM-SC que permiten definir, configurar y gestionar el servicio. La invocación de estos procedimientos se realiza sobre la interfaz xMB-C, del plano de control. El plano de control puede terminar opcionalmente en un Service Creation En-

Figura 2.6: Arquitectura híbrida DASH/FLUTE en LTE. La figura muestra la integración de los elementos que dan soporte a los servicios de streaming 3GP-DASH sobre eMBMS en la arquitectura de referencia de las redes LTE.

vironment Function (SCEF) que expone al proveedor de contenidos los servicios y capacidades de las interfaces de red a través de un API.

El punto de referencia xMB proporciona al proveedor de contenidos capacidad para autenticar y autorizar al BM-SC, crear o modificar un servicio/sesión y entregar contenido al BM-SC para que sea enviado sobre eMBMS. Por su parte, xMB proporciona al BM-SC capacidad para autenticar y autorizar al proveedor de contenidos, notificar a este sobre el estado del uso del servicio y recibir el contenido del proveedor. El contenido se transmite a través de la interfaz xMB-U, del plano de datos.

El estándar del 3GPP [3GP17a] define los procedimientos que permiten crear, actualizar y terminar una sesión o servicio. También define los parámetros de sesión que se pueden configurar para cada servicio. Se debe tener en cuenta que eMBMS no solo da soporte a los servicios de difusión de contenidos, sino que también da soporte a otros servicios como, por ejemplo, de actualización de software. A continuación, se describen los parámetros de sesión que se pueden configurar para definir un servicio de streaming 3GP-DASH sobre eMBMS [3GP17a]:

- Session start: tiempo en que se activa la portadora eMBMS para transmitir datos.
- Sesión stop: tiempo en que se desactiva la portadora eMBMS.
- Max Bitrate: tasa de datos máxima para el transporte del contenido multimedia sobre eMBMS. Se debe definir teniendo en cuenta el bitrate utilizado para codificar el vídeo, así como la sobrecarga que impone la aplicación de las técnicas AL-FEC y las cabeceras que se añaden para el transporte de los datos.
- Max Delay: especifica el retardo máximo que puede existir desde que el BM-SC recibe un paquete hasta que este se entrega al cliente.
- Geographical Area: información relacionada con el área geográfica donde se despliega el servicio. El estándar no define el contenido de este parámetro, que depende del acuerdo particular que establezcan los proveedores de contenidos y el operador de red.
- QoE Reporting: lista de métricas de QoE que pueden ser monitorizadas por el operador de la red para ser reportadas al proveedor de contenidos si este lo desea. La lista de métricas depende del tipo de servicio que se despliegue sobre eMBMS. En el caso de los servicios 3GP-DASH sobre eMBMS, la lista es la siguiente: número de objetos perdidos en la sesión FLUTE durante un intervalo de medida; número de símbolos perdidos en un bloque AL-FEC (i.e., de un segmento) que no ha podido decodificarse correctamente.
- Session Type: indica el tipo de servicio que se va a desplegar sobre eMBMS. Puede adquirir uno de estos valores: streaming (RTP), file, application y transport-mode. Los servicios de streaming 3GP-DASH sobre eMBMS requieren que el valor de Session Type sea application. Solo en ese caso, el BM-SC espera recibir del proveedor de contenidos el fichero MPD y los segmentos multimedia.

- Application Service Description: tipo MIME. Los servicios de streaming 3GP-DASH sobre eMBMS requieren el valor *application/dash+xml*.
- Ingest Mode: indica cómo se transmiten los segmentos multimedia al BM-SC a través de la interfaz xMB-U. Existen dos modos: Pull y Push. Cuando se selecciona el modo Pull, el BM-SC obtiene bajo demanda los segmentos multimedia tan pronto como estén disponibles en el servidor de contenidos. Cuando se selecciona el modo Push, el proveedor de contenidos no espera a que el BM-SC solicite los segmentos multimedia. Tan pronto como estos se generen, se envían al BM-SC.
- Application Entry Point URL: hace referencia a una Uniform Resource Locator (URL) que apunta al fichero MPD. Si se ha seleccionado el modo Pull como Ingest Mode, el BM-SC utiliza esta URL para descargar el MPD mediante una petición HTTP. Si se ha seleccionado el modo Push, el MPD se carga directamente en el BM-SC.
- Unicast Delivery: indica si el contenido multimedia está disponible para poder ser recuperado por unicast. Si su valor es *false* la recuperación unicast de los segmentos no se puede llevar a cabo.

2.4. Arquitecturas de red móvil basadas en la virtualización de las funciones de red

Las arquitecturas de red móvil actuales deben sufrir dos transformaciones fundamentales para satisfacer los requisitos de retardo, ancho de banda y fiabilidad que demandan las futuras aplicaciones [ITU18]. La primera transformación consiste en separar el plano de datos y control de las entidades del núcleo de la red. El 3GPP denomina esta separación como Control and User plane Separation (CUPS) [3GP18b]. La segunda transformación consiste en incorporar paradigmas de red emergentes como la virtualización de las funciones de red (NFV) [CCW⁺12], las redes definidas por software (SDN) [NMN⁺14] o la computación en el borde de la red o Multi-access Edge Computing (MEC) [MB17]. A lo largo de esta sección se explica con más detalle que implican estos procesos de transformación.

2.4.1. Introducción a las redes definidas por software y la virtualización de las funciones de red

Existen diversas razones [GB14] que motivan la aparición de las redes definidas por software y la virtualización de las funciones de red, como, por ejemplo, los altos costes de operación y mantenimiento de las redes actuales, los problemas crecientes en la infraestructura de los centros de datos y la necesidad de acelerar la innovación para satisfacer los requisitos de calidad de servicio.

2.4.1.1. Virtualización de las funciones de red (NFV)

La virtualización de las funciones de red propone desacoplar la implementación software de las funciones de red del hardware sobre el que se ejecutan. Esto permite sustituir el hardware propietario por hardware de propósito general para reducir costes. La virtualización también proporciona mecanismos que permiten flexibilizar el despliegue de nuevos servicios, así como mecanismos que permiten asignar bajo demanda los recursos de almacenamiento, cómputo y red disponibles a las funciones de red. La implementación virtual de una función de red se denomina función de red virtualizada o Virtualised Network Function (VNF). Las VNFs pueden estar compuestas por uno o más elementos software o VNF Component (VNFC) [ETS14a]. Por ejemplo, la implementación virtual del BM-SC podría descomponerse en cinco VNFCs, uno por cada una de las funciones que se definen en la subsección 2.3.6: seguridad, proxy, sesión y transmisión, pertenencia y anuncio del servicio.

Un servicio de red se define como una cadena de funciones de red virtualizadas. El organismo de estandarización European Telecommunications Standards Institute (ETSI) describe un entorno de gestión y orquestación llamado NFV Management and Orchestration (NFV-MANO) [ETS14b] que permite desplegar y gestionar los servicios de red basados en NFV. Para ello, NFV-MANO también debe gestionar los recursos de cómputo, almacenamiento y red de la infraestructura NFV o NFV Infraestructura (NFVI) del operador. El estándar especifica las operaciones que permiten gestionar el ciclo de vida de las VNFs y los servicios de red, así como las operaciones que permiten asignar dinámicamente los recursos de la NFVI a las funciones de red.

La Figura 2.7 muestra las entidades funcionales, puntos de referencia y repositorios de datos que componen NFV-MANO. A continuación, se proporciona una descripción de los elementos más importantes:

Figura 2.7: Plataforma de gestión y orquestación de servicios de red basados en NFV. La figura muestra las entidades funcionales, puntos de referencia y repositorios de datos que componen NFV-MANO.

- VNF Manager (VNFM): es la entidad responsable de gestionar el ciclo de vida de las VNFs. La mayoría de las funciones que realiza son genéricas, es decir, aplican a cualquier VNF. Sin embargo, el estándar de NFV-MANO también indica que el VNFM puede realizar funciones específicas para gestionar una VNF en particular. El comportamiento operacional de una VNF se define en una plantilla o descriptor llamado VNF Descriptor (VNFD), que especifica, entre otras cosas, qué componentes VNFC definen la VNF, qué imagen utiliza la máquina virtual o cuánta memoria/número de CPUs se asignan a la instancia. Durante el proceso de instanciación, se asignan recursos de la NFVI a la VNF a partir de estos datos. En definitiva, el VNFM utiliza este descriptor para instanciar la VNF y gestionar su ciclo de vida. Para crear o actualizar una VNF es necesario cargar en la plataforma de gestión un paquete VNF. Según el estándar, un paquete VNF está compuesto por el descriptor VNFD y la imagen de la máquina virtual.
- Virtualised Infrastructure Manager (VIM): entidad que gestiona los recursos de cómputo almacenamiento y red disponibles en la NFVI. Pueden existir VIMs especializados en la gestión de solo un tipo de recurso (por ejemplo cómputo) o VIMs capaces de gestionar recursos de diferentes tipos. El VIM también se encarga de gestionar los grafos, VNF Forwarding Graph (VNFFG)s, que definen la cadena de VNFs que componen el servicio de red. Para ello, crea y mantiene enlaces, redes, subredes y puertos virtuales. Un buen ejemplo para entender que tipo de funciones realiza un VIM es OpenStack¹, una plataforma de computación en la nube de código abierto que se puede utilizar para desplegar los contenedores y máquinas virtuales que componen las VNFs.
- Network Functions Virtualisation Orchestrator (NFVO): se encarga de gestionar el ciclo de vida del servicio de red, cuyo comportamiento operacional se define en el descriptor Network Service Descriptor (NSD). El despliegue del servicio requiere que el NFVO invoque los procedimientos que define el estándar para instanciar las VNFs que lo componen. Además, se deben encadenar las diferentes VNFs tal y como se especifica en el grafo del servicio (VNFFG). Este grafo se define en el descriptor VNFFG Descriptor (VNFFGD). El NSD hace referencia a los diferentes descriptors que definen el servicio, como por ejemplo, el VNFFGD, los descriptors de las VNFs o los descriptors de los enlaces virtuales o Virtual Link Descriptor (VLD) que las interconectan. Por otro lado, el NFVO es capaz gestionar los recursos de la NFVI a través del VIM (punto de referencia Or-Vi en la Figura 2.7). A través del VIM es capaz de asignar recursos de la NFVI a las VNFs que componen el servicio de red.
- Element Management (EM): se encarga de la gestión de FCAPS, término utilizado en el estándar para hacer referencia a las siguientes funciones: configuración de las funciones

¹“OpenStack,”[Online]. Available: <https://www.openstack.org/>

de red proporcionadas por la VNF, gestión de seguridad, gestión de fallos y monitorización de medidas del rendimiento y uso de la VNF.

- Operations Support System/Business Support System (OSS/BSS): sistemas de soporte a operaciones y negocios [CDLL15]. Mientras los sistemas OSS dan soporte a la provisión de servicios, la configuración de los elementos de la red o al mantenimiento de las redes de comunicaciones, los sistemas BSS se orientan a la tarificación de servicios, la tramitación de pedidos o al establecimiento de acuerdos con clientes.

2.4.1.2. Redes definidas por software (SDN)

Tradicionalmente, la actualización del software de los equipos de red ha estado sujeta a diversas limitaciones. Los fabricantes ofrecen actualizaciones, principalmente, para corregir errores detectados, y con menos frecuencia, para introducir nueva funcionalidad. La configuración de los equipos tampoco es muy flexible, ya que se basa en ficheros de configuración, el uso del protocolo Simple Network Management Protocol (SNMP) [CFSD90] o la interacción del administrador con el equipo a través de una consola y mediante el uso de un lenguaje definido por el propio fabricante. Otro factor clave es la rigidez que presentan los nodos de conmutación, que implementan de forma distribuida el plano de control mediante el uso de protocolos de encaminamiento. Esta rigidez limita la innovación en lo que respecta al despliegue de soluciones que permitan gestionar de una forma más eficiente los flujos de datos.

En este contexto surgen las redes definidas por software (SDN), que se presentan como un paradigma clave para acelerar la evolución de las redes de telecomunicaciones, y en particular, de las redes móviles. Las redes basadas en SDN disponen de mecanismos que facilitan la configuración y gestión de la red, lo que reduce el OPEX. Se trata de redes programables que permiten realizar un control de grano fino de los flujos de tráfico mediante la ejecución de aplicaciones de red.

Como se ha comentado, las funciones del plano de datos y del plano de control se implementan tradicionalmente en los nodos de conmutación. Las redes definidas por software proponen desacoplar ambos planos y centralizar el plano de control a nivel lógico. De esta forma, las funciones del plano de datos se pueden llevar a cabo en conmutadores de caja blanca o white box switches [MHT15], es decir, conmutadores donde existe una marcada separación e independencia entre el hardware y el software que ejecutan. El uso de este tipo de conmutadores reduce el CAPEX.

La arquitectura de las redes SDN se compone de tres capas, como se puede observar en la Figura 2.8. La capa de aplicación, como su nombre indica, es donde residen las aplicaciones de red o aplicaciones SDN, programas software diseñados para realizar funciones como, por ejemplo, de encaminamiento o de filtrado de paquetes. Las aplicaciones SDN se ejecutan sobre la capa de control, es decir, sobre un sistema operativo de red como ONOS² o OpenDaylight³.

²“ONOS,”[Online]. Available: <https://onosproject.org/>

³“OpenDaylight,”[Online]. Available: <https://www.opendaylight.org/>

Figura 2.8: Arquitectura de las redes SDN. Se compone de tres capas: capa de aplicación, capa de control y capa de infraestructura. La figura también muestra cómo se interconectan estas capas a través de las interfaces norte (NBI) y sur (SBI). También muestra cómo se comunican los conmutadores SDN a través de las interfaces este (EBI) y oeste (WBI).

Este sistema operativo o controlador SDN recibe instrucciones de las aplicaciones a través de la interfaz norte o Northbound Interface (NBI). El controlador procesa estas instrucciones y utiliza la interfaz sur o Southbound Interface (SBI) para programar las tablas de reenvío o Forwarding Information Database (FIB) de los conmutadores. Si el protocolo utilizado en la interfaz sur es OpenFlow [MAB⁺08], las tablas de reenvío reciben el nombre de tablas de flujo. Así, la funcionalidad definida por las aplicaciones se implementa en la red, sobre la capa de infraestructura. La figura también muestra la existencia de las interfaces este/oeste (WBI/EBI) que comunican controladores SDN desplegados en diferentes redes o que se utilizan como respaldo.

2.4.2. Arquitecturas de red móvil basadas en NFV y SDN

ETSI, organismo que está liderando las actividades de estandarización de NFV, ha identificado las posibles ventajas de virtualizar el núcleo de la red móvil en un documento que presenta diferentes casos de uso de NFV. En el caso de uso llamado “Virtualization of Mobile Core Network and IMS” [ETS13] se indica que estas ventajas son: mayor eficiencia del uso de la red, mayor disponibilidad del servicio, mayor capacidad para adaptarse a fallos y mayor escalabilidad.

Por otro lado, el 3GPP está abordando la definición de la arquitectura de referencia del núcleo de las redes 5G [3GP18c] y la separación del plano de datos y control (CUPS) de los nodos del EPC. En particular, el 3GPP ha especificado [3GP18b] como se puede aplicar CUPS a las pasarelas S-GW y P-GW, definiendo qué funciones realizan las entidades del plano de control (SGW-C, PGW-C) y qué funciones realizan las entidades del plano de datos (SGW-U,

PGW-U). La aplicación de CUPS hace posible implementar soluciones basadas en SDN para gestionar el tráfico de una forma más eficiente. También permite aplicar soluciones basadas en NFV, por ejemplo, para escalar bajo demanda los recursos de cómputo, almacenamiento y red que se asignan a las entidades de ambos planos si estos se virtualizan en la nube del operador (Capítulo 7).

El estado del arte muestra diversas propuestas para el diseño de una nueva arquitectura del núcleo de la red móvil basada en SDN y NFV [HSMA14, NBGT17, AKA⁺17]. Estas propuestas se pueden clasificar en cuatro enfoques diferentes. El primero, considera que los planos de datos y control están estrictamente acoplados (no aplica CUPS), y que solo se aplican soluciones de tipo NFV para reemplazar el hardware propietario por hardware de propósito general. Existen otros dos enfoques basados en la aplicación de CUPS y la combinación, en mayor o menor medida, de soluciones basadas en SDN y NFV. Ambos enfoques asumen que el plano de control está virtualizado, pero difieren en su propuesta para implementar el plano de datos (virtualizado o no). El cuarto enfoque, denominado en la literatura como “clean-slate”, propone re-diseñar el núcleo de la red sin tener en cuenta los estándares existentes. Este tipo de propuestas están, en general, más orientadas a definir la arquitectura de la red móvil desde la perspectiva de las SDN. Sin embargo, presentan problemas de interoperabilidad con la infraestructura de red ya desplegada.

Por otro lado, han surgido iniciativas que proponen transformar las centrales locales en centros de computación de datos desplegados en el borde de la red. El proyecto Central Office Re-architected as a Datacenter (CORD) [PSA⁺16] combina SDN, NFV y la computación en el borde de la red (MEC) para proporcionar a las centrales locales la capacidad de desplegar dinámicamente servicios que requieren altas necesidades de cómputo, bajos retardos o altos anchos de banda. Aunque inicialmente CORD se diseñó para su aplicación en redes residenciales (R-CORD) y de empresa (E-CORD), ha evolucionado para dar soporte también a las redes móviles (M-CORD). De acuerdo a la fundación Open Networking Foundation (ONF)⁴, M-CORD es un proyecto de código abierto que tiene como objetivo formular soluciones de referencia para el despliegue de las redes 5G. El proyecto está alineado con la propuesta del 3GPP para separar los planos de datos y de control de las entidades funcionales de la red, y aborda la virtualización tanto de la red de acceso radio como del EPC.

El estado del arte muestra que se han publicado otros trabajos que se centran en evaluar arquitecturas del núcleo de la red basadas en NFV/SDN y CUPS [BKH⁺14, BBH⁺17]. Estos trabajos determinan cual es el impacto de incorporar estos paradigmas en el retardo del plano de datos y la carga de la red. A partir de los resultados obtenidos en el análisis, proponen un modelo de optimización que busca encontrar la ubicación óptima de los centros de datos, VNFs y aplicaciones SDN con el objetivo de minimizar el retardo del plano de datos, la carga de la red, y hacer un uso más eficiente de los recursos disponibles en los centros de datos. Sin embargo, los autores de estos trabajos no consideran la virtualización de la red de acceso radio

⁴“ONF,”[Online]. Available: <https://www.opennetworking.org/>

ni la incorporación de elementos de red que permitan dar soporte a los servicios de difusión de contenidos.

2.4.3. Problema de investigación: evolución de la arquitectura de eMBMS combinando NFV y SDN

La evolución de la arquitectura del núcleo de la red móvil está siendo abordada desde diferentes prismas: organismos de estandarización, proyectos de iniciativa privada y la comunidad científica. Sin embargo, la mayoría de las arquitecturas propuestas no incluyen los elementos funcionales que define el 3GPP para el soporte de los servicios eMBMS. Aunque el 3GPP tiene un plan para abordar la evolución de eMBMS, actualmente está centrando sus esfuerzos en diseñar la arquitectura de red 5G para dar soporte a los servicios unicast.

Los operadores de red móvil han obtenido resultados que confirman que las transmisiones broadcast basadas en eMBMS proporcionan diversas ventajas [ALL]. Aunque la tecnología no ha penetrado con éxito en las redes LTE, su adaptación a nuevos paradigmas de red como NFV, SDN o MEC puede hacer crecer el interés de los operadores por desplegar servicios eMBMS sobre las redes 5G. Por ejemplo, gracias a la reducción del CAPEX y el OPEX se deben valorar nuevos modelos de negocio. Esta tesis contribuye a diseñar una nueva arquitectura de eMBMS considerando la combinación de estas dos tecnologías.

El estado del arte muestra algunas propuestas de diseño. Do *et al.* [DNK16] proponen una nueva arquitectura del núcleo de la red basada en SDN. Para ello definen las aplicaciones de red que deben ejecutarse sobre el controlador SDN para dar soporte a los servicios de difusión de contenidos. La funcionalidad que implementan las aplicaciones de red difiere de la que realizan las entidades lógicas definidas por el 3GPP en lo que respecta a su estructura organizativa. Por ejemplo, las funciones de control que realiza el BM-SC y el MBMS GW se agrupan según esta propuesta en una única aplicación SDN llamada Multicast Broadcast Handler. Debido a que la arquitectura propuesta no implementa las interfaces y entidades lógicas definidas por el 3GPP, plantea muchos desafíos para satisfacer la interoperabilidad con las redes existentes.

Chantre y Fonseca [CF17, CF18] definen una nueva arquitectura de eMBMS desde la perspectiva de NFV. Esta arquitectura mantiene las interfaces y entidades lógicas definidas por el 3GPP, por lo que puede interoperar con la infraestructura de red existente. Los autores proponen migrar toda la funcionalidad que ejecutan las entidades de eMBMS (datos y control) a máquinas virtuales que se despliegan en la nube del operador. Sin embargo, la propuesta no considera la aplicación de CUPS a los elementos funcionales de eMBMS, por lo que no está alineada con la visión del 3GPP. Por otro lado, los autores analizan en qué medida es necesario replicar las VNFs que componen los servicios eMBMS para satisfacer diferentes requisitos de fiabilidad. Este tipo de soluciones también podría aplicarse en el marco de las propuestas de la tesis doctoral.

El proyecto 5G-Xcast⁵, ejecutado en el seno del 5G Public-Private Partnership (5GPPP),

⁵“5G-Xcast,”[Online]. Available: <https://5g-xcast.eu/>

valora diferentes alternativas para re-diseñar la arquitectura de eMBMS. Las alternativas propuestas no profundizan en la combinación de SDN y NFV para dar soporte a los servicios de difusión de contenidos. El Capítulo 7 explica como ambas tecnologías pueden integrarse para dar soporte a estos servicios, y en concreto, para facilitar el despliegue de nuevos mecanismos de streaming que mejoren el rendimiento de los servicios 3GP-DASH sobre eMBMS.

2.5. Conclusiones

Tras revisar el estado del arte se ha detectado la necesidad de formular propuestas que permitan satisfacer los requisitos de retardo que demandan las retransmisiones en directo, es decir los servicios de streaming 3GP-DASH sobre eMBMS. Estas soluciones deben evitar, al mismo tiempo, otros problemas que pueden surgir durante la reproducción del contenido multimedia, como son las paradas de servicio.

De la revisión del estado del arte se deduce que las propuestas se pueden realizar con dos enfoques diferentes: el primero consiste en diseñar modelos analíticos que permitan dimensionar adecuadamente el buffer de los reproductores, es decir, modelos que permitan calcular el buffer mínimo que evita las paradas de servicio. El segundo consiste en proponer mecanismos que hagan posible la adaptación asistida por la red, de forma que sea posible conmutar entre diferentes representaciones multimedia durante la recuperación unicast de los segmentos multimedia.

Para evaluar las anteriores soluciones desde la perspectiva de la calidad de la experiencia, es necesario un modelo objetivo de QoE que tenga en cuenta todas las métricas que afectan a la calidad percibida por los usuarios. El estado del arte muestra la dificultad para analizar los servicios de streaming 3GP-DASH sobre eMBMS con esta perspectiva, lo que ha motivado la aplicación de un modelo objetivo que está en proceso de estandarización.

Por otro lado, y considerando la evolución de las arquitecturas de red móvil, resulta necesario adaptar la arquitectura de eMBMS a los cambios evolutivos que están sufriendo las redes de comunicaciones. Las ventajas que proporcionan las redes SDN, NFV y la computación en el borde de la red pueden motivar a los operadores a desplegar servicios basados en eMBMS. Debido a que los organismos de estandarización están centrando sus esfuerzos en la definición de los elementos de red que dan soporte a los servicios unicast, esta tesis contribuye a diseñar la evolución de la arquitectura de eMBMS en el marco de las arquitecturas de red móvil basadas en la virtualización de las funciones de red.

Capítulo 3

Análisis de los servicios de streaming multimedia sobre redes móviles LTE

Establecer una sesión de streaming 3GP-DASH sobre eMBMS implica determinar diferentes parámetros de red y aplicación. Determinar adecuadamente el valor de estos parámetros es necesario, por ejemplo, para proporcionar una buena QoE (Capítulo 5). Para analizar el impacto de los parámetros en el servicio y evaluar las diferentes propuestas que han surgido durante el desarrollo de la tesis, ha sido necesario desarrollar diferentes herramientas. Esto se debe a que el uso comercial de eMBMS es todavía muy reducido, lo que dificulta evaluar las propuestas en un entorno de experimentación real.

Para hacer frente a esta problemática se han usado dos enfoques. El primero se basa en el desarrollo de modelos analíticos y de simulación que permitan analizar los servicios de streaming 3GP-DASH sobre eMBMS. El estado del arte muestra que ninguna de las plataformas existentes [GGTE09, LM11, MZZ13, ZDL⁺14] está orientada al análisis de este tipo de servicios. Por este motivo, el capítulo presenta una herramienta que combina resultados de simulación obtenidos con el simulador de redes OPNET¹ y un modelo analítico desarrollado en MATLAB². La plataforma permite analizar el impacto de múltiples parámetros, tanto de red como de aplicación, en los servicios de streaming 3GP-DASH sobre eMBMS. Una de las ventajas de esta plataforma es que permite extraer resultados considerando una población de usuarios grande, una cuestión clave para analizar servicios basados en transmisiones broadcast.

El segundo enfoque se basa en el desarrollo de escenarios virtuales basados en software de código abierto. En este caso, la plataforma se ha diseñado con el objetivo de que se puedan analizar los servicios de streaming adaptativo sobre HTTP y los servicios de streaming 3GP-DASH sobre eMBMS. La transmisión del contenido sobre la red puede simularse conformando el tráfico multimedia a partir de medidas de ancho de banda, latencia y pérdidas. También es posible conectar esta plataforma a un piloto de demostración o red móvil real. De hecho, este capítulo muestra resultados obtenidos tras la integración de la plataforma con un piloto de demostración de red LTE de una conocida empresa tecnológica del sector.

La estructura del Capítulo es la siguiente: la Sección 3.1 describe los modelos de simulación

¹“Riverbed technologies,”[Online]. Available: <https://www.riverbed.com>

²“Matlab”, [Online]. Available: <http://mathworks.com/products/matlab>

y matemático que se han desarrollado. A continuación, se analiza el servicio de streaming 3GP-DASH sobre eMBMS utilizando esta plataforma. Los resultados obtenidos se presentan en la Sección 3.2. Después, se describe la plataforma basada en escenarios virtuales (Sección 3.3). En este caso, se presentan los resultados que se han obtenido tras integrar la plataforma con el piloto de demostración de red LTE (Sección 3.4). Finalmente, la sección Sección 3.5 proporciona las conclusiones de este trabajo.

Es necesario destacar que las plataformas de experimentación que se describen en este capítulo se han modificado posteriormente para evaluar y validar las diferentes propuestas surgidas durante la tesis. Las modificaciones realizadas se indicarán en los capítulos correspondientes.

3.1. Diseño de un modelo de simulación y un modelo analítico

El simulador de redes OPNET no implementa el servicio de streaming 3GP-DASH sobre eMBMS, es decir, no permite simular el envío unidireccional de ficheros con el protocolo FLUTE ni tampoco la recuperación de los segmentos perdidos mediante el uso de HTTP. Sin embargo, OPNET permite simular la transmisión eMBMS de tráfico genérico desde un servidor hasta N receptores multicast. Además, permite definir diferentes parámetros de red que afectan al servicio, como el Intersite Distance (ISD), el MCS, o el tamaño del área MBSFN.

Las simulaciones permiten calcular la tasa de error de paquetes o Packet Error Rate (PER) que existiría en el canal eMBMS durante la transmisión del contenido multimedia. Así, se ha obtenido el PER para diferentes despliegues del servicio eMBMS, es decir, para diferentes combinaciones de los parámetros de red.

La Tabla 3.1 muestra los parámetros del modelo de simulación. La duración de la simulación es superior a 300 s pero se han despreciado los resultados obtenidos durante el periodo transitorio y se utilizan únicamente los 300 s del periodo estacionario. Se ha considerado una población de 399 usuarios uniformemente distribuidos en áreas MBSFN de 1, 7 o 19 celdas hexagonales. Con respecto a la separación entre eNBs, se ha considerado una separación estándar (ISD) de 500 m para entornos urbanos y de 1732 m para entornos suburbanos. También se han seleccionado varios valores del índice del MCS para considerar diferentes niveles de protección frente a errores a nivel físico.

Se debe tener en cuenta que la tasa máxima de datos del servicio depende del MCS, pero también del ancho de banda del canal (10 MHz) y de la cantidad de recursos radio que se asignan al servicio eMBMS (una subtrama) [3GP17d]. En la Sección 3.2 se discute la relación que existe entre la tasa de datos del servicio y la robustez que proporciona el esquema MCS frente a errores.

Utilizando el simulador se ha obtenido el PER de todos los usuarios conectados a la transmisión eMBMS (399) para diferentes combinaciones de los parámetros del servicio. Por ejemplo, se ha obtenido el PER de los 399 usuarios cuando se selecciona un MCS 15, un ISD de 500 metros y un área MBSFN de 7 celdas.

Parámetro	Valor
Duración de la simulación	300 s
Número de terminales en el área eMBMS	399
Distribución de los usuarios	uniforme
Tamaño del área MBSFN	1 / 7 / 19 eNB
ISD	500 m / 1732 m
Forma geométrica de las celdas	Hexagonal
Número de subtramas asignadas al servicio eMBMS	1
Índice del esquema de modulación y codificación MCS	3 / 4 / 5 / 6 / 7 / 8 / 10 / 15 / 20 / 25 / 28
Ancho de banda del canal	10 MHz

Tabla 3.1: Parámetros del modelo de simulación

Los resultados de PER obtenidos alimentan un modelo analítico que simula el envío de los segmentos multimedia sobre eMBMS tal y como se define en el Capítulo 2. Si el segmento se ha enviado correctamente sobre eMBMS el proceso de decodificación se realiza con éxito y el reproductor puede obtener el segmento directamente de la cache. Si el proceso de decodificación falla, el cliente debe realizar una petición HTTP para recuperarlo utilizando un canal unicast. Así, la probabilidad de que un segmento se pierda durante su transmisión sobre eMBMS y se recupere por HTTP se ha modelado como la probabilidad de fallo del proceso de decodificación AL-FEC. Esta probabilidad se ha denotado como $P(f_{RC})$, y se calcula como

$$P(f_{RC}) = \sum_{n=0}^{k+r} P(f_{RC}|n) \times P(N = n) \quad (3.1)$$

donde $P(f_{RC}|n)$ es la probabilidad de fallo del decodificador AL-FEC en el supuesto de que se hayan recibido n símbolos, y donde $P(N = n)$ es, precisamente, la probabilidad de que se reciban n símbolos.

El calculo de $P(f_{RC}|n)$ es diferente para cada código AL-FEC. Stockhammer *et al.* [FA08] indican que en el caso del código Raptor (código estandarizado por el 3GPP), esta probabilidad puede calcularse como

$$P(f_{RC}|n) = \begin{cases} 1 & \text{if } n < k \\ 0,85 \times 0,567^{n-k} & \text{if } n \geq k \end{cases} \quad (3.2)$$

En caso de utilizar el código RaptorQ, evolución de Raptor, $P(f_{RC}|n)$ puede calcularse como:

$$P(f_{RC}|n) = \begin{cases} 1 & \text{if } n < k \\ 0,01 \times 0,01^{n-k} & \text{if } n \geq k \end{cases} \quad (3.3)$$

Como se puede observar, el código RaptorQ tiene una $P(f_{RC}|n)$ menor que la de su predecesor, pero además, cabe destacar que la complejidad computacional del algoritmo también es menor [BKKP13].

Por otro lado, y dado que cada símbolo se encapsula en un paquete FLUTE, la probabilidad de que se reciban n símbolos se puede calcular a partir de la tasa de error de paquetes (PER). Como muestra la ecuación 3.4, esta probabilidad se ha modelado con una distribución binomial:

$$P(N = n) = \binom{k+r}{n} \times (1 - PER)^n \times (PER)^{(k+r-n)} \quad (3.4)$$

El modelo analítico utiliza los resultados de PER obtenidos en las simulaciones realizadas con OPNET para calcular $P(N = n)$.

Finalmente, y para determinar si el receptor ha decodificado correctamente un segmento, se compara $P(f_{RC})$ con valores de una variable aleatoria v_r que sigue una distribución uniforme. Si $v_r > P(f_{RC})$, se considera que el receptor ha decodificado correctamente el segmento. En caso contrario, se considera que el proceso de decodificación ha finalizado sin éxito [LGSW07].

La plataforma permite analizar la transmisión de los segmentos multimedia sobre eMBMS. Es decir, permite extraer el porcentaje de segmentos que se han transmitido correctamente sobre eMBMS y el porcentaje de segmentos que deben ser recuperados por HTTP. Este análisis se realiza para todos los usuarios conectados al servicio y para diferentes combinaciones de los parámetros de red y aplicación.

Sin embargo, los modelos que se han presentado son insuficientes para analizar como afecta la recuperación unicast de los segmentos multimedia en la calidad percibida por los usuarios. Por este motivo, también se ha desarrollado en MATLAB un simulador de un reproductor DASH. El simulador permite medir varias métricas objetivas de QoE (Capítulo 2), como, por ejemplo, el tiempo de parada del servicio, el retardo inicial de buffering o el bitrate de la representación solicitada (que puede variar si se implementa el mecanismo de recuperación adaptativo propuesto en el Capítulo 6).

El simulador asume que la reproducción comienza cuando el buffer ha almacenado m segmentos, tal y como especifica el campo *minBufferTime* del MPD [ISO14]. Si el modelo analítico determina que un segmento se ha recibido correctamente sobre eMBMS ($v_r > P(f_{RC})$) el nivel del buffer del reproductor aumenta en T_{sgmnt} s, es decir, la duración del segmento multimedia. Sin embargo, si un segmento no se ha recibido correctamente ($v_r < P(f_{RC})$) el buffer se descarga en proporción al retardo que añade la recuperación del segmento por HTTP, y solo cuando se completa la descarga, el nivel del buffer aumenta T_{sgmnt} s. Para modelar el retardo que introduce la recuperación unicast de los segmentos se puede especificar el valor del ancho de banda disponible en el canal unicast.

3.2. Análisis del servicio a partir del modelo de simulación y el modelo analítico

Utilizando el modelo de simulación y el modelo analítico se ha analizado el impacto de diferentes parámetros, tanto de red como de aplicación, en el servicio. Esta sección discute:

- la relación entre la tasa de codificación AL-FEC y la cobertura del servicio,
- la relación entre la tasa de datos y la cobertura del servicio
- y el rendimiento de los códigos Raptor y RaptorQ.

3.2.1. Relación entre la tasa de codificación AL-FEC y la cobertura del servicio

Sea la cobertura del servicio el porcentaje de segmentos entregados correctamente sobre eMBMS, la Figura 3.1 muestra la relación entre la tasa de codificación AL-FEC y la cobertura del servicio para diferentes tamaños del área MBSFN y diferentes valores del índice MCS (ver Tabla 3.1).

En concreto, la Figura 3.1a muestra los resultados obtenidos para un área MBSFN de 7 celdas, un ISD de 500 m y diferentes valores del índice MCS. Los resultados muestran que el porcentaje de segmentos entregados correctamente sobre eMBMS aumenta para valores más robustos del MCS y valores de la tasa de codificación AL-FEC más bajos (es decir, cuando el codificador añade más símbolos de redundancia).

Por otro lado, la Figura 3.1b analiza el impacto del tamaño del área MBSFN en la cobertura del servicio. Los resultados se presentan para el índice 15 del MCS y para áreas MBSFN de 1, 7 y 19 celdas. La Figura 3.1b muestra que si no se utiliza MBSFN (área de 1 celda) solo se puede alcanzar algo más del 80 % de cobertura. Este valor aumenta por encima del 95 % cuando el área MBSFN es de 7 celdas. Los resultados también muestran que la mejora de

(a) Diferentes valores del índice del MCS

(b) Diferentes tamaños del área MBSFN

Figura 3.1: Relación entre la tasa de codificación AL-FEC y la cobertura del servicio para diferentes valores del índice MCS y del tamaño del área MBSFN.

Figura 3.2: Relación entre tasa de datos y cobertura del servicio para diferentes valores del índice MCS y del área MBSFN.

cobertura que se obtiene utilizando áreas MBSFN de 19 celdas no es significativa a no ser que la tasa de codificación AL-FEC sea cercana a 1, es decir, cuando el nivel de protección que se ofrece a nivel de aplicación es bajo. Además, se debe tener en cuenta que la complejidad para sincronizar la transmisión de los eNBs crece para áreas MBSFN más grandes [FLB⁺ 15].

3.2.2. Relación entre la tasa de datos y la cobertura del servicio

La aplicación de las técnicas AL-FEC permite proporcionar más fiabilidad a la transmisión del contenido multimedia sobre eMBMS. Sin embargo, existe un compromiso entre la cobertura y la tasa máxima de datos del servicio. El ancho de banda del canal eMBMS es fijo debido a que el MCS y la cantidad de recursos radio asignados al canal no varía durante la sesión. Como consecuencia, añadir redundancia AL-FEC implica reducir la tasa de datos del servicio. Como se discute a continuación, la relación entre la tasa de datos y la cobertura se ve afectada por diversos parámetros del servicio: el MCS, el tamaño del área MBSFN, el ISD y la duración del segmento multimedia.

- Impacto del MCS y el tamaño del área MBSFN

La Figura 3.2 muestra la relación entre la tasa de datos y la cobertura del servicio para diferentes valores del índice MCS y del tamaño del área MBSFN. La Figura 3.2a muestra los resultados obtenidos para un área MBSFN de 7 eNBs, un ISD de 500 m y diferentes valores del índice del MCS. El uso de esquemas MCS menos robustos (el índice 15 en la figura) permite obtener tasas de datos mayores a costa de reducir la cobertura del servicio. Aunque el porcentaje de retransmisiones unicast se reduce para valores más bajos del índice MCS, la tasa de datos del servicio también se reduce de forma significativa.

Este análisis considera que solo se ha asignado una subtrama (de las diez disponibles) al servicio eMBMS. Podrían obtenerse tasas de datos mayores si se asignan más recursos

(a) Diferentes valores del ISD

(b) Diferentes valores de T_{sgmnt}

Figura 3.3: Tasa de datos del servicio para diferentes valores del ISD y de la duración del segmento multimedia T_{sgmnt} .

radio (i.e., más subtramas) al servicio. Según los estándares del 3GPP [3GP18d], es posible asignar hasta un máximo de seis subtramas a las transmisiones broadcast basadas en MBSFN.

- Impacto del ISD y la duración del segmento multimedia

La Figura 3.3 muestra la relación entre la tasa de datos y la cobertura del servicio para diferentes valores del ISD y de la duración del segmento multimedia (T_{sgmnt}). En concreto, la Figura 3.3a muestra que la cobertura del servicio en entornos suburbanos (ISD de 1732 m) es pobre en comparación con la que se obtiene en entornos urbanos (ISD de 500 m) para el mismo valor del índice MCS. Por ejemplo, para un MCS 15, la cobertura del servicio en un entorno suburbano es inferior al 50 %, mientras que en un entorno urbano es superior al 95 %. Por otro lado, la Figura 3.3b analiza el impacto de la duración del segmento multimedia. Los resultados se muestran para un área MBSFN de 7 celdas, un ISD de 500 m y el índice 15 del MCS. La duración típica de un segmento multimedia es de 2 a 10 s pero también se ha evaluado el rendimiento del decodificador AL-FEC para segmentos de muy larga duración (20 s). Analizando los resultados se concluye que aumentar la duración del segmento por encima de 10 s no proporciona ninguna ventaja significativa en lo que se refiere a la tasa de datos y la cobertura del servicio. De hecho, y como se discutirá en posteriores capítulos, es recomendable utilizar segmentos de corta duración para no incrementar la latencia del servicio y maximizar la QoE. La Figura 3.3b muestra que no hay una degradación excesiva en la cobertura o la tasa de datos del servicio cuando se reduce la duración del segmento.

3.2.3. Análisis comparativo de los códigos Raptor y RaptorQ

Raptor es el código AL-FEC que ha seleccionado 3GPP como estándar para proporcionar fiabilidad al envío de contenido multimedia sobre eMBMS. Sin embargo, existe una evolución del código llamada RaptorQ que es más eficiente, es decir, teóricamente ofrece un mayor nivel de protección frente a errores para la misma cantidad de redundancia (Ecuación 3.3). A continuación, se compara el rendimiento de ambos códigos mediante un análisis muy semejante al que ya se ha presentado.

La Figura 3.4a muestra la relación entre la cobertura del servicio y la tasa de codificación AL-FEC para el índice 25 del MCS y diferentes tamaños del área MBSFN. Los resultados muestran que para una tasa de codificación AL-FEC determinada, RaptorQ obtiene una mayor cobertura del servicio que Raptor, especialmente, cuando la tasa de codificación se acerca a la unidad. Esto demuestra que los códigos RaptorQ requieren menos redundancia para proporcionar la misma fiabilidad. Por ejemplo, para un área MBSFN de 7 celdas, RaptorQ puede

(a) Impacto del tamaño del área MBSFN

(b) Impacto del esquema MCS

(c) Impacto del tamaño del área MBSFN

(d) Impacto de la duración del segmento

Figura 3.4: Análisis comparativo de los códigos Raptor y RaptorQ. La figura muestra resultados que permiten comparar ambos códigos analizando la relación entre la cobertura del servicio, la tasa de codificación AL-FEC y la cobertura del servicio para diferentes valores del tamaño del área MBSFN, del MCS y de la duración del segmento multimedia (T_{sgmnt}).

proporcionar una cobertura del 70 % con un AL-FEC code rate de 0.9. Para cubrir un 70 % de cobertura con el código Raptor sería necesario establecer un AL-FEC code rate de 0.88.

La Figura 3.4b muestra los resultados obtenidos para un área MBSFN de 19 celdas y diferentes valores del índice MCS. Como se puede observar, utilizar RaptorQ permite seleccionar MCS más altos (menos robustos) sin que por ello se degrade la fiabilidad de la transmisión. Por ejemplo, para una tasa de codificación AL-FEC de 0.94, se obtiene una cobertura de aproximadamente el 70 % para el índice 25 del MCS y Raptor o el índice 28 del MCS y RaptorQ. La ventaja de utilizar RaptorQ es que al poder seleccionar el MCS 28, la tasa de datos que puede alcanzarse es mayor. La figura también muestra que para un MCS fijo y tasas de codificación cercanas a 1, la robustez de RaptorQ frente a errores es significativamente mayor que la de Raptor.

La Figura 3.4c muestra la relación entre la tasa de datos y la cobertura del servicio para el índice 25 del MCS y diferentes tamaños del área MBSFN. Los resultados muestran la importancia que tiene combinar MBSFN con el uso de las técnicas AL-FEC. Seleccionar un MCS poco robusto para alcanzar tasas de datos elevadas tiene el inconveniente de que la cobertura del servicio se reduce. Sin embargo, el uso de MBSFN también proporciona fiabilidad a la transmisión sobre eMBMS lo que permite alcanzar coberturas mayores sin renunciar a tasas de datos altas.

La Figura 3.4d muestra que la variación de la duración del segmento no afecta prácticamente a la cobertura o a la tasa de datos del servicio cuando se utiliza el código RaptorQ (derecha). El impacto es visiblemente mayor cuando se utilizan los códigos Raptor (izquierda).

Finalmente, se ha analizado cómo se deben combinar el MCS y el AL-FEC code rate para

Figura 3.5: Análisis comparativo de los códigos Raptor. La figura muestra resultados que permiten analizar cómo se debe combinar el MCS y tasa de codificación AL-FEC con el objetivo de maximizar la tasa de datos del servicio para una cobertura específica.

maximizar la tasa de datos del servicio asumiendo una cobertura específica. La Figura 3.5 muestra este análisis cuando se consideran diferentes tamaños del área MBSFN. Los resultados indican que el uso del código RaptorQ no permite alcanzar tasas de datos significativamente más altas (en comparación a Raptor), salvo en algunos puntos específicos de cobertura. Por ejemplo, en torno al 80 %-85 % de cobertura, y para áreas MBSFN de 7 celdas, se observa que la tasa de datos cae en torno a 500 Kbps si se utiliza el código Raptor, lo que se debe a que se requiere seleccionar un MCS más robusto porque el rendimiento del código Raptor es menor que el de RaptorQ.

3.3. Escenarios virtuales basados en software de código abierto

Para complementar el análisis de los servicios de streaming multimedia se ha desarrollado una segunda plataforma de experimentación basada en la definición de escenarios virtuales. Esta plataforma puede conectarse a un piloto de demostración o una red móvil real, aunque también es posible utilizar medidas de ancho de banda, latencia y pérdidas de paquetes para emular la red.

Los escenarios virtuales se implementan utilizando la herramienta de virtualización Virtual Networks over linux (VNX) [FRB⁺ 16], un proyecto software de código abierto desarrollado en el seno del Departamento de Ingeniería de Sistemas Telemáticos de la Universidad Politécnica de Madrid. El uso de VNX presenta diferentes ventajas, entre las que destaca su flexibilidad y versatilidad, pues permite definir y configurar redes virtuales de forma rápida y sencilla, y desplegar contenedores/máquinas virtuales que simulan servidores o sistemas finales.

El escenario virtual se especifica en un fichero eXtensible Markup Language (XML) que define las máquinas virtuales que lo componen y como se interconectan entre sí. También es posible definir en el fichero XML comandos que permiten configurar y controlar las máquinas virtuales. Por ejemplo, es posible definir un comando que arranque una sesión de streaming basado en DASH entre un reproductor multimedia y el servidor de contenidos.

Para facilitar el despliegue de los escenarios virtuales en equipos con diferentes sistemas operativos es posible utilizar una segunda capa de virtualización. Con el objetivo de portar la plataforma a una empresa del sector para integrarla en un piloto de demostración de una red LTE se ha optado por la siguiente solución: se ha encapsulado la herramienta en una máquina virtual creada con VirtualBox³. VirtualBox permite exportar la máquina virtual a través de una Open Virtual Appliance (OVA) e importarla posteriormente en cualquier sistema anfitrión independientemente de su sistema operativo. El uso de dos capas de virtualización podría afectar al rendimiento de la plataforma, pero en los experimentos realizados se ha comprobado que no existe ninguna limitación en este sentido.

A continuación, se describen los componentes software que están incluidos en la plataforma. Todos ellos pertenecen a proyectos de código abierto y se han combinado convenientemen-

³“VirtualBox,”[Online]. Available: <https://www.virtualbox.org/>

te para emular el servidor de contenidos, el terminal móvil y la red.

3.3.1. Terminal móvil

El terminal del usuario está compuesto por un reproductor DASH, un cliente FLUTE y una cache HTTP. En la plataforma, el terminal se divide en dos partes: una máquina virtual que incluye el cliente FLUTE y la cache HTTP, y un reproductor DASH que se ejecuta desde la máquina anfitrión y que se conecta al escenario definido en VNX para acceder al contenido multimedia. Esto hace posible que los experimentos se puedan llevar a cabo en diferentes tipos de terminales (smartphones, tablets o labtops).

Como cliente FLUTE se ha utilizado el paquete MAD-FCL [PRP⁺], disponible tanto para Linux como para Windows. Como cache HTTP se ha instalado un servidor Squid estándar [SQU]. Finalmente, y con respecto al reproductor DASH, la plataforma incluye dos opciones. La primera es dash.js [DAS], un reproductor escrito en JavaScript y publicado por el DASH Industry Forum (DASH-IF). La segunda opción es un reproductor VLC [VLC] que ejecuta un plugin para permitir sesiones de streaming basado en DASH.

3.3.2. Servidor de contenidos

El servidor de contenidos es responsable de almacenar y entregar el contenido multimedia a los usuarios. Este se puede codificar con H.264 [WSBL03b] a diferentes resoluciones y bitrates utilizando la herramienta ffmpeg [FFm], instalada en la máquina virtual. También es posible cargar en el servidor cualquier contenido almacenado en la máquina anfitrión utilizando un protocolo para la transferencia de ficheros. Para generar los segmentos multimedia y el MPD se ha seleccionado la herramienta MP4Box [GPA]. El servidor de contenidos también incluye un servidor FLUTE para la transmisión multicast de los segmentos multimedia. De nuevo, se ha utilizado el paquete MAD-FCL para proporcionar esta funcionalidad.

3.3.3. Conformador de tráfico

El conformador de tráfico permite emular la interconexión de dos dispositivos finales a través de una red caracterizada mediante los siguientes parámetros de QoS: ancho de banda, retardo, y pérdida de paquetes. Los scripts que se han desarrollado con este objetivo utilizan la disciplina Class-based Queuing (CBQ) para clasificar el tráfico entrante, y las utilidades Linux Traffic Control (TC) y Network Emulation (NetEm), para controlar el ancho de banda, retardo y pérdidas en las interfaces donde se aplica el conformado de tráfico. Los scripts leen de fichero los valores de los parámetros de QoS. Es decir, el fichero especifica la evolución temporal de estos parámetros. También es posible establecer diferentes políticas de conformado para el tráfico UDP y TCP, lo que es útil para emular las características del canal multicast y unicast de forma independiente.

3.3.4. Creación de escenarios con VNX

Utilizando estos componentes se han implementado dos escenarios con VNX (Figura 3.6). El primero de ellos (Escenario A) se ejecuta en un entorno completamente virtualizado, donde se interconectan el terminal móvil y el servidor de contenidos, a través de un conformador de tráfico que emula la red móvil (Figura 3.6a). Este escenario se despliega en una única máquina anfitrión.

El segundo escenario (Escenario B) se ha diseñado con el objetivo de conectar el terminal y el servidor a través de una red móvil real. En este caso, se ha conectado a un piloto de demostración de red LTE de Nokia (Figura 3.6b). El demostrador se compone de un EPC y una red de acceso radio E-UTRAN con tres eNBs. Durante los experimentos se ha utilizado una estación base para dar servicio a los usuarios y las otras dos para generar señales de interferencia. Las estaciones base, equipadas con dos antenas para la transmisión y recepción de la señal radio, incluyen todas las funciones de red necesarias para dar soporte a los servicios unicast, por ejemplo, las relacionadas con la planificación de los recursos radio. La potencia

(a) Escenario A: toda la plataforma se ejecuta sobre un entorno virtualizado. Para simular la transmisión del contenido sobre la red se despliega una máquina virtual que realiza el conformado de tráfico.

(b) Escenario B: integración de la plataforma en un piloto de demostración de red LTE.

Figura 3.6: Escenarios virtuales basados en VNX

de las estaciones base puede regularse manualmente a través de atenuadores, lo que permite simular diferentes condiciones del canal radio. Esto es útil, por ejemplo, para simular la señal que recibe un usuario que está en el borde de la celda.

El EPC y la red de acceso radio se interconectan entre sí a través de interfaces estandarizadas. Los clientes pueden ser smartphones, tablets, o como se representa en la Figura 3.6b, máquinas virtuales que incluyen los componentes software que se han descrito anteriormente. La máquina virtual que actúa como terminal del cliente se conecta a la red de acceso radio utilizando un modem USB con conexión LTE. La máquina virtual que actúa como servidor de contenidos se conecta directamente al P-GW del núcleo de la red móvil a través de una conexión Ethernet.

Como se puede observar comparando ambas figuras, ha sido necesario dividir el Escenario A en dos para integrar la plataforma virtual y el piloto de demostración. En la máquina virtual que se conecta al P-GW se despliega un escenario basado en VNX que emula el servidor de contenidos, y en la máquina virtual donde se conecta el modem USB se despliega otro escenario basado en VNX que emula el terminal del cliente.

Utilizando esta plataforma, y aplicando una metodología para evaluar el servicio, se pueden validar nuevos algoritmos de adaptación DASH u otras soluciones que permitan mejorar la distribución de los contenidos multimedia sobre las redes móviles.

La Figura 3.7 muestra la metodología de evaluación que se propone. La idea consiste en utilizar el Escenario B, que conecta la plataforma virtual a una red móvil real, para analizar el servicio y extraer medidas de ancho de banda, latencia y pérdidas que posteriormente se utilizarán para emular esa misma red en el Escenario A. De esta forma, no hace falta estar co-

Figura 3.7: Metodología para la evaluación de los servicios de streaming multimedia.

nectado físicamente a la red para evaluar nuevos algoritmos de adaptación u otras soluciones, este proceso se puede realizar desde cualquier sistema anfitrión en un entorno virtualizado. Las soluciones propuestas pueden requerir modificaciones en el terminal o el servidor de contenidos. Por ejemplo, puede requerir la implementación de un nuevo algoritmo de adaptación en el reproductor DASH. Tras validar las propuestas sobre el Escenario A, las versiones modificadas de las máquinas virtuales se interconectan de nuevo a la red móvil para validar la solución en un entorno real de pruebas (Escenario B).

Esta metodología se ha aplicado con éxito para validar un nuevo algoritmo de adaptación DASH basado en el método estadístico Adaptive Forgetting Factor (AFF) [ABLP18].

3.4. Uso de la plataforma virtual para analizar los servicios multimedia a partir de mediciones extremo a extremo

A continuación, se presentan los resultados que se han obtenido tras conectar la plataforma virtual al piloto de demostración de la red LTE. El piloto de demostración no permite configurar manualmente la cantidad de recursos radio asignados a cada cliente, por lo que se ha fijado el MCS al índice 3 para analizar el rendimiento del servicio de streaming multimedia en un único reproductor multimedia. Asumiendo que el ancho de banda del canal es de 10 MHz, el índice 3 del MCS proporciona una tasa de datos máxima de 2.85 Mbps cuando se asignan todos los Physical Resource Block (PRB)s (todos los recursos radio) a la conexión unicast del cliente [FLB⁺16]. Se debe tener en cuenta que algunos PRBs están reservados para la transmisión de las señales de control, por lo que el número máximo de PRBs que se puede asignar es ligeramente inferior a 50.

El MPD define cuatro representaciones, todas ellas con una resolución de 720p y diferentes bitrates: 250 Kbps, 500 Kbps, 1000 Kbps y 2000 Kbps. La duración del segmento multimedia es de 2 s. La Figura 3.8 presenta los resultados que se obtienen cuando se desactiva el algoritmo de adaptación del reproductor dash.js. El reproductor siempre solicita la representación de mayor calidad (2 Mbps). Esta representación se envía al reproductor sin que se produzcan paradas del servicio (Figura 3.8b) debido a que el ancho de banda que proporciona el MCS (2.85 Mbps) es mayor que el bitrate de dicha representación. El único evento de parada que se observa ocurre al inicio de la sesión, donde se observa que el retardo acumulado es de 2.5 s (Figura 3.8b).

La Figura 3.9 muestra los resultados obtenidos cuando se activa la lógica de adaptación del reproductor. Como se puede observar, el reproductor conmuta ahora entre las representaciones disponibles (Figura 3.9a) por lo que se produce un efecto de inestabilidad [JSZ14]. Sin embargo, la Figura 3.9b muestra que la ventaja de ejecutar el algoritmo de adaptación es que es posible evitar las paradas de servicio y reducir el retardo inicial de buffering. La Figura 3.9b muestra que, en este caso, el retardo acumulado de parada es de 0.5 s.

Finalmente, la Figura 3.10 muestra resultados que permiten analizar la eficiencia del reproductor, entendiéndose la eficiencia como la capacidad del algoritmo para ajustarse al ancho

(a) Bitrate y nivel del buffer

(b) Retardo. Tiempo de parada del servicio

Figura 3.8: Estabilidad y tiempo de parada cuando no se ejecuta la lógica de adaptación del reproductor.

(a) Bitrate y nivel del buffer

(b) Retardo. Tiempo de parada del servicio

Figura 3.9: Estabilidad y tiempo de parada cuando se ejecuta la lógica de adaptación del reproductor.

de banda disponible. La figura muestra el ancho de banda disponible del canal, la estimación de ancho de banda que hace el algoritmo de adaptación y la representación solicitada por el algoritmo. Los resultados indican que la eficiencia sería mayor si no se ejecutara la lógica de adaptación, pues la diferencia que existe en términos de ancho de banda entre la representación

Figura 3.10: Análisis de la eficiencia cuando se activa o desactiva la lógica de adaptación del reproductor DASH.

de 2 Mbps y el ancho de banda disponible es menor que la diferencia que existe entre la representación solicitada (por el algoritmo de adaptación) y el ancho de banda disponible. La figura muestra que cada vez que se conmuta a una calidad inferior se subestima el ancho de banda, y por tanto, la eficiencia es menor.

3.5. Conclusiones

La comercialización de servicios basados en eMBMS todavía es muy limitada, lo que dificulta la experimentación en entornos de pruebas reales. Esto, junto con la ausencia de otras plataformas alternativas que permitan analizar los servicios de streaming 3GP-DASH sobre eMBMS, ha motivado el diseño y desarrollo de las herramientas que se describen en este capítulo.

Se han adoptado dos enfoques para el desarrollo de las herramientas. El primero, combina un modelo de simulación y un modelo analítico. El segundo, se basa en la creación de escenarios virtuales y la utilización de software de código abierto. Mediante la utilización de las herramientas ha sido posible analizar el impacto de diferentes parámetros que afectan al rendimiento de los servicios de streaming basados en DASH, y en particular, de los servicios de streaming 3GP-DASH sobre eMBMS.

En capítulos posteriores se indican cuales son las modificaciones que han permitido utilizar estas plataformas para evaluar las propuestas que han surgido en el marco de la tesis.

El contenido de este capítulo se ha presentado en varios congresos de carácter internacional [LBF⁺14, FLB⁺15, LAB⁺15, FLB⁺16].

Capítulo 4

Reducción de la latencia en los servicios de difusión de contenidos multimedia

Los servicios de streaming multimedia son sensibles a retardos. El impacto del retardo en la reproducción del contenido es mayor si cabe cuando el contenido multimedia se genera en directo (es decir, en servicios de tipo live streaming). En este capítulo se analiza la cadena de retardos que sufre un segmento multimedia desde su generación en el servidor de contenidos hasta su reproducción en el terminal del cliente. Este análisis ha permitido identificar un problema que surge de combinar el uso de las técnicas AL-FEC con la recuperación de errores mediante HTTP. Si no se informa correctamente a los reproductores acerca de la disponibilidad de los segmentos multimedia, pueden producirse errores de sincronización. La desincronización entre la transmisión multicast y la recuperación de errores por HTTP puede causar el mal funcionamiento del servicio, puesto que puede establecerse la recuperación unicast de segmentos multimedia que, en realidad, van a ser entregados correctamente sobre eMBMS. Para garantizar que los recursos radio se gestionan eficientemente es necesario evitar los errores de sincronización.

La recuperación unicast de los segmentos perdidos puede provocar la parada del servicio si el buffer del reproductor se vacía durante este procedimiento. Para evitar que esto ocurra se debe determinar correctamente el nivel inicial del buffer del reproductor. El retardo inicial de buffering se define como el tiempo que transcurre desde que el cliente inicia la sesión de streaming, hasta que el buffer del reproductor almacena los segundos de vídeo que especifica el fichero MPD y la reproducción puede dar comienzo (Capítulo 2). Aunque un retardo suficientemente largo evitaría las paradas de la reproducción multimedia, el sobredimensionamiento del buffer inicial incrementa innecesariamente la latencia del servicio. Para solucionar este problema, se propone un método para calcular el buffer mínimo que evita las paradas de servicio en un área de cobertura específica.

La reducción de la duración del segmento multimedia permite reducir la latencia del servicio pero puede aumentar la sobrecarga del cliente, del servidor y la infraestructura de red como consecuencia del aumento excesivo del número de transacciones HTTP. En este capítulo se determina si es plausible reducir la duración del segmento para reducir la latencia de los servicios de streaming 3GP-DASH sobre eMBMS.

La estructura del capítulo es la siguiente: la Sección 4.1 presenta el análisis del retardo y explica como pueden evitarse los errores de sincronización. La Sección 4.2 describe un método para calcular el buffer mínimo que evita las paradas del servicio. La Sección 4.3 muestra el cálculo del buffer mínimo en un despliegue específico de la red móvil. La Sección 4.4 muestra que es posible reducir la duración del segmento multimedia sin aumentar los eventos de parada del servicio. Finalmente se analiza la calidad del streaming multimedia (Sección 4.5) y se proporcionan las conclusiones de este trabajo (Sección 4.6).

4.1. Análisis de retardo

Desde su generación en el servidor de contenidos, hasta su reproducción en el terminal del cliente, los segmentos multimedia sufren diferentes retardos. Esta sección identifica los componentes de retardo (representados en la Figura 4.1) que afectan a los servicios de streaming 3GP-DASH sobre eMBMS, y a partir del análisis, se explican los errores de sincronización que pueden ocurrir y cómo pueden evitarse.

4.1.1. Generación de los segmentos multimedia

El análisis de retardo comienza en el instante t_0 , que indica la recepción del primer byte de la señal de video en el servidor de contenidos. Desde ese instante, cada vez que se reciben T_{sgmnt} segundos de video, se genera un nuevo segmento multimedia. La generación del segmento requiere utilizar un códec de video como H.264 o H.265, de forma que pueda transmitirse sobre eMBMS una versión comprimida del contenido original. En el proceso de codificación solo se genera una representación multimedia; la representación que se transmite en el canal eMBMS. El tiempo requerido para fragmentar y codificar el segmento se ha denotado, en la Figura 4.1 como d_{gen} .

Figura 4.1: Análisis de retardo. La figura muestra los retardos sufridos por un segmento multimedia que se envía sobre eMBMS y se recupera por HTTP. La figura también muestra el instante en que el segmento está disponible para ser solicitado (parámetro *availabilityStartTime*).

4.1.2. Transmisión de los segmentos a la red del operador

El proveedor de contenidos debe definir varios parámetros del servicio utilizando el punto de referencia xMB (Capítulo 2). Uno de estos parámetros es el denominado *Ingest Mode*, que puede adquirir uno de estos valores: push o pull. El modo push se utiliza con el objetivo de que los segmentos multimedia se envíen al BM-SC tan pronto como se generen, sin esperar a ninguna solicitud previa. El modo pull, por el contrario, requiere que el BM-SC solicite bajo demanda los segmentos antes de que el proveedor de contenidos proceda a enviarlos. Esto puede añadir un retardo adicional. La Figura 4.1 muestra, como ejemplo, el uso del modo push, pero el uso de uno u otro modo depende del acuerdo entre el proveedor de contenidos y el operador de la red móvil.

4.1.3. Protección de los segmentos multimedia mediante la aplicación de técnicas AL-FEC

El 3GPP propone aplicar las técnicas AL-FEC para proporcionar más fiabilidad al envío de contenido multimedia sobre eMBMS. La aplicación de las técnicas AL-FEC es responsabilidad del operador de red. Durante la definición del servicio, el proveedor de contenidos solo debe especificar el bitrate de la representación multimedia que se transmite sobre eMBMS (ver Capítulo 2).

El uso de un código bloque como Raptor o RaptorQ hace posible fragmentar el flujo multimedia en segmentos (source blocks) que pueden codificarse de forma independiente. Así, el codificador AL-FEC genera datos de redundancia que se envían, junto con los símbolos que componen el *source block*, sobre el canal de eMBMS.

El receptor decodifica correctamente un segmento si se han recibido suficientes paquetes, del *source block* o de redundancia. El estado del arte muestra que el rendimiento del proceso de decodificación mejoraría con el uso de segmentos multimedia de larga duración [BKPP11]. Sin embargo, esto incrementa la latencia del servicio innecesariamente, lo que a su vez, degrada la calidad percibida por los usuarios. En la Figura 4.1 se denota el tiempo requerido para codificar un segmento como d_{cod} , mientras que el tiempo requerido para su decodificación se denota como d_{dec} .

4.1.4. Transporte de los segmentos multimedia utilizando el protocolo FLUTE

El transporte de los segmentos multimedia sobre eMBMS se realiza estableciendo una sesión FLUTE. El cliente FLUTE, ubicado en el terminal del usuario, debe recibir los atributos de la sesión, así como la descripción de las propiedades del fichero que va a transmitirse sobre eMBMS [PWL⁺12]. Esta información se define en un fichero XML llamado File Delivery Table (FDT) que también se envía durante la sesión. Debido a que en un servicio de streaming multimedia en directo los segmentos se generan progresivamente, las propiedades de los ficheros también deben notificarse progresivamente. Un único FDT con la descripción de las propiedades de todos los segmentos no tiene sentido en este contexto, puesto que una de las

propiedades que debe especificarse en el FDT es el tamaño del segmento, dato que se desconoce hasta que el códec de vídeo no ha realizado el proceso de codificación [PWL⁺12]. La solución es enviar una instancia del FDT por cada segmento multimedia que se genere en el servidor. El retardo para generar las instancias del FDT se ha despreciado en el análisis de retardo, ya que se trata de un fichero de metadatos de pequeño tamaño que describe únicamente las propiedades de un segmento multimedia.

Si se aplica una técnica AL-FEC para proporcionar más fiabilidad al envío de los segmentos multimedia, es necesario que la sesión FLUTE transporte un objeto llamado FEC Object Transmission Information (FEC OTI) [WLV07]. La información que transporta este objeto depende del esquema FEC que se utilice. En el caso de Raptor, debe incluir la longitud del segmento y el tamaño del símbolo, parámetros que se requieren durante el proceso de decodificación. Aunque el objeto FEC OTI puede transmitirse directamente en la cabecera de los paquetes FLUTE [PWL⁺12], se recomienda utilizar el FDT para su envío debido a que la sobrecarga que impone es menor.

El tiempo necesario para transmitir un segmento multimedia sobre eMBMS puede calcularse a partir de dos parámetros: el tamaño del segmento y la tasa de transmisión del enlace. La tasa de transmisión del canal eMBMS es fija porque el valor del índice del MCS se determina antes de que la sesión FLUTE comience y no sufre variaciones durante la misma. También es fija porque la cantidad de recursos radio que se asigna al servicio eMBMS permanece constante durante toda la sesión.

El tiempo para transmitir una instancia del FDT (d_{fdt}) es considerablemente menor que el tiempo para transmitir el segmento multimedia (d_{mcast}), por lo que d_{fdt} puede despreciarse en el análisis. El tamaño del FDT es del orden de cientos de bytes, mientras que el tamaño del segmento es un orden de magnitud mayor.

4.1.5. Protección frente a variaciones de retardo

La disponibilidad de los segmentos multimedia se notifica a los reproductores mediante el campo *availabilityStartTime* del MPD, que indica el instante de tiempo (UTC) en que un segmento está disponible en el servidor de contenidos y puede ser enviado al cliente. Lohmar *et al.* [LEF⁺11] analizan el valor que debería tener el campo *availabilityStartTime*, teniendo en cuenta que los diferentes componentes de retardo pueden sufrir variaciones. Por ejemplo, el tiempo empleado en la codificación de los segmentos multimedia no es constante si la codificación que se utiliza es Variable Bit Rate (VBR).

La variación del retardo puede causar errores de sincronización debido a que los reproductores pueden solicitar segmentos multimedia que todavía no están disponibles en el servidor. Por otro lado, si la frecuencia de las peticiones es elevada, puede introducirse una sobrecarga en el servidor que no es despreciable. Para evitar este efecto, Lohmar *et al.* [LEF⁺11] proponen retardar las peticiones de los reproductores multimedia incrementando el valor del parámetro *availabilityStartTime*. Utilizando este margen de seguridad, se garantiza que los segmentos

están disponibles en el servidor de contenidos cuando los reproductores emiten la petición.

El análisis de Lohmar *et al.* está orientado a los servicios de streaming adaptativo sobre HTTP y no a los servicios de streaming 3GP-DASH sobre eMBMS. La combinación de las técnicas AL-FEC y la recuperación por HTTP de los segmentos perdidos hace necesario que, para calcular el valor del campo *availabilityStartTime*, se deba tener en cuenta que si el reproductor solicita un segmento y no está almacenado en la memoria cache intermedia, se descargará a través de una conexión unicast mediante HTTP.

Por tanto, si el campo *availabilityStartTime* no se define correctamente, el reproductor podría solicitar los segmentos multimedia antes de que estén disponibles en la memoria cache. Esto implicaría recuperar por HTTP segmentos, que, en realidad, van a transmitirse correctamente sobre eMBMS. Por lo tanto, la incorrecta definición del campo *availabilityStartTime* deriva en errores de sincronización y, por ende, en el uso ineficiente de los recursos radio.

Para que esto no suceda, la protección frente a variaciones de retardo debe considerar componentes de retardo que son propios de la transmisión broadcast del contenido multimedia, como la codificación y decodificación AL-FEC, o el envío de los segmentos multimedia a la cache de los clientes. La Figura 4.1 muestra este margen de seguridad, denotado como d_{prot} .

4.1.6. Recuperación unicast de los segmentos multimedia

Cuando un reproductor solicita un segmento multimedia que no está disponible en la memoria cache, se establece una conexión HTTP para que pueda ser recuperado. Este procedimiento se realiza a través de una función de tipo *Associated Delivery* del BM-SC (Capítulo 2). Teniendo en cuenta la protección frente a variaciones de retardo que se ha añadido a la cadena de retardos, el establecimiento de la recuperación de errores unicast solo puede ocurrir en dos casos:

- cuando un segmento multimedia no se ha decodificado correctamente, o,
- cuando el segmento multimedia se decodifica correctamente, pero la instancia FDT que describe el segmento se ha perdido durante la transmisión multicast.

La utilización de HTTP para recuperar un segmento añade un retardo adicional (d_{ucast} en la Figura 4.1) que depende de varios factores. Por ejemplo, todos los terminales móviles conectados a servicios unicast compiten por los recursos radio disponibles. Estos se asignan en las estaciones base, donde se ejecuta un algoritmo de planificación de recursos que determina dinámicamente, y en función de diferentes parámetros, el número de PRBs que se asignará a cada conexión. En definitiva, la asignación dinámica del número de PRBs o del esquema de modulación y codificación (MCS) no permite predeterminar el valor del ancho de banda disponible en el canal unicast.

En definitiva, el retardo que añade la recuperación unicast de los segmentos perdidos se desconoce a priori. Si el retardo es mayor que la duración del segmento multimedia, el impacto

de este componente de retardo es especialmente significativo. Es más, puede provocar paradas de servicio si el buffer se vacía antes de que se recuperen por HTTP los segmentos perdidos.

4.1.7. Retardo inicial de buffering

El buffering de video es necesario para evitar paradas de servicio [KR17]. En los servicios de streaming adaptativo sobre HTTP, los reproductores ejecutan un algoritmo de adaptación que combina técnicas de estimación de ancho de banda con métodos de control del buffer para seleccionar la representación multimedia (Capítulo 2). Por tanto, el nivel del buffer es un indicador que utilizan los reproductores para adaptarse a las variaciones del ancho de banda. Por ejemplo, si el nivel del buffer es más bajo que cierto umbral, la lógica de adaptación puede interpretar que no hay suficiente ancho de banda para descargar el segmento y conmutar a una representación multimedia de menor calidad.

En los servicios de streaming 3GP-DASH sobre eMBMS, los requisitos de buffering son diferentes. Esto se debe a dos factores: el primero es que los reproductores no conmutan entre diferentes representaciones porque solo se envía sobre eMBMS una única representación. El segundo factor es la recuperación de los segmentos perdidos por HTTP. Para que no se produzcan paradas de servicio durante este procedimiento es necesario que el nivel de buffer inicial pueda absorber el retardo adicional que introduce la recuperación.

El campo *minBufferTime* del MPD determina el nivel inicial que debe alcanzar el buffer, medido en segundos, para que pueda iniciarse la reproducción. Si el valor del *minBufferTime* es bajo, y la tasa de error del canal es elevada, la recuperación unicast de los segmentos multimedia puede provocar que el buffer se vacíe y la reproducción se interrumpa. Para solucionar este problema, una posible estrategia consiste en calcular el buffer mínimo que se requiere para evitar paradas de servicio en un área de cobertura específica. La siguiente sección, explica un método para realizar este cálculo. El nivel de buffer, medido en segundos, se ha denotado en la Figura 4.1 como d_b .

4.1.8. Tiempo límite de reproducción

La desincronización entre la transmisión broadcast y la recuperación unicast de los segmentos multimedia puede evitarse si el campo *availabilityStartTime* del MPD se define correctamente. Tras analizar la cadena de retardo se determina la necesidad de ajustar el valor del campo *availabilityStartTime* al instante en que se recibe el segmento en la memoria cache, de forma que este pueda ser entregado directamente al buffer del reproductor sin necesidad de que se envíe una petición HTTP.

Así, y de acuerdo al análisis de retardo que se ha presentado, el valor del campo *availabilityStartTime* puede calcularse como:

$$availabilityStartTime = d_{cod} + d_{dec} + d_{mcast} + d_{prot} \quad (4.1)$$

Este instante se ha etiquetado y representado con una línea discontinua en la Figura 4.1. Partiendo del valor del campo *availabilityStartTime* se puede calcular el tiempo límite de reproducción o *PlaybackDeadline* como:

$$PlaybackDeadline = availabilityStartTime + d_b \quad (4.2)$$

donde d_b es el nivel del buffer. En los servicios de streaming multimedia basados en DASH, el nivel inicial del buffer puede comunicarse a los reproductores mediante el campo *minBufferTime*, que como se ha comentado anteriormente, se define como el número de segmentos que debe almacenar el buffer para que la reproducción pueda dar comienzo.

4.2. Método para reducir la latencia y evitar las paradas del servicio

Esta sección muestra un método para calcular el buffer mínimo que permite evitar las paradas de servicio en un área de cobertura específica. El método evita el sobredimensionamiento del buffer inicial, y por consiguiente, el incremento excesivo e innecesario del retardo inicial del servicio.

El buffer mínimo se calcula en función de la tasa de error de paquetes del canal eMBMS (PER), el tiempo de ida y vuelta o Round Trip Time (RTT) y el retardo para la recuperación unicast de los segmentos multimedia. Combinando este método con resultados de simulación de diferentes despliegues de la red móvil, es posible analizar el impacto de diferentes parámetros del servicio en el retardo inicial de buffering.

Como se explica en el Capítulo 2, la aplicación de las técnicas AL-FEC, o el uso del modo de transmisión MBSFN, no garantiza que los segmentos multimedia sean entregados correctamente a los receptores multicast. Por esta razón, el 3GPP propone utilizar HTTP para recuperar aquellos segmentos que no se transmitan correctamente sobre eMBMS. Sin embargo, el ancho de banda disponible en el canal utilizado para la recuperación unicast de los segmentos es desconocido y varía dinámicamente, al contrario de lo que ocurre en el servicio eMBMS, donde el ancho de banda es fijo durante toda la sesión. Monserrat *et al.* [MCFG12] muestran que el ancho de banda máximo asignado al terminal se reduce a medida que aumenta el número de terminales en la celda. Los resultados que presentan en su artículo muestran que el ancho de banda del canal unicast puede ser menor que el bitrate de la representación que se envía por eMBMS. Si esto sucede, el retardo para recuperar el segmento por HTTP será mayor que la duración del propio segmento, lo que podría conllevar la interrupción del servicio si el buffer del reproductor se vacía durante el proceso de recuperación.

Para calcular el buffer mínimo, se realiza un análisis del caso peor, es decir, cuando varios segmentos consecutivos se pierden durante su transmisión sobre eMBMS. El nivel inicial del buffer debe determinarse de forma que la ráfaga de segmentos se recupere por HTTP sin causar la parada del servicio.

El Capítulo 3 describe un modelo analítico que calcula la probabilidad de que un segmento

Figura 4.2: Análisis de retardo. La figura muestra el tiempo que se requiere para recuperar, por HTTP, m segmentos consecutivos.

multimedia no se decodifique correctamente en el receptor. Es decir, la probabilidad de que un segmento se recupere por HTTP. Para calcular el buffer mínimo se ha extendido el modelo analítico definiendo un umbral o *threshold* que representa la probabilidad de que se produzca una parada de servicio. El umbral se relaciona con la probabilidad de que m segmentos consecutivos se pierdan en el canal eMBMS:

$$P(f_{RC})^m < threshold \quad (4.3)$$

Por ejemplo, para garantizar la continuidad de la reproducción durante un largo periodo de tiempo, debe fijarse un valor pequeño del *threshold*. Como se mostrará más tarde, esto requiere aumentar el retardo inicial de buffering. En definitiva, dado un valor del *threshold*, la parada del servicio se evita si el buffer es lo suficientemente largo como para que se pueda recuperar por HTTP una ráfaga de m segmentos perdidos.

La Figura 4.2 muestra, como ejemplo, la recuperación unicast de tres segmentos perdidos durante su transmisión sobre eMBMS. Se asume que se ha establecido una conexión HTTP persistente. De lo contrario, también habría que considerar un retardo adicional asociado al establecimiento de la conexión TCP. La Figura 4.2 muestra que el retardo para recuperar un segmento por HTTP (d_{ucast}) es mayor que la duración del segmento T_{sgmnt} . Para evitar la parada del servicio es necesario ajustar el nivel del buffer al retardo que sufre el último segmento, el que se recupera más tarde. El buffer mínimo necesario puede calcularse, por tanto, como

$$d_b = trec_{(n+2)} - treq_{(n+2)} = RTT + 3d_{ucast} - 2T_{sgmnt} \quad (4.4)$$

donde $treq_{(n+2)}$ es el instante en que se solicita el último segmento, $trec_{(n+2)}$ el instante en

que se recibe, d_{ucast} el retardo para recuperar un segmento por HTTP, y T_{sgmnt} la duración del segmento. Como se puede observar en la figura, en un servicio de streaming multimedia en directo, el reproductor solicita periódicamente los segmentos cada T_{sgmnt} .

En general, si el retardo para recuperar por HTTP un segmento es mayor que su duración ($d_{ucast} > T_{sgmnt}$), el buffer necesario para evitar la parada del servicio durante la recuperación unicast de m segmentos puede calcularse como:

$$d_b = RTT + md_{ucast} - (m - 1)T_{sgmnt} \quad (4.5)$$

Sin embargo, si $d_{ucast} < T_{sgmnt}$, el buffer necesario es:

$$d_b = RTT + d_{ucast} \quad (4.6)$$

Resumiendo, es posible calcular la probabilidad de que un segmento se recupere por HTTP (ecuación 3.1) a partir del PER y de las ecuaciones que modelan el comportamiento de los códigos Raptor (ecuaciones 3.2 y 3.3). Esta probabilidad puede utilizarse para calcular el número de segmentos que se perderían en el canal eMBMS de forma consecutiva para un *threshold* dado, siendo el *threshold* la probabilidad de que el servicio se interrumpa (ecuación 4.3). Finalmente, puede calcularse el buffer mínimo (ecuaciones 4.5 y 4.6) a partir del retardo para la recuperación unicast de los segmentos y el RTT. El valor obtenido se especifica en el campo *minBufferTime* del MPD, que indica el número de segmentos que debe almacenar el buffer antes de que la reproducción pueda dar comienzo.

4.3. Dimensionamiento del buffer para un despliegue específico de la red móvil

Como se ha explicado en la sección anterior, el cálculo del buffer mínimo depende del PER, del RTT y del retardo que añade la recuperación unicast de los segmentos. En esta sección se muestra el cálculo del buffer mínimo para un despliegue específico de la red móvil. La Tabla 4.1 muestra los parámetros de servicio que se han considerado para evaluar el método propuesto.

Aunque el PER depende de varios parámetros de servicio, como, por ejemplo, el AL-FEC

Parámetro	Valor
Cobertura	90 %
Número de terminales	399
Duración del segmento (T_{sgmnt})	2 s
Inter-Site Distance (ISD)	500 m
Índice del esquema de modulación y codificación (MCS)	15
Tamaño del área MBSFN	7 eNBs
AL-FEC code rate	0.84 / 0.86 / 0.88 / 0.9

Tabla 4.1: Parámetros de servicio empleados para determinar el buffer mínimo.

code rate, el MCS o el ISD, la evaluación que se presenta en este capítulo se centra exclusivamente en analizar el impacto del AL-FEC code rate en el buffer mínimo. Otros parámetros que también afectan al PER permanecen fijos. Por ejemplo, se ha seleccionado un área MBSFN de 7 celdas porque, a pesar de que se podría desplegar un área mayor para reducir aun más el PER, la complejidad para sincronizar las estaciones base es mayor. Con respecto al MCS, se ha seleccionado el índice 15 porque proporciona una tasa superior a 1 Mbps, valor que proporciona un buen compromiso entre el ancho de banda asignado al canal eMBMS y la calidad de vídeo [LG12]. Finalmente, y con respecto al ISD, se ha utilizado la distancia de separación estándar para entornos urbanos (500 m). En entornos suburbanos el ISD es mayor (1732 m), lo que degrada el SINR e incrementa el PER.

La Tabla 4.1 también muestra que se va a calcular el buffer mínimo para proporcionar una cobertura del 90 %. Es decir, el nivel inicial del buffer debe evitar paradas de servicio en el 90 % de los terminales, siendo 399 el número total de terminales que se distribuyen en el área MBSFN. La duración de los segmentos multimedia es de dos segundos y el valor del *threshold* es 10^{-6} , lo que equivale a 55 horas de reproducción sin paradas de servicio.

Utilizando las herramientas descritas en el Capítulo 3, se ha calculado, para cada terminal del área MBSFN, la probabilidad de que los segmentos que componen la presentación multimedia se recuperen por HTTP. Después, se ha calculado el buffer mínimo que evita las paradas de servicio en el 90 % de los terminales de acuerdo con el valor del *threshold* seleccionado.

La Figura 4.3 muestra el retardo inicial de buffering (expresado en número de segmentos) en función de la probabilidad $P(f_{RC})$, es decir, la probabilidad de que el segmento no se transmita correctamente sobre eMBMS. El cálculo se realiza para diferentes valores del retardo de

Figura 4.3: Buffer mínimo para evitar paradas de servicio. Las líneas discontinuas muestran la probabilidad de que el proceso de decodificación AL-FEC $P(f_{RC})$ no se lleve a cabo con éxito en el usuario 10 percentil para diferentes valores del AL-FEC code rate.

recuperación unicast y asumiendo un $RTT = 0$. El retardo d_{ucast} se expresa en múltiplos de la duración del segmento (T_{sgmnt}). La figura muestra el buffer mínimo solo si $d_{ucast} > T_{sgmnt}$, es decir, cuando la pérdida de segmentos en el canal multicast implica el aumento del retardo.

En este caso, se va a calcular el buffer mínimo que garantiza la continuidad de la reproducción multimedia en el 90 % de los terminales. Con este propósito, se ha obtenido la probabilidad $P(f_{RC})$ del usuario 10 percentil. Esta probabilidad se representa, en la Figura 4.3, con líneas verticales discontinuas. En particular, se muestran cuatro casos, que se corresponden con cuatro valores distintos del AL-FEC code rate. Por ejemplo, si se utiliza un AL-FEC code rate de 0.84, la probabilidad de que el usuario 10 percentil no reciba correctamente un segmento por eMBMS es 0.0387. Para evitar las paradas del servicio, el buffer mínimo debe ser de 2, 3, 4 y 5 segmentos cuando el retardo de recuperación unicast del segmento es de $1,25 \cdot T_{sgmnt}$, $1,5 \cdot T_{sgmnt}$, $1,75 \cdot T_{sgmnt}$ y $2 \cdot T_{sgmnt}$ respectivamente. Para valores más altos del AL-FEC code rate, la redundancia introducida por el transmisor es menor y la probabilidad de recuperar el segmento por HTTP es mayor. En definitiva, el uso de AL-FEC code rates cercanos a 1 requiere un retardo inicial de buffering mayor.

4.4. Streaming multimedia de bajo retardo

La Sección 4.3 analiza el impacto de la recuperación unicast de los segmentos perdidos en el retardo inicial de buffering. Esta sección analiza la relación que existe entre el retardo inicial de buffering y la duración del segmento multimedia. La hipótesis de partida es que es posible reducir el retardo generando segmentos multimedia de menor duración. Wei *et al.* [WS14] abordaron esta cuestión para un servicio de streaming adaptativo sobre HTTP. Concluyeron que la reducción de la duración del segmento implica un incremento del número de transacciones HTTP, y por tanto, una sobrecarga adicional en clientes, servidor, e infraestructura de la red. En un servicio de streaming 3GP-DASH sobre eMBMS, debe reconsiderarse esta opción, dado que la mayoría de los segmentos se entregan correctamente sobre eMBMS, y HTTP solo se utiliza como mecanismo de recuperación.

Si se define el servicio para garantizar un área de cobertura amplia, la mayoría de los segmentos se transmiten correctamente sobre el canal eMBMS sin la necesidad de que se retransmitan por HTTP. Sin embargo, se debe tener en cuenta que el ancho de banda del canal eMBMS es fijo, por lo que existe un compromiso entre la tasa de datos del servicio y el nivel de protección que proporciona el AL-FEC. La ecuación 4.7 define esta relación:

$$SDR = R_{eMBMS} \cdot CR \quad (4.7)$$

donde $ServiceDataRate(SDR)$ es la tasa de datos del servicio, CR el AL-FEC code rate y R_{eMBMS} el ancho de banda asignado al canal eMBMS. La duración del segmento multimedia afecta al rendimiento de la decodificación AL-FEC (ver Capítulo 3). Por ejemplo, dada una tasa de error fija, reducir la duración del segmento implica un SDR menor, puesto que se requiere

Figura 4.4: Relación entre el AL-FEC code rate, el porcentaje de segmentos recuperados por HTTP y la duración del segmento. La figura muestra, para diferentes duraciones del segmento multimedia, cual es AL-FEC code rate máximo que puede utilizarse sin que se supere el 10 % de retransmisiones HTTP. Estos puntos se han etiquetado en la figura como $MCR_{0.5}$, MCR_1 y MCR_2 .

más redundancia para proporcionar el mismo nivel de protección. De la misma forma, dado un SDR fijo, la tasa de error aumenta si se reduce la duración del segmento, lo que implica un aumento del número de retransmisiones HTTP.

Para determinar si es posible reducir la duración del segmento para reducir el retardo inicial de buffering se deben analizar dos aspectos:

1. La relación entre la duración del segmento y el rendimiento del decodificador AL-FEC.
2. El impacto de la duración del segmento en el retardo inicial de buffering

La Figura 4.4 muestra la relación entre el AL-FEC code rate, el porcentaje de segmentos recuperados por HTTP y la duración del segmento multimedia. Los resultados obtenidos pueden utilizarse para seleccionar el AL-FEC code rate que maximiza el SDR y limita el porcentaje de retransmisiones HTTP al 10 % (90 % de cobertura). La Figura 4.4 muestra que, para segmentos de dos segundos, el AL-FEC code rate no puede ser mayor que 0.90 (MCR_2). Para segmentos de un segundo, 0.86 (MCR_1), y para segmentos de 0.5 segundos, 0.78 ($MCR_{0.5}$).

Como se ha comentado con anterioridad, es recomendable codificar el contenido multimedia a 1 Mbps para su transmisión sobre eMBMS [LG12]. Utilizar segmentos de 0.5 segundos y un AL-FEC code rate de 0.78 da cumplimiento a ese requisito, y además, mantiene las retransmisiones unicast por debajo del 10 % ($MCR_{0.5}$ en la Figura 4.4).

Reducir la duración del segmento a 250 ms afecta de forma significativa al rendimiento del decodificador AL-FEC. Por ejemplo, utilizar un AL-FEC code rate de 0.55 no es suficiente para mantener las retransmisiones unicast por debajo del 10 % (se alcanza un 13,39 %). Pero,

tampoco permitiría codificar el vídeo a 1 Mbps, ya que solo se podría codificar con un bitrate máximo de 0.69 Mbps.

A continuación, se analiza el impacto de la duración del segmento en el retardo inicial de buffering. La Figura 4.5 muestra el nivel de buffer mínimo (en segundos) que evita las paradas de servicio en el 90 % de los terminales si se consideran los valores del AL-FEC code rate y de la duración del segmento que se han identificado anteriormente.

El análisis se realiza aplicando el método que se ha descrito en la Sección 4.2, es decir, teniendo en cuenta que el retardo que añade la recuperación unicast de los segmentos afecta al nivel del buffer. Los resultados muestran que el nivel inicial del buffer es mayor si aumenta el retardo para recuperar los segmentos perdidos por HTTP. Por ejemplo, un retardo de $1,25 \cdot T_{sgmnt}$ requiere un buffer mínimo de diez segundos si se utiliza un AL-FEC code rate de 0.9 y los segmentos multimedia tienen una duración de dos segundos. Para un retardo de $2 \cdot T_{sgmnt}$ el buffer mínimo aumenta por encima de los 30 segundos si se consideran los mismos valores de CR y T_{sgmnt} .

La Figura 4.5 muestra que es posible utilizar segmentos de corta duración para reducir el retardo inicial de buffering. Como se puede observar, el buffer mínimo que se requiere cuando se utilizan segmentos de dos segundos, es significativamente mayor que el que se requiere cuando se utilizan segmentos de 0.5 segundos. Por ejemplo, para un retardo de $1,25 \cdot T_{sgmnt}$, el nivel de buffer inicial se reduce 8 s si en lugar de utilizar segmentos de 2 segundos se utilizan de 0.5 s.

Una vez que se han definido todos los parámetros del servicio, incluida la duración del

Figura 4.5: Impacto de la duración del segmento en el retardo inicial de buffering.

segmento, se puede generar la representación multimedia y el fichero MPD. A través del MPD es posible comunicar el retardo inicial de buffering a los reproductores mediante el campo *minBufferTime*.

4.5. Evaluación de la calidad multimedia

Con el objetivo de entender la relación que existe entre las paradas de servicio, el retardo inicial de buffering y la calidad de la transmisión multimedia, se han usado dos enfoques diferentes. El primero consiste en calcular el PSNR para comparar la señal de vídeo original con la recibida por el terminal. El segundo, relaciona el número de eventos de parada de cada cliente con los resultados obtenidos tras aplicar un modelo subjetivo para la evaluación de la QoE [HSH⁺11].

En los servicios de streaming sobre UDP, los fotogramas de vídeo que recibe el reproductor pueden estar distorsionados por diferentes factores, como, por ejemplo, la codificación del vídeo o la pérdida de paquetes que se produce durante la transmisión del contenido por el canal. En un servicio de streaming 3GP-DASH sobre eMBMS, el PSNR solo puede verse afectado por la codificación del vídeo y/o los eventos de parada del servicio. Aunque el PSNR no se ha diseñado específicamente para evaluar el impacto de las paradas de servicio en la calidad percibida, otros autores [KO13] ya han utilizado un modelo basado en el PSNR para entender cual sería su impacto.

El modelo de Kumar y Oyman [KO13] asume que los segmentos multimedia se componen de un número entero de fotogramas y que, si un segmento se pierde durante su transmisión sobre eMBMS, el reproductor mostrará por pantalla el último fotograma del último segmento que se recibió correctamente. Es decir, el modelo asume que los segmentos que no se reciben antes del tiempo límite de reproducción (*PlaybackDeadline*) se descartan y no son reproducidos.

Para aplicar este modelo, y evaluar la propuesta que se realiza en este capítulo desde la perspectiva de la calidad de la experiencia, se ha utilizado el código publicado por la UIT-T

Figura 4.6: Y-PSNR en función del usuario percentil cuando el retardo para recuperar los segmentos por HTTP es de $2 \cdot T_{sgmnt}$

para calcular el PSNR [ITU10]. El código se ha integrado con la plataforma descrita en el Capítulo 3, que combina modelos analíticos y de simulación. Cabe destacar que el cálculo del PSNR se realiza considerando solo la componente de luminancia (Y-PSNR). Sin embargo, se ha comprobado que la visión es más sensible al brillo que a otros factores [Bin15], por lo que este cálculo refleja bien la calidad percibida por los usuarios.

La Figura 4.6 muestra el Y-PSNR en función del usuario percentil cuando la duración del segmento multimedia es de 0.5 s. Para mostrar con claridad el impacto de las paradas de servicio en el Y-PSNR, se ha representado el caso de mayor retardo, es decir, cuando el retardo para recuperar los segmentos por HTTP es $2 \cdot T_{sgmnt}$. La Figura 4.6 muestra que es posible garantizar la continuidad de la reproducción multimedia en el usuario 10 percentil si el nivel de buffer es de 6 segundos. Esto es consistente con los resultados que se muestran en la Figura 4.5, donde se observa que es necesario un buffer de 6 segundos para evitar las paradas del servicio (segmentos de 0.5 s y $CR = 0,78$). Si el nivel del buffer es inferior a los seis segundos, se producen interrupciones del servicio que degradan la calidad percibida.

El segundo enfoque que se ha abordado consiste en utilizar los resultados obtenidos por Hoßfeld *et al.* [HSH⁺11] para evaluar la calidad del streaming multimedia. Estos autores aplican un modelo subjetivo de QoE que permite relacionar los eventos de parada de servicio con la calidad percibida. Los resultados obtenidos indican que la calidad percibida se degrada severamente cuando se producen tres o más eventos de parada.

Siguiendo este enfoque, se ha utilizado la plataforma descrita en el Capítulo 3 para calcular

Figura 4.7: Porcentaje de usuarios afectados por los eventos de parada en función del nivel del buffer: 2, 4 y 6 segundos. El retardo que impone la recuperación de segmentos por HTTP es de $2 \cdot T_{sgmnt}$.

el número de eventos de parada que se detecta en cada reproductor. La Figura 4.7 muestra la relación entre el nivel inicial del buffer y el porcentaje de usuarios que perciben 0, 1, 2, 3 o más de eventos de parada. Los resultados muestran que un buffer de dos segundos es demasiado corto, puesto que, como se muestra en la figura, el 30.4 % de los usuarios experimentaría tres o más paradas. Utilizar un buffer de cuatro segundos mejora este resultado, pero el 8.5 % de los usuarios sigue sin obtener niveles satisfactorios de QoE. Finalmente, la Figura 4.7 muestra que es posible evitar la degradación de la calidad percibida en el 95 % de los usuarios, si se utiliza un buffer de seis segundos.

4.6. Conclusiones

Los estándares del 3GPP [3GP17a, 3GP17c] definen el servicio de streaming 3GP-DASH sobre eMBMS, servicio que combina de forma novedosa diferentes mecanismos, protocolos y técnicas para el streaming multimedia. El contenido se codifica como indica el estándar de DASH, pero no con el objetivo de generar diferentes versiones del mismo contenido, sino para fragmentar el flujo multimedia en segmentos de la misma duración.

Para proporcionar más fiabilidad al envío de los segmentos sobre eMBMS, el 3GPP propone aplicar técnicas de corrección de errores a nivel de aplicación (AL-FEC). Sin embargo, y debido a errores de transmisión, es posible que los segmentos multimedia no se reciban en los terminales incluso si se utilizan este tipo de códigos. En ese caso, se propone utilizar HTTP para que los segmentos perdidos se puedan recuperar a través de una conexión unicast.

La combinación de todos estos elementos afecta a la latencia del servicio. Hasta la fecha, no se había realizado un análisis de retardo como el que se presenta en este capítulo, describiendo todos los retardos que puede sufrir un segmento con formato DASH, desde que se genera en el servidor de contenidos hasta que se reproduce en el terminal del cliente. Esto ha permitido detectar posibles errores de sincronización que se pueden resolver determinando adecuadamente el campo *availabilityStartTime* del MPD, que indica instante de tiempo a partir del cual un segmento está disponible para ser enviado.

El capítulo también describe un método que permite reducir al mínimo el retardo inicial de buffering sin afectar a la continuidad de la reproducción multimedia. Además, se discute la posibilidad de reducir aun más la latencia mediante el uso de segmentos multimedia de corta duración. Esta solución puede aplicarse en el marco de los servicios 3GP-DASH sobre eMBMS porque la mayor parte de los segmentos se entrega correctamente durante su transmisión multicast, y por tanto, no se introduce una sobrecarga significativa en clientes, servidor o infraestructura de la red por el aumento de las transacciones HTTP.

El impacto de las soluciones propuestas en el retardo del servicio es significativo, pero el impacto en el diseño de los dispositivos móviles es mínimo, puesto que solo se requiere la correcta definición de los parámetros del servicio y del fichero MPD.

Las soluciones propuestas también se han evaluado desde la perspectiva de la QoE. Para

ello, se han abordado dos enfoques: utilizar un modelo basado en el cálculo del PSNR, y utilizar resultados de test subjetivos de QoE. Sin embargo, y como se explica Capítulo 5, ambos enfoques presentan limitaciones, por lo que se propone utilizar un nuevo modelo objetivo de QoE que está en proceso de estandarización.

El contenido de este capítulo se ha publicado en un artículo situado en el Q1 del JCR [LBP17b].

Capítulo 5

Análisis de los servicios de difusión de contenidos desde la perspectiva de la QoE

En el Capítulo 4 se ha adoptado un doble enfoque para evaluar la calidad del streaming multimedia. El primer enfoque, que consiste en calcular el Y-PSNR, no parece ser muy apropiado porque se basa en la comparación pixel-a-pixel entre el fotograma original y el fotograma recibido. Debido a que el modelo no está específicamente diseñado para analizar el impacto de las paradas de servicio en la calidad percibida, se propuso un segundo enfoque.

Este segundo enfoque utiliza resultados de test subjetivos de QoE obtenidos en el contexto de un servicio de streaming convencional sobre HTTP [HSH⁺11]. Los autores relacionan el número de eventos de parada con la calidad percibida. Para evaluar el servicio de streaming 3GP-DASH sobre eMBMS se ha utilizado esta misma relación, calculando el número de eventos de parada que tiene lugar en cada reproductor. Sin embargo, este enfoque también podría ser inadecuado, puesto que los resultados que han obtenido Hoßfeld *et al.* se centran en los servicios de streaming adaptativo sobre HTTP.

En este capítulo se propone aplicar otro modelo para analizar el servicio desde la perspectiva de la calidad de la experiencia. En particular, el capítulo analiza el impacto de diferentes parámetros de servicio (tanto de red como de aplicación) en la calidad percibida. La evaluación de la QoE permite determinar el valor más adecuado de estos parámetros. La Figura 5.1 resume

Figura 5.1: Definición del servicio de streaming 3GP-DASH sobre eMBMS. La figura muestra un diagrama de bloques que expresa cómo se puede evaluar el servicio desde la perspectiva de la QoE, y como esta evaluación puede utilizarse para definir correctamente diferentes parámetros de red y aplicación.

esta idea, que será explicada con más detalle en las siguientes secciones.

Aunque podría utilizarse un modelo de QoE subjetivo, preparar y realizar las encuestas es costoso, requiere mucho tiempo. Por esta razón se ha optado por aplicar un modelo de QoE objetivo. En particular, un modelo NR, que como se explicó en el Capítulo 2, proporciona una correlación entre los parámetros de red o aplicación y la calidad percibida.

El modelo video Mean Opinion Score (vMOS) [HUA], que está siendo estandarizado en el seno del grupo de trabajo 12 del UIT-T, es de tipo NR (Capítulo 2). vMOS utiliza la misma escala de puntuaciones que se muestra en la Tabla 2.1 (MOS) pero su nombre hace referencia a un modelo de evaluación objetiva de QoE, no a una escala. vMOS surge por la necesidad de proporcionar un estándar centrado en el usuario que permita medir la calidad de los servicios de streaming con independencia del escenario de red o de las características de los dispositivos de los usuarios. El modelo vMOS no se ha concebido específicamente para analizar los servicios de streaming 3GP-DASH sobre eMBMS, pero en este capítulo se explica que es posible aplicarlo en este contexto, y que su uso permite definir el servicio desde la perspectiva de la QoE especificando diferentes parámetros de red y aplicación.

La estructura del capítulo es la siguiente: la Sección 5.1 presenta vMOS, nuevo estándar para medir la calidad de los servicios de streaming multimedia. La Sección 5.2 describe el análisis del servicio desde la perspectiva de la calidad de la experiencia, y explica, como a partir de este análisis pueden determinarse diferentes parámetros del servicio. Finalmente, la Sección 5.3 proporciona las conclusiones de este trabajo.

5.1. Hacia la estandarización de un nuevo modelo para medir la calidad de los servicios de streaming multimedia

El modelo vMOS puede utilizarse para analizar los servicios de streaming 3GP-DASH sobre eMBMS por dos razones. Primero, porque proporciona un marco general para medir la calidad de los servicios de streaming independientemente de la red que da soporte al servicio o del dispositivo que reproduce el contenido. Segundo, porque las métricas de QoE en que se basa, son las que afectan a la calidad que perciben los usuarios de los servicios de streaming 3GP-DASH sobre eMBMS: calidad técnica del vídeo, paradas del servicio y retardo inicial de buffering.

La solución propuesta por Huawei, llamada U-vMOS fue aprobada por el UIT-T y constituye una pieza clave para la estandarización del modelo vMOS. U-vMOS puede calcularse a partir de tres factores [HUA, PCWT16]: sQuality, sStalling y sLoading.

sQuality determina en qué grado la calidad técnica del vídeo afecta a la calidad percibida. El modelo U-vMOS considera varios parámetros relacionados con la codificación del vídeo para medir esta calidad: bitrate, codec de vídeo, codec profile, y resolución.

Por otro lado, sLoading y sStalling miden, respectivamente, el impacto del retardo inicial de buffering y las paradas de servicio en la calidad percibida. En este caso, su cálculo se realiza utilizando diferentes umbrales. Por ejemplo, si el retardo inicial de buffering es menor que

Figura 5.2: Diagrama de bloques de la plataforma experimental utilizada para medir la calidad percibida por los usuarios

0.1 s, el valor que adquiere sLoading es máximo. A medida que el retardo aumenta, el factor sLoading se reduce hasta alcanzar su valor mínimo, que se obtiene cuando el retardo inicial de buffering es mayor a 10 s. Respecto al factor sStalling, su cálculo depende del porcentaje de tiempo en que la reproducción ha estado detenida. Si el porcentaje de parada es superior al 15 %, el valor de sStalling se reduce significativamente.

U-vMOS predice la calidad percibida estableciendo una relación entre todas estas métricas objetivas de QoE. Para validar el modelo de predicción sus diseñadores han realizado test subjetivos de QoE. De esta forma, se garantiza que la puntuación que se obtiene aplicando el modelo U-vMOS refleja adecuadamente la calidad que sería percibida por los usuarios en la realidad.

5.2. Análisis del servicio de streaming 3GP-DASH sobre eMBMS utilizando el modelo U-vMOS

Esta sección muestra cómo se pueden definir los parámetros del servicio desde la perspectiva de la calidad de la experiencia. Para ello, se analiza el servicio de streaming 3GP-DASH sobre eMBMS utilizando el modelo U-vMOS.

Aplicar el modelo U-vMOS ha requerido extender la plataforma de experimentación descrita en el Capítulo 3, combinando el modelo analítico y de simulación, con una calculadora de U-vMOS a la que se puede acceder desde Internet¹. Así, la Figura 5.2 muestra que la plataforma experimental se compone ahora de tres bloques: transmisión eMBMS, reproductor DASH y la calculadora U-vMOS. Los bloques transmisión eMBMS y reproductor DASH permiten obtener el tiempo promedio de parada en cada reproductor para diferentes combinaciones de los parámetros de aplicación (e.g., retardo inicial de buffering, bitrate y AL-FEC code rate) y red (e.g., MCS y ancho de banda del canal unicast). Esto hace posible introducir el valor de las métricas de QoE que considera el modelo objetivo (Sección 5.1) en la calculadora U-vMOS.

Los parámetros de red que configuran la transmisión eMBMS (Tabla 5.1) se han ajustado para que el ancho de banda del canal permita el envío de vídeo codificado con H.265 a 1080p, junto con los datos de redundancia que genera el codificador AL-FEC. Esta decisión se debe a que resultados preliminares mostraban que, según el modelo U-vMOS, no es posible alcanzar una buena QoE si el vídeo se codifica con H.264 a 720p.

Si se asume que el ancho de banda del canal es 10 MHz, que solo se asigna una subtrama al servicio eMBMS, y que se selecciona el índice 25 del MCS, el ancho de banda máximo del

¹“Calculadora de U-vMOS,”[Online]. Available: <http://www.mbbllab.com:9090/mobilemos/index.php?rDsite/calculation>

Parámetro	Valor
Número de terminales	399
Inter-Site Distance (ISD)	500 m
Tamaño del área MBSFN	7 eNBs
Ancho de banda del canal	10 MHz
Subtramas reservadas al servicio eMBMS	1
Índice del esquema de modulación y codificación (MCS)	25
R_{eMBMS}	2.8 Mbps
$BW_{unicast}$	0.5 / 1.1 / 1.7 / 2.2 Mbps
AL-FEC code rate	0.4 / 0.5 / 0.6 / 0.7 / 0.8
Duración del video	300 s
Codec de video	H.265
Resolución del video	1080p
Bitrate de la representación B_{eMBMS}	1.1 / 1.3 / 1.6 / 1.8 / 2.1 Mbps
Duración del segmento (T_{sgmnt})	0.2 / 0.5 / 1 / 2 s
retardo inicial de buffering (mBT)	0.5 / 1 / 2 / 4 / 6 s

Tabla 5.1: Parámetros de red y aplicación considerados para analizar el servicio desde la perspectiva de la calidad de la experiencia

canal eMBMS es 2.8 Mbps. El Capítulo 4 explica que la tasa de datos del servicio depende del ancho de banda del canal y del AL-FEC code rate (ecuación 4.7). Esto significa que la tasa de datos también se relaciona con el tiempo de parada del servicio a través del AL-FEC code rate (si no se protege el envío de los segmentos, el tiempo de parada puede aumentar). El análisis que se detalla a continuación considera varios valores del bitrate de la representación eMBMS (desde 1.1 Mbps a 2.2 Mbps), que se corresponden con varios valores del AL-FEC code rate (desde 0.4 a 0.8). Tal y como muestra la Tabla 5.1, se usan valores fijos para otros parámetros del servicio como el número de terminales, el ISD o el tamaño del área MBSFN.

5.2.1. Relación entre la tasa de datos del servicio y las retransmisiones HTTP

Parámetros relativos a la codificación del vídeo, como el bitrate o la duración de los segmentos, afectan a la fiabilidad de la transmisión del contenido multimedia sobre eMBMS. La Figura 5.3 muestra el porcentaje de segmentos que deben ser recuperados por HTTP en función de la tasa de datos del servicio. Esta se relaciona con el AL-FEC code rate y el ancho de banda del canal eMBMS como muestra la ecuación 4.7. Para un ancho de banda de 2.8 Mbps, y una sobrecarga del 5 %, el AL-FEC code rate debe ajustarse a 0.4, 0.5, 0.6, 0.7 y 0.8 para obtener las siguientes tasas de datos: 1077 Kbps, 1346 Kbps, 1615 Kbps, 1884 Kbps y 2153 Kbps.

Los resultados se muestran para diferentes valores de la duración del segmento (0.2 s, 0.5 s, 1 s y 2 s). La Figura 5.3 muestra que el porcentaje de segmentos recuperados por HTTP es mayor si la duración del segmento es menor. Reducir la duración de 2 s a 0.5 s hace que

Figura 5.3: Relación entre la tasa de datos del servicio, la duración del segmento multimedia (T_{sgmnt}) y el porcentaje de segmentos que deben ser recuperados con HTTP.

el porcentaje de retransmisiones unicast aumente en un 2 % aproximadamente. El incremento es significativamente mayor cuando se utilizan segmentos de 0.2 s. Esto se debe a que la codificación AL-FEC es menos eficiente cuando se utilizan bloques de datos más pequeños, pero también, porque el codec de vídeo H.265 es menos eficiente para segmentos de corta duración [HPW⁺16].

Hooft *et al.* [HPW⁺16] discuten que es posible utilizar segmentos de corta duración, sin afectar a la calidad percibida por los usuarios, si se asume una sobrecarga del bitrate en la codificación con H.265. Con el objetivo de limitar esta sobrecarga, recomiendan utilizar segmentos con una duración mayor o igual a 0.5 s. Por esta razón se han descartado los segmentos de 0.2 s en los siguientes análisis.

La Figura 5.3 también muestra que el porcentaje de segmentos recuperados por HTTP aumenta para tasas de datos mayores. Esto es lógico, puesto que el ancho de banda del canal eMBMS es fijo y el codificador AL-FEC puede añadir menos datos de redundancia a los segmentos multimedia. Por ejemplo, para segmentos de 0.5 s, la Figura 5.3 muestra que el porcentaje de retransmisiones unicast es del 6 % para un tasa de datos del servicio de 1077 Kbps (SR_{1077}). El porcentaje de retransmisiones aumenta hasta un 15 % cuando la tasa de datos del servicio es 2153 Kbps (SR_{2153}).

La recuperación de los segmentos por HTTP también puede tener un impacto significativo en la sobrecarga de los canales ascendente y descendente de la red móvil [3GP17a]. Esto depende del número de segmentos que se recuperan por HTTP, lo que a su vez, depende del número de usuarios que hay en el área eMBMS. Para el escenario analizado, es decir, considerando 399 terminales distribuidos uniformemente en un área MBSFN de 7 celdas, un 15 %

de retransmisiones unicast equivalente a 9 clientes por celda solicitando al mismo tiempo segmentos por HTTP. Si la capacidad del canal ascendente y descendente es, por ejemplo, 20 y 30 Mbps respectivamente, y el tamaño de las peticiones HTTP es de 550 bytes, la recuperación de errores por HTTP impondría una sobrecarga en el canal ascendente del 0.44 % y de un 65 % en el canal descendente. Reducir esta sobrecarga se plantea como un paso futuro que debe resolver la investigación postdoctoral.

5.2.2. Relación entre el retardo inicial de buffering, el bitrate de la representación eMBMS y el tiempo de parada del servicio

Como se discutió en el Capítulo 4, la principal causa de las interrupciones del servicio es la necesidad de recuperar por HTTP aquellos segmentos que no se han transmitido correctamente sobre eMBMS. Es decir, el retardo que añade la recuperación unicast de los segmentos puede hacer que el buffer del reproductor se vacíe y se produzca una interrupción del servicio. También se ha discutido en este capítulo que el bitrate de la representación eMBMS determina el porcentaje de segmentos recuperados por HTTP. Ambos factores afectan a la frecuencia y duración de los eventos de parada. A continuación, se analiza la relación entre el tiempo promedio de parada, el retardo inicial de buffering y el bitrate de la representación eMBMS.

La Figura 5.4 representa el tiempo promedio de parada (399 terminales) para diferentes valores del retardo inicial de buffering (mBT igual a 2 s, 4 s y 6 s) y diferentes valores de la duración del segmento (T_{sgmnt} igual 0.5 s, 1 s y 2 s). Con el objetivo de analizar el impacto del retardo inicial de buffering en el tiempo promedio de parada se ha utilizado un ancho de banda

Figura 5.4: Relación entre retardo inicial de buffering (mBT), bitrate de la representación eMBMS (B_{eMBMS}) y el tiempo promedio de parada del servicio. Los resultados se obtienen para diferentes valores de la duración del segmento multimedia (T_{sgmnt}). El ancho de banda unicast ($BW_{unicast}$) es igual a 500 Kbps.

unicast de 500 Kbps, menor que el bitrate de la representación de menor calidad (Tabla 5.1).

La Figura 5.4 muestra que el tiempo promedio de parada se reduce cuando se utilizan bitrates más bajos. Esto se debe a que el codificador AL-FEC puede añadir más datos redundancia para proteger el envío de los segmentos multimedia, lo que aumenta la probabilidad de que estos se entreguen correctamente sobre la sesión FLUTE.

La Figura 5.4 también muestra que es posible reducir el tiempo promedio de parada utilizando segmentos de menor duración. Por ejemplo, para un retardo inicial de buffering de 4 s, el tiempo promedio de parada es menor si se utilizan segmentos de 0.5 s en lugar de segmentos de 1 s o 2 s. El mismo comportamiento se observa cuando el retardo inicial de buffering es de 6 s.

Los resultados obtenidos muestran que el comportamiento es diferente si el retardo inicial de buffering es de 2 s. La Figura 5.4 muestra que si se codifica el video a 2153 Kbps, el tiempo promedio de parada es ligeramente superior cuando se utilizan segmentos de 0.5 s. La falta de ancho de banda y el bajo nivel del buffer no permiten que la recuperación de los segmentos por HTTP se realice a tiempo. Además, el porcentaje de segmentos recuperados por HTTP es mayor cuando la duración del segmento es menor, tal y como mostraba la Figura 5.3.

Por lo tanto, la conclusión es que, en lo que respecta al tiempo promedio de parada, no hay mucha diferencia entre utilizar segmentos de 0.5 s, 1 s o 2 s. Es por ello que los siguientes análisis se centrarán únicamente en segmentos de 0.5 s, ya que su uso permite reducir la latencia del servicio, y como se verá más adelante, mejorar la calidad percibida.

5.2.3. Definición de los parámetros de servicio a partir de la evaluación de la QoE

Aunque se ha analizado la relación entre el retardo inicial de buffering, el bitrate de la representación eMBMS y el tiempo promedio de parada, no está claro en qué medida estos parámetros afectan a la calidad percibida por los usuarios. Utilizar el modelo U-vMOS permite analizar el servicio de streaming 3GP-DASH sobre eMBMS desde la perspectiva de la calidad de la experiencia. Así, es posible determinar correctamente diferentes parámetros del servicio, tanto de red, como de aplicación.

El enfoque que se propone para analizar el servicio puede ser útil para los proveedores de contenidos, que deben definir el servicio especificando diferentes parámetros relacionados con la codificación del contenido multimedia (Capítulo 2). También es útil para los operadores de red, responsables de determinar los parámetros de red que dan soporte a la transmisión de dicho contenido sobre la red móvil.

Para analizar el servicio se consideró inicialmente la codificación de vídeo con H.264 a una resolución de 720p. Sin embargo, los resultados preliminares mostraban que, según el modelo U-vMOS, no es posible proporcionar una buena QoE bajo estas condiciones. Es por ello que se optó finalmente por analizar la transmisión de vídeo codificado con H.265 a una resolución de 1080p.

La Figura 5.5 muestra el U-vMOS promedio para diferentes valores del retardo inicial de

buffering: 2 s (Figura 5.5a), 1 s (Figura 5.5b), y 0.5 s (Figura 5.5c). Los resultados obtenidos confirman que los servicios de streaming en directo son muy sensibles a retardos. Un retardo inicial de buffering de 1 s o 2 s es demasiado elevado como para proporcionar puntuaciones altas de U-vMOS. Sin embargo, un retardo inicial de 0.5 s proporciona una alta calidad de la experiencia siempre que el ancho de banda unicast sea mayor que el bitrate de la representación eMBMS. Seufert *et al.* [SES⁺15] apuntan que, en general, la latencia del servicio tiene menor importancia que las paradas de servicio o la calidad técnica del vídeo en lo que respecta a la calidad percibida. Sin embargo, en el contexto de un servicio de streaming 3GP-DASH sobre eMBMS, reducir el retardo inicial de buffering no implica un incremento significativo del tiempo promedio de parada, básicamente, porque la mayoría de los segmentos se entregan correctamente sobre eMBMS. Por lo tanto, y en este contexto, es recomendable reducir el retardo inicial de buffering para mejorar la calidad percibida por los usuarios.

La Figura 5.5 también muestra el impacto del bitrate en el U-vMOS. El ancho de banda unicast ($BW_{unicast}$) se ha seleccionado de forma que siempre exista un caso en el que su valor sea mayor que el bitrate de la representación que se envía sobre eMBMS. En ese caso, no se producen paradas del servicio porque los segmentos se recuperan por HTTP a tiempo. La Figura 5.5 muestra que la puntuación U-vMOS más alta se obtiene para la representación de mayor calidad ($B_{eMBMS} = 2153$ Kbps), y cuando $BW_{unicast} > B_{eMBMS}$ (en la figura, 2200 Kbps). Usar representaciones con menor bitrate en ese caso implica obtener peor QoE para el mismo ancho de banda. Aunque no se producen paradas de servicio la reducción del bitrate degrada la calidad percibida.

Para un bitrate fijo, la Figura 5.5 muestra que se obtienen peores puntuaciones del U-vMOS para anchos de banda unicast menores. Esto se debe a que el tiempo de parada es mayor cuando crece la diferencia entre el bitrate de la representación eMBMS y el ancho de banda unicast. Sin embargo, la Figura 5.5c muestra que para un retardo inicial de buffering de 0.5 s, y un bitrate de 2153 Kbps, se obtiene la misma puntuación para diferentes valores del ancho de banda unicast. Esto ocurre porque no hay suficiente ancho de banda o datos en el buffer del reproductor para

Figura 5.5: Puntuación U-vMOS para diferentes retardos de buffering inicial: 2 s (a), 1 s (b) y 0.5 s (c). Los resultados se obtienen para diferentes valores del ancho de banda unicast ($BW_{unicast}$) y del bitrate de la representación enviada por eMBMS (B_{eMBMS}).

Figura 5.6: Análisis del impacto del retardo de buffering inicial en la calidad percibida en función del usuario percentil. El bitrate de la representación eMBMS (B_{eMBMS}) es 2153 Kbps, mientras que el ancho de banda unicast ($BW_{unicast}$) es de 1100 Kbps. Los resultados se muestran para diferentes valores del retardo inicial de buffering (mBT).

que los segmentos perdidos en el canal eMBMS se recuperen por HTTP a tiempo.

Para un retardo inicial de buffering de 0.5 s se obtiene cuatro veces una puntuación superior a 4, por lo que la pregunta que cabría hacerse es, ¿cual es el bitrate más adecuado para codificar un video que se transmite sobre eMBMS? Para el despliegue del servicio que se considera en esta evaluación, la Figura 5.5c muestra que un bitrate de 1077 Kbps proporciona puntuaciones más altas para anchos de banda menores. Esto se debe a que el codificador AL-FEC puede introducir más redundancia para proteger los segmentos, reduciéndose así, el porcentaje de retransmisiones unicast. Sin embargo, no es posible alcanzar una puntuación de 4 en ningún caso.

Proporcionar una buena QoE ($U-vMOS > 4$) requiere reducir el retardo inicial de buffering (0.5 s) y codificar el contenido multimedia utilizando el máximo bitrate posible (2153 Kbps). Sin embargo, cuando el ancho de banda unicast cae por debajo del bitrate de la representación eMBMS, los eventos de parada son más frecuentes y la calidad percibida se degrada.

Para finalizar, la Figura 5.6 analiza el impacto del retardo inicial de buffering en la calidad percibida. Las puntuaciones de U-vMOS se expresan en función del percentil de usuarios. Los resultados se muestran para vídeo codificado a 2153 Kbps (la representación de mayor calidad) y un ancho de banda unicast de 1100 Kbps. La Figura 5.6 muestra que, si el retardo inicial de buffering es 0.5 s, la puntuación que otorga el 35 % de usuarios a la calidad del servicio es mala. A partir de ese punto, la condición de canal de los clientes mejora, es decir, la transmisión sobre eMBMS es más fiable, por lo que se requieren menos retransmisiones unicast. La consecuencia de esto es que la calidad percibida mejora. De hecho, la puntuación es máxima cuando todos los segmentos se reciben correctamente sobre el canal eMBMS.

La Figura 5.6 también muestra que para un retardo inicial mayor, el porcentaje de usuarios que valora mal el servicio se reduce. La figura muestra que esto es posible si se incrementa el retardo inicial a 2 s. Aumentar el retardo a 4 s no tiene efectos significativos. El número de usuarios que otorgan una mala puntuación no es mucho menor. Es más, también causa una reducción de la calidad percibida en otros usuarios que antes daban puntuaciones más altas.

En definitiva, los clientes con buena condición de canal pueden beneficiarse de un retardo de buffering inicial bajo, que proporciona la QoE máxima. Por el contrario, clientes con mala condición de canal requieren retardos de buffering inicial mayores para que su percepción sobre el servicio mejore. Esto se debe a que con ello se evitan las paradas del servicio.

5.3. Conclusiones

Este capítulo muestra que el modelo vMOS, en proceso de estandarización, puede utilizarse para analizar un servicio de streaming 3GP-DASH sobre eMBMS desde la perspectiva de la calidad de la experiencia. Utilizando el modelo U-vMOS, aprobado por el UIT-T como pieza clave para la estandarización de vMOS, es posible entender el impacto de diferentes parámetros del servicio, como el bitrate o el retardo inicial de buffering, en la calidad percibida.

Un análisis basado en la QoE como el que se describe en este capítulo, puede aplicarse a cualquier despliegue del servicio para comprender cómo deben determinarse los parámetros que lo definen. Del análisis del escenario que se ha planteado en este capítulo se han extraído conclusiones interesantes. Por ejemplo, los resultados obtenidos muestran que se puede mejorar la QoE utilizando representaciones de vídeo codificadas con un bitrate menor. La puntuación máxima que puede alcanzarse no es tan alta, pero el promedio es mayor.

En el Capítulo 4 se propone un método para calcular el buffer mínimo que evita las paradas del servicio. Desde la perspectiva de la QoE, se ha mostrado que retardos de buffering mayores permiten mejorar la calidad percibida en los clientes donde se recuperan más segmentos por HTTP. Sin embargo, el incremento del retardo también reduce la calidad percibida en los clientes donde no existe la necesidad de utilizar HTTP para recuperar los segmentos.

En definitiva, la metodología que se propone en este capítulo permite definir los servicios de streaming 3GP-DASH sobre eMBMS especificando parámetros relacionados con la codificación del vídeo y con la transmisión del contenido sobre eMBMS. Por ejemplo, permite determinar el bitrate o el retardo de buffering inicial más adecuados para maximizar la calidad percibida.

El contenido de este capítulo se ha publicado en un artículo situado en el Q1 del JCR [LBC⁺17].

Capítulo 6

Difusión de contenidos multimedia asistida por la red

Los servicios de streaming 3GP-DASH sobre eMBMS son sensibles a retardos. De hecho, los resultados presentados en el Capítulo 5 vienen a confirmar que el retardo inicial de buffering es un parámetro que afecta significativamente a la calidad de la experiencia.

Dimensionar adecuadamente el retardo inicial de buffering es una solución para evitar las paradas de servicio. El Capítulo 4 propone un método para calcular el buffer mínimo que evita las paradas de servicio. Los clientes con mala condición de canal pueden beneficiarse de experimentar retardos mayores para obtener una QoE mejor, puesto que son estos clientes los que más segmentos solicitan por HTTP, y los que sufren más paradas de servicio. Sin embargo, aumentar el retardo en clientes con buena condición de canal, es decir, en clientes que reciben correctamente los segmentos por eMBMS, es innecesario e implica degradar su calidad percibida.

Los requisitos de buffering de los servicios de streaming adaptativo sobre HTTP son diferentes. Los reproductores multimedia se adaptan al ancho de banda disponible conmutando entre diferentes representaciones, y esto, puede evitar que el buffer del reproductor se vacíe. En un servicio de streaming 3GP-DASH sobre eMBMS esta opción no se ha considerado debido a que el canal multicast solo transporta una representación multimedia. El bitrate de esta representación se ajusta al ancho de banda del canal eMBMS, cuyo valor es constante durante toda la sesión debido a que, para este servicio, los parámetros de red no varían.

El ancho de banda del canal unicast no puede determinarse, ya que en este caso, los parámetros de red sí varían de forma dinámica. El MCS se selecciona dinámicamente para maximizar la tasa de transmisión dada una probabilidad de error. La cantidad de recursos radio (PRB) que se asignan a un canal unicast también varía dinámicamente. Las estaciones base ejecutan un algoritmo de planificación que asigna los recursos a los diferentes canales a partir de los reportes de calidad que envían los terminales y otra información de capas superiores, como el nivel del buffer del reproductor u otros parámetros de QoS.

El problema es que, cuando se recuperan los segmentos por HTTP, el ancho de banda disponible para el cliente puede ser menor que el bitrate de la representación que se envía sobre eMBMS, lo que añade un retardo adicional que puede agotar el buffer del reproductor si no se

ha dimensionado adecuadamente.

En este capítulo se propone un mecanismo que da soporte a la recuperación adaptativa de errores en un servicio de streaming 3GP-DASH sobre eMBMS. El diseño de este mecanismo considera las nuevas tendencias de streaming adaptativo basado en SAND (Capítulo 2). El mecanismo hace posible que el reproductor multimedia pueda conmutar a la representación que más se ajusta al ancho de banda disponible en el canal unicast durante la recuperación de los segmentos por HTTP. Aunque la conmutación entre las representaciones puede introducir cierta inestabilidad en la reproducción, es posible evitar las paradas del servicio sin necesidad de aumentar el retardo inicial de buffering, lo que podría mejorar la calidad de la experiencia en todos los clientes (Capítulo 5).

El resto del capítulo se organiza de la siguiente manera: la Sección 6.1 explica, mediante un ejemplo, cómo se produce una parada del servicio. La Sección 6.2 describe el fichero de metadatos que utiliza el reproductor para conmutar entre las representaciones disponibles. En la Sección 6.3 se detalla el mecanismo propuesto para la recuperación adaptativa de errores. La Sección 6.4 presenta los resultados obtenidos durante la evaluación de la propuesta, y finalmente, la Sección 6.5 discute las conclusiones de este trabajo.

6.1. Continuidad de la reproducción multimedia

Las paradas de servicio pueden producirse como consecuencia del retardo que añade la recuperación unicast de los segmentos multimedia. La Figura 6.1 ilustra cómo se producen las paradas del servicio mediante un ejemplo. Los segmentos se envían periódicamente (cada T_{sgmnt}) a los receptores multicast. El tiempo esperado de llegada del segmento a la cache del terminal se ha denotado en la Figura 6.1 como t_{mcast} .

El ejemplo muestra que un reproductor obtiene directamente de la cache los segmentos 1, 2 y 3, segmentos que han podido decodificarse correctamente. La petición de dichos segmentos se produce en los instantes $t_{req}(1-3)$, es decir, coincidiendo con su disponibilidad en la cache. El Capítulo 4 explica como debe definirse el campo *availabilityStartTime* del MPD para evitar la desincronización entre la transmisión multicast y la recuperación unicast de los segmentos.

Cada vez que se recibe un segmento, el nivel del buffer se incrementa en T_{sgmnt} s. La reproducción no se inicia hasta que no se alcanza el nivel que especifica el campo *minBufferTime* del MPD. En la Figura 6.1, el valor de *minBufferTime* es tres veces la duración del segmento ($3 \cdot T_{sgmnt}$), por lo que la reproducción comienza en el instante $t_{req}(3)$.

Cuando se ha reproducido el segmento 1, comienza la reproducción del segmento 2, y al mismo tiempo, se solicita el segmento 4. La Figura 6.1 muestra que el segmento 4 no se transmite correctamente sobre eMBMS, por lo que debe ser recuperado por HTTP. Debido a que el retardo para recuperar un segmento por HTTP es, en este ejemplo, mayor a dos veces la duración del segmento multimedia ($2 \cdot T_{sgmnt}$), el segmento 4 se recibe en el instante t_{ucast} , lo que causa la parada del servicio. Es decir, el buffer del reproductor se vacía antes de que pueda

Figura 6.1: Ejemplo de parada del servicio. La figura muestra que el retardo para recuperar por HTTP los segmentos multimedia puede ser mayor que el nivel del buffer. En este ejemplo, la recuperación por HTTP del segmento número 4 provoca la parada del servicio.

recibirse el segmento 4.

6.2. MPD híbrido para distribuciones unicast y multicast

Para dar soporte a la recuperación adaptativa de errores es necesario que los metadatos que se envían a los reproductores antes del inicio de la sesión, definan la existencia de varias representaciones alternativas.

El 3GPP define un fichero de metadatos llamado User Service Bundle Description (USBD) que describe el servicio eMBMS [3GP17a]. En el caso particular de un servicio de streaming 3GP-DASH sobre eMBMS, el fichero USBD puede contener dos elementos: *mediaPresentationDescription* y *appService*. El primero, es una referencia a un fichero MPD que solo describe la representación que se envía sobre eMBMS. El segundo también hace referencia a un MPD, pero en este caso, describe todas las representaciones disponibles, la que se envía por eMBMS y las que están disponibles para su acceso por unicast.

Para identificar qué representación se envía sobre eMBMS y que representaciones se pueden acceder mediante HTTP, el campo *appService* se combina con otros dos elementos: *broadcastAppService* y *unicastAppService*. El primero almacena la URL base de la representación que se envía por eMBMS. El segundo, la URL base de la representación que se puede acceder mediante HTTP. Por tanto, si hay varias representaciones alternativas, el USBD puede incluir varios campos *unicastAppService*. De la misma forma, si la representación que se envía por eMBMS también puede descargarse por HTTP, su URL base aparecerá tanto en el elemento *broadcastAppService* como en el elemento *unicastAppService*.

Los reproductores multimedia diseñados en cumplimiento con las especificaciones de la última versión del estándar utilizan el elemento *appService* para acceder al contenido. El elemento *mediaPresentationDescription* solo se utiliza en reproductores que no cumplen con estas

especificaciones, o si el elemento *appService* no está incluido en el USBD.

Una solución alternativa a la que define el 3GPP es que esta identificación se realice directamente sobre el fichero MPD, denominado de aquí en adelante *MPD híbrido*. La primera versión del estándar de DASH no especificaba ninguna forma de añadir nuevos elementos de información en el MPD, por lo que esta solución requería proponer la extensión del estándar. Sin embargo, una nueva versión del estándar de DASH, publicada en el año 2014 [ISO14], define un nuevo campo llamado *SupplementalProperty* que puede utilizarse para dar soporte a esta funcionalidad. Este campo permite definir un atributo de carácter obligatorio que especifica el esquema (*@schemeIdUri*) y un atributo opcional que especifica su valor (*@value*). Así, es posible definir un esquema llamado, por ejemplo, *accessTech* que permita indicar cual es la representación que se envía por eMBMS. La Figura 6.2 muestra cómo se definiría el MPD según esta propuesta. Como se puede observar, la representación de 500 Kbps tiene la etiqueta “multicast” como valor del esquema *accessTech*. Esta solución alternativa es más compacta que la solución definida por el estándar ya que proporciona la misma funcionalidad mediante el uso de una única etiqueta. También es fácil de implementar, puesto que solo se requiere modificar el reproductor DASH para que sea capaz de procesar un nuevo campo del MPD.

El USBD se definió originalmente para responder a necesidades propias de escenarios de handover. Los procesos de handover garantizan la continuidad del servicio cuando un terminal móvil conmuta entre diferentes redes de acceso (handover vertical) o diferentes estaciones base (handover horizontal). La Figura 6.3 muestra escenarios de handover que requieren la conmutación entre los servicios de streaming unicast y multicast. Esta conmutación es posible gracias al uso del fichero USBD. Por ejemplo, cuando un terminal abandona el área eMBMS, pierde la señal de la transmisión broadcast y el reproductor solo puede acceder al contenido desde la red móvil utilizando HTTP. El USBD también permite que esta conmutación se produzca cuando el terminal se conecta a un punto de acceso Wi-Fi o a una femtocelda. Tras conectarse al nuevo punto de acceso, el reproductor deja de recibir el contenido por eMBMS y comienza a recibirlo

```
<Period id="1"> <!-- Video -->
<AdaptationSet mimeType="video/mp4" codecs="avc1.4D401F" frameRate="30000/1001"
segmentAlignment="true" startWithSAP="1">
<BaseURL>video/</BaseURL>
<SegmentTemplate timescale="90000" initialization="$Bandwidth%/init.mp4v"
media="$Bandwidth%/$Time$.mp4v">
<SegmentTimeline> <S t="0" d="180180" r="432"/> </SegmentTimeline>
</SegmentTemplate>

<Representation id="1" width="320" height="240" bandwidth="250000">
</Representation>

<Representation id="2" width="640" height="480" bandwidth="500000">
<SupplementalProperty schemeIdUri="urn:dash:scheme:accessTech"
value="multicast"/>
</Representation>

<Representation id="3" width="960" height="720" bandwidth="1000000">
</Representation>
```

Figura 6.2: MPD para la identificación de la representación eMBMS mediante el nuevo esquema *accessTech*.

Figura 6.3: Escenarios de handover. La figura muestra diferentes escenarios donde se produce un handover y donde el uso del fichero USBD permitiría conmutar entre servicios de streaming unicast y multicast.

por unicast vía HTTP. La Figura 6.3 indica que estos procesos de handover pueden ocurrir en doble dirección. Es decir, el terminal también puede abandonar la conexión unicast para recibir el contenido por eMBMS.

Un escenario especial de handover es el que se representa en la Figura 6.4. En este caso, el terminal se mueve entre dos áreas eMBMS: el área eMBMS 1, que se corresponde con el despliegue del servicio en un entorno suburbano, y el área eMBMS 2, que se corresponde con el despliegue del servicio en un entorno urbano. En un entorno suburbano la tasa máxima de datos del servicio es significativamente menor que en un entorno urbano para el mismo nivel de cobertura (Capítulo 3). Si en ambas áreas eMBMS se envía la misma representación multimedia, el handover entre áreas puede introducir el problema de la discontinuidad del servicio. Para que esto no ocurra se pueden adoptar dos estrategias:

- enviar la misma representación eMBMS en todas las áreas. El contenido multimedia se codifica ajustando el bitrate al área de menor ancho de banda.
- transmitir en cada área una representación eMBMS distinta. El bitrate de la representación dependerá, en cada caso, del ancho de banda que se haya reservado al servicio eMBMS. En la Figura 6.4, el bitrate de la representación que se envía en el área 1 es mayor que el bitrate de la representación que se envía en el área 2.

La segunda estrategia mejora la calidad percibida por los usuarios porque se maximiza el bitrate en cada área. Sin embargo, requiere notificar al reproductor que, tras el handover, la representación eMBMS ha cambiado. De lo contrario se produce un error de sincronización. El terminal se mueve a otra área pero el reproductor sigue solicitando los segmentos que se describen en el fichero MPD que tiene. Como los segmentos que llegan a su cache pertenecen a otra representación tras el handover, se establece la recuperación por HTTP de los segmentos que ha solicitado. La consecuencia de esto es que, para ese cliente, el servicio eMBMS no funciona correctamente, y se produce un uso indebido de los recursos radio.

Figura 6.4: Handover entre diferentes áreas eMBMS. La figura muestra un proceso de handover horizontal entre dos áreas eMBMS: el área 1 representa el despliegue del servicio en un entorno suburbano. El área 2, en un entorno urbano.

Para evitar este problema se propone utilizar el campo *minimumUpdatePeriod* del MPD. El estándar de DASH define que puede utilizarse para que los reproductores comprueben si el MPD ha sufrido cambios. El valor de *minimumUpdatePeriod* establece con qué frecuencia debe descargar el MPD un reproductor para detectar posibles cambios. Su uso es especialmente útil en servicios de streaming multimedia en directo, por ejemplo, para añadir spots publicitarios.

Si un usuario se mueve dentro del área eMBMS, su reproductor no detecta ningún cambio en el MPD que pueda relacionarse con un proceso de handover. Sin embargo, cuando se mueve a otra área, recibe una versión diferente del MPD, donde se indica que la representación que ahora se envía por eMBMS tiene otro bitrate. Esto se indica, por ejemplo, mediante el uso de la etiqueta “multicast” del esquema *@accessTech*. En la nueva versión del MPD, la etiqueta se especifica en otra representación.

Una desventaja de esta propuesta es que los usuarios que no cambian de área eMBMS pueden descargarse innecesariamente múltiples versiones del MPD. Esto es especialmente significativo si el valor de *minimumUpdatePeriod* es bajo. El uso de la cabecera ETag de HTTP solucionaría este problema. La cabecera ETag se utiliza para informar a los clientes sobre cambios de los recursos web. Cuando un reproductor solicita por primera vez el MPD, el servidor añade la cabecera ETag en la respuesta con un determinado valor. El cliente almacena el valor de la cabecera junto con el MPD. Si este fichero no ha cambiado en el servidor, cuando se solicita de nuevo, el cliente recibe un mensaje “HTTP 304 Not Modified”. Si el usuario se mueve a otra área eMBMS verá que la cabecera ETag de la respuesta tiene otro valor, y se descargará la nueva versión del MPD.

6.2.1. Análisis del impacto de los procesos de handover en la continuidad de los servicios de streaming multimedia

A continuación, se analiza el impacto de los procesos de handover en la continuidad de los servicios de streaming multimedia. El análisis se centra en procesos de handover vertical entre eMBMS y una tecnología de acceso alternativa donde se realiza streaming adaptativo sobre

Figura 6.5: Representación de video seleccionada antes y después de la conmutación entre el servicio de streaming 3GP-DASH sobre eMBMS y un servicio de streaming adaptativo sobre HTTP. El handover se produce en el segundo 50 de la reproducción. Los resultados se muestran para diferentes retardos de handover (HO Lat.).

HTTP. Para ello, se ha modificado la plataforma que se describe en el Capítulo 3. En particular, se ha modificado el reproductor *dash.js* para que sea capaz de procesar el nuevo esquema *accessTech*. La conmutación entre el servicio 3GP-DASH sobre eMBMS y el servicio de streaming adaptativo sobre HTTP se realiza sobre esta plataforma de forma manual, simplemente pulsando un botón. La Figura 6.5 muestra la representación de vídeo que se selecciona antes y después de la conmutación entre los servicios. El handover se inicia en el segundo 50 de la reproducción y finaliza tras un periodo de tiempo conocido como handover latency o retardo de handover (HO Lat. en la Figura 6.5). La Figura 6.5 muestra que el reproductor solicita la representación etiquetada en el MPD como “multicast” siempre y cuando el terminal esté conectado al servicio de streaming 3GP-DASH sobre eMBMS. Esta representación, como muestra la Figura 6.2, es la que tiene asociado el índice 2 y está codificada a 500 Kbps. Conmutar a otra tecnología de acceso permite seleccionar otras representaciones disponibles. La Figura 6.5 muestra que tras la finalización del handover, el reproductor selecciona representaciones de mayor calidad (índices 3 y 4). Esto sucede porque el ancho de banda disponible en la nueva red acceso es mayor.

La Figura 6.6 muestra el impacto del handover vertical en el buffer del reproductor. El nivel del buffer oscila entre 8 y 10 segundos cuando el reproductor está conectado al servicio de streaming 3GP-DASH sobre eMBMS. Durante el proceso de handover el nivel del buffer se reduce en proporción al retardo de handover. Tras finalizar este procedimiento, vuelve al estado original, conmutando de nuevo entre 8 y 10 segundos. Esto ocurre, porque en este caso, el ancho de banda disponible en la nueva red de acceso es mayor que el que se ha reservado para el servicio eMBMS. Dado que también podría ser menor (su valor no se conoce de antemano),

Figura 6.6: Impacto de un handover vertical en el buffer del reproductor. El handover se produce en el segundo 50 de la reproducción. Los resultados se muestran para diferentes retardos de handover (HO Lat.).

el algoritmo debería tener en cuenta, que tras un handover, se debe seleccionar la representación de menor calidad. Así, el reproductor puede estimar el ancho de banda disponible y seleccionar la representación que más se ajusta a la estimación.

Aunque el USBD, o la solución alternativa que se ha propuesto, puede utilizarse para dar soporte a los procesos de handover, también puede utilizarse para dar soporte a la recuperación adaptativa de los segmentos por HTTP. Es decir, su uso en los servicios de streaming 3GP-DASH sobre eMBMS permite seleccionar la representación que mejor se ajusta al ancho de banda unicast durante el procedimiento de recuperación de los segmentos.

6.3. Mecanismo de recuperación de errores adaptativo

El uso del MPD híbrido permite diferenciar la representación que se envía sobre eMBMS de otras representaciones alternativas que se pueden acceder por HTTP. Cuando un reproductor utiliza HTTP para recuperar un segmento, debe seleccionar la representación que más se ajusta al ancho de banda del canal unicast. El problema es que el valor del ancho de banda unicast no puede determinarse, puesto que depende de varios parámetros de red que varían dinámicamente.

En general, los algoritmos de adaptación que se han propuesto para servicios de streaming adaptativo sobre HTTP [SES⁺15] se basan en técnicas pasivas para la estimación del ancho de banda (ver Capítulo 2). Sin embargo, no parece adecuado aplicar estas técnicas a los servicios de streaming 3GP-DASH sobre eMBMS, puesto que la descarga de segmentos por HTTP se produce eventualmente, y por tanto, no se tienen suficientes datos para realizar estimaciones precisas del ancho de banda del canal. Por ejemplo, supongamos que se recupera un segmento por HTTP y que se estima el ancho de banda utilizando el tiempo de descarga de dicho segmen-

to. Posteriormente, se transmiten correctamente varios segmentos multimedia sobre eMBMS. Durante ese tiempo, no hay nuevos datos para estimar el ancho de banda unicast, por lo que cuando se recupera otro segmento por HTTP, la estimación previa puede no ajustarse a las condiciones de ancho de banda actuales. Esto es muy probable en las redes inalámbricas, donde la variación del ancho de banda es mayor. Para suplir la ausencia de datos se podrían utilizar medidas de otras aplicaciones que se estén ejecutando en el terminal, pero esto aumentaría la complejidad de la solución, puesto que se requiere la interacción entre el reproductor y dichas aplicaciones.

Otros trabajos proponen utilizar técnicas activas para la estimación del ancho de banda [YCH14, LBK15]. La desventaja de estas técnicas es que se basan en inyectar tráfico a la red que no tiene un propósito específico salvo el de la propia estimación. Es decir, requieren consumir ancho de banda para estimar el ancho de banda, lo que no es de por sí muy eficiente.

De la revisión del estado del arte se deduce que es difícil estimar con precisión el ancho de banda desde la perspectiva del cliente de streaming. Los servicios de streaming adaptativo basados en SAND (Capítulo 2) no utilizan estimadores, sino que adaptan la transmisión del streaming multimedia mediante la actuación de elementos intermedios de la red que disponen de información del ancho de banda disponible. Por ejemplo, El Essaili *et al.* [ESS⁺15] proponen adaptar el streaming multimedia a partir de información que proporcionan los algoritmos de planificación de recursos que ejecutan las estaciones base. En ese caso, la selección de la representación no se basa en una estimación, sino en información que permite determinar el ancho de banda que se asigna a cada canal unicast.

Aunque teóricamente este tipo de soluciones mejora las prestaciones de los servicios de streaming adaptativo, en la práctica son difíciles de implementar. La rigidez de las arquitecturas de red móvil clásicas dificulta la modificación de los elementos funcionales de la red para desplegar ágilmente los mecanismos de soporte a los servicios de streaming adaptativo basados en SAND (Capítulo 7).

6.3.1. Adaptación asistida por el BM-SC

En un servicio de streaming 3GP-DASH sobre eMBMS la recuperación de los segmentos perdidos se realiza a través de una función de tipo *Associated Delivery* del BM-SC (Capítulo 2). En esta sección se propone modificar dicha función para que el reproductor pueda recibir una representación alternativa a la que se envía por eMBMS, más adecuada al ancho de banda disponible en el canal unicast. Es decir, lo que se propone es que el BM-SC ejecute un algoritmo de adaptación que seleccione la representación multimedia a partir de medidas reales de ancho de banda obtenidas a través de la información de la planificación de los recursos radio y el valor de otros parámetros físicos como el MCS.

En las arquitecturas de red móvil actuales, la estación base integra software y hardware propietario en un mismo equipo. Modificar el software que ejecuta la estación base para que envíe información del ancho de banda al BM-SC parece costoso, ya que requiere que el fabrican-

Figura 6.7: Mecanismo de recuperación de errores adaptativo, basado en la modificación de una función de tipo *Associated Delivery* del BM-SC. La figura muestra como función selecciona una representación diferente a la enviada por eMBMS, más adecuada al ancho de banda unicast, para la recuperación de los segmentos perdidos.

te desarrolle nuevas versiones de estos equipos incorporando esta funcionalidad. Modificar la función *Associated Delivery* del BM-SC para que ejecute el algoritmo de adaptación plantea el mismo problema. Además, conviene destacar que la complejidad aumenta cuando el operador ha adquirido equipos de diferentes fabricantes.

En definitiva, las arquitecturas de red móviles clásicas no permiten desplegar ágilmente nuevos servicios, y en concreto, mecanismos como el que se propone. El Capítulo 7 aborda este problema proponiendo que los servicios eMBMS evolucionen combinando soluciones basadas en NFV y SDN.

La Figura 6.7 muestra el procedimiento que se debe ejecutar para que se establezca, según la propuesta, la recuperación unicast de los segmentos que se pierden en el canal eMBMS. Aunque la recuperación unicast también puede ocurrir por errores de sincronización, el Capítulo 4 describe como es posible evitar que esto suceda.

La Figura 6.7 muestra que el proveedor de contenidos utiliza el punto de referencia xMB-U (Figura 2.4) para enviar los segmentos multimedia al BM-SC. El estándar especifica dos métodos diferentes para realizar este envío: *Push* o *Pull* (ver Capítulo 2). Sin embargo, no contempla que los segmentos perdidos por eMBMS puedan retransmitirse por HTTP con otro bitrate, y es por ello que se propone modificar la configuración del servicio para que se pueda indicar esta información. Así, sería posible implementar las siguientes estrategias de envío:

- Utilizar el modo *Push* para enviar todas las representaciones multimedia que ha generado el proveedor de contenidos (la que se envía sobre eMBMS y las que están disponibles para la recuperación adaptativa de errores). La ventaja de esta estrategia es que es posible reducir el retardo evitando la solicitud bajo demanda del contenido. Sin embargo, también hay que considerar que la mayor parte de los segmentos se entrega correctamente sobre eMBMS, por lo que es posible que el proveedor envíe, innecesariamente, segmentos que pertenecen a representaciones que no van a ser seleccionadas durante la recuperación unicast de errores.

- Utilizar el modo *Push* para enviar la representación eMBMS (que siempre se transporta sobre la red) y utilizar el modo *Pull* cuando se establece la recuperación adaptativa de errores. Aunque esto evita enviar y almacenar contenido multimedia en el BM-SC de forma innecesaria, tiene la desventaja de que puede aumentar el retardo durante la recuperación unicast de errores.
- Utilizar el modo *Pull* para enviar todas las representaciones multimedia al BM-SC bajo demanda.

La selección de una u otra estrategia se debe especificar durante la definición del servicio. El proveedor, utilizando el canal de control xMB-C indicaría, entre otros parámetros de la sesión, el método de envío o *Ingest Mode* de cada una de las representaciones que se definen en el fichero MPD. La Figura 6.7 muestra, como ejemplo, que el proveedor de contenidos envía los segmentos de la representación eMBMS al BM-SC utilizando el modo *Push* (1), y los segmentos de otras representaciones alternativas utilizando el modo *Pull* (4-5).

La recuperación de los segmentos por HTTP se realiza a través de una función de tipo *Associated Delivery* del BM-SC, que ejecuta un algoritmo de adaptación que selecciona la representación más adecuada al ancho de banda unicast. El reproductor siempre solicita los segmentos que pertenecen a la representación eMBMS (3), pero si el ancho de banda unicast es menor que el bitrate de la representación eMBMS, la función *Associated Delivery* selecciona una representación de menor calidad. Si, por el contrario, el ancho de banda unicast es mayor o igual que el bitrate de la representación eMBMS, el algoritmo de adaptación no modifica la URL solicitada por el reproductor. Es decir, se entrega por unicast la misma representación que se envía por eMBMS.

Para que la recuperación adaptativa de errores sea posible, las estaciones base deben mandar información del ancho de banda al BM-SC. Como se explicará en el Capítulo 7, las arquitecturas de red móvil basadas en virtualización proporcionan la flexibilidad necesaria para modificar las unidades Base Band Unit (BBU) con el objetivo de que puedan reportar esta información. Una vez que el algoritmo de adaptación determina la representación más adecuada, el BM-SC debe entregar el segmento perdido al cliente (6). Como ya se ha comentado, la Figura 6.7 muestra que el BM-SC solicita el segmento al proveedor utilizando el modo *Pull* (4-5).

El mecanismo que se propone sigue el enfoque que se presenta en la Figura 2.2b, donde se observa que el cliente no ejecuta la lógica de adaptación, sino que es un elemento de red intermedio (DANE) el que se encarga de seleccionar la representación multimedia en su nombre. El objetivo del algoritmo de adaptación es evitar paradas de servicio durante la retransmisión unicast de los segmentos y reducir el retardo inicial de buffering. A continuación, se describen dos posibles diseños del algoritmo de adaptación: buffer-aware y buffer-unaware.

6.3.1.1. Algoritmo *buffer-unaware*

El algoritmo de adaptación *buffer-unaware* selecciona la representación multimedia que se enviará durante la recuperación de los segmentos perdidos a partir del ancho de banda unicast. El algoritmo es capaz de evitar las paradas del servicio si el ancho de banda disponible es menor que el bitrate de la representación eMBMS y mayor que el bitrate de la representación alternativa de menor calidad.

La Figura 6.8a muestra que el BM-SC selecciona la representación multimedia en función del número y bitrate de las representaciones que se definen en el MPD híbrido, y del ancho disponible en el canal unicast. Durante la definición del servicio, el proveedor de contenidos envía al BM-SC el fichero MPD. Además, debe configurar la sesión especificando el valor del parámetro *Application Entry Point URL*, que en el caso particular de un servicio de streaming 3GP-DASH sobre eMBMS, apunta, precisamente, al fichero MPD.

Como se ha comentado anteriormente, el algoritmo de adaptación puede utilizar información del ancho de banda disponible gracias la comunicación que existe entre el BM-SC y las unidades BBU. En definitiva, la representación seleccionada por el algoritmo *buffer-unaware* ($B_{unaware}$) se puede calcular como

$$\begin{aligned} B_{unaware} &= \operatorname{argmax}_k B_k \\ &\text{subject to } B_k < BW_{unicast} \end{aligned} \quad (6.1)$$

donde B_k es el bitrate de la k-ésima representación y $BW_{unicast}$ es el ancho de banda disponible en el canal unicast. El objetivo de esta selección es garantizar que el retardo para recuperar un segmento por HTTP sea menor que la duración de dicho segmento, de forma que se complete su descarga antes de que finalice la reproducción del segmento actual. Esto será posible siempre que el MPD incluya una representación cuyo bitrate sea menor que el ancho de banda disponible en el canal unicast.

6.3.1.2. Algoritmo *buffer-aware*

La Figura 6.8b muestra otro posible diseño del algoritmo de adaptación. Este diseño tiene en cuenta que el buffer del reproductor puede almacenar varios segmentos antes de que la sesión de streaming de comienzo. La idea es que el reproductor envía, junto con la petición del segmento, su nivel del buffer en segundos (BL). El algoritmo de adaptación utiliza el estado del buffer para intentar solicitar la representación enviada por eMBMS siempre que sea posible. Incluso si el ancho de banda del canal unicast es menor que el bitrate de la representación eMBMS, es posible evitar la conmutación a una calidad inferior si la descarga se realiza antes de que el buffer del reproductor se vacíe.

En definitiva, el algoritmo que se propone selecciona la representación enviada por eMBMS (B_{eMBMS}) si el retardo que añade la recuperación por HTTP de los segmentos (d_{ucast}) es menor

Figura 6.8: Algoritmos de adaptación ejecutados en la función *Associated Delivery* del BM-SC: *buffer-unaware* y *buffer-aware*

que el nivel del buffer del reproductor. Si esto no ocurre, el algoritmo *buffer-aware* no tiene en cuenta el estado del buffer y se ejecuta el algoritmo de adaptación *buffer-unaware*. Para calcular $d_{unicast}$ se puede asumir que el componente de retardo más significativo es el asociado a la transmisión unicast del segmento multimedia. Es decir, $d_{unicast}$ puede calcularse como:

$$d_{unicast} = \frac{B_{eMBMS} \cdot T_{sgmnt}}{BW_{unicast}} \quad (6.2)$$

Así, la representación seleccionada por el algoritmo de adaptación *buffer-aware* (B_{aware}) puede calcularse como:

$$B_{aware} = \begin{cases} B_{eMBMS}, & d_{unicast} \leq BL \\ B_{unaware}, & d_{unicast} > BL \end{cases} \quad (6.3)$$

6.4. Prestaciones del mecanismo de recuperación de errores adaptativo

Para evaluar los algoritmos de adaptación se ha utilizado la plataforma descrita en el Capítulo 3, que combina resultados de simulación con un modelo analítico. El modelo analítico se ha extendido implementando los algoritmos de adaptación *buffer-unaware* y *buffer-aware*. Utilizando esta plataforma, se pueden extraer medidas de varios parámetros objetivos de QoE. En este caso se ha medido la eficiencia, la inestabilidad y las paradas del servicio (Capítulo 2). Hasta ahora no era necesario medir la inestabilidad, pero la recuperación adaptativa de errores introduce este efecto en la reproducción del contenido.

Figura 6.9: Relación entre el AL-FEC code rate y el porcentaje de segmentos de 2 s recuperados por HTTP.

6.4.1. Relación entre la tasa de codificación AL-FEC y las retransmisiones HTTP

En el Capítulo 4 se ha analizado la relación entre la tasa de codificación AL-FEC, el porcentaje de segmentos recuperados por HTTP y la duración del segmento multimedia. En este caso, los resultados se han obtenido solo para segmentos de dos segundos (Figura 6.9). Este análisis es suficiente para ilustrar que el mecanismo propuesto evita las paradas del servicio sin aumentar el retardo inicial de buffering.

La Figura 6.9 muestra que es posible utilizar tasas de codificación AL-FEC más bajas para proporcionar una mayor capacidad de corrección frente a errores. Es decir, para tasas de codificación más bajas, el porcentaje de segmentos que se recupera por HTTP es menor.

Sin embargo, la tasa de datos del servicio depende de la tasa de codificación AL-FEC y del ancho de banda del canal eMBMS. A su vez, el ancho de banda del canal eMBMS depende de la cantidad de recursos radio asignados al servicio eMBMS y del índice del MCS. La Tabla 6.1

Parámetro	Valor
Duración de la reproducción	300 s
Número de terminales	399
Duración del segmento (T_{sgmnt})	2 s
Inter-Site Distance (ISD)	500 m
Índice del esquema de modulación y codificación (MCS)	15
Tamaño del área MBSFN	7 eNBs
Tasa de codificación AL-FEC	[0.72-0.92]
$BW_{unicast}$	[0.3-1] Mbps
Video bitrate	0.25 / 0.5 / 1 Mbps
$minBufferTime$	2 / 3 / 4 segmentos

Tabla 6.1: Parámetros de servicio para evaluar el mecanismo de recuperación de errores adaptativo

Figura 6.10: Tiempo de parada de servicio para diferentes valores del ancho de banda unicast ($BW_{unicast}$) y del campo *minBufferTime* (mBT). Los resultados se obtienen para un AL-FEC code rate de 0.78.

muestra que para la evaluación se ha considerado el índice 15 del MCS y el uso de una única subtrama para la transmisión eMBMS. Con estos valores, el ancho de banda que se obtiene es de 1.4 Mbps.

También se ha asumido que el contenido multimedia se codifica a 1 Mbps, bitrate que proporciona un buen compromiso entre el ancho de banda asignado al canal de eMBMS y la calidad del vídeo [LG12]. Para garantizar este bitrate, y según los resultados que se muestran en la Figura 6.9, se debe utilizar una tasa de codificación AL-FEC mayor o igual a 0.78.

6.4.2. Análisis de la continuidad de la reproducción multimedia

El mecanismo de recuperación de errores adaptativo permite reducir el retardo inicial de buffering y evitar las paradas de servicio (subsección 6.4.3). Sin embargo, en un servicio de streaming 3GP-DASH sobre eMBMS estándar, la recuperación no es adaptativa.

Utilizando la plataforma de experimentación se ha simulado el envío de un vídeo de 300 segundos codificado con un bitrate de 1 Mbps. Esta representación se envía sobre el canal eMBMS a los múltiples receptores. Si algún segmento se pierde durante su transmisión se recupera por HTTP en la misma calidad.

La Figura 6.10 muestra el valor promedio del tiempo de parada del servicio (expresado como porcentaje) para diferentes valores del ancho de banda unicast ($BW_{unicast}$) y diferentes valores del nivel inicial del buffer (*minBufferTime*). Los resultados muestran que para un *minBufferTime* de dos segmentos, el tiempo promedio de parada es, aproximadamente, un 3 % del tiempo total de reproducción. Es posible reducir este valor incrementando el *minBufferTime*, pero el aumento del retardo inicial de buffering degrada la calidad percibida (Capítulo 5).

La figura también muestra que a medida que el ancho de banda unicast aumenta, y se ajusta en mayor medida al bitrate de la representación eMBMS, el tiempo de parada es menor.

Figura 6.11: Representación de la inestabilidad para los dos algoritmos de adaptación propuestos: *buffer-unaware* (BUP) y *buffer-aware* (BAP). La inestabilidad se calcula para diferentes valores del ancho de banda unicast ($BW_{unicast}$) y del minBufferTime (mBT). Los resultados se obtienen para un AL-FEC code rate de 0.78.

6.4.3. Rendimiento de los algoritmos de adaptación

El mecanismo de recuperación de errores adaptativo permite, en el contexto de los servicios de streaming 3GP-DASH sobre eMBMS, evitar las paradas de servicio gracias a que los reproductores conmutan entre diferentes representaciones multimedia. El mecanismo, sin embargo, introduce un efecto de inestabilidad en la reproducción que también puede degradar la calidad percibida. A continuación, se presenta un análisis del rendimiento de los algoritmos de adaptación *buffer-unaware* y *buffer-aware* utilizando las siguientes métricas objetivas de QoE: inestabilidad y calidad de vídeo promedio.

Para el análisis se ha considerado que, además de la representación enviada por eMBMS (1 Mbps), el MPD híbrido define dos representaciones alternativas (250 Kbps y 500 Kbps) que pueden solicitarse durante la recuperación de los segmentos por HTTP.

La Figura 6.11 muestra la inestabilidad que introducen los algoritmos *buffer-aware* (BAP) y *buffer-unaware* (BUP) para diferentes valores del *minBufferTime* y del ancho de banda unicast ($BW_{unicast}$). El primer valor del ancho de banda que se analiza es 300 Kbps, ligeramente superior al bitrate de la representación de menor calidad. El ancho de banda máximo es 1 Mbps, que coincide con el bitrate de la representación eMBMS. Por encima de 1 Mbps no tiene sentido evaluar la propuesta, puesto que la recuperación unicast de los segmentos no produce retardos adicionales que puedan afectar a la continuidad de la reproducción multimedia.

Los resultados muestran que la inestabilidad es constante cuando se utiliza el algoritmo *buffer-unaware*. La razón es que, cada vez que se pierde un segmento, el algoritmo selecciona una representación diferente a la que se envía sobre eMBMS.

Utilizar el algoritmo *buffer-aware* permite reducir la inestabilidad, especialmente para valores del *minBufferTime* de tres y cuatro segmentos. Además, los resultados muestran que la inestabilidad se reduce progresivamente hasta el 0 % a medida que aumenta el ancho de banda unicast.

La Figura 6.11 también muestra que la inestabilidad que introduce el algoritmo *buffer-aware* puede ser mayor que la que introduce el algoritmo *buffer-unaware* si el *minBufferTime* es pequeño (dos segmentos). Esto se debe a que, en este caso, solicitar la representación eMBMS hace que el buffer se agote rápidamente, por lo que se debe conmutar a otra calidad cuando se recupera el siguiente segmento.

Como se puede observar, la inestabilidad depende del número y bitrate de las representaciones que se definen en el MPD. Esto explica la reducción de la inestabilidad cuando el ancho de banda unicast es igual a 500 Kbps. Debido a que el MPD no define una representación intermedia entre la codificada a 500 Kbps y la codificada a 1 Mbps, la inestabilidad puede incrementarse cuando el ancho de banda es ligeramente superior a 500 Kbps. Sin embargo, cuando el ancho de banda adquiere valores cercanos a 1 Mbps, el buffer es lo suficientemente grande ($mBT > 2$ sgmnt) como para compensar el retardo que añade el envío de la representación eMBMS, y esto, permite reducir el número de conmutaciones.

La Figura 6.12 muestra el valor de la calidad de vídeo promedio para diferentes valores del *minBufferTime* y del ancho de banda unicast. La figura muestra que, en el peor caso representado ($BW_{unicast} = 300$ Kbps y *minBufferTime* = 1 segmento), la calidad de vídeo promedio es aproximadamente 0.98. Por lo tanto, no se aprecia una degradación significativa de la calidad con respecto al máximo que podría alcanzarse si todos los segmentos se entregaran correctamente sobre eMBMS (calidad de vídeo promedio igual a 1). Esto es consistente con los resultados que se muestran en la Figura 6.9, que indican que solo un 2 % de segmentos se recupera por HTTP cuando se utiliza una tasa de codificación AL-FEC de 0.78. Es decir, la mayor parte de los segmentos se entrega correctamente sobre eMBMS.

La Figura 6.12 muestra que utilizar un *minBufferTime* de cuatro segmentos permite incrementar la calidad de vídeo promedio hasta un valor muy cercano 1. Sin embargo, esta mejora no justifica incrementar la latencia del servicio ocho segundos, y más, cuando se ha concluido que el retardo inicial de buffering afecta significativamente al servicio (Capítulo 5). De los

Figura 6.12: Representación de la calidad de vídeo promedio para los algoritmos de adaptación propuestos: *buffer-unaware* (BUP) y *buffer-aware* (BAP). La calidad de vídeo promedio se calcula para diferentes valores del ancho de banda unicast ($BW_{unicast}$) y del *minBufferTime* (mBT). Los resultados se obtienen para un AL-FEC code rate de 0.78.

resultados también se concluye que la calidad de vídeo promedio que se obtiene cuando se utiliza el algoritmo *buffer-aware* no es mucho mejor que la que se obtiene cuando se utiliza el algoritmo *buffer-unaware*.

Para concluir, ambos algoritmos de adaptación cumplen con el objetivo de evitar las paradas del servicio. El algoritmo *buffer-aware* puede ser útil para minimizar la inestabilidad y maximizar la calidad de vídeo promedio. Sin embargo, y a tenor de los resultados, no parece razonable aumentar el retardo inicial de buffering porque la mejora de inestabilidad y calidad de vídeo promedio que se obtiene no es muy significativa. En el escenario que se ha considerado para realizar el análisis la mayoría de los segmentos se entregan correctamente sobre eMBMS. Además, la complejidad de esta solución es mayor, ya que requiere que los reproductores envíen su nivel de buffer cada vez que solicitan un segmento por HTTP. Esto implica realizar modificaciones en los reproductores multimedia. El algoritmo *buffer-unaware* tiene la ventaja de minimizar el retardo inicial de buffering sin requerir cambios en los reproductores. Además, su uso no afecta significativamente a la inestabilidad o a la calidad de vídeo promedio.

6.5. Conclusiones

Sobredimensionar el buffer del reproductor para evitar paradas del servicio incrementa la latencia y degrada la calidad percibida. Este capítulo propone una alternativa, que consiste en dar soporte a la recuperación adaptativa de errores, de forma que un reproductor pueda conmutar entre diferentes representaciones multimedia cuando se recuperan los segmentos por HTTP.

Basar la adaptación en la utilización de técnicas activas o pasivas para la estimación del ancho de banda no es adecuado en el marco de los servicios de streaming 3GP-DASH sobre eMBMS. Por esta razón, se ha propuesto una solución que sigue las nuevas tendencias de streaming adaptativo basado en SAND. Esta solución requiere modificar el BM-SC para que reciba de las unidades BBU información del ancho de banda asignado a los terminales. También requiere modificar el BM-SC para que implemente el algoritmo de adaptación.

Esto plantea un reto desde la perspectiva de las arquitecturas de red móvil clásicas, que no proporcionan mecanismos flexibles para desplegar ágilmente nuevas soluciones. El Capítulo 7 explica como la combinación de la virtualización de las funciones de red y las redes definidas por software flexibiliza el despliegue del mecanismo de recuperación de errores que se propone en este capítulo.

Con respecto a los algoritmos de adaptación propuestos, los resultados obtenidos muestran que el algoritmo *buffer-aware* puede utilizarse para mejorar la inestabilidad y la calidad de vídeo promedio. Sin embargo, el algoritmo *buffer-unaware* tiene la ventaja de que es más fácil de implementar y, además, proporciona retardos menores. Como se explica en el Capítulo 5, los servicios de difusión de contenidos son muy sensibles a retardos, por lo que puede ser recomendable seleccionar esta opción.

El contenido de este capítulo se ha publicado en un artículo situado en el Q4 del JCR [LBP17a]. Además, se han realizado dos ponencias en congresos de ámbito internacional [LBP15, BLAP15].

Capítulo 7

Arquitecturas de red móvil basadas en NFV y SDN para la difusión de contenidos multimedia

Desplegar mecanismos para dar soporte a los servicios de streaming multimedia basados en SAND es costoso desde la perspectiva de las arquitecturas de red móvil clásicas. El mecanismo de recuperación de errores que se ha propuesto en el Capítulo 6 requiere modificar algunas de las entidades funcionales definidas por el 3GPP, como el BM-SC o las estaciones base. Sin embargo, las arquitecturas de red móvil actuales son rígidas y no proporcionan la capacidad de desplegar ágilmente nuevos servicios. La aparición de nuevos paradigmas de red como NFV y SDN permite definir arquitecturas de red móvil más flexibles, y acelerar el proceso de innovación a través del despliegue ágil de mecanismos como el que propone el Capítulo 6 para reducir la latencia en los servicios de streaming 3GP-DASH sobre eMBMS.

La definición de las redes 5G se basa en la aplicación de dos transformaciones. La primera consiste en separar los planos de datos y de control de las entidades funcionales de la red móvil (i.e., aplicar CUPS). La segunda consiste en incorporar soluciones que combinen las tecnologías NFV y SDN en la arquitectura de red. La aplicación de ambas transformaciones permite controlar y gestionar las redes móviles de una forma más flexible, más eficiente y a un coste menor [VIP15].

Actualmente, los esfuerzos de estandarización del 3GPP se centran en analizar como estas transformaciones pueden dar soporte a los servicios unicast, i.e., a servicios basados en transmisiones punto-a-punto. El 3GPP también se plantea abordar en el futuro la evolución de la arquitectura de eMBMS [5GP], llamada de aquí en adelante virtual MBMS (vMBMS). Sin embargo, no existe a día de hoy ningún estándar que especifique como se pueden combinar CUPS y NFV/SDN en este contexto. Por otro lado, el estado del arte (Capítulo 2) muestra que existen trabajos de investigación que proponen diseñar la evolución de la arquitectura de eMBMS considerando estas tecnologías. Sin embargo, estas propuestas consideran NFV y SDN de forma independiente y, en algunos casos, no garantizan la interoperabilidad de las arquitecturas de red móvil basadas en virtualización con las arquitecturas de red móvil clásicas.

En este capítulo se discute el diseño de una nueva arquitectura de soporte a los servicios eMBMS considerando la combinación de CUPS y NFV/SDN. El diseño sigue el mismo enfoque de otros proyectos como M-CORD [MCO], que parten de la arquitectura de red móvil

LTE para especificar cuales son las transformaciones que deben sufrir las redes móviles para proporcionar una mayor programabilidad, flexibilidad, y eficiencia. Además, este enfoque de diseño garantiza las necesidades de interoperabilidad que no se consideran en el estado del arte.

Tras analizar diferentes estrategias de diseño, se aborda el despliegue ágil de mecanismos que permitan dar soporte a los servicios de streaming multimedia basados en SAND. En particular, se explica como se puede desplegar en vMBMS el mecanismo de recuperación de errores adaptativo que se ha propuesto para reducir la latencia de los servicios de streaming 3GP-DASH sobre eMBMS.

El resto del capítulo se organiza de la siguiente manera: la Sección 7.1 discute cuatro enfoques diferentes para definir la evolución de la arquitectura de eMBMS desde la perspectiva de CUPS, NFV y SDN. La Sección 7.2 analiza cómo se puede desplegar sobre vMBMS el mecanismo de recuperación de errores adaptativo que se propone en el Capítulo 6. Finalmente, la Sección 7.3 proporciona las conclusiones de este trabajo de investigación.

7.1. Diseño de arquitecturas de red móvil basadas en NFV y SDN para la difusión de contenidos

El despliegue y gestión del ciclo de vida de las funciones de red ha sufrido muchos cambios con la aparición de la virtualización de las funciones de red y las redes definidas por software. Según estos paradigmas, las funciones de red que pertenecen al plano de datos (por ejemplo, las que realizan el P-GW o el S-GW) se pueden implementar de dos formas distintas. La primera, de acuerdo a los principios de NFV, consiste en desacoplar la implementación software de la función de red del hardware sobre el que se ejecuta. Esto permite desplegar elementos como el P-GW o el S-GW sobre hardware de propósito general, más barato. La segunda estrategia, basada en SDN, consiste en implementar las funciones de red como aplicaciones que se ejecutan sobre un sistema operativo de red. Estas aplicaciones, a través del controlador SDN, programan los dispositivos de conmutación de paquetes para que los flujos de tráfico se gestionen conforme a las especificaciones de las funciones de red. Stallings [Sta15] indica que ambas estrategias están relacionadas, ya que las aplicaciones SDN se pueden considerar como componentes software de las VNFs (VNFCs). Por ejemplo, una aplicación SDN que implementa el plano de datos del P-GW puede formar parte de una VNF que incluye, además, otras funciones del plano de control (subsección 7.1.4).

A partir de estas ideas, han surgido diferentes enfoques para diseñar la arquitectura de la red móvil (Capítulo 2), combinando, en mayor o menor medida, soluciones basadas en NFV y SDN. Siguiendo estos enfoques, la sección discute como abordar la evolución de la arquitectura de eMBMS.

7.1.1. vMBMS: virtualización del plano de control e implementación del plano de datos sobre hardware propietario

El 3GPP ha definido como separar los planos de datos y de control de las entidades P-GW y S-GW del núcleo de la red móvil. En el estándar de CUPS [3GP18b], el 3GPP identifica qué funciones pertenecen al plano de datos (PGW-U, SGW-U) y qué funciones pertenecen al plano de control (PGW-C, SGW-C). También define nuevas interfaces que comunican las funciones de ambos planos. La interfaz Sxb, que conecta el PGW-U y el PGW-C, y la interfaz Sxa, que conecta el SGW-U y el SGW-C. Estas interfaces permiten que el plano de control gestione el plano de datos.

Otros proyectos, como por ejemplo M-CORD, también abordan la evolución de la red móvil considerando el estándar de CUPS. Sin embargo, no analizan como aplicar la separación del plano de datos y de control a las entidades funcionales que dan soporte a los servicios de difusión de contenidos.

CUPS no puede aplicarse a la entidad MCE de la arquitectura de eMBMS, o a otras entidades de la arquitectura de LTE como el MME, porque solo implementan funciones del plano de control. La Figura 7.1 muestra que, en este caso, la propuesta es desplegar estas entidades como VNFs en la nube del operador. Pero CUPS si puede aplicarse a las entidades BM-SC y MBMS GW, de la misma forma que se aplica a las pasarelas P-GW y S-GW. Debido a que realizan funciones tanto del plano de datos como del plano de control, su funcionalidad se puede separar en dos componentes (Figura 7.1): uno que solo implementa las funciones del plano de datos (BMSC-U, MBMS GW-U) y otro que solo implementa las funciones del plano de control (BMSC-C, MBMS GW-C). Las funciones del plano de datos se encargan de encaminar los datos de vMBMS. Las funciones del plano de control son responsables de gestionar las funciones del plano de datos [3GP18b], y de participar en diferentes procedimientos de control definidos por el 3GPP, como es la gestión de la sesión, la definición del servicio, la gestión de los recursos radio o el control de admisión.

La Figura 7.1 muestra que el primer enfoque que se ha considerado para diseñar vMBMS se basa en virtualizar las funciones del plano de control (BMSC-C, MBMS GW-C) e implementar las funciones del plano de datos (BMSC-U, MBMS GW-U) sobre equipos basados en hardware propietario. Se asume que los componentes de ambos planos están conectados por interfaces como las que define el 3GPP para las entidades P-GW y S-GW (interfaces Sxa y Sxb). De aquí en adelante, la interfaz que conecta el BMSC-C y el BMSC-U se denominará Sxc, mientras que la interfaz que conecta el MBMS GW-C y el MBMS GW-U se denominará Sxd. A pesar de que se definen nuevas interfaces que no existían en las redes LTE, la interoperabilidad se sostiene tal y como indica el estándar de CUPS [3GP18b].

Aplicar CUPS a las entidades de eMBMS hace posible que los planos de control y datos escalen de forma independiente. En general, las funciones del plano de control tienen altos requisitos de retardo, especialmente las funciones que se encargan de procesar los mensajes de señalización. Las funciones del plano de datos, por el contrario, demandan más ancho de banda

para procesar el tráfico de los usuarios. CUPS permite escalar dinámicamente los recursos que se asignan a las funciones de red de ambos planos, con el objetivo de satisfacer estos requisitos de una forma más eficiente.

Por otro lado, la Figura 7.1 muestra la red de acceso radio, que como ya se ha explicado en el Capítulo 2, solo se compone de estaciones base o eNBs. Tradicionalmente, las estaciones base incluyen una unidad de banda base o BBU y un cabezal de radio remoto o Remote Radio Header (RRH). El RRH se encarga de transmitir la señal radio a los terminales móviles, mientras que la BBU realiza otras funciones, como por ejemplo, aquellas que están relacionadas con la gestión de los recursos radio.

Cloud-Ran (C-RAN) es una propuesta de red de acceso radio para las redes 5G que se basa en el paradigma de la computación en la nube. C-RAN permite agregar todos los recursos computacionales que utilizan las estaciones base en un único pool de recursos [WZHW15]. De esta forma, los RRHs permanecen distribuidos geográficamente, tal y como ocurre en las redes LTE, mientras que las BBUs se despliegan como VNFs en la infraestructura NFV del operador [PFK⁺17]. Esto permite asignar bajo demanda los recursos disponibles de la NFVI a las BBUs que forman el pool.

La virtualización de la red de acceso radio requiere enviar el tráfico de los usuarios a la nube para que sea procesado. El estado del arte muestra que el despliegue de las BBUs se realiza en el borde de la red para reducir retardos [WZHW15]. Así, las BBUs y los RRH se conectan a través del fronthaul mediante enlaces que proporcionan anchos de banda muy altos y retardos muy bajos [CCY⁺15]. Basta *et al.* [BBH⁺17] indican que es posible reducir aún más el retardo

Figura 7.1: Propuesta de diseño de la arquitectura vMBMS. La funcionalidad del MCE se mueve a la infraestructura de computación en la nube. Los planos de datos y control del MBMS GW y del BM-SC se desacoplan. Las funciones del plano de datos (MBMS GW-U y BMSC-U) se ejecutan sobre hardware propietario y las del plano de control (MBMS GW-C y BMSC-C) se despliegan como VNFs en la nube.

si no se virtualiza el plano de datos de las entidades del EPC. Sin embargo, no consideran que la red de acceso radio esté basada en la computación en la nube, y por tanto, asumen que las estaciones base incluyen la unidad de procesamiento BBU y el cabezal de radio remoto RRH. Es decir, en el escenario que consideran, el tráfico de los usuarios no tiene que procesarse en el pool de BBUs que se despliega en la NFVI.

La Figura 7.1 muestra como se integraría la propuesta de Basta *et al.* [BBH⁺17] en este escenario. El plano de datos de las estaciones base (eNBs) atraviesa el backhaul para llegar al núcleo de la red móvil, tal y como ocurre en las redes LTE. El plano de control se encamina a la NFVI desplegada en el borde de la red, donde se encuentra la implementación virtual de las funciones del plano de control del núcleo de la red. Como se puede observar, las VNFs de control están unidas a los equipos que implementan el plano de datos través del backhaul, lo que se debe a que el plano de control debe gestionar el plano de datos.

7.1.2. vMBMS: virtualización del plano de datos y de control

La Figura 7.2 muestra otra estrategia para diseñar la arquitectura de vMBMS. Como puede observar, se mantiene la separación del plano de datos y de control, al igual que en el enfoque anterior. La diferencia en este caso, es que tanto las funciones del plano de datos como las funciones del plano de control se implementan sobre una plataforma de computación en la nube.

Migrar toda la funcionalidad de una entidad lógica a un entorno de gestión de VNFs hace posible que dicha entidad se implemente como una única VNF compuesta por dos VNFCs. Así lo indica el estándar de la ETSI [ETSI14a], que afirma que una VNF puede estar compuesta por dos VNFCs: un componente que se encarga de las funciones del plano de datos y otro que se encarga de las funciones del plano de control. Por ejemplo, el P-GW se podría implementar como una VNF compuesta por dos VNFCs: un VNFC actuaría como PGW-U y el otro como PGW-C.

El estándar de la ETSI también define los mecanismos que permiten escalar los recursos que se asignan a los VNFCs. Es decir, define mecanismos que permiten que las funciones del plano de control y del plano de datos escalen (y evolucionen) de forma independiente si se implementan como VNFCs diferentes. Esto es coherente con el principio de CUPS.

Sin embargo, no es posible abordar esta estrategia para implementar el plano de datos y de control de las entidades del EPC porque, según la ETSI, las interfaces que conectan los VNFCs de una VNF no están bien definidas, no son estables y pueden diferir entre diferentes proveedores de VNF. El 3GPP ha estandarizado las interfaces Sxa y Sxb, por lo que es necesario que las funciones del plano de datos y de control se implementen como dos VNFs independientes. Esta visión también es coherente con el estándar de CUPS puesto que es posible realizar operaciones de escalado sobre una VNF incluso si solo contiene un VNFC.

Así, la Figura 7.2 muestra que todas las funciones de las entidades del EPC (control y datos) se implementan como VNFs en la nube del operador. Es decir, junto a las VNFs de control que

Figura 7.2: Propuesta de diseño de la arquitectura de vMBMS. La funcionalidad del MCE se mueve a la infraestructura de computación en la nube. Los planos de datos y control del MBMS GW y del BM-SC se desacoplan. Tanto las funciones del plano de datos (MBMS GW-U y BMSC-U) como las del plano de control (MBMS GW-C y BMSC-C) se despliegan en la nube.

se identifican en el enfoque anterior se definen las siguientes VNFs: PGW-U, SGW-U, MBMS GW-U, BMSC-U.

El BMSC-U se diferencia del resto de entidades del plano de datos en que es la única entidad funcional de la arquitectura de vMBMS que incluye componentes software que dan servicio extremo a extremo a los usuarios. El BMSC-U actúa como almacén de contenidos e incluye componentes software que son necesarios para dar soporte a los servicios de broadcast, como, por ejemplo, un servidor FLUTE, que se utiliza para el envío unidireccional de ficheros sobre el canal multicast. Virtualizar el plano de datos del BM-SC hace posible instalar, actualizar y desinstalar estos componentes de una forma muy flexible. En definitiva, gracias a la virtualización del BMSC-U, es posible desplegar ágilmente el mecanismo de recuperación de errores que se ha propuesto en el Capítulo 6, tal y como se describe en la Sección 7.2.

Por otro lado, la Figura 7.2 muestra que la virtualización de la BBU incluye los planos de datos y de control. El 3GPP no ha discutido la necesidad de aplicar CUPS a las entidades de la red de acceso radio, por lo que en este caso se considera que la BBU se despliega como una única VNF compuesta por dos VNFC (control y datos).

Estas transformaciones de la arquitectura requieren que el tráfico de los usuarios atraviese la red de transporte para llegar a la nube del operador. Como señalan Basta *et al.* [BKH⁺14], esto añade un retardo adicional, por lo que, actualmente, proyectos como M-CORD están considerando aprovechar las capacidades de la computación en el borde la red (MEC) para desplegar las entidades funcionales de la arquitectura. Otra diferencia con respecto al enfoque de la subsección 7.1.1 es que los planos de datos y control no están limitados por las características de los equipos basados en hardware propietario para escalar.

Con una visión similar a la que se presenta en esta subsección, Chantre y Fonseca [CF18]

proponen virtualizar las entidades de la arquitectura de eMBMS. Sin embargo, su propuesta no considera la aplicación de CUPS, ni analiza como se pueden combinar soluciones de tipo NFV y SDN. A continuación, se describen dos enfoques que si abordan la combinación de estas tecnologías.

7.1.3. vMBMS: VNFs conectadas por una red SDN

La Figura 7.3 muestra como se pueden combinar soluciones basadas en SDN y NFV para dar soporte a los servicios de difusión de contenidos en las redes móviles. La diferencia entre este enfoque y el anterior es que la comunicación entre las VNFs se soporta a través de una red SDN y no mediante modelos tradicionales de conmutación de paquetes.

La propuesta sigue los patrones de diseño que especifica la ETSI para aplicar soluciones de tipo SDN a las arquitecturas de red basadas en NFV [ETS15, ETS17]. La figura muestra que las VNFs se despliegan sobre la infraestructura NFV del operador a través de un conjunto de máquinas virtuales. El VIM gestiona un controlador SDN denominado *Infrastructure SDN controller* (I-SDNC) que se encarga de conectar las máquinas virtuales para que las VNFs puedan comunicarse. Es decir, el controlador solo proporciona servicios de conectividad en la infraestructura NFV pero no se encarga de implementar ninguna función de red de las entidades lógicas de la arquitectura de la red móvil. Las funciones del plano de datos y del plano de

Figura 7.3: Propuesta de diseño de la arquitectura de vMBMS. Todas las funciones de red se virtualizan en la infraestructura NFVI del operador. El controlador SDN (I-SDNC) se encarga de proporcionar conectividad entre las máquinas virtuales que implementan las VNFs.

control están virtualizadas tal y como describe la subsección 7.1.2.

El proyecto M-CORD también considera este enfoque en su arquitectura. El controlador SDN ONOS ejecuta las aplicaciones “VTN” y “Fabric”, que se encargan de implementar las redes virtuales que conectan las VNFs y de gestionar el encaminamiento de los flujos de tráfico en una red de tipo leaf-spine. Sin embargo, aunque M-CORD combina soluciones de tipo NFV y SDN, estas soluciones no aplican en el contexto de los servicios de difusión de contenidos.

Por último, y dado que este enfoque de diseño considera la incorporación de soluciones de tipo SDN en la arquitectura de red, cabe destacar que los nodos de conmutación de la red de transporte también podrían estar basados en SDN [ASR15, RBD⁺14].

La incorporación de las SDNs en las arquitecturas de red facilita la gestión y configuración de las redes y permite realizar un control de grano fino de los flujos de tráfico. El enfoque que presenta esta subsección requiere encaminar los flujos de datos a los centros de datos para que sean procesados por las VNFs. Por tanto, el rendimiento que se obtiene en términos de retardo es semejante al del enfoque anterior. La virtualización de las funciones del plano de datos y de control, desacopladas por la aplicación de CUPS, permite que ambos planos escalen de forma independiente. Sin embargo, la adopción de soluciones basadas en SDN también puede plantear problemas de escalabilidad debido a que el control de la red se centraliza a nivel lógico. Para proteger la red SDN frente a fallos del controlador, se han propuesto arquitecturas distribuidas de controladores de red [BSM18].

7.1.4. vMBMS: aplicaciones SDN para implementar el plano de datos

La Figura 7.4 muestra la última estrategia que se considera para definir vMBMS. Esta propuesta también sigue los patrones de diseño que especifica la ETSI para aplicar soluciones de tipo SDN a las arquitecturas de red basadas en NFV [ETS15, ETS17]. En este caso, y a diferencia del enfoque anterior, se consideran dos controladores. El primero es el denominado *Infrastructure controller* (I-SDNC) cuyas funciones ya se han explicado en la subsección 7.1.3. El segundo, denominado *Tenant SDN controller* (T-SDNC) se encarga de interactuar con las VNFs para tomar decisiones sobre el encaminamiento de los flujos de datos. Es decir, el controlador T-SDNC se encarga de ejecutar las aplicaciones de red de las VNFs para implementar las funciones del plano de datos de las entidades lógicas de la red.

El control de la red que realiza ONOS¹ en el proyecto CORD se basa en este enfoque. ONOS ejecuta las aplicaciones “VTN” y “Fabric” para gestionar la infraestructura NFV, pero también otras aplicaciones de red como “vRouter”, que implementa las funciones que se llevaban a cabo en los Broadband Network Gateway (BNG) tradicionales. Sin embargo, M-CORD no define vMBMS.

En la Figura 7.4 se ha usado la notación C⁺ para hacer referencia a las VNFs que contienen aplicaciones SDN que gestionan el plano de datos. Como se puede observar, las entidades P-GW-U, S-GW-U y MBMS GW-U no aparecen en la Figura 7.4. El procesamiento del tráfico

¹“ONOS,”[Online]. Available: <https://onosproject.org/>

se realiza en los conmutadores y la gestión del tráfico en las VNFs (C⁺). Como se discutirá más adelante, la implementación del BM-SC es diferente debido a que es la única entidad que da servicio extremo a extremo a los usuarios.

El 3GPP indica que el punto de referencia *MI* (Figura 2.4) es una interfaz del plano de datos que conecta el MBMS GW y las estaciones base. A través de esta interfaz, y utilizando IP multicast, los datos de vMBMS se envían a todas las estaciones base que forman parte del área vMBMS. Nuestra propuesta es que esta funcionalidad, que se ejecutaba en el MBMS GW-U (según enfoques anteriores), se implemente como aplicación SDN y se ejecute en el MBMS GW-C⁺. Es decir, la aplicación SDN se encargará de programar los conmutadores SDN, haciendo llamadas a la interfaz norte del controlador T-SDNC, para que se realice IP multicast de los datos del servicio vMBMS a las estaciones base que pertenecen a una misma área. Existen diversas propuestas sobre IP multicast en redes SDN que se podrían aplicar en este contexto [IMA18].

Por otro lado, el MBMS GW-C⁺ implementa las funciones que ya incluía el MBMS GW-C. Es decir, las funciones que dan soporte a los diferentes procedimientos de control que se ejecutan durante la provisión del servicio vMBMS, como por ejemplo, procedimientos para la gestión de la sesión o para el control de admisión. Con este propósito, el MBMS GW-C⁺

Figura 7.4: Propuesta de diseño de la arquitectura de vMBMS. La figura muestra la existencia de dos controladores SDN. El primero, denominado *Infrastructure SDN controller* (I-SDNC), proporciona servicios de conectividad a las VNFs. El segundo, denominado *Tenant SDN controller* (T-SDNC), implementa las funciones del plano de datos de las entidades lógicas de la red móvil.

se conecta a las entidades MME y BMSC-C, que también participan en el intercambio de los mensajes de señalización. El controlador I-SDNC proporciona esta conectividad.

A pesar de que el BM-SC realiza funciones del plano de datos y de control, su diseño es diferente al del MBMS GW o al del P-GW porque contiene diversos componentes software que se utilizan para dar servicio a los usuarios (por ejemplo, un servidor web). La Figura 7.4 muestra que la propuesta en este caso consiste en mantener una VNF en la nube del operador (BMSC-U*) que contiene estos componentes software. El resto de funciones del plano de datos se implementa sobre la red de conmutación SDN de la infraestructura NFV del operador, tal y como se ha propuesto para otras entidades del plano de datos como el PGW-U. Las funciones del plano de control se implementan en el BMSC-C, como en enfoques anteriores.

Como puede observar, el rendimiento en términos de retardo no varía con respecto a enfoques anteriores, puesto que los datos deben procesarse en la nube (en las BBU's y el BMSC-U*). Tampoco se reduce la flexibilidad para desplegar nuevos servicios, NFV-MANO permite asignar los recursos de la NFVI al BMSC-U* bajo demanda y proporciona mecanismos que permiten modificar fácilmente esta VNF.

Los proveedores de contenidos utilizan el punto de referencia *xMB-U* para enviar al operador de red (en este caso al BMSC-U*) los datos que se van a transmitir sobre vMBMS. El contenido se envía a través de la red SDN a todas las BBU's que pertenecen al área vMBMS. Sin embargo, los datos que se envían sobre la red móvil pueden perderse por errores de transmisión. En ese caso, pueden recuperarse mediante conexiones unicast. Para establecer la ruta unicast que siguen los datos transmitidos por vMBMS cuando se produce la recuperación de errores, se propone una nueva aplicación de red que se incluye como VNFC en el BMSC-C⁺. Además, el BMSC-C⁺ realiza las funciones del plano de control que fueron descritas en el Capítulo 2.

Como resumen, la Figura 7.5 muestra los VNFC que incluyen las VNFs MBMS GW-C⁺, BMSC-U* y BMSC-C⁺. En la práctica, cada VNFCs es un contenedor de virtualización que se despliega a través de NFV-MANO, y que se construye a partir de los recursos disponibles de la infraestructura NFV del operador. El MBMS GW-C⁺ incluye dos VNFCs. Uno implementa las funciones relacionadas con los procedimientos de control, el otro, es una aplicación de red que realiza IP multicast de los datos de vMBMS a las estaciones base que pertenecen a la misma área. Ambos VNFCs están conectados entre sí para que las funciones del plano de control gestionen las funciones del plano de datos, tal y como se ha explicado anteriormente. De la misma forma, el BMSC-C⁺ incluye dos componentes VNFC. Uno implementa todas las funciones relacionadas con los procedimientos de control, el otro, es una aplicación SDN que se utiliza para establecer la ruta que siguen los datos de vMBMS durante la recuperación de errores unicast. La figura muestra que el BMSC-U* también incluye varios VNFCs, uno por cada componente software que se requiere para dar servicio a los usuarios. El BMSC-C⁺ y el BMSC-U* se interconectan a través de una nueva interfaz, denominada *Sxc*.

La principal ventaja del diseño que aquí se presenta es que es posible programar el plano de datos de vMBMS a través de una red SDN. Además, la arquitectura sigue proporcionando

Figura 7.5: Composición de las VNFs. La figura muestra los diferentes componentes software (VNFC) que se incluyen en las VNFs MBMS GW-C⁺, BMSC-C⁺ Y BMSC-U*. Cada componente se ejecuta sobre un contenedor que se instancia a través de NFV-MANO utilizando los recursos de la NFVI.

la capacidad de modificar ágilmente los componentes que se utilizan para dar servicio extremo a extremo a los usuarios gracias a la existencia del BMSC-U* como VNF. El rendimiento en términos de retardo es semejante al del enfoque anterior, puesto que sigue siendo necesario procesar los datos de los usuarios en la nube del operador. Sin embargo, la escalabilidad puede verse algo más comprometida porque aumenta la carga de señalización para programar los conmutadores SDN. La red SDN realiza dos funciones en este caso: proporciona conectividad entre las diferentes VNFs e implementa funciones del plano de datos sobre los nodos de conmutación.

7.1.5. Comparación de las diferentes estrategias

Desde la perspectiva del operador de la red, la selección de una u otra estrategia depende de múltiples factores. En primer lugar, del coste. La separación del plano de datos y control que propone el 3GPP exige que los operadores de red inviertan en nuevo equipamiento de red, ya sea propietario (enfoque 7.1.1) o de propósito general y basado en SDN y NFV (enfoques 7.1.2, 7.1.3 y 7.1.4). Sin embargo, y según Valencia *et al.* [VIP15], el OPEX y el CAPEX se reducen con el uso de tecnologías basadas en SDN y NFV, por lo que, atendiendo únicamente a este parámetro, las estrategias más adecuadas serían aquellas donde existe cooperación entre soluciones de tipo NFV y SDN.

Otro parámetro que se debe considerar es la flexibilidad, entendiéndose esta como la capacidad de desplegar ágilmente nuevos servicios de red o mejoras en los servicios de red. Las arquitecturas clásicas, basadas en hardware propietario, presentan muchas limitaciones a este respecto (ver Capítulo 2). La flexibilidad que introducen los paradigmas de red basados en SDN y NFV es significativamente mayor. Lin *et al.* [LWHL18] indican que las arquitecturas de red que combinan soluciones de tipo SDN y NFV son las que proporcionan una mayor flexibilidad. En el marco de la tesis doctoral, esto se correspondería con el enfoque 7.1.4.

Uno de los parámetros más importantes para definir las redes 5G es la latencia. Las arquitecturas de red móvil 5G deben dar soporte a nuevos servicios con estrictos requisitos de retardo como, por ejemplo, servicios para la industria 4.0, servicios de realidad aumentada/virtual, o el Internet táctil [DML⁺17]. La virtualización de las funciones de red hace necesario procesar el tráfico de los usuarios en los centros de datos (enfoques 7.1.2, 7.1.3 y 7.1.4) lo que puede au-

mentar el retardo [BKH⁺14]. Sin embargo, la tendencia de desplegar grandes centros de datos ha cambiado y está dando paso al paradigma de la computación en el borde de la red (MEC), que lleva las capacidades de cómputo cerca de los usuarios. En definitiva, el rendimiento con respecto al retardo de las soluciones que se han planteado anteriormente depende de múltiples factores, como el número y ubicación de los centros de datos [LWHL18] [BKH⁺14] o la utilización o no de soluciones de tipo CDN [YLML10].

CUPS se propone con el objetivo de que los planos de datos y control puedan escalar y evolucionar de forma independiente. Todos los enfoques que se han propuesto cumplen con esa premisa. Sin embargo, no todos proporcionan mecanismos para escalar las funciones de red de forma flexible. El enfoque 7.1.1 propone virtualizar las funciones del plano de control en la nube del operador, por lo que estas funciones son fácilmente escalables a través del gestor de VNFs. El plano de datos, sin embargo, está limitado por las características de los equipos propietarios que se despliegan en la red. Los enfoques 7.1.2 y 7.1.3 proponen migrar a la nube todas las funciones de red, tanto del plano de control como del plano de datos, por lo que la flexibilidad para realizar operaciones de escalado se extiende a ambos planos. Finalmente, el enfoque 7.1.4 propone implementar el plano de datos en los equipos de conmutación basados en SDN (con la excepción del BM-SC). Sin embargo, la escalabilidad puede verse comprometida en este caso por el aumento de la carga de señalización que impone la centralización a nivel lógico del plano de control.

7.2. Despliegue del mecanismo de recuperación de errores adaptativo en arquitecturas de red móvil basadas en virtualización

Esta sección aborda el soporte del mecanismo de recuperación de errores adaptativo propuesto en el Capítulo 6, diseñado para evitar que los clientes de streaming sufran paradas en la reproducción o experimenten largos retardos, asumiendo que las funciones de red de eMBMS han sido virtualizadas en la nube del operador. Las arquitecturas 7.1.2, 7.1.3 y 7.1.4, basadas en NFV, proporcionan los mecanismos de red necesarios para desplegar ágilmente este mecanismo. La arquitectura 7.1.1, por el contrario, mantiene la implementación del plano de datos sobre equipos de red basados en hardware propietario, por lo que, para desplegar este mecanismo, sería necesario que los fabricantes lanzasen nuevas versiones de sus productos incluyendo esta funcionalidad. En este caso, la rigidez de la arquitectura de red dificulta el despliegue del mecanismo de recuperación de errores propuesto.

En un servicio de streaming 3GP-DASH sobre eMBMS, la recuperación de los segmentos perdidos se realiza a través de una función que se ejecuta en el BM-SC (Capítulo 2). El Capítulo 6 propone modificar esta función para que, durante la recuperación de los segmentos perdidos, el reproductor pueda recibir una representación alternativa a la que se envía por eMBMS, más adecuada al ancho de banda disponible en el canal unicast. Es decir, se propone que el BM-SC ejecute un algoritmo de adaptación que seleccione la representación más adecuada a partir del ancho de banda unicast y otros campos del fichero MPD (Capítulo 6).

La virtualización del BM-SC en la arquitectura de vMBMS hace posible modificar fácilmente esta función de red para que implemente el mecanismo de recuperación de errores. En las arquitecturas 7.1.2 y 7.1.3 este mecanismo se desplegaría en el BMSC-U, mientras que en la arquitectura 7.1.4, se desplegaría en el BMSC-U*. Es decir, el mecanismo debe residir en el componente del BM-SC en el que se implementan las funciones de tipo *Associated Delivery* y de tipo *eMBMS Delivery* [3GP17a]. Estas funciones contienen los elementos software que se utilizan para dar servicio extremo a extremo a los usuarios, como, por ejemplo, un servidor web o un servidor FLUTE.

Para dar soporte al mecanismo de recuperación de errores adaptativo en la arquitectura de vMBMS se requieren dos modificaciones. La primera, consiste en modificar la implementación virtual del BM-SC para que incluya el algoritmo de adaptación propuesto. La segunda, consiste en modificar la BBU de la estación base para que esta envíe información del ancho de banda unicast al BM-SC. La BBU se encarga de configurar el valor de varios parámetros del nivel físico que determinan el ancho de banda unicast que se asigna a cada cliente. Estos parámetros son el MCS, el número de PRBs o el número de antenas utilizadas para la transmisión y recepción de la señal [FLB⁺15]. Según la nueva arquitectura de red de acceso radio C-RAN, la BBU se virtualiza en la nube del operador, lo que facilita modificar su implementación para poner a disposición del BM-SC el valor de estos parámetros (o el ancho de banda, si es que tiene registrado su valor). Así, el algoritmo de adaptación puede utilizar esta información para seleccionar la representación multimedia.

Conviene tener en cuenta que la mayoría de los segmentos se entregan correctamente sobre vMBMS y que solo un porcentaje de ellos se recupera por HTTP. Solo en estos casos se ejecuta el algoritmo de adaptación, por lo que no es necesario mandar al BM-SC información del ancho de banda de todos los usuarios conectados a vMBMS. El BM-SC, que recibe las peticiones HTTP de los segmentos, puede solicitar a las BBUs solo el ancho de banda de aquellos clientes que inician la recuperación adaptativa de errores. Dado que un área eMBMS se compone de varias estaciones base, el BM-SC solo realizaría esta petición a las BBUs que gestionan clientes que han solicitado segmentos por HTTP.

Este nuevo procedimiento se ilustra en la Figura 7.6. El BMSC-U (arquitecturas 7.1.2 y 7.1.3) o BMSC-U* (arquitectura 7.1.4) recibe la petición de un terminal (UE_i) para recuperar por HTTP un segmento que se ha perdido durante su transmisión sobre vMBMS (1). Este solicita a la BBU que corresponda, aquella donde el usuario está conectado (BBU_n), información del ancho de banda del canal unicast que transportará el segmento (2). La BBU envía esta información al BM-SC (3), que ejecuta el algoritmo de adaptación (4) a partir del ancho de banda, del fichero MPD y, opcionalmente, del nivel del buffer del reproductor (BL). Finalmente, el BM-SC responde al UE con el segmento que mejor se adapte al ancho de banda del canal unicast (5).

Como se puede observar en la figura, el procedimiento requiere la comunicación entre las entidades BM-SC y BBU. Los estándares del 3GPP que especifican las redes LTE no definen

Figura 7.6: Procedimiento para la recuperación de errores en el entorno NFV. La figura muestra la comunicación entre el terminal (UE), el BMSC-U/U* y el BBU para establecer la recuperación unicast de un segmento perdido durante su transmisión sobre vMBMS. El algoritmo tiene en cuenta el ancho de banda del canal unicast del cliente i -ésimo, las representaciones disponibles en el MPD y el buffer del reproductor BL.

ninguna interfaz que una el eNB y el BM-SC directamente. Si el 3GPP no define esta interfaz cuando aborde el diseño de la arquitectura de vMBMS, sería necesario modificar la implementación virtual de las entidades funcionales que unen las VNFs correspondientes a la BBU y el BM-SC para que el procedimiento pueda ejecutarse. Por ejemplo, y considerando la arquitectura 7.1.2, la petición de ancho de banda que realiza el BMSC-U a la BBU debe atravesar el PGW-U y el SGW-U para llegar a la estación base. Sin embargo, tanto la BBU como el BMSC-U son entidades que están virtualizadas en la nube del operador, por lo que podría evitarse este camino si existiera una interfaz que conectara directamente la BBU y el BMSC-U.

A continuación, la sección detalla como se realiza el envío multicast de los segmentos multimedia sobre la arquitectura de vMBMS, y como se lleva a cabo la recuperación de errores adaptativa. Para ello, se tomará como referencia la arquitectura 7.1.4, que proporciona la mayor flexibilidad al menor coste (subsección 7.1.5).

7.2.1. Distribución de segmentos multimedia en arquitecturas de red móvil basadas en NFV y SDN

La Fig 7.7 muestra el procedimiento de inicio de sesión que se debe llevar a cabo sobre la arquitectura vMBMS 7.1.4 para que se puedan enviar los segmentos multimedia codificados según el estándar de DASH. Inicialmente, el proveedor de contenidos utiliza la interfaz xMB-C (Figura 2.4) para solicitar la creación de un nuevo servicio (1), en este caso, un servicio de streaming 3GP-DASH sobre vMBMS. El Capítulo 2 indica los parámetros que el proveedor de contenidos debe especificar para definir el servicio.

La entidad BMSC-C⁺ procesa esta solicitud y envía un mensaje de petición de inicio de sesión al MBMS GW-C⁺ (2). Este mensaje transporta el valor de diferentes parámetros [3GP17c], como por ejemplo: parámetros de QoS, tiempo en que se iniciará la transmisión de los datos, duración estimada de la transmisión o parámetros relacionados con la formación del área vMBMS. El MBMS GW-C⁺ almacena el valor de estos parámetros, asigna la dirección IP multicast que se utilizará para transportar los datos sobre vMBMS y emite una petición de

Figura 7.7: Procedimiento de inicio de sesión vMBMS. La figura muestra los mensajes de señalización que se intercambian los elementos de la red para que se puedan enviar los segmentos multimedia sobre vMBMS.

inicio de sesión al MME que incluye esta dirección IP y los parámetros ya mencionados (3).

El MME almacena el valor de estos parámetros y emite otro mensaje de petición de inicio de sesión para enviárselos a las BBUs que forman parte del área vMBMS (4). Es entonces cuando se establecen los recursos radio necesarios para transmitir los datos sobre vMBMS (5). Es decir, se determina el número de subtramas que se asignarán al servicio vMBMS y el esquema de modulación y codificación (MCS) que se empleará durante la transmisión de la señal. Una vez finalizada esta operación, las BBUs notifican al MME que la sesión vMBMS se ha establecido con éxito (6) y se unen al grupo multicast utilizando la dirección IP que se transmite en el mensaje de petición de inicio de sesión (7).

En el momento en que el MME recibe la notificación de, al menos, una BBU, puede indicar al MBMS GW-C⁺ que la sesión vMBMS se ha establecido éxito (8). Es entonces cuando el MBMS GW-C⁺ ejecuta la aplicación SDN que permite transmitir los datos a las estaciones base del área vMBMS mediante IP multicast (9). Para ello, la aplicación programa la red de conmutación SDN a través del controlador, y mediante un protocolo de la interfaz sur como OpenFlow.

Tras lo anterior, el MBMS GW-C⁺ envía una respuesta al BMSC-C⁺ para indicarle que la sesión se ha establecido con éxito (10). El BMSC-C⁺ recibe esta respuesta e indica al proveedor de contenidos que el servicio de streaming 3GP-DASH sobre vMBMS se ha creado con éxito (11) y que puede proceder a enviar el contenido multimedia a la entidad BMSC-U* para que sea transmitido sobre vMBMS. El Capítulo 6 discute las diferentes estrategias que puede adoptar el proveedor de contenidos para realizar el envío al BMSC-U*. La figura muestra, como ejemplo, que el proveedor ha utilizado el modo Push (12). El transporte de los segmentos se realiza sobre la interfaz xMB-U (Figura 2.4). Finalmente, se entregan los segmentos multimedia a los clientes a través de la red de acceso radio (13).

Figura 7.8: Recuperación adaptativa de errores en arquitecturas de red móvil basadas en SDN y NFV. La figura muestra el procedimiento que se lleva a cabo para dar soporte a la recuperación unicast de un segmento perdido durante su transmisión sobre vMBMS.

7.2.2. Recuperación adaptativa de errores en arquitecturas de red móvil basadas en NFV y SDN

La Figura 7.8 muestra el procedimiento que se lleva a cabo cuando se establece la recuperación unicast de los segmentos multimedia. Cuando el cliente DASH del terminal de usuario (UE) solicita un segmento, primero comprueba si dicho segmento se encuentra disponible en una memoria cache que almacena los segmentos que han sido transmitidos correctamente sobre vMBMS. Si el segmento no se encuentra en cache, el reproductor realiza una petición HTTP al servidor para solicitar su retransmisión sobre un canal unicast (1). Entonces, el BMSC-U* solicita a la BBU información del ancho de banda unicast asignado a dicho terminal (2). Con esta información (3), el BMSC-U* ejecuta el algoritmo de adaptación que se propone en el Capítulo 6 (4).

Una vez seleccionada la representación multimedia que se va a entregar al cliente, y que puede ser de una calidad igual o inferior a la transmitida sobre vMBMS, el BMSC-U* utiliza la interfaz Sxc para indicarle al BMSC-C+ que debe programar la red de conmutadores SDN para encaminar el flujo de tráfico hasta la BBU que gestiona al terminal que inicio la recuperación (5). Esta programación se realiza gracias a la aplicación SDN que contiene el BMSC-C+ (6), identificada en la subsección 7.1.4.

Tras la instalación de los flujos se envía una notificación al BMSC-U* (7), de nuevo a través de la interfaz Sxc, para que solicite al proveedor de contenidos el segmento perdido (8). Esta solicitud solo se realiza si el algoritmo de adaptación ha seleccionado una representación multimedia diferente a la enviada sobre vMBMS. Como se puede observar en la figura, en este

caso se ha utilizado el modo Pull, que puede ser una estrategia adecuada si la mayor parte de los segmentos se transmiten correctamente sobre vMBMS (Capítulo 6).

Finalmente, el BMSC-U* recibe el segmento del proveedor de contenidos (9) y se lo envía al terminal que inició la recuperación a través de la red de conmutadores (10).

7.3. Conclusiones

Este capítulo analiza diferentes enfoques para definir la arquitectura de soporte a los servicios de difusión de contenidos en las redes móviles basadas en SDN y NFV. El uso combinado de estas tecnologías permite desplegar ágilmente, y a un coste menor, soluciones que mejoran los servicios que se proporcionan sobre la red móvil, y en particular, los servicios de streaming multimedia.

El diseño de vMBMS tiene en cuenta las especificaciones del 3GPP para desacoplar el plano de control y de datos de las entidades funcionales de la red móvil. El 3GPP indica que la aplicación de CUPS permite que los planos de datos y control escalen y evolucionen de forma independiente, y además, es necesaria para poder aplicar soluciones basadas en SDN.

En este capítulo también se aborda la evolución de la arquitectura de eMBMS teniendo en cuenta la necesidad que existe de proporcionar interoperabilidad entre las futuras arquitecturas de red móvil, basadas en tecnologías de virtualización, y las redes LTE actuales. Es por ello que se han considerado las mismas interfaces y entidades lógicas definidas por el 3GPP para LTE, tal y como han hecho proyectos como el de CORD. Aun así, y tal y como se indica en el Capítulo 8, las transformaciones que aquí se proponen se podrían aplicar considerando los estándares del 3GPP en redes móviles 5G.

Finalmente, en el capítulo se explica cómo se puede desplegar el mecanismo de recuperación de errores adaptativo descrito en el Capítulo 6 sobre la arquitectura de vMBMS. En este caso, el análisis se centra en la arquitectura que proporciona la mayor flexibilidad al menor coste.

El contenido de este capítulo ha sido publicado, parcialmente, mediante la realización de varias ponencias en congresos de carácter nacional [LB18] e internacional [MMJ⁺18, LBF⁺17, LBRP⁺17, LBF⁺18]. También se ha participado en la propuesta de una nueva arquitectura de red residencial basada en SDN y NFV [FFB⁺16], presentada en un workshop de carácter internacional.

Capítulo 8

Conclusiones y Trabajos futuros

8.1. Conclusiones generales

Los servicios de difusión de contenidos multimedia imponen estrictos requisitos de retardo y ancho de banda que deben satisfacerse mediante soluciones que permitan distribuir el contenido de una forma más eficiente. Los servicios eMBMS surgieron en las redes móviles con este objetivo. La posibilidad de enviar el mismo contenido multimedia a múltiples usuarios al mismo tiempo, enviando una copia de los datos sobre un canal común, resulta muy adecuado cuando se realizan retransmisiones en directo.

Sin embargo, la revisión del estado del arte revela la necesidad de proponer mecanismos orientados a reducir el retardo del servicio y garantizar la continuidad de la reproducción multimedia. Es por esta razón que la tesis doctoral propone dos mecanismos diferentes que responden a esta necesidad. El primero de ellos, descrito en el Capítulo 4, consiste en un método que permite calcular el buffer mínimo que evita las paradas de servicio en la mayor parte de los clientes de streaming. El impacto de la solución en el retardo es muy significativo, y además, no requiere grandes cambios en el diseño de los dispositivos móviles u otros elementos de la red.

El retardo y la continuidad de la reproducción son factores clave de los servicios de streaming multimedia. Para entender mejor su impacto en la calidad de la experiencia se ha utilizado un modelo objetivo de evaluación que está en proceso de estandarización. El Capítulo 5 muestra que, partiendo de un despliegue específico de la red móvil, la metodología propuesta permite determinar el valor de diferentes parámetros del servicio, como el retardo de buffering inicial, con el objetivo de garantizar un determinado nivel de calidad percibida.

La aparición de mecanismos que permiten adaptar la transmisión del streaming multimedia mediante la actuación de elementos intermedios de la red abre nuevas oportunidades (Capítulo 2). La ventaja de este enfoque es que la adaptación se puede realizar a partir de medidas del ancho de banda disponible y no mediante estimaciones realizadas por el cliente, lo que implica una mayor precisión a la hora de seleccionar la representación multimedia. La desventaja es que requiere realizar modificaciones en los elementos de la red para que ejecuten la lógica de adaptación y puedan asistir a los reproductores. La tesis propone adaptar el uso de este tipo de mecanismos a los servicios de difusión de contenidos multimedia. La idea consiste

en modificar los elementos funcionales de la arquitectura de vMBMS para que ejecuten la lógica de adaptación. Así, es posible seleccionar la representación multimedia que mejor se adapta al ancho de banda disponible durante la recuperación unicast de errores. Los resultados obtenidos muestran que aplicar esta propuesta permite reducir el retardo del servicio sin perder la continuidad de la reproducción multimedia.

Desplegar este mecanismo de recuperación de errores es costoso desde la perspectiva de las arquitecturas de red clásicas, puesto que estas arquitecturas no proporcionan la capacidad de desplegar nuevos servicios ágilmente. El uso de soluciones basadas en la virtualización de las funciones de red y las redes definidas por software promete solventar esta deficiencia. El problema es que, a día de hoy, no se ha definido cómo puede el servicio eMBMS aprovechar las ventajas de esas tecnologías. Esta tesis doctoral propone cuatro enfoques diferentes para diseñar vMBMS, la evolución de la arquitectura de eMBMS (Capítulo 7). Del análisis se concluye que cuanto mayor es la integración entre las tecnologías SDN y NFV, mayor es la agilidad que proporciona la arquitectura para desplegar nuevos servicios y más se reduce el CAPEX y el OPEX. El Capítulo 7 explica cómo se puede desplegar el mecanismo de corrección de errores unicast sobre la arquitectura de vMBMS.

Se puede afirmar que la tesis formula propuestas que contribuyen al diseño de los servicios de difusión de contenidos en las arquitecturas de red móvil 4G y posteriores, por lo que cumple con el objetivo general y los objetivos específicos definidos en el Capítulo 1. Como muestran los resultados obtenidos, la aplicación de estas propuestas permite mejorar los servicios multimedia mediante la reducción de la latencia y garantizando la continuidad de la reproducción multimedia.

8.2. Trabajos futuros

Además de profundizar en algunas de las líneas de investigación ya abiertas, el trabajo futuro debe centrarse en abrir nuevas líneas de investigación que permitan aumentar el impacto de la contribución que realiza esta tesis al diseño de los servicios multimedia en las arquitecturas de red móvil basadas en SDN y NFV. A continuación, se ofrece una visión sobre los trabajos de investigación que se pueden acometer en el futuro.

- Considerando la combinación de soluciones basadas en NFV y SDN, el Capítulo 7 analiza diferentes enfoques para diseñar la evolución de la arquitectura de soporte a los servicios de difusión de contenidos en las redes móviles. Se ha proporcionado una comparación cualitativa de algunas propiedades de la arquitectura, pero también se podría realizar una comparación cuantitativa. Con este propósito, se planea diseñar y desarrollar un entorno de experimentación que permita determinar cual es el enfoque más adecuado para un determinado escenario de red. Utilizando esta plataforma de experimentación, también se pueden realizar pruebas de concepto que permitan analizar cómo se puede desplegar el mecanismo de recuperación de errores propuesto en el Capítulo 6 en un

entorno de pruebas real.

- Las arquitecturas de red móvil basadas en SDN y NFV plantean diversos desafíos para garantizar bajos retardos. De hecho, una problemática que está siendo analizada actualmente, es el incremento de retardo que se introduce por la necesidad de mandar el tráfico de los usuarios a los centros de datos para que sea procesado. En este sentido, se propone extender el análisis realizado en el Capítulo 4 para que considere la incorporación de nuevos paradigmas de red como SDN, NFV y MEC.
- En esta tesis doctoral se han formulado propuestas de diseño de vMBMS analizando qué transformaciones debería sufrir la arquitectura de la red LTE para proporcionar la capacidad de desplegar nuevos servicios ágilmente. Este enfoque está alineado con el de otros proyectos como M-CORD. Sin embargo, el 3GPP ya ha publicado estándares que especifican el núcleo de la red 5G. Los estándares no abordan como dar soporte a los servicios eMBMS, por lo que es necesario analizar cómo se pueden integrar las propuestas de la tesis doctoral en este contexto.
- Para validar las propuestas surgidas durante el desarrollo de la tesis doctoral se ha desarrollado un modelo analítico que realiza algunas asunciones que simplifican el problema. Aunque el modelo desarrollado ha sido útil para validar las propuestas, resulta conveniente abordar su revisión y adaptarlo a las características de las arquitecturas de red móvil basadas en virtualización.
- La computación en el borde la red es una evolución de la computación en la nube que busca desplegar los centros de datos en ubicaciones más cercanas a los usuarios finales: las centrales locales. La aplicación de soluciones basadas en este nuevo paradigma puede reducir la latencia de los servicios multimedia, y en particular, de las retransmisiones en directo, ya que es posible acceder al contenido desde nodos de una red de distribución de contenidos virtualizada que se despliegan en el borde de la red. Por esta razón, se propone abrir una nueva línea de investigación que valore la posibilidad de desplegar el servicio vMBMS en las centrales locales definidas por software, y que evalúe soluciones que permitan hacer un uso eficiente de los nodos virtuales de la red de distribución de contenidos.
- Network slicing es un nuevo paradigma de red que posibilita el despliegue de servicios mediante la creación de redes virtuales aisladas sobre una infraestructura de red común. El aislamiento de las redes virtuales o slices permite aplicar nuevas soluciones orientadas a la provisión de la calidad de servicio extremo a extremo. Como trabajo futuro, se propone analizar el impacto del Network Slicing en los servicios de difusión de contenidos. Para ello, se definirá un slice que integre las soluciones que se proponen en esta tesis doctoral para garantizar la continuidad de la reproducción multimedia y reducir el retardo de los servicios de streaming multimedia sobre vMBMS.

8.3. Divulgación de resultados

Las propuestas de esta tesis surgen como resultado del trabajo de investigación realizado durante la participación del doctorando en dos proyectos de carácter nacional y una red de excelencia. La mayor parte de los resultados obtenidos durante la realización de la tesis han sido publicados como artículos en revistas indexadas en el Journal Citation Reports (JCR) y como ponencias en congresos de ámbito nacional e internacional. A continuación, se proporciona más información sobre las publicaciones y los proyectos de investigación en los que el autor ha participado.

8.3.1. Proyectos de investigación subvencionados

- Nombre: LTextreme (Optimización de Servicios Multiusuario y Multimedia sobre LTE y LTE-Advanced)

Identificador: IPT-2012-0525-430000

Fecha de inicio - Fecha de fin: Mayo 2013 - Diciembre 2015

Entidad Financiadora: Ministerio de Economía y Competitividad. Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (INNPACKTO)

Investigador Principal: Encarnación Pastor Martín

Resumen: se diseña y desarrolla una arquitectura de sistema para la optimización de los servicios multiusuario y multimedia ofrecidos sobre las redes móviles LTE. Las actividades del grupo de investigación se centran en proponer nuevas soluciones que permitan mejorar aspectos claves de dichos servicios, tales como la continuidad de la reproducción multimedia o la latencia de las retransmisiones en directo.

- Nombre: ElasticNetworks (Nuevos Paradigmas De Redes Elásticas Para Un Mundo Radicalmente Basado En Cloud Y Fog Computing)

Identificador: TEC2015-71932-REDT

Fecha de inicio - Fecha de fin: Diciembre 2015 - Noviembre 2018

Entidad financiadora: Financiado por el Ministerio de Economía y Competitividad. Redes de excelencia

Investigador Principal: David Larrabeiti López

Responsable en la UPM: David Fernández Cambrónero

Resumen: los estrictos requisitos de latencia y ancho de banda de algunas aplicaciones hacen necesario rediseñar la red de transporte para que esta sea ágil, flexible y programable. El proyecto pretende resolver el problema tecnológico que supone construir una red de muy altas prestaciones soportada por la conectividad con la nube (Cloud) o el entorno (Fog).

- Nombre: GREDOS (Redes definidas por Software para la prestación y gestión de servicios en entornos residenciales y de campus)

Identificador: TEC2015-67834-R

Fecha de inicio - Fecha de fin: Enero 2016 - Diciembre 2018

Entidad financiadora: Ministerio de Economía y Competitividad. Programa estatal de investigación, Desarrollo e Innovación orientada a los Retos de la Sociedad.

Investigador Principal: David Fernández Cambronero

Resumen: el proyecto GREDOS surge tras la aparición de las redes definidas por software (SDN), la virtualización de las funciones de red (NFV) y el cloud computing. Su objetivo es mejorar la prestación y gestión de servicios en entornos residenciales y de campus. Esto incluye los servicios multimedia y el Internet de las cosas.

8.3.2. Publicaciones

- Revistas indexadas en el JCR (2017)

1. M. Aguayo, L. Bellido, C. M. Lentisco and E. Pastor, “DASH Adaptation Algorithm Based on Adaptive Forgetting Factor Estimation” *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1224-1232, May 2018.
JCR: 3.977 (Q1)
2. C. M. Lentisco, L. Bellido, J. C. Cuellar Q., E. Pastor and J. L. Arciniegas H., “QoE-Based Analysis of DASH Streaming Parameters Over Mobile Broadcast Networks” *IEEE Access*, vol. 5, pp. 20684-20694, 2017.
JCR: 3.557 (Q1)
3. C. M. Lentisco, Luis Bellido, and Encarna Pastor, “Seamless Mobile Multimedia Broadcasting Using Adaptive Error Recovery” *Mobile Information Systems*, vol. 2017, Article ID 1847538, 11 pages, 2017.
JCR: 0.958 (Q4)
4. C. M. Lentisco, L. Bellido and E. Pastor, “Reducing Latency for Multimedia Broadcast Services Over Mobile Networks” *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 173-182, Jan. 2017. JCR: 3.977 (Q1)

- Comunicaciones y ponencias presentadas a congresos

1. X. Masip, E. Marín, A. Jurnet, P. Reviriego, R. Flores, C. M. Lentisco, D. Fernández, L. Bellido, S. Gómez, E. Calle, G. Otero, J. A. Hernández and D. Larrabeiti “Elastic Networks Thematic Network Results II: Edge Computing and Wireless Support,” in *Proc. International Conference on Transparent Optical Networks (ICTON)*, Bucharest, 2018, pp. 1-4.
2. C. M. Lentisco, L. Bellido, D. Fernández, R. Flores, and E. Pastor, “Virtual Multimedia Broadcast Multicast Service Architecture,” *International Workshop on Elastic Networks Design and Optimization*, Valencia, Spain, Jun. 2018.

3. C. M. Lentisco, L. Bellido, “Multimedia streaming over mobile networks based on virtualization,” *II edición simposio: cuéntanos tu tesis*. Universidad Politécnica de Madrid, Madrid, Spain, Mar. 2018.
4. C. M. Lentisco, L. Bellido, J. Ruiz-Piñar, R. Flores, E. Pastor, and D. Fernández, “Multimedia Streaming Service over 5G networks,” *International Workshop on Elastic Networks Design and Optimization*, Castelldefels, Spain, Oct. 2017
5. C. M. Lentisco, L. Bellido, R. Flores, D. Fernández, and E. Pastor, “Leveraging edge computing for broadcast as a service in 5G networks,” *International Workshop on Elastic Networks Design and Optimization*, Vilanova, Spain, Jun. 2017.
6. A. de la Fuente, C. M. Lentisco, L. Bellido, R. Pérez Leal, A. G. Armada, and E. Pastor, “End to end measurements of multimedia streaming over LTE” in *Proc. European Conference on Networks and Communications (EuCNC)*, Athens, Greece, Jun. 2016, pp. 57-61.
7. R. Flores, D. Fernández, L. Bellido, C. M. Lentisco, E. Pastor, J. Ruiz, O. Walid, “A user-centric cloud-based SDN/NFV architecture for Residential Networks,” *Int. Workshop on Elastic Networks Design and Optimization*, Cartagena, Spain, May. 2016.
8. L. Bellido, C. M. Lentisco, M. Aguayo, E. Pastor. “Supporting handover between LTE video broadcasting and unicast streaming”. in *Proc. International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST)*, Cambridge, United Kingdom, Sept. 2015.
9. C. M. Lentisco, M. Aguayo, L. Bellido, E. Pastor, D. De-Antonio-Monte, A. García Bolívar. “A virtualized platform for analyzing LTE broadcast services”. in *Proc. European Conf. on Networks and Communications (EuCNC)*, Paris, France, Jul. 2 2015, pp. 517-521.
10. A. de la Fuente, C. M. Lentisco, L. Bellido, R. Pérez Leal, E. Pastor y A. García Armada, “Analysis of the impact of FEC techniques on a multicast video streaming service over LTE”. in *textitProc. European Conference on Networks and Communications (EuCNC)*, Paris, France, Jul. 2015, pp. 224-228.
11. C. M. Lentisco, L. Bellido, E. Pastor. “Application level mechanisms for service continuity in an LTE multimedia broadcast service,” in *Proc. IEEE International Symposium on Consumer Electronics (ISCE)*, Madrid, Spain, Jun. 2015.

12. C. M. Lentisco, L. Bellido, A. de la Fuente, E. Pastor, R. P. Leal and A. G. Armada, “A model to evaluate MBSFN and AL-FEC techniques in a multicast video streaming service,” in *Proc. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Larnaca, 2014, pp. 691-696.
13. C. M. Lentisco, L. Bellido, E. Pastor, A. de la Fuente, “Unicast and multicast steaming services over LTE”. in *Proc. European Conference on Networks and Communications (EuCNC)*, Bologna, Italy, Jun 2014.

- Trabajo Fin de Grado

1. Eloy Ramírez Hernanz, “Desarrollo y aplicación de herramientas para el análisis del envío anticipado de video mediante HTTP/2 en servicios de streaming multimedia,” *Trabajo Fin de Grado*, Tutor: C. M. Lentisco, Ponente: Luis Bellido Triana, Jul. 2016

Lista de acrónimos

3GPP 3rd Generation Partnership Project.

5GPPP 5G Public-Private Partnership.

AFF Adaptive Forgetting Factor.

AL-FEC Application Layer Forward Error Correction.

ATSC Advanced Television Systems Committee.

BBU Base Band Unit.

BM-SC Broadcast Multicast Service Centre.

BNG Broadband Network Gateway.

C-RAN Cloud-Ran.

CBQ Class-based Queuing.

CDN Content Delivery Network.

CORD Central Office Re-architected as a Datacenter.

CUPS Control and User plane Separation.

DASH Dynamic Adaptive Streaming over HTTP.

E-UTRAN Evolved Universal Terrestrial Radio Access Network.

EM Element Management.

eMBMS evolved Multimedia Broadcast and Multicast Service.

eNB evolved Node B.

EPC Evolved Packet Core.

ETSI European Telecommunications Standards Institute.

FDT File Delivery Table.

FEC OTI FEC Object Transmission Information.

FIB Forwarding Information Database.

FLUTE File Delivery over Unidirectional Transport.

FR Full Reference.

HSS Home Subscriber Server.

ISD Intersite Distance.

LTE Long Term Evolution.

MBMS GW Multimedia Broadcast Multicast Service Gateway.

MBSFN Multicast/Broadcast over Single Frequency Network.

MCE Multi-cell/multicast Coordination Entity.

MCS Modulation and Coding Scheme.

MEC Multi-access Edge Computing.

MME Mobility Management Entity.

MOS Mean Opinion Score.

MPD Media Presentation Description.

NetEm Network Emulation.

NFV Network Functions Virtualization.

NFV-MANO NFV Management and Orchestration.

NFVI NFV Infrastructure.

NFVO Network Functions Virtualisation Orchestrator.

NR No Reference.

NSD Network Service Descriptor.

ONF Open Networking Foundation.

OSS/BSS Operations Support System/Business Support System.

OVA Open Virtual Appliance.

P-GW Packet Data Network Gateway.

PCRF Policy and Charging Rules Function.

PER Packet Error Rate.

PRB Physical Resource Block.

PSNR Peak Signal-to-Noise Ratio.

QoE Quality of Experience.

QoS Quality of Service.

ROUTE Real-Time Object Delivery over Unidirectional Transport.

RRH Remote Radio Header.

RTT Round Trip Time.

S-GW Serving Gateway.

SAND Server and Network Assisted DASH.

SCEF Service Creation Environment Function.

SDN Software Defined Networks.

SDR Service Data Rate.

SINR Signal to Interference plus Noise Ratio.

SNMP Simple Network Management Protocol.

TC Linux Traffic Control.

U-vMOS User, Unified, Ubiquitous-Mean Opinion Score for Video.

UDP User Datagram Protocol.

UIT Unión Internacional de Telecomunicaciones.

UMTS Universal Mobile Telecommunications System.

URL Uniform Resource Locator.

USBD User Service Bundle Description.

VBR Variable Bit Rate.

VIM Virtualised Infrastructure Manager.

VLD Virtual Link Descriptor.

vMBMS virtual MBMS.

vMOS video Mean Opinion Score.

VNF Virtualised Network Function.

VNFC VNF Component.

VNFD VNF Descriptor.

VNFFG VNF Forwarding Graph.

VNFFGD VNFFG Descriptor.

VNFM VNF Manager.

VNX Virtual Networks over linuX.

VQM Video Quality Metric.

XML eXtensible Markup Language.

Bibliografía

- [3GP17a] *Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs*, 3GPP TS 26.346 v14.2.0, 2017.
- [3GP17b] *Transparent End-to-End Packet-Switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP*, 3GPP TS 26.247 v14.1.0, 2017.
- [3GP17c] *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description (Release 15)*, 3GPP TS 23.246 v15.0.0, 2017.
- [3GP17d] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures*, 3GPP TS 36.213 v14.4.0, 2017.
- [3GP18a] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, 3GPP TS 36.300 v15.2.0, 2018.
- [3GP18b] *Architecture enhancements for control and user plane separation of EPC nodes*, 3GPP TS 23.214 v15.2.0, 2018.
- [3GP18c] *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2*, 3GPP TS 23.501 v15.3.0, 2018.
- [3GP18d] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 15)*, 3GPP TS 36.331 v15.4.0, 2018.
- [5GP] “5G-Xcast: Mobile core network,” 5GPPP, White Paper, Accessed: Jan. 2019 [Online]. Available: http://5g-xcast.eu/wp-content/uploads/2018/06/5G-Xcast_D4.1_v1.0.pdf.
- [ABC⁺10] R. Agusti, F. Bernardo, F. Casadevall, R. Ferrús, J. Pérez, and O. Sallent, “Arquitectura del sistema LTE,” in *LTE: Nuevas tendencias en comunicaciones móviles*, Madrid, España: Fundación Vodafone España, 2010, ch. 2, pp. 53–111.

- [ABG14] A. Ahmed, L. M. Boulahia, and D. Gaiti, “Enabling Vertical Handover Decisions in Heterogeneous Wireless Networks: A State-of-the-Art and A Classification,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 776–811, 2014.
- [ABLP18] M. Aguayo, L. Bellido, C. M. Lentisco, and E. Pastor, “DASH Adaptation Algorithm Based on Adaptive Forgetting Factor Estimation,” *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1224–1232, May 2018.
- [AGH⁺15] V. K. Adhikari, Y. Guo, F. Hao, V. Hilt, Z. Zhang, M. Varvello, and M. Steiner, “Measurement Study of Netflix, Hulu, and a Tale of Three CDNs,” *IEEE/ACM Transactions on Networking*, vol. 23, no. 6, pp. 1984–1997, Dec. 2015.
- [AKA⁺17] I. Alawe, A. Ksentini, Y. H. Aoul, P. Bertin, and A. Kerbellec, “On evaluating different trends for virtualized and SDN-ready mobile network,” in *Proc. IEEE International Conference on Cloud Networking (CloudNet)*, Sept. 2017, pp. 1–6.
- [ALL] “Lessons learned from trials and early deployments,” LTE Broadcast Alliance, White Paper, Accessed: Jan. 2019 [Online]. Available: <https://lte-broadcast-alliance.com/>.
- [AM14] A. Anttonen and A. Mämmelä, “Interruption Probability of Wireless Video Streaming With Limited Video Lengths,” *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 1176–1180, Jun. 2014.
- [ASR15] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, “Software-defined networking in cellular radio access networks: potential and challenges,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 150–156, Jan. 2015.
- [ATS17] *ATSC Standard: Signaling, Delivery, Synchronization, and Error Protection*, ATSC A/331:2107, 2017.
- [BBH⁺17] A. Basta, A. Blenk, K. Hoffmann, H. J. Morper, M. Hoffmann, and W. Kellerer, “Towards a Cost Optimal Design for a 5G Mobile Core Network Based on SDN and NFV,” *IEEE Transactions on Network and Service Management*, vol. 14, no. 4, pp. 1061–1075, Dec. 2017.
- [Bin15] B. Bing, “Assessing and Enhancing Video Quality,” in *Next-Generation Video Coding and Streaming*, Hoboken, New Jersey, USA: John Wiley and Sons, 2015, ch. 5, pp. 165–182.
- [BKH⁺14] A. Basta, W. Kellerer, M. Hoffmann, H. J. Morper, and K. Hoffmann, “Applying NFV and SDN to LTE Mobile Core Gateways, the Functions Placement Problem,” in *Proc. Workshop on All Things Cellular: Operations, Applications, & Challenges (AllThingsCellular)*, 2014, pp. 33–38.

- [BKBP11] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, “AL-FEC for streaming services over LTE systems,” in *Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Oct. 2011, pp. 1–5.
- [BKBP13] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, “Embracing RaptorQ FEC in 3GPP multicast service,” *Wireless Networks*, vol. 19, no. 5, pp. 1023–1035, Jul. 2013.
- [BLAP15] L. Bellido, C. M. Lentisco, M. Aguayo, and E. Pastor, “Supporting handover between LTE video broadcasting and unicast streaming,” in *Proc. International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST)*, Sep. 2015, pp. 329–334.
- [BPT15] M. Belshe, R. Peon, and M. Thomson, *Hypertext Transfer Protocol Version 2 (HTTP/2)*, IETF RFC 7540, May 2015.
- [BRZS17] D. Bhat, A. Rizk, M. Zink, and R. Steinmetz, “Network Assisted Content Distribution for Adaptive Bitrate Video Streaming,” in *Proc. ACM on Multimedia Systems Conference (MMSys)*, Taipei, Taiwan, Jun. 2017, pp. 62 – 75.
- [BSA⁺12] H. Bobarshad, M. Schaar, A. H. Aghvami, R. S. Dilmaghani, and M. R. Shikh-Bahaei, “Analytical Modeling for Delay-Sensitive Video Over WLAN,” *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 401–414, Apr. 2012.
- [BSM18] F. Bannour, S. Souihi, and A. Mellouk, “Distributed SDN Control: Survey, Taxonomy, and Challenges,” *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, 2018.
- [CB15] S. S. Chaudhari and R. C. Biradar, “Survey of Bandwidth Estimation Techniques in Communication Networks,” *Wireless Personal Communications*, vol. 83, no. 2, pp. 1425–1476, Jul. 2015.
- [CCPM13] L. D. Cicco, V. Caldaralo, V. Palmisano, and S. Mascolo, “ELASTIC: A Client-Side Controller for Dynamic Adaptive Streaming over HTTP (DASH),” in *Proc. International Packet Video Workshop*, San Jose, CA, USA, Dec. 2013, pp. 1–8.
- [CCW⁺12] M. Chiosi, D. Clarke, P. Willis, A. Reid, J. Feger, and M. B. et al., “Network functions virtualisation: An introduction, benefits, enablers, challenges and call for action,” in *SDN and OpenFlow World Congress*, Darmstadt, Germany, Oct. 2012, pp. 22–24.
- [CCY⁺15] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, “Cloud RAN for Mobile Networks—A Technology Overview,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, 2015.

- [CCZ⁺17] G. Cofano, L. D. Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo, “Design and Performance Evaluation of Network-assisted Control Strategies for HTTP Adaptive Streaming,” *ACM Transactions on Multimedia Computing Communications and Applications*, vol. 13, no. 3s, pp. 42:1 – 42:24, Aug. 2017.
- [CDLL15] L. M. Contreras, P. Doolan, H. Lønsethagen, and D. R. López, “Operational, organizational and business challenges for network operators in the context of SDN and NFV,” *Computer Networks*, vol. 92, no. 2, pp. 211 – 217, 2015.
- [CF17] H. D. Chantre and N. L. S. Fonseca, “Redundant placement of virtualized network functions for LTE evolved Multimedia Broadcast Multicast Services,” in *Proc. IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.
- [CF18] H. Chantre and N. L. S. Fonseca, “Reliable Broadcasting in 5G NFV-Based Networks,” *IEEE Communications Magazine*, vol. 56, no. 3, pp. 218–224, Mar. 2018.
- [CFSD90] J. Case, M. Fedor, M. Schoffstall, and J. Davin, *A Simple Network Management Protocol (SNMP)*, IETF RFC 1157, May 1990.
- [CIS] “Cisco Visual Networking Index: Forecast and Trends, 2017-2022,” Cisco, San Jose, CA, USA, White Paper, Accessed: Jan. 2019 [Online]. Available: <https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html>.
- [CLFT14] M. Claeys, S. Latre, J. Famaey, and F. D. Turck, “Design and Evaluation of a Self-Learning HTTP Adaptive Video Streaming Client,” *IEEE Communications Letters*, vol. 18, no. 4, pp. 716–719, Apr. 2014.
- [COLM11] N. Cardona, J. J. Olmos, M. G. Lozano, and J. F. Monserrat, “El Núcleo de red - EPC,” in *3GPP LTE: Hacia la 4G móvil*, Barcelona, España: Marcombo, 2011, ch. 3.
- [CPG⁺13] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, “Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.
- [CWZ15] Y. Chen, K. Wu, and Q. Zhang, “From QoS to QoE: A Tutorial on Video Quality Assessment,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.
- [DAS] DASH-IF forum, “dash.js JavaScript Reference Client,” Accessed: Jan. 2019 [Online]. Available: <https://reference.dashif.org/dash.js/>.

- [DM11] S. N. Donthi and N. B. Mehta, “An Accurate Model for EESM and its Application to Analysis of CQI Feedback Schemes and Scheduling in LTE,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3436–3448, Oct. 2011.
- [DML⁺17] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami, “Internet of skills, where robotics meets AI, 5G and the Tactile Internet,” in *Proc. European Conference on Networks and Communications (EuCNC)*, Jun. 2017, pp. 1–5.
- [DNK16] T. X. Do, V. Y. Nguyen, and Y. Kim, “SDN-based mobile packet core for multicast and broadcast services,” *Wireless Networks*, Dec. 2016, <https://doi.org/10.1007/s11276-016-1433-6>
- [ERI] “Ericsson Mobility Report,” Ericsson, Kista, Stockholm, Sweden, White Paper, Accessed: Jan. 2019 [Online]. Available: <https://www.ericsson.com/en/mobility-report/reports/november-2018>.
- [ESS⁺15] A. E. Essaili, D. Schroeder, E. Steinbach, D. Staehle, and M. Shehada, “QoE-Based Traffic and Resource Management for Adaptive HTTP Video Delivery in LTE,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 988–1001, Jun. 2015.
- [ETS13] *Network Functions Virtualisation (NFV); Use Cases*, ETSI GS NFV 001 v1.1.1, 2013.
- [ETS14a] *Network Function Virtualisation (NFV); Virtual Network Functions Architecture*, ETSI GS NFV-SWA 001 v1.1.1, 2014.
- [ETS14b] *Network Function Virtualisation (NFV); Management and Orchestration*, ETSI GS NFV-MAN 001 v1.1.1, 2014.
- [ETS15] *Network Function Virtualisation (NFV); Ecosystem; Report on SDN usage in NFV Architectural Framework*, ETSI GS NFV-EVE 005 v1.1.1, 2015.
- [ETS17] *Network Function Virtualisation (NFV) Release 3; Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework*, ETSI GR NFV-EVE 012 v3.1.1, 2017.
- [FA08] B. Furht and S. A. Ahson, “Application Layer Forward Error Correction for Mobile Multimedia Broadcasting,” in *Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T and MediaFLO*, Boston, MA, USA: Auerbach Publications, 2008, ch. 2.

- [FFB⁺16] R. Flores, D. Fernández, L. Bellido, C. M. Lentisco, E. Pastor, J. Ruiz-Piñar, and O. Walid, “A User Centric Cloud-Based SDN/NFV Architecture for Residential Networks,” in *International Workshop on Elastic Networks Design and Optimization*, Cartagena, Spain, May. 2016.
- [FFBG12] I. Fez, F. Fraile, R. Belda, and J. C. Guerri, “Analysis and Evaluation of Adaptive LDPC AL-FEC Codes for Content Download Services,” *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 641–650, Jun. 2012.
- [FFm] FFMpeg, “Download FFMpeg,” Accessed: Jan. 2019 [Online]. Available: <https://www.ffmpeg.org/download.html>.
- [FG14] I. Fez and J. C. Guerri, “An Adaptive Mechanism for Optimal Content Download in Wireless Networks,” *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 1140–1155, Jun. 2014.
- [FLB⁺15] A. Fuente, C. M. Lentisco, L. Bellido, R. P. Leal, E. Pastor, and A. G. Armada, “Analysis of the impact of FEC techniques on a multicast video streaming service over LTE,” in *Proc. European Conference on Networks and Communications (EuCNC)*, Jun. 2015, pp. 219–223.
- [FLB⁺16] A. Fuente, C. M. Lentisco, L. Bellido, R. P. Leal, A. G. Armada, E. Pastor, and A. G. Bolívar, “End to end measurements of multimedia streaming over LTE,” in *Proc. European Conference on Networks and Communications (EuCNC)*, Jun. 2016, pp. 57–61.
- [FRB⁺16] D. Fernández, F. J. Ruiz, L. Bellido, E. Pastor, O. Walid, and V. Mateos, “Enhancing Learning Experience in Computer Networking through a Virtualization-Based Laboratory Model,” *International Journal of Engineering Education*, vol. 32, no. 6, pp. 2569–2584, Dec. 2016.
- [GB14] P. Goransson and C. Black, “Why sdn?” in *Software Defined Networks: A Comprehensive Approach*, Amsterdam, Netherlands: Elsevier, 2014, pp. 21–36.
- [GGTE09] G. Gómez, J. P. Gonzalez, C. A. Torres, and J. Entrambasaguas, “A Real-Time End-to-End Testbed for Evaluating the Performance of Multimedia Services,” in *Future Multimedia Networking*, Berlin Heidelberg: Springer, 2009, pp. 212–217.
- [GPA] GPAC Multimedia Open Source Project, “MP4Box,” Accessed: Jan. 2019 [Online]. Available: <http://gpac.wp.mines-telecom.fr/mp4box/>.
- [HBP⁺18] J. Hooft, C. Boom, S. Petrangeli, T. Wauters, and F. Turck, “An HTTP/2 push-based framework for low-latency adaptive streaming through user profiling,” in

- Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Apr. 2018, pp. 1–5.
- [HHB⁺15] R. Huysegems, J. Hooft, T. Bostoen, P. R. Alface, S. Petrangeli, T. Wauters, and F. Turck, “HTTP/2-Based Methods to Improve the Live Experience of Adaptive Streaming,” in *Proc. ACM International Conference on Multimedia (MM)*, Brisbane, Australia, Oct. 2015, pp. 541–550.
- [HPW⁺16] J. Hooft, S. Petrangeli, T. Wauters, R. Huysegems, P. R. Alface, T. Bostoen, and F. Turck, “HTTP/2-Based Adaptive Streaming of HEVC Video Over 4G/LTE Networks,” *IEEE Communications Letters*, vol. 20, no. 11, pp. 2177–2180, Nov. 2016.
- [HPW⁺18] J. Hooft, S. Petrangeli, T. Wauters, R. Huysegems, T. Bostoen, Filip, and Turck, “An HTTP/2 Push-Based Approach for Low-Latency Live Streaming with Super-Short Segments,” *Journal of Network and Systems Management*, vol. 26, no. 1, pp. 51–78, Jan. 2018.
- [HSBP13] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, “Internet video delivery in youtube: From traffic measurements to quality of experience,” in *Data Traffic Monitoring and Analysis*, Berlin, Heidelberg: Springer, 2013, pp. 264–301.
- [HSH⁺11] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, “Quantification of YouTube QoE via Crowdsourcing,” in *Proc. IEEE International Symposium on Multimedia (ISM)*, Dec. 2011, pp. 494–499.
- [HSMA14] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, “NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC),” *IEEE Network*, vol. 28, no. 6, pp. 18–26, Nov. 2014.
- [HUA] “Technical White Paper on Mobile Bearer Network Requirements for Mobile Video Services,” Huawei, Shenzhen, Guangdong China, White Paper, Accessed: Jan. 2019 [Online]. Available: <https://www.huawei.com/en/industry-insights/technology/white-papers/moving-video>.
- [IMA18] S. Islam, N. Muslim, and J. W. Atwood, “A Survey on Multicasting in Software-Defined Networking,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 355–387, 2018.
- [ISO14] *Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segments formats*, ISO/IEC 23009-1, 2014.
- [ISO17] *Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 5: Server and network assisted DASH (SAND)*, ISO/IEC 23009-5, 2017.

- [ITU10] *Reference Algorithm for Computing Peak Signal to Noise Ratio of a Processed Video Sequence With Compensation for Constant Spatial Shifts, Constant Temporal Shift, and Constant Luminance Gain and Offset*, ITU-T J.340, 2010.
- [ITU17] *Vocabulary for performance, quality of service and quality of experience*, ITU-T P.10/G.100, 2017.
- [ITU18] *Requirements of the IMT-2020 network*, ITU-T Y.3101, 2018.
- [JSZ14] J. Jiang, V. Sekar, and H. Zhang, “Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming With Festive,” *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 326–340, Feb. 2014.
- [KO13] U. Kumar and O. Oyman, “QoE evaluation for video streaming over eMBMS,” in *Proc. International Conference on Computing, Networking and Communications (ICNC)*, Jan. 2013, pp. 555–559.
- [KR17] J. Kurose and K. Ross, “Multimedia Networking,” in *Computer Networking: A Top-Down Approach*, Strand, London, UK: Pearson, 2017, ch. 7.
- [LAB⁺15] C. M. Lentisco, M. Aguayo, L. Bellido, E. Pastor, D. De-Antonio-Monte, and A. G. Bolívar, “A virtualized platform for analyzing LTE broadcast services,” in *Proc. European Conference on Networks and Communications (EuCNC)*, Jun. 2015, pp. 512–516.
- [LB18] C. M. Lentisco and L. Bellido, “Multimedia streaming over mobile network based on virtualization,” in *II edición del simposio: cuéntanos tu tesis*, Universidad Politécnica de Madrid, Mar. 2018.
- [LBC⁺17] C. M. Lentisco, L. Bellido, J. Cuellar, E. Pastor, and J. Arciniegas, “QoE-Based Analysis of DASH Streaming Parameters Over Mobile Broadcast Networks,” *IEEE Access*, vol. 5, pp. 20 684–20 694, 2017.
- [LBF⁺14] C. M. Lentisco, L. Bellido, A. Fuente, E. Pastor, R. P. Leal, and A. G. Armada, “A model to evaluate MBSFN and AL-FEC techniques in a multicast video streaming service,” in *Proc. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2014, pp. 691–696.
- [LBF⁺17] C. M. Lentisco, L. Bellido, R. Flores, D. Fernández, and E. Pastor, “Leveraging Edge Computing for Broadcast as a Service in 5G Networks,” in *International Workshop on Elastic Networks Design and Optimization*, Vilanova, Spain, Jun. 2017.

- [LBF⁺18] C. M. Lentisco, L. Bellido, D. Fernández, R. Flores, and E. Pastor, “Virtual Multimedia Broadcast Multicast Service Architecture,” in *International Workshop on Elastic Networks Design and Optimization*, Valencia, Spain, Jun. 2018.
- [LBK15] K. Lazic, I. Basicovic, and J. Kovacevic, “Bandwidth estimation in adaptive video streaming over HTTP,” in *Proc. IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2015, pp. 62–65.
- [LBP15] C. M. Lentisco, L. Bellido, and E. Pastor, “Application level mechanisms for service continuity in an LTE multimedia broadcast service,” in *Proc. International Symposium on Consumer Electronics (ISCE)*, Jun. 2015, pp. 1–2.
- [LBP17a] C. Lentisco, L. Bellido, and E. Pastor, “Seamless Mobile Multimedia Broadcasting Using Adaptive Error Recovery,” *Mobile Information Systems*, vol. 2017, pp. 1–11, Feb. 2017, <https://doi.org/10.1155/2017>
- [LBP17b] C. M. Lentisco, L. Bellido, and E. Pastor, “Reducing Latency for Multimedia Broadcast Services Over Mobile Networks,” *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 173–182, Jan. 2017.
- [LBRP⁺17] C. M. Lentisco, L. Bellido, J. Ruiz-Piñar, R. Flores, E. Pastor, and D. Fernández, “Multimedia Streaming Service over 5G Mobile Broadcast Networks,” in *International Workshop on Elastic Networks Design and Optimization*, Castelldefels, Spain, Oct. 2017.
- [LEF⁺11] T. Lohmar, T. Einarsson, P. Fröjd, F. Gabin, and M. Kampmann, “Dynamic adaptive HTTP streaming of live content,” in *Proc. IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Jun. 2011, pp. 1–8.
- [LG12] D. Lecompte and F. Gabin, “Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements,” *IEEE Communications Magazine*, vol. 50, no. 11, pp. 68–74, Nov. 2012.
- [LGSW07] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, “Reliable Multimedia Download Delivery in Cellular Broadcast Networks,” *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 235–246, Mar. 2007.
- [LM11] B. Lewcio and S. Moller, “A testbed for QoE-based multimedia streaming optimization in heterogeneous wireless networks,” in *Proc. International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec. 2011, pp. 1–9.

- [LNN⁺18] H. T. Le, T. Nguyen, N. P. Ngoc, A. T. Pham, and T. C. Thang, “HTTP/2 Push-Based Low-Delay Live Streaming Over Mobile Networks With Stream Termination,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2423–2427, Sept. 2018.
- [LSWS07] M. Luby, A. Shokrollahi, M. Watson, and T. Stockhammer, *Raptor forward error correction scheme for object delivery*, IETF RFC 5053, Oct. 2007.
- [LSWS11] M. Luby, A. Shokrollahi, M. Watson, and T. Stockhamer, *RaptorQ Forward Error Correction Scheme for Object Delivery*, IETF RFC 6330, Aug. 2011.
- [Lub12] M. Luby, “Best practices for mobile broadcast delivery and playback of multimedia content (BMSB),” in *Proc. IEEE international Symposium on Broadband Multimedia Systems and Broadcasting*, Seoul, South Korea, Jun. 2012, pp. 1–7.
- [LWHL18] Y. Lin, C. Wang, C. Huang, and Y. Lai, “Hierarchical CORD for NFV Datacenters: Resource Allocation with Cost-Latency Tradeoff,” *IEEE Network*, vol. 32, no. 5, pp. 124–130, Sept. 2018.
- [LZG⁺14] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, “Probe and Adapt: Rate Adaptation for HTTP Video Streaming At Scale,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 719–733, Apr. 2014.
- [MAB⁺08] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “OpenFlow: Enabling Innovation in Campus Networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, Apr. 2008.
- [MB17] P. Mach and Z. Becvar, “Mobile Edge Computing: A Survey on Architecture and Computation Offloading,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.
- [MCC17] Y. Ma, Y. Chen, and J. Chen, “SDN-enabled network virtualization for industry 4.0 based on IoTs and cloud computing,” in *Proc. International Conference on Advanced Communication Technology (ICACT)*, Feb. 2017, pp. 199–202.
- [MCFG12] J. F. Monserrat, J. Calabuig, A. Fernandez, and D. Gomez, “Joint Delivery of Unicast and E-MBMS Services in LTE Networks,” *IEEE Transactions on Broadcasting*, vol. 58, no. 2, pp. 157–167, Jun. 2012.
- [MCO] “M-CORD as an Open Reference Solution for 5G Enablement,” CORD project, White Paper, Accessed: Mar. 2019 [Online]. Available: <http://opencord.org/wp-content/uploads/2017/02/MWC-M-CORD-WP-Short-Final.pdf>.

- [MHT15] A. W. Manggala, Hendrawan, and A. Tanwidjaja, “Performance analysis of white box switch on Software Defined Networking using open vSwitch,” in *Proc. International Conference on Wireless and Telematics (ICWT)*, Nov. 2015, pp. 1–7.
- [MMJ⁺18] X. Masip, E. Marín, A. Jurnet, P. Reviriego, R. Flores, C. M. Lentisco, D. Fernández, L. Bellido, S. Gómez, E. Calle, G. Otero, J. A. Hernández, and D. Larrabeiti, “Elastic Networks Thematic Network Results II: Edge Computing and Wireless Support,” in *Proc. International Conference on Transparent Optical Networks (ICTON)*, Jul. 2018, pp. 1–4.
- [MZZ13] D. Munaretto, M. Zanforlin, and M. Zorzi, “Performance evaluation in ns-3 of a video delivery framework for next generation cellular networks,” in *Proc. IEEE International Conference on Communications Workshops (ICC)*, Jun. 2013, pp. 642–646.
- [NBGT17] V. Nguyen, A. Brunstrom, K. Grinnemo, and J. Taheri, “SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1567–1602, 2017.
- [NMN⁺14] B. A. A. Nunes, M. Mendonca, X. N. Nguyen, K. Obraczka, and T. Turletti, “A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.
- [OS12] O. Oyman and S. Singh, “Quality of experience for HTTP adaptive streaming services,” *IEEE Communications Magazine*, vol. 50, no. 4, pp. 20–27, Apr. 2012.
- [PCWT16] W. Pan, G. Cheng, H. Wu, and Y. Tang, “Towards QoE assessment of encrypted YouTube adaptive video streaming in mobile networks,” in *Proc. IEEE/ACM International Symposium on Quality of Service (IWQoS)*, Jun. 2016, pp. 1–6.
- [PFK⁺17] S. Peng, J. O. Fajardo, P. S. Khodashenas, B. Blanco, F. Liberal, C. Ruiz, C. Turayagenda, M. Wilson, and S. Vadgama, “QoE-Oriented Mobile Edge Service Management Leveraging SDN and NFV,” *Mobile Information Systems*, vol. 2017, no. 3961689, pp. 1–14, Jun. 2017, <https://doi.org/10.1155/2017/3961689>
- [PRP⁺] S. Peltotalo, V. Roca, L. Poderico, N. Lamb, and G. Mendola, “Mad project’s home page,” Accessed: Jan. 2019 [Online]. Available: <http://mad.cs.tut.fi/>.
- [PSA⁺16] L. Peterson, A. Shabibi, T. Anshutz, S. Baker, A. Bavier, S. Das, J. Hart, G. Palukar, and W. Snow, “Central office re-architected as a data center,” *IEEE Communications Magazine*, vol. 54, no. 10, pp. 96–101, Oct. 2016.

- [PW04] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sept. 2004.
- [PWL⁺12] T. Paila, R. Walsh, M. Luby, V. Roca, and R. Lehtonen, *FLUTE - File Delivery over Unidirectional Transport*, IETF RFC 6726, Nov. 2012.
- [Ram16] E. Ramírze, “Desarrollo y aplicación de herramientas para el análisis del envío anticipado de video mediante HTTP/2 en servicios de streaming multimedia,” in *Trabajo Fin de Grado*, Jul. 2016, Tutor: C. M. Lentisco, Ponente: L. Bellido
- [RBB⁺16] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, “Mobile network architecture evolution toward 5G,” *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, May 2016.
- [RBD⁺14] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, “Cloud technologies for flexible 5G radio access networks,” *IEEE Communications Magazine*, vol. 52, no. 5, pp. 68–76, May 2014.
- [RDS15] A. Roy, P. De, and N. Saxena, “Location-based social video sharing over next generation cellular networks,” *IEEE Communications Magazine*, vol. 53, no. 10, pp. 136–143, Oct. 2015.
- [RRA⁺16] D. Z. Rodríguez, R. L. Rosa, E. C. Alfaia, J. I. Abrahão, and G. Bressan, “Video Quality Metric for Streaming Service Using DASH Standard,” *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 628–639, Sept. 2016.
- [Rue13] D. Rueda, “Análisis de las alternativas de funcionamiento de la red LTE para la entrega de servicios de IPTV con QoS,” *Avances: Investigación en Ingeniería*, vol. 10, no. 2, pp. 35–48, Dic. 2013.
- [SES⁺15] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, “A Survey on Quality of Experience of HTTP Adaptive Streaming,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.
- [SQU] SQUID: squid-cache.org, “Squid: Optimising Web Delivery,” Accessed: Jan. 2019 [Online]. Available: <http://www.squid-cache.org/Versions/>.
- [Sta15] W. Stallings, “NFV Functionality,” in *Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud*, Strand, London, UK: Pearson, 2015, ch. 8.
- [SUS16] K. Spiteri, R. Urgaonkar, and R. K. Sitaraman, “BOLA: Near-optimal bitrate adaptation for online videos,” in *Proc. IEEE International Conference on Computer (INFOCOM) Communications*, Apr. 2016, pp. 1–9.

- [TC12] E. Tan and C. T. Chou, “A Frame Rate Optimization Framework for Improving Continuity in Video Streaming,” *IEEE Transactions on Multimedia*, no. 3, pp. 910–922, Jun. 2012.
- [THB08] A. Takahashi, D. Hands, and V. Barriac, “Standardization activities in the ITU for a QoE assessment of IPTV,” *IEEE Communications Magazine*, vol. 46, no. 2, pp. 78–84, Feb. 2008.
- [TKCJ16] T. Taleb, A. Ksentini, M. Chen, and R. Jantti, “Coping With Emerging Mobile Social Media Applications Through Dynamic Service Function Chaining,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2859–2871, Apr. 2016.
- [TLW⁺18] B. Tan, J. Lu, J. Wu, D. Zhang, and Z. Zhang, “Toward a Network Slice Design for Ultra High Definition Video Broadcasting in 5G,” *IEEE Wireless Communications*, vol. 25, no. 4, pp. 88–94, Aug. 2018.
- [VIP15] E. H. Valencia, S. Izzo, and B. Polonsky, “How will NFV/SDN transform service provider opex?” *IEEE Network*, vol. 29, no. 3, pp. 60–67, May 2015.
- [VLC] VLC: VideoLan Organization, “VLC media player,” Accessed: Jan. 2019 [Online]. Available: <http://www.videolan.org/vlc/>.
- [WLV07] M. Watson, M. Luby, and L. Vicisano, *Forward Error Correction (FEC) Building Block*, IETF RFC 5052, Aug. 2007.
- [WS14] S. Wei and V. Swaminathan, “Low Latency Live Video Streaming over HTTP 2.0,” in *Proc. Network and Operating System Support on Digital Audio and Video Workshop (NOSSDAV)*, Singapore, Singapore, Mar. 2014, pp. 37:37–37:42.
- [WSBL03a] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [WSBL03b] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [WSM⁺16] G. K. Walker, T. Stockhammer, G. Mandyam, Y. Wang, and C. Lo, “ROUTE/-DASH IP Streaming-Based System for Delivery of Broadcast, Broadband, and Hybrid Services,” *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 328–337, Mar. 2016.

- [WZHW15] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, “Cloud radio access network (C-RAN): a primer,” *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan. 2015.
- [XZG18] Z. Xu, X. Zhang, and Z. Guo, “QoE-driven Adaptive K-Push for HTTP/2 Live Streaming,” *IEEE Transactions on Circuits and Systems for Video Technology*, Jun. 2018, Early Access.
- [YCH14] D. Yun, K. Chung, and J. Hong, “Efficient bandwidth estimation for http adaptive streaming,” in *Proc. International Conference on Information Networking (ICOIN)*, Feb. 2014, pp. 464–468.
- [YL15] C. Yang and Y. Liu, “On Achieving Short Channel Switching Delay and Playback Lag in IP-Based TV Systems,” *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 1096–1106, Jul. 2015.
- [YLML10] H. Yin, X. Liu, G. Min, and C. Lin, “Content delivery networks: a bridge between emerging applications and future IP networks,” *IEEE Network*, vol. 24, no. 4, pp. 52–56, Jul. 2010.
- [YLNS17] M. B. Yahia, Y. L. Louedec, L. Nuaymi, and G. Simon, “When HTTP/2 rescues DASH: Video frame multiplexing,” in *Proc. IEEE Conference on Computer Communications Workshops (INFOCOM)*, May 2017, pp. 677–682.
- [ZDL⁺14] Q. Zheng, H. Du, J. Li, W. Zhang, and Q. Li, “Open-LTE: An Open LTE simulator for mobile video streaming,” in *Proc. IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, Jul. 2014, pp. 1–2.