

TESIS DOCTORAL

2016

**ANÁLISIS DE PRESTACIONES PARA SERVICIOS
MULTIMEDIA MASIVOS Y SU APLICACIÓN A
SISTEMAS DE DISTRIBUCIÓN DE VÍDEO
ADAPTATIVOS**

ALBERTO ÁLVAREZ GONZÁLEZ

INGENIERO DE TELECOMUNICACIÓN

**PROGRAMA DE DOCTORADO EN TECNOLOGÍAS
INDUSTRIALES**

GABRIEL DÍAZ ORUETA

ROBERTO GARCÍA FERNANDEZ

TESIS DOCTORAL

2016

**ANÁLISIS DE PRESTACIONES PARA SERVICIOS MULTIMEDIA
MASIVOS Y SU APLICACIÓN A SISTEMAS DE DISTRIBUCIÓN DE
VÍDEO ADAPTATIVOS**

ALBERTO ÁLVAREZ GONZÁLEZ

INGENIERO DE TELECOMUNICACIÓN

**PROGRAMA DE DOCTORADO EN TECNOLOGÍAS
INDUSTRIALES**

GABRIEL DÍAZ ORUETA

ROBERTO GARCÍA FERNANDEZ

Agradecimientos

A mis directores de Tesis, Gabriel y Roberto, por su sabio consejo, su paciencia y dedicación. Sin su tutela hubiese sido impensable llegar hasta aquí.

A todo el grupo de investigación DMMS, que ha contribuido a hacer crecer este trabajo, en especial a Xabiel, por haberme dado la oportunidad de trabajar en él todos estos años y a Laura por ser un apoyo imprescindible dentro del él.

A mi familia, por su apoyo incondicional.

Resumen

En los últimos años, la proliferación de servicios de vídeo en Internet, la expansión del mercado a un público masivo y un aumento notable de las calidades de los flujos multimedia involucrados son hechos constatados. La tecnología avanza más rápido que el mercado y con frecuencia surgen nuevos paradigmas en el diseño y la ejecución de los servicios de vídeo. Las compañías realizan grandes esfuerzos por adaptarse a la demanda creciente e incorporar las tecnologías emergentes, dedicando tiempo y esfuerzo a evaluar, y mejorar, las prestaciones de sus servicios.

La popularidad de los servicios de vídeo adaptativos es sólo un ejemplo de las posibilidades que se abren con las mejoras en la tecnología, en este caso de codificadores escalables. Se habilitan los diseños de nuevos servicios basados en estos paradigmas escalables y se abre un nuevo ciclo de vida para éstos.

Pero la evaluación de las prestaciones de los servicios es un concepto muy amplio que esconde varios retos. La primera contrariedad de la evaluación de los servicios multimedia en situaciones heterogéneas es la enorme dificultad de disponer de los escenarios adecuados para la reproducción de las condiciones deseadas. En muchos casos, la evaluación de servicios debe realizarse de distintas formas según la fase de su desarrollo. Entre la simulación inicial y el despliegue real del servicio el proceso de evaluación pasa por distintas fases que implican el conocimiento y la aplicación de técnicas diferentes.

Además, las medidas realizadas sobre el desempeño del servicio deben realizarse en base a algún criterio objetivo, acreditado y comparable. En este sentido las métricas de calidad de vídeo han surgido de forma natural evolucionando desde los mismos modelos que se venían aplicando a las imágenes. Otras técnicas más avanzadas han ido incorporando las características intrínsecas del contenido de vídeo, más allá de la pura definición de secuencias de imágenes, para reflejar con un criterio más adecuado la calidad del servicio multimedia.

Como paso natural a la evaluación del servicio multimedia, la opinión del usuario ha sido incorporada como una métrica altamente importante de cara a considerar todos los factores determinantes en el servicio extremo a extremo, o *end-to-end*.

En esta tesis se hacen una serie de contribuciones en los tres ejes principales descritos en este resumen. Para la evaluación de servicios se han explorado técnicas de simulación, en particular centrándose en simulación en tiempo real y se ha propuesto una

metodología para la construcción de entornos de experimentación enfocada en servicios masivos.

Para la evaluación de los servicios en las primeras etapas se recurre con frecuencia a las métricas de calidad objetivas. En este trabajo se han investigado y desarrollado herramientas que permiten la medición de estas métricas en el ámbito de los flujos multimedia escalables y adaptativos.

Finalmente se han realizado dos estudios de campo con la participación de más de 170 usuarios en total, evaluando distintos aspectos de la calidad percibida de los sistemas de multimedia en general y vídeos adaptativos en particular. El trabajo original concluye las aportaciones realizando una serie de propuestas para la ejecución de las fases de diseño y evaluación de los sistemas multimedia adaptativos.

Esta tesis doctoral se enmarca en una línea de investigación sobre sistemas multimedia abierta en el grupo de investigación DMMS de la Universidad de Oviedo, en colaboración según el acuerdo marco con el grupo del Departamento de Ingeniería Eléctrica Electrónica y Control de la UNED.

Abstract

In recent years, the proliferation of Internet video services, expanding the market to a mass audience and a significant increase in the quality of multimedia streams involved are established facts. Technology advances faster than the market and new paradigms often arise in the design and implementation of video services. Companies make great efforts to adapt to the growing demand and incorporate emerging technologies, devoting time and effort to evaluate and improve the performance of their services.

The popularity of adaptive video services is just one example of the possibilities open thanks to improvements made in technology, in this case of scalable encoders. New service designs based on scalable paradigms are enabled and a new life cycle for these services opens.

But the evaluation of the performance of services is a very broad concept that hides several challenges. The first setback of the evaluation of multimedia services in heterogeneous situations is the enormous difficulty of having the appropriate settings for playing the desired conditions. In many cases, the evaluation of services must be performed in different ways depending on the stage of development. Between the initial simulation and actual deployment of the service, the assessment process goes through various stages that involve knowledge and application of different techniques.

In addition, measurements on service performance should be made based on some objective criteria, accredited and comparable. In this sense, the video quality metrics have emerged, naturally evolved from the same models that had applied to images. More advanced techniques have been incorporating the intrinsic characteristics of video content, beyond the pure definition image sequences to reflect a more appropriate criterion quality multimedia service.

As a natural step to assessing the multimedia service, user opinion has been incorporated as a highly important metric face to consider all the factors in the service end-to-end or end-to-end.

In this thesis a series of contributions are made in three main areas described in this summary. For the evaluation of services simulation techniques, including focusing on real-time simulation have been explored. A methodology for building experimental environments focused on mass services has been proposed.

For the evaluation of services in the early stages are used frequently objective quality metrics. In this paper we have researched and developed tools for measuring these metrics in the field of adaptive and scalable multimedia streams.

Finally there have been two field studies involving more than 170 users in total, evaluating different aspects of perceived quality of multimedia systems in general and adaptive

videos in particular. The original work concludes making a series of proposals for the implementation of design and evaluation phases of adaptive multimedia systems.

This thesis is part of a research on multimedia systems open with DMMS research group at the University of Oviedo, in cooperation with the group of the Department of Electronics and Electrical Engineering Control at UNED.

Índice general

1	Introducción y objetivos	1
1.1	Introducción.....	1
1.2	Marco conceptual.....	4
1.3	Objetivos y metodología de investigación.....	5
1.4	Estructura del contenido.....	8
2	Estado del arte	11
2.1	Evaluación de servicios: simulación en tiempo real	11
2.1.1	Herramientas de simulación en tiempo real	15
2.2	Métricas de calidad	15
2.2.1	Métricas de QoS.....	17
2.2.2	Métricas de QoE objetivas	18
2.2.2.1	Métricas de QoE objetivas en SVC	24
2.2.3	Métodos de QoE subjetivos	25
2.2.3.1	Evaluaciones subjetivas de servicios multimedia.	27
3	Herramientas, técnicas y metodología de emulación de servicios	31
3.1	Introducción.....	31
3.2	NS-3 en la evaluación de servicios masivos. Extensión distribuida	35
3.2.1	Extensión de NS-3 para clientes distribuidos	36
3.2.2	Evaluación: Caso de estudio distribuido frente a centralizado.....	42
3.2.2.1	Configuración de los experimentos	45
3.2.2.2	Discusión de resultados	47
3.3	Límites de NS-3 y OPNET para la simulación en tiempo real.....	54
3.3.1	Red de acceso NS-3.....	57
3.3.2	Backbone OPNET	60
3.3.3	Escenario conjunto	63
3.4	Metodología para la construcción de escenarios de simulación	67
3.4.1	Diseño de la metodología	67
3.4.1.1	Fase 1, Especificación del SEUT	70
3.4.1.2	Fase 2, Caracterización de los requisitos del testbed.....	70
3.4.1.3	Fase 3, Asignación de recursos	73
3.4.1.4	Fase 4, Análisis del experimento.....	74
3.4.2	Aplicación de la metodología	75
3.5	Contribuciones del capítulo	77
4	La calidad de experiencia objetiva en los sistemas adaptativos	79
4.1	Introducción.....	79
4.2	Métrica cPSNR. PSNR de transmisión en SVC.....	81
4.2.1	Validación de la métrica cPSNR	83
4.3	El framework para estimar QoE en SVC	88
4.3.1	El sistema adaptativo.....	90
4.4	El camino hacia el decodificador adaptativo	90
4.4.1	Trasfondo	90
4.4.2	La solución diseñada	92
4.4.3	Caso de estudio.....	94
4.4.4	Método de estimación de calidad de vídeo FR	97
4.4.4.1	Resultados	99

4.5	Contribuciones del capítulo.....	103
5	Calidad de experiencia subjetiva.....	105
5.1	Experimentos subjetivos end-to-end.....	106
5.1.1	Metodología de los experimentos.....	107
5.1.2	Entorno y equipamiento.....	108
5.1.3	Contenidos.....	110
5.1.4	Población.....	113
5.1.5	Análisis de los resultados.....	113
5.2	Experimentos subjetivos en sistemas adaptativos.....	128
5.2.1	Retos de los sistemas adaptativos.....	130
5.2.2	Metodología de los experimentos.....	134
5.2.3	Entorno y equipamiento de el experimento subjetivo.....	136
5.2.4	Contenidos del experimento subjetivo.....	137
5.2.5	Población.....	139
5.2.6	Análisis de los resultados.....	140
5.2.6.1	Objetivo I.....	142
5.2.6.2	Objetivo II.....	144
5.2.6.3	Objetivo III.....	146
5.2.6.4	Objetivo IV.....	148
5.2.6.5	Objetivo V.....	151
5.2.7	Correlación con métricas objetivas.....	152
5.3	Contribuciones del capítulo.....	157
6	Conclusiones.....	159
7	Trabajo futuro.....	163
8	Publicaciones derivadas de la Tesis.....	167
	Bibliografía.....	173
	Glosario de términos.....	185
	ANEXO.....	189
	Curriculum Vitae.....	189

Lista de figuras

Figura 1-1: Ciclo PDCA y ámbito de aplicación de la investigación	2
Figura 2-1: Diagrama básico del servicio multimedia.....	11
Figura 2-2: Entornos de evaluación de servicios, simulación vs simulación en tiempo real.	14
Figura 2-3: Calidad de Servicio vs Calidad de Experiencia.....	16
Figura 2-4: Clasificación de métricas de QoE.....	17
Figura 2-5: Clasificación de métricas de QoE según la cantidad de información de referencia	20
Figura 2-6: Algoritmo de cálculo de la métrica MOVIE, fuente [52].....	23
Figura 2-7: Instalaciones en el grupo Multimedia Signal Processing Group (Lausanne, Suiza), colaborando en los experimentos de estandarización internacional (Call for Proposal for High Efficiency Video Coding (HEVC))	26
Figura 3-1: Dimensiones del universo de herramientas de experimentación	33
Figura 3-2: Esquema de simulación en tiempo real con NS-3	35
Figura 3-3: Evolución distribuida del modelo de emulación de NS-3	37
Figura 3-4: Detalle de la configuración de la extensión distribuida en NS-3.....	39
Figura 3-5: El problema del medio compartido en la extensión distribuida.....	42
Figura 3-6: Configuración del experimento "single-machine"	44
Figura 3-7: Configuración del experimento "distributed-machine"	44
Figura 3-8: Desglose de experimentos	45
Figura 3-9: Consumo de CPU en una realización temporal de los experimentos con NS-3 (20 clientes, <i>High</i>)	48
Figura 3-10: Caudal en una realización temporal de los experimentos con NS-3 (20 clientes, <i>High</i>)	49
Figura 3-11: Evolución del consumo global de CPU en los experimentos con NS-3.....	50
Figura 3-12: Evolución del caudal en los experimentos con NS-3	51
Figura 3-13: Evolución de consumo del equipo en los experimentos con NS-3.....	52
Figura 3-14: Evolución del <i>jitter</i> en los experimentos con NS-3.....	53
Figura 3-15: Esquema de simulación en tiempo real con OPNET	55
Figura 3-16: Topologías del caso de estudio combinado.....	57
Figura 3-17: <i>jitter</i> en NS-3 simulado en tiempo real versus emulado	58
Figura 3-18: PLR en NS-3 simulado en tiempo real versus emulado.....	59
Figura 3-19: MOS en NS-3 simulado en tiempo real	60
Figura 3-20: PLR en OPNET simulado en tiempo real versus emulado.....	61
Figura 3-21: <i>jitter</i> en OPNET simulado versus emulado	62
Figura 3-22: MOS en OPNET simulado versus emulado.....	63
Figura 3-23: Escenario compuesto de simulación en tiempo real sobre arquitectura del <i>testbed</i>	64
Figura 3-24: Métricas de QoE medidas en el servicio simulado en tiempo real	65
Figura 3-25: Métricas de rendimiento en la simulación en tiempo real, CPU	66

Figura 3-26: Métricas de rendimiento en la simulación en tiempo real, Eventos por segundo.....	66
Figura 3-27: Fases de la metodología de construcción de entornos de evaluación.....	68
Figura 3-28: Flujograma de la metodología de construcción de entornos de evaluación.....	69
Figura 3-29: Secuencia de acciones de la fase de análisis	75
Figura 3-30: Fotogramas de la secuencia real de un instante en la evaluación del servicio masivo mediante la aplicación de la metodología.	76
Figura 4-1: Tipos de análisis en la transmisión de vídeo	80
Figura 4-2: Correlación entre métricas objetivas de transmisión y caída de calidad (MOS_drop).....	84
Figura 4-3: Métricas de distorsión de transmisión por contenido según la tasa de pérdidas, con la métrica <i>end-to-end</i> como referencia.....	85
Figura 4-4: Modos de sincronización disponibles en DMMSAF	87
Figura 4-5: Comparativa del rendimiento de DMMSAF frente a Evalvid	87
Figura 4-6: Esquema de cancelación de errores en el decodificador adaptativo	93
Figura 4-7: Caso de estudio para análisis de VQ en el sistema adaptativo	94
Figura 4-8: Escenario 1 (ramp-up), fuente [131].....	96
Figura 4-9: Escenario 2 (ramp-down) , fuente [131]	96
Figura 4-10: Escenario 3, fuente [131]	96
Figura 4-11: Comparativa de métricas normalizadas en dos escenarios.....	101
Figura 4-12: Representación temporal de VQ reales y estimadas.....	102
Figura 5-1: Aspecto de la aplicación desarrollada para los experimentos subjetivos <i>end-to-end</i>	109
Figura 5-2: Instantáneas de los contenidos del experimento subjetivo.....	110
Figura 5-3: Patrón rebuffering largo [IM-SL].....	111
Figura 5-4: Patrón rebuffering ráfaga [IM-SB]	111
Figura 5-5: Pérdidas audio largas [IM-CAL]	112
Figura 5-6: Pérdidas video largas [IM-CVL].....	112
Figura 5-7: Pérdidas audio ráfaga [IM-CAB].....	112
Figura 5-8: Pérdidas vídeo ráfaga [IM-CVB].....	112
Figura 5-9: Asincronía vídeo/audio en el inicio [IM-AS].....	112
Figura 5-10: Asincronía vídeo/audio en el final [IM-AE].....	112
Figura 5-11: Baja calidad de imagen en el inicio [IM-PS]	113
Figura 5-12: Baja calidad de imagen en el final [IM-PE]	113
Figura 5-13: Resultados MOS según el tipo de distorsión.....	114
Figura 5-14: Resultados MOS según el tipo de contenido.....	116
Figura 5-15: Resultados MOS para IM-S según el tipo de contenido	117
Figura 5-16: Resultados MOS para IM-CA según el tipo de contenido.....	119
Figura 5-17: Resultados Test <i>Tukey</i> para IM-CAL	120
Figura 5-18: Resultados MOS para patrón de pérdidas en el vídeo.	122
Figura 5-19: Resultados MOS para patrón de asincronía en audio-vídeo	123
Figura 5-20: Resultados MOS para patrón de pixelación en vídeo.....	125
Figura 5-21: Test <i>Tukey</i> para grupos IM-PE	126

Figura 5-22: Resultados subjetivos para distintos tipos de contenido: (a) <i>cartoons</i> , (b) <i>action</i> , (c) <i>sports</i> , (d) <i>news</i>	127
Figura 5-23: Esquema de evaluación SCACJ con categorías modificadas	135
Figura 5-24: Diagrama del procedimiento del experimento subjetivo	136
Figura 5-25: Interfaz del software desarrollada para el experimento subjetivo	137
Figura 5-26: Fotograma de las secuencias del experimento subjetivo	138
Figura 5-27: Resultados CMOS agrupados por objetivos	141
Figura 5-28: Resultados CMOS agrupados por objetivos y edad	141
Figura 5-29: Histogramas CMOS para objetivo I	143
Figura 5-30: Histogramas CMOS para objetivo II	145
Figura 5-31: Test de <i>Tukey HSD</i> para diferencias CMOS con PLR 3%	146
Figura 5-32: Histogramas CMOS para objetivo III	147
Figura 5-33: Histogramas CMOS para objetivo IV	149
Figura 5-34: CMOS según experiencia con servicios multimedia	150
Figura 5-35: Histogramas CMOS para objetivo V	151
Figura 5-36: Resumen gráfico de la regresión para la métrica PSNR	154
Figura 5-37: Resumen gráfico de la regresión para la métrica SSIM	154
Figura 5-38: Resumen gráfico de la regresión para la métrica VQM	155
Figura 5-39: Correlación de CMOS con métricas SSIM Y VQM para el objetivo II	157

Lista de tablas

Tabla 2—1 Correspondencia PSNR - MOS.....	21
Tabla 3—1 Características de la plataforma de evaluación	45
Tabla 3—2 Ejemplo del error al simular escenarios grandes en NS-3	47
Tabla 4—1 Coeficientes de correlación Pearson para cPSNR, PSNR_100dB y SSIM con MOS_drop.....	84
Tabla 4—2 Coeficientes de correlación Pearson para cPSNR, PSNR_100dB con PLR .	85
Tabla 4—3 Características de la secuencia city.yuv codificada con H.264-SVC.....	95
Tabla 4—4 Métricas VQ de las capas de la secuencia original.....	97
Tabla 4—5 Ecuaciones del estimador de VQ para las distintas métricas.....	98
Tabla 4—6 VQ promedio para el escenario 1	99
Tabla 4—7 VQ promedio para el escenario 2.....	99
Tabla 4—8 VQ promedio para el escenario 3.....	99
Tabla 5—1 Resultados de la encuesta final de los experimentos subjetivos <i>end-to-end</i>	108
Tabla 5—2 Detalles del equipo usado en los experimentos	109
Tabla 5—3 Detalles del monitor usado en los experimentos	109
Tabla 5—4 Descripción de los contenidos del experimento subjetivo.....	111
Tabla 5—5 Parámetros estadísticos para los distintos patrones de distorsión.....	115
Tabla 5—6 Parámetros estadísticos para distintos tipos de contenido	116
Tabla 5—7 Parámetros estadísticos para distorsión IM-S	117
Tabla 5—8 Parámetros estadísticos para distorsión IM-CA.....	120
Tabla 5—9 Parámetros estadísticos para distorsión IM-CV.....	121
Tabla 5—10 Parámetros estadísticos para distorsión IM-A	124
Tabla 5—11 Parámetros estadísticos para distorsión IM-P	126
Tabla 5—12 Distribución de objetivos y factores clave de éxito en el sistema adaptativo	132
Tabla 5—13 Resultados de la encuesta final del experimento subjetivo.....	136
Tabla 5—14 Características de las secuencias originales en el experimento subjetivo .	138
Tabla 5—15 Resultados t-test para usuarios con experiencia multimedia limitada a televisión	150
Tabla 5—16 Resultados t-test para distintos saltos de calidad.....	150
Tabla 5—17 Estadístico R-cuadrado para los modelos de regresión de las métricas objetivas.....	155
Tabla 5—18 Resultados de la correlación lineal entre logaritmo de métricas objetivas y resultados subjetivos por objetivo.....	156

Tesis Doctoral

1 Introducción y objetivos

1.1 Introducción

La mayoría de los trabajos de investigación en el ámbito de las tecnologías de distribución de vídeo en Internet, éste incluido, comienzan relatando la enorme popularidad de los servicios de distribución multimedia y remarcando la importancia de cualesquiera soluciones propuestas. En realidad, echando la vista atrás, los avances han sido verdaderamente notorios. Desde los inicios de la transmisión de flujos de contenidos multimedia a través de las redes de paquetes, originalmente concebidas para la transmisión de datos discretos, ha habido una serie de importantes contribuciones que han conseguido mejorar con mucho todos los elementos involucrados en esta distribución.

Una vez alcanzado un nivel importante de popularización, los servicios de vídeo en Internet han perseguido aumentar la calidad que ofrecen a sus clientes. Si bien en un principio la filosofía era presentar las imágenes en ventanas de tamaño reducido, la tendencia actual es que éstas se incrementen junto con el número de fotogramas o *frames* por segundo, que ya es comparable al de la televisión convencional. Sin embargo, en este camino a la mejora de la calidad todavía hay diversos escollos que vencer. El aumento del ancho de banda requerido para transmitir estos vídeos de mayor calidad provoca en muchos casos que la reproducción de los mismos se corte para recargar los *buffers* en los que los clientes almacenan los *frames* previamente a ser presentados al usuario. En general esta situación es provocada porque la red carece en determinados momentos de capacidad para lograr una transmisión fluida. Los servicios multimedia en Internet comúnmente no cuentan con sistemas que le permitan garantizar la calidad de servicio por lo que su dependencia de la capacidad de las redes por las que atraviesa la información es muy importante.

Con el objetivo de solucionar los problemas de calidad derivados del aumento de la resolución de los vídeos y del incremento de usuarios, se ha trabajado en los últimos años en diversos elementos que podrían hacer que los servicios multimedia en Internet sean más flexibles ante problemas puntuales de capacidad en las redes de comunicaciones.

En el primer eslabón de la cadena de generación, distribución y consumo de contenidos multimedia, ha habido un gran avance en el desarrollo de codificadores de vídeo que ha permitido no sólo optimizar la cantidad de información que es necesario transmitir, sino también se ha visto mejorada la gestión de la fragmentación orientada a la recuperación ante los errores causados por una transmisión imperfecta sobre redes de paquetes. Al ritmo del mercado, los desarrollos en codificadores se van implantando en los sistemas de distribución de masas. En términos de versatilidad y eficiencia, los avances en la tecnología *Scalable Video Coding* (SVC) ofrecen a los creadores de sistemas multimedia un gran abanico de posibilidades [1]. SVC es el nombre para la extensión escalable del codificador H.264/MPEG-4 *Advanced Video Coding* (AVC).

Las tecnologías subyacentes para el transporte del contenido multimedia se han desarrollado especialmente enfocadas en el ámbito multimedia como *Real Time Streaming Protocol* (RTSP) o han evolucionado desde otros usos, adaptándose a los nuevos paradigmas que imponen los contenidos multimedia MPEG-DASH (*Dynamic Adaptive Streaming over HTTP*) [2] estandarizada a finales de 2011.

Durante los años de trabajo con grupo de investigación de Sistemas de Distribución Multimedia (DMMS), hemos tenido la oportunidad de analizar varias de las tecnologías involucradas en la transmisión de contenidos multimedia. Se han analizado tecnologías, se han construido soluciones y se han realizado contribuciones al estado del arte de la materia en forma de publicaciones en revistas y congresos.

Aplicando la estrategia del ciclo de Deming, o ciclo PDCA [3], al proceso de mejora continua en el que está inmerso, por su propia naturaleza, el crecimiento de los sistemas de distribución multimedia, podemos entender la necesidad de establecer mecanismos sólidos para la evaluación y la medición de los sistemas y servicios desarrollados. En este sentido los trabajos de esta Tesis suponen un aporte claramente focalizado en la fase de verificación o evaluación del ciclo de Deming, Figura 1-1

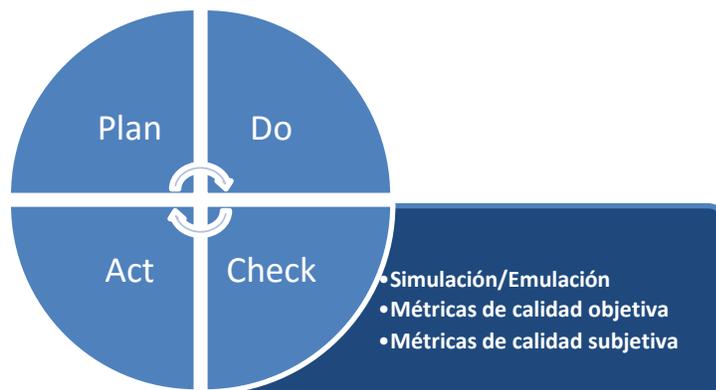


Figura 1-1: Ciclo PDCA y ámbito de aplicación de la investigación

Sea cual fuere el objetivo de estudio, siempre es necesario disponer de las herramientas necesarias para una correcta experimentación. Además es necesario conocer cómo aplicar las distintas soluciones propuestas en la literatura a cada caso de estudio. En este ámbito se han desarrollado trabajos de investigación de tecnologías de simulación. El conocimiento de las herramientas de simulación ha conducido a profundizar en otros conceptos como simulación en tiempo real y emulación.

Los conceptos de simulación, simulación en tiempo real y emulación han sido estudiados ampliamente durante esta investigación. La primera aproximación para la evaluación de los servicios multimedia estudio es la simulación completa del entorno bajo estudio. Mediante diseños existentes o desarrollos específicos todos los componentes del servicio son representados en los términos que exija la tecnología de simulación. Esta fase supone una alternativa altamente escalable, sin embargo, su precisión depende de la eficacia del modelado de sus componentes y puede que algunas partes tengan que ser implementadas de nuevo una vez fuera del modelo de simulación.

La necesidad de incorporar realismo a las simulaciones conduce a la exploración de las técnicas de simulación en tiempo real. Gracias a los avances realizados en los simuladores convencionales se habilita la inclusión de elementos reales en las simulaciones. Usuarios, servicios, sistemas bajo estudio o cualquier elemento real puede ser incorporado a la simulación a través de esta tecnología. Esta nueva técnica incorpora el tiempo como una restricción adicional a la simulación, reduciendo la escalabilidad. En este nuevo paradigma se centran parte de los trabajos realizados en esta investigación. Dos de los simuladores más populares, NS-3¹ como herramienta libre y comunitaria y OPNET Modeler (ahora parte de Riverbed y renombrado a Riverbed Modeler)² en la opción comercial, incorporan la tecnología necesaria para permitir la simulación añadiendo elementos reales en la topología simulada. En ambos casos, encontrar los límites por debajo de los cuales los resultados del experimento son confiables resulta de vital importancia. Al detectarse la carencia de sistematización en este campo, se formula una metodología para la construcción de escenarios escalables para la evaluación de servicios para la diversidad de situaciones que pueden encontrarse los investigadores a la hora de evaluar sus nuevos servicios. Las técnicas referenciadas en la metodología permiten aumentar el grado de escalabilidad de las simulaciones para llegar a servicios masivos.

De una forma natural, y sin abandonar la misma fase de la estrategia de mejora continua, el foco del trabajo pasa de la experimentación en sí, a la evaluación de los resultados. En el campo de la experiencia multimedia, las métricas tradicionales de calidad de servicio heredadas de los otros servicios han sido extendidas para incorporar progresivamente

¹ <https://www.nsnam.org/>

² <http://www.riverbed.com/products/steelcentral/opnet.html?redirect=opnet>

características exclusivas de los servicios multimedia. En esta parte se distinguen dos enfoques principales, por un lado las métricas objetivas de calidad de vídeo centradas en la medida de factores cuantificables del servicio y, por otro lado, la nueva realidad que incorpora aspectos subjetivos en la evaluación extremo a extremo (o *end-to-end*) de la experiencia multimedia, involucrando experimentos con usuarios reales. Pero la clasificación de métricas no es binaria y cada vez más trabajos buscan hallar la fórmula que relacione de forma definitiva las métricas objetivas con la experiencia final de usuario medida en términos subjetivos.

Las métricas objetivas de calidad de vídeo son la primera aproximación lógica surgida para la evaluación de los servicios multimedia. Suponen la adaptación de las métricas de calidad de servicio incorporando de distintas formas la información del contenido multimedia o las características especiales de estos flujos. En la realidad actual además, incorporar los servicios adaptativos basados en SVC es un beneficio conocido. Se deben adaptar las métricas a esta nueva situación. En esta Tesis se trabajo con las métricas objetivas más conocidas y se proporcionan los mecanismos para aplicar estas métricas a los estudios específicos realizados con codificadores adaptativos. Como resultado del estudio de las métricas objetivas se construyen herramientas y se formulan métricas derivadas que extienden la aplicación de las métricas objetivas a diversas situaciones, especialmente centradas en la particularidad de los sistemas escalables basados en SVC.

Las métricas subjetivas constituyen un intento por plasmar de forma cuantitativa la valoración que los usuarios hacen del sistema multimedia en su totalidad. En las evaluaciones subjetivas se incluyen todos los aspectos de la calidad percibida realmente por los usuarios. Es, en definitiva, la mejor valoración posible del sistema *end-to-end*. Sin embargo, la realización de test subjetivos requiere una inversión en tiempo y recursos que no siempre es posible. En el marco de esta Tesis se realizan dos experimentos subjetivos con alto grado de participación, para estudiar factores de calidad relevantes en sistemas multimedia en general, y en los servicios adaptativos en particular. Para aliviar la exigencia de realización de *tests* subjetivos, se busca encontrar una relación entre ambos tipos de métricas, objetivas y subjetivas, contrastando los resultados obtenidos en los experimentos subjetivos con las métricas objetivas analizadas para el caso especial de los servicios adaptativos.

1.2 Marco conceptual

Un servicio multimedia tiene el objetivo de proporcionar un contenido en una determinada forma a unos espectadores conectados al servicio a través de redes de telecomunicaciones. En los foros más técnicos, es frecuente encontrar visiones simplificadas de los servicios multimedia que no llegan a representar todos los agentes involucrados. El concepto de servicio multimedia va mucho más allá de la tecnología que lo habilita. En un servicio multimedia existen varios factores que dan forma a la industria multimedia y de cuya coordinación depende el caso de éxito del servicio.

Esta tesis se enfoca en la evaluación de los servicios multimedia. La ubicuidad de los servicios en el universo de Internet pone de manifiesto la relevancia de la investigación en nuevos y mejores servicios. Además la penetración actual de las tecnologías de acceso de banda ancha abre el mercado a una escala global y masiva. El desarrollo de nuevos servicios pasa necesariamente por la definición, prueba y evaluación del desempeño del nuevo servicio, pero estas fases cada vez son más cortas. La lección aprendida es que el éxito no puede predecirse, por lo que el *time-to-market* del producto mínimo viable tiene que ser inmediato, y con los menores cambios posibles en la infraestructura existente.

El éxito del servicio viene determinado por los usuarios. Los clientes del servicio deben estar satisfechos con la experiencia global. El concepto de experiencia global es amplio, y dependiendo del servicio puede incluir: la facilidad de descubrir nuevos servicios, la sencillez en su uso o el coste (si es aplicable), ajustado a sus cualidades. Indiscutiblemente, el servicio debe generar una experiencia positiva en el usuario. Precisamente la identificación de las mejoras obtenidas en los servicios supone en sí mismo un reto. Existen muchas variables que cuantifican el desempeño de un servicio, pero es la evaluación subjetiva la que proporciona una medición *end-to-end* de la verdadera utilidad del servicio. En el concepto de experiencia multimedia tienen cabida también aspectos variados como la calidad de audio, la relación audio-video o incluso aspectos sociológicos y del entorno. En este caso el análisis se concentra fundamentalmente en el aspecto puramente visual de la calidad, asociado al contenido de vídeo, que, por otra parte, es el componente de los servicios multimedia que más recursos consume.

En términos operativos, los agentes involucrados en el desarrollo y la gestión de los servicios multimedia necesitan la realimentación de métricas objetivas que permitan evaluar cuantitativamente el rédito del servicio. En este aspecto las métricas de video objetivas y su correcta alineación con las evaluaciones subjetivas suponen una herramienta muy útil para los operadores.

1.3 Objetivos y metodología de investigación

El objeto principal de la presente Tesis es la preparación de los entornos de experimentación, análisis y medida destinados a servicios multimedia. Del foco principal y en base a las contribuciones realizadas durante los últimos años al estado del arte en este campo, se distinguen tres grandes ejes sobre los que se articula esta tesis, cada cual con sus objetivos particulares: herramientas, métricas objetivas y métricas subjetivas.

En la primera fase de la investigación se fija el objetivo de experimentar con las herramientas de emulación o simulación en tiempo real para determinar sus rangos de funcionamiento, su precisión y su habilidad para trabajar con servicios masivos en el

ámbito multimedia. Las distintas herramientas existentes en el mundo académico o el mercado tienen ámbitos de trabajo diferentes y propósitos variados. Desde las herramientas de simulación tradicionales, a los *testbed* más completos o dispuestos mundialmente, existe un gran abanico de posibilidades disponibles para la elección de un sistema de evaluación. La sistematización de la construcción de entornos de evaluación por otro lado es inexistente.

Los objetivos particulares en lo relativo a las herramientas son:

- Analizar los entornos de evaluación de prestaciones de servicios multimedia, especialmente de vídeo, a través de herramientas conocidas como simuladores y emuladores.
- Describir el alcance de las soluciones y las alternativas de escalabilidad existentes.
- Proponer una metodología de construcción de topologías para la evaluación de servicios masivos

Las métricas de calidad objetiva aplicadas a los sistemas de distribución *streaming* forman parte de los criterios de evaluación, tanto del desempeño como de los resultados del experimento realizado. Las métricas objetivas son muy populares por su bajo coste de aplicación. La aplicación de las métricas objetivas más populares a los servicios de *streaming* adaptativos presenta algunas particularidades. Partiendo de las dificultades para obtener un decodificador adaptativo operativo incluso para secuencias alteradas por los errores de transmisión, se encuentra que además el uso de las métricas más utilizadas en estos casos presentan indeterminaciones, como es el caso de la métrica *Peak Signal-to-Noise Ratio* (PSNR). Además, las métricas de calidad objetiva más reputadas son generalmente del tipo de referencia completa, *Full Reference* o FR, por lo que su aplicación en tiempo real no es posible. Esto es un inconveniente en la aplicación de las métricas de calidad objetiva como criterios de decisión en sistemas adaptativos.

Los objetivos particulares en este bloque incluyen:

- Desarrollar un decodificador compatible con los flujos adaptativos distorsionados por causas de congestión.
- Proponer y contrastar una solución para la indeterminación de la métrica PSNR aplicada a los sistemas de *streaming* adaptativos.
- Proponer un método de estimación de métricas FR basado en información contextual que permita su aplicación a los algoritmos de decisión de los sistemas adaptativos.

Las métricas de calidad subjetiva son el método de evaluación más preciso para la evaluación de los servicios multimedia. Contar con la opinión de una masa suficiente de usuarios permite extrapolar importantes resultados sobre el desempeño del servicio. Sin embargo la evaluación subjetiva de cada servicio creado es inviable. Por esto se trabaja

en el estudio subjetivo de ciertos criterios base que permitan construir multitud de servicios siguiendo unas pautas definidas según los criterios subjetivos. En el transcurso de esta tesis se realizan dos estudios subjetivos con un alto grado de participación.

Los objetivos particulares respecto a las métricas de calidad subjetiva:

- Analizar los criterios base de los servicios multimedia desde una perspectiva global mediante un estudio subjetivo
- Comparar un servicio adaptativo con uno tradicional y determinar los umbrales de tolerancia y preferencia de los usuarios mediante un estudio subjetivo.
- Analizar la correlación de las métricas objetivas con las preferencias indicadas por los usuarios del sistema adaptativo

Para conducir la investigación realizada hemos aplicado las técnicas de investigación en ingeniería. El método científico establece las fases principales a seguir para reconocer el objetivo, una vez comprendida la realidad a la que nos enfrentamos, y para seguir los pasos necesarios para alcanzar los objetivos definidos.

En primer lugar, de una investigación inicial se identifican los problemas presentes en el ámbito de estudio. Los objetivos principales son el reflejo de los problemas identificados. La exploración de las bases de datos de conocimiento científicas permite profundizar en el estudio del problema y enfocar correctamente las soluciones propuestas. A continuación se formulan las hipótesis y objetivos concretos a demostrar. Condicionado por lo anterior, se diseñan y ejecutan los experimentos necesarios para confirmar o desmentir las hipótesis planteadas

En la fase de análisis de los resultados éstos son procesados estadísticamente para refutar la significación de las conclusiones obtenidas. La fase final concluye con la exposición de las conclusiones

A lo largo de la memoria de la tesis se aportan detalles técnicos de la realización de los experimentos para facilitar su reproducibilidad. Particularmente importante en el caso de los experimentos subjetivos, se disponen al servicio de la comunidad los contenidos procesados junto a sus evaluaciones para posibilitar el contraste de los resultados descritos.

Las fuentes de conocimiento exploradas son compiladas en una herramienta de bases de datos bibliográficas³ que permite la integración directa de las referencias en los principales procesadores de textos.

³ <https://www.zotero.org/>

1.4 Estructura del contenido

La estructura de la tesis está adaptada a los bloques temáticos sobre los que se asientan las contribuciones realizadas. El documento se estructura en 8 capítulos. Los primeros capítulos se reservan a la introducción y el estado del arte de la técnica. En la parte central del documento se desarrollan los tres bloques temáticos diferenciados que concentran las contribuciones resultado de la investigación realizada. En la última parte de la memoria se resumen las conclusiones, trabajos futuros y aportaciones realizadas.

En el Capítulo 1 se lleva a cabo una introducción a la temática de la tesis y se define el marco conceptual en el que se emplaza. Se definen los objetivos generales y particulares, describiendo la metodología de investigación utilizada. Por último se detalla la estructura de la tesis.

En el Capítulo 2 se realiza un recorrido por el estado del arte de la materia en los tres bloques temáticos que se incluyen en esta tesis. La primera parte se dedica a la exploración de los conceptos de simulación, simulación en tiempo real y emulación y su estado actual. Seguidamente se explora la situación de las métricas objetivas de calidad multimedia. En la última parte se estudia el estado de la técnica de las métricas subjetivas de calidad y las metodologías de experimentación subjetiva.

El Capítulo 3 se dedica a las herramientas y técnicas de experimentación de servicios. Se explora en detalle la técnica de simulación en tiempo real con los dos simuladores más conocidos que admiten dicha técnica. Se evalúa su rendimiento para un caso de estudio particular y se postula la metodología de construcción de entornos de experimentación de servicios masivos.

En el Capítulo 4 se revisan las métricas objetivas de calidad, en particular de calidad de vídeo. Se describen los problemas asociados a éstas métricas al ser aplicadas a servicios adaptativos. Se detalla la construcción de un decodificador compatible con secuencias adaptativas afectadas por pérdidas de transmisión. Se formula una métrica adaptada para la medición de calidad objetiva en secuencias adaptativas distorsionadas. Finalmente se propone un método de estimación de métricas FR basadas en información contextual para sistemas adaptativos.

En el Capítulo 5 se exploran las métricas de calidad subjetiva. Se describe la realización y los resultados de los dos experimentos subjetivos realizados. El primero de ellos en términos de calidad subjetiva de servicios multimedia desde el punto de vista global (*end-to-end*) y el segundo desde un punto de vista específico de servicios adaptativos, utilizando una metodología de comparación por pares con respecto a un servicio convencional (no adaptativo).

En la parte final del documento, el Capítulo 6 se dedica a recoger las conclusiones de esta tesis. El Capítulo 7 enuncia los trabajos futuros derivados de la línea de

investigación iniciada. En el Capítulo 8 se relacionan las publicaciones derivadas de esta tesis junto con una breve descripción de las mismas.

2 Estado del arte

En el campo de los servicios multimedia, en un periodo de tiempo reducido, se han producido grandes avances en todos los aspectos. Dada la importancia de la evaluación de dichos servicios la tecnología que da soporte a las evaluaciones y las métricas han evolucionado de igual forma.

La base de la evaluación de los servicios es la utilización de escenarios bien sean reales o representaciones de la realidad. En base a esto, la primera parte de nuestro repaso al estado de la técnica y los trabajos relacionados se centra en los métodos y sistemas de evaluación de servicios.

Las métricas específicas de calidad de servicios multimedia evolucionan de las métricas de servicio ya conocidas. En un segundo bloque se revisan las métricas objetivas y los trabajos relacionados en el ámbito de los servicios multimedia. Puesto que los experimentos subjetivos suponen la única forma realista de evaluar el rendimiento global del servicio multimedia en todo su ámbito, dedicaremos esfuerzo también a revisar el estado de la tecnología en ese aspecto y los trabajos relacionados.

2.1 Evaluación de servicios: simulación en tiempo real

Un servicio multimedia puede ser descompuesto en tres elementos básicos. En un extremo, el ente productor y adaptador de contenidos (el servidor). En el otro extremo, el consumidor o reproductor de los contenidos (cliente) y entre medias la red de transporte que soporta dicho servicio. La Figura 2-1 es un esquema simplificado de un servicio multimedia clásico.

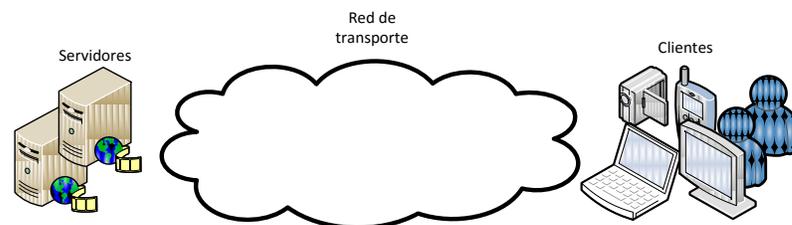


Figura 2-1: Diagrama básico del servicio multimedia

Tradicionalmente y llegando hasta nuestros días, la simulación de redes continúa siendo una opción popular en muchos ciclos de desarrollo de los servicios multimedia. En las primeras fases de la evaluación de un nuevo servicio es frecuente recurrir a la simulación completa de los sistemas. En este tipo de simulaciones, todos los elementos del servicio son descritos en el lenguaje de simulación que corresponda. Generalmente las herramientas de simulación proporcionan los mecanismos para describir la mayor parte de los componentes del servicio de forma analítica. El propio comportamiento de los hipotéticos usuarios del servicio puede ser representado por una función probabilística en base a un determinado comportamiento esperado. En trabajos como [4], los autores realizan un modelo del comportamiento de los usuarios de un servicio de vídeo bajo demanda.

En este tipo de evaluaciones, aumentar el tamaño del servicio evaluado o su precisión suponen aumentar la complejidad de la evaluación y, generalmente, aumentar el tiempo de simulación. En la medida de lo posible, las técnicas de computación paralela han permitido aumentar la capacidad de los sistemas de simulación [5] sin incrementar el tiempo. Otra opción habilitada en este tipo de evaluaciones, dado que todos los elementos son representaciones virtuales de elementos reales, es controlar el avance del tiempo. En simulaciones puramente virtuales, donde los elementos no están relacionados con el reloj de tiempo real se permite completar grandes simulaciones en tiempos de espera razonables.

Sin embargo, la precisión y el realismo de las simulaciones puras han sido puestos en duda en muchas ocasiones [6][7][8]. Es necesario desarrollar modelos para todas las capas de red, desde el que emula el nivel físico hasta la capa de aplicación, y esto genera mucha incertidumbre. El ajuste fino de los parámetros del modelo en particular condiciona mucho los resultados, aunque algunos autores confirman que pueden llegar a ser bastante precisos. En [9], los autores comparan los simuladores NS-2, OPNET Modeler y QualNet con una implementación real inalámbrica en entornos interiores y exteriores concluyendo que los modelos, tanto de la capa física como de propagación, bien parametrizados, pueden ser bastante similares a los resultados reales.

Al margen de la disputa por la precisión de los resultados, parece claro que la evaluación de ciertos aspectos ligados al comportamiento real de dispositivos, aplicaciones o usuarios reales, no es posible hasta llegar a la fase del despliegue del servicio real.

En ciertas fases del desarrollo del servicio, el uso del máximo número de elementos reales es clave. Precisamente porque en estas fases se evalúa el rendimiento del servicio y también su implementación. El esfuerzo realizado para modelar ambos aspectos en un entorno de simulación, además de costoso, puede comprometer los resultados. Por ejemplo, algunos autores reconocen que la simulación de redes no es precisa ya que no modela adecuadamente los recursos *hardware* o los procesos *software* y su sobrecarga en el sistema [10]. En este trabajo, los autores proponen una extensión para simuladores de propósito general que habilita la inclusión de modelos de ejecución

de procesos a través de trazas obtenidas en sistemas reales. Para mejorar la precisión de las simulaciones, la inclusión de estos modelos debe ser considerada.

La obtención de los modelos es una tarea costosa que además debe ajustarse tras cada iteración en el desarrollo del servicio real. Algunos modelos han sido obtenidos de forma efectiva a través de la observación del servicio real [11]. No obstante, la inclusión de modelos complejos de aplicaciones o protocolos puede ser una tarea delicada, como reconocen algunos autores para el caso de un modelo de un servicio multimedia en OPNET Modeler [12].

Los avances en las herramientas de simulación están reduciendo el salto entre la simulación pura y la implementación real del servicio en todas las fases del desarrollo. Mediante la técnica de simulación en tiempo real, las aplicaciones reales y su tráfico pueden ser introducidos en un entorno simulado ahorrando el esfuerzo de su modelado. El esfuerzo de modelado se reduce enormemente puesto que los modelos de red son reutilizables y los del servicio o clientes no son necesarios. Los autores en [13] comentan la motivación tras el concepto de simulación en tiempo real y lo comparan con otros métodos de evaluación de redes, entre ellos la simulación.

En la figura 2-2 se representa la diferencia de enfoque entre simulación pura y simulación en tiempo real. En la simulación pura todos los componentes son modelados mientras que en una configuración típica de simulación en tiempo real la red de transporte es simulada y el resto son componentes reales.

Las extensiones practicadas en reputados simuladores de red, como NS-3 u OPNET Modeler han permitido integrar éstos en escenarios reales, inyectando paquetes hacia y desde la red real [14][15][16]. Un problema común de éstas herramientas es que el caudal máximo alcanzable está limitado por el consumo de recursos y el rendimiento del simulador, lo que supone un barrera importante [17][18].

Varias técnicas han sido estudiadas para reducir el impacto de la limitación en el caudal de las herramientas de simulación en tiempo real. Las técnicas de *partición de la topología* [19], *modelado de tráfico mediante flujos* [20] o *virtualización del tiempo real* [21] son algunos ejemplos. La metodología desarrollada durante el transcurso de esta Tesis [22] propone los pasos necesarios para obtener entornos de emulación masivos combinando varias de éstas técnicas.

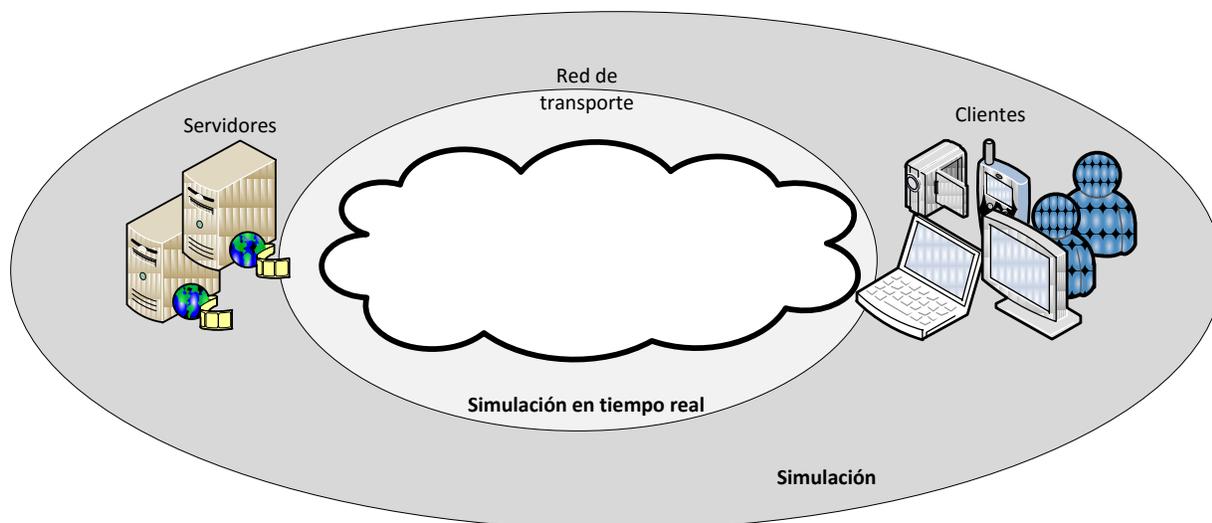


Figura 2-2: Entornos de evaluación de servicios, simulación vs simulación en tiempo real.

Otras aproximaciones proponen utilizar la infraestructura de Internet para llevar a cabo evaluaciones de servicios con un realismo mayor. PlanetLab [23] ofrece una plataforma para la configuración de simulaciones en tiempo real, ofreciendo la posibilidad de disponer de emplazamientos remotos conectados mediante Internet. Este tipo de soluciones adolece de limitaciones para servicios de gran consumo de recursos, debido a la limitación en los enlaces. Varias extensiones a PlanetLab proponen soluciones como el uso de procesadores de red para alcanzar mayor capacidad de procesamiento [24]. En contrapartida, la escalabilidad es aún menor. En ambos casos la repetibilidad es baja, debido a que no es posible controlar las condiciones del entorno durante el experimento.

En un esfuerzo por mejorar la escalabilidad con un detrimento reducido en el realismo, otros trabajos proponen emplear pasarelas o “*flakeways*” que pueden alterar paquetes de tráfico reales que las atraviesan, pudiendo producir alteraciones como pérdidas o retardos [25]. Algunas implementaciones de este concepto permiten manejar un caudal importante de tráfico, al implementar nodos virtuales y enlaces que básicamente mueven imágenes de paquetes en un *kernel* personalizado de Linux [26] o usando equipos especializados de alto rendimiento [27]. Sin embargo, la configuración de estos modelos requiere la parametrización del comportamiento de la red bajo estudio en términos como PLR, retardos o tasa de bits erróneos (BER). La flexibilidad se reduce enormemente para desplegar topologías variadas y modelos complejos.

Dada la heterogeneidad de técnicas y entornos de experimentación posibles, es necesario centrar el objetivo de trabajo en alguna de las técnicas y decantarse por los entornos correspondientes. El enfoque sobre las técnicas de simulación en tiempo real es una apuesta por explorar las capacidades de una tecnología prometedora. La simulación en tiempo real parece ofrecer las ventajas del longevo mundo de la simulación y las dosis necesarias de realidad que supone incluir elementos reales en las evaluaciones. En el campo de la experimentación de servicios multimedia novedosos, tales como los de

tecnologías escalables, resulta interesante combinar experimentados simuladores de red con la información real de aplicaciones y protocolos en desarrollo. La elección de las herramientas de simulación para este objetivo es el primer paso.

2.1.1 Herramientas de simulación en tiempo real

En los entornos simulados, la elección de una herramienta de trabajo puede estar condicionada por el rendimiento, la escalabilidad o la disponibilidad de modelos. Algunos trabajos han tratado de comparar el rendimiento de los simuladores más comunes [28]. Para aquellos en los que la extensión de simulación en tiempo real ha sido habilitada, el rendimiento interno de la simulación es aún más importante.

NS-3 y OPNET Modeler son dos de los simuladores más reputados entre la comunidad. Las capacidades como simuladores de ambos han sido extensamente evaluadas. Por nombrar solamente algunos trabajos, los autores en [12] utilizan OPNET para simular toda clase de tráfico en un servicio de vídeo bajo demanda en una red hospitalaria y en [29] los autores utilizan OPNET para evaluar el rendimiento de varios protocolos en una red de un campus universitario.

El posterior desarrollo de las tecnologías de emulación (*Tapbrige* y SITL para NS-3 y OPNET respectivamente) han posicionado a ambas herramientas como candidatos idóneos para evaluar servicios en entornos mixtos, incorporando elementos reales a las simulaciones. La investigación usando estos módulos es menos extensa. En el trabajo realizado en [30] los autores evalúan una aplicación de vídeo usando el módulo SITL de OPNET. Sin embargo, no proporcionan ninguna referencia respecto a la confiabilidad de los resultados. Dada la restricción temporal que impone el módulo SITL es importante contrastar el rendimiento del simulador antes de extraer conclusiones sobre los resultados. En el trabajo realizado en [31] con NS-3, los autores sí comparan el experimento simulado con un despliegue real, dando pistas sobre el rendimiento de NS-3. Sin embargo, este es un resultado colateral. No se conocen trabajos que profundicen en el rendimiento específico de las extensiones de simulación en tiempo real de las herramientas NS-3 u OPNET.

En esta investigación se eligen ambas herramientas, NS-3 y OPNET Modeler como medio para la construcción de entornos de evaluación enfocados en la exploración de la técnica de simulación en tiempo real.

2.2 Métricas de calidad

Las métricas son el medio necesario para el análisis del rendimiento de los servicios multimedia. La distribución del contenido digital a través de las redes de telecomunicaciones se ve afectado por pérdidas en distintos niveles de la cadena. Siguiendo el ciclo de vida del servicio, desde el proceso de codificación hasta el proceso de visualización se producen pérdidas. En la generación el vídeo es codificado para

comprimir la información a transmitir, lo que ocasiona pérdidas. En el proceso de transmisión, el vídeo sufre pérdidas, retrasos y retardos variables que perjudican la forma en que la información es decodificada. Aunque en el proceso de recepción es posible corregir o compensar algunos errores, la reproducción final se ve afectada por la suma de los errores acumulados desde la generación del contenido.

En los sistemas de telecomunicación tradicionales, las métricas de calidad de servicio (QoS) están centradas en la medida de los parámetros de red: pérdida de paquetes, retraso, *jitter*, etc. El problema con este enfoque es que, por la heterogeneidad de los servicios actuales, las métricas de QoS han dejado de representar el rendimiento global del servicio. Cada vez de forma más frecuente, la calidad de los servicios está influida por características de capas superiores como codificación, compresión o la naturaleza del contenido en sí mismo.

En el ámbito de los servicios multimedia, las métricas de QoS son extendidas para incorporar otros aspectos específicos de las características del medio en particular, el usuario y el contexto. La calidad de la experiencia (QoE) es considerado el sustituto natural a la popular QoS pues se sabe que la experiencia del usuario final incluye factores que no eran considerados en el concepto de QoS. La QoE pretende incluir factores que abarcan aspectos muy diversos. Entre muchos factores, las métricas de QoE buscan evaluar el impacto de la falta de sincronización entre audio y vídeo, la influencia del tipo de contenido multimedia en la valoración final del usuario, la respuesta del usuario ante las distorsiones producidas por los procesos de codificación y decodificación y, por supuesto, los errores introducidos como resultado de la transmisión sobre una red de comunicaciones propensa a errores [32]. En la figura 2-3 se presenta las etapas clásicas en el ciclo de vida del servicio y la diferencia entre los conceptos de QoS y QoE.

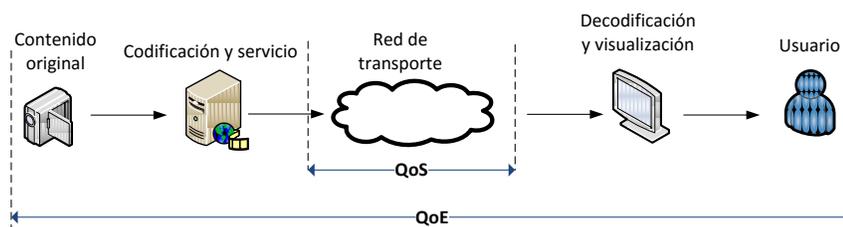


Figura 2-3: Calidad de Servicio vs Calidad de Experiencia

Según se recoge en el estándar de la Unión Internacional de Telecomunicaciones (ITU) ITU-T P.910, Apéndice I, el concepto de QoE representa la aceptabilidad general de una aplicación o servicio, tal como es percibida subjetivamente por el usuario final. Incluye todos los efectos *end-to-end* en el sistema (cliente, red, infraestructura del servicio) y puede estar influenciado por las expectativas del usuario y el contexto. Por tanto, la QoE es determinada de forma subjetiva por el usuario final y puede ser distinta para otro usuario.

Sin embargo, por razones prácticas, la QoE es generalmente estimada mediante medidas objetivas. Existe una necesidad creciente por modelar y llegar a estimar la QoE basándose en indicadores de rendimiento objetivos (*objective KPI*). Además conseguir establecer las relaciones adecuadas entre las causas de las pérdidas en QoE puede permitir gestionar los sistemas para poder garantizar los niveles adecuados de QoE.

Las métricas de QoE, estrictamente hablando se reducen a las dos formas de representar la valoración subjetiva, en una escala absoluta MOS (*Mean Opinion Score*) o en una escala relativa DMOS (*Differential Mean Opinion Score*). Las métricas de QoE son obtenidas a través de los experimentos subjetivos, guiados por las metodologías existentes y definidos en las recomendaciones ITU-T P.910. Las métricas de QoE objetivas son en realidad predictores de las métricas subjetivas. El trabajo necesario para construir y validar nuevas métricas de QoE objetivas incluye la calibración de éstas con las métricas MOS o DMOS, Figura 2-4.

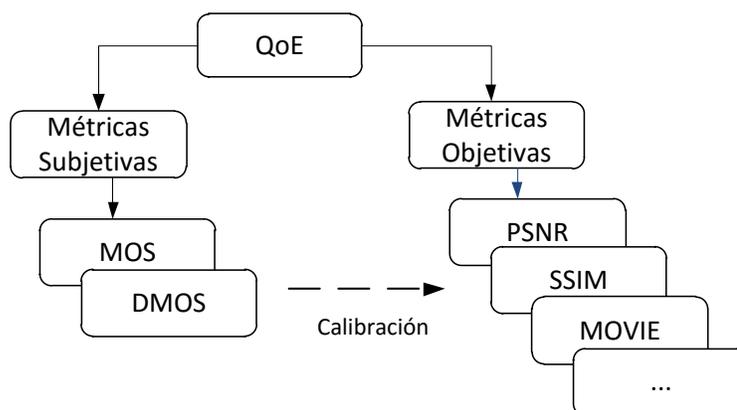


Figura 2-4: Clasificación de métricas de QoE

Seguidamente se repasan las métricas principales de QoS, las métricas objetivas de QoE y los métodos subjetivos principales para la obtención de las métricas subjetivas.

2.2.1 Métricas de QoS

Las métricas básicas de los servicios y aplicaciones en Internet tienen la misión de evaluar los aspectos de calidad de servicio generalmente acordados con los usuarios. Entre otros parámetros, las métricas de QoS evalúan la disponibilidad, la latencia, el ancho de banda y los tiempos de fallo y recuperación [33]

La tasa de paquetes perdidos (PLR) y el *jitter* (variación del retardo) son las métricas de QoS de primer nivel usadas comúnmente en las evaluaciones de servicios multimedia [34], [35].

Las métrica de pérdidas de paquetes, PLR, se obtiene como la relación entre el total de paquetes perdidos en un experimento y la cantidad total de paquetes transmitidos, según la ecuación 2.1.

$$PLR = \frac{packetlost}{totalpakcets} \times 100 \quad (2.1)$$

El *jitter* es la medida de la variación del retardo, tal como se define en [36], el *jitter* para un elemento i es una media móvil del retardo (D) para el paquete actual, ecuación 2.2.

$$J_i = J_{i-1} + \frac{(|D(i-1, i)| - J_{i-1})}{16} \quad (2.2)$$

Aún hoy en día continúa siendo importante mantener las primitivas métricas de QoS y se siguen realizando estudios para relacionarlas con las métricas de QoE [37].

2.2.2 Métricas de QoE objetivas

Las métricas de calidad objetiva pueden servir a múltiples propósitos dentro de la arquitectura del servicio multimedia. Entre otras aplicaciones sirven como evaluación de nuevos *codecs*, para la planificación del esquema de transmisión o el dimensionamiento de la red, como una media de la provisión de calidad entre los extremos del servicio o como una métrica del rendimiento del servicio dentro de la red.

Las métricas de calidad objetiva son clasificadas atendiendo a diversos criterios. De acuerdo a [38] se establece una primera clasificación en función de las características de la información considerada por la métrica.

- *Media Layer Metrics*. Utilizan la información del contenido. Serán posteriormente clasificadas según la cantidad de información empleada como referencia en referencia completa, reducida o nula. En el ámbito de los servicios multimedia y dentro de esta categoría, los autores en [39] hacen una distinción adicional entre métricas de imagen y métricas de vídeo.
- *Parametric packet-layer*. Predicen la calidad a partir de la información de las cabeceras de los paquetes del servicio lo que aligera la complejidad de cálculo pero imposibilita los análisis contextuales tan importantes en QoE. Un ejemplo es el trabajo con ITU-T P.564 para servicios de voz.
- *Parametric planning*. Combina el conocimiento de la codificación de la información y la información del rendimiento del servicio en la red para explicar la QoE. Por ejemplo la métrica E-model para VoIP o ITU-R G. 1070 para videoconferencia
- *Bitstream-layer*. Una versión intermedia entre las dos anteriores que utiliza tanto información de nivel de paquete como la información del servicio en la red.
- *Hybrid*. Combina los modelos anteriores

Las métricas más utilizadas corresponden a la categoría primera y reciben, como se menciona, distintas clasificaciones. Una de las posibles clasificaciones atiende a la cantidad de información de referencia que se tiene en cuenta a la hora de evaluar el rendimiento de servicio en el extremo del receptor. En general la precisión de la métrica decrece a medida que se reduce la información de la fuente original de contenido que se utiliza en recepción para evaluar la calidad percibida, figura 2-5.

Se habla de métricas de referencia completa o *Full Reference* (FR) cuando el contenido original es utilizado para establecer una comparación directa con el contenido multimedia recibido. Estas métricas son de aplicación habitual en laboratorios, estudios específicos de rendimiento, ensayos puntuales o evaluación de *codecs*. Requieren un procesamiento más costoso, entre otras cosas porque se supone una alineación espacio-temporal de ambos contenidos que no siempre es posible. En sistemas adaptativos por ejemplo encontramos la primera limitación del uso de este tipo de métricas, que requieren compromisos a la hora de comparar la secuencia adaptada con un contenido original que puede ser de distinto en tamaño en la resolución temporal (menos *frames per second*) o espacial (menor tamaño del *frame*).

Las métricas *No Reference* (NR) utilizan solo la información de la secuencia recibida para establecer la calidad percibida. Son perfectas para la aplicación en entornos reales pero gozan de menor reputación.

Las métricas *Reduced Reference* (RR) utilizan solo una parte —pequeña— de la información. Algún tipo de procesamiento establece qué metadatos se extraen del contenido original para ser utilizados en la medición de la calidad percibida. Ofrecen una solución de compromiso entre las dos anteriores. Para ser utilizadas en sistemas reales se requiere la adaptación del servicio que incluya la meta-información en la transmisión.

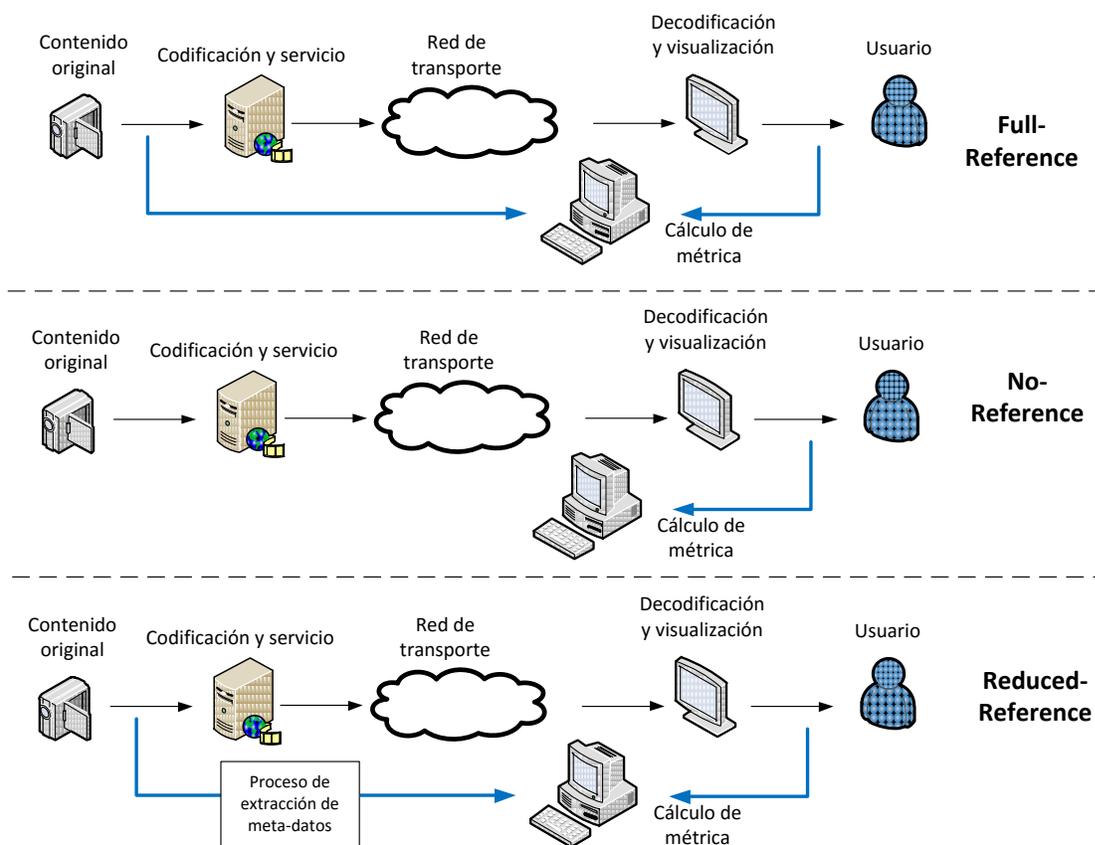


Figura 2-5: Clasificación de métricas de QoE según la cantidad de información de referencia

Existen en la literatura multitud de métricas de los distintos tipos. No es objeto de este trabajo realizar una evaluación de todas ellas. Entre las más utilizadas se encuentran las métricas *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity Index* (SSIM) y más recientemente, el índice *MOTION-based Video Integrity Evaluation* (MOVIE). Estas tres métricas han sido elegidas recurrentemente durante la evaluación de los resultados de esta investigación. Las tres pertenecen a la categoría de métricas FR. La métrica PSNR es una de las métricas de imagen cuyo uso se ha extendido rápidamente a la evaluación de secuencias de vídeo. Las métricas SSIM y MOVIE incorporan aspectos específicos del vídeo como las relaciones estructurales que evalúa SSIM o las dependencias *inter-frame* que considera MOVIE.

PSNR es la métrica objetiva, con frecuencia asimilada de forma directa a QoE, que ha sido desarrollada con la intención de emular la impresión de la calidad de imagen del sistema visual humano [40]. Los valores de PSNR se obtienen para cada par de imágenes de la secuencia multimedia (*frames*), una imagen original y la otra imagen una vez transmitida mediante el servicio bajo estudio. Cada par de imágenes, original y distorsionada, se comparan aplicando la fórmula original de la ecuación 2.3. Las

imágenes son comparadas en su formato nativo, una vez decodificadas. Normalmente el formato asumido es 4:2:0. Es un convenio asumido que el valor de PSNR se computa exclusivamente para el valor de luminancia de la imagen, ya que el sistema de visión humano (HVS) es más sensible a la distorsión de la iluminación que a los cambios de color [41].

$$PSNR_{dB} = 20 \log_{10} \left(\frac{V_{peak}}{\sqrt{MSE}} \right) \quad (2.3)$$

Donde $V_{peak} = 2^k - 1$, $k = \text{número de bits por pixel (componente luminancia)}$. MSE , definido en la ecuación 2.4 es el valor del error cuadrático medio. MSE es la forma de estimar la diferencia entre los pares de frames (x,y).

$$MSE = \frac{1}{N_{col} N_{row}} \sum_{i=1}^{N_{col}} \sum_{j=1}^{N_{row}} [Y_S(i,j) - Y_D(i,j)]^2 \quad (2.4)$$

$Y_S(n,i,j)$ es el valor de luminancia del pixel (i,j) en el frame n del vídeo enviado. $Y_D(n,i,j)$ es el equivalente en el vídeo recibido.

Los valores de PSNR son normalmente clasificados de acuerdo a una escala subjetiva MOS que ha sido calibrada por varios autores en [42]. Como resultado, se obtiene una forma de histograma con la calidad de las imágenes yendo desde malo hasta excelente. Los niveles y sus equivalencias en PSNR se muestran en la Tabla 2-1

Tabla 2—1 Correspondencia PSNR - MOS

PSNR (dB)	MOS
> 37	5 (Excellent)
31 – 37	4 (Good)
25 – 31	3 (Fair)
20 – 25	2 (Poor)
< 20	1 (Bad)

La métrica PSNR para la evaluación de vídeo es la métrica objetiva más popular conocida. En efecto, PSNR ha sido usada como criterio de referencia para evaluar nuevas métricas subjetivas, como por ejemplo, en los trabajos del grupo VQEG. De estos trabajos se obtiene precisamente la estandarización del cálculo de PSNR (ITU-T J.340). Sin embargo, muchos trabajos han discutido que PSNR ha demostrado tener un comportamiento deficiente cuando distintos tipos de distorsión son aplicados a la fuente original. En efecto, *Mean Square Error* (MSE) y en consecuencia PSNR, tiene una esfera de valores equivalentes para distorsiones que ocasionan calidades percibidas relativamente diferentes [43]. Algunos trabajos proponen mejoras a las métricas objetivas

como PSNR a base de incorporar conceptos como factores de ponderación o aspectos del HVS [44]–[47].

La métrica SSIM tiene una base técnica similar a la PSNR, incorpora sin embargo aspectos que la PSNR no llega a modelar. Para cada par de imágenes (nuevamente en origen y en destino), tiene en cuenta niveles de luminancia (luma o señal Y), crominancia (información de color) y distorsión estructural de la imagen, según la ecuación 2.5. Si se compara con una referencia no distorsionada, la métrica ofrece una medida cuantitativa de la calidad de la segunda imagen [48]. Aunque los valores de esta métrica son obtenidos para cada *frame* de forma independiente (al igual que PSNR) con frecuencia es admitido que el promedio de valores a lo largo de la secuencia multimedia es un buen indicador de la calidad del contenido global, siempre y cuando todos los *frames* de la secuencia tengan la misma importancia.

$$SSIM(x, y) = l(x, y)c(x, y)s(x, y) \quad (2.5)$$

Los cambios de luminancia, contraste y estructurales son expresados por $l(x, y)$, $c(x, y)$ y $s(x, y)$, respectivamente. Cada valor se calcula con las ecuaciones 2.6, 2.7 y 2.8:

$$l(x, y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} \quad (2.6)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (2.7)$$

$$s(x, y) = \frac{\sigma_{xy}}{\mu_x^2 + \mu_y^2} \quad (2.8)$$

Y el valor de SSIM general se computa de acuerdo a la ecuación 2.9:

$$SSIM(x, y) = \frac{4\mu_x\mu_y\sigma_{xy}}{(\mu_x^2 + \mu_y^2)(\sigma_x^2 + \sigma_y^2)} \quad (2.9)$$

Donde μ_x , μ_y son las medias de los *frames* (x,y), σ_x , σ_y son las desviaciones estándar de los *frames* (x,y) y σ_{xy} es la correlación de los *frames* (x,y).

Se han desarrollado modelos matemáticos que relacionan las métricas PSNR y SSIM para ciertos tipos de imágenes y distorsiones [49]. Los resultados de la aplicación de la métrica SSIM a las secuencias de vídeo de distintas bases de datos arrojan resultados mejor correlados con las métricas subjetivas que los resultados de otras métricas, tal como muestran los resultados discutido en [50]. Algunas métricas buscan extender la información estructural incorporando aspectos como la textura de la imagen [51].

También enfocado en analizar la información estructural, el índice MOVIE [52], es una métrica de tipo FR. Los autores demuestran que el índice MOVIE arroja una correlación

muy buena con los estudios subjetivos publicados por el grupo *Video Quality Experts Group* (VQEG), usando la librería *FRTV Phase 1* como banco de pruebas. En efecto los autores defienden que este algoritmo supera en precisión a varios algoritmos respaldados por el propio grupo VQEG

Para el cálculo de índice MOVIE, las secuencias evaluadas (referencia y test) son descompuestas en canales espacio-temporales utilizando filtros del tipo Gabor. La calidad espacial utiliza un método inspirado en los cálculos de SSIM. El índice temporal se obtiene usando información de la actividad en la secuencia de referencia. Finalmente los índices temporal y espacial se combinan para obtener un índice único, conocido como índice MOVIE. En la Figura 2-6 se muestra un diagrama de bloques del cálculo de la métrica MOVIE.

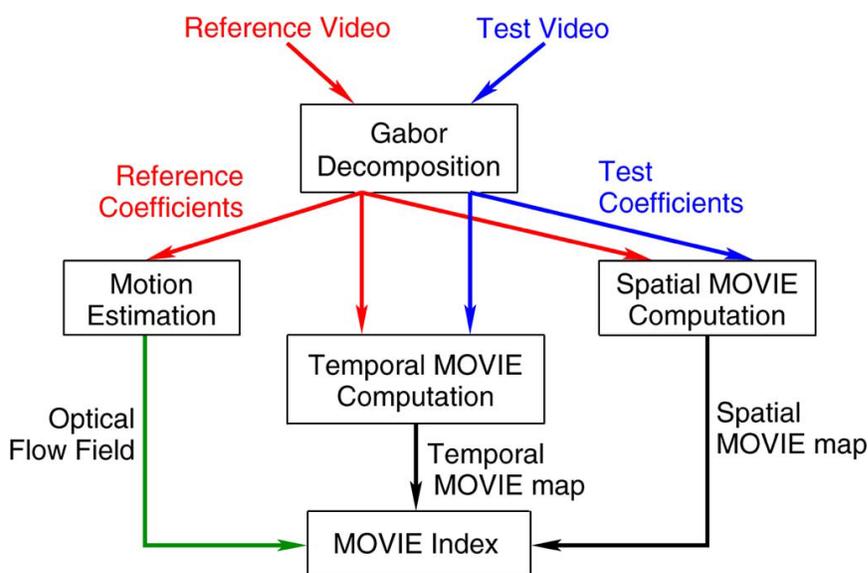


Figura 2-6: Algoritmo de cálculo de la métrica MOVIE, fuente [52]

En una comparativa de las métricas, PSNR, SSIM y MOVIE, los autores demuestran que MOVIE arroja unos resultados superiores sobre las anteriores [53].

Los métodos de estimación objetiva de la calidad percibida se revelan como un sustituto muy plausible de los complejos test subjetivos. Ciertamente, las métricas objetivas actuales no pueden sustituir completamente a las evaluaciones subjetivas, pero su comportamiento es prometedor. Aunque la mayor parte de los esfuerzos se dirigen a la evaluación objetiva de la calidad de vídeo percibida, los autores remarcan la necesidad de considerar también otros aspectos tan relevantes como el audio o la sincronización entre audio y vídeo [54].

La aplicación de las métricas en los experimentos realizados durante esta investigación no está exenta de problemas. Especialmente cuando se tratan de aplicar en la evolución

de la calidad de los servicios multimedia adaptativos. En el Capítulo 4 se profundiza en las contribuciones realizadas sobre las métricas de calidad objetivas.

2.2.2.1 Métricas de QoE objetivas en SVC

Existe un interés creciente en la medida de la mejora de calidad añadida por la nueva tecnología de *streaming* adaptativo. Las métricas objetivas aplicadas a este tipo de tecnologías son una solución muy adoptada.

En los primeros trabajos enfocados en este campo, debido al estado primitivo de las fases del desarrollo de los servicios adaptativos, es frecuente encontrar evaluaciones objetivas realizadas emulando el comportamiento del sistema adaptativo con otro codificador. En el trabajo descrito en [55], los autores utilizan secuencias de vídeo de distintas resoluciones, emulando la escalabilidad espacial de un hipotético sistema adaptativo, para medir mediante PNSR y SSIM la calidad resultante. En este trabajo los autores realizan un análisis interesante de cómo se compensa la pérdida de paquetes con la pérdida de resolución en distintos experimentos. Sin embargo, no existe un proceso real de adaptación y no utilizan el *codec* H.264 SVC.

Pero la aplicación directa de las métricas de calidad objetiva más habituales al contenido adaptativo no está exenta de problemas. Los autores en [56] ponen de manifiesto las limitaciones de MSE (o PSNR) para SVC. Proponen una nueva métrica con la PSNR como base que incorpora aspectos de la variación temporal y espacial así como la tasa de *frames*. Los autores dejan fuera del estudio la influencia de las pérdidas de paquetes en su métrica. El estudio es prometedor, incluso ha sido validados subjetivamente, aún así es un estudio más próximo a la evaluación de la codificación que a la evaluación *end-to-end* (que incluya las pérdidas de transmisión) que es necesaria para los servicios multimedia.

Precisamente por la falta de un decodificador versátil, capaz de trabajar con flujos adaptados y con pérdidas en la fase de transmisión, algunos autores han buscado la evaluación de este tipo de servicios a través del análisis a nivel de paquete. Los autores en [57] resuelven esta dependencia del decodificador obteniendo una métrica derivada del análisis a nivel de paquetes. Analizando la lista de paquetes recibidos un algoritmo determina la secuencia que es supuestamente decodificada y consultando una base de datos previamente construida con la calidad de cada bloque enviado, pueden estimar la calidad recibida. Es un aproximación ingeniosa, sin embargo se están asumiendo comportamientos del decodificador que no han evaluado.

La influencia del decodificador es además relevante por la existencia de distintos algoritmos de compensación de error ante las pérdidas de transmisión. La tecnología escalable por sus características incrementa las posibilidades en la compensación de los errores de transmisión. En el diseño de un servicio adaptativo cobra relevancia la protección que brinda la propia tecnología de capas [58]. Aún así, encima de esta capacidad nativa de compensación de errores, trabajos como el expuesto en [59]

proponen algoritmos de compensación que mejoran en varios aspectos a los competidores. Los autores en [60] evalúan distintos algoritmos de cancelación sobre el *software* decodificador de referencia *Joint Scalable Video Model* (JSVM). En este trabajo los autores investigan la calidad objetiva a través de PSNR y SSIM de secuencias con distintos ratios de pérdidas en las capas base y en las capas de mejora. Es un enfoque más correcto. Sin embargo, su estudio se limita a un tipo de escalabilidad y además no incluyen la variación dinámica de las capas a lo largo del tiempo, que en un sistema adaptativo real, se produce de forma frecuente.

Contar con un sistema adaptativo real, que utilice el *codec* nativo en esta tecnología (H.264 SVC) y contar con el decodificador capaz de reproducir las secuencias recibidas, tengan o no pérdidas no es una tarea sencilla, sobre todo porque el codificador JSVM no admite la reproducción de contenido extraído de entornos reales, donde las pérdidas ocurren con carácter aleatorio. Varios autores han construido herramientas para facilitar el análisis de este tipo de secuencias, pero o bien se basan en trazas textuales que simulan la transmisión [61] o bien simulan las pérdidas también sobre trazas pero además con otro *codec* [62]. Afortunadamente, comienzan a aparecer implementaciones más flexibles del decodificador que ya permiten evaluar secuencias más realistas. Los autores en [63] proponen la utilización del decodificador *openSVC* para estimación de la QoE a través de en una red neuronal entrenada con parámetros de QoS. Es muy interesante la propuesta del uso de este decodificador y será un foco importante de trabajo en nuestra investigación.

2.2.3 Métodos de QoE subjetivos

La única forma de evaluación de la calidad de experiencia percibida, estrictamente hablando, es la elaboración de test subjetivos, presentando los contenidos a un determinado número de usuarios, en unas condiciones establecidas (véase la Figura 2-7 como ejemplo) y recogiendo sus evaluaciones. El documento ITU-T P.910 contiene las recomendaciones de la ITU para realizar test subjetivos de evaluación de calidad percibida.

Entre los métodos propuestos se incluyen comparaciones implícitas como *Double Stimulus Continuous Quality Scale* (DSCQS), comparaciones explícitas como *Double Stimulus Impairment Scale* (DSIS), o evaluaciones absolutas como *Single Stimulus* (SS), *Single Stimulus Continuous Quality Evaluation* (SSCQE) o *Absolute Category Rating* (ACR). Las escalas de evaluación propuestas para las diferentes metodologías son construidas combinando distintas propiedades: absoluta o relativa, continua o discreta, numérica o adjetival [64].



Figura 2-7: Instalaciones en el grupo Multimedia Signal Processing Group (Lausanne, Suiza), colaborando en los experimentos de estandarización internacional (Call for Proposal for High Efficiency Video Coding (HEVC))

El trabajo en [65] constituye un repaso del trabajo del grupo de expertos en estimación de la calidad de vídeo VQEG erigido desde los grupos ITU-T e ITU-R. Entre las tareas que realizan se incluyen planes de tests, que indican los pasos requeridos para realizar la validación de métricas objetivas, incluyendo aspectos sobre el contenido, la configuración de los experimentos subjetivos hasta las técnicas estadísticas de procesamiento de datos. De los cinco test de validación completados hasta el momento, los dos primeros, *VQEG FR Television Phase I (FRTV-I)* [66] y *VQEG FR Television Phase II (FRTV-II)* [67] cubren los aspectos relativos a las métricas FR, donde se tiene en cuenta la información de origen completa en la evaluación. Los métodos RR y NR incluyen una referencia parametrizada o ninguna referencia al contenido original respectivamente. Son métodos que permiten la evaluación en tiempo real pero tienen mayor complejidad [68][69]. El tercer proyecto, *VQEG RRNR-TV* [70] estudia la estandarización de los métodos RR y NR, rechazando todos los modelos NR propuestos y estandarizando algunos RR. El cuarto proyecto se orienta a evaluar modelos objetivos de calidad de experiencia apropiados para aplicaciones multimedia y vídeo digital, *VQEG Multimedia Phase I (MM-I)* [71]. Como resultado, cinco recomendaciones han sido publicadas. Recientemente completado, el proyecto *VQEG High Definition Television (HDTV)* [72] se dirige a la validación de modelos que predicen la calidad subjetiva de televisión de alta definición. En este proyecto se evalúan métricas FR, RR y NR. Como resultado, dos modelos FR y RR respectivamente fueron estandarizados. Futuros trabajos planificados en el grupo incluyen una segunda fase para *VQEG Multimedia (MM-II)* que incluirá la consideración de otros aspectos más allá del vídeo en la evaluación de la calidad percibida y trabajos incluyendo contenido 3D.

La elección de una metodología particular y una escala concreta depende de los requisitos del experimento y el propósito del estudio. En cualquier caso, la realización de test subjetivos requiere una inversión en tiempo y recursos que no siempre es posible. A pesar de ello, en el ámbito de los servicios adaptativos, un campo muy específico de los servicios multimedia y con gran cantidad de condicionantes que pueden tener una influencia crucial en la calidad percibida, se han realizado varios estudios subjetivos.

2.2.3.1 Evaluaciones subjetivas de servicios multimedia.

Para evaluar la calidad percibida por los usuarios finales es imprescindible realizar estudios con usuarios reales. Su valor es doble, ya que sirven al mismo tiempo para calibrar las métricas objetivas que pretenden desbancar a los costosos experimentos. La mayoría de los estudios existentes de la calidad multimedia percibida establecen el foco en la calidad resultante del proceso de codificación de la información de audio o vídeo, y solo algunos tienen en cuenta los efectos de la transmisión con pérdidas sobre éstos. Pocos estudios analizan el servicio multimedia desde una perspectiva global, tal cual como lo percibe el usuario.

La calidad globalmente percibida de un servicio multimedia depende de muchos factores. Estos factores incluyen problemas con la transmisión y procesado y la reproducción del contenido. En algunos trabajos se estudia la experiencia de usuario *end-to-end*, pero solo considerando un número concreto de distorsiones [73] o siendo evaluado con una población pequeña [74].

Algunos trabajos exploran la calidad percibida de servicios populares como *Youtube* o *Netflix*. Mayormente centrado en el intercambio entre el retraso inicial y los eventos de *rebuffering* que aquejan al popular servicio *Youtube*, el trabajo en [75] describe estudios subjetivos realizados en laboratorio con 36 participantes, respaldado por un estudio paralelo de *crowdsourcing*. De acuerdo a sus resultados, los retrasos iniciales son preferidos por el 90% de los participantes en el estudio. No consideran aspectos como el interés en contenido en los tiempos de espera, algo determinante en este tipo de servicio. El estudio realizado en [76] incorpora además la frecuencia de los eventos de *rebuffering* como un factor. En particular los autores encuentran que la frecuencia de estos eventos es el factor de más peso en la variación del índice MOS obtenido. Su estudio con tan solo 10 participantes, está basado en la calidad temporal de vídeos HTTP. En el trabajo en [77] se analizan las interrupciones de *rebuffering* en otros escenarios, incluyendo los límites de tolerancia a partir de los cuales los usuarios evalúan el servicio como inaceptable. Un enfoque similar es descrito en [78], donde los autores analizan distintos patrones de *rebuffering* en un entorno móvil. Su estudio se limita a una sola distorsión y el número de usuarios participantes aún es menor que los estudios subjetivos realizados durante el transcurso de nuestra investigación.

Por otro lado, la influencia del contenido en las evaluaciones subjetivas no es tenida en cuenta en muchos trabajos. Los autores en [79] establecen un método para clasificar los contenidos según ciertos parámetros característicos. Esta clasificación es usada en una métrica objetiva combinada con factores como pérdidas o *frame-rate*. del mismo modo, los autores en [80] realizan un estudio teniendo en cuenta el contenido, pero basado principalmente en métricas objetivas como SSIM o VQM. Otros autores evalúan la deseabilidad del contenido y su influencia en la calidad percibida y concluyen que los usuarios valoran más positivamente los contenidos que eran más deseables [81]. Otro factor que condiciona los resultados de los análisis de calidad percibida es la posición de la distorsión analizada en la secuencia multimedia o la experiencia del usuario, que

puede afectar su percepción en el presente. Así por ejemplo los autores en [82] realizan un estudio subjetivo para evaluar los efectos del "efecto memoria" aunque enfocado en el campo de las búsquedas web.

En definitiva, hay varios trabajos relacionados explorando diferentes factores y cómo éstos afectan individualmente a la calidad percibida. Ninguno de los trabajos descritos analiza de forma conjunta todas las distorsiones y factores para realizar una evaluación de la calidad percibida *end-to-end* con un número de usuarios suficientemente representativo. En el capítulo 5 se describe el trabajo realizado en esta investigación para obtener la evaluación subjetiva del efecto combinado de varios de estos factores en la calidad global con un número de participantes superior al de los trabajos vistos.

Poniendo el foco sobre una tecnología en particular, en el campo de la tecnología escalable, y en concreto, sobre los sistemas adaptativos y sus estrategias para la adaptación se han realizado varios estudios subjetivos. El trabajo en [83] proporciona un resumen completo del estado del arte referente a estudios subjetivos y objetivos empleando diferentes técnicas de *streaming* adaptativo, incluyendo la tecnología SVC.

La mayoría de los estudios subjetivos están relacionados con evaluaciones del comportamiento de *codecs*, incluyendo así la evaluación de las diferentes capas de calidad por separado [84]. Con este enfoque, algunos autores han realizado estudios sobre el codificador escalable. En el trabajo presentado en [85], los autores explican su estudio subjetivo realizado con 20 usuarios y 12 secuencias de vídeo diferentes comparando la calidad del *codec* escalable H.264/SVC con su precursor H.264 Advanced Video Codec (H.264/AVC). El nuevo codificador escalable ofrece resultados con una calidad similar a costa de un ligero incremento en la sobrecarga de procesamiento y la complejidad del decodificador. Otros autores también llevan a cabo estudios subjetivos comparando el *codec* H.264/SVC con Xvid y H.264/AVC [86]. En su trabajo los autores utilizan hasta 5 secuencias con una población de 21 usuarios. Otros trabajos se enfocan en el ámbito de los dispositivos móviles, como unos posibles destinatarios de la tecnología adaptativa. Los autores en [87] comparan también AVC con SVC en el entorno de dispositivos móviles.

En el mismo contexto, los autores en [88] presentan un estudio subjetivo investigando los efectos de la escalabilidad multidimensional de SVC con 6 secuencias de partida y 30 usuarios. En este estudio los autores adoptan la metodología DSCQS y concluyen que las preferencias de los usuarios en cuanto a las dimensiones de escalabilidad seleccionadas en los servicios adaptativos son dependientes del contenido. Este hecho también será corroborado por nuestros estudios subjetivos con poblaciones mucho más numerosas. Los autores en [89] llevan a cabo un estudio subjetivo con 28 participantes para evaluar el impacto de la red y de los parámetros de codificación en la calidad final, usando distintas alternativas de compensación de errores en decodificación basadas en tecnología SVC. Otros se centran en las diferentes alternativas a SVC para llevar a cabo la adaptación de los contenidos, por ejemplo usando la tecnología DASH [90], con un experimento subjetivo con 24 participantes. Los resultados muestran que los usuarios

valoran más positivamente los cambios graduales en la calidad de los contenidos que los cambios bruscos.

Las metodologías subjetivas también autorizan la evaluación comparativa de distintas opciones en los sistemas escalables. Los trabajos realizados en [91] utilizan la metodología de comparación por pares (PC o *paired comparison*) para analizar diferentes opciones de escalabilidad con un resultado en términos de tasa binaria similar. Emplean en sus estudios 3 secuencias origen distintas evaluadas por 16 usuarios. En este trabajo los autores proponen también un método de interpretación de los resultados comparativos. Este nuevo enfoque de análisis es interesante y estará presente en nuestros trabajos, cuando realizamos la comparación subjetiva entre un sistema adaptativo real y un sistema convencional, desarrollada en el Capítulo 5.

Si bien estos estudios superan el número mínimo de participantes establecido por las recomendaciones de la ITU-T P.910, 15 usuarios, la representatividad de estos estudios con un número tan bajo de participantes es controvertida. Por esto, otros trabajos proponen una metodología basada en los estudios subjetivos no intrusivos, realizados en entornos domésticos y más numerosos, siguiendo la filosofía de *crowdsourcing*. Los autores en [92] han realizado un experimento entregando el contenido de la prueba, dirigida a la evaluación del impacto de la degradación de los sistemas adaptativos en películas completas, en formato DVD a 38 hogares. Con este método la población asciende a los 100 usuarios. Después de visionar los contenidos los participantes deben realizar un cuestionario. El trabajo de los autores incluye una comparativa de esta novedosa metodología con las de enfoque clásico (SS, ACR). En sus conclusiones enuncian que, según sus resultados, los usuarios acostumbrados al visionado de vídeos online son más tolerantes a los saltos de calidad en el vídeo. Este resultado ha sido corroborado por el segundo de nuestros experimentos subjetivos realizado en una población igual de numerosa pero mediante los métodos clásicos. De hecho este resultado es más valioso al proceder de un estudio de metodología tradicional, ya que los trabajos basados en *crowdsourcing* requieren por su naturaleza poblaciones más numerosas, pues están sujetos a filtrados adicionales para garantizar la robustez de los resultados [93], [94].

A la luz del estado del arte actual, podemos concluir que aún existen situaciones que precisan de evaluaciones subjetivas que describan cuantitativamente las preferencias de los usuarios en lo que se refiere a los sistemas adaptativos.

3 Herramientas, técnicas y metodología de emulación de servicios

El presente capítulo explora los conceptos de simulación y emulación, particularmente la simulación en tiempo real, que se definen más adelante. Se profundiza especialmente en los simuladores en tiempo real y las técnicas que se aplican a las herramientas existentes para mejorar aspectos como la escalabilidad o el rendimiento

En este capítulo se detalla la técnica de simulación propuesta para extender el funcionamiento del simulador NS-3 como simulador distribuido. Se analiza y compara el rendimiento de las herramientas NS-3 y OPNET en la construcción de un escenario de servicios masivos. Finalmente se desarrolla la metodología propuesta para la construcción de escenarios de evaluación de servicios masivos.

3.1 Introducción

De acuerdo a la definición básica de simulación formulada por Robert Shannon [95], la simulación es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos - para el funcionamiento del sistema. Las técnicas de simulación aplicadas en ámbito de las comunicaciones y los servicios de Internet han dado lugar a una extensa literatura que busca enfocar desde puntos de vista variados la complejidad inherente a los procesos de la transmisión de contenido en las redes públicas.

Dada la definición de simulación pura, se trata de una realidad modificada, disgregada o incompleta que busca proporcionar respuestas a determinados aspectos para los que fueron diseñadas. En la escala de realismo y de acuerdo a la definición proporcionada, la simulación es la parte más baja. En la dimensión de escalabilidad, la simulación pura se encuentra en una posición favorable. Precisamente por la versatilidad de intercambiar complejidad por escalabilidad éstos pueden suponerse conceptos "casi" intercambiables. En el ámbito de la simulación, técnicas como la simulación paralela en sistemas de memoria compartida o distribuida o recursos como la abstracción en la representación de los flujos de tráfico evaluados permite elevar la dimensión del experimento de forma sencilla. Con todas las técnicas de simulación, los límites del experimento pueden

elevarse hasta 500 millones de paquetes por segundo en redes de hasta 200.000 nodos [96].

En el otro extremo del universo bidimensional definido sobre la experimentación, las evaluaciones de sistemas reales suponen el extremo más fiable pero el menos escalable. En el ámbito de esta tesis en particular, la realidad es escalable en tanto en cuanto se incrementen paralelamente los recursos. Sin embargo, se ha optado por dejar constante el coste, dentro de un intervalo razonable, para analizar las soluciones y alternativas existentes en la literatura.

Entre los dos extremos del universo se encuentran toda una variedad de técnicas y soluciones que resuelven muchos de los problemas encontrados durante cada una de las fases de diseño, desarrollo y evaluación de un servicio multimedia. La definición más completa del concepto de emulación se encuentra desarrollada en [96]. La emulación consiste en evaluar protocolos y aplicaciones sobre un entorno de red simulado en tiempo real. Existen tres variantes diferenciadas por la forma en que se gestiona la parte simulada. La primera incluye modificaciones en el núcleo (kernel) del sistema operativo para modificar las condiciones del intercambio de la información entre aplicaciones. La segunda aproximación involucra el desarrollo de software que intercepta las llamadas al sistema para capturar la información y modificar las condiciones del intercambio del tráfico entre aplicaciones. La última de las aproximaciones, más moderna, consiste en la utilización de simuladores en tiempo real, capaces de integrar aplicaciones reales en entornos simulados. Este criterio de clasificación fue introducido también por Razan Beuran en [97]. Así pues, la simulación en tiempo real se define como la forma de emulación según la cual se permite combinar entidades reales con partes simuladas en un entorno de experimentación, de tal forma que las partes simuladas y reales comparten el mismo reloj que sincroniza el funcionamiento del conjunto [13].

Las dos primeras formas de emulación permiten la simulación de varias características, pero no admiten la representación del tráfico competente que puede haber en una red real. El retardo del tráfico evaluado y en especial su variación en función del tiempo puede depender de las interacciones con otros tráficos presentes en el entorno. En la simulación en tiempo real, flujos sintéticos pueden ser generados en el interior del simulador para competir con el tráfico real. Las condiciones del tráfico simulado son además repetibles

En el desarrollo de esta investigación, los conceptos de simulación en tiempo real y emulación han sido usados de forma separada. En particular, en el estudio comparativo de los simuladores NS-3 y OPNET se utiliza el término emulación para referirse a las técnicas de evaluación que agrupan la modificación del *kernel* y el *software* que intercepta las llamadas al sistema. Atendiendo a esta distinción, en el emulador, la parte simulada se limita a los enlaces entre los nodos o más específicamente, el comportamiento del tráfico entre dichos nodos. En este caso, los nodos son elementos discretos que carecen de un modelo de ejecución y no están incluidos en la simulación..

En cambio, se utiliza de forma genérica el término de simulación en tiempo real cuando la topología, o al menos una parte, es simulada.

Los *testbeds*, entendidos como maquetas a escala de los sistemas reales sobre los que se pretende desplegar un nuevo servicio, suponen la forma más fiable de evaluación de un servicio en desarrollo. *Testbeds* aislados desarrollados de forma específica permiten evaluar una tecnología real sobre una realidad escalada. Sin embargo, se encuentran generalmente fuera del alcance del presupuesto de el investigador medio [98]. Si se habla de *testbeds* distribuidos, la comunidad tiene a su disposición algunas soluciones que permiten compartir la infraestructura distribuida globalmente entre múltiples simulaciones simultáneas [99]. Sin embargo la limitación de una infraestructura que envejece rápido, dada la velocidad de desarrollo de la tecnología, condiciona la versatilidad de estas técnicas.

Gráficamente, las herramientas de experimentación generalmente se asientan sobre una función como en la figura 3-1. Las técnicas de mejora aplicadas sobre las herramientas existentes buscan aumentar bien el realismo o escalabilidad o, deseablemente, ambas al mismo tiempo. La simulación en tiempo real se representa específicamente de entre las variantes de emulación como una de las más relevantes, que será ampliamente utilizada a lo largo de este capítulo de la tesis.

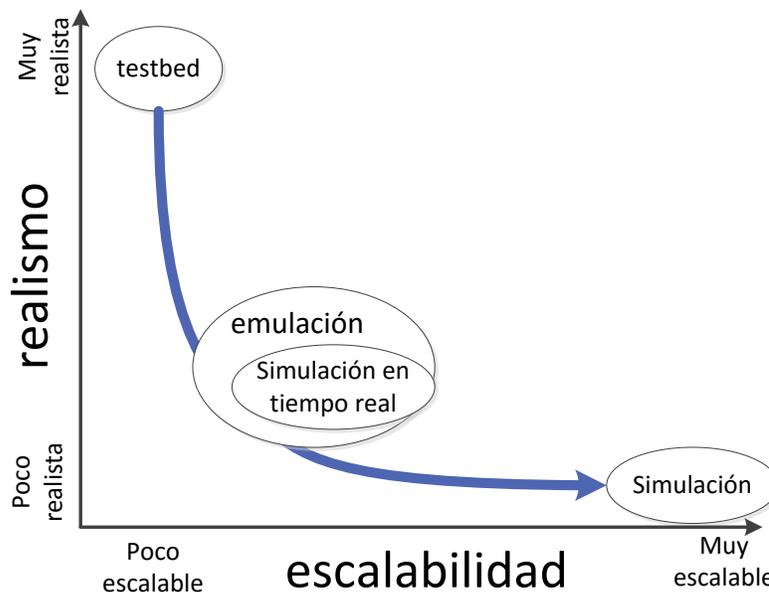


Figura 3-1: Dimensiones del universo de herramientas de experimentación

Cuando hablamos de la técnica de simulación en tiempo real para la evaluación de servicios masivos, preocupa especialmente la escalabilidad que puede ofrecer. Identificar

los rangos de operación de las herramientas elegidas va a permitir obtener resultados dentro de la zona de trabajo óptimo de las herramientas de simulación en tiempo real, evitando comportamientos anómalos de los resultados como consecuencia de la saturación de los recursos de la herramienta. La elección de la plataforma de trabajo y la configuración de los elementos adyacentes no es una tarea menor. La correcta configuración de las herramientas de trabajo y la identificación de los indicadores de su rendimiento es una forma de asegurar que los resultados obtenidos corresponden a una fiel representación de los resultados esperables en un entorno real, si el modelo ha sido construido de forma correcta.

En esta fase de la investigación se fija el objetivo de experimentar con las herramientas de simulación en tiempo real para determinar sus rangos de funcionamiento, su precisión y su habilidad para trabajar con servicios masivos en el ámbito multimedia. Las distintas herramientas existentes en la comunidad y también las disponibles en el mercado tienen ámbitos de trabajo diferentes y propósitos variados. Desde las herramientas de simulación tradicionales, a los *testbed* más completos o dispuestos mundialmente, existe un gran abanico de posibilidades disponibles para la elección de un sistema de evaluación.

Los simuladores con la capacidad de simulación en tiempo real ofrecen un punto de entrada óptimo para la experimentación rápida y económica de los distintos escenarios requeridos para la evaluación de un servicio multimedia. En particular, NS-3, el simulador libre heredero de NS-2, y OPNET, el simulador comercial por excelencia son las herramientas sobre las que se centra la investigación. NS-3 es uno de los simuladores más conocidos en el ámbito de licencias abiertas (GNU). OPNET por su parte tiene una larga experiencia en el ámbito comercial demostrando sus atribuciones en todos los ámbitos, desde el académico al militar.

La investigación en este área ha sido especialmente prolija en la proposición de herramientas y técnicas para apoyar la experimentación de servicios de cualquier índole. Muchos trabajos han propuesto nuevas herramientas o se han ocupado de su relación y clasificación para facilitar el trabajo de los investigadores, ya sea de forma genérica [28], [97] o enfocados en distintos campos, [100][101]. Otros han propuesto técnicas para extender el funcionamiento de las herramientas existentes. Sin embargo, se echa en falta una sistematización de la aplicación de las herramientas y técnicas de emulación que permitan a los investigadores comenzar a desarrollar experimentos de forma segura, obteniendo unos resultados no condicionados por una utilización incorrecta o más allá de los límites de la herramienta o técnica elegida

Es por este motivo que, tras la experiencia adquirida con las investigaciones realizadas en el campo, se propone una metodología para la construcción de escenarios para la experimentación de servicios. Esta metodología facilita la labor de creación de escenarios asesorando al investigador en las técnicas y herramientas más adecuadas en función de la definición de los requisitos iniciales del experimento.

Para evidenciar, primero el desempeño de las herramientas analizadas, y luego el beneficio alcanzado con la aplicación de la metodología propuesta, los resultados de las evaluaciones se apoyan en las métricas de servicio más relevantes.

3.2 NS-3 en la evaluación de servicios masivos. Extensión distribuida

Hoy por hoy NS-3, sucesor del popular simulador NS-2, es el simulador más conocido del ámbito libre. El simulador NS-2 ya fue pionero en la extensión a entornos emulados mediante el trabajo realizado Mahernholz et.al. [14]. NS-3 soporta de forma nativa la emulación usando un programador de eventos de simulación en tiempo real y adaptadores de red virtuales integrados en el entorno de simulación. Los denominados *TapBridge* (el puente realizado entre un interfaz físico y un interfaz virtual *software*), realizan la función de comunicación entre los nodos reales y los nodos simulados (o nodos fantasma) en el interior de la simulación. De esta forma, las aplicaciones reales pueden intercambiar tráfico real a través de una topología totalmente simulada.

Esta estructura puede configurarse de forma virtual en el interior del mismo equipo. Así, dos procesos corriendo en el mismo procesador pueden simular ser conectados a través de una topología de red compleja de forma totalmente transparente para esos procesos. Para la evaluación de servicios mayores y más complejos, se puede incrementar el número de interfaces virtuales y la complejidad de la red sin coste aparente, Figura 3-2.

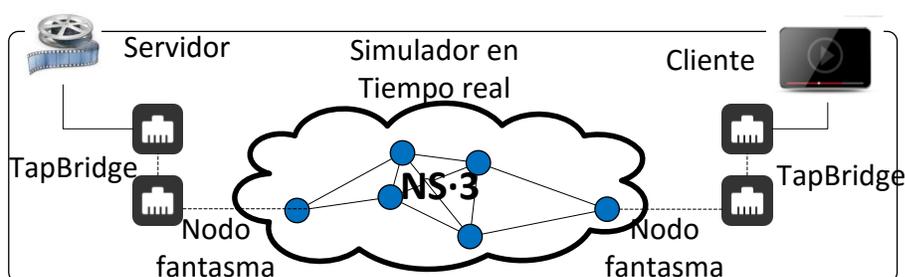


Figura 3-2: Esquema de simulación en tiempo real con NS-3

Obviamente, la capacidad de los recursos comprometidos tendrá mucho que decir en este incremento de complejidad y cantidad de elementos participantes en la simulación. Por un lado, la correcta identificación de los límites existentes ayudará a construir entornos adecuados para las emulaciones y por otro lado establece los límites a la validez de los resultados del experimento. En otras palabras, permitirá distinguir cuando los resultados son debidos al comportamiento real del servicio o cuando han sido comprometidos por los fallos de la plataforma sobre la que se están evaluando.

Analizando la solución propuesta por NS-3 para la emulación de servicios reales, y particularmente, servicios multimedia, se revelan algunos inconvenientes. En primer lugar los recursos de la plataforma son compartidos entre el propio emulador y las aplicaciones reales conectadas a la red virtual. El rendimiento de las aplicaciones condiciona el crecimiento de la envergadura del experimento.

La separación de las aplicaciones en un entorno distribuido, tal como ya se ha construido para NS-2 [14], alivia este problema. En contrapartida, surgen otros problemas nuevos que deben ser considerados. En una aproximación distribuida, todo el tráfico intercambiado por las aplicaciones debe obligatoriamente pasar por los canales adecuados de la topología virtual y la conexión entre las aplicaciones distribuidas y los nodos virtuales introduce, por razones físicas, un retardo y un límite de capacidad inherente a las líneas de conexión entre la maqueta distribuida. Por ejemplo, la emulación de una red Gigabit Ethernet es incompatible con una maqueta con aplicaciones distribuidas conectadas por medio de una topología Ethernet 10/100. Es inevitable acometer la evaluación de plataformas limitadas pues siempre va a existir una restricción que no se puede salvar aún a costa de incrementar el coste tecnológico de la solución.

En este apartado se acomete la evaluación de NS-3 como emulador. En particular, se desarrollan los mecanismos que habilitan una emulación distribuida, para más tarde realizar una evaluación comparativa del rendimiento de NS-3 en ambos escenarios. Se ha elegido un servicio de distribución de vídeo sobre una red de área como aplicación clásica con una alta demanda de recursos, suficiente para colocar a la plataforma en una situación comprometida. Entonces, el rendimiento del simulador, y el comportamiento del servicio en sí mismo son evaluados de forma comparativa entre las soluciones centralizada y la propuesta distribuida. Las métricas de rendimiento *hardware* son importantes en este ámbito, ya que se evalúa el efecto que la plataforma de emulación tiene sobre el soporte físico. Aspectos como la tasa de uso del procesador del equipo (CPU), la memoria ocupada o el estado de congestión de los enlaces son observados. Al mismo tiempo, no se debe perder de vista el objetivo inicial, la evaluación del servicio, para lo que las métricas clásicas de calidad de servicio son consideradas (tasas de paquetes perdidos, *jitter*).

3.2.1 Extensión de NS-3 para clientes distribuidos

La comunidad de NS-3 proporciona la metodología a seguir para construir un escenario de simulación en tiempo real centralizado. Construyendo la topología deseada usando el lenguaje de la herramienta, se pueden conectar las aplicaciones reales (cliente y/o servidor) a los nodos interfaz ubicados en el interior de la topología. Con el desarrollo de la solución propuesta, denominada extensión para clientes distribuidos, se habilita también el uso de otras aplicaciones ubicadas fuera de la máquina destinada a la emulación propiamente dicha.

Como el objetivo es proporcionar una integración transparente para el propio *software* emulador, la extensión utiliza túneles UDP para transportar las tramas entre el simulador y cada aplicación. La configuración no requiere operaciones complejas pero si impone una serie de requisitos. En primer lugar todos los equipos involucrados en la emulación deben disponer un la versión adecuada del sistema operativo *Linux* y el *Kernel* personalizado para esta extensión. La descripción de la maqueta o entorno de emulación y la distribución de clientes en la misma deben ser descritos en un lenguaje específicamente creado con formato XML

El *software* NS-3 proporciona dos clases de emulación distintas. Las dos aproximaciones son denominadas "*TapBridge*" y "*EmulatedNetDevice*". En particular la tecnología detrás de "*EmulatedNetDevice*" tiene como propósito permitir que objetos simulados sean ubicados en el seno de una topología real. En cambio, el concepto detrás de "*TapBridge*" tiene relación con la definición de emulación y permite la introducción de tráfico de aplicaciones reales en una topología simulada.

Los modos de operación que NS-3 habilita para el objeto "*TapBridge*" son varios. Los modos más relevantes para esta investigación son el modo "*LocalDevice*" y el modo "*BridgedDevice*". En el modo "*LocalDevice*", NS-3 permite la creación de interfaces virtuales a los que se pueden vincular aplicaciones reales. En este modo la configuración de los interfaces en gobernada íntegramente por el propio simulador. En el modo "*BridgedDevice*", NS-3 permite la conexión con interfaces virtuales, tales como "*Linux Bridges*" creados de forma externa. Este modo es especialmente indicado para el caso de conexión con máquinas virtuales como clientes, donde las máquinas virtuales pueden requerir el control de los interfaces. El modo "*BridgedDevice*" es el modo indicado para realizar una extensión distribuida, aprovechando los interfaces virtuales y siendo extendidos mediante túneles UDP hacia otros sistemas o máquinas físicas dentro de la misma maqueta. El objetivo de la extensión distribuida puede verse reflejado gráficamente en la Figura 3-3.

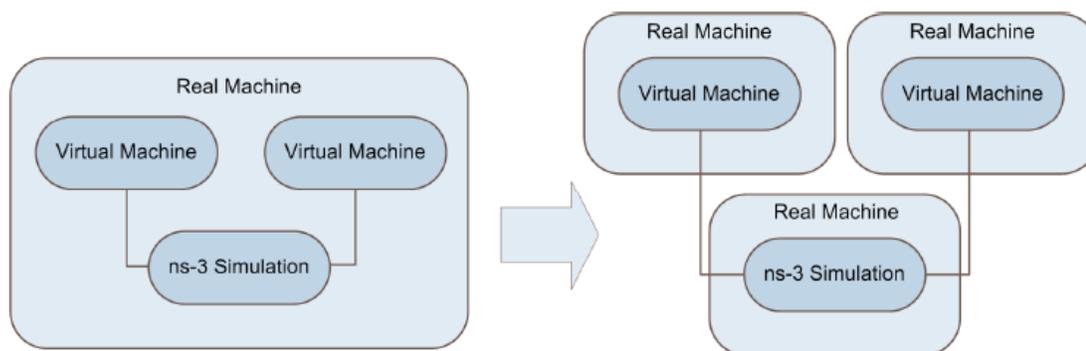


Figura 3-3: Evolución distribuida del modelo de emulación de NS-3

En el modo "*BridgedDevice*", los objetos de la simulación se conectan a un extremo de los dos interfaces interconectados por un puente o "*bridge*" de red. Estos objetos dentro de la simulación reciben en nombre de nodos fantasma. En el otro extremo del puente se conecta la aplicación o cliente virtualizado. Para asegurar una conexión transparente entre el extremo del "*bridge*" y los clientes remotos -- en la propuesta de extensión distribuida, situados en otro punto de la topología física de la red del laboratorio-- se encapsula todo el enlace en datagramas UDP. El tráfico UDP se encauza a través de la red real hasta el host remoto que alberga los clientes, donde se realiza el proceso inverso y el contenido de los paquetes de transporte es entregado a los clientes. De esta forma, para los efectos de la simulación, el nodo cliente está conectado directamente al nodo fantasma dentro de la simulación.

El proceso responsable de realizar el transporte de la información entre el nodo fantasma y las aplicaciones de cliente distribuidas es el denominado "*taptunnel*". El diseño del proceso tiene su inspiración en el trabajo realizado en [102]. En contraposición a la propuesta del trabajo de Mahrenholz, el proceso desarrollado para esta extensión utiliza un solo túnel UDP para cada cliente virtualizado remoto y no realiza ninguna transformación ni traducción de las direcciones IP de los paquetes de información original. En el trabajo desarrollado para NS-2, la traducción era un requisito debido al esquema de direccionamiento interno del simulador, sin embargo, NS-3 es totalmente compatible con el esquema de direccionamiento IPv4 con lo que las traducciones son innecesarias.

En el caso de esta extensión realizada para NS-3, al contrario que la desarrollada para NS-2, no se requieren modificaciones internas en el código de los nodos fantasma del simulador. Esta particularidad facilita la integración de esta extensión con las nuevas versiones o actualizaciones de NS-3. En la Figura 3-4, se muestra el esquema de configuración descrito para la extensión distribuida. En esta configuración, los clientes remotos son agnósticos de la configuración de la topología, por lo que actúan como clientes clásicos en una red IPv4.

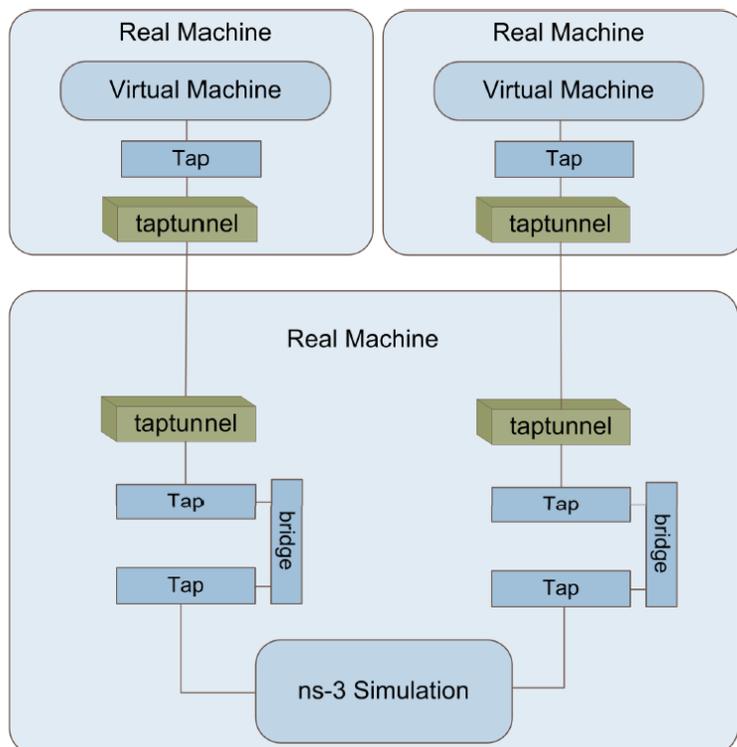


Figura 3-4: Detalle de la configuración de la extensión distribuida en NS-3

En virtud de las capacidades expuestas por la extensión distribuida desarrollada para NS-3, se permite la deslocalización de aplicaciones de cliente en hosts remotos. Estos hosts deben estar conectados de forma local (por restricciones en el direccionamiento de NS-3) al host que alberga el simulador. Además, si el consumo de recursos lo autoriza, varios clientes pueden convivir en el mismo equipo físico, por lo que puede reducirse el tamaño de la maqueta física de laboratorio para el mismo número de aplicaciones en la topología emulada.

La extensión distribuida desarrollada no se limita al uso de clientes en sistemas operativos completos y permite también emplear técnicas de virtualización. Tecnologías de virtualización, como VMWare⁴ o *User Mode Linux* (UML)⁵, son compatibles con la extensión distribuida. Los usuarios

En función del nivel de virtualización elegido por el usuario, hay algunas particularidades de la comunicación entre las aplicaciones de cliente cuando se disponen varias en el mismo host físico. En el caso en el que las aplicaciones que se ejecuten sobre un

⁴ <http://www.vmware.com/es>

⁵ <http://usermodelinux.org/>

sistema operativo completo con varios interfaces de red virtuales deben ser capaces de elegir el interfaz adecuado, el que corresponda según la configuración de la topología simulada. Para esto las aplicaciones de cliente pueden modificar la opción "SO_BINDTOEVICE" en la creación de los sockets de comunicaciones. Sin embargo, en muchos casos no es posible acceder al código fuente de la aplicación de cliente para modificar su comportamiento. Para estos casos, la librería "*libsocktap*" desarrollada para el emulador NEMAN [103] permite vincular los sockets a interfaces sin modificar el código fuente de la aplicación cliente. En el otro extremo, si las aplicaciones se encapsulan en un sistema virtualizado, bien sea completo o a nivel de red exclusivamente, no es necesario ningún mecanismo de selección de interfaz, ya que a todos los efectos el proceso cliente sólo conoce un interfaz.

Un problema añadido a la convivencia de varios interfaces virtuales, configurados en una red local, dentro del mismo host físico, es que deben evitar la comunicación directa entre ellos. Este es un comportamiento por defecto impuesto por el sistema operativo, ya que en la mayoría de las situaciones es la forma correcta de proceder. Por el contrario, en esta configuración se desea forzar el tráfico de los interfaces virtuales a través del simulador remoto. Considerando que, generalmente, el sistema operativo Linux es la plataforma preferida para la construcción de maquetas de redes, en los equipos destinados a albergar esta funcionalidad puede aplicarse un parche al Kernel del sistema operativo para habilitar la selección del camino por defecto en este tipo de comunicaciones. El parche "*sendtoself*" desarrollado por Ben Geagr habilita este particular comportamiento [104]. Con la nueva opción de configuración añadida al aplicar el parche al kernel, denominada "*loop*", se puede elegir que los interfaces virtuales localmente conectados fuercen el envío de datos hacia el exterior. Consecuentemente, las tablas de rutas deben considerar los distintos orígenes para operar de forma correcta. Las reglas de enrutamiento deben ser cuidadosamente escritas en el proceso de configuración de la extensión distribuida para que el tráfico se clasifique y reparta apropiadamente [105].

Incluso considerando el parche y la correcta configuración de las reglas de enrutamiento, existe un problema cuando dos clientes que pertenecen a dos redes IP distintas conviven en el mismo host. La consulta a la tabla de rutas para el encaminamiento de paquetes entre redes distintas tiene una prioridad superior a la consulta de tabla de rutas para los interfaces definidos localmente. Es por esto que en condiciones normales, los paquetes que se envían con destino una interfaz virtual local, pero en el otra red IP distinta, encuentran primero la tabla que define los interfaces virtuales y, por tanto, ignoran la ruta por defecto establecida en la tabla de enrutamiento remoto. Para sobrepasar este límite es necesario una modificación adicional en el kernel del sistema operativo Linux. Modificando la prioridad de la tabla de rutas locales en el código fuente del kernel se permite anteponer las reglas de encaminamiento externo antes que las locales.

Con las modificaciones realizadas en el kernel del sistema operativo Linux y las herramientas que manipulan las rutas de encaminamiento se permite redirigir el tráfico de

los clientes desplegados en la misma máquina en cualquier configuración hacia una topología emulada en el cliente remoto, de forma transparente para las aplicaciones, por mediación de los túneles UDP. Sin embargo, esta extensión también impone una serie de condiciones que deben ser tenidas en cuenta al trabajar con ciertas aplicaciones.

En el caso de que la aplicación bajo estudio requiera el uso de protocolos dinámicos de encaminamiento, típicos, por ejemplo, de redes móviles ad-hoc, la gestión de las tablas de enrutamiento es parcialmente diferente. Una de las características intrínsecas de los protocolos de encaminamiento dinámico es que las rutas en los nodos varían constantemente, lo que entronca frontalmente con la definición estática de rutas definida para la operación de la extensión distribuida. Los clientes distribuidos en este tipo de aplicaciones deben tener su propia pila de protocolos (*TCP/IP stack*), independiente de la pila de protocolos del host físico. Los protocolos de encaminamiento dinámicos serán aplicados sobre los interfaces virtualizados de cada cliente. La virtualización puede articularse del modo deseado, desde una virtualización completa como la anteriormente mencionada solución de VMWare o UML, hasta herramientas más ligeras que permiten virtualizar únicamente la pila de protocolos de red (VirtNET⁶).

En otros casos, existen escenarios de evaluación que requieren un medio compartido. El concepto de medio compartido supone que varios nodos en la simulación comparten el canal de comunicaciones. Es el caso por ejemplo de la simulación de un escenario inalámbrico con varios nodos conectados a un punto de acceso. Debido al encapsulamiento de los paquetes de datos en datagramas UDP utilizado en la extensión distribuida, se produce un efecto de multiplicación de paquetes que debe ser tenido en cuenta. En el mundo real, cuando un nodo envía un mensaje, todos los nodos en su rango de alcance (el dominio de *broadcast*) reciben el mismo paquete, pero el canal solo se ocupa una vez, con el mensaje original. Sin embargo, en el entorno emulado, un nodo envía un mensaje éste es encapsulado y enviado primero al nodo que lo representa dentro de la topología emulada (nodo fantasma), ocupando una vez el medio físico de conexión entre el cliente remoto y la máquina que alberga el simulador. Tras el debido procesamiento, el nodo reenvía una copia del mensaje a todos los nodos que pertenecen al mismo dominio de *broadcast*. En consecuencia, el tráfico generado por un nodo que incide sobre la red física de transporte entre los nodos y el simulador es multiplicado por el número de vecinos del nodo origen en la topología simulada. Este es un efecto colateral que tiene que ser considerado en la fase de dimensionamiento de la plataforma para simulación en tiempo real. El ancho de banda de la red que conecta al simulador con los nodos de clientes remotos debe ser suficiente para soportar la carga extra generada por esta situación. El problema puede ser aliviado añadiendo inteligencia en los nodos simulados de forma que se compruebe que cada paquete se envía sólo al nodo destino, una solución similar a la adoptada en [102]. Sin embargo, no se ha considerado

⁶ http://www.ntkernel.com/?Products:Network_and_Security:VirtNet

esta opción ya que modifica abiertamente el comportamiento de los nodos alejándolos de su comportamiento real. Por ejemplo un nodo inalámbrico que, aun estando en el dominio de colisión del paquete enviado, no escucha el paquete enviado, dando por desocupado el canal. En la Figura 3-5 se describe gráficamente este concepto. Un único mensaje enviado en la topología simulada (a), implica varios mensajes en la red física que soporta la simulación (b).

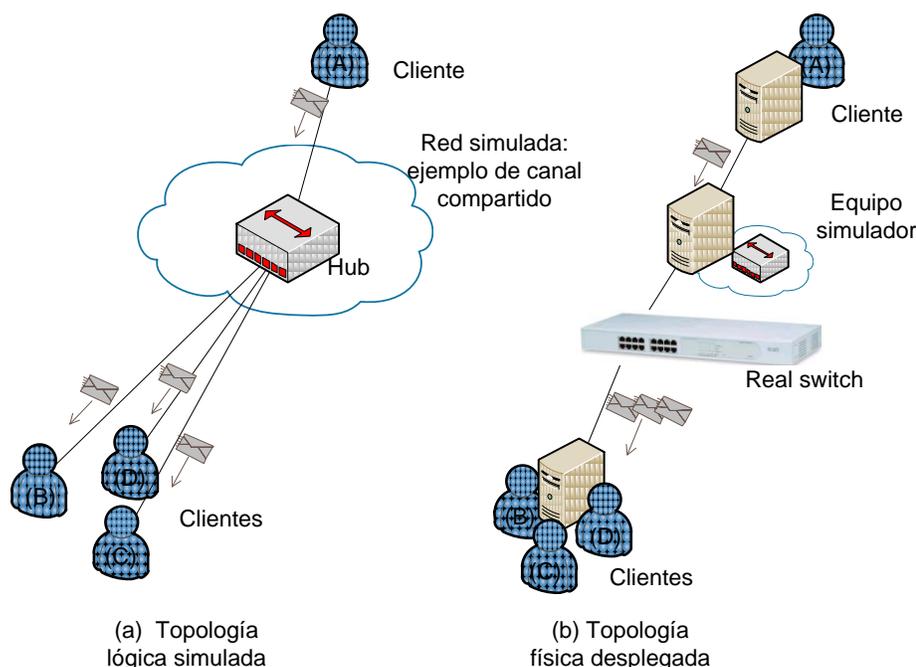


Figura 3-5: El problema del medio compartido en la extensión distribuida.

3.2.2 Evaluación: Caso de estudio distribuido frente a centralizado

Existe un interés creciente en la técnica de simulación en tiempo real para la evaluación de servicios. En este caso se ha elegido NS-3 para ejemplificar las ventajas y limitaciones de ésta técnica. Además, mediante el desarrollo de la solución distribuida se habilita un mecanismo más para comparar el rendimiento de las distintas técnicas de evaluación. Mediante el análisis del caso de estudio se cumplen dos objetivos. Por un lado se verifica la operación del sistema diseñado de distribución de aplicaciones en equipos físicos distintos del que alberga el simulador. Al mismo tiempo se contrasta el funcionamiento de NS-3 como simulador en tiempo real, con la identificación de los indicadores de rendimiento que garantizan unos resultados de emulación no comprometidos por fallos de rendimiento de las herramientas utilizadas en la emulación.

Para el análisis de la extensión distribuida y NS-3 como simulador en tiempo real se ha seleccionado un caso de estudio clásico, compuesto por un servicio de distribución de

vídeo sobre una red de área local. En esta topología, un servidor de red se conecta a un número creciente de clientes a través de redes locales y conmutadores de nivel 2 (*switches*) en una red Ethernet 10/100 Mbps. La topología de red, compuesta por dos niveles de conmutadores de nivel 2, está simulada utilizando la tecnología NS-3. En NS-3 se configuran también los enlaces como canales CSMA (*Carrier Sense Multiple Access*) con 16 nanosegundos de retardo y un caudal límite de 100Mbps, lo que correspondería a un cableado UTP categoría 5e de 3 metros de longitud.

Existen razones de peso para la elección del caso de estudio como tal. En primer lugar es muy factible que existan implementaciones de este mismo caso de estudio en la realidad. Por ejemplo una empresa que ofrece a sus empleados un servicio formativo a través de la red interna de su sede. Si el número de empleados es alto, el servicio se torna de alta demanda y el rendimiento se ve comprometido. Los encargados de su gestión pueden desear evaluar el servicio antes de desplegarlo. Desde el punto de vista técnico, otra razón para elegir un servicio de vídeo *streaming* es la alta demanda de recursos de este tipo de aplicaciones lo que permite estresar más fácilmente el sistema para encontrar las limitaciones en la plataforma de emulación que con otros servicios. Además es totalmente viable replicar la topología en una maqueta para comparar el rendimiento del servicio emulado con un servicio completamente real. Finalmente, los clientes de este tipo de servicios ejercen un consumo de recursos elevado, lo que pone de manifiesto rápidamente la diferencia entre la extensión distribuida y la alternativa de emulación estándar. La escalabilidad puede ser evaluada en dos direcciones, incrementando el número de clientes por cada conmutador o incrementando el número de conmutadores.

Para ofrecer la comparativa entre la extensión distribuida propuesta y la aproximación clásica se realizan un experimento compuesto de dos escenarios. En el primero de los escenarios, se configura una simulación en tiempo real clásica, denominada "*single-machine*", Figura 3-6. En esta configuración toda la topología se encierra en el mismo *host* físico. En el segundo escenario se aplica la extensión distribuida desarrollada en este capítulo para crear un escenario "*distributed-machine*", donde la topología simulada (Figura 3-7 (a)) se distribuye en una topología física distribuida (Figura 3-7 (b)). Como se adivina en la figura, los clientes se reparten entre varios equipos clientes, la simulación se distribuye a otro equipo y el servidor a un tercero, conectados por una red dedicada.

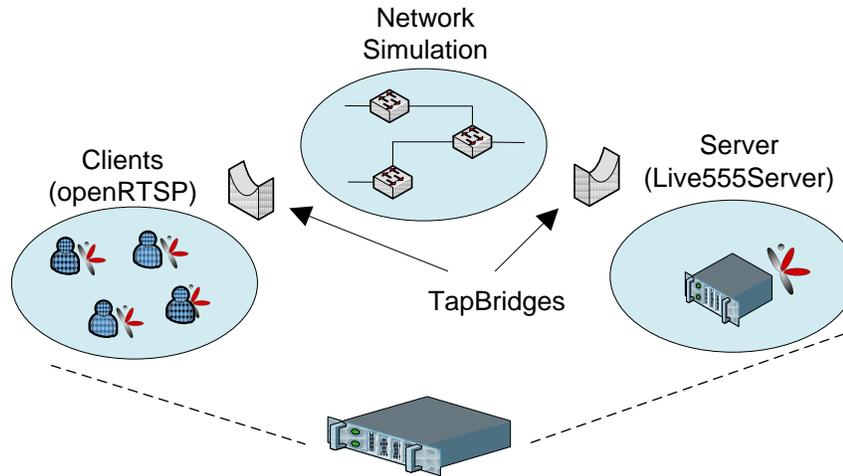


Figura 3-6: Configuración del experimento "single-machine"

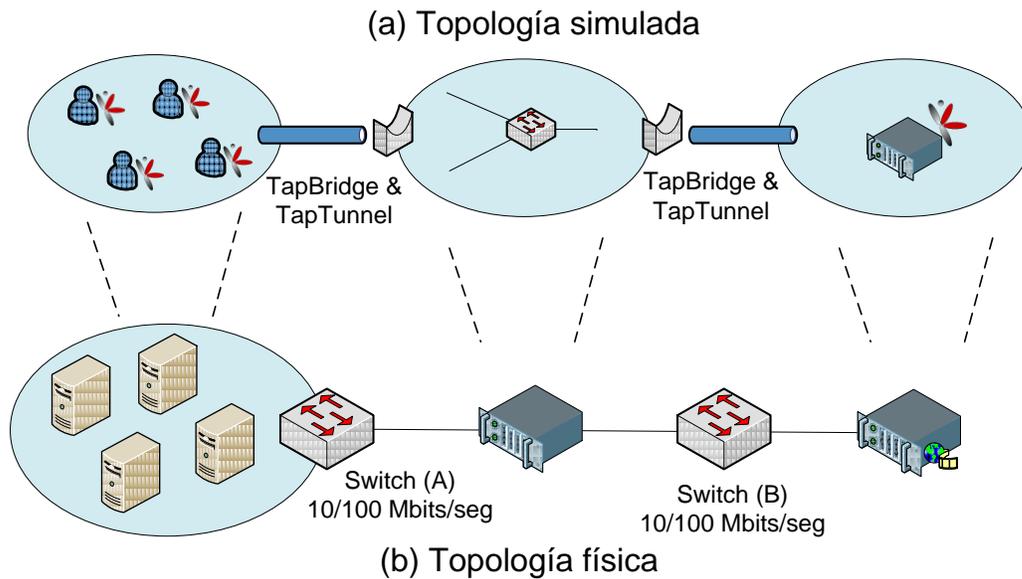


Figura 3-7: Configuración del experimento "distributed-machine"

Ambos escenarios son estresados mediante la inclusión progresiva de nuevos clientes del servicio bajo estudio. Ambos escenarios utilizan la misma la versión del simulador NS-3 (versión 3.5). El servicio se compone de un servidor (*live555*) y varios clientes (*openRTSP*). Todos los componentes *software* utilizados son de libre acceso, accesible a cualquier que tenga intención de replicar el experimento. El protocolo de transmisión del

servicio de vídeo es RTSP/RTP y el vídeo de muestra empleado (*highway_qcif.yuv*), extraído de una conocida librería de contenido⁷, está codificado con MPEG-2 a dos tasas diferentes. La secuencia denominada "*Low-Quality*" (*Low*) es el resultado de codificar el contenido a una tasa de 250 Kbps y mismo contenido codificado a una tasa de 1Mbps da lugar a la secuencia denominada "*High-Quality*" (*High*). Las secuencias tienen una longitud de 79 segundos. Dado que el objetivo es la evaluación del rendimiento del escenario y el servicio de vídeo a nivel del tráfico de red que genera, para evitar el consumo de recursos de procesamiento de vídeo, los *frames* del vídeo solicitado son descartados en cliente. Cada experimento de 120 segundos se repite 3 veces para garantizar la estabilidad de los resultados.

Combinando los escenarios descritos anteriormente, *single-machine* (*S-M*) y *distributed-machine* (*Distr.*) con las distintas calidades de codificación, *Low* y *High*, se obtienen cuatro experimentos, resumidos en la Figura 3-8.

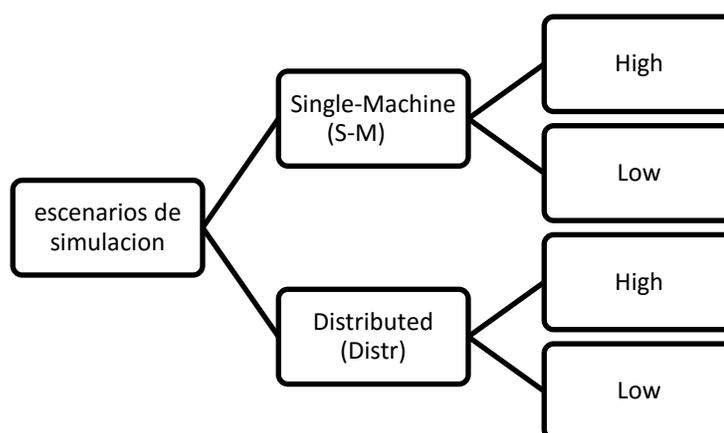


Figura 3-8: Desglose de experimentos

3.2.2.1 Configuración de los experimentos

En el experimento del escenario *S-M* se utiliza un equipo físico Dell PowerEdge 860 con un procesador Intel Xeon Dual Core a 2.40GHz, 1Gb RAM y dos interfaces de red Ethernet Gigabit. El sistema operativo es Ubuntu Server 8.04 (32-bit). Los detalles del *hardware* y *software* del entorno de evaluación se detallan en la Tabla 3-1.

Tabla 3—1 Características de la plataforma de evaluación

⁷ <http://trace.eas.asu.edu/yuv/>

Modelo (CPU, MEM)	Sistema Operativo.	Simulador
DELL r210 (double core duo @2.4GHz, 8Gb)	Ubuntu 9.10	NS-3 v. 3.8
DELL r210 (double quad core@2.3GHz, 16Gb)	RedHat Enterprise Linux 4.0	OPNET v. 14.5

El procedimiento de ejecución es idéntico para todos los escenarios. En primer lugar el simulador en tiempo real es instanciado, creando los interfaces virtuales. Los procesos de cliente son vinculados al interfaz virtual correspondiente (*tap*). En el interfaz (*tap*) que en la topología representa el servidor. Los clientes y el servidor estarán entonces conectados a través de la topología de red diseñada en el simulador NS-3. Los resultados de la realización son monitorizados a nivel de red a través de un *sniffer* (*tcpdump*) y a nivel de proceso a través de los analizadores del sistema (*sar* y *pidstat*). En las evaluaciones realizadas se presta especial atención a la sincronización de las herramientas de medida para no incurrir en errores frecuentemente encontrados en las medidas realizadas sobre la red [106].

En el escenario distribuido, la configuración del equipo que contiene el simulador es la misma. El mismo equipo *hardware* es utilizado para el experimento completo en el caso del primer escenario y exclusivamente para el simulador en el segundo de los experimentos. Además en este segundo escenario se usa la configuración diseñada en la extensión distribuida para NS-3. A través de la extensión desarrollada se conectan los nodos simulador con los clientes reales ubicados en otros equipos físicos dentro de la misma maqueta de red del laboratorio. Los clientes se distribuyen en una granja de equipos de uso general (Pentium III con 512 Mb RAM y un interfaz Ethernet). El Sistema Operativo de estos equipos es Ubuntu Server 8.44 (32-bits). Para ubicar el servidor de vídeo se utiliza un equipo Dell PowerEdge idéntico al usado para albergar el simulador NS-3. Todos los equipos de la maqueta física de red sobre la que se asienta el escenario están conectados por una red Ethernet 10/100 dedicada de corta distancia.

Dado que los recursos consumidos por los clientes no son elevados, un mismo equipo físico puede contener varios clientes lógicos. Esto permite escalar rápido el número de clientes. Dado que varios clientes comparten el mismo interfaz de red, éste puede modelarse como un conmutador, de ahí que la topología simulada para el experimento "*distributed-machine*" solo contenga un nivel de conmutadores, en contraste con la topología representada para el caso "*single-machine*". Los conmutadores del nivel inferior están representados por los propios interfaces de red de los equipos distribuidos.

En este nuevo escenario, las condiciones del experimento se adaptan para reflejar fielmente las mismas condiciones ambientales consideradas en el escenario simple. Los clientes son iniciados siguiendo la misma secuencia, y los monitores de recursos son reubicados para recoger la misma información ahora distribuida entre distintos equipos físicos. Los equipos son sincronizados entre sí para garantizar una similitud elevada entre los dos experimentos, de tal forma que los resultados obtenidos son comparables.

3.2.2.2 Discusión de resultados

Para cada experimento se ejecutan 3 realizaciones idénticas y se promedian los resultados de las tres realizaciones. En cada realización se introduce un número incremental de clientes y se monitorizan las métricas de rendimiento elegidas.

Se han observado dos comportamientos indeseados del simulador NS-3 durante la realización de los experimentos. En primer lugar, a medida que se incrementa el número de usuarios en el escenario, es más probable que la simulación se aborte antes incluso de iniciarse. Con más de 30 clientes por escenario, el número de errores se incrementa progresivamente de tal forma que, con número más altos, se torna inviable obtener resultados válidos (o finalizar la simulación). El número máximo evaluado es 60.

La justificación conocida para este error se refiere a la gestión de los recursos utilizados por la herramienta. Cada nodo *TabBridge* se gestiona mediante un hilo independiente y la concurrencia de todos agota los recursos del sistema. El error se produce sin embargo en el proceso de construcción del escenario, antes incluso de introducir tráfico real de los servicios externos que sobrecargue el sistema. En la Tabla 3—2 se recoge el error mostrado en los casos de fallo encontrados.

Tabla 3—2 Ejemplo del error al simular escenarios grandes en NS-3

```
Program received signal SIGSEGV, Segmentation
fault.
[Switching to Thread 1216416080 (LWP 31319)]
0x00002aae1ffd2491 in
std::_Rb_tree_rebalance_for_erase ()
from /usr/lib/libstdc++.so.6
```

Otro de los problemas detectados con el simulador, es que aunque crea muchos hilos independientes para gestionar los nodos *TapBridge*, todos los hilos utilizan el mismo procesador, no permitiendo la distribución de la carga entre los dos procesadores físicos disponibles en el equipo utilizado. Este es un impedimento claro al crecimiento de las simulaciones.

Para continuar aumentando el rendimiento de simulación por esta vía habría que analizar soluciones al problema de la gestión multi-hilo. La comunidad de NS-3 es consciente de esta limitación y existe un proyecto que persigue una solución para este límite [107]. En el momento de la realización de esta parte del trabajo, el proyecto no estaba concluido por lo que no se pudo evaluar la mejora de rendimiento con las mejoras introducidas en esta propuesta.

Para comprender el comportamiento de la simulación, analizamos en primer lugar una realización temporal del experimento con 20 clientes comparando los dos escenarios "*single-machine*" y "*distributed-machine*" para el vídeo de alta calidad (*High*). Como era esperable, existe un ahorro en el consumo de recursos de CPU en el escenario

distribuido. En ambos experimentos y a excepción de la diferencia en el valor medio, el patrón de uso de la CPU es muy similar y está muy relacionado con la codificación del vídeo. Hay, sin embargo, un pico de tráfico al inicio del experimento, en torno al segundo 8, mucho más marcado en el experimento *S-M*. Este no es un comportamiento esperable en el servicio. Su aparición parece estar relacionada con el consumo de CPU puntual del escenario *S-M* en ese instante, superior al 70%. El servidor ve afectada su capacidad de envío porque coexiste el proceso servidor con el proceso de simulación y no es hasta que el simulador reduce su consumo de CPU, que el servidor tiene oportunidad de enviar la información atrasada. Este efecto resalta la importancia de vigilar celosamente las métricas de rendimiento físico antes de extraer conclusiones sobre las métricas del servicio.

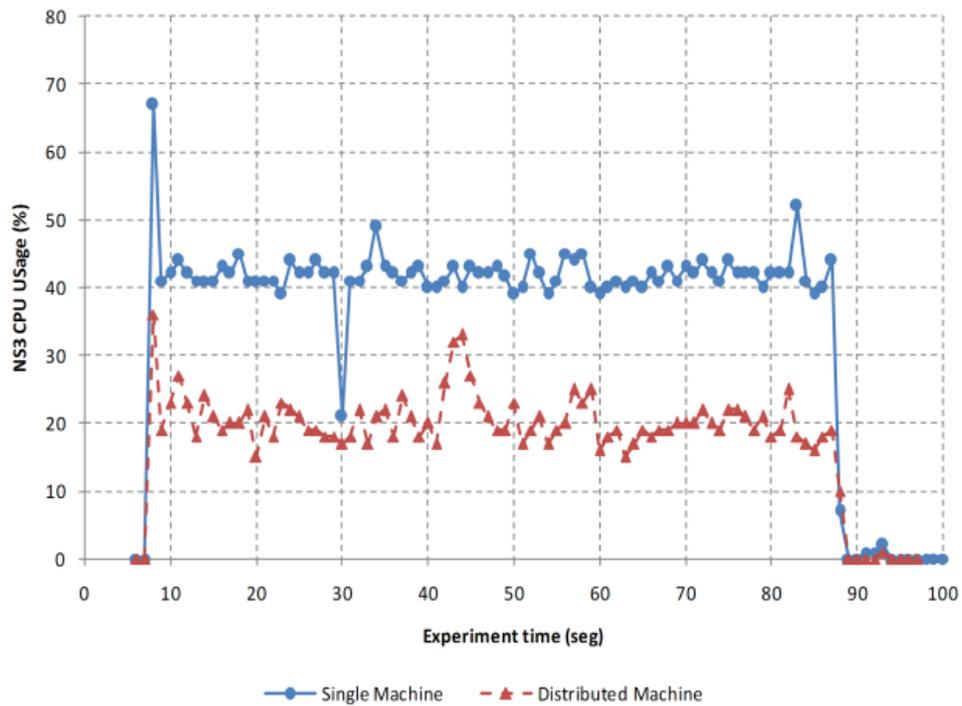


Figura 3-9: Consumo de CPU en una realización temporal de los experimentos con NS-3 (20 clientes, *High*)

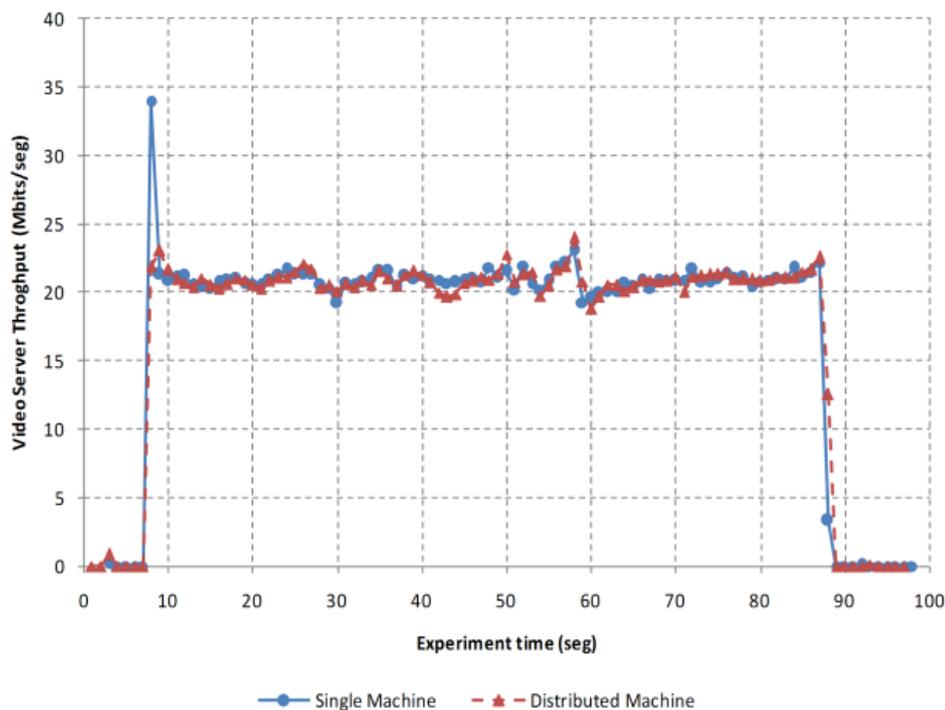


Figura 3-10: Caudal en una realización temporal de los experimentos con NS-3 (20 clientes, *High*)

En la Figura 3-11 se representa el valor promedio de utilización de CPU en función del número de clientes en el experimento para las 3 realizaciones realizadas de cada experimento. La relación entre ambos es prácticamente lineal. Esto no es una sorpresa ya que a medida que aumenta el número de clientes, sube el número de paquetes y por tanto el consumo del simulador. Hay una reducción significativa del consumo de CPU para el caso *Distr.*, aunque el tráfico involucrado en la simulación es el mismo para ambos casos, según la Figura 3-12. La diferencia del consumo del proceso NS-3 entre ambos evidencia que la gestión del interfaz no es igual en ambos casos. Para el escenario *S-M*, el simulador gestiona íntegramente los interfaces, mientras que en el caso *Distr.*, inyecta tráfico directo a un interfaz externo.

A la vista de los resultados la implementación del modelo de trabajo en NS-3 para el segundo caso, cuando inyecta tráfico directo a un interfaz real, es más eficiente que cuando NS-3 es quien crea y gestiona el interfaz virtual. Por estos motivos, el número máximo de clientes alcanzados en las distintas configuraciones es diferente. En el caso distribuido, ligeramente mayor. Los errores comentados para el caso *S-M* hacen imposible finalizar los experimentos con muchos clientes. Se aprecia un efecto de saturación por encima de 40 clientes en ambas figuras en el caso *Distr.* El motivo de la bajada de rendimiento está relacionado con los cambios de contexto entre la gran cantidad de hilos creados para el proceso. Una vez más, este no es un efecto esperable

en este servicio en situación real, donde el tráfico se debería incrementar de forma proporcional al número de clientes. Hemos encontrado por tanto un límite de rendimiento no avisado por el simulador.

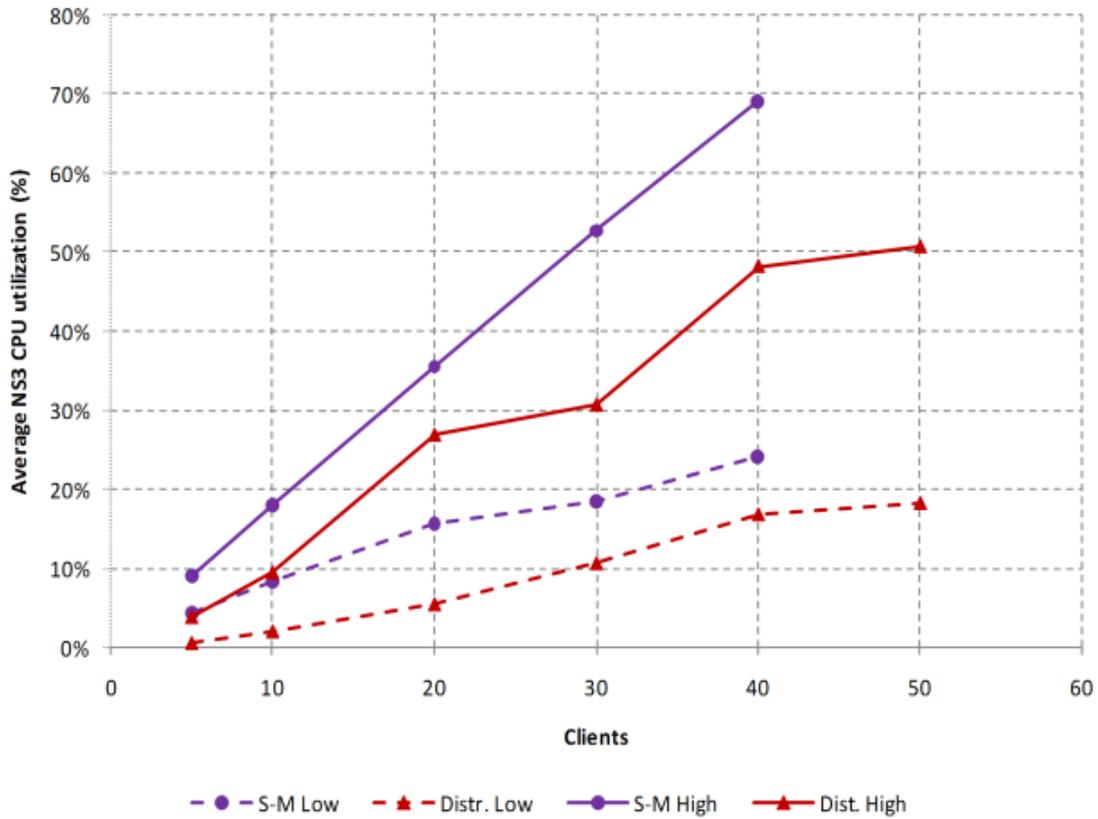


Figura 3-11: Evolución del consumo global de CPU en los experimentos con NS-3

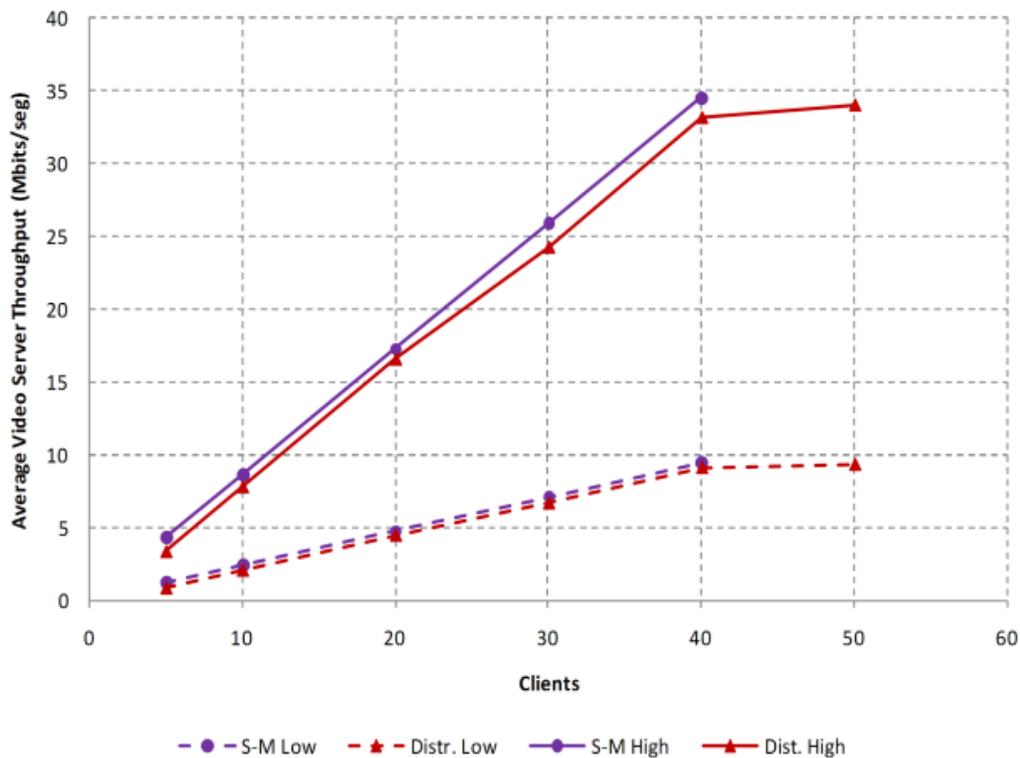


Figura 3-12: Evolución del caudal en los experimentos con NS-3

Finalmente se aporta el consumo de CPU general de la máquina. La Figura 3-13 incorpora el consumo de todos los procesos de sistema, incluido el simulador. Es especialmente relevante para analizar el complemento al consumo de recursos que implican los procesos de audio/vídeo en el caso de la simulación *S-M* y el proceso diseñado en este trabajo "*taptunnel*" para habilitar la extensión distribuida. Los clientes y el servidor de vídeo elegidos para estos experimentos son considerablemente ligeros, en el sentido de que no suponen una carga excesiva de procesamiento. Además el vídeo en los clientes no se almacena ni reproduce, sino que se descarta. Por el contrario, los procesos *taptunnel* reenvían toda la información que reciben, lo que supone una carga de trabajo para CPU considerable. Para el caso del vídeo de baja calidad, el proceso *taptunnel* eleva el gráfico de consumo global del equipo del escenario distribuido sobre el consumo de toda la emulación para el caso *S-M*. Sin embargo para el caso de vídeo de alta calidad, el elevado consumo del simulador pesa más sobre el uso intensivo de CPU de los procesos *taptunnel* del escenario gemelo distribuido.

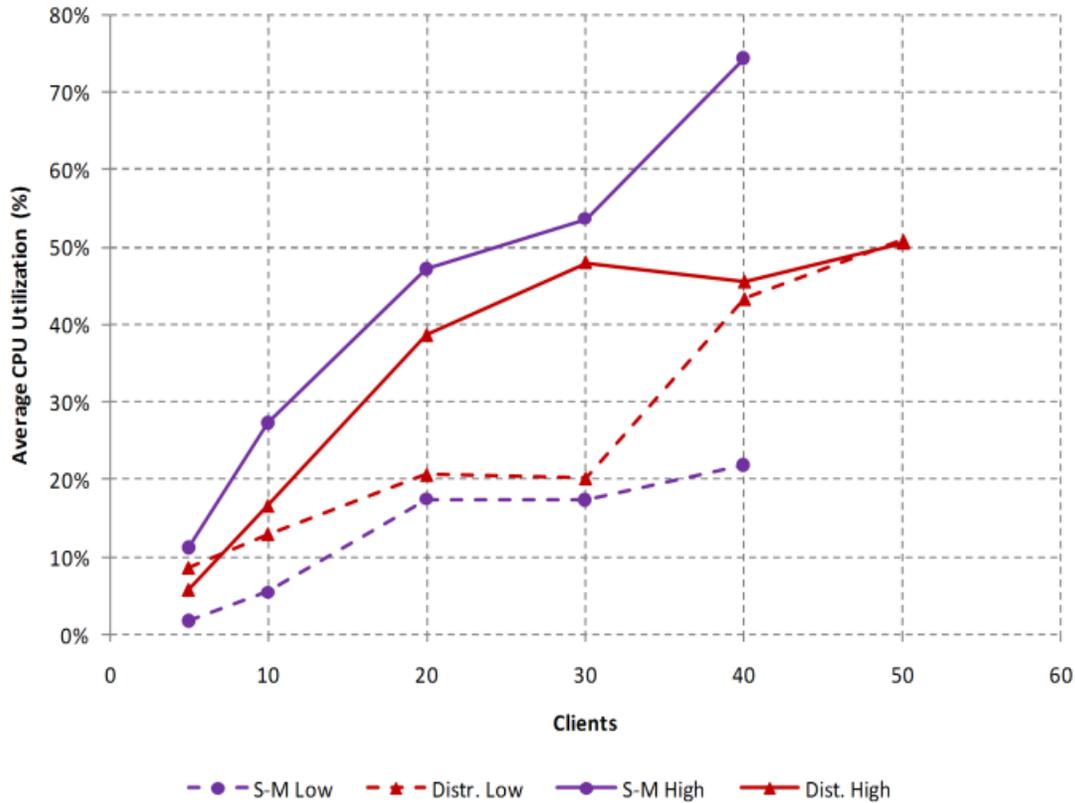


Figura 3-13: Evolución de consumo del equipo en los experimentos con NS-3

Si la cantidad de tráfico en la red se mantiene en límites bajos, una simulación en el modelo *S-M* puede ser mejor opción que la versión *Distr.* En cambio, si las aplicaciones consumidoras fuesen de un consumo elevado de recursos, ya no se podría mantener esta premisa.

Pasando a observar las métricas de servicio, se hace especial hincapié en la métrica del *jitter*. El *jitter* de los paquetes RTP es una métrica muy significativa para los servicios multimedia. Para un número fijo de clientes, la Figura 3-14 representa la variación media del retardo para un cliente al azar. El primer efecto apreciable es la falta de puntos para el caso *S-M*, consecuencia de los fallos en las simulaciones que se mencionaron al inicio de este apartado.

Es importante destacar cómo el *jitter* es siempre inferior en el escenario distribuido, para cualquier calidad del vídeo y/o número de clientes, tal como se observa en la Figura 3-14. En un primer momento este resultado está fuera de lo esperado, ya que al introducir una red externa adicional en el caso distribuido, con un proceso que encapsula y desencapsula paquetes, aparece un retardo adicional, aunque pequeño. La variación de este retardo también podría verse afectada y sin embargo la mejora del rendimiento, en términos de reducción de consumo de CPU, parece favorecer al *jitter* que se reduce para

el escenario distribuido, compensando el retardo adicional por la red externa. El efecto del *jitter* en los servicios multimedia constituye una importante métrica del servicio, cuyo desempeño condiciona enormemente el desarrollo de protocolos y aplicaciones. En los capítulos siguientes haremos hincapié en el tratamiento que estas técnicas de experimentación hacen del *jitter*.

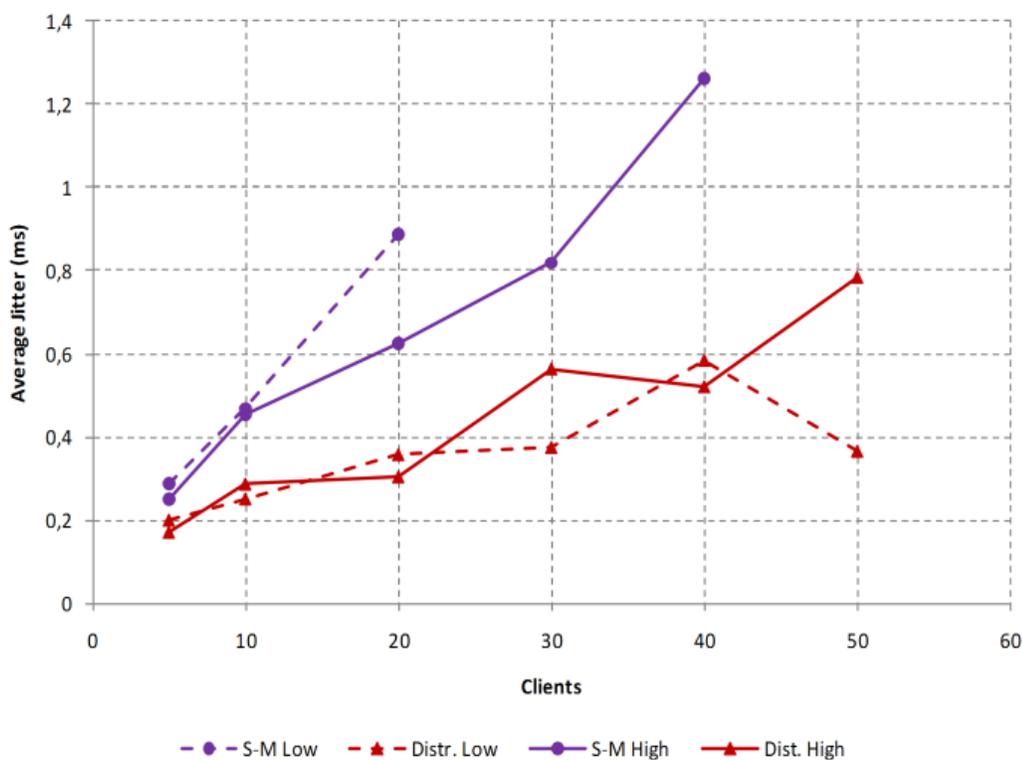


Figura 3-14: Evolución del *jitter* en los experimentos con NS-3

A la vista de los resultados presentados hasta el momento, a mayor consumo de CPU en la máquina que alberga el simulador, más comprometidos se ven los resultados. Ya que el *jitter* en el servicio modelo empleado en estos experimentos es esperable que sea bajo, se considera que los escenarios distribuidos ofrecen resultados más verosímiles. El factor limitante en NS-3 no está directamente relacionado con el tráfico sino más bien con el número de nodos. Este hecho revela una gestión ineficiente de los objetos *TapBridge* por parte de NS-3. Otros simuladores utilizan otras tecnologías para el intercambio de tráfico entre entidades reales y virtuales, por lo que la investigación debe continuar explorando otras posibilidades.

3.3 Límites de NS-3 y OPNET para la simulación en tiempo real.

En virtud de los primeros trabajos de investigación realizados en el universo de las herramientas de emulación, especialmente centrados hasta ahora en el popular simulador NS-3, se ha experimentado cómo la escalabilidad está muy ligada a la fuerza bruta. La capacidad mostrada por NS-3 está directamente relacionada con la capacidad física del equipo, y particularmente, un único procesador, que gestiona toda la arquitectura simulada dentro del escenario emulado. Esto por una parte es un problema que, a priori, tiene una solución fácil (el incremento de la capacidad del equipo a base de coste) pero no supone una alternativa válida en muchas situaciones.

Sin abandonar el estudio de NS-3 como herramienta base para las simulaciones en tiempo real, introducimos una alternativa comercial, el popular simulador OPNET Modeler (ahora parte de Riverbed y renombrado a Riverbed Modeler)⁸. Después de analizar de forma independiente las capacidades de ambas herramientas como simuladores en tiempo real, construiremos un escenario que permitirá ejemplificar las posibilidades combinatorias para sacar el mejor partido de cada una de las herramientas. Se trabajará además en mantener controlada la inversión en infraestructura de soporte de los experimentos de forma que las conclusiones obtenidas sean de aplicación a infraestructuras basadas en *commodity hardware*. El trabajo descrito en esta parte ha sido recientemente publicado en [108].

La técnica de simulación en tiempo real es habilitada en OPNET gracias a la inclusión del módulo *System in the Loop (SITL)*. Este módulo ha sido, y es, ampliamente utilizado en los campos tanto académico como industrial desde hace más de 15 años [109]. Particularmente en la industria militar este módulo se utiliza ampliamente [110], [111]. Un esquema básico del concepto de simulación en tiempo real con OPNET se muestra en la Figura 3-15.

⁸ <http://www.riverbed.com/products/steelcentral/opnet.html?redirect=opnet>

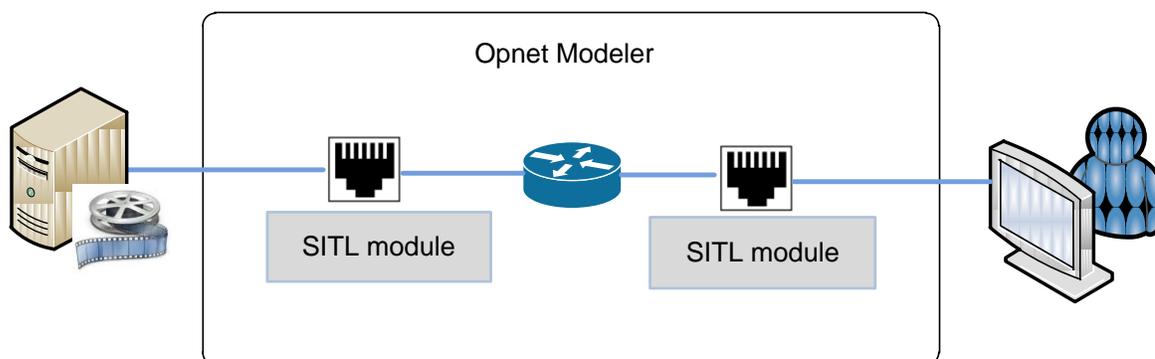


Figura 3-15: Esquema de simulación en tiempo real con OPNET

OPNET es el simulador comercial que goza de una gran popularidad dada su longevidad y su versatilidad [112]. El módulo SITL ubica a OPNET a la vanguardia de los simuladores en tiempo real. Aprovechar la potencia de su núcleo con la posibilidad de extraer el tráfico simulado al mundo real y viceversa supone una ventaja competitiva importante respecto a otros simuladores, NS-3 por ejemplo con un recorrido menor.

En el contexto de los servicios multimedia, ambas herramientas, tanto NS-3 como OPNET ofrecen una versátil solución para realizar simulaciones en tiempo real. El estudio individual del rendimiento de OPNET como herramienta de simulación en tiempo real, tal como se hizo anteriormente para NS-3, es necesario para sentar las bases necesarias que permitan construir un experimento combinado con ambas herramientas.

En esta ocasión y como complemento a los trabajos desarrollados en la primera parte del capítulo con NS-3, se incorpora una nueva técnica de comparación de los resultados del experimento. Aparte de controlar los indicadores de rendimiento en la simulación en tiempo real, las métricas de calidad de servicio son comparadas con los resultados obtenidos en una composición puramente emulada, según el concepto definido por los autores en [113]. La emulación pura se realiza mediante la simulación de enlaces entre nodos virtualizados. A diferencia de la simulación en tiempo real, se ahorra la sobrecarga de la simulación de los modelos de ejecución de los nodos y el trabajo de virtualización de los paquetes de datos. En su lugar se utiliza el modelado de colas para emular el comportamiento de la infraestructura directamente sobre el servicio real completo [114]. Como ésta técnica no está limitada por el rendimiento de la herramienta, se puede utilizar como referencia del comportamiento esperado del servicio y detectar en qué momento los resultados de simulación en tiempo real están condicionados por el rendimiento del simulador.

Es por esto que retomamos el ejemplo de NS-3 simulado en tiempo real con la topología red de acceso (LAN tipo edificio), completándola con su contrapartida emulada. A continuación se hace una configuración similar pero no idéntica para OPNET. No se utilizan topologías idénticas en ambos experimentos ya que las condiciones intrínsecas

de ambos entornos no permiten su comparación en condiciones equitativas. De hecho se compara el rendimiento de los escenarios de cada herramientas de forma individual con su contrapartida emulada. A continuación se aprovecha la combinación de ambas herramientas para construir un escenario global mayor, combinando la red de acceso tipo edificio más la red de núcleo o *backbone*, y explorar las capacidades globales.

El objetivo particular de esta sección es evaluar la viabilidad de la simulación de un servicio multimedia en tiempo real para un escenario clásico. Una particularidad importante es que continuamos manteniendo el objetivo de adaptarse a una plataforma *hardware* de simulación basada en equipamiento de propósito general, conocido como "*commodity hardware*".

Para replicar una red compleja compuesta por topologías radicalmente distintas en las partes de acceso y núcleo, configuramos distintos simuladores en tiempo real para cada tipo de topología, Figura 3-16. Aprovechando la flexibilidad y la automatización facilitada por el interfaz de programación (API) de NS-3, se complementa el trabajo del apartado anterior con la simulación en NS-3 de varias redes de acceso, con la topología ya vista. Se usa un marco referencial basado en la emulación pura de enlaces entre interfaces virtuales creadas dentro de un equipo, imitando la topología simulada en NS-3.

Para la parte del núcleo de la red, con una topología clásica en anillo, se ha elegido como plataforma de simulación OPNET. La capacidad de OPNET de introducir flujos de tráfico sintéticos para contender con el tráfico real emulado, resulta útil para evaluar los servicios en el contexto de una red de núcleo congestionada.

Los primeros experimentos están encaminados a determinar el rendimiento de las herramientas elegidas en el contexto de la topología seleccionada, cada una en su propio campo. De esta forma se espera obtener los límites por debajo de los cuales los resultados del experimento son confiables.

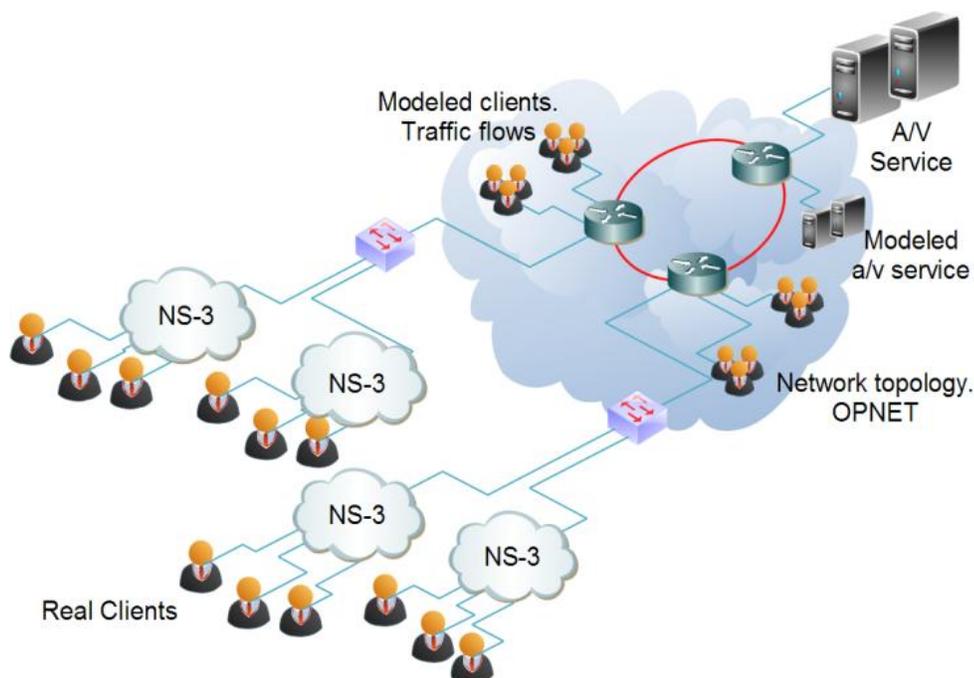


Figura 3-16: Topologías del caso de estudio combinado

El servicio multimedia evaluado se compone de un servidor de vídeo formado por dos servidores físicos (Darwin Streaming Server 5.5.5) conectados a un balanceador de carga real. Los clientes son en este caso VLC v.1.0.2. La secuencia de vídeo⁹ ha sido codificada con MPEG4 - Part 2 a una tasa de 1Mbps, 24 *frames* por segundo y un tamaño de 640x260 (SD).

Los resultados de las métricas de *hardware* (CPU), QoS (PLR, *jitter*) y QoE (MOS) son comparadas entre los escenarios simulados en tiempo real y su contrapartida emulada. Se entiende que los resultados de la simulación serán correctos mientras que mantengan un grado importante de similitud con los resultados de la misma topología emulada.

3.3.1 Red de acceso NS-3

En la topología de la red de acceso, similar a la desarrollada en el apartado anterior, se utilizan un total de 5 conmutadores de nivel 2 (*switches*), 4 de los cuales imitan a los *switches* de planta y uno más para interconectar las plantas hacia fuera. Nuevamente los objetos *TapBridge* son empleados para conectar la topología simulada con el entorno real.

⁹ <http://www.bigbuckbunny.org>

En su alternativa emulada se utilizan *Linux Bridges* aislados mediante la herramienta de virtualización ligera *Linux Containers* (LXC) para crear interfaces aisladas, que se conectan mediante herramientas de conformado de tráfico (*Traffic Shaping tools*). En esta configuración, los nodos *Linux Bridge* permiten replicar el comportamiento de un conmutador y las herramientas de conformado de tráfico limitan el caudal de los enlaces al mismo valor al que se configura en la topología simulada.

Ambos escenarios, que llamaremos R-T simulado y emulado, son evaluados mediante la inyección progresiva de sesiones del servicio multimedia, que se distribuyen de forma automática y equitativa entre los conmutadores de la red de acceso.

A medida que se incrementa el número de flujos en el sistema simulado en tiempo real, se aprecia cómo sus resultados empiezan a diferenciarse de los obtenidos por la representación emulada. En este punto sabemos que estamos alcanzando el límite físico impuesto por la herramienta de simulación en tiempo real, en este caso NS-3.

Tal como ya se experimentó anteriormente, el *jitter*, medido en los paquetes reales no se ve afectado por la simulación en tiempo real. En cambio sí que es esperable que en un servicio real se modifique, cuando se alcanzan niveles de congestión. En el escenario simulado en tiempo real sí se observa un efecto de dispersión de los valores de *jitter*, Figura 3-17.

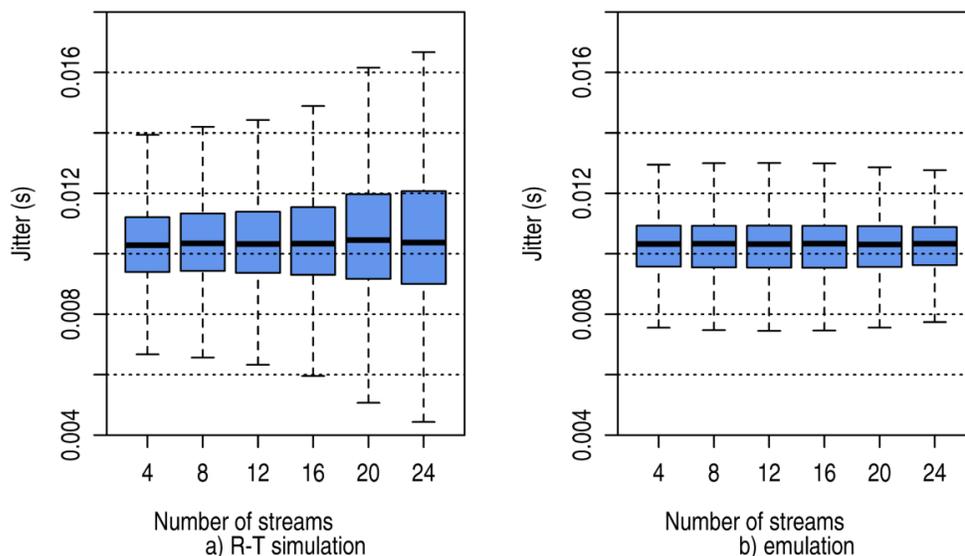


Figura 3-17: *jitter* en NS-3 simulado en tiempo real versus emulado

Un incremento rápido del PLR es detectado cuando el número de flujos en el escenario simulado en tiempo real supera ligeramente la veintena. En una observación más

detallada, ya a partir de 10 flujos se observa un PLR de 0.01%, Figura 3-18. Las métricas de QoE, en este caso el MOS usado como mapeo de la PSNR, muestran detalles específicos de los servicios multimedia que las métricas de calidad de servicio no destacan tan notablemente. Aunque el PLR no se considera realmente relevante hasta que llega al 0.1% (con 16 flujos), lo que coincide con un consumo del 80% de CPU, los resultados de una secuencia de MOS indican problemas antes de llegar a ese límite. La métrica MOS clasifica frames en la escala de 1 a 5 siendo 5 la máxima calidad. Según se muestra en la Figura 3-19, por encima de 12 flujos comienzan a aparecer un número significativo de *frames* de vídeo dañados, por tanto con calidades inferiores a 5. Por el contrario, en el escenario emulado, todos los *frames* mantienen la calidad máxima, por lo que no se incluyen en la figura.

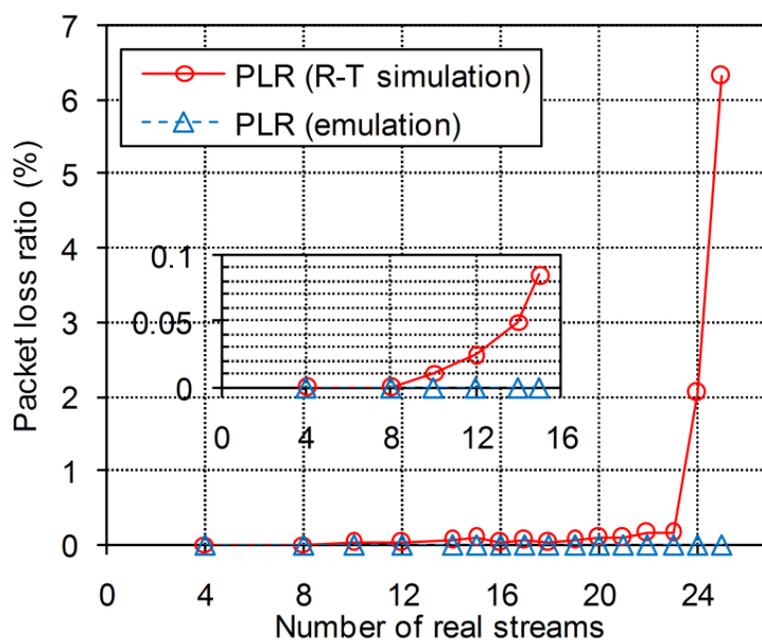


Figura 3-18: PLR en NS-3 simulado en tiempo real versus emulado

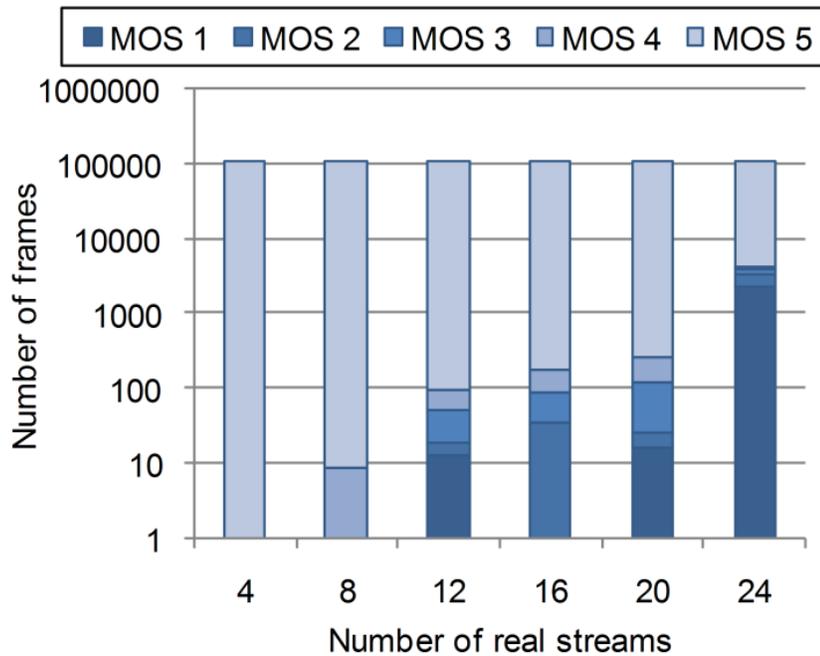


Figura 3-19: MOS en NS-3 simulado en tiempo real

Claramente, con todos los indicadores considerados, una topología de estas características simulada en tiempo real sobre el *hardware* definido y con NS-3 no es capaz de asumir más de 10 flujos, o los resultados observados en el servicio objetivo del estudio se verán comprometidos. Las métricas de memoria o eventos por segundo en el simulador no son relevantes ya que no aportan información sobre la bajada de rendimiento del simulador.

3.3.2 Backbone OPNET

En el caso de OPNET se utiliza una topología típica de redes de "*backbone*". La topología consiste en un anillo de red Gigabit con una dirección de tráfico determinada, en este caso en el sentido de las agujas del reloj. La topología simulada se conecta a los extremos reales mediante instancias del objeto SITL. Las aplicaciones reales y el extremo servidor son conectados a los *routers* simulados mediante los módulos SITL de forma distribuida a lo largo de los distintos puntos del anillo.

En la contrapartida emulada, se conectan varios equipos físicos mediante links reales Gigabit y mediante tablas de rutas se configura una dirección única para el tráfico idéntica a la de la topología simulada. Al igual que en todos los escenarios evaluados, se inyectan de forma progresiva flujos del servicio real. Empezando por 16 flujos y terminando por 72, se van disminuyendo los pasos de incremento a medida que nos acercamos a la región de saturación de la herramienta de simulación.

Las diferencias en PLR para esta topología y herramientas se aprecian por primera vez a partir de 60 flujos, donde se alcanza el límite de PLR del 0.01% como se ve en la Figura 3-20.

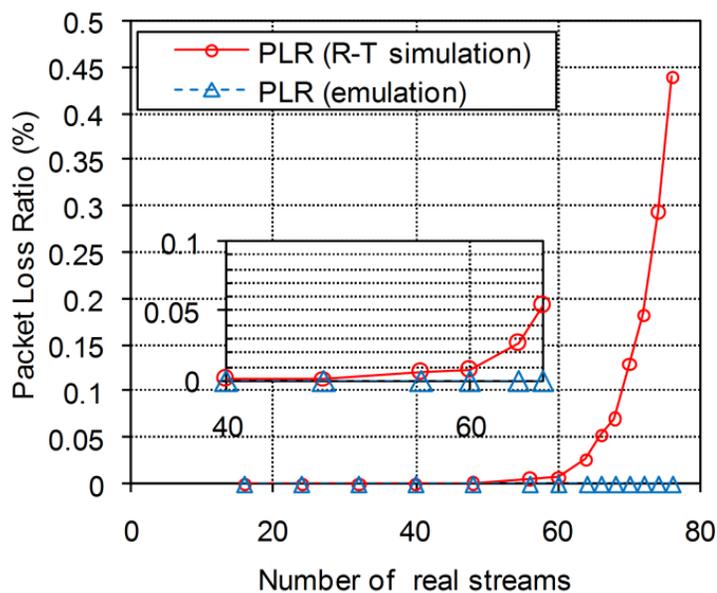


Figura 3-20: PLR en OPNET simulado en tiempo real versus emulado

El *jitter* observado en el servicio real, a diferencia de lo que ocurre con NS-3, no sufre variación alguna. No se altera de ninguna forma el comportamiento del *jitter* en el experimentos simulado con respecto al emulado, tal como se muestra en la Figura 3-21. Una vez más la memoria y los eventos por segundo programados en el simulador no se alteran a pesar de la bajada de rendimiento evidente en las métricas de servicio.

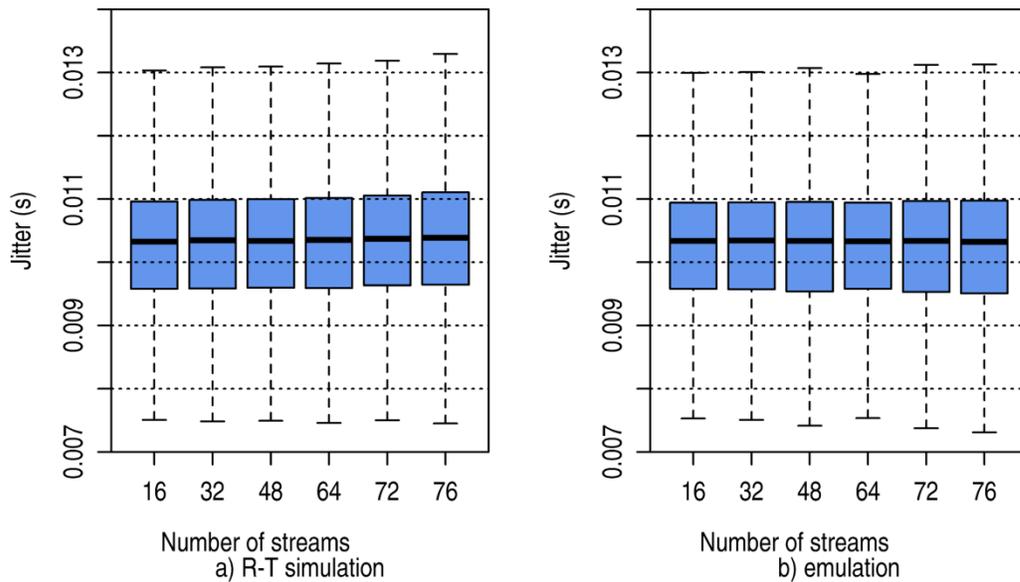


Figura 3-21: jitter en OPNET simulado versus emulado

Nuevamente la métrica del MOS es la más sensible de las elegidas. Por encima del 50% de CPU (en una sola CPU), lo que equivale a 48 flujos, comienzan a clasificarse *frames* entre los niveles más bajos de MOS, lo que significa un deterioro de la imagen decodificada que se debe exclusivamente a un rendimiento deficiente de la plataforma de evaluación (conjunto simulador-entorno), según se aprecia en la Figura 3-22. Los valores de MOS para los *frames* del caso emulado permanecen en el máximo grado de su escala independientemente de la carga (hasta 76 flujos al menos), por lo que no se incluyen en la figura.

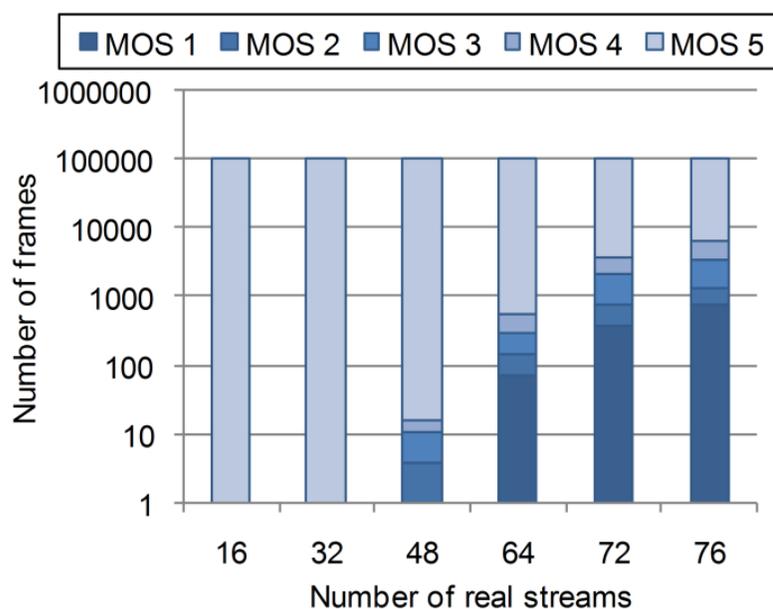


Figura 3-22: MOS en OPNET simulado versus emulado

Considerando todas las métricas en conjunto, con el simulador, la topología y plataforma de simulación seleccionados, el límite admisible se fija en 60 flujos. Por encima de este valor los resultados de la evaluación del servicio estarán terriblemente afectados por el rendimiento de la plataforma tornándose inválidos.

3.3.3 Escenario conjunto

La construcción de un escenario compuesto de simulación debe tener en cuenta los límites independientes de cada herramienta. Para aumentar el alcance y sobreponerse a ciertos límites de las herramientas pueden aplicarse técnicas como la desarrollada por el autor en el caso de NS-3 y la extensión distribuida. Este tipo de distribuciones permiten separar partes de la simulación de forma transparente sin más que configurar adecuadamente los interfaces con el mundo real proporcionados por la herramienta en particular (*TapBridge* en NS-3 y *SITL* en OPNET). La distribución puede aplicarse también a nivel de proceso, dada la incompatibilidad de los simuladores en tiempo real con el procesamiento multi-hilo.

Combinando los dos escenarios analizados de forma separada hasta ahora se construye una nueva simulación con 6 procesos de NS-3 representando 6 redes de acceso de tipo edificio, conectados a través de un único proceso de OPNET representando el núcleo de la red, con la topología en anillo ya analizada. De esta forma fijamos el número de sesiones de la aplicación real en 60. En la Figura 3-23 se representa la topología simulada en tiempo real vinculada a la arquitectura *hardware* que compone el entorno de evaluación. La red propia de la maqueta de simulación en tiempo real está compuesta de enlaces dedicados entre la granja de clientes, el equipo que alberga la simulación del

núcleo de la red y el servicio representado por dos servidores balanceados. Una red paralela e independiente se emplea para labores de control y monitorización durante la realización de los experimentos.

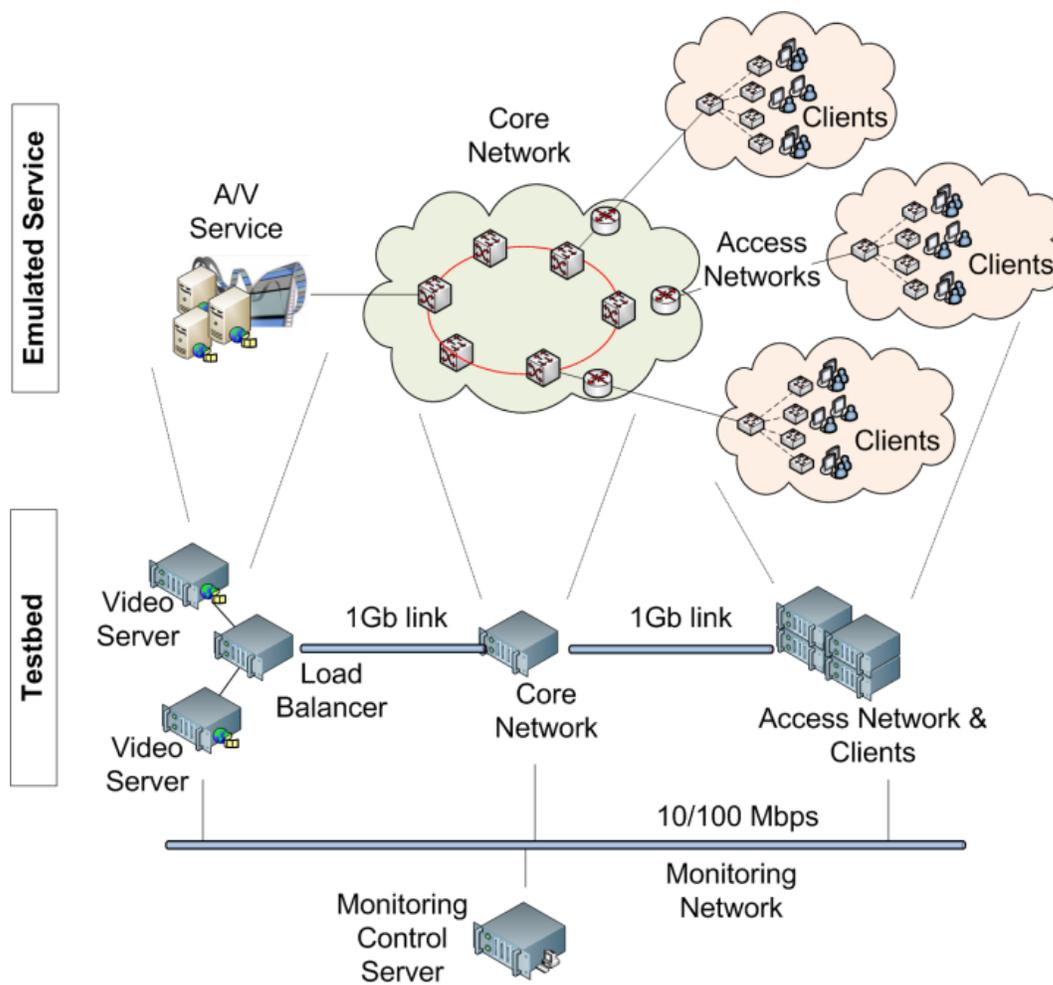


Figura 3-23: Escenario compuesto de simulación en tiempo real sobre arquitectura del *testbed*

Para añadir tráfico de fondo en el núcleo de la red que compita por recursos con los flujos reales atravesando la topología simulada, hemos añadido en OPNET flujos sintéticos modelados a semejanza de los reales, mediante las opciones que proporciona el simulador. Se añaden flujos de forma progresiva hasta que se aprecian los efectos de la congestión sobre el servicio real. Este es un recurso que permite evaluar el servicio en tiempo real ante situaciones de congestión que no se puede alcanzar simplemente con flujos del servicio real, ya que antes de llegar a este límite los resultados se ven comprometidos por el rendimiento de la plataforma y los límites ya encontrados.

Los resultados de esta simulación demuestran que al incrementar el número de sesiones en tráfico de fondo dentro del núcleo de la red se aprecian efectos negativos sobre las

métricas de QoE. En esta ocasión se incluye, además de la clasificación de *frames* en la escala MOS, la métrica SSIM. El incremento por progresivo de los flujos de fondo perjudica seriamente al servicio real (60 flujos) cuando se alcanza la cifra de 1048 flujos sintéticos. En el último experimento las métricas de MOS y SSIM muestran una mejoría ficticia que es consecuencia de la disminución de *frames* recibidos en el servicio real (obsérvese la escala logarítmica de la Figura 3-24).

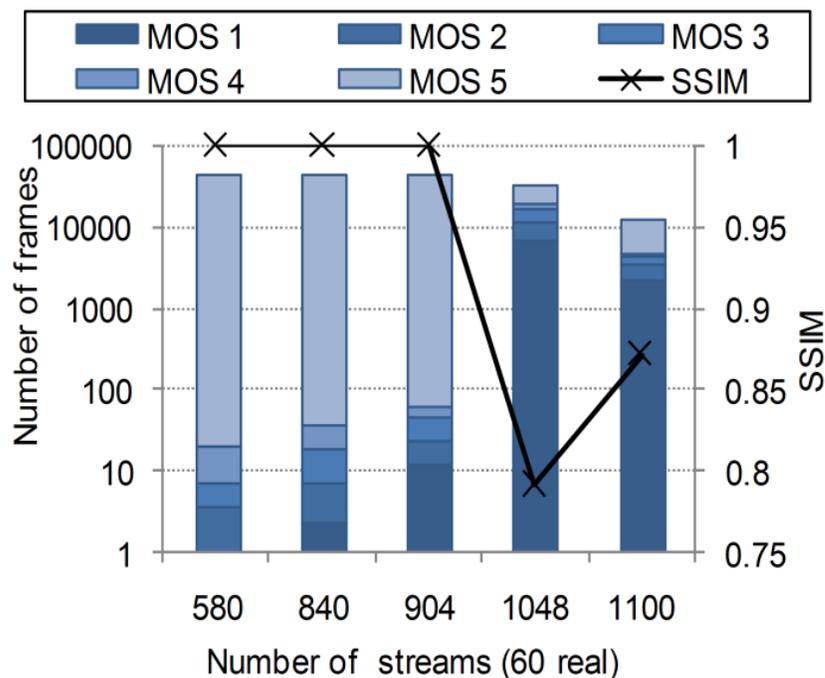


Figura 3-24: Métricas de QoE medidas en el servicio simulado en tiempo real

Al haber pre-establecido los límites físicos de cada herramienta, podemos confiar en que los resultados observados en el servicio no están afectados por una deficiencia en el rendimiento de ningún componente de soporte del servicio. Sin embargo, es recomendable continuar monitorizando las métricas de rendimiento de la plataforma a lo largo de los experimentos. En este caso, la carga de CPU en los experimentos se mantiene dentro de los límites admisibles, reduciéndose cuando el servicio real se congestiona y por tanto es incapaz de iniciar nuevas sesiones de vídeo, Figura 3-25. De hecho la métrica de eventos por segundo en ambos simuladores se reduce paralelamente a la reducción en consumo de ciclos de CPU, Figura 3-26.

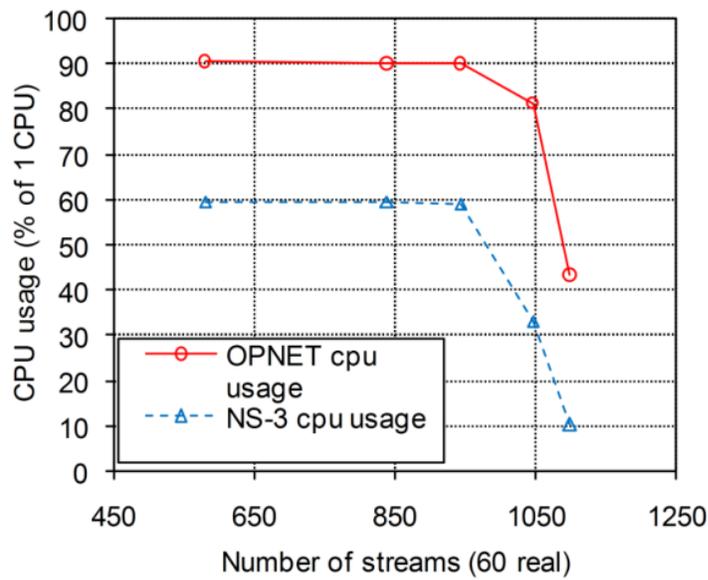


Figura 3-25: Métricas de rendimiento en la simulación en tiempo real, CPU

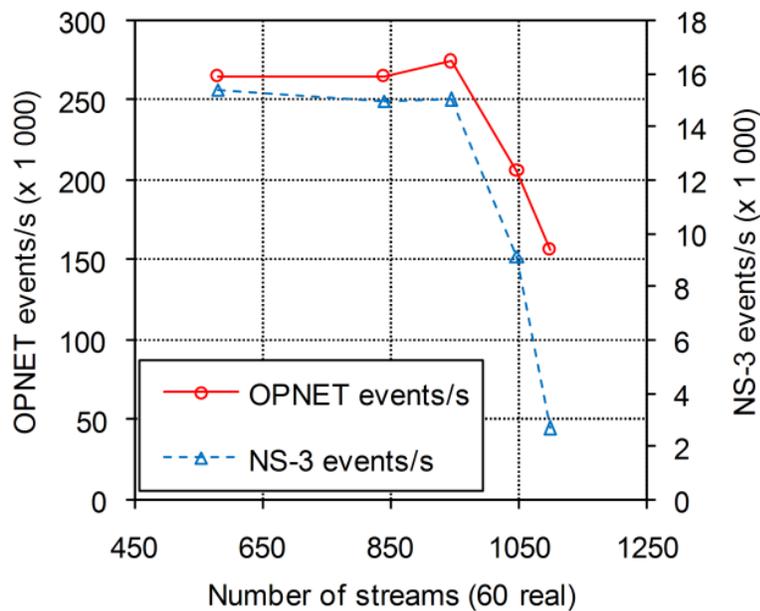


Figura 3-26: Métricas de rendimiento en la simulación en tiempo real, Eventos por segundo

Ha quedado evidenciado que sí es posible aprovechar el conocimiento del rendimiento de las herramientas para construir un escenario de simulación en tiempo real viable para un elevado número de flujos de tráfico o clientes. Es una cuestión de ajuste y de la correcta elección de las técnicas de escalado aplicables en este tipo de evaluaciones.

3.4 Metodología para la construcción de escenarios de simulación

Parece claro que la simulación en tiempo real es una alternativa interesante para la evaluación de servicios. Otras técnicas de simulación pura, emulación o directamente evaluación de servicios en entornos reales (escalados o no) complementan el universo de posibilidades para la evaluación.

Sin embargo, no hay una sistematización del proceso que el investigador interesado debe seguir para llegar a construir el entorno de evaluación de servicios adecuado para sus necesidades. Nosotros mismos hemos llegado a construir un escenario compuesto por distintas herramientas y técnicas, pero no ha sido un proceso formal y desde luego no directo. Los errores en el proceso, errores fruto de la inexperiencia, retrasan el desarrollo.

Nuestra propuesta es suplir esta necesidad con el desarrollo de una metodología de cuatro fases diseñada específicamente para la construcción de entornos de experimentación de servicios. Partiendo de un conjunto de objetivos para la evaluación de un servicio masivo (SEvice Under sTudy, o SEUT), esta metodología guía al investigador hacia los pasos a tomar para lograr un entorno fiable que ofrezca resultados controlados y realistas.

Como apoyo al proceso de la metodología, se han seleccionado las técnicas más relevantes, extraídas del análisis de patrones comunes entre el estado del arte de la materia. Cuando es posible, estas técnicas ayudan al investigador a hacer un reparto más eficiente de los recursos disponibles hasta lograr cumplir los objetivos de sus requerimientos.

3.4.1 Diseño de la metodología

La metodología propuesta consta de 4 fases. Especificación del sistema bajo estudio (*SEUT specification*), caracterización de los requerimientos del entorno (*testbed requirements characterization*), asignación de los recursos (*testbed resources allocation*) y análisis del experimento (*experiment analysis*), como se indica en la Figura 3-27.

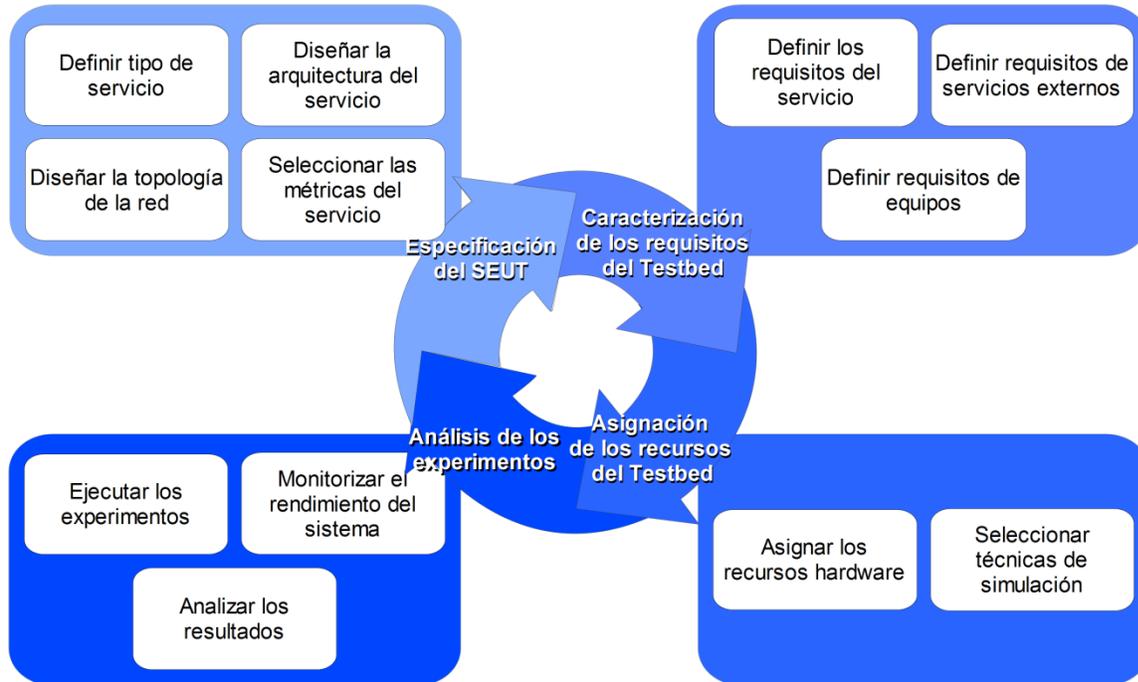


Figura 3-27: Fases de la metodología de construcción de entornos de evaluación

Los requisitos del entorno de evaluación repasan los componentes del escenario que va a ser construido. Los recursos del *testbed* son los recursos *hardware* y *software* disponibles para realizar la evaluación. En la fase de experimentación propiamente dicha, tanto el servicio como los recursos del *testbed* son escrupulosamente vigilados para certificar la fiabilidad de los resultados. Una vez los resultados arrojan luz en uno u otro sentido sobre los objetivos iniciales, el ciclo puede volver a iniciarse ajustando los objetivos, según el ciclo de mejora continua.

La representación gráfica de la metodología se ha realizado teniendo en cuenta los requerimientos formales que impone la notación *Business Process Model and Notation* (BPMN v2)¹⁰. El diagrama de la figura 3-28 explica gráficamente los procesos secuenciales y las transiciones entre los procesos. Además incluye en la notación pertinente las entradas y salidas esperadas. Para cada fase de la metodología, las técnicas que son aplicables se numeran desde la técnica 1 (T1) hasta la técnica 17 (T17).

¹⁰ <http://www.visual-paradigm.com/>

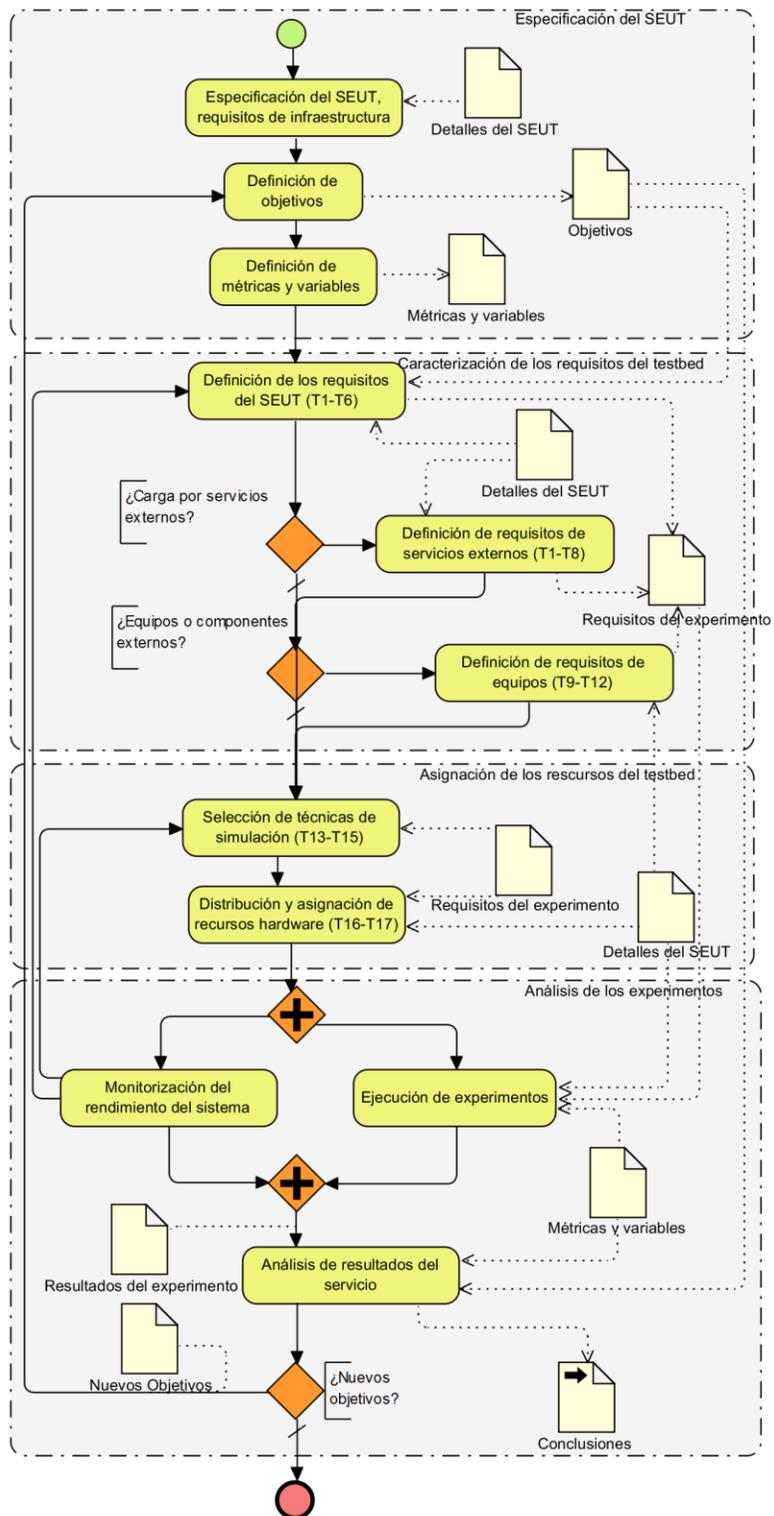


Figura 3-28: Flujograma de la metodología de construcción de entornos de evaluación

3.4.1.1 Fase 1, Especificación del SEUT

En la caracterización del servicio deben incluirse aspectos tales como la definición del tipo de servicio evaluado (por ejemplo, vídeo bajo demanda o servicio web) y su arquitectura (número de servidores, distribución del contenido, etc.). La infraestructura objetivo debe ser también explicitada en esta fase (redes de cable, inalámbricas, MANET, etc.). Por último, se definen los objetivos del experimento.

Una vez que el servicio ha sido descrito y los objetivos establecidos, se pueden definir las métricas e indicadores de rendimiento del servicio que se desea supervisar para evaluar el cumplimiento de los objetivos descritos en la infraestructura elegida.

Las métricas incluyen en esta fase métricas de red (consumo de ancho de banda, medidas de la congestión, etc.), métricas QoS (retraso *end-to-end*, *jitter*, tasa de pérdida de paquetes, cumplimiento de SLA), y métricas QoE,(PSNR, SSIM, MOS, etc.)

3.4.1.2 Fase 2, Caracterización de los requisitos del testbed

Esta fase establece los requisitos del experimento que deben ser atendidos por la plataforma de evaluación o *testbed* construido. Los requisitos del servicio evaluado (SEUT) son definidos y, opcionalmente, se pueden añadir al entorno de la evaluación cargas externas, otros servicios o dispositivos específicos para la evaluación conjunta.

Existen varios enfoques disponibles para realizar la emulación del servicio. El investigador debe definir qué camino elegir para aplicar las técnicas adecuadas a cada caso. El servicio puede ser modelado, puede ser capturado para ser reproducido después en la simulación, o puede ser evaluado usando las aplicaciones en tiempo real. En cualquiera de los casos, éstas técnicas pueden ser combinadas según las necesidades del investigador. Hay que considerar que para mantener el concepto de emulación definido, al menos alguna parte del servicio evaluado deben utilizar una parte no modelada, es decir, real. El mismo enfoque es aplicable a los usuarios del servicio. Los usuarios pueden ser reales o modelados mediante generadores de carga automáticos.

A continuación se describen más en detalle las técnicas aplicables a ésta fase, de acuerdo a la notación introducida en el diagrama de flujo de la metodología.

T1. Modelos de simulación en tiempo real.

Esta técnica enuncia el modelo de trabajo ya conocido. Mediante la técnica de simulación en tiempo real, el servicio real se integra con los modelos simulados del resto del entorno de evaluación. El modelo simulado puede abarcar incluso partes del propio servicio. Los simuladores de eventos discretos (DES) con los desarrollos que han permitido la inclusión de interfaces al mundo real y la restricción de tiempo real son las tecnologías que implementan esta técnica. La mayor limitación de esta técnica está en el consumo de recursos.

T2. Trazas de tráfico pre-grabadas

El servicio real puede ser muestreado para ser reproducido luego dentro de una emulación. Esta técnica permite una gran repetitividad y ahorra esfuerzos en el proceso de modelado del servicio. Sin embargo, dado que se trata de una muestra de tiempo limitado y comportamiento determinista, no es posible evaluar todas las situaciones como con el uso de un servicio real. Las implementaciones ofrecidas en el mercado para esta técnica distinguen entre flujos analíticos, que extraen las características del modelo del servicio de la traza grabada, o flujos a nivel de paquete, que replican paquete por paquete el flujo en el entorno emulado cuando son aplicados. Estas implementaciones ofrecen distintos niveles de rendimiento que pueden ser determinantes según las necesidades del investigador.

T3. Implementación real del servicio

Esta técnica propone la utilización del servicio real completo en el escenario evaluado. Es apropiada para servicios complejos, cuyo coste de modelado es alto, o cuyo comportamiento no es posible capturarlo mediante una traza grabada de tiempo limitado. Mejora la fiabilidad del experimento a costa de una carga computacional mucho mayor y una exigencia de recursos muy estricta.

T4. Aplicaciones distribuidas

Cuando se trabaja con cierto tipo de servicios, las aplicaciones involucradas pueden requerir una capacidad de procesamiento elevada. Puede resultar interesante para el investigador recurrir a técnicas de distribución de aplicaciones para dispersarlas entre la infraestructura, utilizando métodos de comunicación transparente para conectarlas con el resto de la simulación. Hemos visto el desarrollo de esta técnica para NS-3 en la primera parte de este capítulo. Hay que considerar que la infraestructura extra para comunicar las aplicaciones con el resto del entorno pueden introducir problemas adicionales.

T5. Usuarios Reales

Según esta técnica, una muestra de la población es elegida para interactuar con el SEUT mientras se realizan la evaluación del servicio. Los experimentos involucrando a usuarios reales están muy lejos de ser escalables y además difícilmente logran la repetitividad. Usar usuarios reales esta generalmente vinculado a utilizar implementaciones reales del servicio (T3).

T6. Generadores de carga de servicio.

El comportamiento de los usuarios del servicio puede ser caracterizado estadísticamente y modelado en el seno de un generador de carga. Luego las aplicaciones del servicio pueden ser controladas por el generador de carga que simula el comportamiento de un usuario. Esta técnica hace posible generar cargas escalables sobre un SEUT. Sin embargo, el modelado del comportamiento de los usuarios no es trivial y requiere un análisis estadístico detallado del uso que usuarios reales hacen del servicio real.

Frecuentemente, en las situaciones reales, otros tipos de tráfico presentes en la infraestructura de comunicaciones afectan de forma directa al servicio evaluado. Dos

técnicas adicionales ofrecen al investigador la posibilidad de considerar esta característica en la construcción de su *testbed*.

T7. Tráfico de fondo

El tráfico de fondo consiste en tráfico analíticamente modelado que tiene el objetivo de evaluar el efecto del tráfico existente en el servicio evaluado. En las simulaciones, este tráfico puede ser dispuesto entre dos nodos como flujos analíticos, o puede modelarse como una carga extra en un determinado enlace. Dependiendo del tipo software de simulación es posible considerar tráfico híbrido, con un porcentaje de tráfico explícito y otro de flujos analíticos. Los flujos analíticos son menos precisos que el tráfico explícito, pero ofrece mejor rendimiento global, abstrayendo detalles en favor de la mejora de rendimiento. En el trabajo en [115] se desarrollan modelos de tráfico individual y agregado que permiten llegar al desarrollo de una herramienta estratégica de planificación. Una de las conclusiones resalta cómo los modelos de tráfico a ráfagas que se observa en internet requieren soluciones adaptadas. El cálculo mediante técnicas de análisis de tráfico real es la opción más idónea. Sin embargo esta medida es complicada en las fases de planificación estratégica. Y más dentro de de las fases de desarrollo del servicio en cuestión.

T8. Generadores de tráfico

Un generador de tráfico puede ser utilizado para incorporar tráfico con unos parámetros determinados en la simulación. Los escenarios evaluados son más reales al contar con otros tipos de tráfico, no solo del SEUT. En una situación real el SEUT rara vez tendrá en exclusiva la dedicación de la red.

El modelado de escenarios puede requerir la inclusión de dispositivos en la infraestructura evaluada. Dependiendo de los objetivos definidos, pueden ser necesarios distintos niveles de abstracción. Las técnicas relacionadas con los dispositivos en la simulación son:

T9. Modelos de dispositivos o protocolos detallados

Algunas herramientas de simulación ofrecen la posibilidad de implementar modelos detallados de dispositivos o protocolo. Generalmente aplicable a simuladores DES. Esta técnica puede utilizarse para realizar experimentos evitando el despliegue del componente modelado de forma real. Sin embargo está limitado a un modelado preciso del componente o protocolo. El consumo de recursos de la simulación con restricciones de tiempo real del servicio sobre éstos dispositivos o protocolos modelados, pone en entredicho los resultados y supone un límite para la escalabilidad para esta técnica.

T10. Aproximación de caja negra

Esta técnica se beneficia de la abstracción de los detalles del funcionamiento interno del componente o protocolo. Permite la combinación de dispositivos interconectados que modifican el tráfico real que les atraviesa, introduciendo retrasos variables o pérdidas aleatorias. Esta técnica no permite considerar las interacciones del tráfico del SEUT con

otros tipos de tráfico ya que las condiciones de pérdida de paquetes o retardos dependen con frecuencia de la influencia de otros flujos competidores en la red evaluada.

T11. Virtualización de nodos

Esta técnica consiste en la utilización de máquinas virtuales u otras alternativas ligeras de virtualización para representar nodos en la red evaluada. Técnicas ligeras de virtualización [116], como la virtualización de la pila de protocolos de red, exclusivamente, en oposición a la virtualización del sistema completo, popular por las herramientas como VMWare¹¹ o Virtualbox¹². A través de esta técnica, varios nodos virtuales puede desplegarse en el mismo equipo físico. Es especialmente interesante para el desarrollo de experimentos en el ámbito del enrutamiento.

T12. Dispositivos de red reales

En ocasiones es necesario introducir en la evaluación dispositivos físicos reales que forman parte de los objetivos de evaluación, tales como un modelo concreto de un nuevo *router*. Si estudiar el rendimiento y desempeño del dispositivo es un objetivo, esta técnica materializa este concepto.

3.4.1.3 Fase 3, Asignación de recursos

Habiendo sentado los objetivos y requerimientos para construir el entorno de evaluación, el último paso antes de la experimentación en sí es el aprovisionamiento y la distribución de los recursos *software* y *hardware*.

Las técnicas de esta fase se reparten en técnicas que permiten elegir los aspectos de la simulación (T13 -- T15) y las técnicas que se relacionan con la asignación y distribución de los recursos *hardware* (T16 -- T17)

T13. Simuladores distribuidos

Esta técnica sugiere la partición de la simulación en diferentes simuladores. Un investigador tiene la opción de considerar utilizar distintas herramientas para diferentes partes del mismo experimento. La comunicación entre las particiones debe ser, preferiblemente, encapsulada de un modo similar al que se realizó en la extensión distribuida explicada en la primera parte del capítulo. Con el encapsulamiento del tráfico se evita realizar traducciones de la información entre los distintos lenguajes o protocolos que hablen los interfaces intermedios. Una limitación de esta técnica es que la simulación debe prestarse a la partición. Pueden darse problemas, por ejemplo, relacionados con el descubrimiento en los protocolos de encaminamiento si la topología no admite una partición natural.

¹¹ <http://www.vmware.com/es>

¹² <https://www.virtualbox.org/>

T14. Co-simulación

La co-simulación o simulación cooperativa [117] es una técnica de simulación que permite que componentes individuales sean simulados por distintas herramientas que se ejecutan simultáneamente e intercambian información de forma colaborativa. Similares a la co-simulación son la simulación de federados [118] y la simulación paralela de eventos discretos (PDES) [119]. A diferencia de la partición propuesta en la técnica de simuladores distribuidos, esta aproximación requiere *hardware* y *software* específico. De todas formas la partición de la topología es un problema complejo ya que deben mantenerse las comunicaciones entre las partes escindidas tan bajas como sea posible.

T15. Virtualización del tiempo

Esta técnica utiliza la virtualización del tiempo como medida para aumentar la escalabilidad de la simulación en tiempo real. El tiempo real se extiende a voluntad para todos los componentes involucrados en la emulación, de esta forma se permite procesar más eventos por unidad de tiempo. Sin embargo, presenta una complejidad elevada e imposibilita la utilización de otras técnicas como usuarios o dispositivos reales.

T16. Partición de la topología

Esta técnica permite la asignación de piezas disjuntas de la topología de la red simulada a distintos procesos dentro de la misma máquina. Favorece la escalabilidad cuando el consumo de recursos se vuelve un límite al crecimiento. Las particiones pueden aplicarse a cualquier *software* de simulación de forma transparente, aprovechando siempre la partición natural de las topologías de red, si existe. Tiene el inconveniente de que las particiones son inconscientes del resto de la topología. Además, redes complejas o del tipo malla son difícilmente particionables.

T17. Testbed mundial

Esta técnica sugiere el uso de sistemas distribuidos mundialmente, disponibles para que los usuarios evalúen sus servicios y aplicaciones utilizando Internet como medio de transporte. Algunas plataformas como PlanetLab ofrecen muchas posibilidades en este ámbito. Esta práctica presenta algunas limitaciones en cuanto a repetibilidad y controlabilidad de los experimentos. Además las conexiones existentes entre los nodos de la red en cuestión puede tener capacidades limitadas según el objetivo del estudio. No es, por ejemplo, apropiado para experimentos con servicios no tolerantes al retardo.

3.4.1.4 Fase 4, Análisis del experimento

Hemos definido los requisitos y las técnicas, estamos en disposición de proceder a ejecutar la evaluación y capturar los resultados. En esta última fase hay dos procesos paralelos, mientras que la simulación se realiza, se mantiene un proceso de monitorización de las variables de rendimiento elegidas para controlar el progreso de la simulación en todo momento. En última instancia, se analizan los resultados, Figura 3-29.

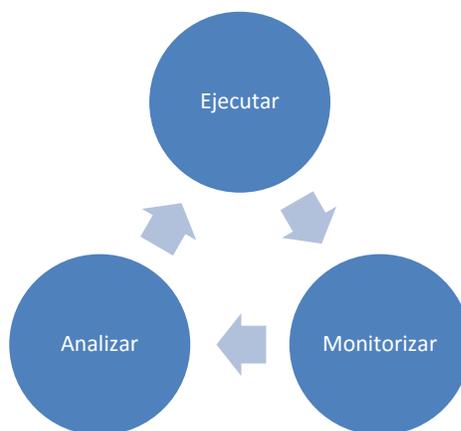


Figura 3-29: Secuencia de acciones de la fase de análisis

Con el fin de controlar que los resultados no se vean influenciados por los límites *hardware* o *software* impuestos por la infraestructura de la evaluación, se monitorizan paralelamente los recursos del entorno. Basándose en nuestra experiencia, el consumo de CPU y memoria son los recursos *hardware* que más influencia negativa pueden imponer en los resultados. Los recursos *software* medidos en cualesquiera métrica disponible, por ejemplo, la cantidad de eventos ejecutados por el simulador por unidad de tiempo, son también recomendables para garantizar una evaluación correcta del SEUT.

Durante el proceso de monitorización, cualquier problema detectado permite redirigir el foco de nuevo a la fase de asignación de recursos o a la fase de selección de técnicas, para renegociar las condiciones de la evaluación. Si el problema persiste pese a haber iterado entre estas fases, puede ser necesario renegociar los requisitos del experimento.

Toda vez que el experimento haya finalizado de forma satisfactoria de acuerdo a las métricas *hardware* y *software* elegidas, las métricas del SEUT pueden ser analizadas con total confianza. El análisis de los resultados generalmente requiere un tratamiento estadístico que garantice la certidumbre de los experimentos.

De este proceso se extraen las conclusiones sobre el rendimiento del SEUT, que puede llevar a redefinición de objetivos, para continuar de nuevo con el proceso desde el inicio tantas veces como sea necesario.

3.4.2 Aplicación de la metodología

Siguiendo el proceso descrito por la metodología, revisamos el procedimiento para construir el escenario de evaluación para el servicio multimedia sobre la topología de red descrita en el apartado anterior.

Especificación del SEUT. El servicio, la topología y las métricas a utilizar se han definido ya con anterioridad. Se pretende alcanzar la evaluación de hasta 1000 usuarios

Caracterización de los requisitos del testbed. La primera intención será evaluar un servicio totalmente real para todos los 1000 usuarios (T3). El comportamiento de los usuarios es modelado mediante un generador de carga (T6). Existe tráfico de fondo que incrementa la carga en el núcleo de la red (T7). Además de los modelos de simulación (T9) se incorporan componentes reales como el balanceador de carga o los servidores de vídeo (T12).

Asignación de recursos del testbed. En una segunda iteración de la metodología o de forma directa para los usuarios expertos, se revela necesario particionar el experimento (T16). Además también se combinan distintas tecnologías para la simulación (T13).

Análisis del experimento. Tras comenzar los primeros experimentos, la obligada monitorización de los recursos revela los primeros problemas: la partición de los clientes y la red del núcleo (T16) es necesaria. Sin embargo, la reasignación de los recursos es aún insuficiente para completar el experimento. Por esto, es necesario revisar los requisitos del servicio. Se opta por utilizar un 5% del tráfico del servicio real y el resto modelado de forma analítica en el núcleo (T2). En estas condiciones las métricas de rendimiento autorizan a realizar el estudio de las métricas del servicio. Una vez constatado que el servicio tal cual se ha definido se congestiona cuando hay más de 860 usuarios, se puede reformular el servicio, cambiando la distribución de los servidores de vídeo, por ejemplo, y volver a ejecutar la metodología desde el inicio para continuar con los experimentos. En la Figura 3-30 se muestran capturas de un instante en uno de los flujos reales del servicio durante la realización de los experimentos. Se observa la degradación del servicio evaluado a medida que se introducen flujos compitiendo por los recursos.

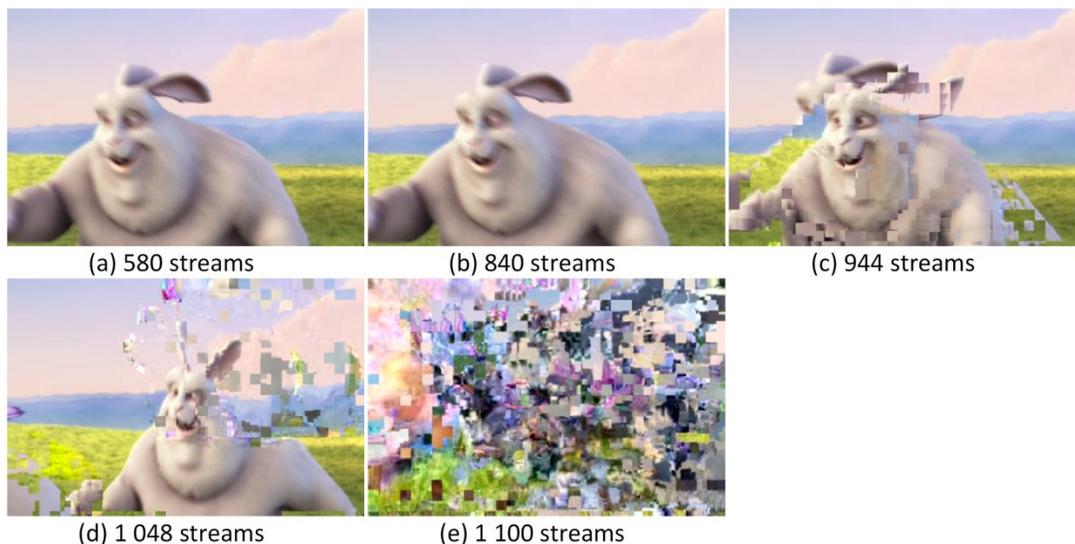


Figura 3-30: Fotogramas de la secuencia real de un instante en la evaluación del servicio masivo mediante la aplicación de la metodología.

Los resultados de la evaluación realizada en el escenario conjunto del apartado anterior serían obtenidos, tras la aplicación de la metodología, de una forma más rápida y procedimental. Innegablemente se ha llegado a las mismas conclusiones y resultados, pero el procedimiento enunciado facilita el recorrido y ahorra la improvisación en la construcción de entornos de evaluación.

3.5 Contribuciones del capítulo

En general, los simuladores en tiempo real, y en particular los utilizados durante nuestra investigación, han demostrado una capacidad limitada para proveer de soluciones de evaluación a los entornos masivos. Una sola instancia de NS-3 se muestra incapaz de gestionar más de 10 flujos simultáneos de un servicio típico de demanda media, en una de las topología más simples y convencionales. En el caso de utilizar OPNET esta capacidad se eleva hasta los 60 flujos, de nuevo condicionado a la topología elegida y los recursos *hardware* a su disposición.

La primera contribución se materializa en la extensión propuesta para dotar a NS-3 de capacidad de simulación distribuida alivia ligeramente la carga de rendimiento del procesador que se encarga de la simulación. De los estudios realizados con el simulador NS-3 se concluye además que el límite de rendimiento viene impuesto más por el número de enlaces o nodos en la topología simulada que por el volumen de tráfico en sí mismo. El consumo de recursos elevado condiciona los resultados de la simulación, haciendo que se pierda precisión en el resultado.

El experimento desarrollado para evaluar los límites de simulación en tiempo real con las dos populares herramientas constituye una contribución al estado del arte, ya que ningún estudio se enfoca claramente en evaluar los límites reales del rendimiento de la herramienta antes de atacar directamente las evaluaciones objetivo. Es importante matizar el rango de funcionamiento válido para cada herramienta. Mientras se garantice una zona operacional dentro de los límites de funcionamiento correcto de las herramientas, los resultados pueden ser aceptados como fiables. Sin embargo no existe una métrica única y clara que sea indicativa de la reducción del rendimiento y la confiabilidad de los resultados. Desde luego, ninguna de las herramientas evaluadas incluye este indicador interno como *output* en la simulación. Ese indicador, en forma de eventos procesados por unidad de tiempo, ha sido introducido de forma manual para esta investigación, pero sería una propuesta interesante a incorporar en la línea de desarrollo principal de ambas herramientas

Nuestros experimentos han demostrado que métricas QoS, como PLR, u otras métricas QoE (PSNR, SSIM, MOS) son igualmente indicadas para evaluar el rendimiento del servicio siempre que los rangos de funcionamiento hayan sido acotados. Sin embargo, extraer resultados fiables sobre el *jitter* no es posible. En ninguno de los dos simuladores en tiempo real se ha podido constatar una modificación del *jitter* en los flujos reales por incidencia de la congestión de los enlaces simulados hasta que la tasa de pérdidas es

fatalmente alta. Este hecho invalida el uso de estos simuladores para, por ejemplo, aplicaciones que usan el *jitter* como una métrica de congestión para algoritmos adaptativos en el entorno de servicios adaptables, tal como se propone en [120].

La simulación en tiempo real escala de forma limitada. Con el desarrollo de la primera metodología de construcción de entornos de evaluación de servicios masivos hemos contribuido a favorecer el crecimiento controlado de los entornos, pero siempre a costa de asumir ciertas concesiones en alguna de las dimensiones de realismo o escalabilidad. En contrapartida, es una tecnología económica y flexible que permite un trabajo ágil en las primeras fases de desarrollo de una tecnología, protocolo o servicio.

4 La calidad de experiencia objetiva en los sistemas adaptativos

En este capítulo se describen las contribuciones realizadas en el ámbito de las métricas de calidad objetivas. Las métricas de calidad objetivas son el método empleado para representar la calidad percibida de un servicio multimedia. Con asiduidad las métricas objetivas representan y hasta sustituyen a las evaluaciones subjetivas, por el elevado coste de éstas. Las métricas objetivas más relevantes fueron enunciadas en el apartado correspondiente del estado del arte, pero como se evidencia en este capítulo, existen muchos inconvenientes que limitan la aplicación de las métricas de calidad objetiva a secuencias resultantes de servicios multimedia adaptativos. En este capítulo se desarrolla y contrasta la adaptación propuesta sobre la métrica PSNR para representar la calidad subjetiva de una secuencia adaptativa. Se construye también una herramienta marco para decodificar y procesar secuencias resultantes de sistemas adaptativos y se postula un método de estimación de métricas objetivas basado en información reducida aplicable a los servicios adaptativos en tiempo real.

4.1 Introducción

Con el desarrollo de servicios multimedia en Internet y el rápido crecimiento de los usuarios accediendo a estos servicios, la industria se esfuerza por mantener unos niveles de calidad aceptables para la comunidad de usuarios finales. En el pasado, las métricas de calidad de servicio (QoS) constituían el estándar de medida para los servicios en general. Entre las métricas de servicio clásicas se cuentan métricas como PLR, BER o *jitter*, entre otras.

Pero con el tiempo, los agentes del servicio multimedia comienzan a demandar la valoración de aspectos que las métricas orientadas a otros flujos de datos no eran capaces de evaluar, principalmente la consideración del usuario final en el servicio evaluado. Más recientemente, las métricas de QoE surgen como una extensión natural y necesaria de las métricas de servicio. Los flujos multimedia incluyen características complejas que influyen en la evaluación que el usuario final hace del servicio *end-to-end*. Las métricas de QoE, con indicadores como el MOS son ciertamente deseables, aunque generalmente suponen una inversión en tiempo que en muchas ocasiones las hace inviables.

Es por esto que surgen las comúnmente referenciadas como métricas objetivas de QoE. Las métricas objetivas de QoE son estimaciones de la calidad subjetiva basándose en alguna observación física del flujo multimedia. Las métricas de QoE son luego contrastadas con las métricas subjetivas, como es el caso de PSNR y su popular escala de correspondencia con MOS [42].

Las métricas de QoE tienen la misión de evaluar todos los aspectos del servicio multimedia: audio, vídeo y la integración audio-vídeo en el servicio en particular. Sin embargo, la calidad de vídeo es generalmente la parte que mayor reto supone ya que es la parte más voluminosa del flujo multimedia y es más sensible a errores en la transmisión. De ahí que los mayores esfuerzos se dediquen al desarrollo de métricas de calidad de vídeo objetivas.

Las métricas de calidad de vídeo objetivas tienen su origen en la evaluación de los algoritmos de codificación con pérdidas [121]. Estos algoritmos modifican la calidad del contenido fuente para adaptarlo mejor a las características del medio de transmisión. En el proceso se produce una pérdida de información que se evalúa a través de dichas métricas. Lo mismo ocurre para el proceso de decodificación. A una secuencia comprimida mediante un algoritmo se le aplica un algoritmo recíproco para obtener un contenido visual que tiene una calidad determinada. De una forma global, se pueden establecer comparaciones de la calidad original y la calidad decodificada entre los extremos de la cadena de distribución del servicio. Esta es una medida habitualmente conocida como análisis *end-to-end*. Por último, es posible que resulte interesante realizar un análisis exclusivamente del contenido transmitido. En la Figura 4-1 se representan los tipos de análisis según su alcance.

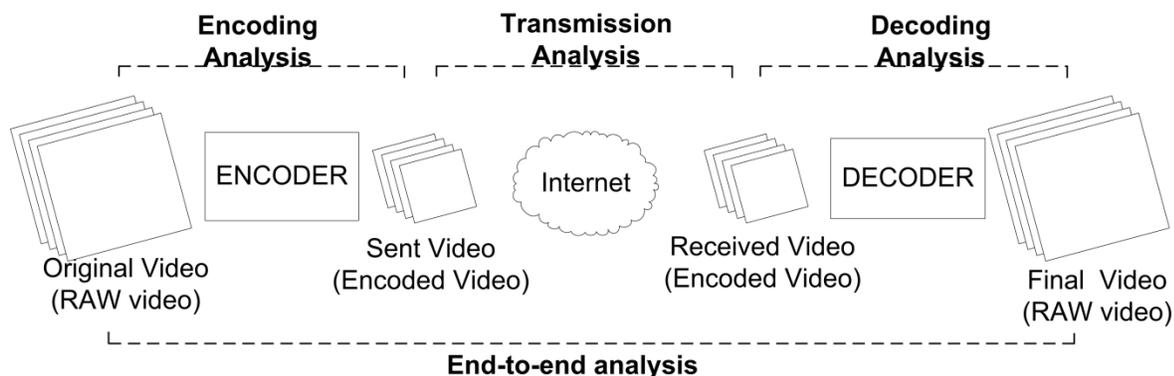


Figura 4-1: Tipos de análisis en la transmisión de vídeo

El problema específico en servicios de contenidos adaptativos se manifiesta cuando al hacer un análisis *end-to-end* existe una pérdida de información importante que no es debida a errores en la transmisión, sino que forma parte de la definición del servicio. En los servicios adaptativos se elige arbitrariamente (en función de algún tipo de algoritmo) eliminar capas del contenido codificado. Al comparar la secuencia decodificada final con

la original, la ausencia de información debida al proceso de adaptación, se mezcla con el efecto de las pérdidas debidas a la incidencia de los errores en la transmisión. En una evaluación *end-to-end* no pueden aislarse las pérdidas de calidad debidas a errores en la transmisión.

Y es que, en el caso de analizar un servicio en desarrollo, un protocolo o una topología determinada, es necesario conocer la disminución de calidad debida al proceso de transmisión de forma exclusiva. Es en estos casos cuando es relevante el análisis de transmisión, a través del cual se puede obtener la disminución de calidad objetiva debida exclusivamente al efecto del proceso de entrega del contenido, sea adaptado o no, a través de la red elegida y mediante el protocolo seleccionado.

Para permitir que algunas métricas objetivas, en particular PSNR admita el cálculo de las pérdidas de transmisión, deben realizarse algunas modificaciones, o asumir algunas concesiones. Para resolver esta indeterminación y resultado de nuestra investigación, se introduce la métrica cPSNR.

En el ámbito de la tecnología escalable, existe además un problema añadido. Ante la falta de un decodificador estable capaz de trabajar con secuencias reales que hayan sufrido pérdidas de transmisión, los trabajos de análisis de calidad multimedia existentes se limitan a imitar el comportamiento de secuencias escalables con otros *codecs* o a trabajar con trazas en vez de con contenido real. En el transcurso de la investigación se realiza un esfuerzo por construir un sistema de decodificación realmente operativo y funcional que admita todo tipo de secuencias, distorsionadas o no. Este hito posibilita los análisis subjetivos sobre contenido escalable real.

4.2 Métrica cPSNR. PSNR de transmisión en SVC

El uso tradicional de la métrica PSNR era la evaluación de los algoritmos de compresión de imágenes o esquemas de codificación para secuencias de vídeo. Trabajos posteriores propusieron la extensión de la métrica PSNR para incluir la evaluación de los errores resultantes de la transmisión del contenido sobre redes propensas a errores [44], [122]. Estos trabajos ayudaron a extender la consideración de PSNR hasta una métrica *end-to-end*. La comunidad ha tomado como convenio utilizar un *frame* original no distorsionado, antes de codificar, para calcular los valores de PSNR en las evaluaciones *end-to-end*. Esta consideración alivia el problema de los valores nulos de MSE que llevan a valores de PSNR indefinidos, en los casos en los que la transmisión del *frame* en particular no haya generado ningún error. Esto es debido a que la gran mayoría de los codificadores de vídeo actuales introducen pérdidas de calidad en la secuencia codificada, con respecto a la secuencia original. Por lo tanto, comparar un *frame* de una secuencia original, no comprimida con el *frame* recibido en el extremo de la red correspondiente, nunca podrá arrojar valores indefinidos de PSNR.

Sin embargo, el convenio de inclusión de las pérdidas de codificación en el estudio de las pérdidas ejercidas por una transmisión incorrecta, enmascara los resultados de la medida de las pérdidas de transmisión. En algunos casos, es necesario el estudio de las pérdidas en la calidad objetiva como consecuencia exclusiva de la transmisión. Es el caso por ejemplo de la tecnología SVC. En este caso particular la tecnología de transmisión por capas permite que el contenido enviado no se corresponda con la secuencia de imágenes original en cuanto a calidad se refiere. En tal caso la medida de la calidad del sistema de transmisión mediante la comparación de la secuencia recibida con la secuencia original resulta incorrecto. Las capas no transmitidas influyen en la calidad global si se mide respecto a la fuente original y esta pérdida de calidad no tiene nada que ver con las pérdidas ocasionadas en la transmisión.

Por el contrario, para evaluar de forma aislada las pérdidas de transmisión se considera la PSNR entre las secuencias recibida y realmente transmitida. En este caso pueden producirse valores nulos de MSE que llevan a indefiniciones en el caso de PSNR. Algunos trabajos han tratado de mitigar esta incompatibilidad asignando un valor fijo de PSNR (100dB) a los casos con valor nulo en MSE [123]. Esta decisión puede considerarse arbitraria e imprecisa, ya que no se han evaluado otras alternativas.

En este sentido se propone la creación de la métrica cPSNR o PSNR continua. Esta métrica, a diferencia de la PSNR original, tiene la característica de ser capaz de medir exclusivamente los errores resultantes de la transmisión del contenido. La cPSNR mantiene el mismo rango de escala logarítmico tan conocido de la escala PSNR. Al mantener la escala, la correspondencia lineal con la escala MOS sigue siendo válida.

Para computar la métrica cPSNR, para cada par de fotogramas, antes y después del proceso de transmisión, se obtiene el valor de MSE sobre el componente de luminancia. Los valores obtenidos para cada par son promediados para la secuencia completa, obteniendo un único valor (\overline{MSE}) que representa toda la secuencia. Posteriormente se transforma el valor a la escala PSNR según la ecuación 4.1.

$$cPSNR = 20 \log_{10} \left(\frac{V_{peak}}{\sqrt{\overline{MSE}}} \right) \quad (4.1)$$

El proceso intermedio de evaluar el promedio de la distancia euclídea entre las secuencias antes de aplicar el logaritmo permite evitar los valores indefinidos en el caso de evaluar secuencias que no han sufrido pérdidas en la transmisión de alguno de sus *frames*. Nótese que esta métrica no es de aplicable en secuencias que no han sufrido ninguna pérdida en transmisión, en ninguno de los frames. Se puede continuar utilizando la métrica cPSNR en la medida de la calidad *end-to-end* tal como se hace con la PSNR original.

4.2.1 Validación de la métrica cPSNR

Para evaluar el desempeño de la métrica propuesta se utiliza una base de datos pública de secuencias evaluadas subjetivamente [124]. En esta base de datos, los autores proporcionan la puntuación subjetiva de 78 secuencias obtenidas en dos laboratorios diferentes. Las secuencias se corresponden a 6 secuencias originales. Las secuencias originales son popularmente conocidas, ya que forman parte de las secuencias clásicas usadas en las evaluaciones de este tipo de contenidos¹³. Cada una tiene distintas características espaciales y temporales, además de representar una temática diferente (*foreman, hall, mobile, mother, news, paris*). Las secuencias originales son procesadas para generar secuencias corruptas de acuerdo a unos patrones fijos de pérdidas de transmisión con diferentes ratios. Las tasas de pérdida de paquete van desde el 0.1% hasta el 10%. La puntuación subjetiva proporcionada ha sido obtenida mediante la metodología SS. Según este procedimiento las secuencias se presentan de forma individual al sujeto participante, sin incluir la referencia original. Afortunadamente, los autores incorporaron la evaluación de la secuencia original como una secuencia más del conjunto, sin notificar su condición. Este hecho hace posible conocer la caída de calidad subjetiva por la inclusión de las pérdidas de transmisión. Restando la calidad subjetiva de la secuencia original y la secuencia con pérdidas se obtiene un valor de MOS diferencial (MOS_drop) vinculado a la calidad perdida por la transmisión. Este nuevo valor MOS_drop será utilizado para evaluar la correlación de las métricas de medición de calidad objetiva de transmisión.

Procesando las secuencias de la base de datos con los algoritmos definidos, se obtienen tres métricas diferentes: cPNSR, PSNR convencional con la corrección de 100dB para valores nulos de MSE y SSIM. Las tres métricas se comparan con los indicadores subjetivos representados en forma de MOS_drop obtenidos de las valoraciones subjetivas de la base de datos.

Los resultados indican que la métrica cPSNR tiene, en general, un mejor índice de correlación con la calidad subjetiva (Figura 4-2 a) que la alternativa, PSNR con la compensación de 100dB (Figura 4-2 b) y que la métrica SSIM (Figura 4-2 c). Los coeficientes de correlación de Pearson para cada contenido son presentados en la tabla 4-1.

¹³ <https://media.xiph.org/video/derf/>

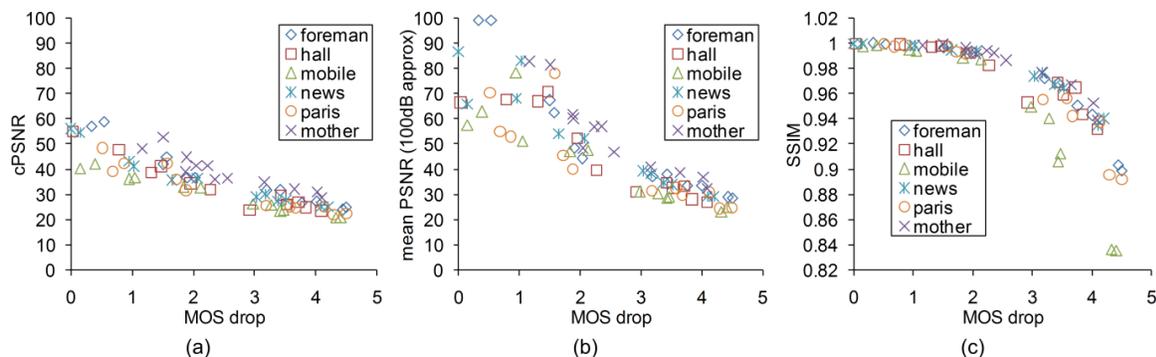


Figura 4-2: Correlación entre métricas objetivas de transmisión y caída de calidad (MOS_drop)

Tabla 4—1 Coeficientes de correlación Pearson para cPSNR, PSNR_100dB y SSIM con MOS_drop

Content	cPSNR	mean PSNR (100 dB approx)	SSIM
<i>foreman</i>	-0.967	-0.929	-0.898
<i>hall</i>	-0.965	-0.952	-0.925
<i>mobile</i>	-0.995	-0.963	-0.910
<i>mother</i>	-0.931	-0.922	-0.960
<i>news</i>	-0.971	-0.978	-0.912
<i>paris</i>	-0.959	-0.885	-0.938

Según el tipo de contenido pueden encontrarse algunas excepciones en la relación entre las métricas. La secuencia "news" presenta resultados más parecidos a la medida subjetiva con la métrica objetiva PSNR convencional con el ajuste de 100dB. En el caso de la secuencia "mother", ésta presenta mejor índice de correlación con la medida subjetiva para la métrica SSIM. Estas excepciones vienen a confirmar la influencia del tipo de contenido en las evaluaciones subjetivas. Algo que ya fue enunciado en algunos trabajos [44] y sugiere que el tipo de contenido debería ser incorporado a las métricas objetivas de alguna forma.

Particularmente interesante es también la relación entre las métricas objetivas que miden la distorsión de transmisión y, en este caso particular, PLR, que es el factor principal de las distorsiones de transmisión para las secuencias sintéticas generadas en la base de datos. En la figura 4-3 se presentan los resultados para las métricas cPSNR y PSNR (compensación 100dB) junto a la métrica PSNR *end-to-end* convencional (sin compensación) como referencia. La primera observación evidente que las métricas objetivas de medida de la calidad de transmisión se aproximan a la métrica PSNR *end-to-end* cuando las pérdidas son altas. En estas condiciones la calidad objetiva (o en particular, la falta de calidad) por las pérdidas, constituye el mayor factor, el más influyente, en la calidad final.

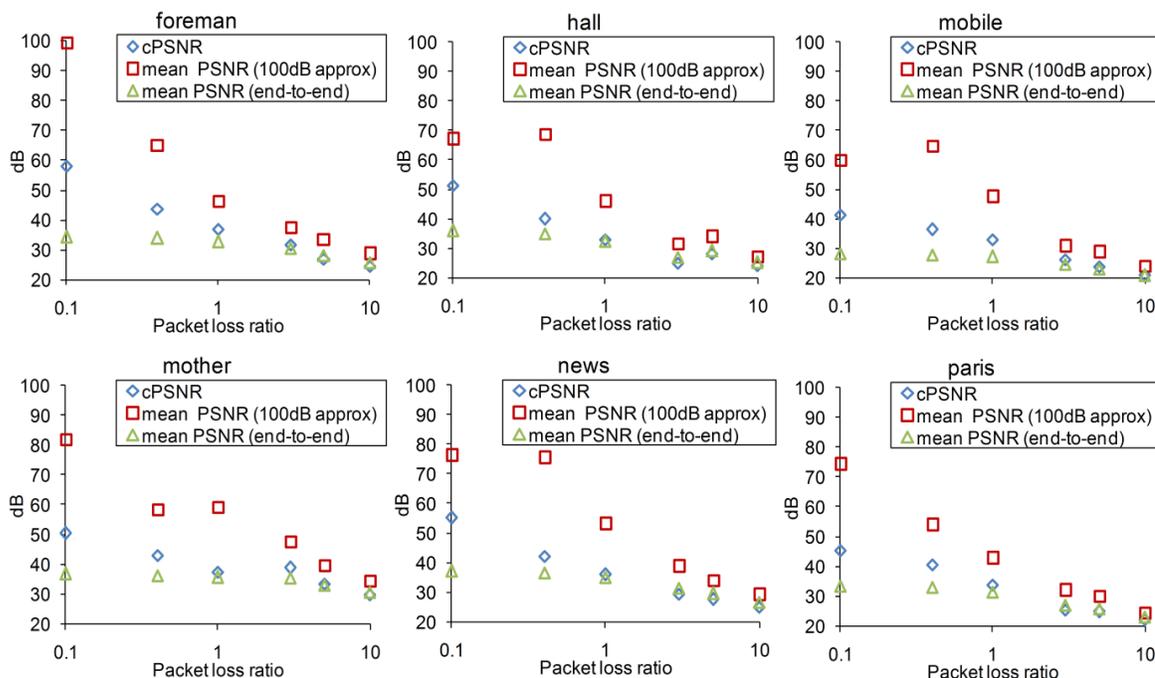


Figura 4-3: Métricas de distorsión de transmisión por contenido según la tasa de pérdidas, con la métrica *end-to-end* como referencia.

Sin embargo, cuando las pérdidas son bajas aparece una diferencia mayor entre las dos métricas de calidad objetiva de transmisión. La utilización de un valor arbitrario, y presuntamente elevado, para compensar las singularidades de la ecuación de PSNR convencional cuando se mide la calidad de transmisión, hace que se sobreestime la calidad. En este caso, los coeficientes de correlación mostrados en la Tabla 4-2 demuestran que para pérdidas bajas, cPSNR mejora a la medición de PSNR con la compensación de 100dB. Pueden encontrarse pequeñas excepciones nuevamente en función del tipo de contenido. Ciertamente, el tipo del contenido tiene mucho que ver en el efecto que las pérdidas ejercen sobre la calidad objetiva percibida. En cualquier caso, la métrica cPSNR supone una simplificación y una mejora sobre la alternativa estudiada.

Tabla 4—2 Coeficientes de correlación Pearson para cPSNR, PSNR_100dB con PLR

Content	cPSNR	mean PSNR (100 dB approx)
<i>foreman</i>	-0.787	-0.714
<i>hall</i>	-0.726	-0.794
<i>mobile</i>	-0.874	-0.841
<i>mother</i>	-0.839	-0.827
<i>news</i>	-0.769	-0.834
<i>paris</i>	-0.823	-0.777

La métrica cPSNR ha sido incluida en la herramienta marco desarrollada para analizar y comparar las calidades de transmisión. La herramienta *Distributed Multimedia Systems Group Analysis Framework (DMMSAF)* ha sido desarrollada en el marco de esta Tesis para incorporar distintos aspectos clave en la evaluación de la calidad de transmisión de contenidos.

En la literatura existen distintos trabajos que compilan las métricas existentes en *frameworks* o herramientas, proporcionando a la comunidad de un vehículo para las medidas de QoE de forma estandarizada. La plataforma Evalvid [125] es una herramienta que permite calcular las métricas PNSR y SSIM en secuencias con formato *raw*. En otro trabajo [126], los autores desarrollan una herramienta *software* que incluye un estimador de calidad de experiencia para entorno reales, presentando los resultados de la simulación que respalda su propuesta. Otro *framework*, basado esta vez en los criterios de escalabilidad y precisión ha sido desarrollado por los autores en [127]. Los autores integran varias métricas en *framework*, que es capaz de intercambiar precisión por escalabilidad en el cálculo de las métricas. Una extensión de Evalvid ha sido presentada en [125]. Esta última contribución permite trabajar con transmisión de vídeo adaptativo utilizando trazas textuales del tráfico que habilitan la reconstrucción del vídeo en el destino. Sin embargo, no permite trabajar con vídeos reales ya que la manipulación se limita exclusivamente al uso de trazas y referencia al contenido original.

En este sentido, la herramienta DMMSAF incorpora los mecanismos para permitir trabajar con secuencias reales, codificadas o no. Con el afán de permitir el análisis de la calidad de transmisión del contenido escalable, DMMSAF incorpora mecanismos de sincronización configurables entre la secuencia original y la recibida tras el proceso de transmisión. En la Figura 4-4 se muestran los modos disponibles. En el primer caso, figura 4-4 (a), las secuencias son tratadas sin asumir ningún tipo de sincronización. Es el mismo modo de trabajo que impone Evalvid y es, generalmente, la forma de trabajar con secuencias *raw*, que no tienen ninguna información sobre la composición temporal de la secuencia. En el segundo caso, figura 4-4 (b), la herramienta obliga una sincronización sin ningún algoritmo de cancelación de error. Los *frames* que no tienen recíproco en recepción son excluidos del resumen de la métrica en cuestión. Este caso es especialmente útil en el caso de analizar secuencias de vídeo escalable, por ejemplo SVC, en las que un contenido original puede contener *frames* que no han sido transmitidos de forma intencionada. En el último caso, figura 4-4 (c), la herramienta fuerza un método de corrección por copia de *frame* (el más sencillo).

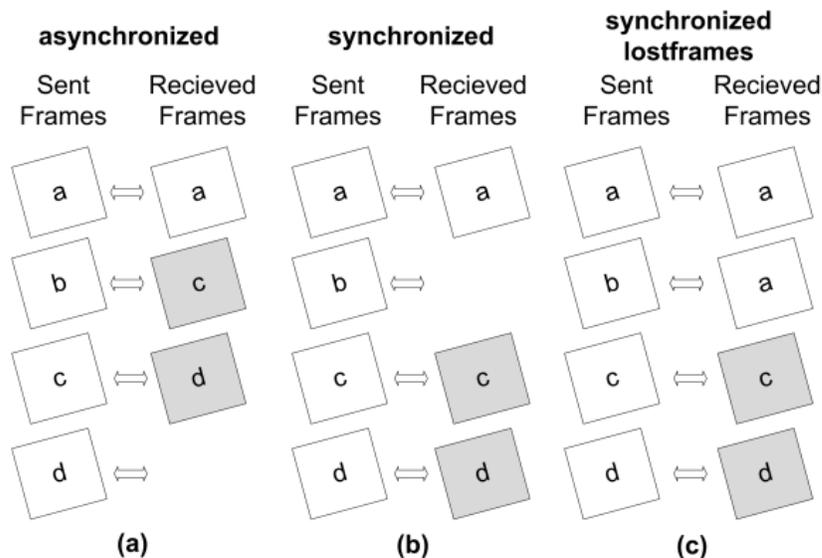


Figura 4-4: Modos de sincronización disponibles en DMMSAF

El rendimiento global de la herramienta DMMSAF ha sido comparado con el desempeñado por Evalvid. Para distintas calidades de vídeo y formatos de vídeo, se ha comparado el rendimiento en términos de tiempo de computación para ambas herramientas. En general DMMSAF intercambia operaciones de uso de disco por procesamiento en memoria, mientras que Evalvid hace un uso intensivo de disco. En términos generales, el rendimiento de la herramienta DMMSAF es similar al de Evalvid, simplificando el proceso de análisis de secuencias sin incurrir en mayores costes de computación, Figura 4-5.

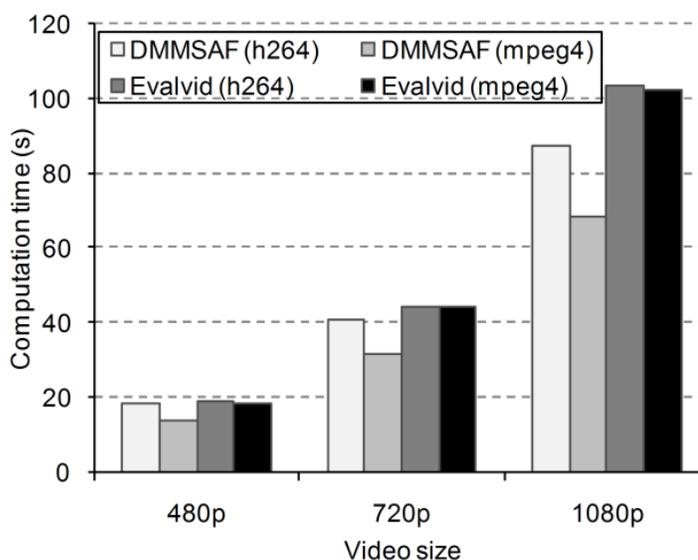


Figura 4-5: Comparativa del rendimiento de DMMSAF frente a Evalvid

4.3 El framework para estimar QoE en SVC

Hoy en día, SVC es considerada una de las tecnologías de codificación de vídeo más avanzadas. Gracias a sus características, SVC es capaz de brindar una representación por capas de un vídeo dentro de un mismo flujo de datos. Las capas ofrecen diferentes calidades, en términos de resolución temporal y espacial y en la dimensión de calidad. Las capas en un mismo flujo o *stream* se pueden agregar o quitar para obtener *streams* de calidades adaptativas. El estándar SVC ha originado numerosas ideas novedosas en el campo de la comunicación de vídeo. Se ha visto como una tecnología muy flexible para aplicaciones multimedia. El enfoque por capas ha demostrado ser adecuado para un mercado creciente en servicios diferenciados dirigidos para entornos específicos. La misma secuencia codificada se puede adaptar fácilmente para diferentes tamaños de pantalla diferentes calidades o cuando, por ejemplo, los dispositivos de destino están limitados en la potencia de procesamiento. En la tesis desarrollada en [128], el autor investiga los mecanismos de transporte y métodos de señalización entre la cache y el servidor así como estrategias de cacheo que contribuyen a mejorar la calidad global del sistema escalable. Dentro de las estrategias de transporte y métodos de señalización, la tesis doctoral desarrollada en [129] profundiza sobre la estimación del ancho de banda disponible, implementa y evalúa los mecanismos de adaptación basados en estimación no intrusiva que mejores resultados ofrecen en términos de calidad de experiencia.

Hasta la actualidad, se han propuesto una variedad de sistemas adaptativos basados en SVC que aprovechan el potencial de la codificación por capas. La combinación de la tecnología SVC, o cualquier otro *codec* escalable [130], con los algoritmos de estimación de la congestión de la red, permite construir sistemas adaptativos que son capaces de adaptar el contenido a las condiciones de la red en cada momento. Esta técnica ayuda a evitar la mayoría de las pérdidas de paquetes. Algunas soluciones utilizan la retroalimentación del cliente para estimar congestión de la red y actuar en el servidor [131]; otros implementan servidores proxy que realizan la adaptación [132]. Como resultado, la calidad *end-to-end* es potencialmente mejorada.

Como ocurre con cualquier nuevo *codec* o tecnología de transmisión desarrollados, sus beneficios deben ser evaluados de diferentes maneras. Ambos procesos, codificación y transmisión imponen una distorsión en el vídeo resultante. Las métricas de Calidad de Vídeo (VQ) son las métricas más útiles para cuantificar la distorsión de vídeo. Las métricas VQ están relacionadas con la QoE, ya que tienen la intención de medir la calidad de vídeo, incluyendo la evaluación que el usuario hace de la distorsión. La mayoría de las métricas VQ útiles incluyen medidas objetivas que proporcionan una estimación de la calidad subjetiva esperada. Las métricas de referencia completa (FR) se cree que son las métricas más confiables. En el lado negativo, estos indicadores requieren acceso a la secuencia intacta original, algo que en los servicios en tiempo real resulta complicado. Las métricas PSNR y SSIM son ejemplos de métricas objetivas de referencia completa (FR). Aunque se originan en la evaluación de la calidad de imagen también se han utilizado ampliamente para evaluar la calidad del vídeo. Métricas más

avanzadas como MOVIE [52] introducen además consideraciones temporales y espaciales. Por lo tanto, se cree que es más precisa para las secuencias de vídeo que las métricas simplemente heredadas de evaluación de la calidad de imagen, tales como PSNR.

Sin embargo, el uso de métricas FR con los sistemas adaptativos SVC no es sencillo. Además de los procesos habituales de corrección de errores presentes en la mayoría de los *codecs* modernos, las adaptaciones temporales y espaciales tienen que ser compensadas antes de extraer las métricas VQ. La evaluación de VQ en SVC en la literatura está básicamente restringida a analizar de forma independiente la calidad de diferentes capas o su combinación. Otros trabajos tratan de emular el comportamiento de la adaptación o usan una simulación de la adaptación basada en trazas. Ciertamente, la medición de VQ en flujos adaptativos reales que comprenden tanto las pérdidas debidas a errores de transmisión como variaciones dinámicas de capas de acuerdo a las condiciones de la red es un elemento clave para evaluar el desempeño de los sistemas adaptativos.

En primer lugar se propone una solución para medir métricas VQ en flujos reales adaptativos basados en SVC. La solución propuesta hace uso de una implementación abierta del decodificador SVC, más flexible que la implementación *software* JSVM). El sistema implementa un mecanismo de cancelación de errores y sincronización de la escalabilidad temporal simple, apoyándose en la información temporal contenida en los paquetes del protocolo *Real Time Protocol* (RTP) del *stream*. Para ejemplificar la utilidad del sistema diseñado para la medición de las métricas VQ en flujos adaptativos, se utiliza el sistema adaptativo diseñado en [131].

La calidad resultante también se obtiene a través de una propuesta de estimación de las métricas FR, comparando los resultados con los calculados a través de la herramienta diseñada. La propuesta de estimación de métricas FR para SVC se basa en el análisis de la calidad individual de cada capa. A cada capa del vídeo original se le asigna un valor VQ medio. La tabla de valores de VQ para cada capa es enviada junto al flujo de vídeo con una sencilla modificación de las cabeceras del protocolo de transmisión. Una solución análoga, para la información de los atributos de imagen se ha visto en [133] para resolver los problemas de los métodos FR en el cálculo de nitidez en imágenes naturales capturadas por cámaras digitales.

La información de las calidades individuales de cada capa, combinada con la información de la capa decodificada en cada momento y la tasa de pérdidas en recepción permiten estimar de forma relativamente fiel la métrica FR. Para demostrar este punto se comparan las métricas VQ calculadas en la forma original (calculadas en diferido con acceso a la secuencia original) y la estimación por el método propuesto.

4.3.1 El sistema adaptativo

El sistema desarrollado en [131] utiliza una estimación no intrusiva, usando el propio flujo de vídeo, del ancho de banda de la red para alimentar un algoritmo de selección de capas. El resultado es que proporciona un flujo adaptado a las condiciones del momento para cada usuario.

Los paquetes RTP que alcanzan el cliente se procesan y analizan para calcular métricas clásicas como tasa de pérdida de paquetes o *jitter* y otras métricas novedosas como la linealidad de los instantes de recepción de paquetes RTP. El sistema adaptativo utiliza los paquetes de retroalimentación disponibles en el protocolo RTCP para enviar la información medida por el cliente al servidor. Específicamente se utilizan paquetes RTCP-APP. Cuando los paquetes RTCP-APP llegan al servidor, éste puede estimar el ancho de banda para el cliente en ese momento.

Cuando se detecta una situación de congestión ($PLR > 1\%$) el sistema decrementa la tasa de transmisión eliminando las capas de más alta calidad del flujo enviado. Por el contrario, cuando no hay congestión, la tasa de envío puede ser incrementada añadiendo más capas al flujo que está siendo transmitido al cliente.

El algoritmo de selección de capas y su rendimiento está fuera del alcance de esta tesis. Utilizamos el sistema adaptativo como caso de uso para demostrar las propuestas en cuanto al sistema de decodificación y las métricas VQ estimadas. Las futuras revisiones del algoritmo de selección de capas se podrán ver beneficiadas por los avances realizados en las herramientas y métricas estudiadas aquí.

4.4 El camino hacia el decodificador adaptativo

4.4.1 Trasfondo

Las características definitorias de la tecnología SVC imponen un reto en la utilización de las métricas de calidad de vídeo en su forma original. En su nacimiento, al ser una tecnología relativamente reciente, las primeras implementaciones eran escasas y de funcionalidad limitada. Una primera propuesta que se realiza es el estudio de las posibilidades de la tecnología basada en capas a través de los *codecs* existentes, como H.264/AVC. Han sido varios los estudios subjetivos sobre este popular codificador, por ejemplo, [134]. Como H.264/SVC es la extensión escalable de su cercano pariente (H.264/AVC), ambos comparten muchas características. Posiblemente por esta razón varios autores han considerado emular el comportamiento adaptativo con secuencias codificadas con H.264/AVC de forma independiente, variando aspectos como la resolución espacial [135]. En este caso, las medidas de VQ pueden ser realizadas sin ningún inconveniente sobre los flujos reales. Sin embargo, es importante involucrar el *codec* y las herramientas adecuadas a la tecnología en las evaluaciones realizadas.

Por un periodo largo de tiempo, el *software* de JSVM fue la única alternativa de codificador disponible para obtener secuencias que cumplieren con el estándar SVC. De forma equivalente, el decodificador es capaz de revertir el proceso y generar una secuencia reproducible en un reproductor adaptado. Las propuestas de decodificadores proliferan y surgen alternativas basadas en paralelismo que prometen mejorar el rendimiento [136], pero la disponibilidad de implementaciones es limitada. La primera implementación de un decodificador disponible es el de referencia, JSVM. Sin embargo, este decodificador no admite secuencias distorsionadas, tales como las normalmente resultantes de la transmisión de contenido a través de redes no fiables. El decodificador JSVM no tiene capacidades de cancelación de errores. Se eliminaron desde la versión 9.10. Aún así, el decodificador JSVM permite realizar evaluaciones de calidad objetiva y subjetiva, siempre y cuando no haya pérdidas en la secuencia recibida. Los autores en [56] proponen una métrica de vídeo teniendo en cuenta varios aspectos de la secuencia codificada, pero si considerar la transmisión del vídeo y sus efectos sobre la calidad. Parámetros generales de los contenidos, como complejidad espacial y temporal han sido criterios importantes en muchas evaluaciones de codificadores. También SVC ha sido evaluado en términos de las complejidades temporales y espaciales, proporcionando valiosas bases de datos de secuencias caracterizadas subjetivamente [84]

Más adelante, los mecanismos de transporte fueron apareciendo progresivamente en los estudios con SVC. La primera versión del formato de paquete RTP para SVC fue publicado por la ITU en 2006 [137]. Las implementaciones de este estándar en los clientes y servidores son escasas. Dado que las implementaciones de RTP para SVC son escasas y el decodificador de referencia, poco flexible, algunos autores optan por construir plataformas de evaluación basadas en trazas, en vez de en flujos reales. De hecho, las herramienta proporcionadas por JSVM incluyen la posibilidad de obtener trazas de secuencias y reconstruirlas luego a partir de la traza modificada. Esta aproximación ha sido ampliamente utilizada con simulaciones. Los autores en [62] usan una estructura simplificada de RTP para simular una secuencia recibida a través de la reconstrucción originada en una traza modificada. Una aproximación similar es seguida en [61]. En cualquier caso, los modelos basados en trazas aún carecen del punto de precisión que aporta el sistema real.

Ciertamente, la implementación real del decodificador es una parte importante en el esquema de transmisión. En particular, los mecanismos de corrección o compensación de errores en la decodificación suponen un factor a tener muy en cuenta en la evaluación de la calidad de la experiencia *end-to-end*. Prueba de la relevancia del proceso de decodificación en la calidad es la variedad de trabajos recogiendo las distintas técnicas existentes y sus efectos sobre la calidad final. En particular, en el ámbito de la tecnología SVC trabajos como [58], [59], [89], [138]–[141] proponen una gran variedad de soluciones para este problema. Los algoritmos de cancelación no están, sin embargo, presentes en el decodificador de referencia.

Dadas las continuas restricciones presentes en el *software* de referencia, otras implementaciones alternativas van cobrando protagonismo. El proyecto OpenSVC [142]

proporciona una implementación más flexible del decodificador. *OpenSVC decoder* permite trabajar con secuencias distorsionadas de una forma mucho más flexible que el decodificador de referencia JSVM.

4.4.2 La solución diseñada

La evaluación de métricas de calidad en servicios multimedia adaptativos es un reto considerando las dificultades encontradas en la práctica. Cuando se pretende trabajar con las herramientas disponibles se encuentran implementaciones incompletas o deficiencias que dan lugar a las soluciones propuestas. Sin embargo, ninguna solución de las vistas ofrece una solución versátil para la medida de calidad sobre un sistema adaptativo real, con un *software* de decodificación en tiempo real y un entorno de experimentación emulado.

Nuestra propuesta demuestra cómo *OpenSVC decoder* puede ser usado como base para desarrollar un decodificador adaptativo. El decodificador diseñado admite flujos con cambios dinámicos en las capas transmitidas y pérdidas en la transmisión, tal como ocurre en los entornos reales.

Los sistemas adaptativos imponen unas restricciones adicionales a los sistemas de cancelación de errores. En la implementación actual de *OpenSVC decoder*, cuando una determinada capa no puede decodificarse, el algoritmo de decodificación busca la capa base del flujo para decodificarla. Si también la capa base se ha corrompido, el decodificador no puede corregir este *frame* en particular. En un sistema en tiempo real esto puede suponer que se congela la imagen, manteniendo el *frame* anterior durante más tiempo, o que se muestra un *frame* neutro (normalmente verde). En los sistemas adaptativos, además, el hecho de reducir las capas en la dimensión temporal, significa que el decodificador debe compensar la tasa de representación para mantener en pantalla los *frames* el tiempo necesario. En el caso de la escalabilidad espacial, la técnica de interpolación es la más sencilla para mantener el mismo tamaño de imagen visualizada independientemente de la capa recibida. En el sistema adaptativo usado como caso de estudio [131], la escalabilidad espacial no es utilizada. Para facilitar el cálculo de las métricas de calidad, la capa que está siendo decodificada para cada *frame* debe ser registrada. La Figura 4-6 muestra el proceso de cancelación de errores de una secuencia con una reducción de una capa en la dimensión temporal después del tercer *Group of Pictures (GOP)*. Para ocultar las variaciones en la capa temporal el decodificador debe conocer la información temporal asociada al flujo. En el sistema estudiado, la información temporal se almacena en el campo PTS (*Presentation Time Stamp*) del protocolo RTP. El decodificador reconoce el tiempo de presentación de cada *frame* o *Access Unit (AU)* y realiza la operación de copia de *frame* cuando el PTS es más largo que la duración del *frame*.

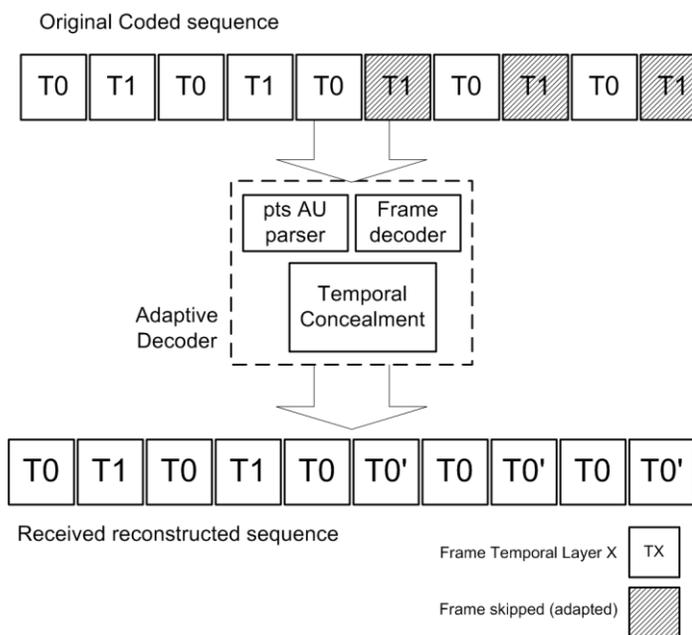


Figura 4-6: Esquema de cancelación de errores en el decodificador adaptativo

En la operación habitual del protocolo RTP, los tiempos de presentación se obtienen después de la primera sincronización, en el momento en que los mensajes de tipo *Sender Report (SR)* son intercambiados entre cliente y servidor. En los sistemas adaptativos este protocolo debe ser lo más prematuro posible, para comenzar la sincronización tan pronto como sea posible. El servidor usado en el estudio actual, basado en la librería *live555*, ha sido modificado para incorporar una sincronización rápida [143]. Con esta modificación, el retraso en la transmisión del primer mensaje RTCP SR se reduce de forma que es posible obtener tiempos de presentación desde el inicio del flujo para garantizar la sincronización temprana.

Además, las librerías de *live555* en nuestro ejemplo han sido modificadas para recibir y almacenar la información temporal que se gestiona de forma separada al vídeo. De acuerdo la especificación H.264, la información temporal ha sido eliminada del *codec*, precisamente para permitir el desarrollo de aplicaciones basadas en capas. Los *muxers* de vídeo como MP4 tienen el objetivo de registrar la información temporal de los flujos, algo que ya ha sido estudiado [144] [145], aunque hasta la fecha no se han encontrado implementaciones. Consecuentemente, el decodificador basado en *OpenSVC* ha sido modificado para considerar esta información temporal externa, en forma de trazas textuales.

4.4.3 Caso de estudio

Una vez construido el decodificador compatible con los flujos de vídeo adaptativo que presentan pérdidas en la transmisión, se elabora un caso de estudio para analizar los resultados de la calidad de vídeo *end-to-end*. Los resultados del sistema adaptativo serán comparados con los obtenidos en el mismo escenario anulando en el sistema el algoritmo de adaptación, es decir, con un sistema de transmisión convencional. Distintos escenarios de emulación creados en NS-3 son intercalados entre el servidor y el cliente, para forzar las condiciones deseadas en la red. El trabajo se apoya en los desarrollos hechos por los autores en [131]. La calidad de vídeo es evaluada en términos de PSNR, SSIM y MOVIE. Las tres métricas son del tipo FR, con lo que las secuencias recibidas requieren de la comparación con la secuencia original para computar la métrica correspondiente. Además, se propone un método de estimación, que será descrito más adelante, para obtener las métricas considerando exclusivamente la secuencia en recepción. La figura 4-7 es un esquema de la configuración del caso de estudio.

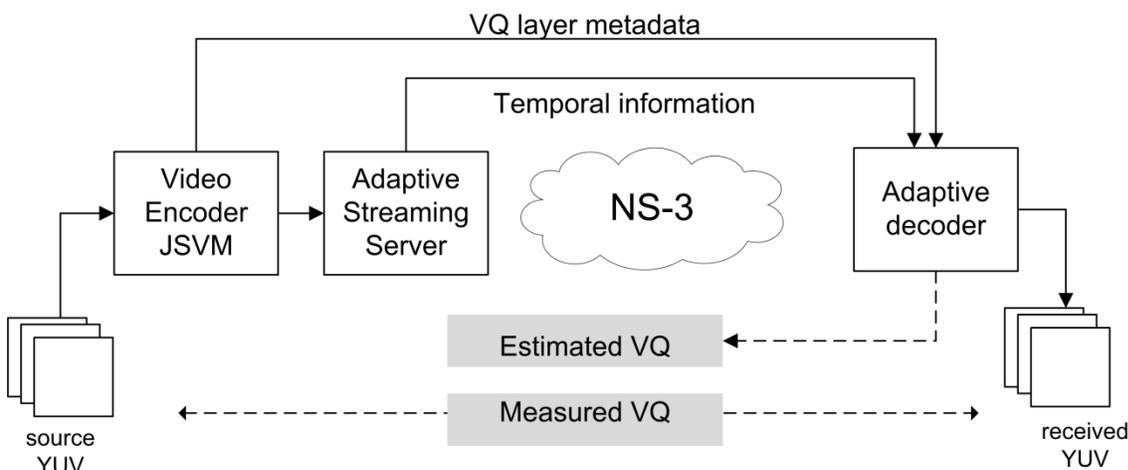


Figura 4-7: Caso de estudio para análisis de VQ en el sistema adaptativo

La secuencia de vídeo (*city.yuv*), cuyas características se muestran en la Tabla 4-3 se repite 6 veces para obtener una secuencia más larga y que permita que el sistema adaptativo pase por varios estados. La secuencia original es codificada usando el codificador de referencia, JSVM, resultando en una secuencia formada por dos capas temporales (T1, T2) y dos capas de calidad (D0, D1). El esquema de codificación elegido es *Coarse Grain Scale* (CGS). Las tasas de transmisión mostradas en la Tabla 4-3 han sido medidas en la red, incluyendo la parte correspondiente al protocolo. Representan una medida de los recursos consumidos por cada capa a nivel de red.

Tabla 4—3 Características de la secuencia city.yuv codificada con H.264-SVC

Parámetro	Valor
Frame Rate máximo	30 fps
Duración	60 segundos
Esquema de codificación	CGS
Capas temporales	T0, T1
Capas de calidad	D0, D1
BitRate	Capa id 0: T0D0 2.1Mbps Capa id 1: T1D0 2.9Mbps Capa id 2: T0D1 4.9Mbps Capa id 3: T1D1 6.3Mbps

El protocolo de transmisión es RTSP, utilizando un servidor basado en *live555* con las modificaciones descritas en apartados anteriores para habilitar la decodificación de la secuencia adaptativa con pérdidas. En la parte del cliente, *OpenSVC decoder*, se añaden las modificaciones complementarias a las del servidor, para procesar la información extra y la cancelación de error por la técnica de FC (*frame copy*).

Los escenarios emulados en NS-3 reproducen condiciones de red variables. Una red punto a punto conecta los extremos. La disponibilidad de ancho de banda se modifica añadiendo tráfico compitiendo por los recursos fijos del enlace punto a punto. Se han evaluado tres patrones de variación diferentes. Un escenario de crecimiento (*ramp-up*) de disponibilidad de ancho de banda, Figura 4-8. Un escenario donde la disponibilidad se decreta (*ramp-down*), Figura 4-9 y un escenario con cambios de ambos sentidos, Figura 4-10. Junto a la variación de los anchos de banda disponibles en cada escenario, se incluye la tasa resultante del vídeo transmitido, y en caso de existir, las pérdidas de paquetes incurridas. En el caso del sistema no adaptativo se elige una tasa de partida según las condiciones iniciales y se mantiene independientemente de las variaciones en la disponibilidad de la red (Figuras 4-8, 4-9 (a) y 4-10 (a)). En el sistema adaptativo, el sistema reacciona antes los cambios de ancho de banda disponible y adapta la secuencia transmitida, reduciendo las pérdidas en los casos en los que se produce una disminución de la disponibilidad de ancho de banda en la red. Figuras 4-8, 4-9 (b) y 4-10 (b).

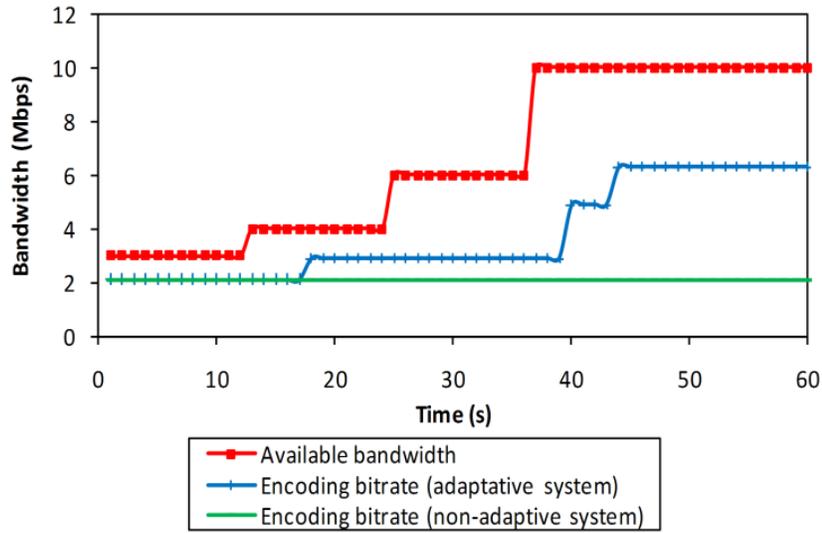


Figura 4-8: Escenario 1 (ramp-up), fuente [131]

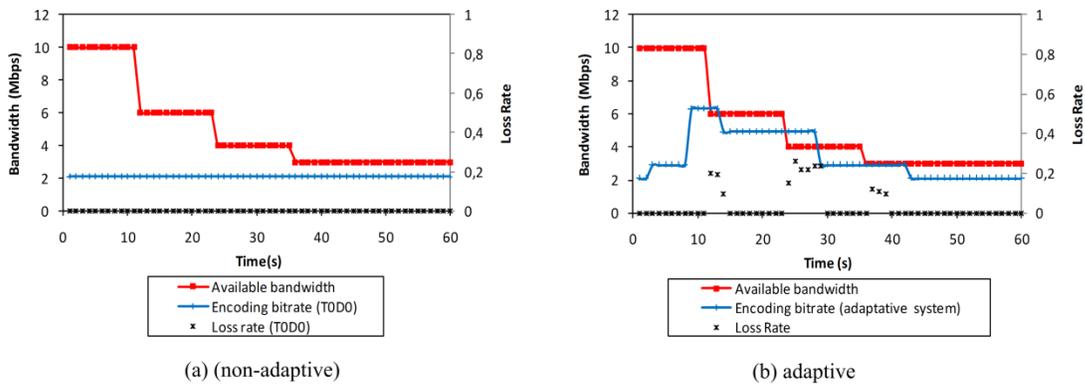


Figura 4-9: Escenario 2 (ramp-down) , fuente [131]

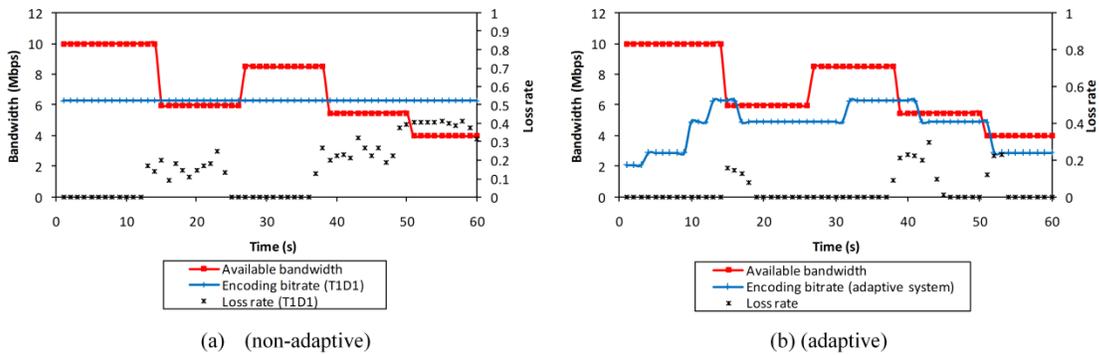


Figura 4-10: Escenario 3, fuente [131]

La calidad de las secuencias recibidas en los tres escenarios será calculada mediante las métricas PSNR, SSIM y MOVIE, usando las herramientas desarrolladas en [146]. Seguidamente, se procesa la meta-información de las secuencias recibidas, junto con la descripción de la calidad individual de las capas y la tasa de pérdidas para obtener una estimación de las métricas FR en recepción, sin necesidad de la referencia a la secuencia original. El método de estimación será descrito a continuación.

4.4.4 Método de estimación de calidad de vídeo FR

La metodología estándar para calcular las métricas FR requiere comparar la secuencia recibida, incluyendo todos los posibles errores resultado de la codificación y transmisión, con la secuencia original antes de ser codificada. Esta calidad se conoce como calidad *end-to-end*. El rendimiento de algunas métricas del tipo FR según el tipo de distorsión ha sido discutido en [147]. En cualquier caso, las métricas FR normalmente imponen un procesamiento offline, ya que se requiere un proceso costoso computacionalmente con acceso a una secuencia original que no tiene porqué ser conocida en el receptor.

Por otro lado, la utilización de métricas de referencia reducida (RR) o nula (NR) es una alternativa interesante. La necesidad de construir modelos de métricas NR ya fue expresada en otros trabajos [148]. Sin embargo, estas métricas tienen normalmente una credibilidad inferior a la acreditada para las métricas FR [121]. El método de estimación propuesto aquí permite aproximar las métricas FR reduciendo la información necesaria en el receptor a una simple descripción de la secuencia original. Además, el método de estimación no se limita a una sola métrica.

La información necesaria para que el estimador procese la métrica FR es la calidad media de las capas para la métrica en cuestión. Esta meta-información puede ser obtenida fácilmente en el proceso de codificación. Normalmente la calidad de la secuencia codificada es ya un producto de la codificación para la mayoría de codificadores. La estructura con esta meta-información puede ser compartida en alguno de los mensajes del protocolo de sesión de la transmisión. En el receptor, el decodificador puede utilizar la meta-información para asignar la calidad a la capa que esta decodificando realmente. En la Tabla 4-4 se muestran los valores medios de las métricas seleccionadas para las capas de la secuencia utilizada en el caso de estudio.

Tabla 4—4 Métricas VQ de las capas de la secuencia original

Capa de calidad	PSNR	SSIM	MOVIE
Capa id 0	33.2436	0.786842	0.0002166007
Capa id 1	42.7406	0.988498	6.594448e-08
Capa id 2	34.0604	0.788384	0.0002184489
Capa id 3	44.7599	0.992176	2.663533e-08

En recepción, la calidad de la secuencia recibida, en particular, de la capa más alta recibida puede ser igual o menor que la de la secuencia original. Las pérdidas sufridas

durante el proceso de transmisión pueden ocasionar desde errores en los frames decodificados hasta la necesidad de recurrir a una capa inferior en el proceso de decodificación. Por tanto, las pérdidas deben ser incluidas en el estimador de la calidad recibida de algún modo. El algoritmo propuesto para estimar la calidad se expresa como una función F de la calidad media de la capa decodificada en cada instante (ecuación 4.2). Esa función F depende de la métrica en cuestión. Para las tres métricas más relevantes las ecuaciones se muestran en la tabla 4-5. La utilización de una función lineal para relacionar las pérdidas con las métricas de calidad ha sido discutido en [149], donde el autor argumenta que la falsa linealidad observada puede ser debida a la utilización de modelos simplificados de pérdidas, basados en Gilbert, en los experimentos. En este caso el modelo de pérdidas es impuesto por impuesto por el simulador por lo que no se tiene esta dependencia y sin embargo se observa un buen ajuste de la métrica estimada considerando esa relación lineal con las pérdidas.

$$VQ_{\text{metric}_{\text{estimated}}} = F \cdot VQ_{\text{metric}_{\text{LAYER}}} \quad (4.2)$$

Tabla 4—5 Ecuaciones del estimador de VQ para las distintas métricas

PSNR	SSIM	MOVIE
$\alpha = 0.65$	$\alpha = 0.42$	$\alpha = 0.04$
$PSNR_{\text{estimated}} = F \cdot PSNR_{\text{LAYER}}$	$SSIM_{\text{estimated}} = F \cdot SSIM_{\text{LAYER}}$	$MOVIE_{\text{estimated}} = F \cdot MOVIE_{\text{LAYER}}$
$F = \begin{cases} 1, & PLR = 0 \\ \alpha \cdot (1 - PLR), & PLR > 0 \end{cases}$	$F = \begin{cases} 1, & PLR = 0 \\ \alpha \cdot (1 - PLR), & PLR > 0 \end{cases}$	$F = \begin{cases} 1, & PLR = 0 \\ \frac{1}{\alpha \cdot (1 - PLR)}, & PLR > 0 \end{cases}$

Según el modelo propuesto, en la ausencia de pérdidas, el estimador asume que la calidad de la secuencia recibida en ese instante es igual a la calidad media de la capa original que está siendo decodificada. En caso de existir pérdidas, la pérdida de calidad es una función de la tasa de pérdidas ponderada por un factor que tiene relación con el rango de la escala que utiliza cada métrica. El factor (α) se calibra una vez, en un escenario con pérdidas, para cada métrica y se puede generalizar después. Además, los factores de ponderación pueden ser calculados indistintamente en un escenario adaptativo o convencional. La calibración de los factores de ponderación ha sido realizada mediante un proceso de minimización del valor cuadrático medio del error (RMSE). Nótese que la métrica MOVIE tiene una relación inversa con las pérdidas, ya que la escala de la métrica MOVIE tiene una escala inversa con respecto a PSNR o SSIM.

4.4.4.1 Resultados

La diferencia entre los sistemas adaptativos y los convencionales es clara en términos de QoS. De hecho, el propio sistema adaptativo se retroalimenta de los parámetros de QoS para evaluar el estado de congestión del entorno y adaptarse a las condiciones. En este sentido es evidente que los parámetros de QoS mejoran en el sistema adaptativo. Diferencias en la tasa de pérdidas ya han sido incluidas en las figuras 4-9 y 4-10 Sin embargo, el impacto de los sistemas de adaptación en las métricas VQ no ha sido evaluado.

Como resultado de la aplicación del *framework* de análisis desarrollado, el trabajo realizado para elaborar un decodificador compatible con las secuencias con pérdidas, se puede presentar la comparativa entre las principales métricas de VQ para los dos sistemas, adaptativo y convencional. Los valores de las métricas, obtenidos como un promedio a lo largo de la secuencia son mostrados en las tablas 4-6, 4-7 y 4-8.

Tabla 4—6 VQ promedio para el escenario 1

	Medido		Estimado	
	No adaptativo	adaptativo	No adaptativo	adaptativo
PSNR	33.9137	39.0829	33.2542	40.0578
SSIM	0.78718	0.91395	0.78707	0.91929
MOVIE	0.000216	0.00012	0.000216	0.000075

Tabla 4—7 VQ promedio para el escenario 2

	Medido		Estimado	
	No adaptativo	adaptativo	No adaptativo	No adaptativo
PSNR	33.9137	28.1581	33.2542	32.4621
SSIM	0.78718	0.55574	0.78707	0.71400
MOVIE	0.000216	0.00091	0.000216	0.000884

Tabla 4—8 VQ promedio para el escenario 3

	Medido		Estimado	
	No adaptativo	adaptativo	No adaptativo	adaptive
PSNR	29.7728	32.6664	29.0834	33.3890
SSIM	0.58784	0.72295	0.55348	0.7156
MOVIE	0.00087	0.00050	0.00097	0.000837

Para analizar de forma conjunta la variación de la calidad medida se compilan las métricas en una misma representación gráfica para dos de los escenarios. Previamente es necesario normalizar las mismas debido a las diferentes escalas de medición, según la ecuación 4.3. En todos los casos, la utilización del sistema adaptativo ha supuesto un

incremento de la calidad objetiva medida. Así se refleja, además, en la métrica medida y en su versión estimada según el algoritmo propuesto (véase en la figura 4-11 que PSNR y SSIM son superiores para los escenarios adaptativos y MOVIE es inferior).

$$\widehat{VQ}' = \frac{\sum_{i=1}^N VQ_i}{\max(VQ)} \quad (4.3)$$

Todos los valores promedio mostrados corroboran la correlación existente entre las métricas medida y estimada. Analizando temporalmente las secuencia se aprecia cómo la métrica estimada es capaz de replicar las variaciones temporales de las métricas medidas. En la Figura 4-12 se muestra la representación temporal de la calidad de vídeo en función del tiempo para el escenario 3. Las gráficas se han diezmado, usando un filtro IIR de factor 8 (ya que la métrica MOVIE se calcula para *frame* en posiciones múltiplos de 8) para poder ser representadas de forma clara. Las ecuaciones propuestas permiten computar las métricas que de otra forma sería costoso evaluar, en especial en el caso de la métrica MOVIE, con un coste computacional elevado.

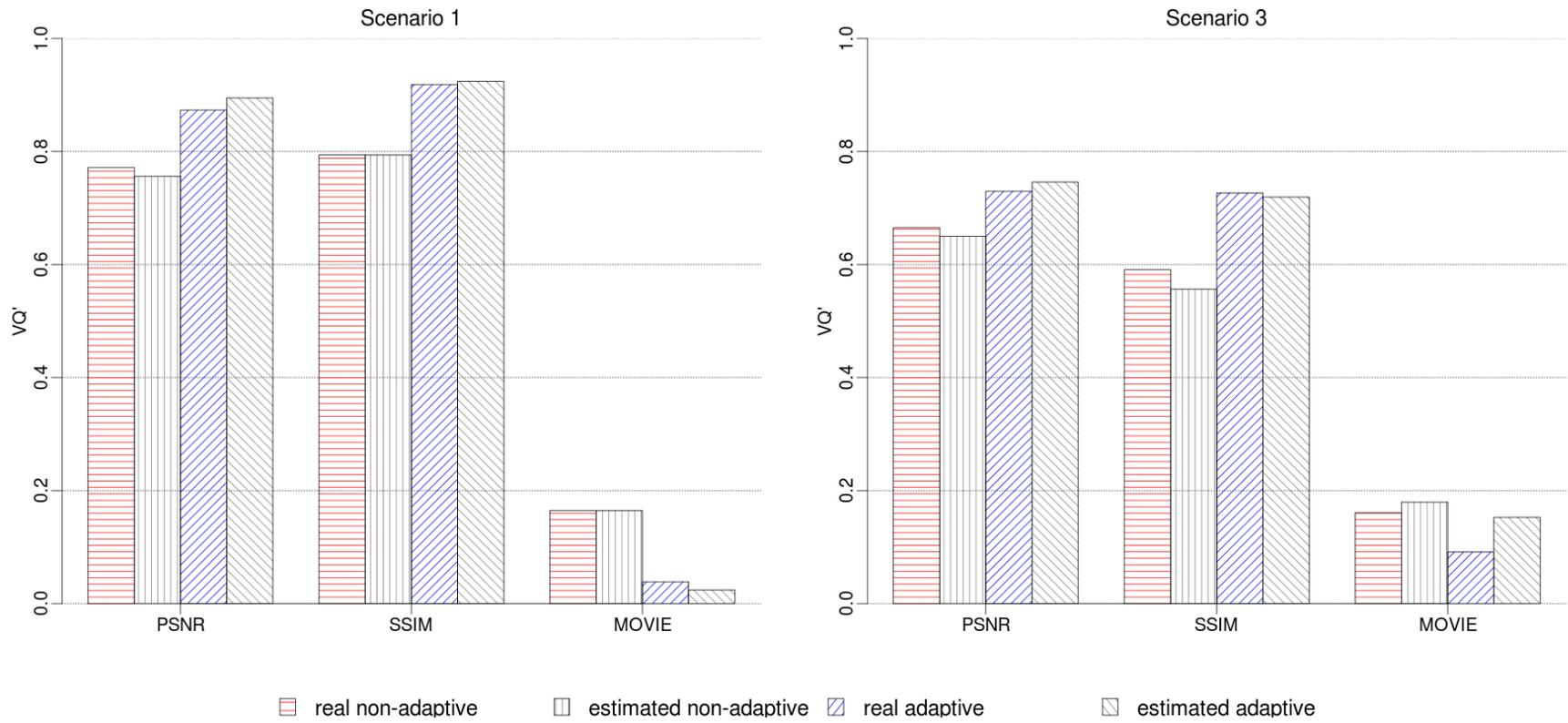


Figura 4-11: Comparativa de métricas normalizadas en dos escenarios

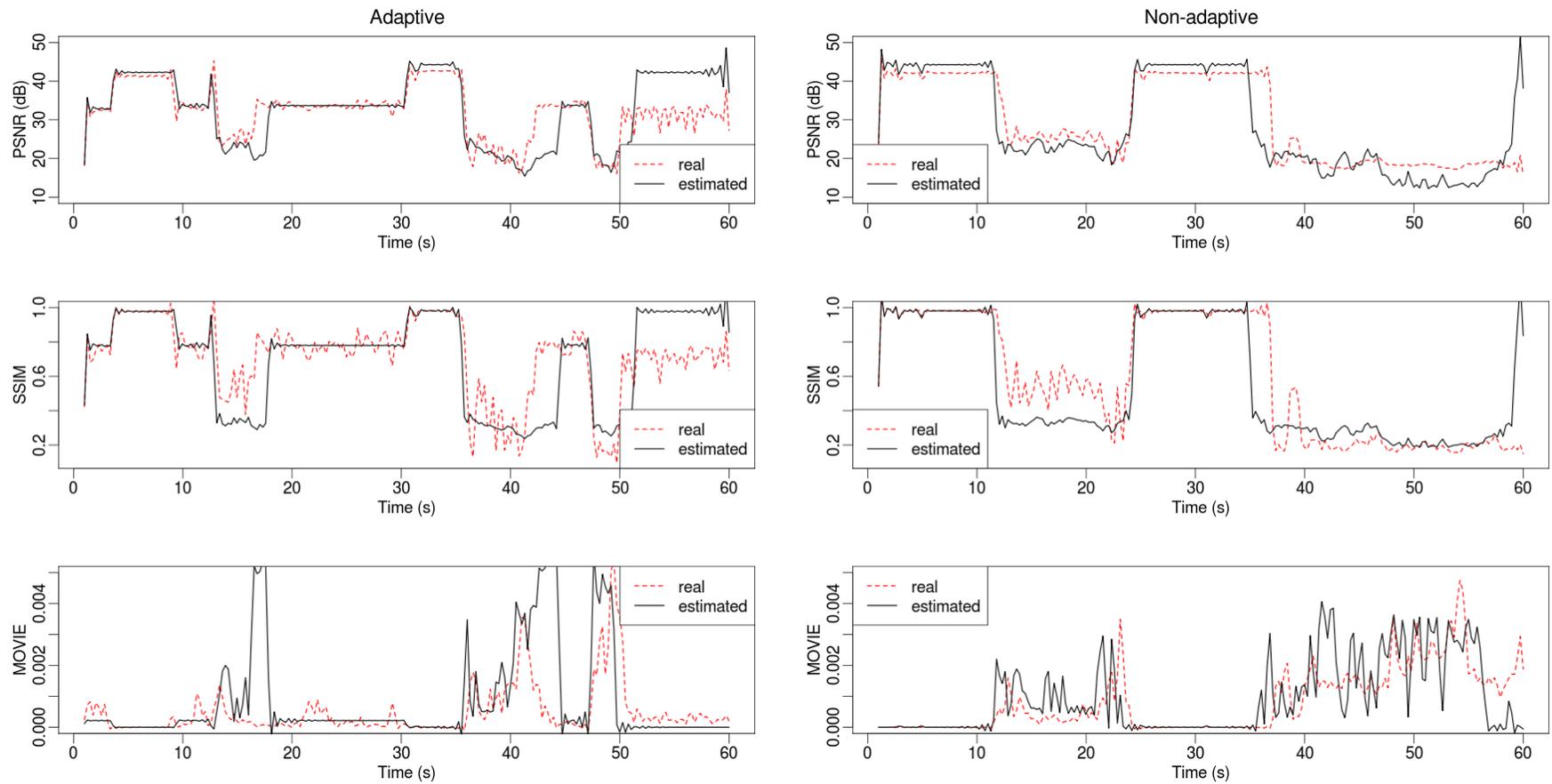


Figura 4-12: Representación temporal de VQ reales y estimadas

4.5 Contribuciones del capítulo

En las evaluaciones de los servicios multimedia se ha experimentado que la funcionalidad de las herramientas disponibles es importante para dar soporte a los estudios que se realizan. En el ámbito de la tecnología escalable es más importante dada la particularidad de su estructura de codificación, transmisión y decodificación. Los esfuerzos dedicados a implementar una estructura de programas, capaces de trabajar con secuencias reales escalables reales y en cualquier escenario posibilita la realización de análisis más precisos.

En este capítulo se describen varias contribuciones en el ámbito de las métricas objetivas enfocadas en sistemas adaptativos:

En primer lugar se construye un decodificador flexible, capaz de trabajar con secuencias reales que resultan dañadas como consecuencia de su transmisión por redes propensas a errores. El decodificador basado en *openSVC decoder* constituye un avance necesario para la evaluación de la métrica cPSNR pero no exclusivamente. Se abre la puerta a la experimentación con sistemas adaptativos reales, y no exclusivamente trazas de simulación, en situaciones de congestión sin que la aparición de pérdidas inhabilite la decodificación satisfactoria.

Para permitir la aplicación de la métrica PSNR como evaluador de las pérdidas de transmisión exclusivamente se define la métrica cPSNR, que mejora las asunciones consideradas en la literatura hasta la fecha. Este hecho se ha contrastado mediante la correlación con los resultados subjetivos públicos proporcionados por una popular base de datos de evaluaciones subjetivas.

Integrando la métrica desarrollada y otras ya existentes, se ha desarrollado la herramienta DMMSAF, que incorpora los mecanismos que permiten trabajar con secuencias reales, codificadas o no. Con el afán de habilitar el análisis de la calidad de transmisión del contenido escalable, DMMSAF incorpora mecanismos de sincronización configurables entre la secuencia original y la recibida tras el proceso de transmisión.

También se ha desarrollado un método de estimación de métricas FR en servicios adaptativos. El método de estimación habilita la utilización en tiempo real de las métricas FR más populares convirtiéndolas en métricas RR. Con la información de la calidad de cada capa transmitida al inicio de la transmisión y las modificaciones realizadas en el decodificador desarrollado, se permite estimar en recepción las métricas FR más conocidas.

Por último, se contrastan los resultados de las métricas estimadas (RR) con las métricas medidas (FR) equivalentes. Para ello se emplea un caso de estudio utilizando un sistema adaptativo real y con la ayuda del decodificador desarrollado.

De forma global, se concluye que las métricas objetivas de calidad deben ajustarse adecuadamente al entorno que miden y prueba de ello es la mejora introducida con la métrica cPSNR, que arroja resultados muy bien correlados con las métricas subjetivas. Los avances realizados en esta fase de la tesis permiten realizar mejores evaluaciones de servicios adaptativos.

5 Calidad de experiencia subjetiva

El universo de servicios multimedia en Internet es cada vez más extenso y variado. El rendimiento de los servicios sobre una red tan masificada está amenazado continuamente. Construir servicios de alto rendimiento pero sobre todo capaces de generar una buena experiencia en los usuarios es el principal objetivo perseguido por los operadores de dichos servicios.

Todos los problemas que se producen en el video transmitido a través de las redes afectan a la calidad percibida. No hay una idea clara del nivel de distorsión introducido en la calidad percibida por cada tipo de artefacto (o factor de perturbación). Para este objetivo, deben analizarse diferentes situaciones basándose en la opinión de los usuarios.

En la primera parte de este capítulo se desarrollan los experimentos subjetivos realizados para evaluar la influencia de distintos tipos de distorsiones sobre distintos contenidos de forma subjetiva. Hasta 100 usuarios con edades comprendidas entre los 22 y los 68 años participan en este experimento. Este estudio constituye uno de los más numerosos entre los experimentos subjetivos realizados bajo las directrices establecidas en las recomendación ITU-R BT.500-12 y ITU-R P.910. Es ciertamente mucho más numeroso que el mínimo recomendado en dichas publicaciones, establecido en 15 usuarios. La heterogeneidad de la población seleccionada proporciona una robustez mayor a los resultados. La calidad percibida evaluada por la población seleccionada aporta conocimiento sobre el efecto de artefactos como las pérdidas, la falta de sincronismo entre audio y vídeo o eventos de *rebuffering*. Dicho conocimiento es de utilidad en el diseño de servicios multimedia que deben priorizar aquellos aspectos que los usuarios evalúan más positivamente ante situaciones de congestión.

La introducción de la tecnología adaptativa, particularmente SVC, en la construcción de los servicios multimedia contribuye a mejorar la QoE partiendo de una situación comprometida. La adaptación de los contenidos de forma dinámica durante las situaciones adversas añade un factor más de ajuste que los servicios pueden utilizar para paliar la congestión. Ante una situación de congestión existe un umbral de tolerancia a partir del cual el usuario preferirá ver un contenido adaptado que un contenido distorsionado. Del mismo modo, existe un rango de indiferencia dentro del cual, subjetivamente hablando, no interesa aumentar la calidad del sistema adaptativo puesto que los usuarios no lo aprecian. En la segunda parte del capítulo se compara de forma

subjetiva un sistema adaptativo con un sistema convencional. El objetivo es dar respuesta a los factores de tolerancia e indiferencia que guían en la construcción de mejores sistemas adaptativos. Un total de 75 candidatos fueron invitados a realizar los test subjetivos de esta segunda fase, con edades comprendidas entre los 22 y los 68 años. Una vez más, el estudio subjetivo realizado incluye una población heterogénea y muy numerosa, mucho más alta que la de los trabajos existentes o que la recomendación.

Como contribución final, se analiza también la correlación de las métricas objetivas FR más comunes con las elecciones de los usuarios en la comparación de los sistemas adaptativos. El objetivo es determinar la capacidad de las métricas de representar las decisiones subjetivas.

5.1 Experimentos subjetivos end-to-end.

La falta de mecanismos garantes de QoS en los servicios multimedia en general, no exime de la aparición de degradaciones en la QoE. Puesto que la QoE no se puede garantizar, existe una necesidad de evaluar la influencia de los distintos factores, degradaciones, que afectan negativamente a la experiencia de usuario, para reducir su frecuencia de aparición en el servicio *end-to-end*.

Existen muchos tipos de degradaciones: errores en la decodificación, asincronía entre audio y vídeo o parones en la reproducción del servicio multimedia, entre otros. Conocer qué degradaciones son más relevantes en la calidad percibida es importante para realizar un mejor diseño y configuración del servicio multimedia. Para evaluar la calidad percibida es necesario realizar experimentos con usuarios reales, de forma que las distintas situaciones planteadas son evaluadas subjetivamente.

La calidad percibida global en la experiencia multimedia está relacionada con muchos factores. Entre los factores se incluyen problemas relacionados con el proceso de transmisión o con la reproducción de los contenidos. Sin embargo, desde un punto de vista *end-to-end*, el impacto de las distintas deficiencias en la QoE solo ha sido analizada para ciertos factores degradantes específicos [73] o con una población pequeña [74].

Los estudios subjetivos para la evaluación de nuevas tecnologías de codificación son los únicos que han sido tradicionalmente realizados usando un número ciertamente elevado de participantes [150]. Aparte de los estudios de el grupo VQEG (VQEG-RF y VQEG-HD), el estudio subjetivo realizado en esta Tesis y descrito en este apartado es el más numeroso entre los que analizan la degradación del audio y vídeo ante distintas fuentes de distorsión. Según el trabajo publicado en [151] que analiza hasta 13 bases de datos públicas de estudios subjetivos, se extrae que nuestro trabajo, entre las evaluaciones subjetivas, es el que mayor porcentaje de participantes femeninas presenta, y de un rango de edad más amplio.

Una reciente contribución, no mencionada en el citado trabajo es la realizada en [152]. En él los autores realizan estudios subjetivos con 41 usuarios y 184 secuencias distorsionadas con *jitter*, retrasos o pérdidas. Sin embargo, no hay referencia a la localización de dichas distorsiones en la secuencia o análisis de la influencia del contenido

En esta parte del trabajo, analizaremos la importancia que el usuario proporciona a distintos aspectos del servicio multimedia, evaluando la calidad percibida de un contenido procesado para incorporar pérdidas, eventos de *rebuffering* o asincronía entre el audio y el vídeo. Además, hemos tenido en cuenta el tipo de contenido y la influencia de la posición y duración de la degradación dentro de las secuencias. La población invitada, hasta 100 participantes con edades comprendidas entre 20 y 68 años, proporciona una evaluación general de la experiencia multimedia que puede ser utilizada para priorizar aspectos que influyen más negativamente la calidad global de experiencia en el diseño de los servicios multimedia.

5.1.1 Metodología de los experimentos

La elección de la metodología subjetiva está condicionada por el propósito del experimento. Del catálogo de metodologías subjetivas propuestas por la ITU-T, referenciadas en el estado del arte de este trabajo, se elige para este experimento la metodología SSACJ (*Single Stimulus Adjectival Categorical Judgement*) [153]. Según esta metodología, las secuencias se presentan de forma individual, cada secuencia corresponde a un contenido diferente con un procesamiento distinto. Tras la visualización de cada secuencia los participantes son interrogados acerca de la calidad de la imagen, la calidad del audio y la calidad global según su experiencia, usando una escala de categorías.

La metodología SSACJ y la escala seleccionadas tienen el objetivo de evaluar la influencia de los distintos tipos de degradaciones y su relación con el tipo de contenido. La escala de 5 categorías es mapeada en la escala MOS para indicar la tolerancia de los usuarios hacia los distintos problemas que se producen en el servicio cuando llega al cliente, por tanto, *end-to-end*.

Siguiendo las pautas recomendadas proporcionadas por la metodología subjetiva, los participantes son instruidos para la realización de los experimentos. Una sesión de entrenamiento explica el procedimiento, incluyendo la explicación de la herramienta *software* diseñada para los experimentos y la descripción de la escala de evaluación. Después de la sesión de entrenamiento individual, los usuarios completan sus datos personales y un cuestionario acerca de su experiencia en el universo multimedia. La información personal solicitada incluye edad, sexo, educación, experiencia con sistemas multimedia (solo TV, video en Internet, o vídeo en alta definición (HD) en Internet), dispositivos usados para visualizar el contenido multimedia (TV, PC, Tablet, Smartphone).

Para mantener la duración del experimento contenida, se establecen grupos de control diferentes de forma que cada grupo de control visualiza un subconjunto de las secuencias tratadas. El test se compone de 8 secuencias diferentes, cuyo orden de presentación se elige de forma aleatoria entre usuarios, para evitar que los efectos del cansancio o el aprendizaje se reflejen en los resultados. Cada secuencia tiene una duración aproximada de 2 minutos. Los usuarios pueden repetir la visualización de las secuencias tantas veces como deseen. Esta posibilidad permite ofrecer una segunda visualización a un usuario indeciso y al mismo tiempo evitar una evaluación incompleta por la ocurrencia de distracciones ajenas al experimento. Tras la visualización, el usuario elige en la escala presentada la calidad percibida para cada secuencia, audio, vídeo y experiencia global. La duración de la prueba total, incluyendo la votación asciende a los 23.15 minutos de media. Al final del experimento los usuarios reciben un cuestionario (escala Likert) para valorar la dificultad y la duración del experimento. Los resultados se muestran en la Tabla 5.1. En general los experimentos no son considerados como difíciles. La consideración negativa más generalizada es la duración del experimento.

Tabla 5—1 Resultados de la encuesta final de los experimentos subjetivos *end-to-end*

Escala Likert (1-5)	Valor promedio (desviación estándar)
Dificultad general (1 muy fácil – 5 muy complejo)	1.47 (0.71)
Duración del test (1 muy corto – 5 muy largo)	2.76 (0.8)
Complejidad de la aplicación (1 muy fácil– 5 muy compleja)	1.34 (0.83)
Sistema de votación (1 muy sencillo – 5 muy complejo)	1.66 (0.98)

5.1.2 Entorno y equipamiento

En la comunidad existen diversas herramientas *software* que implementan varias de las metodologías de experimentación subjetiva. Entre otras. ACRVQWin¹⁴, VQLab¹⁵ o MSU VQMT¹⁶. Sin embargo, ninguna de ellas cumple todos los requisitos definidos para el experimento propuesto. En consecuencia, una herramienta específica para dar soporte a la evaluación subjetiva es desarrollada en el transcurso de la investigación. El *software* se desarrolla utilizando Qt¹⁷ y el *framework* Phonon¹⁸, una plataforma multimedia

¹⁴ <https://www.acreo.se/acrvqwin>

¹⁵ <http://www.semaca.co.uk/content9552.html?produs=2>

¹⁶ http://compression.ru/video/quality_measure/video_measurement_tool_en.html

¹⁷ <http://www.qt.io/>

¹⁸ <http://doc.qt.io/qt-4.8/phonon-module.html>

compatible con distintos sistemas operativos. La apariencia de la aplicación se ha modificado de acuerdo a las recomendaciones de la metodología subjetiva (Figura 5-1).

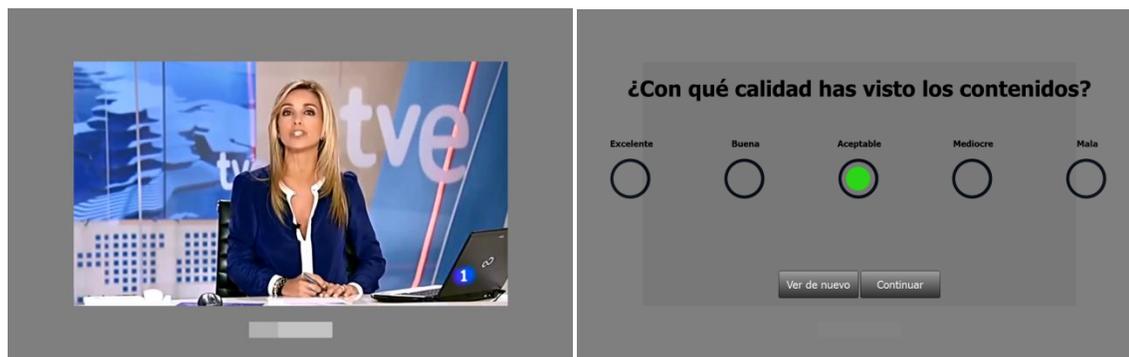


Figura 5-1: Aspecto de la aplicación desarrollada para los experimentos subjetivos *end-to-end*

Para la reproducción de los contenidos durante el experimento se utiliza un ordenador de alto rendimiento y un monitor de 23 pulgadas. El detalle del equipamiento se proporciona en las tablas 5.2 y 5.3. De acuerdo con las condiciones del experimento recomendadas en la metodología, los experimentos tienen lugar en una oficina privada y tranquila, con unas condiciones de iluminación media. La distancia del usuario al monitor se mantiene entre 1.5 y 2 metros.

Tabla 5—2 Detalles del equipo usado en los experimentos

Placa Base	ASUS P8H61-M LE-USB3
Procesador	Intel Core i7
Gráficos	ATI HD6570 1GB DDR3 PCI-E PowerColor
RAM	8 GB DDR3
SSD	128 GB Series 830 Notebook Kit Samsung
Sistema Operativo	Windows 7 Home Premium 64-bit

Tabla 5—3 Detalles del monitor usado en los experimentos

Monitor LCD	Philips led modelo 236V3L
Tamaño de la diagonal	23 pulgadas
Resolución	1920x1080 (Full HD)
Contraste dinámico	5000000:1
Tiempo de respuesta	5 ms
Interfaz	D-Sub, DVI & HDMI
Ángulo de visión	170/160°

5.1.3 Contenidos

Se seleccionan diferentes contenidos originales para generar el contenido procesado que será incluido en el test. Los contenidos usados en el estudio incluyen noticias, deportes, dibujos y secuencias de acción. Estas categorías se eligen en base a las diferentes categorías clásicas en los contenidos multimedia y se analiza el efecto de los diferentes defectos en cada tipo de contenido. Una secuencia adicional se usa en la sesión de entrenamiento. Existen dos secuencias por cada tipo de contenido (noticias, dibujos, etc.). En la figura 5-2 se muestra una instantánea de cada secuencia. Los índices de complejidad espacial y temporal se muestran en la Tabla 5.4

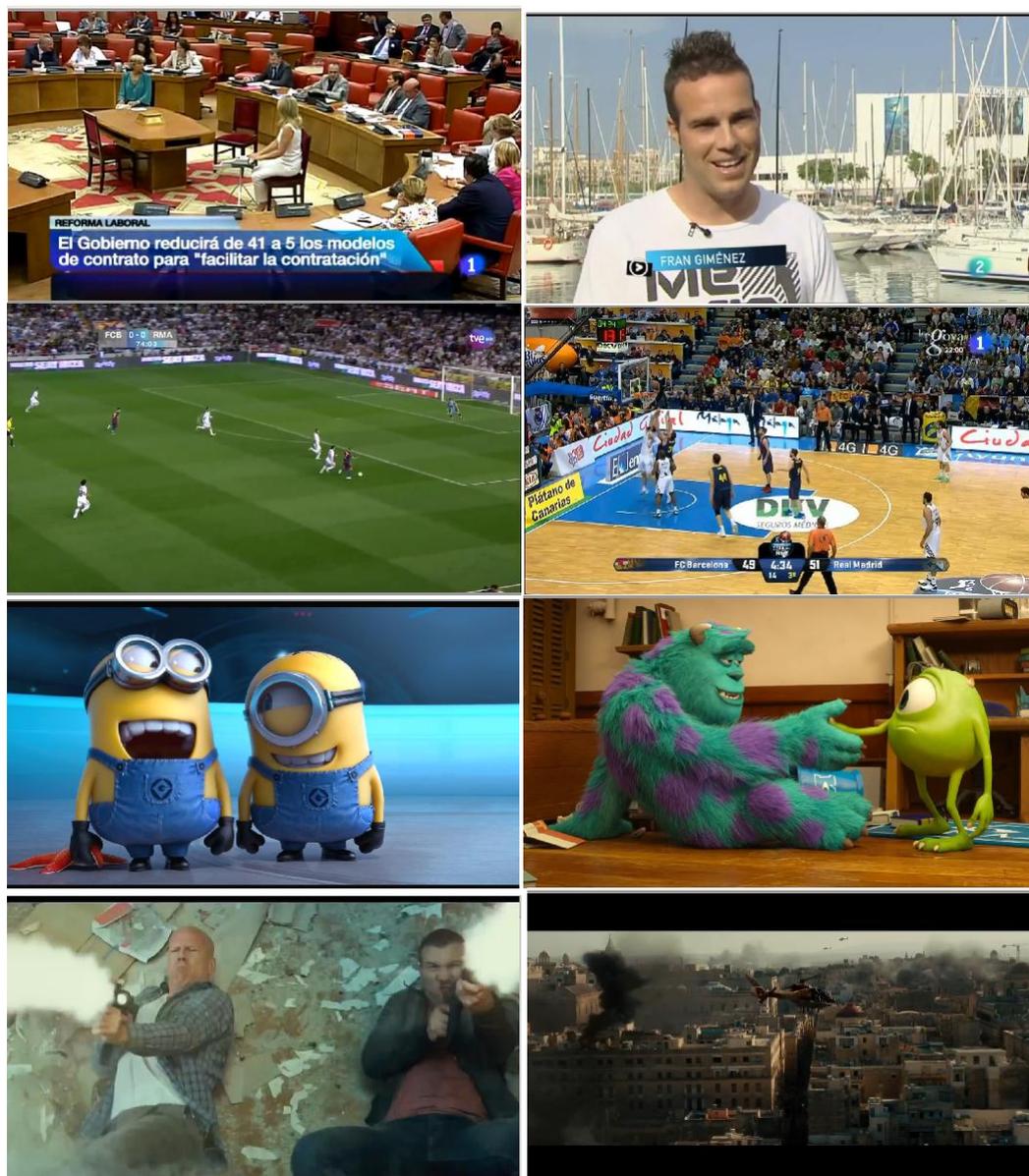


Figura 5-2: Instantáneas de los contenidos del experimento subjetivo

Tabla 5—4 Descripción de los contenidos del experimento subjetivo

Contenido	Índice SI	Índice TI	Detalles [resolución@frameRate, duración]
News1	33.8629	91.5806	1280 x 720@25 fps, 2'28"
News2	26.1874	84.9127	1280 x 720@25 fps, 2.17
Sports1	19.3815	78.9768	1280 x 720@25 fps, 2'
Sports2	25.8234	81.7493	1280 x 720@25 fps, 2' 17"
Cartoons1	23.2876	77.4565	1280 x 720@24 fps, 2'25"
Cartoons2	30.7635	87.9173	1280 x 720@24 fps, 2'18"
Action1	42.9343	93.8401	1280 x 720@25 fps, 2'18"
Action2	34.8185	90.8393	1280 x 720@24 fps, 2'19"

Cada contenido es procesado considerando diez modelos de distorsión, *impairment models (IM)* diferentes. Los tipos de distorsión comprenden eventos de *rebuffering* o *stalling (S)*, cortes en el audio o el vídeo (C), asincronía entre audio y vídeo (A) y pérdidas de calidad en la imagen (P). Además, en los modelos se introducen los siguientes modificadores. Según el tipo de distorsión, pérdidas o eventos de *rebuffering* pueden desarrollar patrones a ráfagas o *burst (B)*, patrones largos (L). Las pérdidas o la asincronía pueden ocurrir al inicio (S) o al final (E) de la secuencia. De la combinación de estos factores se obtienen los diez modelos de distorsión que se presentan a continuación.

- **Eventos de rebuffering (stalling) [IM-S].** Las secuencias afectadas por este patrón de distorsión comienzan con un proceso de carga de 5 segundos. Luego, a intervalos de 30 segundos, tiene lugar un nuevo evento de carga o *rebuffering*. Durante el proceso de *rebuffering*, la reproducción del audio y del video se detiene para continuar después del evento en el mismo punto. En este tipo distorsiones no se produce pérdida de información. Se establecen dos tipos de patrones en relación a la distribución de la duración de los eventos de carga. De larga duración [IM-SL], Figura 5-3, con una duración de 10 segundos o en ráfaga [IM-SB], Figura 5-4, con cortes de duraciones de 1 o 2 segundos. En ambos casos la suma total de los periodos de *rebuffering* es de 35 segundos

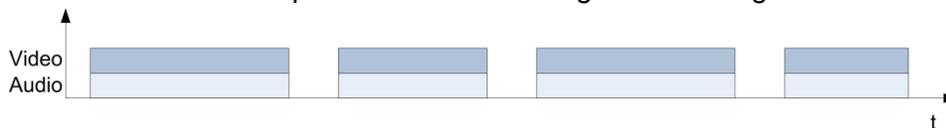


Figura 5-3: Patrón rebuffering largo [IM-SL]



Figura 5-4: Patrón rebuffering ráfaga [IM-SB]

- **Cortes en audio o video [IM-C].** Al contrario que en el caso anterior, con este patrón se introducen pérdidas en el vídeo [IM-CV] o en el audio [IM-CA] aunque nunca en los dos al mismo tiempo. Los cortes se distribuyen en intervalos de 30 segundos y pueden corresponderse a patrones largos de 10 segundos [IM-CAL, IM-CVL] (Figura 5-5 y 5-6) o a patrones de ráfaga [IM-CAB, IM-CVB] (Figura 5-7 y 5-8) también con una duración total de 10 segundos.

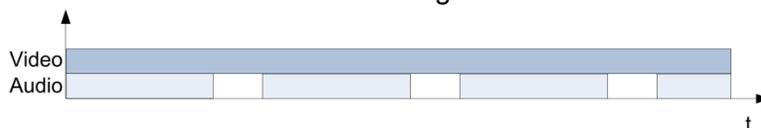


Figura 5-5: Pérdidas audio largas [IM-CAL]

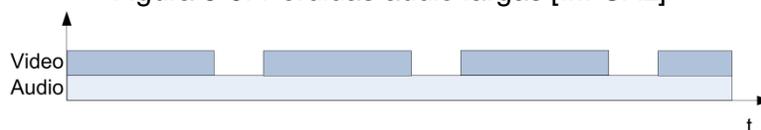


Figura 5-6: Pérdidas video largas [IM-CVL]

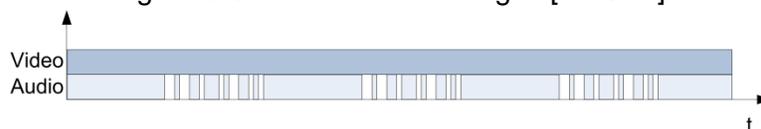


Figura 5-7: Pérdidas audio ráfaga [IM-CAB]

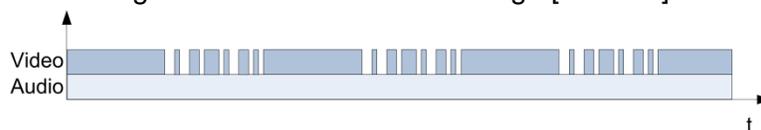


Figura 5-8: Pérdidas video ráfaga [IM-CVB]

- **Asincronía entre audio y vídeo [IM-A].** En este patrón el audio es retrasado un segundo con respecto al vídeo. La falta de sincronía puede ocurrir al inicio de la secuencia [IM-AS] (Figura 5-9) o al final [IM-AE] (Figura 5-10).

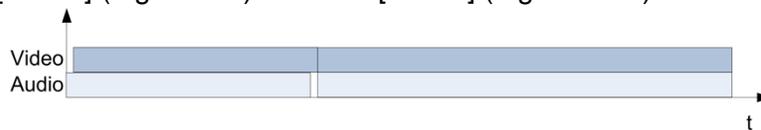


Figura 5-9: Asincronía vídeo/audio en el inicio [IM-AS]



Figura 5-10: Asincronía vídeo/audio en el final [IM-AE]

- **Baja calidad de imagen (pixelación) [IM-P].** La imagen del vídeo sufre los efectos del pixelado durante un intervalo total de 40 segundos, distribuidos en ráfagas,

bien al principio [IM-PS] (Figura 5-11) o en la segunda mitad de la secuencia [IM-PE] (Figura 5-12).

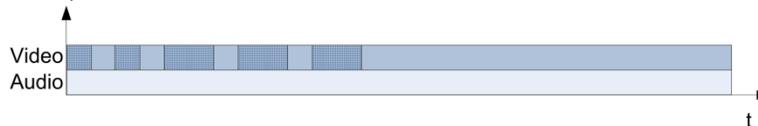


Figura 5-11: Baja calidad de imagen en el inicio [IM-PS]

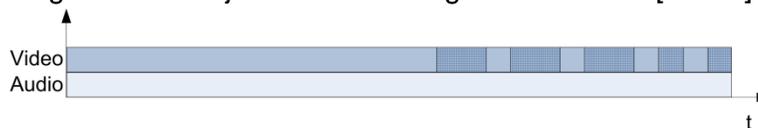


Figura 5-12: Baja calidad de imagen en el final [IM-PE]

5.1.4 Población

En total, 100 usuarios, 62 hombres y 38 mujeres, participaron en el estudio subjetivo. La edad de los participantes está comprendida entre 20 y 68 años, con un valor medio de 39.8 y una desviación típica de 13.08. De esta población, el 43% es menor de 35 años.

En términos de experiencia con servicios multimedia, el 20% de la población declara que sólo consume televisión convencional, mientras que el 61% admite ver habitualmente vídeos a través de Internet y un 19% ser habituales consumidores de contenido en alta definición. Los dispositivos más habituales para consumir los contenidos son televisión (31.85%) y PC (31.05%) seguidos de Smartphone (21.37%) y Tablet (15.73%).

Dada la heterogeneidad y masa de la población de nuestro estudio, éste se sitúa entre los más completos y variados con respecto a la población, excluyendo los experimentos de *crowdsourcing*. De hecho los trabajos basados en *crowdsourcing* requieren por su naturaleza poblaciones más numerosas, pues están sujetos a filtrados adicionales para garantizar la robustez de los resultados [93], [94].

5.1.5 Análisis de los resultados

Las ocho secuencias fuente incluidas en este experimento se agrupan en cuatro categorías. Los resultados son expresados en función de estas categorías. Los resultados medios de MOS obtenidos de la evaluación subjetiva de cada participante son procesados estadísticamente para determinar las verdaderas diferencias entre distintas categorías y tipos de distorsión según los resultados subjetivos. Se han usado los procedimientos estadísticos más comunes, de acuerdo a lo sugerido en [154]. Estos test requieren precondiciones con respecto a la normalidad y homocedasticidad de las muestras. En primer lugar se comprueba la normalidad de los resultados con test de *Shapiro-Wilk* y de homocedasticidad con el correspondiente test de *Barlett*. Cuando los datos cumplen ambas condiciones los resultados se comparan mediante tests *ANOVA* de una muestra. Si la homocedasticidad no se cumple, se usan tests *Kruskal-Wallis*, al igual

que en los casos de una marcada falta de normalidad en los datos. Como caso particular de *Kruskal-Wallis* se utiliza el test de *Wilcoxon* para dos muestras. Finalmente, ante la existencia de diferencias, se utilizan test como *Student-t* o *Tukey* con intervalos de confianza al 95% para realizar comparativas dos a dos.

En términos generales, los resultados globales considerando únicamente los diferentes tipos de distorsiones (sin considerar la influencia del tipo de contenido) muestran que el tipo de distorsión más positivamente evaluado es la desincronización de audio/vídeo, figura 5-13. En este caso, los participantes han evaluado las secuencias del grupo IM-A con una media de 3.55 en la escala MOS (véanse detalles de los parámetros estadísticos en la tabla 5-5). Es conocido que incluso en condiciones de evaluaciones subjetivas en entornos relajados, entornos reales, los usuarios son capaces de detectar asincronías tan leves como 240 ms [92]. Por tanto, el retraso entre audio/vídeo de un segundo utilizado en nuestros experimentos es inicialmente considerada como una distorsión altamente molesta. Sin embargo, los resultados subjetivos no corroboran nuestra hipótesis.

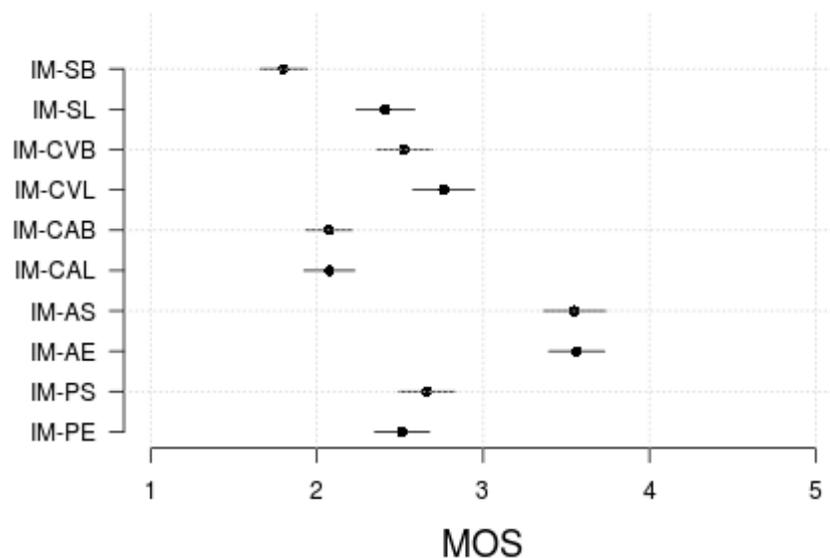


Figura 5-13: Resultados MOS según el tipo de distorsión

Tabla 5—5 Parámetros estadísticos para los distintos patrones de distorsión

	Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
IM-SB	1.80	0.72	1.64	1.96
IM-SL	2.41	0.90	2.21	2.61
IM-CVB	2.52	0.86	2.33	2.72
IM-CVL	2.76	0.94	2.55	2.97
IM-CAB	2.07	0.71	1.92	2.23
IM-CAL	2.07	0.78	1.90	2.25
IM-AS	3.55	0.95	3.34	3.76
IM-AE	3.56	0.85	3.37	3.75
IM-PS	2.66	0.87	2.47	2.86
IM-PE	2.51	0.84	2.32	2.70

En el otro extremo, de acuerdo a los resultados obtenidos, el tipo de distorsión más negativamente evaluado son los eventos de *rebuffering* en formato ráfaga (IM-SB), con los valores medios de MOS más bajos. En cambio, los eventos de *rebuffering* más largos (IM-SL) no son evaluados tan negativamente por los usuarios. En este sentido, se corroboran las conclusiones en [155] que también apoyaban que un evento de *rebuffering* largo es preferible a varios cortos y frecuentes.

Comparando entre los modelos de distorsión propuestos, no hemos encontrado diferencias significativas entre asincronía en el audio (IM-A) o deficiencias en la calidad (IM-P) emplazadas en el inicio o el final de la secuencia ($p\text{-value}>0.1$). Tampoco hay diferencias significativas entre los modelos de cortes en el audio propuestos (IM-CA). Sin embargo, los patrones de distorsión de cortes en el vídeo (IM-CV) y *rebuffering* (IM-S) sí que presentan diferentes resultados según el modelo analizado.

La figura 5-14 y la Tabla 5-6 muestran los valores MOS y los parámetros estadísticos de cada caso relacionándolos con el tipo de contenido, sin considerar la influencia del tipo de patrón de distorsión empleado. Encontramos que los usuarios son más tolerantes a los problemas en la reproducción cuando los contenidos son de tipo dibujos animados, con el mayor valor medio de MOS (2.86). Los autores en [156] corroboran este fenómeno. Además los contenidos de deportes y noticias son los contenidos más afectados por las distorsiones, según la calidad percibida reportada. Un test de Kruskal-Wallis revela que las medias de los grupos de contenido son significativamente diferentes ($p\text{-value}=0.00145$).

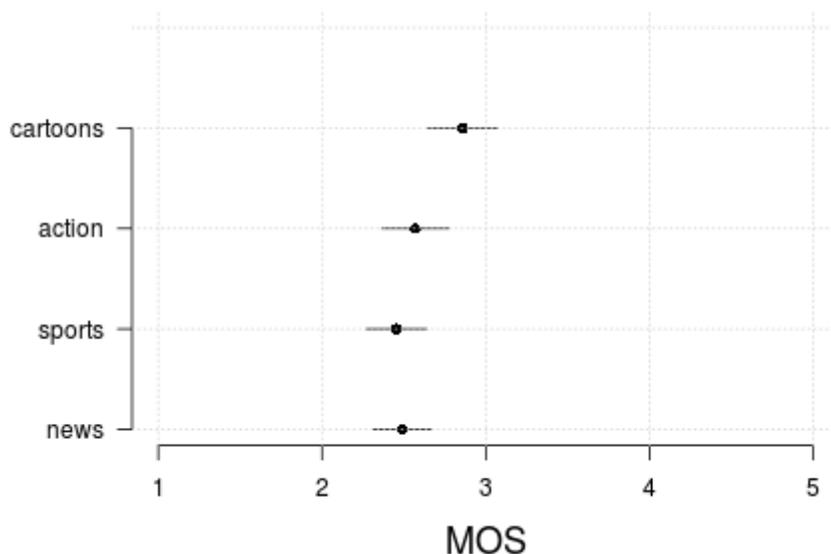


Figura 5-14: Resultados MOS según el tipo de contenido

Tabla 5—6 Parámetros estadísticos para distintos tipos de contenido

	Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Cartoons	2.86	1.08	2.71	3.01
Action	2.57	1.04	2.42	2.71
Sports	2.45	0.94	2.32	2.59
News	2.49	0.91	2.36	2.62

Para el análisis de la relación entre los diferentes tipos de contenido y los patrones de distorsión, en la figura 5-15 se presentan los valores MOS medios para el tipo de distorsión IM-S, el peor valorado por los usuarios. En la parte superior, los resultados con el patrón de distorsión de eventos de *rebuffering* tipo ráfaga (IM-SB) y en la parte inferior los correspondientes a eventos *rebuffering* de larga duración (IM-SL). Por columnas los tipos de contenido asociado a cada resultado. En las leyendas de las figuras se adjunta la información de valores medios, medianas y factores de asimetría de la distribución de resultados (*skwneess*) con el siguiente formato: *stats:(media, median, skewness)*. El resumen de los parámetros estadísticos de estos resultados se muestra en la tabla 5-7.

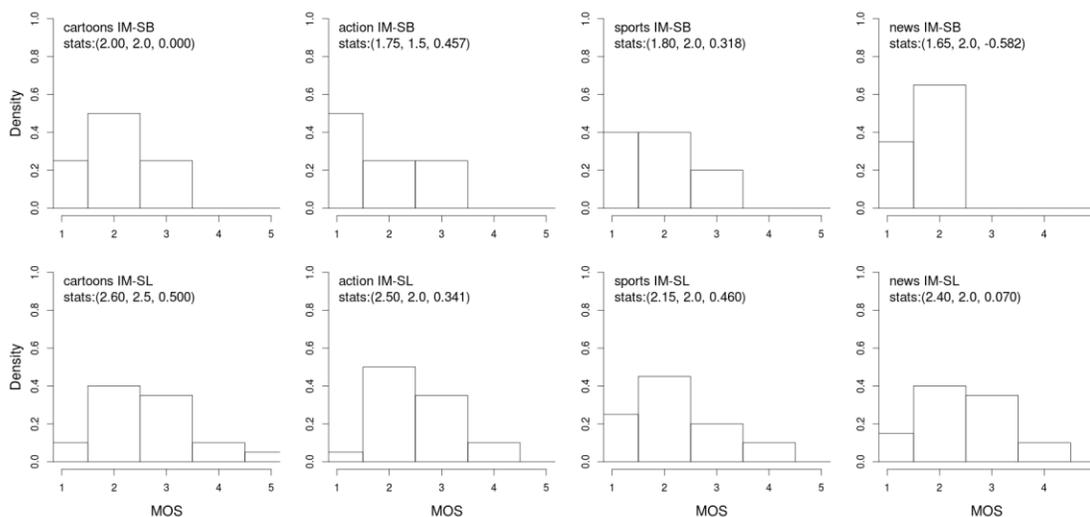


Figura 5-15: Resultados MOS para IM-S según el tipo de contenido

Tabla 5—7 Parámetros estadísticos para distorsión IM-S

		Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Eventos rebuffering en ráfaga (IM-SB)	Cartoons	2.00	0.72	1.66	2.34
	Action	1.75	0.85	1.35	2.15
	Sports	1.80	0.77	1.44	2.16
	News	1.65	0.49	1.42	1.88
Eventos de rebuffering largo (IM-SL)	Cartoons	2.60	0.99	2.13	3.07
	Action	2.50	0.76	2.14	2.86
	Sports	2.15	0.93	1.71	2.59
	News	2.40	0.88	1.99	2.81

Según los resultados, aunque las duraciones globales de los tiempos de *rebuffering* son idénticas, las secuencias son evaluadas más negativamente si los eventos ocurren en ráfagas. Los resultados de *Wilcoxon* muestran que para el tipo de contenido deportivo, las diferencias entre los patrones ráfagas o continuo no son relevantes.

Los patrones de distorsión basados en audio son más negativamente evaluados según el tipo de contenido. El análisis detallado para el patrón de distorsión IM-A es representado en la figura 5-16. Los valores de la fila superior corresponden a las distribuciones de resultados MOS para las secuencias con patrón a ráfagas (IM-CAB) mientras que la fila inferior corresponde al patrón continuo (IM-CAL). Según los resultados gráficos, por la media y asimetría de los histogramas, el contenido basado en noticias se ve más

afectado por estos patrones de distorsión, igual que el caso de contenido de acción y pérdidas continuas de audio. En general los cortes de audio en ráfagas no son tan negativamente evaluados como los cortes continuos. En la Tabla 5-8 se muestra el resumen de los parámetros estadísticos relevantes para el grupo analizado. En virtud de los resultados no existen diferencias estadísticamente significativas entre ambos patrones de distorsión de audio en función del tipo de contenido ($p\text{-value} > 0.1$). Sin embargo, sí que existe una diferencia para el patrón IM-CAL en función del tipo de contenido. En particular, según el test de *Tukey*, excepto en los pares *sports-action* y *sports-news*, las medias no son diferentes (incluyen el cero en el intervalo de confianza), véase Figura 5-17.

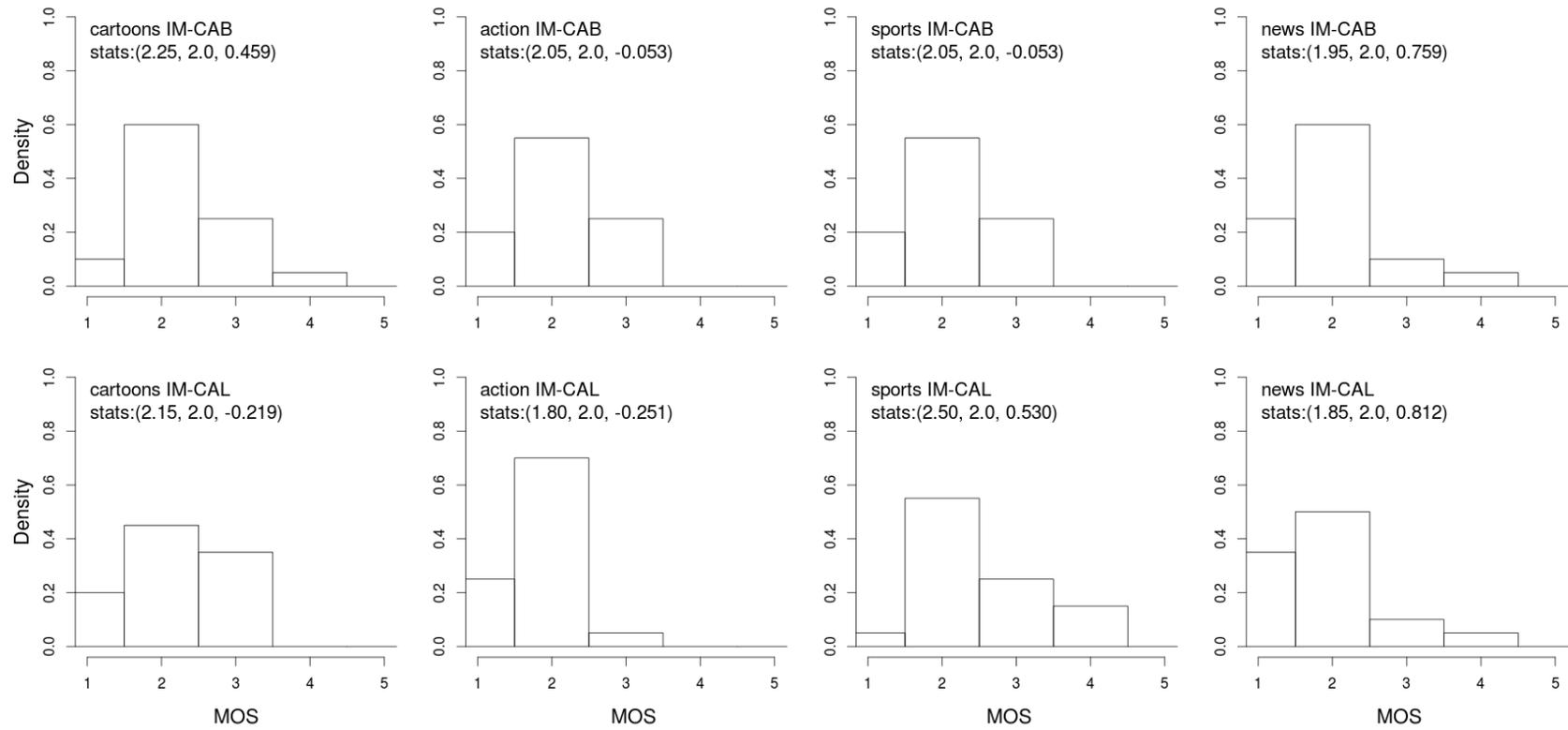


Figura 5-16: Resultados MOS para IM-CA según el tipo de contenido

Tabla 5—8 Parámetros estadísticos para distorsión IM-CA

		Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Patrón pérdidas audio en ráfaga (IM-CAB)	Cartoons	2.25	0.72	1.91	2.58
	Action	2.05	0.69	1.73	2.37
	Sports	2.05	0.69	1.73	2.37
	News	1.95	0.76	1.59	2.30
Patrón pérdidas de audio largas (IM-CAL)	Cartoons	2.15	0.74	1.80	2.50
	Action	1.80	0.52	1.55	2.04
	Sports	2.50	0.83	2.11	2.89
	News	1.85	0.81	1.47	2.23

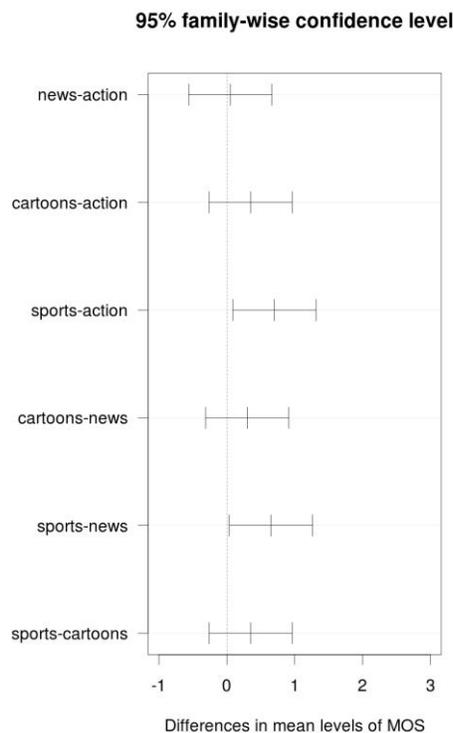


Figura 5-17: Resultados Test *Tukey* para IM-CAL

Los mismos contenidos analizados para el caso de distorsión en el audio, presentan, en el caso de tipos de contenido de noticias y acción, mayores valores subjetivos cuando el patrón de distorsión se corresponde con pérdidas en el vídeo, figura 5-18. Este resultado adelanta que este tipo de contenido es más sensible a pérdidas en el audio que en el vídeo. En la Tabla 5-9 se muestran los parámetros estadísticos relevantes del análisis de los patrones de pérdidas en vídeo (IM-CVB, IM-CVL). Los test de dos contraste de dos

muestras revelan que los resultados para los patrones en el contenido de tipo acción son estadísticamente diferentes ($p\text{-value}=0.04$)

Tabla 5—9 Parámetros estadísticos para distorsión IM-CV

		Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Pérdidas de video en ráfaga (IM-CVB)	Cartoons	2.75	0.79	2.38	3.12
	Action	2.30	0.80	1.92	2.67
	Sports	2.30	1.03	1.82	2.78
	News	2.75	0.72	2.41	3.08
Pérdidas de video largas (IM-CVL)	Cartoons	2.45	0.83	2.06	2.84
	Action	2.95	1.15	2.41	3.49
	Sports	2.55	0.94	2.11	2.99
	News	3.10	0.72	2.76	3.44

El concepto de asincronía entre audio y vídeo es, de acuerdo a los resultados obtenidos, augustamente tolerado por los usuarios. En el caso de las noticias la consideración es un poco más restrictiva. En este caso los errores entre la sincronización de audio y vídeo son más obvios para el espectador. En la figura 5-19 y la Tabla 5-10 se muestran en destalle los resultados para este grupo. No se encuentran diferencias en la calidad percibida por los usuarios entre una asincronía que se produce al inicio de la secuencia o al final. Otros autores evaluaron y modelaron el resultado de la QoE como consecuencia de la diferencia espacial entre audio y vídeo [157]. En este caso profundizamos el estudio para añadir las diferencias según el tipo de contenido. En este sentido, los test de contraste de Kruskal-Wallis muestran que para ambos modelos de distorsión (IM-AS e IM-AE) los resultados son significativamente diferentes ($p\text{-value} < 0.01$). Un test de *Tukey* para los casos de distorsión IM-AS corrobora que solo para el caso de los pares "cartoon-news" y "cartoon-sports" las medias pueden considerarse diferentes. En el caso de IM-AE lo mismo ocurre para los pares "sports-news" y "cartoon-action", sus medias son las únicas que puede decirse que no son iguales.

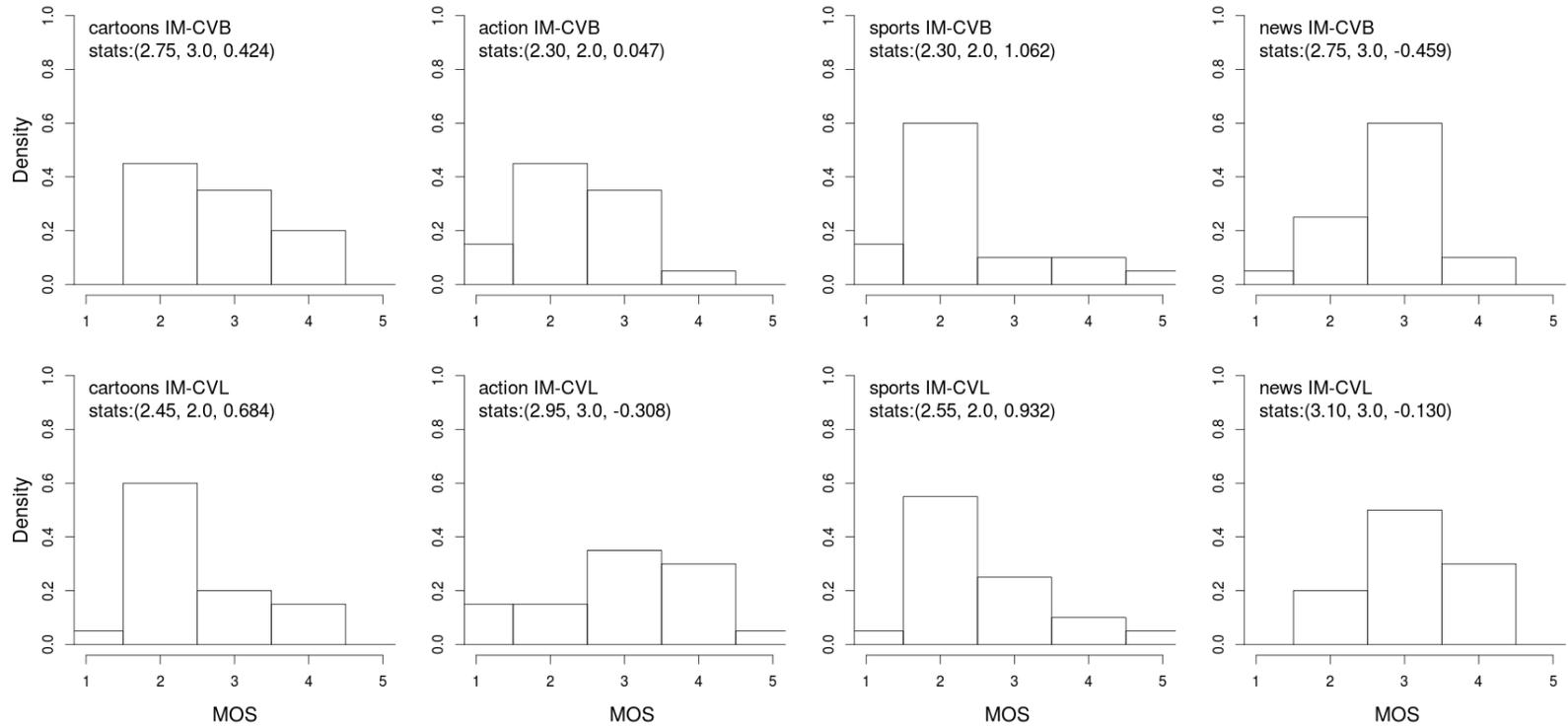


Figura 5-18: Resultados MOS para patrón de pérdidas en el vídeo.

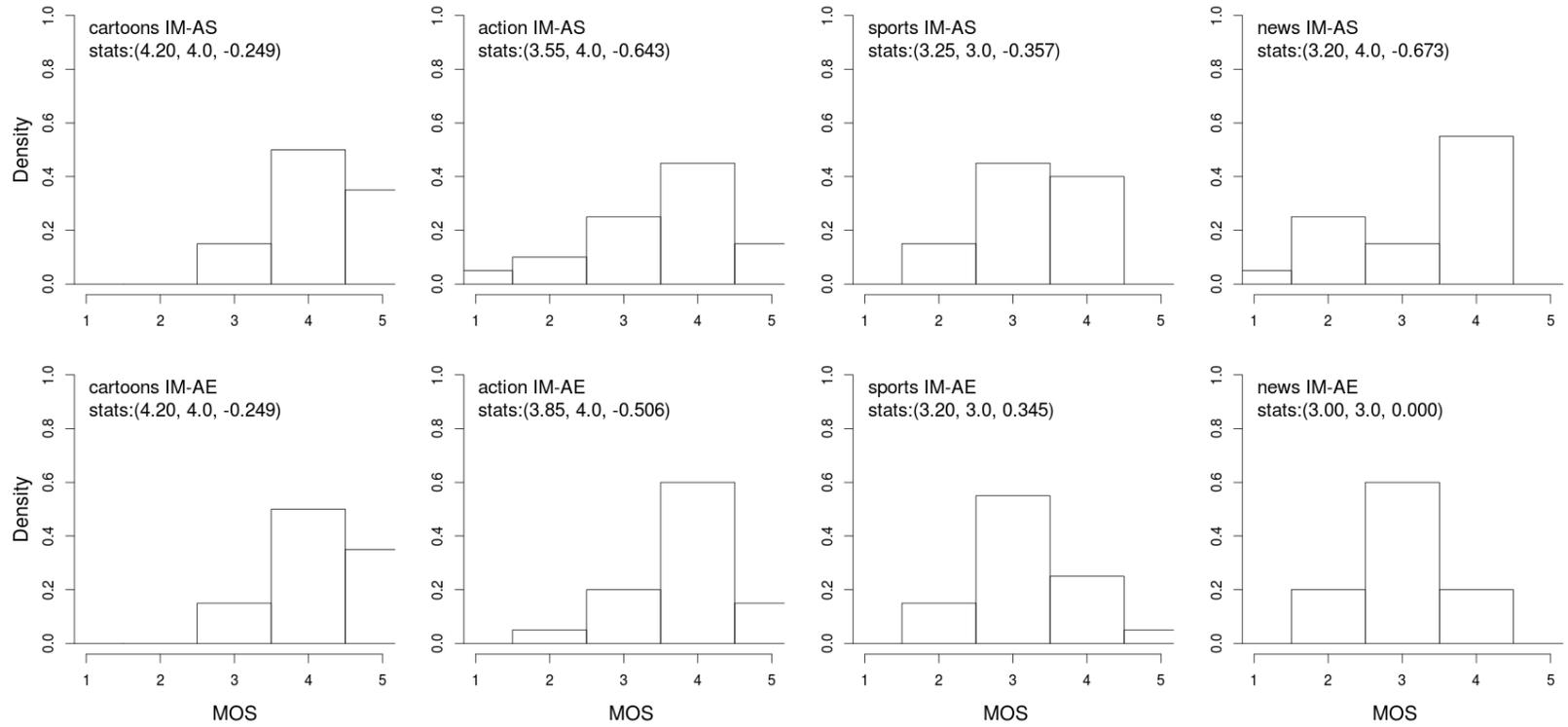


Figura 5-19: Resultados MOS para patrón de asincronía en audio-vídeo

Tabla 5—10 Parámetros estadísticos para distorsión IM-A

		Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Asincronía en la primera mitad (IM-AS)	Cartoons	4.20	0.70	3.87	4.53
	Action	3.55	1.05	3.06	4.04
	Sports	3.25	0.72	2.91	3.58
	News	3.20	1.00	2.73	3.67
Asincronía en la segunda mitad (IM-AE)	Cartoons	4.20	0.70	3.87	4.53
	Action	3.85	0.74	3.50	4.20
	Sports	3.20	0.77	2.84	3.56
	News	3.00	0.65	2.70	3.30

Para el último de los grupos objetivo, el patrón de errores o pixelados en el contenido de vídeo, se muestran los resultados subjetivos en la Figura 5-20. En la Tabla 5.11 se resumen los parámetros estadísticos más relevantes. En función de los resultados y desde el punto de vista gráfico, para el contenido del tipo "cartoons" este tipo de patrón de distorsión tiene una valoración subjetiva superior al resto de los contenidos. Se puede afirmar que los usuarios son más tolerantes a este tipo de distorsión en este tipo de contenido, mientras que son más exigentes cuando se visualizan contenidos del tipo noticias o deportes. Al igual que para el subgrupo de análisis de distorsión en asincronía entre audio y vídeo, no se encuentran diferencias según la localización de las distorsiones, al inicio o al final ($p\text{-value} > 0.1$). En este sentido, el resultado se contrario a lo postulado por otros autores, que afirman que la localización de distorsiones de un segundo en vídeos cortos tiene impacto en la calidad [158]. El estudio sin embargo es poco concluyente ya que los autores no informan del número de participantes en su estudio. Por otro lado, si que se hallan diferencias entre los resultados para el caso de distorsiones situadas al final de la secuencia (IM-PE) según el tipo de contenido, pero un test de *Tukey* solo respalda las diferencias significativas para el caso "cartoon-sports", Figura 5-21.

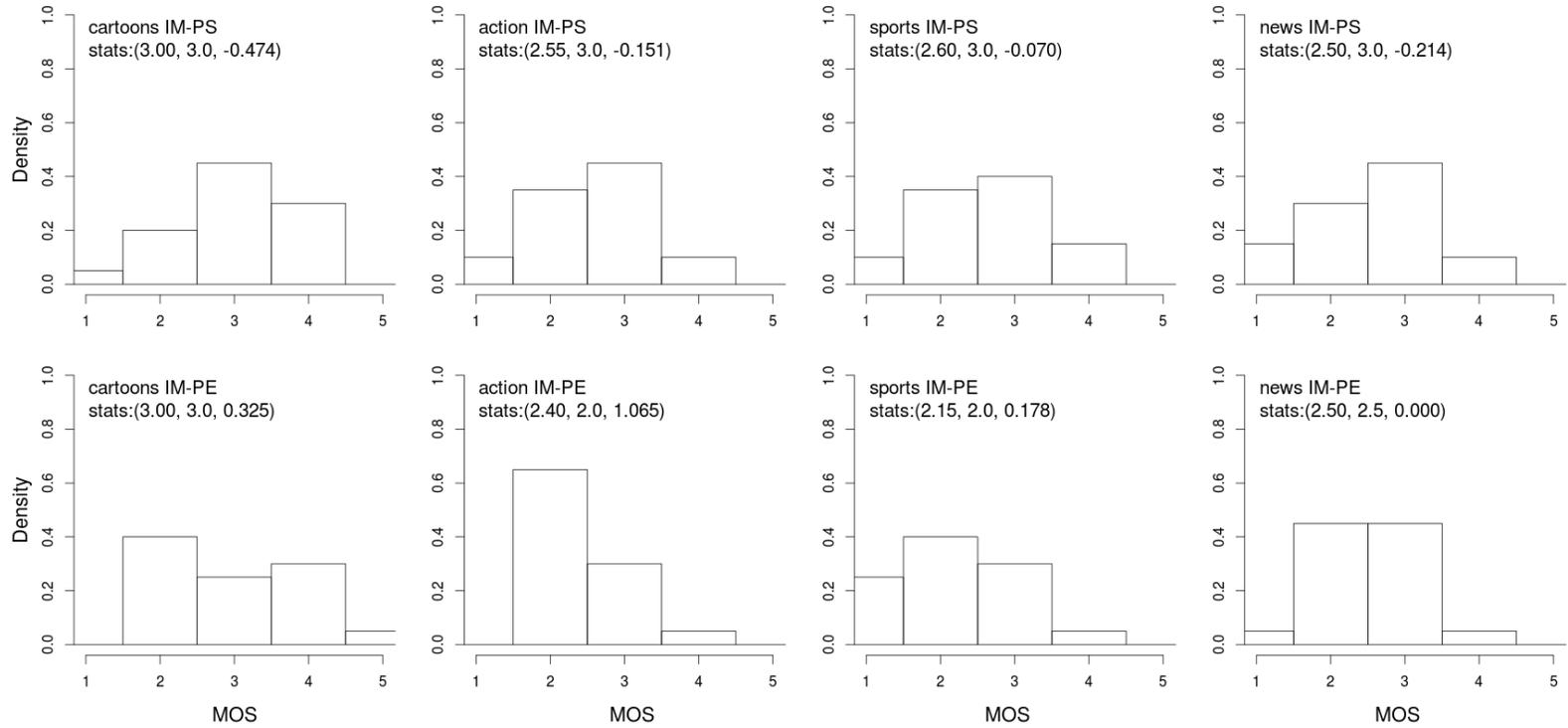
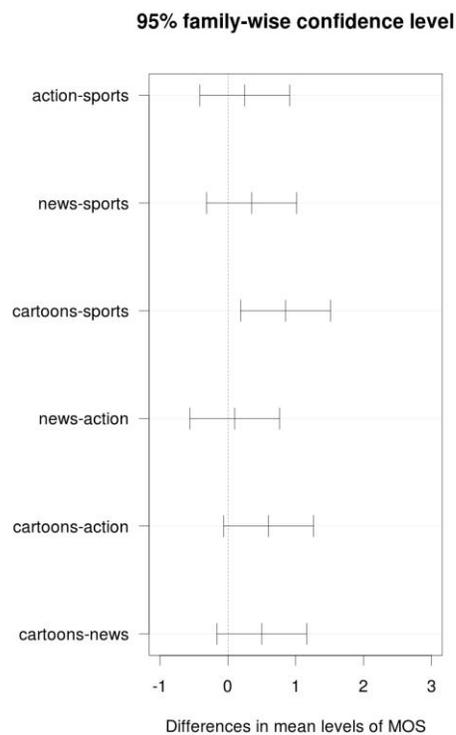


Figura 5-20: Resultados MOS para patrón de pixelación en vídeo

Tabla 5—11 Parámetros estadísticos para distorsión IM-P

		Valor promedio	Desviación estándar	Intervalo de confianza 1	Intervalo de confianza 2
Pixelación durante la primera mitad (IM-PS)	Cartoons	3.00	0.86	2.60	3.40
	Action	2.55	0.83	2.16	2.94
	Sports	2.60	0.88	2.19	3.01
	News	2.50	0.89	2.08	2.92
Pixelación durante la segunda mitad (IM-PE)	Cartoons	3.00	0.97	2.54	3.46
	Action	2.40	0.60	2.12	2.68
	Sports	2.15	0.87	1.74	2.56
	News	2.50	0.69	2.18	2.82

Figura 5-21: Test *Tukey* para grupos IM-PE

Para resumir los titulares extraídos de los resultados y enunciar las recomendaciones a realizar para los servicios de *streaming* en función del tipo de contenido, se incluye la figura 5-22, que visualmente recapitula los valores medios de MOS para cada tipo de contenido en función del modelo de distorsión (IM). Para todos los contenidos el factor al que los usuarios son más sensibles es IM-S, es decir, eventos de *rebuffering*. Para todos

los contenidos los usuarios se muestran más sensibles a los eventos de *rebuffering* cortos. Los contenidos de tipo "cartoon" no se ven perjudicados por la distorsión de asincronía entre audio y vídeo o por el efecto pixelación. Los contenidos de tipo "action" y "sports" son los más perjudicados subjetivamente por los efectos de cortes en la reproducción del audio y los eventos de *rebuffering* y además en el tipo de contenido "news" la distorsión en la sincronía entre audio y está subjetivamente muy mal valorada.

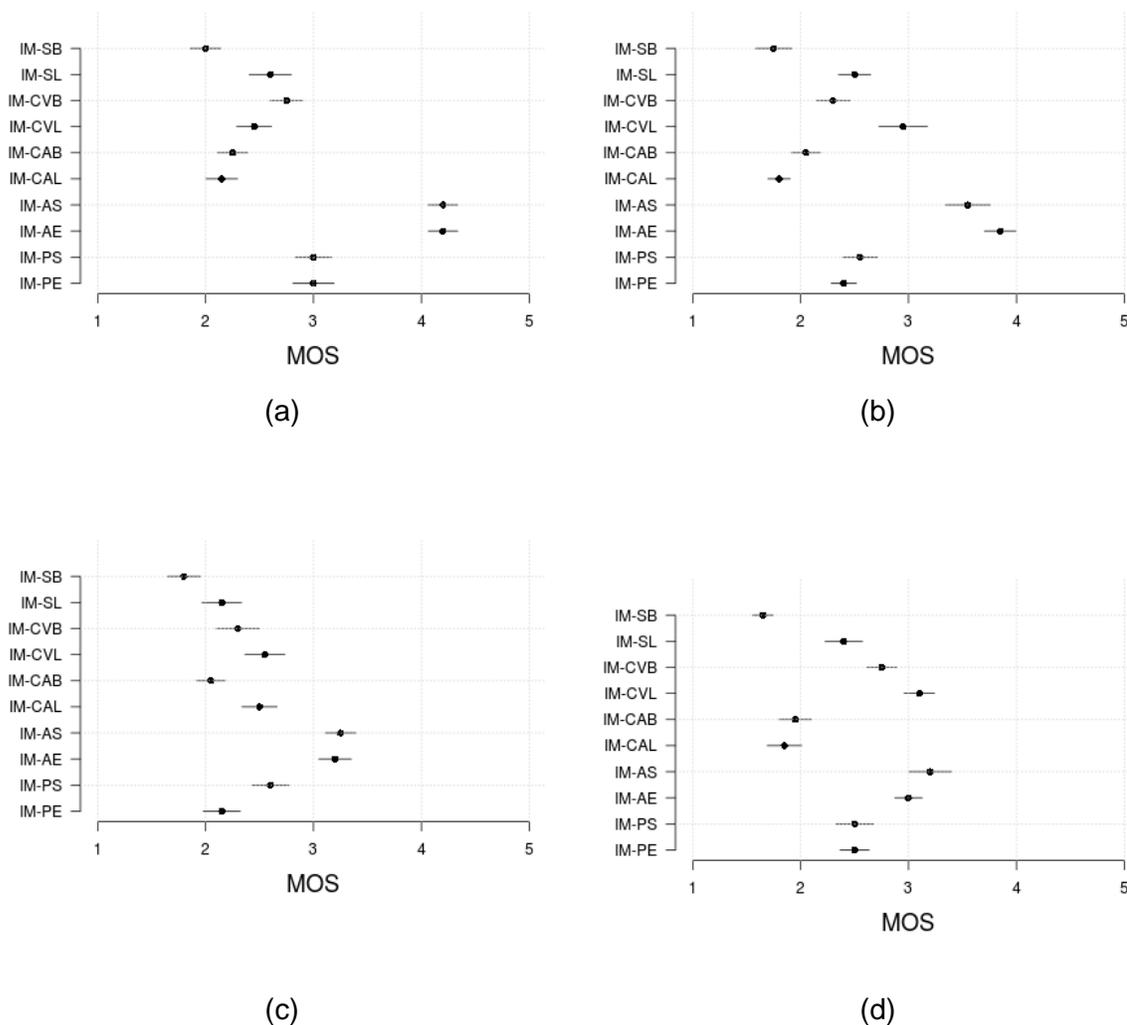


Figura 5-22: Resultados subjetivos para distintos tipos de contenido: (a) *cartoons*, (b) *action*, (c) *sports*, (d) *news*

En el transcurso del experimento, sólo 12 participantes recurrieron a la posibilidad de repetir la reproducción del contenido. La opción de repetición de la secuencia actual exclusivamente es obligatoria para no involucrarse en comparativas entre videos vistos y que se modifiquen los resultados por el efecto de aprendizaje o comparación. Las características de los vídeos repetidos son muy diferentes, si bien el tipo de distorsión IM-

S es la más repetida. Considerando por tipo de contenido el de tipo "*action*" fue el más repetido seguido del de tipo "*news*". Como apunte final, la edad media de los usuarios que usaron la opción de repetición de vídeo es de 48.25 años.

Profundizando en el análisis por edad, dividiendo a los participantes en dos grupos: por encima de 35 años y por debajo, los resultados muestran que no existen diferencias significativas entre las evaluaciones subjetivas globales para ambos grupos ($p\text{-value} > 0.1$). En este sentido y tomando otros factores socio-demográficos en este análisis, tampoco se han hallado diferencias significativas. Si se considera la experiencia como factor, los usuarios con experiencia exclusiva con contenido a través de TV adjudican valores subjetivos con una mayor media global (2.79), mientras que los usuarios que acreditan experiencia con contenido en alta definición y en Internet son menos tolerantes con las distorsiones, reportando un valor medio de 2.41 global. El factor de experiencia tiene además una relación con la edad. En la población de la prueba, los usuarios con experiencia en contenido en alta definición tienen una edad media menor (32 años) que los que solo tienen experiencia con TV (52 años). Además el 75% de los usuarios que admiten experiencia en contenido en alta definición pertenecen a un perfil técnico o científico. Estas diferencias en la edad y el perfil demográfico pueden estar relacionadas con las diferencias en los resultados subjetivos. Según hemos podido demostrar, los usuarios con experiencia en contenido de alta definición tienen unas expectativas más altas, lo que les hace menos tolerantes a los diferentes patrones de distorsión.

5.2 Experimentos subjetivos en sistemas adaptativos

En contraposición con la tecnología tradicional *best-effort*, la novedad introducida en sistemas adaptativos como SVC supone una alternativa muy prometedora. La posibilidad de adaptar los contenidos de forma dinámica a las condiciones de una red cada vez más cambiantes, mientras se maximiza la calidad percibida por los usuarios se ha vuelto un requisito imprescindible para asegurar el éxito del servicio. En efecto, la tecnología SVC ha sido determinante en la construcción de servicios modernizados cuyo rango de aplicación varía desde videoconferencia hasta el tradicional vídeo bajo demanda.

Estos nuevos servicios son capaces de variar las características del contenido multimedia no solamente realizando una clasificación de los usuarios, como realizaba en su día la tecnología Surestream¹⁹ de RealNetworks antes de la aparición de los codificadores escalables [159]. La adaptación de contenido basado en SVC es complementario a otras estrategias de adaptación basadas en QoS como las propuestas en [160].

¹⁹ <https://www.it.iastate.edu/services/real/information>

En la construcción de los algoritmos de toma de decisión a la hora de elegir el tipo de escalabilidad a implementar en cada caso aún existen muchas discrepancias [161]. Varios estudios se esfuerzan en analizar las percepciones subjetivas para apoyar la construcción de las estrategias de adaptación y, paralelamente, buscar métricas optimizadas de QoE que permitan realizar adaptaciones que maximicen la calidad percibida [162]. Sin embargo, la mayor parte de los estudios subjetivos son herencia de los estudios de rendimiento de los distintos codificadores, que han sido evolucionados para asumir la evaluación de la calidad de las distintas capas [84]. El trabajo en [83] constituyen un repaso al estado del arte de los estudios subjetivos y objetivos con relación a la tecnología de distribución de vídeo escalable en general, y en particular con SVC. A la luz de las conclusiones arrojadas en estos trabajos, existen aún muchas situaciones que requieren una caracterización en términos subjetivos

Ciertas cuestiones importantes, como si los sistemas adaptativos mejoran la calidad percibida, en qué situaciones y hasta qué grado, están aún abiertas a discusión. Los índices de tolerancia y umbrales de indiferencia con respecto a los sistemas adaptativos deben ser determinados para cada tipo de contenido, condiciones del sistema o características del usuario, por nombrar algunos factores. La respuesta a estas preguntas son los factores clave para la construcción de algoritmos de adaptación y constituyen una base imprescindible para construir y optimizar las redes de distribución de contenido para servicios realmente masivos [163]. El trabajo en [164] proporciona un repaso profundo de los factores que afectan la QoE. Desarrollan una aplicación para medir QoE en el popular servicio *Youtube* que utilizan para medir la QoE con diferentes proveedores en redes de acceso. Proponen un algoritmo (SARA) para mejorar la QoE aunque dedicado exclusivamente a servicios DASH.

Mientras otros trabajos se enfocan en obtener la evaluación de la calidad para diferentes capas o codificaciones de forma individual, en este punto de la investigación, nuestro trabajo busca comparar el rendimiento de los sistemas adaptativos y no adaptativos en términos subjetivos. El objetivo es dar respuesta a esos factores de tolerancia e indiferencia que guían en la construcción de mejores sistemas adaptativos.

La evaluación subjetiva comparativa entre un sistema tradicional y un sistema adaptativo permite esclarecer la preferencia de los usuarios hacia uno u otro sistema. Para la evaluación de los dos sistemas en paralelo se recurre a la metodología *Stimulus Comparison Adjectival Categorical Judgement* (SCACJ) [153], un tipo de *Comparison Category Rating* (CCR). Esta metodología permite evaluar la preferencia de los usuarios antes los sistemas en diferentes situaciones, desde alta disponibilidad a situaciones de alta congestión en la red. Como los sistemas reaccionan de forma diferente a las condiciones, la, potencialmente diferente, respuesta de los usuarios proporciona valiosa información para la construcción y mejora de los sistemas adaptativos.

Para evaluar las reacciones de los usuarios ante distintas condiciones de adaptación se construye una librería de pares de secuencias de vídeo. Las secuencias son fragmentos cortos que contienen la reacción de un sistema de transmisión de vídeo ante situaciones

variadas de la red, desde situaciones de congestión hasta situaciones de excedente de ancho de banda. Otro grupo de secuencias, correspondiente a la reacción de un sistema no adaptativo a las mismas condiciones de red, completa la librería.

Las secuencias son presentadas en parejas, según la metodología, a un grupo de 75 usuarios, con edades comprendidas entre los 23 y los 66 años. Los resultados son compilados en una base de datos que se ha publicado junto con los resultados de su estudio [165]. La población convocada para este estudio está de nuevo muy por encima de lo que es sugerido en las recomendaciones ITU-R BT.500-12 y ITU-R P.910. La heterogeneidad de la población es mayor de la encontrada normalmente en estudios de este tipo realizados en campus universitarios, que suelen centrarse en poblaciones jóvenes de perfil técnico, sesgando los resultados. Los resultados del experimento desarrollado aquí, con una población de edad y perfil heterogéneo, arrojan resultados más extrapolables a un sistema real.

En términos generales, los resultados muestran que los usuarios tienen una predilección clara por los sistemas adaptativos en la mayoría de situaciones. Sin embargo, se han identificado ciertas regiones de indiferencia o márgenes de tolerancia. En algunas situaciones los sistemas adaptativos no implican una mejora notable con respecto a los sistemas convencionales, lo que alimenta a los críticos con los sistemas adaptativos [55]. También se identifican algunas diferencias por edad o perfil del usuario que son discutidas en profundidad.

5.2.1 Retos de los sistemas adaptativos

Los servicios multimedia en Internet son altamente sensibles a las condiciones de congestión o pérdidas. Por esa razón los algoritmos de adaptación del contenido a las condiciones de transmisión gozan de tanta reputación [166]. Más particularmente, la adaptación del contenido en tiempo real es un aspecto clave de la tecnología SVC. Las dimensiones de escalabilidad proporcionadas por SVC incluyen temporal, espacial y de calidad, permitiendo la modificación de la secuencia adaptada en cualquiera de las dimensiones según las condiciones del momento. La combinación de la tecnología escalable y los algoritmos de estimación y decisión de adaptación, construyen los sistemas multimedia adaptativos.

Cuando se detecta una situación de congestión cualquier sistema adaptativo busca disminuir la tasa de transmisión asociada al contenido, el *bitrate* del flujo de vídeo. En una codificación por capas, esto supone que las capas superiores, consideradas capas de mejora, no son transmitidas. En el caso contrario, cuando el sistema detecta una situación de ancho de banda excedente, puede intentar ocuparlo transmitiendo más capas de mejora hacia el cliente. El algoritmo de decisión que establece cómo y cuando se cambia el número y tipo de capas a enviar es un aspecto clave en los sistemas de

este tipo. Cada vez se proponen más alternativas y todas buscan maximizar la calidad percibida, o QoE.

Precisamente, la evaluación de la calidad percibida ante distintas decisiones de adaptación, con respecto a un sistema no adaptativo, es la parte central de este capítulo. Los resultados contribuyen a comprender el efecto de las decisiones de adaptación en la experiencia del usuario y por tanto apoyar en la construcción de sistemas adaptativos basados en la maximización de la QoE. En el tipo de comparaciones analizadas, el sistema no adaptativo, mantiene siempre la misma capa durante la duración del experimento mientras que el sistema adaptativo reacciona a las condiciones de la red. Tras un periodo fijo de observación y estimación de la capacidad disponible, el sistema modifica el número de capas hacia arriba o abajo en alguna de las direcciones de adaptación ofrecidas por el codificador por capas.

Se definen diferentes objetivos que responden a factores de éxito críticos para los sistemas adaptativos. Cada objetivo analiza un cambio en una sola dimensión (temporal o de calidad). El contenido procesado para construir la librería de secuencias busca responder las preguntas clave realizadas para cada objetivo. Distintos contenidos originales han sido procesados según los patrones de adaptación definidos para obtener hasta 45 pares de secuencias. Una pareja de secuencias está formada por la secuencia de referencia, correspondiente a un sistema no adaptativo, y la secuencia adaptativa, emulando el resultado de un sistema adaptativo en el mismo escenario. Todas las secuencias son mostradas al total de la población del estudio. La descripción de los escenarios con los objetivos y las secuencias creadas se muestra en la Tabla 5-12.

Tabla 5—12 Distribución de objetivos y factores clave de éxito en el sistema adaptativo

Objetivo	Tipo de escenario	Tipo de adaptación	Número de secuencias	Factor crítico de éxito identificado
I	Escenario congestionado	Decrementar capas temporales o de calidad	6	Mejor dirección (temporal o de calidad) hacia la cual el algoritmo de adaptación debería decrementar capas cuando hay situaciones de congestión severas. Los pares de secuencias asignados a este incluyen cambios de una capa temporal o varias de calidad.
II	Escenario congestionado	Decrementar capas temporales	16	Límite de tolerancia ante pérdidas pequeñas que justifica un evento de adaptación Las secuencias para este objetivo incluyen secuencias con pérdidas entre el 1 y el 10%. La versión adaptativa reduce una capa temporal para compensar las pérdidas. La dimensión de calidad se mantiene alta para evitar su influencia en los resultados
III	Escenario congestionado	Decrementar capas temporales	6	Preferencia entre secuencias severamente afectadas por pérdidas debidas a problemas de red o secuencias adaptadas en la dimensión temporal. Las secuencias preparadas para este objetivo incluyen pérdidas severas y adaptaciones de varias capas temporales de una vez.
IV	Escenario no congestionado	Incrementar capas de calidad	8	Impacto y nivel de sensibilidad ante el incremento de capas de calidad. Justificación del coste adicional por ancho de banda consumido frente a mantener el nivel de calidad inicial. Las secuencias para este objetivo incluyen cambios de capas de calidad ascendente entre 1 y 3 capas, mientras que las capas temporales (altas) son constantes
V	Escenario no congestionado	Incrementar capas temporales	12	Impacto en el usuario del cambio de tasa reproducción (capas temporales) frente a mantener una tasa constante inicial. Las secuencias seleccionadas incluyen adaptaciones de 1 a 4 capas temporales manteniendo constante y a un nivel alto la capa de calidad

Más detalladamente, los objetivos responden a la necesidad de clasificar las secuencias para dar respuesta a una o varias cuestiones de las decisiones de adaptación.

Objetivo I: El primer objetivo es determinar la mejor dirección, es decir, escoger la dimensión de adaptación más adecuada, desde el punto de vista de QoE, que un algoritmo adaptativo debe tomar a la hora de adaptarse a situaciones de congestión severa en la red. Las dimensiones consideradas comprenden la dimensión temporal y de calidad. Cuando los cambios en la secuencia son de un salto pequeño, ha sido documentado que es preferible decrementar la tasa de la secuencia eliminando capas en la dimensión de calidad [83]. Se asume que cambios pequeños en la calidad son menos apreciables que reducir la tasa de reproducción (*frame rate*) por la mitad (el mínimo salto temporal admisible), aunque ambos cambios resulten en un *bitrate* similar. Sin embargo, en presencia de una situación de congestión severa, el impacto del salto de varias capas de calidad necesario para compensar las pérdidas debe ser analizado de nuevo.

Objetivo II: Algunos algoritmos de adaptación utilizan la tasa de pérdidas como un criterio disparador del evento adaptación. El segundo objetivo es revisar el umbral de pérdidas que, desde un punto de vista subjetivo, justifica un cambio en la dimensión temporal. Este umbral define la tolerancia de los usuarios a las pérdidas frente a una secuencia adaptada temporalmente, y por lo tanto sin pérdidas.

Objetivo III: La adaptación en la dimensión temporal se percibe normalmente como una solución pobre en términos de QoE. Este objetivo busca dar respuesta a la preferencia de los usuarios ante una secuencia con pérdidas severas y una secuencia que ha sido adaptada en la dimensión temporal.

Objetivo IV: Los servicios adaptativos ofrecen la posibilidad de incrementar la tasa de transmisión, incrementando el número de capas transmitidas, cuando se detectan condiciones de ancho de banda favorables. Con este objetivo se busca dar respuesta, desde el punto de vista de QoE, a si el incremento de capas en la dimensión de calidad es justificada, o los usuarios prefieren una secuencia de calidad constante desde el inicio. También responde a si los incrementos de calidad son advertidos, o los usuarios muestran indiferencia por uno u otro sistemas, con lo cual el incremento del coste asociado a transmitir una secuencia de mayor calidad no estaría justificado.

Objetivo V: Tal como ocurre en la dimensión de calidad, los sistemas adaptativos también pueden incrementar el número de capas en la dimensión temporal ante situaciones favorables. Este objetivo busca evaluar la calidad percibida por el usuario al comparar secuencias con una tasa constante con secuencias que aumentan su *frame-rate* ante situaciones favorables.

Varios modelos de sistemas adaptativos han sido propuestos en la literatura [167]–[169]. Aunque todos ellos son diferentes en muchos aspectos, todos ellos reaccionan de la misma forma ante variaciones en las condiciones de la red, modificando el número de capas cuando hay una variación apreciable en el ancho de banda estimado en la red. El

sistema desarrollado en [131], realiza las estimaciones del ancho disponible en la red del cliente cada 5 segundos. El servidor actualiza entonces la cantidad de niveles temporales o de calidad transmitidos a cada cliente en respuesta a esta estimación. Aunque los autores incluyen experimentos con distintos intervalos para la estimación y la adaptación, este intervalo fue considerado adecuado. Es lo suficientemente lento para evitar que el sistema adaptativo reaccione ante situaciones circunstanciales de congestión o ancho de banda excedente y suficientemente rápido como para seguir los cambios de capacidad de la red de más largo recorrido. Este modelo sirve de modelo de sistema adaptativo para construir los patrones de adaptación de las secuencias procesadas para este capítulo.

5.2.2 Metodología de los experimentos

En este caso, buscando la comparación entre los sistemas de transmisión de vídeo tradicionales con los sistemas adaptativos, los usuarios son interrogados acerca de la preferencia hacia uno u otro sistema en cada caso de los presentados bajo diferentes circunstancias. La metodología SCACJ descrita en [153] responde perfectamente a las necesidades particulares de este experimento. La metodología SCACJ establece que los usuarios observadores evalúan la relación entre dos secuencias presentadas de forma simultánea en términos semánticos, es decir, en una escala categórica. En este caso, las secuencias corresponden a una secuencia adaptativa y otra secuencia no adaptativa en unas condiciones específicas. Según SCACJ, el observador proporciona la existencia (igual o diferente) la dirección y el grado (menos, más) de diferencias perceptibles entre los sistemas comparados.

Este método establece 7 categorías, puntuadas numéricamente como el intervalo [-3,3], para indicar cuál de los dos estímulos presentados es preferido y en qué grado. En este caso, la escala presentada en los experimentos realizados difiere ligeramente de la sugerida en la metodología. Las categorías propuestas por ITU-T resultaron complicadas de entender en los primeros entrenamientos del experimento, ya que forzaban a los participantes a interpretar cuidadosamente la pregunta de "Cómo han visto la secuencia A con respecto a la secuencia B" cuando habían perdido la referencia visual de las secuencias. En este sentido, hemos propuesto una escala simétrica en el supuesto que resulte válida la propiedad transitiva en la evaluación de las dos secuencias. Decir que "A es mucho peor que B" es equivalente a decir que "B es mucho mejor que A". Con la nueva escala de categorías propuesta, los participantes en los experimentos de calibración consideraron la escala de votación mucho más intuitiva al votar hacia el lado en el que habían visto la mejor secuencia. El cambio en las categorías de la escala no supone sin embargo ninguna modificación de la asignación numérica subyacente. Las categorías de la nueva escala también tienen la capacidad de indicar la existencia, dirección y grado de la diferencia observada por los usuarios. En la Figura 5-23 se muestra la nueva relación de categorías y su asignación numérica para una representación esquemática del experimento

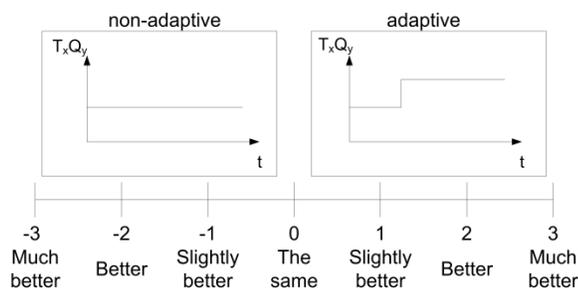


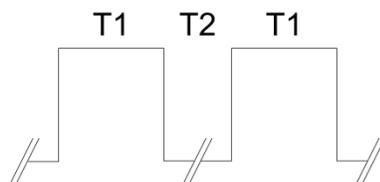
Figura 5-23: Esquema de evaluación SCACJ con categorías modificadas

La escala de comparación puede ser directamente traducida a la escala *Comparative Mean Opinion Score* (CMOS) [170]. CMOS es una escala muy sensible tradicionalmente usada en evaluaciones de codificadores de audio. En los experimentos subjetivos realizados durante la elaboración de este trabajo, la preferencia de los usuarios hacia el sistema adaptativo está identificada por valores positivos de la escala CMOS, mientras que los valores negativos indican una preferencia hacia el sistema no adaptativo o convencional.

Las metodologías para la realización de experimentos subjetivos indican los pasos necesarios para llevar a cabo los experimentos. De acuerdo a las recomendaciones de SCACJ, los participantes reciben las instrucciones del experimento por escrito. Los usuarios son instruidos para concentrarse en la evaluación global de la calidad percibida en cada par de secuencias presentadas y se les ruega que voten hacia el lado que les haya gustado más. Las sesiones de entrenamiento comienzan con una descripción del tipo de evaluación, el sistema de votación y la descripción de la escala subjetiva. Se utiliza una secuencia de entrenamiento diferente a las incluidas en el test y se avisa a los participantes que, aunque la secuencia de entrenamiento representa el tipo de distorsiones que van a visualizar en las secuencias siguientes, ésta no se corresponde con los límites de la escala de medición.

Tras la sesión de entrenamiento, los usuarios rellenan un cuestionario que incluye la información de contacto y algunos detalles personales. La información personal solicitada incluye edad, sexo, educación, experiencia con sistemas multimedia (solo TV, video en Internet, o vídeo HD en Internet), dispositivos usados para visualizar el contenido multimedia (TV, PC, Tablet, Smartphone).

El núcleo del experimento incluye 45 pares de secuencias, cada uno de 19 segundos de duración, alternados con periodos de votación de duración ilimitada, controlada por el usuario. En la figura 5-24 se describe gráficamente la secuencia del experimento subjetivo



T1 = stimulus evaluation 19 s
T2 = voting time

Figura 5-24: Diagrama del procedimiento del experimento subjetivo

Para evitar que los efectos de la fatiga y el aprendizaje influyan los resultados del experimento, el orden de los pares de secuencias en cada experimento es alterado de forma aleatoria. También se altera de forma aleatoria la posición de la secuencia de referencia (no adaptativa) en la pantalla, pudiendo aparecer a derecha o izquierda indistintamente. Tras finalizar la sesión de evaluación subjetiva, los usuarios son interrogados acerca de la dificultad general de experimento, la dificultad con respecto a la escala de votación, su opinión acerca de la duración del experimento así como comentarios adicionales sobre sus impedimentos visuales, si existiesen. Esta encuesta final consta de 4 escalas de Likert, cuyos resultados se resumen en la tabla 5-13. En general los usuarios perciben una dificultad baja aunque consideran que la duración es relativamente alta. El experimento se desarrolla en una sola sesión de duración media de 21.14 ± 4.97 minutos. Para agradecer su colaboración, los usuarios fueron obsequiados con un pequeño regalo.

Tabla 5—13 Resultados de la encuesta final del experimento subjetivo

Escala Likert (1-5)	Valor promedio (desviación estándar)
Dificultad general (1 muy fácil – 5 muy complejo)	1.96 (0.98)
Duración del test (1 muy corto – 5 muy largo)	2.81 (0.83)
Complejidad de la aplicación (1 muy fácil– 5 muy compleja)	1.29 (0.69)
Sistema de votación (1 muy sencillo – 5 muy complejo)	1.37 (0.63)

5.2.3 Entorno y equipamiento de el experimento subjetivo

El *software* utilizado para la realización de los experimentos se desarrolla de nuevo de forma específica para cumplir las especificaciones del experimento. La apariencia de la aplicación se ha modificado de acuerdo a las recomendaciones de la metodología subjetiva. En la figura 5-25 se muestran las capturas de pantalla de la aplicación, en sus ventanas de evaluación y votación.

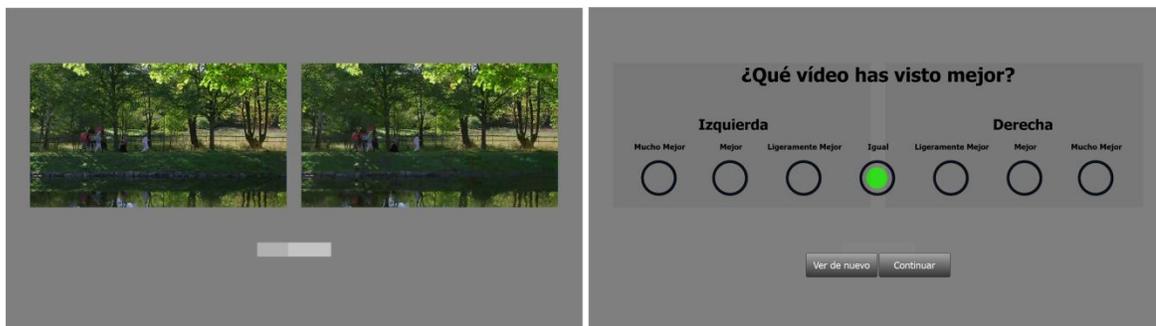


Figura 5-25: Interfaz del software desarrollada para el experimento subjetivo

Para realizar el experimento, el *hardware* requerido para presentar dos secuencias en crudo (sin compresión) en el monitor al mismo tiempo incluye un ordenador de alto rendimiento con un monitor de 23 pulgadas. Las características del equipo utilizado, el mismo que en los experimentos anteriores ya fueron detallados en las Tablas 5-2 y 5-3. El experimento fue realizado en una oficina privada, con una iluminación comprendida en el rango de 70 a 100lux, dentro del rango recomendado en [171]. Los usuarios fueron colocados a una distancia entre 1.5 y 2 metros del monitor.

5.2.4 Contenidos del experimento subjetivo

El contenido del experimento incluye varias secuencias originales de distinta temática. Para mantener la duración del experimento en los límites recomendados por la ITU BT-500, se han seleccionado cuatro secuencias originales: *factory*, *tractor*, *marathon* y *touchdown*. En las sesiones de entrenamiento se utiliza una secuencia distinta. En la Tabla 5-14 se describen las características del contenido por medio de los índices de complejidad temporal y espacial (SI, TI). En el caso de la secuencia *marathon* deben indicarse dos valores de complejidad temporal, con y sin el cambio de escena contenido en la secuencia. Las secuencias han sido seleccionadas para representar el mayor rango posible del espacio bidimensional de complejidad. Las secuencias *touchdown* y *tractor* tienen índices de complejidad similares, ambos bajos. La secuencia *factory* tiene el índice de complejidad temporal más alto y la secuencia *marathon* tiene el mayor índice de complejidad espacial entre la secuencias del experimento. En la Figura 5-26 se muestra un *frame* de cada secuencia.

Tabla 5—14 Características de las secuencias originales en el experimento subjetivo

Contenido	Índice SI	Índice TI	Detalles [resolución@frameRate, duración]
Factory	21.2556	55.3439	854x480@30 fps, 570 frames, 19"
Tractor	17.8137	23.9409	854x480@25 fps, 475 frames, 19"
Marathon	23.9810	68.16 (31.8 sin cambios de escena)	854x480@50 fps, 950 frames, 19"
Touchdown	18.0624	27.0181	854x480@30 fps, 570 frames, 19"



Figura 5-26: Fotograma de las secuencias del experimento subjetivo

El contenido codificado con el codificador de referencia, JSVM, se genera con 5 capas temporales y 4 capas de calidad usando el esquema CGS. A continuación se utiliza el *software* SVEF [62] para generar las secuencias procesadas. En el proceso de la investigación, se trabaja en extender la forma de trabajar del *software* SVEF para incorporar un modelo de pérdidas al sistema de emulación, algo de lo que carece, inicialmente, el *software* SVEF. El modelo de pérdidas incluido responde a un proceso de Bernoulli simple con la probabilidad de pérdidas como parámetro de entrada al modelo [172]. Considerando que lo que se pretende modelar es el momento exacto de la adaptación, justo cuando nos encontramos en el punto medio de la secuencia, tras 5 segundos de análisis, se asume que la secuencia en cuestión pasa por un estado de pérdidas (de los dos estados en un modelo de Gilbert). En esta situación, un modelo de Bernoulli simple es suficiente para modelar las pérdidas en el estado degradado,

suponiendo que los evento de pérdida de paquete dentro de este estado no estén correlados.

La traza resultante del modelo de Bernoulli aplicado a cada caso se utiliza para procesar la secuencia en el subsistema de emulación de la herramienta SVEF. Dentro de la misma herramienta, al recomponer la secuencia y, en particular, las NAL, si una NAL está afectada por la pérdida de uno o más paquetes, ésta se descarta por completo. Finalmente se decodifica la secuencia resultante a través de las herramientas del decodificador de referencia, que, a pesar de no permitir trabajar con secuencias reales afectadas por pérdidas, proporciona herramientas para reconstruir secuencias alteradas a través de trazas y la secuencia original. Todas las secuencias procesadas consideran un único tipo de escalabilidad en cada caso, ya sea temporal o de calidad, pero nunca ambos juntos.

Para componer los pares de secuencias que serán presentados a los usuarios se emparejan secuencias resultantes de un comportamiento no-adaptativo y las secuencias resultado del sistema adaptativo en idénticas condiciones. En el caso de la secuencia procesada para emular un sistema adaptativo, las adaptación tiene lugar tras los 5 segundos iniciales, por lo que las dos secuencias del par son iguales durante los 5 primeros segundos. Los 45 pares de secuencias se agrupan según el criterio de los objetivos definidos en la metodología del experimento, aunque alguna secuencias sirven a varios objetivos. El reparto de secuencias en los objetivos ya fue visto en la Tabla 5-12.

Para todos los objetivos, las secuencias fueron seleccionadas para cubrir el rango disponible del espacio bidimensional de complejidad (temporal y espacial). En total se han usado 14 veces la secuencia *tractor*, 13 veces la secuencia *factory*, 9 veces la secuencia *marathon* y 8 veces la secuencia *touchdown*.

Para evitar la fatiga de los usuarios y por tanto contener la influencia de ésta en los resultados del experimento, se reparten las secuencias de forma que los objetivos más simples son evaluados con menos secuencias y se reserva mayor número de secuencias para los objetivos más amplios.

5.2.5 Población

Un total de 75 candidatos fueron invitados a realizar los test subjetivos. Todos los usuarios demostraron capacidad visual, normal o corregida, suficiente para realizar los experimentos. Además se realizó un test de percepción del color, a través de los diagramas de Ishihara. Dos de los usuarios resultaron tener deficiencias en la percepción del color, aunque sus resultados no han sido excluidos del experimento. El porcentaje hombres en la muestra es del 76% frente al 24% de mujeres. La edad de los participantes está comprendida entre los 22 y los 66 años, con una media de 37.01. El 49% de los participantes es menor de 35 años.

Ante el cuestionario realizado sobre sus preferencias y experiencia con servicios multimedia, el 17.3% declara que sólo acostumbra a ver contenido multimedia a través del televisor. El 48% son usuarios habituales de contenido multimedia en Internet y el 34% que lo son de contenido en Internet y en alta definición. Entre los que ven contenido multimedia en Internet, el 55.8% lo hace en el PC, seguido de un 29.8% que lo visualiza en un Smartphone y un 14.4% suele visualizarlo en una Tablet.

5.2.6 Análisis de los resultados

Los datos obtenidos de la realización de las evaluaciones subjetivas, son procesadas para analizar cada objetivo planteado. La consistencia de las votaciones de cada usuario es corroborada. Un votación consistente de los usuarios implica que éstos mantendrán un mismo criterio ante situaciones similares. En particular, para evaluar la consistencia de los usuarios en este experimento se comparan las puntuaciones asignadas a dos pares de secuencias con condiciones equivalentes. Tan solo un 2.7% de la población muestra poco criterio en sus evaluaciones si se considera que dos secuencias que aumentan dos niveles de calidad, ya sea en la dimensión temporal o en la de calidad, deben recibir mejor puntuación que la secuencia que se mantiene constante.

Los resultados obtenidos en forma del indicador CMOS son procesados y agrupados por objetivos, Figura 5-27. En la figura se incluyen los valores medios y los intervalos de confianza al 95%. Todos los resultados fueron evaluados estadísticamente para determinar la relevancia de los mismos frente a la hipótesis nula, mediante un test Student-t para una muestra simple. Solo 4 pares no pueden ser considerados estadísticamente diferentes a la hipotética media 0 ($p\text{-value} > 0.01$). Estos pares se resaltan en la Figura 5-27. Los resultados positivos indican una preferencia hacia el sistema adaptativo, mientras que los negativos indican preferencia por el sistema no-adaptativo.

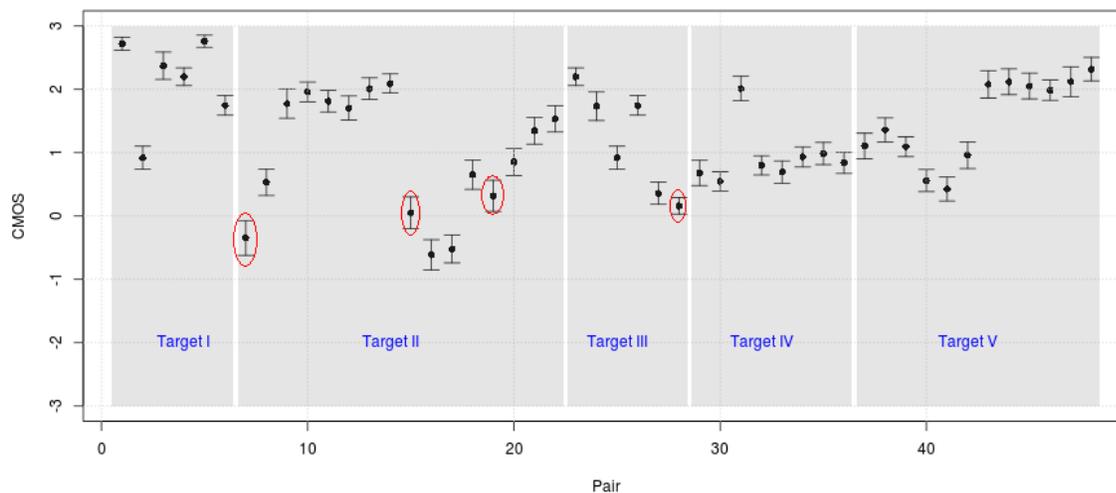


Figura 5-27: Resultados CMOS agrupados por objetivos

Existe una tendencia de repetición de experimentos con la edad. La gente más longeva repite las secuencias con más frecuencia que la población joven del experimento. Además, las evaluaciones de la población mayor de 35 años son más contenidas, mientras que los valores de la población menor son más extremos, tanto en valor positivo como negativo, un comportamiento observable en los valores medios de la Figura 5-28

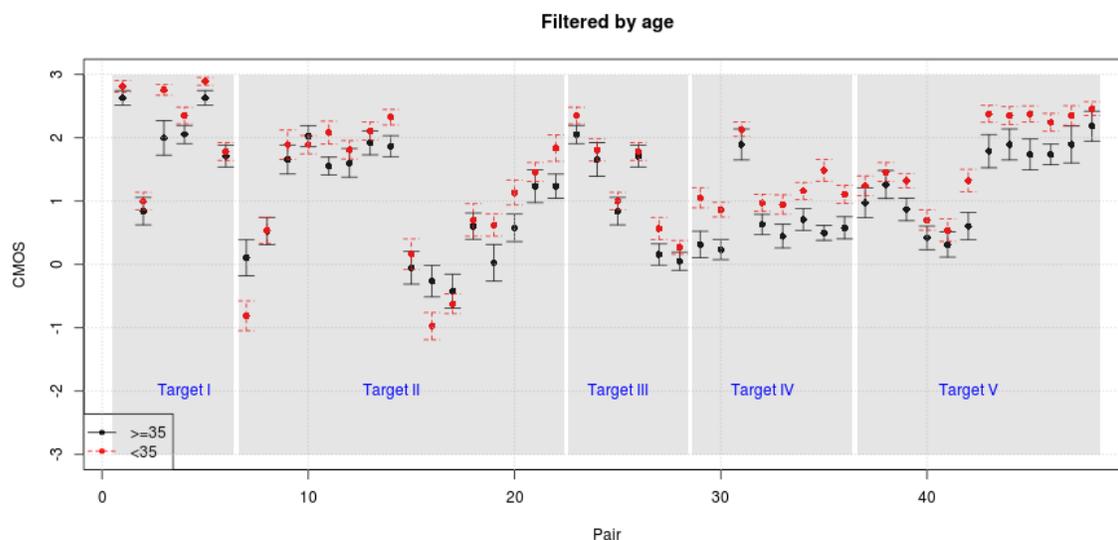


Figura 5-28: Resultados CMOS agrupados por objetivos y edad

Seguidamente, se detallan los resultados particularizando por cada uno de los objetivos definidos en el planteamiento del experimento. Para analizar cada objetivo se obtienen los histogramas normalizados de las valoraciones subjetivas proporcionadas por los usuarios. En las figuras se incluye una leyenda con la información de las secuencias evaluadas en el par en particular. Entre los datos de la leyenda se encuentra el tipo de contenido (nombre de la secuencia), las características de ambas secuencias del par —la resultante de un sistema adaptativo (secuencia *adapt*) y la secuencia correspondiente a un sistema no adaptativo (secuencia *ref*)— e información estadística de la distribución. Para ambas, la información tiene la siguiente sintaxis: (*frame-rate*, *bitrate*, *PLR*). En el caso de la secuencia adaptada, estos valores se declaran dos veces, antes y después del evento adaptación. La información estadística presentada en la leyenda sigue el siguiente formato (*mean*, *median*, *swkeness*) que indican el valor medio, la mediana y la medida de asimetría de la distribución o histograma.

5.2.6.1 Objetivo I

Cuando las pérdidas son altas, en torno al 40%, los resultados subjetivos respaldan que el sistema adaptativo es preferible. Los valores medios de los histogramas de resultados demuestran la inclinación subjetiva hacia el sistema adaptativo frente a la referencia no adaptada, figura 5-29. Evaluando la elección entre la adaptación de varias capas temporales o una capa de calidad, cuando ambas suponen un salto de tasa de transmisión similar, se puede inferir de los resultados subjetivos que los usuarios aceptan mejor cambios en calidad (los valores medios son más altos). Además, los factores de asimetría son siempre negativos, evidencia cuantitativamente que la masa representativa de los votos subjetivos se concentra en el lado positivo del histograma. En particular, los valores del factor de asimetría para los pares con adaptación en la dimensión temporal toma valores cercanos a -0.5, que puede ser asumido como distribuciones simétricas. En este caso, los votos se distribuyen equitativamente alrededor de la media.

Los valores de los t-test de dos muestras para evaluar las diferencias estadísticamente significativas entre la media de las muestras para cada pareja (adaptación en la dimensión temporal con relación a la adaptación en la dimensión de calidad) muestran que no puede garantizarse dicha diferencia solamente para el caso de la secuencia *factory* ($p\text{-value} > 0.01$). La influencia del contenido en la decisión de la dimensión de adaptación con pérdidas altas debe ser considerada.

En virtud de los resultados, es indiscutible que los usuarios valoran más positivamente una degradación en la dimensión de calidad que en la dimensión temporal. Esta observación es corroborada para secuencias con menor complejidad temporal, como tractor, o para secuencias con mayores índices de complejidad temporal o espacial. En este grupo de secuencias el par con menor complejidad temporal y una adaptación de 3 niveles temporales fue el par más repetido por los usuarios (hasta en 8 ocasiones).

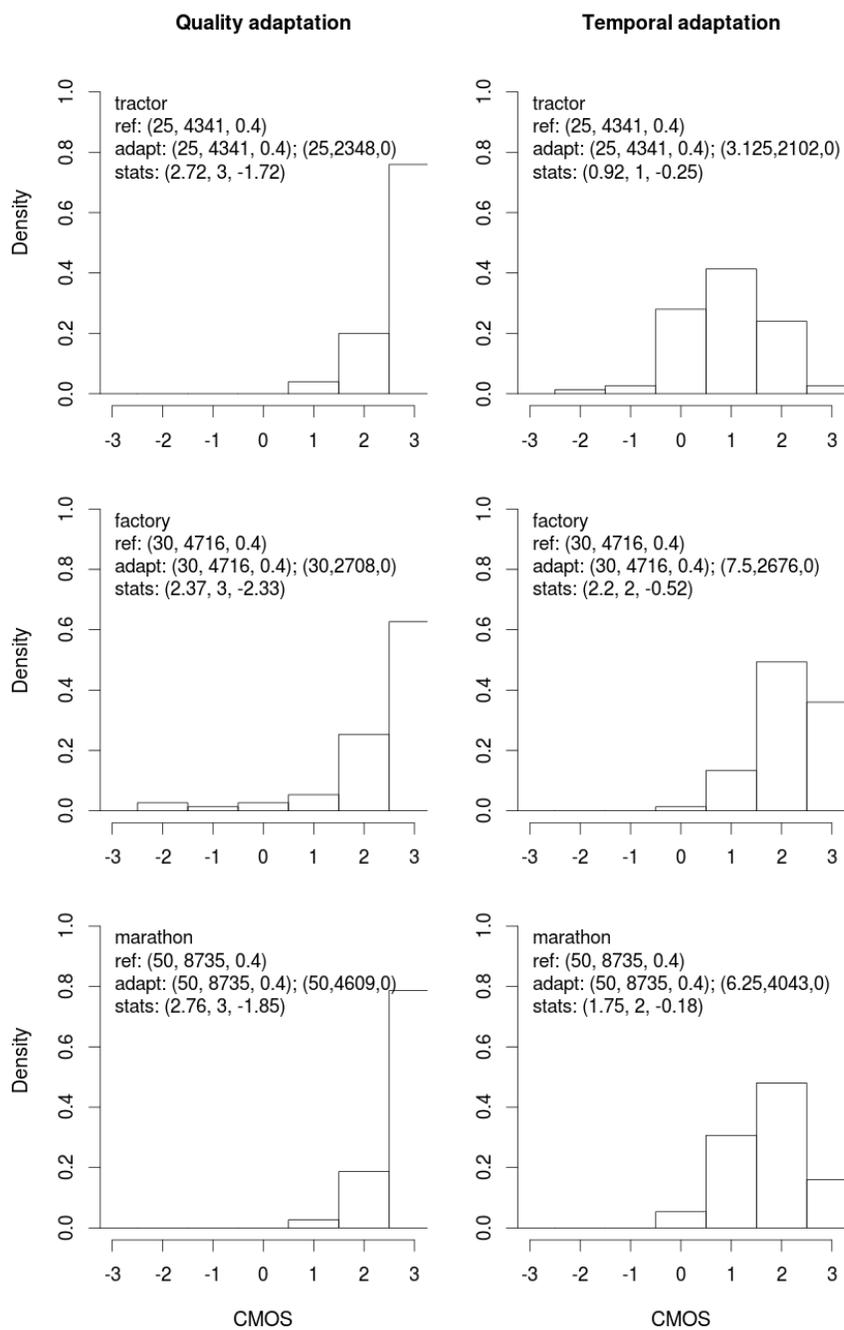


Figura 5-29: Histogramas CMOS para objetivo I

5.2.6.2 Objetivo II

Cambiando de objetivo de análisis, se valoran los resultados en el caso de pérdidas de menor magnitud. Observando la distribución de los votos en histogramas para la primera columna del figura 5-30, los resultados de las votaciones indican que los usuarios difícilmente eligen el sistema adaptativo en presencia de pérdidas alrededor del 1%. De acuerdo a los resultados para la secuencia *marathon* el sistema adaptativo es mejor mientras que, de acuerdo a los resultados para la secuencia *factory* el sistema no-adaptativo es mejor. Ambos contenidos tienen índices de complejidad similares (altos) por lo que se confirma que para el mismo tipo de contenido la elección entre los sistemas no está clara. Los t-test realizados con la hipótesis nula revelan que las otras dos secuencias *tractor* y *touchdown* no son estadísticamente distintos de cero. Es decir, ambos sistemas son valorados igual en este aspecto.

Para un 3% de pérdidas la tendencia ya se inclina más claramente hacia el sistema adaptativo. Aún así, en función del tipo de contenido o sus características se tienen algunas particularidades. Para la secuencia *tractor* en la segunda columna de la figura 5.8, con un nivel temporal bajo de inicio, los usuarios eligen antes el sistema no adaptativo (valor medio -1). De forma similar, para la secuencia *factory* los resultados indican que existen una preferencia débil hacia el sistema adaptativo, estadísticamente significativa, pero con un valor CMOS bajo. Se puede inferir que el umbral de 3% de pérdidas está condicionado por las características del contenido, más particularmente por el *frame-rate* inicial de las secuencias. Para secuencias con *frame-rate* bajo, reducir aún más la calidad en la dimensión temporal, no es preferible a convivir con las pérdidas en la secuencia original.

Para el 5% y el 10% de pérdidas siempre se obtienen resultados positivos de CMOS (preferencia por sistema adaptativo), aunque el *frame-rate* de la secuencia de partida influye en la distribución de los resultados. Para *frame-rate* bajos, los valores están más dispersos en torno a la media positiva (se dan desviaciones estándar mínimas de 1.04 y máximas de 1.27). En cambio, para secuencias con *frame-rate* altos los valores se apilan en el extremo positivo de la escala CMOS (desviaciones mínima de 0.78 y máxima de 0.97). Los factores de asimetría de las distribuciones para tasas de *frame-rate* bajos indican que los histogramas son mayormente simétricos, no así para los *frame-rate* altos.

Comúnmente en la práctica, los sistemas adaptativos utilizan los umbrales de pérdidas para condicionar las decisiones de adaptación. Basándose en los resultados de nuestros experimentos subjetivos se puede decir que los umbrales de tolerancia a las pérdidas deben estar en torno al 3% de forma genérica. Pero en función del *frame-rate* inicial este umbral puede ajustarse más adecuadamente.

Del grupo de secuencias que se asocian a este objetivo, la más repetida por los usuarios fueron los pares *factory* y *touchdown* con pérdidas de 1% y 5% respectivamente. Ambos pares fueron repetidos 18 veces. Al final los resultados para ambos pares de secuencias son de -0.61 y 0.32 respectivamente.

Un test Kurskal-Wallis revela que las medias en el grupo de las cuatro secuencias del 3% son diferentes. Un test posterior de Tukey HSD (*honest significant difference*) muestra que excepto para la combinación *marathon-touchdown* todas las medias son significativamente distintas, Figura 5-31.

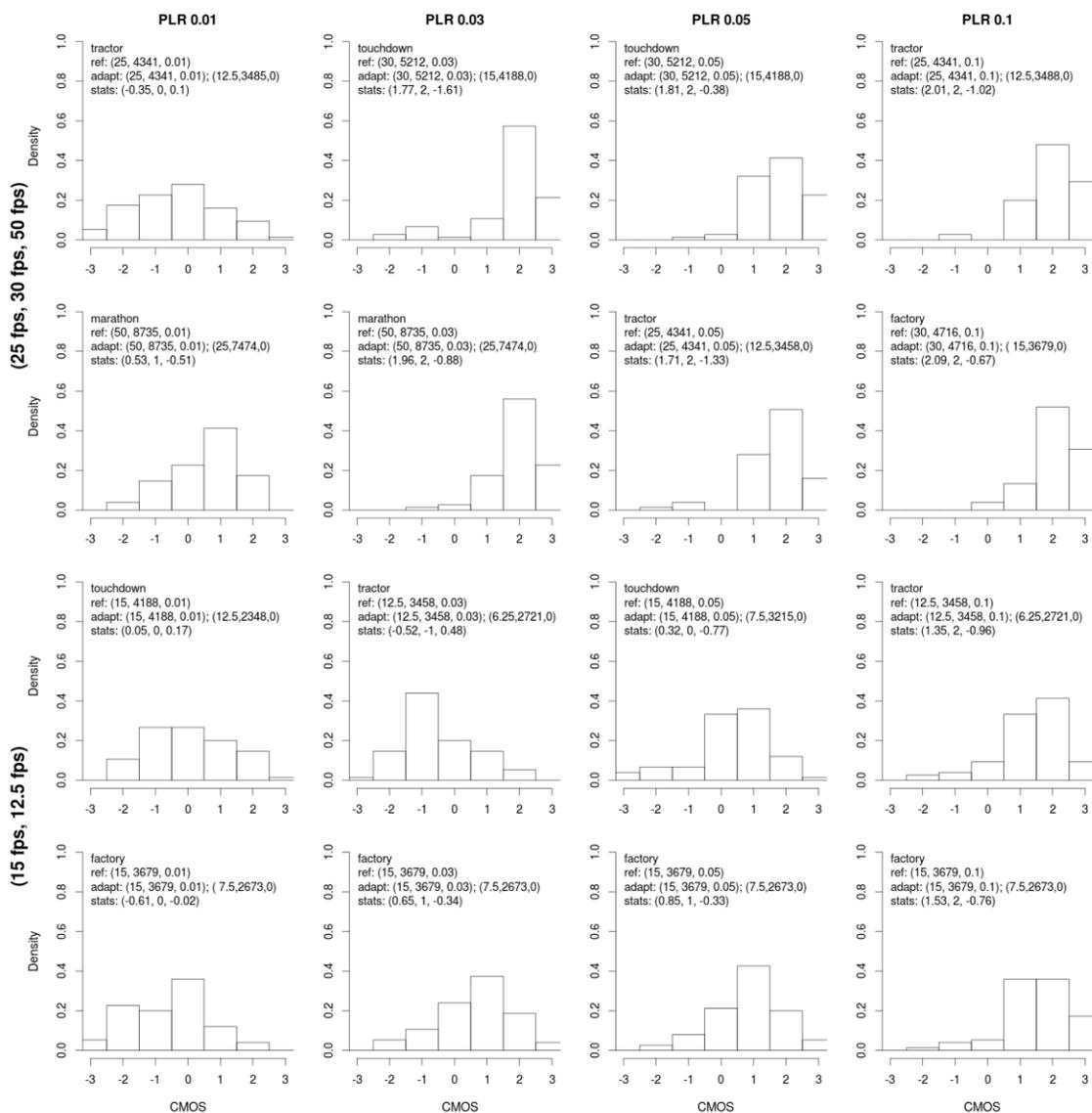


Figura 5-30: Histogramas CMOS para objetivo II

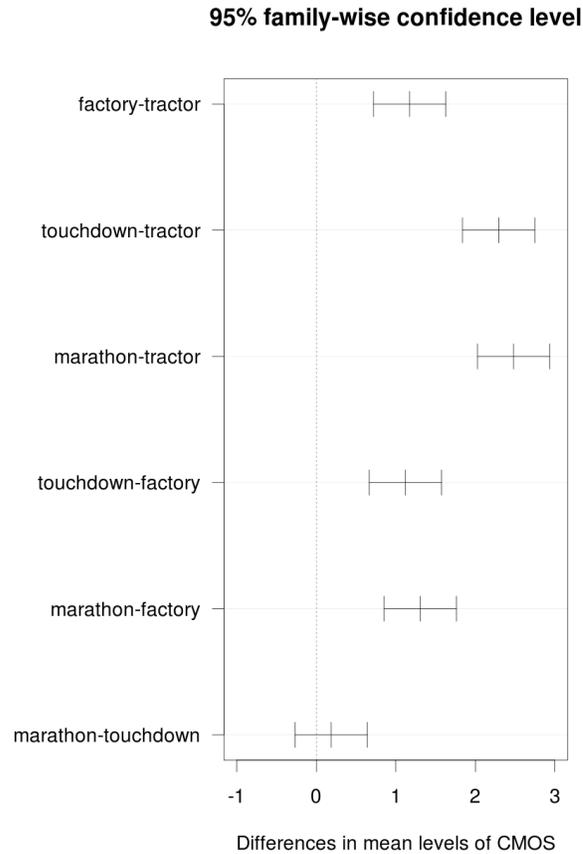


Figura 5-31: Test de *Tukey* HSD para diferencias CMOS con PLR 3%

5.2.6.3 Objetivo III

Los resultados subjetivos para este grupo se agrupan en la figura 5-32. En escenarios congestionados, compara reducciones de calidad en la dimensión temporal. De acuerdo a los dos primeros histogramas, ante la reducción de dos niveles temporales, primera fila de la figura 5-32 (motivados por pérdidas de 40% y 30% respectivamente) los sistemas adaptativos supone una mejor alternativa. Los factores de asimetría indican que la secuencia tractor concentra la mayor parte de las votaciones en el extremo positivo.

Para los casos de una reducción de tres niveles temporales (fila central de la figura 5-32), aún existe una preferencia por el sistema adaptativo, tal como indican los valores medios de CMOS. Sin embargo, a medida que se aumenta el número de niveles temporales eliminados en la fase de adaptación, los usuarios dejan de apreciar las ventajas del sistema adaptativo con respecto a asumir las pérdidas sufridas por el sistema convencional. Los test estadísticos con la hipótesis nula indican que la media de CMOS en el último caso (Figura 5-32 derecha abajo) no es estadísticamente diferente de cero.

Todas estas evaluaciones están condicionadas por las técnicas de cancelación del error y compensación de las reducciones de niveles temporales utilizadas, basadas en copia de *frame*. Para otras técnicas, el experimento debe ser repetido.

De las secuencias asociadas a este objetivo, la más repetida es el último caso de la Figura 5-32, precisamente la secuencia con una media nula, estadísticamente hablando. Esta secuencia fue repetida en 12 ocasiones.

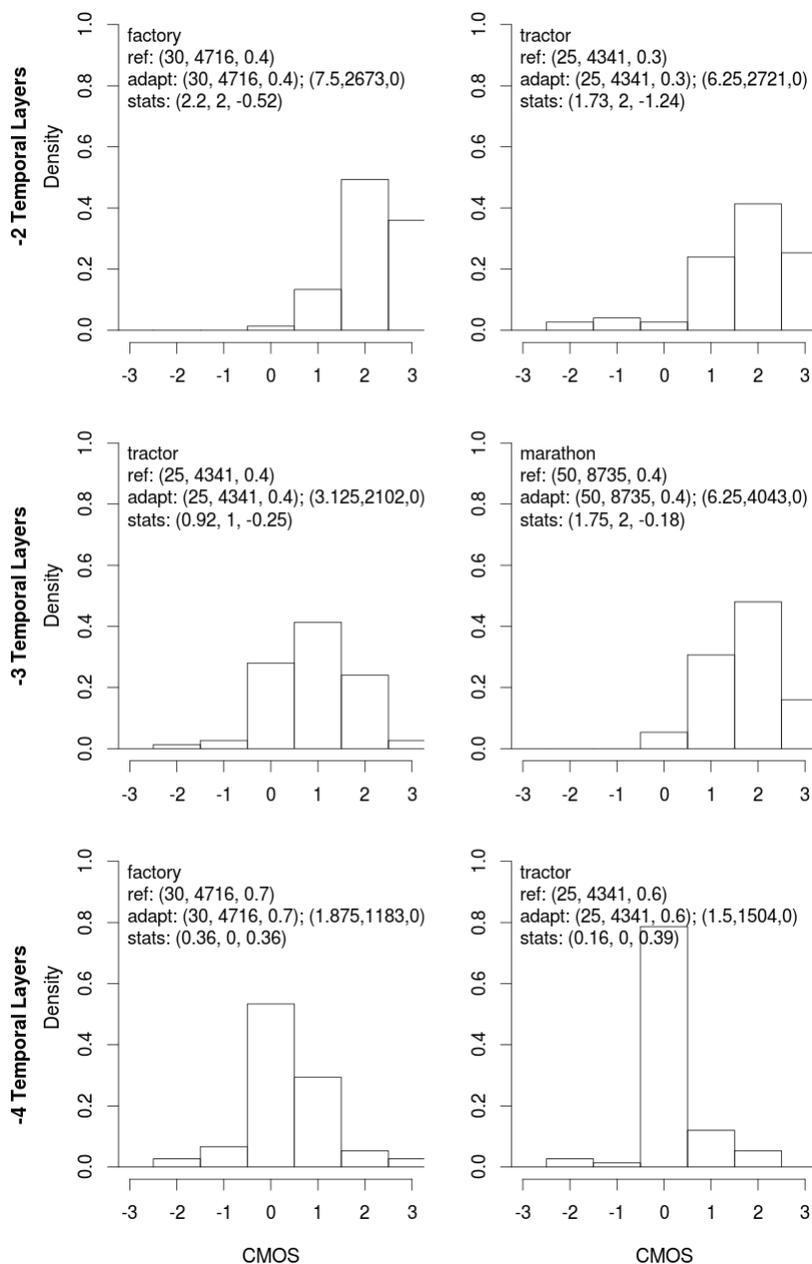


Figura 5-32: Histogramas CMOS para objetivo III

5.2.6.4 Objetivo IV

En los escenarios inversos, de adaptación hacia niveles de calidad superiores, los resultados objetivos indican una preferencia generalista hacia el sistema adaptativo, en comparación con el sistema no adaptativo, como era esperable. Según los resultados presentados en la figura 5-33, todos los casos evaluados reportan valores medios de CMOS positivos, y aunque son estadísticamente distintos de cero, su valor absoluto es pequeño por lo que la preferencia subjetiva no es fuerte. Los factores de asimetría de las distribuciones indican que son mayoritariamente simétricas.

En estas condiciones el coste de incrementar niveles de calidad en las secuencias adaptativas es, al menos, disputable. El coste de duplicar la tasa de transmisión requerida para una secuencia con una capa de calidad adicional puede no estar justificado por una ligera preferencia subjetiva (ligero incremento de QoE). La única excepción encontrada en el subconjunto de secuencias del objetivo es el par marathon (tercer histograma de la primera fila, figura 5-33). Este caso se antoja más sensible a los cambios de calidad. Tiene una distribución fuertemente asimétrica (hacia el lado positivo). Como particularidad encontramos que esta secuencia tiene los índices de complejidad temporal y espacial más altos entre los contenidos del experimento, por lo que se beneficia más del incremento de calidad.

Los participantes que acreditan experiencia con servicios multimedia en Alta Definición detectan más fácilmente las mejoras en la calidad que los usuarios que limitan su experiencia multimedia a la Televisión, como se muestra en la figura 5-34. Para estos últimos usuarios, los valores de CMOS y los p-values correspondientes a los t-test con la hipótesis nula se adjuntan en la Tabla 5-15. De acuerdo a los resultados, todos los casos excepto uno, no son estadísticamente distintos de la hipótesis nula para un nivel de significancia del 1%. La única excepción corresponde al par de la secuencia touchdown con un incremento dos capas de calidad en la secuencia adaptativa. Además este par en particular fue el más repetido del grupo, hasta en 14 ocasiones.

Realizando análisis estadísticos de la verdadera diferencia entre los valores CMOS medios para incrementos de 1 o 2 capas de calidad en la misma secuencia, se observa que las diferencias no son estadísticamente significativas, tabla 5-16.

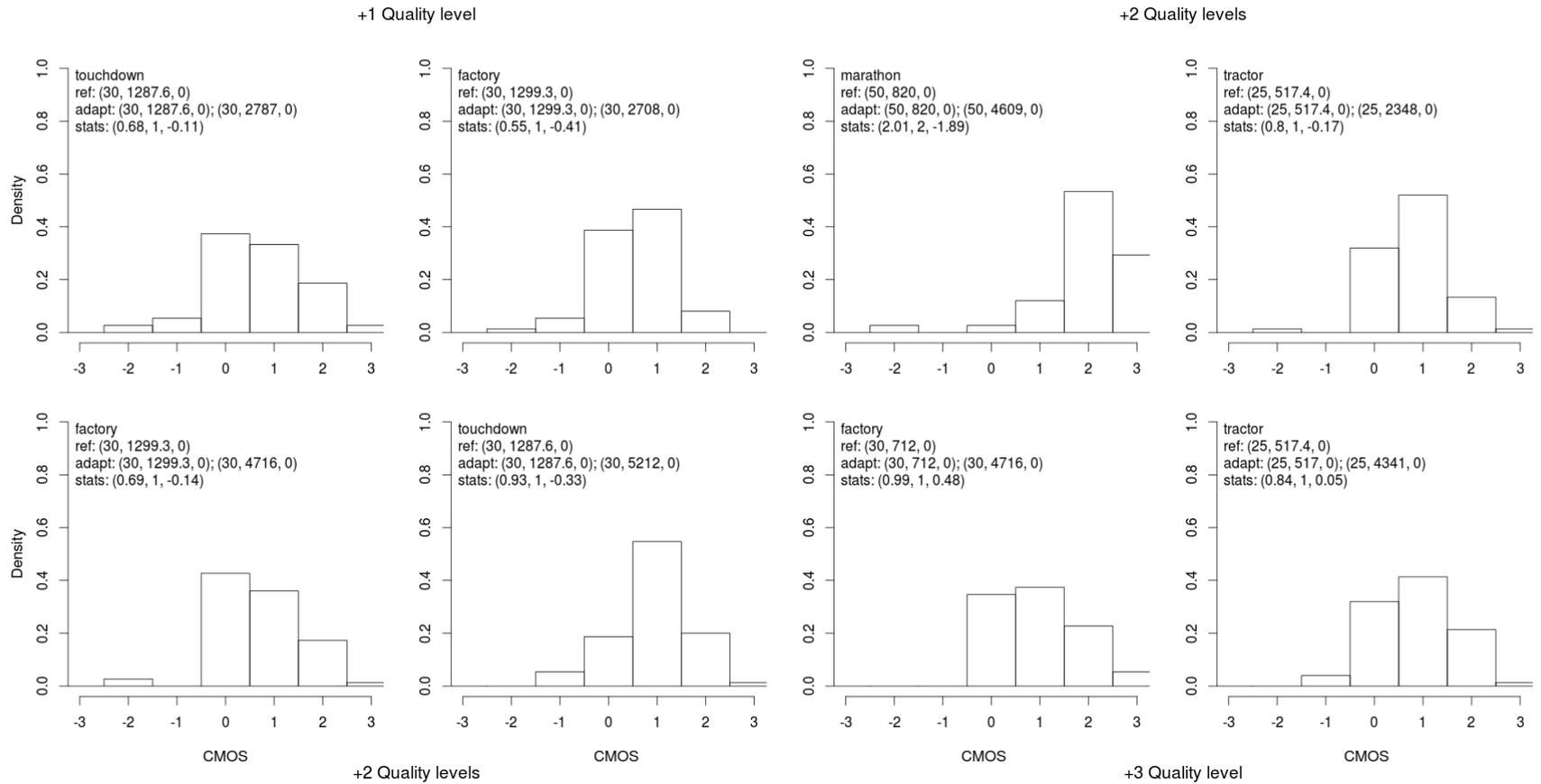


Figura 5-33: Histogramas CMOS para objetivo IV

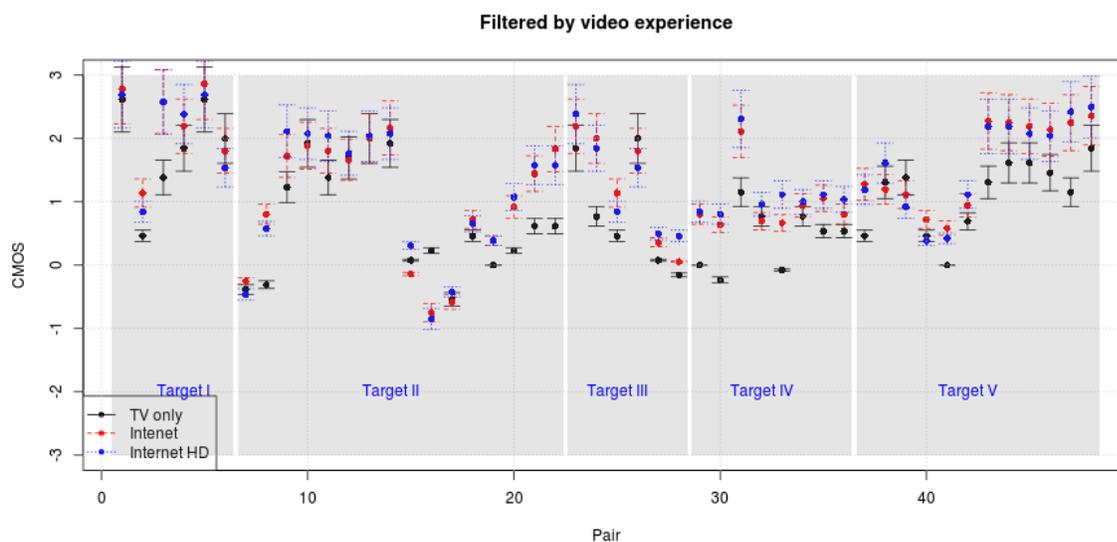


Figura 5-34: CMOS según experiencia con servicios multimedia

Tabla 5—15 Resultados t-test para usuarios con experiencia multimedia limitada a televisión

Pares de secuencias	CMOS promedio	t-test p-valor (alpha 0.01)
Touchdown +1Q	0	1
Factory +1Q	-0.23	0.34
Marathon +2Q	1.15	0.02
Tractor +2Q	0.78	0.04
Factory +2Q	-0.08	0.79
Touchdown +2Q	0.77	0
Factory +3Q	0.54	0.02
Tractor +3Q	0.54	0.89

Tabla 5—16 Resultados t-test para distintos saltos de calidad

Pares de secuencias	CMOS promedio (+1Q)	CMOS promedio (+2Q)	CMOS promedio (+3Q)	Two-sample t-test p-valor
Touchdown (+1Q vs +2Q)	0.68	0.93	-	0.09359*
Factory (+1Q vs +2Q)	0.55	0.69	-	0.287
Tractor (+2Q vs +3Q)	-	0.80	0.84	0.7639

* las medias no son estadísticamente diferentes

5.2.6.5 Objetivo V

En el caso de escenarios de aumento de calidad, y agrupando los resultados para los incrementos en la dimensión temporal, se observa una preferencia subjetiva clara por el sistema adaptativo. Los valores medios de CMOS para el total de la población son siempre positivos y los test estadísticos revelan la diferencia con la media nula es estadísticamente significativa. Además el valor medio de CMOS crece con el número de capas temporales añadidas por el sistema adaptativo, y progresivamente se apila en el extremo positivo de la escala subjetiva, modificando el valor de asimetría de la distribución. La figura 5-35 resume los resultados para este objetivo.

Para los cambios en la dimensión temporal analizados para los usuarios con la experiencia multimedia limitada a la TV, se encuentra algún caso aislado cuyo CMOS no es significativamente distinto de cero para niveles de significancia del 1%. Estos pares están asociados a incrementos de capas temporales de una sola capa. Aún así, no son una diferencia sobre la tendencia general.

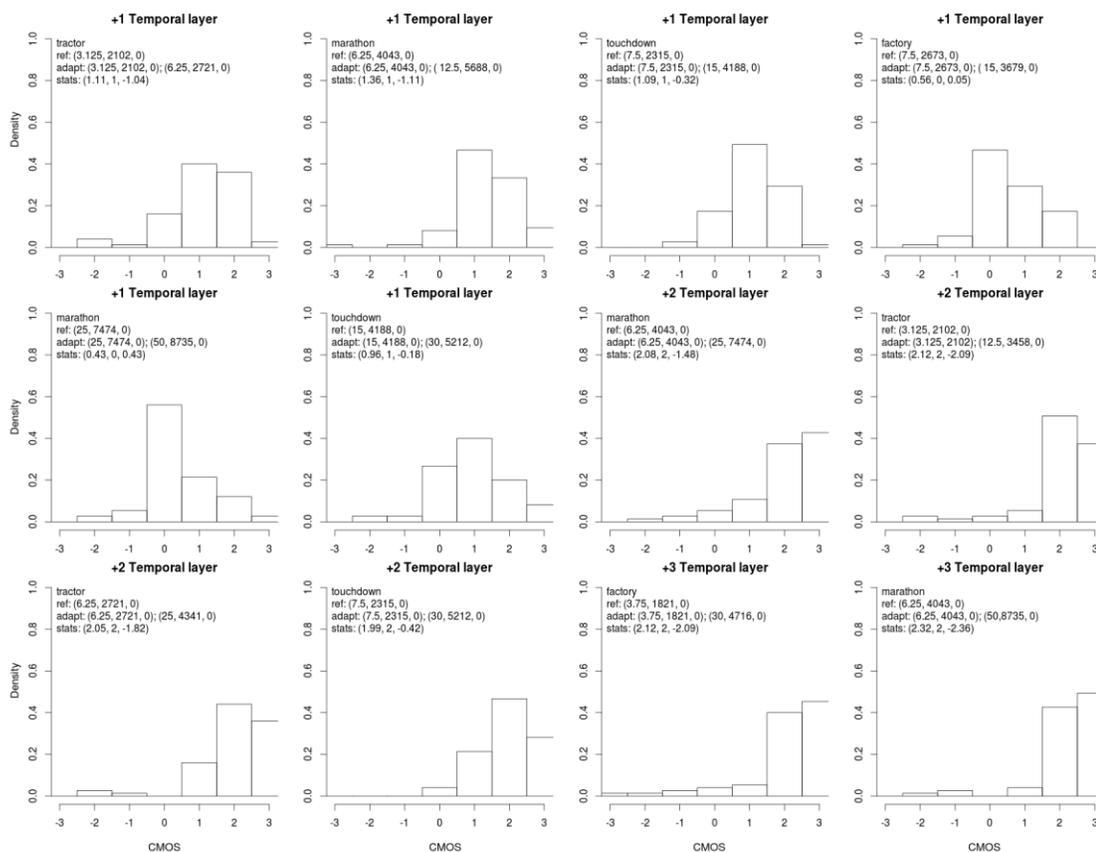


Figura 5-35: Histogramas CMOS para objetivo V

5.2.7 Correlación con métricas objetivas

La obtención de las métricas subjetivas es, como hemos visto, un proceso costoso y largo. El estudio de la correlación entre las métricas objetivas, más fáciles de calcular, y los resultados subjetivos es un recurso obligado. Para corroborar el rendimiento de las métricas FR más relevantes se evalúa su correlación con los resultados subjetivo del experimento realizado con sistemas adaptativos.

A medida que florece el desarrollo de nuevas métricas y algoritmos de medida de la calidad objetiva, hay un número creciente de trabajos que buscan relacionar dichas medidas con parámetros de servicio (QoS) fácilmente observables. En esta Tesis hemos visto cómo es posible predecir las métricas objetivas en sistemas adaptativos a partir de la información de las capas recibidas y la calidad original de cada capa.

La siguiente etapa es pues, relacionar la medida de la métrica objetiva con la verdadera calidad subjetiva (QoE), experimentada por el usuario del servicio, y por extensión, con los indicadores de QoS originales. Una de las teorías más aceptadas, el método IQX, establece que el mejor modelo para relacionar QoS y QoE es el exponencial [173]. Sin embargo, aún quedan varias cuestiones por aclarar. Por ejemplo, no se conoce la influencia de todos los parámetros en la QoE. Existen muchos parámetros que son raramente investigados, como son la interacción del audio/vídeo, el interfaz de usuario o la consciencia que el usuario tiene de la calidad.

Aprovechando la disposición de los resultados subjetivos del experimento subjetivo I realizado, podemos buscar la relación directa entre las métricas objetivas y la QoE siendo agnósticos de todos los parámetros de QoS. Al disponer de los resultados de una evaluación subjetiva mediante una metodología de doble estímulo sobre la elección entre el sistema adaptativo y no-adaptativo en cada par de secuencias se pueden obtener correlaciones directas de las métricas objetivas con las decisiones humanas.

Se utilizan las métricas FR más relevantes, PSNR, SSIM y VQM para evaluar su habilidad de modelar y reproducir las decisiones humanas. Una primera aproximación a la relación entre estas métricas objetivas y observaciones subjetivas fue obtenida, para dos secuencias y 12 participantes [174]. Su conclusión se resume en que para una tasa (*bitrate*) dada, las métricas objetivas producen resultados diferentes en función del contenido mientras que la evaluación subjetiva (en términos de MOS) permanecía similar. Nuestro trabajo, sin embargo, aprovecha los resultados subjetivos de 48 pares de secuencias evaluados por 75 usuarios para revelar una relación exponencial entre las decisiones humanas (CMOS) y las métricas objetivas.

Siguiendo la misma estructura usada en los experimentos subjetivos realizados, se computan las métricas FR (PSNR, SSIM y VQM) para los pares de secuencias. Dado que el resultado de los experimentos subjetivos es expresado en término de CMOS, para obtener un valor objetivo comprable, se calculan las métricas FR para ambas secuencias

del par y a continuación se obtiene la diferencia entre la métrica objetiva para la secuencia adaptada y la no adaptada.

Tal como fue explicado anteriormente, valores positivos en CMOS indicaban una preferencia subjetiva hacia el sistema adaptativo. Análogamente, mediante la diferenciación de las métricas objetivas, un valor positivo en el resultado de la diferencia para las métricas PSNR o SSIM indica una mejor evaluación objetiva de la secuencia adaptada con respecto a la no adaptada. En el caso de VQM, al tener una escala inversa, las diferencias positivas indican una mejor evaluación objetiva de la secuencia de referencia.

Para evaluar la correlación entre las métricas subjetivas y objetivas, primero se realiza una transformación lineal para corregir el signo en VQME. La relación exponencial buscada se calcula mediante la correlación lineal entre los votos subjetivos y el logaritmo de las series de métricas objetivas.

En las Figura 5-36, 5-37 y 5-38 se representan los resultados de la regresión lineal aplicada sobre la transformación logarítmica mencionada y los valores CMOS. En virtud de los resultados gráficos, PSNR es, de las tres métricas evaluadas, la que tiene una peor correlación con las decisiones subjetivas. La métrica PSNR ha sido criticada en numerosas ocasiones como poco adecuada para modelar la QoE [175][176]. Los residuos de los tres modelos no muestran ningún patrón observable. Las medianas de los residuos de los 3 modelos son cercanas a cero.

Las métricas VQM y SSIM arrojan resultados más aproximados según esta propuesta de modelo exponencial, sin embargo ambas métricas destacan un par en concreto como valor atípico. En ambos modelos, el par número 5, correspondiente al a secuencia *marathon* con una adaptación de una capa en la dimensión de calidad comparada con una secuencia afectada por un 40% de pérdidas. De acuerdo a los resultados puramente objetivos, la adaptación mejora la calidad notablemente con respecto a la calidad de la secuencia no adaptada. La misma secuencia, pero con una adaptación en la dimensión temporal (par número 6) mejora "objetivamente" en un grado menor. Otros contenidos con similar complejidad espacial (por ejemplo la secuencia *factory*, par número 3) no destaca tampoco con una diferencia notable en las métricas objetivas de las secuencias adaptada y convencional. Por todo esto, la diferencia entre las métricas objetivas SSIM y VQM para esta secuencia en particular es atribuido a el cambio de escena, presente solamente en la secuencia *marathon*. El efecto de los cambios de escena en el seno de una secuencia adaptada, ya sea en la dimensión temporal o de calidad, debería ser considerada en profundidad desde un punto de vista subjetivo y también objetivo.

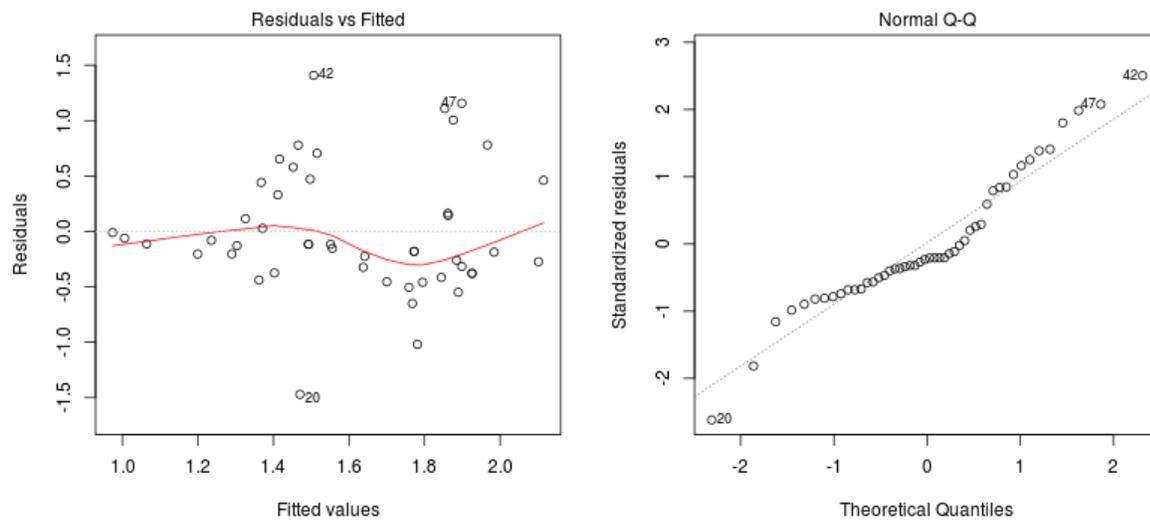


Figura 5-36: Resumen gráfico de la regresión para la métrica PSNR

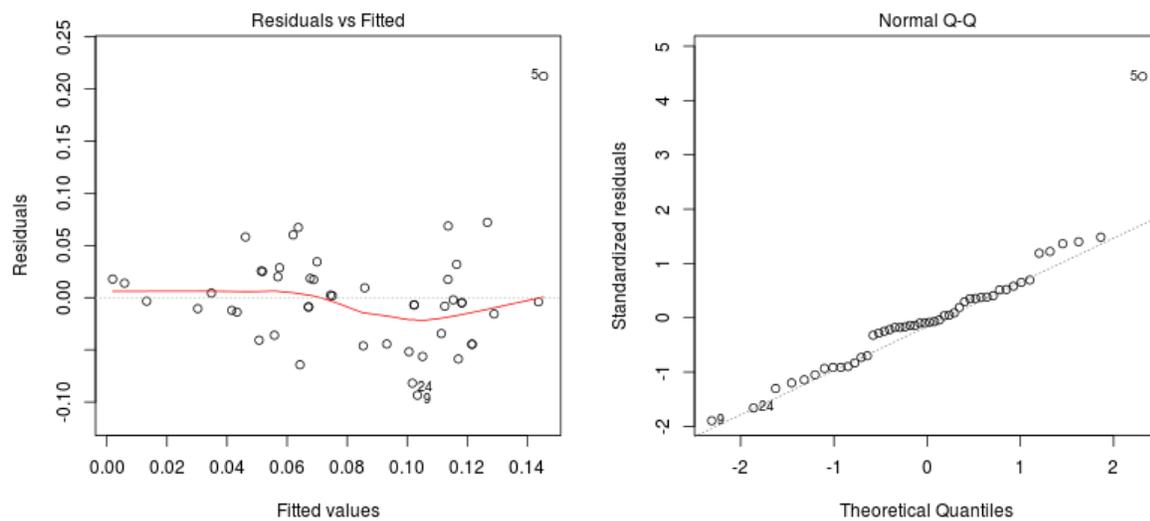


Figura 5-37: Resumen gráfico de la regresión para la métrica SSIM

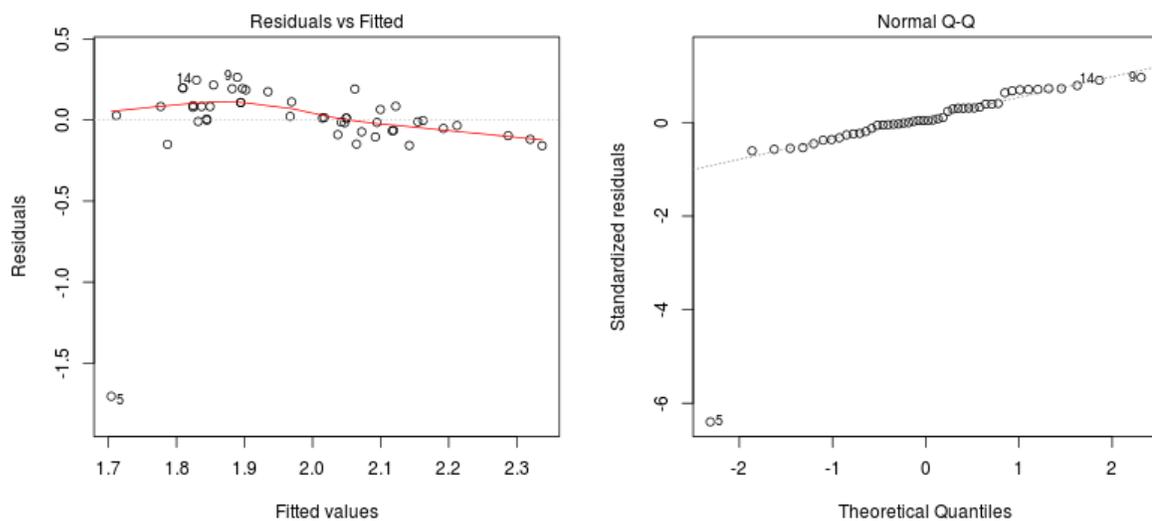


Figura 5-38: Resumen gráfico de la regresión para la métrica VQM

Los valores R-cuadrado obtenidos para cada modelo se presentan en la tabla 5-17. A pesar de los buenos resultados gráficos, los valores de correlación son bajos, ya que las predicciones del comportamiento humano producen normalmente valores más bajos en este tipo de correlaciones [177].

Tabla 5—17 Estadístico R-cuadrado para los modelos de regresión de las métricas objetivas

Métrica	Coefficiente R cuadrado múltiple
PSNR	0.2055
SSIM	0.3461
VQM	0.2489

Los resultados globales han demostrado una buena correlación entre los valores CMOS y la diferencia de métricas objetivas. Si se analizan las relaciones por objetivo, Tabla 5-18, los mejores resultados de correlación son proporcionados por la métrica VQM mientras que los peores índices de correlación corresponden a la métrica PSNR. El resultado de la correlación con un modelo lineal, sin la transformación logarítmica, empeora de media un 0.76% los índices de correlación. Analizado por objetivo además, los objetivos con menos pares de secuencias son más difíciles de correlar. Por otra parte, las características de las secuencias en los objetivos también determina diferencias en las correlaciones.

Para el objetivo I, que incluye el par número 5 mencionado antes como valor atípico, la correlación excluyendo este par se incrementa hasta 0.81 y 0.79 para las métricas SSIM y VQM respectivamente.

La correlación de SSIM para el objetivo III es débil debido fundamentalmente a la influencia del algoritmo de cancelación de error (*frame-copy*) en esta métrica. La preferencia subjetiva se inclina predominantemente hacia la secuencia adaptada, pero la métrica SSIM da valores muy bajos a estas secuencias.

En el objetivo II la correlación de las métricas SSIM y VQM es débil. Como refleja la figura 5-39 para la relación entre CMOS y el logaritmo de la diferencia entre las métricas (Delta SSIM y Delta VQM), no existe una función clara entre ambos conjuntos de datos. Determinar el umbral de pérdidas que los usuarios están dispuestos a tolerar es incluso más difícil basándose en las métricas objetivas que en las métricas subjetivas. En particular, los experimentos subjetivos para determinar dicho umbral incluyen otros factores no tenidos en cuenta en las métricas de calidad objetivos, como el tipo de contenido o las condiciones iniciales. Modelar ese tipo de comportamientos en las métricas objetivas es complejo. En el caso de PSNR en el objetivo IV la correlación es muy pequeña. Dado que la métrica PSNR no considera las dependencia *inter-frame*, el beneficio de incrementar la tasa de transmisión o *frame-rate* en las secuencias adaptadas no es reflejado en esta métrica.

Tabla 5—18 Resultados de la correlación lineal entre logaritmo de métricas objetivas y resultados subjetivos por objetivo

Grupos (pares por objetivo)	PSNR subjetivo-objetivo	SSIM subjetivo-objetivo	VQM subjetivo-objetivo
Target I	0.7391	0.627604	-0.5476
Target II	0.459	0.6306	-0.6607
Target III	0.63123	0.51276	-0.85571
Target IV	0.2008	0.8728	-0.7988
Target V	0.5328	0.68202	-0.7534

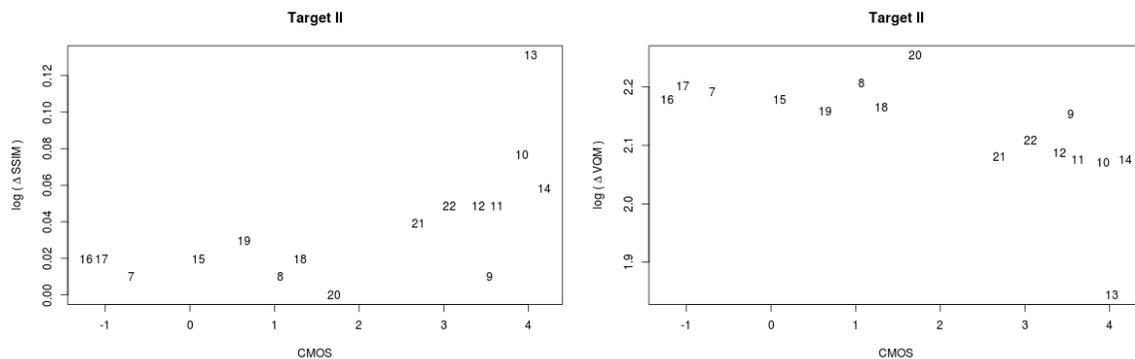


Figura 5-39: Correlación de CMOS con métricas SSIM Y VQM para el objetivo II

5.3 Contribuciones del capítulo

Los servicios multimedia en Internet están expuestos a muchos problemas derivados principalmente del proceso de transmisión. Aunque existen muchas propuestas para evitar o paliar los efectos de las degradaciones producidas en la transmisión, por ejemplo [178], no todas las tecnologías actuales implementan todos los mecanismos necesarios y en último término, se producen degradaciones que afectan negativamente a la experiencia del usuario.

Una contribución destacada de los dos estudios subjetivos realizados durante el desarrollo de esta Tesis es el volumen y la heterogeneidad de la población que se involucra en ambos estudios, sumando 175 participantes. Por un lado, supera con mucho el límite inferior recomendado por ITU-T, pero aumenta significativamente también las poblaciones encontradas en la literatura especializada. En términos de heterogeneidad, la diversión de edades, perfiles y experiencia con servicios multimedia evita que los resultados obtenidos puedan inducir a error debido al sesgo de considerar solo poblaciones especializadas, con un rango de edad similar y mismo perfil, como en [73].

El primer experimentos subjetivo contribuye aportando conclusiones transversales a los servicios multimedia en general. Se verifica que los eventos de *rebuffering* son un factor muy influyente en la degradación de QoE. En cambio la asincronía entre audio y vídeo, aunque pudiera considerarse molesto, es evaluado como el factor degradante menos relevante. En relación al contenido, los usuarios toleran mejor las degradaciones en contenidos del tipo "cartoons". En cambio, ciertas degradaciones, como las pérdidas de audio tienen especial incidencia en la QoE de algunos tipos de contenido como "news" o "action". No se han encontrado diferencias entre las degradaciones producidas al inicio o al final de la secuencia, probablemente porque la longitud de las secuencias no es suficiente como para que la evaluación se viese influenciada por el efecto memoria,

según el cual los usuarios evalúan más negativamente las degradaciones que han visto más recientemente.

Al profundizar en los estudios subjetivos centrados en la evaluación de sistemas adaptativos, se pone rápidamente de manifiesto que la QoE resulta beneficiada por las mejoras introducidas en estos sistemas, pero no está claro dónde y cómo actuar en cada caso. Nuestra contribución constituye un análisis muy detallado los factores clave que pueden ser utilizados en el diseño de los algoritmos de decisión necesarios en los sistemas adaptativos. En la metodología seguida en el segundo experimento se definen objetivos claros para identificar dichos factores clave para realizar el ajuste fino de los sistemas adaptativos en términos de QoE. Cuando existe excedente de ancho de banda en la red, los sistemas adaptativos son claramente preferidos por los usuarios. En el otro extremo, cuando la degradación en la transmisión es alta, los sistemas adaptativos también son marcadamente preferidos por los usuarios. Sin embargo cuando existen situaciones de congestión moderada la tendencia no es tan clara y en ocasiones se invierte. Hemos determinad cómo, para tasas bajas de PLR, los sistemas tradicionales obtienen mejores datos en QoE. En términos generales, la población valora mejor la adaptación en la dimensión de calidad que la reducción de niveles temporales. Se han visto diferentes tendencias en función del grupo de edad. La población más joven (por debajo de 35 años) es menos moderada en sus evaluaciones mientras que la población mayor de 35 años suele proporcionar evaluaciones más próximas al valor central de indiferencia.

Como última contribución, se cierra el círculo de la evaluación de servicios multimedia relacionando las métricas subjetivas con las métricas objetivas. En particular, la metodología subjetiva por comparación aplicada al experimento subjetivo centrado en SVC permite evaluar un aspecto distinto a las métricas objetivas. Se permite evaluar su capacidad de tomar decisiones en el mismo sentido que las decisiones subjetivas. De esta forma se habilitan métricas objetivas como indicadores del rendimiento subjetivo del sistema.

6 Conclusiones

Hoy en día, el afán por proporcionar más y mejores servicios hace que se avance continuamente en el desarrollo de nuevas tecnologías, mayores infraestructuras y nuevos paradigmas. El mundo de los servicios adaptativos es un ejemplo más de la aplicación de las nuevas ideas al mercado más exigente. En el ciclo de desarrollo natural de los servicios multimedia hemos encontrado que la fase de evaluación constituye una etapa clave.

En esta Tesis se han realizado diversas contribuciones a los tres bloques clave descritos en la fase de evaluación del proceso de desarrollo de servicios multimedia: simulación, evaluación objetiva y evaluación subjetiva. Se han satisfecho todos los objetivos propuestos al inicio de esta Tesis.

Para analizar las prestaciones de servicios multimedia, especialmente en las etapas iniciales de su desarrollo, se recurre con frecuencia a las técnicas de simulación. Además de las técnicas clásicas de simulación se han investigado otras alternativas, como la emulación y, particularmente, la simulación en tiempo real. La simulación en tiempo real habilita la construcción de entornos de evaluación flexibles y económicos sobre los que desplegar el servicio real. En este tipo de entornos, la evaluación debe ser cuidadosamente vigilada ya que, como se ha visto, pueden producirse desviaciones en los resultados no relacionadas con el servicio propiamente dicho, sino con problemas de rendimiento en el entorno de evaluación. Hemos determinado con precisión los límites existentes para los simuladores NS-3 y OPNET en su funcionamiento como simuladores en tiempo real. Encontrar los límites, por debajo de los cuales, los resultados del experimento son confiables resulta de vital importancia, como paso previo a cualquier evaluación de servicios.

Como recurso ante los problemas de rendimiento que se evidencian en los emuladores, la comunidad ha aportado una serie de técnicas que buscan incrementar el rendimiento de los entornos de evaluación. La construcción de los entornos de evaluación puede resultar compleja al tener que manejar una gran variedad de técnicas. Al detectarse la carencia de sistematización en este campo, hemos formulado una metodología específica para la construcción de escenarios escalables destinados a la evaluación de servicios multimedia.

Concluyendo el primer bloque, se evidencia que la simulación en tiempo real constituye una técnica versátil y ágil para la construcción de entornos de evaluación, pero su escalabilidad está limitada. En este tipo de evaluación, el efecto del rendimiento de la herramienta y su entorno es vital para obtener resultados confiables de la evaluación realizada. Con la aplicación de la metodología desarrollada se mejora la escalabilidad, gracias a la ejecución controlada y el uso guiado de las distintas técnicas de apoyo a la evaluación. Sin una metodología como la propuesta, el investigador se enfrenta a una incertidumbre mayor en la construcción de entornos de evaluación confiables. Sin embargo, a pesar de las propuestas analizadas o la metodología propuesta, la técnica de simulación en tiempo real se considera de escalabilidad limitada y primordialmente aplicable a las primeras fases de desarrollo del servicio.

Independientemente del entorno de evaluación, la monitorización de los resultados del servicio se realiza mediante métricas objetivas. En los servicios multimedia las métricas objetivas buscan incorporar todos los aspectos posibles del flujo multimedia que las métricas de servicio exclusivamente dejan fuera. Cuando los servicios multimedia se especializan, como es el caso de los servicios adaptativos, las métricas deben adaptarse consecuentemente.

En el campo de las métricas objetivas, las implicaciones prácticas de este bloque están estrechamente relacionadas con las contribuciones aportadas. En primer lugar, hemos desarrollado un decodificador que habilita el trabajo con secuencias adaptativas que incluso han sufrido distorsiones en el proceso de transmisión. Después, hemos contribuido desarrollando distintas métricas objetivas de aplicación a las secuencias adaptativas.

Por un lado, la modificación propuesta en la métrica cPSNR habilita la evaluación de la calidad de transmisión en sistemas adaptativos manteniendo la escala de una métrica muy popular y conocida. La correlación de la métrica modificada con las alternativas existentes ha sido contrastada usando una librería pública de evaluaciones subjetivas. Esta métrica se integra, junto con otras utilidades y métricas en una herramienta de análisis (DMMSAF) desarrollada específicamente para facilitar los trabajos de análisis de secuencias adaptadas.

Además, se ha postulado un método de estimación de métricas FR en entornos adaptativos que permite la inclusión de estas métricas como indicadores de calidad en tiempo real. Este método de estimación se ha revelado muy interesante, y está abierto a la inclusión de más métricas, precisando solamente la calibración de las mismas.

Como conclusión de este segundo bloque de la Tesis, todos estos avances han ayudado a mejorar la forma en que se evalúan los servicios multimedia adaptativos. Las herramientas construidas facilitan el trabajo, pero las métricas y métodos de estimación propuestos abren muchas posibilidades dentro del ámbito de los servicios adaptativos. En particular, la inclusión de la estimación de métricas FR como factores de peso en la

decisión de adaptación puede suponer un avance en la calidad *end-to-end* del servicio prestado como no se había planteado hasta el momento.

Como las métricas objetivas deben respaldarse con evaluaciones subjetivas, este es un campo que no puede dejarse fuera en esta investigación. En el tercer bloque de esta Tesis, se desarrollan los trabajos realizados en el campo de la experimentación subjetiva. Además se ha cuidado especialmente de involucrar a una población extensa y heterogénea que de una mayor representatividad a los resultados obtenidos. En total, se ha consultado a más de 175 individuos, de edades comprendidas entre los 20 y los 68 años, de perfiles diferentes y con variada experiencia en el entorno multimedia. Esta población supone una diferencia notable con las encontradas hasta el momento en la literatura por su volumen y pluralidad.

Como primera conclusión, los experimentos subjetivos realizados desde una perspectiva global demuestran cómo, la ponderación de los distintos factores que alteran la calidad del servicio, depende de forma directa del tipo de contenido. Las métricas de QoS quedan definitivamente relegadas a una segunda división y cada vez es más importante involucrar métricas de QoE completas para representar todos los aspectos del servicio. Además, la detección de los factores más influyentes en la calidad global, según el tipo de contenido, permite a los diseñadores del servicio enfocar las soluciones hacia el factor degradante que más perjudica su servicio.

La identificación, desde un punto de vista subjetivo, de los factores clave para disparar los eventos de adaptación, es determinante para los algoritmos de decisión imprescindibles en los sistemas adaptativos. Sin embargo, aún es necesario extraer los factores objetivos que puedan ser realmente aplicados a un algoritmo real. En la última parte de la Tesis se establece la correlación de las métricas objetivas con los experimentos subjetivos realizados en el campo de sistemas adaptativos. Esto ha permitido identificar los indicadores objetivos que tienen importancia en los sistemas adaptativos. Concluimos que los indicadores objetivos son capaces de seguir de forma satisfactoria la dirección de las decisiones subjetivas.

En conclusión, el trabajo realizado constituye una aportación múltiple en la fase de medición del ciclo de mejora continua del servicio multimedia. Desde las contribuciones realizadas para mejorar el diseño de entornos de evaluación, pasando por las métricas propuestas para medir de forma efectiva las degradaciones del servicio en el entorno diseñado y llegando hasta las claves encontradas para guiar el desarrollo de los servicios, gracias a la realización de los experimentos subjetivos.

7 Trabajo futuro

El trabajo desarrollado en esta Tesis supone un paso adelante en el camino de mejora continua de los procesos de evaluación y medición de servicios multimedia. Especialmente en el campo de las tecnologías adaptativas, donde existe un importante nicho de desarrollo. Con todo, existen algunas líneas de crecimiento para continuar los trabajos desarrollados en esta Tesis.

La evaluación de la técnica de simulación en tiempo real debe ser extendida, incorporando el estudio de otras herramientas. La mejora de los procesos de simulación para habilitar la ejecución paralela supondría un avance importante en la escalabilidad de esta técnica. La aplicación de nuevos modelos de simulación sobre los simuladores ya utilizados en esta Tesis permitiría incrementar la escalabilidad de la simulación. En este sentido el desarrollo del simulador NS-3 es dinámico debido a la existencia de una comunidad de usuarios activa, por lo que la revisión continua de los avances en este simulador es necesaria.

Como hemos visto, existe una reciprocidad entre el rendimiento del simulador y la precisión de los resultados. Es de vital importancia mantener la vigilancia continua del rendimiento de las herramientas utilizadas. En la búsqueda de aliviar la complejidad computacional de las simulaciones, uno de los caminos de desarrollo más interesantes lo constituye la aplicación de las técnicas de simulación orientada a flujos sobre el simulador NS-3. Obtener modelos efectivos de los servicios bajo estudio u otros tráficos competentes permitiría el crecimiento de la escalabilidad de forma controlada.

Un aspecto que continua preocupando acerca del realismo de las evaluaciones de servicios es la modelización del entorno de Internet dentro del simulador. Las técnicas de emulación alivian en parte esta preocupación incorporando aspectos reales del servicio bajo estudio, pero aún se desconoce en muchos aspectos el comportamiento del entorno y el tráfico existente en la red Internet. En este sentido conviene considerar las plataformas de colaboración como PlanetLab. La evaluación del rendimiento del servicio bajo estudio en este tipo de plataformas permitiría completar el conocimiento del impacto del entorno real de Internet sobre el servicio desarrollado y viceversa.

Por otra parte, sería interesante profundizar en la implementación de la metodología desarrollada en un sistema experto que proporcione recomendaciones de forma automatizada o incluso que aporte sugerencias en el diseño del entorno de evaluación. El

sistema experto podría ser entrenado analizando los diferentes recursos consumidos en el entorno de tal forma que la medida retroalimente el proceso de selección de técnicas de simulación.

En el campo de las métricas objetivas, existe un largo camino aún para obtener la métrica objetiva definitiva. El primer paso es continuar incorporando aspectos audiovisuales a las métricas de QoE objetivas que permitan incrementar la precisión de estas métricas. Según las propias recomendaciones de la ITU, cualquier modelo de QoE debe incluir el desarrollo de la métrica objetiva basada en parámetros medibles y respaldada por un estudio subjetivo involucrando una población suficiente. Es necesario mantener la vigilancia continua de los avances en las métricas objetivas para incorporar en los estudios futuros.

La propuesta de estimación de métricas FR en recepción a partir de la información reducida proporcionada por el servidor supone un interesante campo de estudio y aplicación de nuevas métricas objetivas que puedan desarrollarse. Es necesario continuar la evaluación con diferentes secuencias en diferentes escenarios para validar que el método de estimación es aplicable a otros entornos. Además resultaría interesante evaluar el comportamiento dinámico de esta estimación para secuencias largas, valorando la inclusión de un protocolo de intercambio de los metadatos relativos a la calidad de codificación de capa cada cierto intervalo, también a definir. En este sentido se prevé un crecimiento muy necesario de la herramienta de análisis desarrollada (DMMSAF).

En cuanto a las evaluaciones subjetivas, existe un vasto camino por recorrer. En el futuro más inmediato, se pretende utilizar los resultados subjetivos obtenidos para evaluar las métricas objetivas. Realizar correlaciones de las métricas objetivas existentes con resultados subjetivos específicos permitirá validar o incluso proponer mejoras sobre las métricas existentes. Además, de esta forma se rentabilizan los experimentos subjetivos.

En un futuro próximo existen muchas preguntas por resolver que requieren el planteamiento de nuevos y más extensos experimentos subjetivos. Experimentos que involucren entornos más reales, considerando incluso diferentes dispositivos y tamaños de pantalla dentro del mismo experimento.

Muchas de las sugerencias proporcionadas por los usuarios entrevistados en el transcurso de los experimentos desarrollados deben ser consideradas. Por ejemplo abrir la posibilidad de seleccionar los contenidos como parte del experimento, garantizaría de esta forma que el interés por la evaluación se mantuviese constante. Muchos usuarios enfatizan la necesidad de incluir una escala subjetiva adicional que permita evaluar el desinterés acumulado hasta el punto crítico que supone el abandono de la visualización actual dentro de la prueba, un comportamiento muy contemporáneo y habitual en los servicios multimedia en Internet.

En este sentido una técnica de experimentación como *crowdsourcing* puede abrir nuevos horizontes que son limitados en la evaluación tradicional de test presenciales. Aunque se ha argumentado en contra de los experimentos de tipo *crowdsourcing* a lo largo de esta Tesis, las propuestas de los usuarios recogidas se pueden enfocar en este campo. La principal ventaja es la posibilidad de evaluar el servicio bajo estudio en una población potencialmente más numerosa a un coste menor. En contrapartida, se pierde el control de los detalles del experimento y los resultados tienden a ser menos confiables. Algunos de los problemas inherentes a los experimentos de este tipo pueden aliviarse incidiendo sobre el diseño del experimento, pero otros aún escapan al control del evaluador.

8 Publicaciones derivadas de la Tesis

Como resultado de las investigaciones desarrolladas en esta Tesis se producen 9 publicaciones en distintos congresos y revistas. A continuación se relacionan incluyendo una breve descripción de cada una

El primer trabajo explorando la técnica de simulación en tiempo real se presenta en el congreso internacional *Simutools*.

A. Alvarez, R. Orea, S. Cabrero, X. G. Pañeda, R. García, y D. Melendi, «Limitations of network emulation with single-machine and distributed ns-3», en *Proceedings of the 3rd International Conference on Simulation Tools and Techniques, Torremolinos, Spain, 2010, pp. 1-9.*

Abstract: Research on large-scale internet services requires an extensive evaluation prior to deployment. A good analysis must include tests over large networks, using real devices and a considerable number of users. However, how to test in these scenarios with many users is an open question. Network emulation can be a good alternative before real deployments, which are complex and expensive. In this paper, we examine the new ns-3 network simulator/emulator in order to determine its capacity in the evaluation of large scale services. For that purpose, a real client/server video service is deployed over an emulated network. The service is progressively scaled up by increasing the number of clients on a single machine. In addition, we have extended ns-3 to support a distributed architecture for network nodes, thus, we repeat the experiments with a distributed set-up. Advantages, disadvantages, possibilities and limitations of both approaches are thoroughly discussed.

El trabajo que explora los límites y la integración de los dos populares simuladores, NS-3 y OPNET, en tiempo real ha sido publicado en la revista *Journal of Simulación*.

A. Álvarez, L. Pozuenco, X. G. Pañeda, R. García, D. Melendi, S. Cabrero, y G. D. Orueta, «Limits for the real-time simulation of video services over commodity hardware», *J. Simul.*, jun. 2015. [JCR 0.58 (2014)]

Abstract: The results from event-driven simulations with video services can present limitations because of the cost of developing accurate models. Creating a video

server model precise in how streams are transmitted is generally as expensive as the construction of the real server, thus several simplifications are generally introduced to reduce the effort. In this context, the lack of accuracy could be devastating if metrics such as jitter are imprecise. One potential alternative to solve these issues is the combination of real services with Real-Time (R-T) capable network simulators. This paper evaluates R-T simulation capabilities of two of the most popular tools, NS-3 and OPNET to support the performance evaluation of video services using a commodity hardware testbed. Taking a video on demand service as a test case, we have studied the use of NS-3 and OPNET in situations where their more profitable characteristics, such as instance creation simplicity or the use of analytic traffic, could produce the best results. The performed tests show how NS-3 is able to replicate an access network topology with up to 10 MPEG-4 video streams or OPNET has been able to replicate a ring core network with up to 60 of the same streams without affecting important metrics as important as jitter. These results point out the validity of R-T simulation in the context of services that have streams as a main component if their strengths are taken into account.

Fruto de la experiencia adquirida en el campo de la simulación y las técnicas de escalabilidad, se postula una metodología de construcción de entornos de evaluación que ha sido publicada en la revista *IEEE Magazine*.

A. Alvarez, R. Garcia, S. Cabrero, X. G. Paneda, D. Melendi, y R. Orea, «In pursuit of massive service emulation: a methodology for testbed building», *IEEE Commun. Mag.*, vol. 49, n.o 9, pp. 162 -168, sep. 2011. [JCR 3.785]

Abstract: Every day, more services in the Internet (Youtube, Facebook, etc.) are being massively used, while network operators struggle to maintain their performance. Since simulations are not always accurate and evaluation prior to deployment is a must, it is often convenient to build an emulation testbed to test these services in more realistic environments. This methodology proposes four steps to guide researchers in this task. Before running the experiments and gathering results, it is necessary to specify the service under test, define the testbed requirements, and allocate the resources properly. For this reason, a set of techniques are proposed to solve the most common challenges at every step, such as adding real users or selecting the best simulation technique for our goals. The purpose is to efficiently employ the available resources without compromising the realism and accuracy of results. The application of the methodology to a case study, a high-quality audio and video ondemand service over a communication network, shows how resource constraints can be overcome to accomplish our goals.

En el bloque temático de la calidad objetiva se han realizado dos publicaciones. La primera de ellas, describiendo el framework desarrollado y la métrica cPSNR, fue publicado en el congreso internacional *Ubiconet*.

Alberto Alvarez, Sergio Cabrero, Xabiel García Pañeda, Roberto García, David Melendi, Rafael Orea, «A Flexible QoE Framework for Video Streaming Services», en Workshop on Ubiquitous Computing and Networks, 2011.

Abstract: QoE metrics are a must in new service evaluations as they have shown to outperform traditional QoS. This work presents a flexible framework for measuring different types of QoE analysis and capable of synchronizing reference sequence with distorted videos for avoiding erroneous coupling. We define cPSNR metric, a measurement system suitable for performing the evaluation of transmission impairments exclusively, from the perceived quality perspective. The metric is validated with a subjective video database showing a promising correlation with both packet loss ratio and subjective quality.

La segunda publicación en el ámbito de las métricas objetivas se publica en la revista *International Journal of Business Data Communications and Networking*. En este trabajo se publica el método de estimación de métricas FR.

A. Álvarez, L. Pozueco, S. Cabrero, X. G. Pañeda, R. García, D. Melendi, y G. Díaz, «A Framework to Measure and Estimate Video Quality in SVC Real-Time Adaptive Systems»:, *Int. J. Bus. Data Commun. Netw.*, vol. 10, n.o 1, pp. 47-64, 2014.

Abstract: Effectively adapting the content to network conditions in real-time is an important matter in best-effort networks like the Internet. Scalable Video Coding (SVC) is an interesting alternative to implement such systems. However, some problems of the performance evaluation of SVC based adaptive systems have not been solved. We review the main efforts directed to measure video quality on SVC related systems and discuss the limitations of each one. This paper elaborates a framework to measure video quality metrics in real adaptive SVC based streams. An estimation method for full reference video quality metrics is proposed. This method reduces reference information required and it is able to provide real-time accurate results simply using metadata regarding the video quality of the reference layers. The video quality of several streams that have been generated using a real-time adaptive system is first measured with the elaborated framework and then estimated with the proposed method.

En la temática de las evaluaciones subjetivas, el trabajo realizado en esta tesis ha dado lugar a cuatro publicaciones. Los primeros resultados se publican en 2013 en el congreso *Jitel*. También se publican otros tres trabajos en las revistas *Computer Communications*, *Network Protocols and Algorithms* y *New Review of Hypermedia and Multimedia*.

En el trabajo presentado en congreso Jitel se postulan los primeros resultados obtenidos de la evaluación subjetiva realizada comparando un sistema adaptativo y uno convencional.

Alberto Álvarez, Laura Pozueco, Sergio Cabrero, Xabiel García Pañeda, Roberto García, David Melendi, Gabriel Díaz Orueta, «Evaluaciones subjetivas de servicios streaming adaptativos vs no-adaptativos.», Jornadas de Ingeniería Telemática (JITEL), Granada, España., 2013.

Resumen: El análisis subjetivo de sistemas streaming adaptativos es un proceso clave para optimizar algoritmos que permitan una adaptación basada en medidas de calidad de la experiencia (QoE). En este trabajo abordamos el estudio comparativo de un sistema streaming adaptativo frente a un sistema streaming tradicional (no adaptativo) desde el punto de vista subjetivo. Se evaluará en qué situaciones y condiciones un sistema adaptativo mejora la calidad percibida por el usuario y cuáles son los umbrales de tolerancia a pérdidas que marquen el inicio del proceso de adaptación. Para ello se plantean diferentes situaciones de disponibilidad de ancho de banda en la red y diferentes decisiones de adaptación. Los resultados del estudio, en el que han participado 75 usuarios, resuelven cuestiones clave para el diseño de sistemas adaptativos y muestran que la adaptación mejora la experiencia de usuario en la mayoría de las condiciones evaluadas.

Un análisis completo de los resultados subjetivos obtenidos en el primer experimento es publicado en el revista *Computer Communications*, también en 2013.

A. Alvarez, L. Pozueco, S. Cabrero, X. G. Pañeda, R. Garcia, D. Melendi, y G. D. Orueta, «Subjective evaluation of critical success factors for a QoE aware adaptive system», *Comput. Commun.*, vol. 36, n.o 15-16, pp. 1608-1620, sep. 2013. [JCR 1.079]

Abstract: For many years video content delivery has established itself as the killer application. Improving QoE on adaptive streaming is focusing many efforts in the quest for optimized methods and metrics to allow a QoE driven adaptation. Questions such as whether adaptive systems based on Scalable Video Coding improve subjective quality and in which situations or to what degree are still open issues. Tolerance and indifference thresholds for each type of content, conditions or viewer category, with regard to adaptive systems are critical success factors that are yet unresolved. We compare the performance of a complete adaptive system with the traditional, i.e. non-adaptive, approach in subjective terms. Results of surveying 75 participants show that the adaptation improves QoE under most of the evaluated conditions. Tolerance thresholds for triggering adaptation events have been identified. Users accustomed to Internet video are more critical than users that only watch TV. The under 35 year old subset among the available population is generally more satisfied with the adaptive system than the older subset.

El trabajo presentado en el congreso Jitel fue seleccionado para su publicación en la revista internacional *Network Protocols and Algorithms*. Invitados por el comité del congreso *Jitel*, se publica el siguiente trabajo, donde además se explora la relación entre la preferencia por un sistema de streaming adaptativo u otro convencional desde el punto de vista subjetivo y objetivo.

A. Álvarez, L. Pozueco, S. Cabrero, X. G. Pañeda, R. García, D. Melendi, y G. D. Orueta, «Adaptive Streaming: A subjective catalog to assess the performance of objective QoE metrics», *Netw. Protoc. Algorithms*, vol. 6, n.o 2, pp. 123-136, jun. 2014.

Abstract: Scalable streaming has emerged as a feasible solution to resolve users' heterogeneity problems. Scalable Video Streaming (SVC) is the technology that has served as the definitive impulse for the growth of streaming adaptive systems. Systems need to incorporate user's Quality of Experience (QoE) in the layer switching algorithms. This paper characterizes the relation between human decisions and objective metrics that could be mapped into systems. We have performed extensive subjective experiments to corroborate the preference towards adaptive systems when compared to traditional non-adaptive systems. The resulting subjective scores are correlated with most relevant Full Reference (FR) objective metrics. We obtain an exponential relationship between human decisions and the same decisions expressed as a difference of objective metrics. A strong correlation with subjective scores validates objective metrics to be used as aid in the adaptive decision taking algorithms to improve overall systems performance. Results show that, among the evaluated objective metrics, PSNR is the metric that provide worse results in terms of reproducing the human decisions.

Por último, recientemente se han publicado los resultados del experimento subjetivo *end-to-end* en la revista *New Review of Hypermedia and Multimedia*.

L. Pozueco, A. Álvarez, X. G. Pañeda, R. García, D. Melendi, y G. D. Orueta, «Subjective video quality evaluation of different content types under different impairments», *New Review of Hypermedia and Multimedia*, 1-28, 2016 [JCR 0.586]

Abstract: Nowadays, access to multimedia content is one of the most demanded services on the Internet. However, the transmission of audio and video over these networks is not free of problems that negatively affect user experience. Factors such as low image quality, cuts during playback or losses of audio or video, among others, can occur and there is no clear idea about the level of distortion introduced

in the perceived quality. For that reason, different impairments should be evaluated based on user opinions, with the aim of analyzing the impact in the perceived quality. In this work, we carried out a subjective evaluation of different types of impairments with different types of contents, including news, cartoons, sports and action movies. A total of 100 individuals, between the ages of 20 and 68, participated in the subjective study. Results show that short-term rebuffering events negatively affect the quality of experience (QoE) and that desynchronization between audio and video is the least annoying impairment. Moreover, we found that the content type determines the subjective results according to the impairment present during the playback.

Bibliografía

- [1] J. Hoebeke, I. Moerman, B. Dhoedt, y P. Demeester, «An overview of mobile ad hoc networks: applications and challenges», *Journal-Communications Network*, vol. 3, n.º 3, pp. 60–66, 2004.
- [2] I. Sodagar, «The mpeg-dash standard for multimedia streaming over the internet», *IEEE MultiMedia*, n.º 4, pp. 62–67, 2011.
- [3] W. E. Deming y J. N. Medina, *Calidad, productividad y competitividad: la salida de la crisis*. Ediciones Díaz de Santos, 1989.
- [4] R. García, X. G. Pañeda, D. Melendi, y V. Garcia, «Probabilistic analysis and interdependence discovery in the user interactions of a video news on demand service», *Computer Networks*, vol. 53, n.º 12, pp. 2038–2049, 2009.
- [5] S. Bhatt, R. Fujimoto, A. Ogielski, y K. Perumalla, «Parallel simulation techniques for large-scale networks», *Communications Magazine, IEEE*, vol. 36, n.º 8, pp. 42–47, 1998.
- [6] K. Pawlikowski, H. D. . Jeong, y J. S.R Lee, «On credibility of simulation studies of telecommunication networks», *IEEE Communications Magazine*, vol. 40, n.º 1, pp. 132–139, 2002.
- [7] K. Pawlikowski, «Simulation studies of telecommunication networks and their credibility», en *Proceedings. 13th European Simulation Multiconference, ESM*, 1999, vol. 99, pp. 349–355.
- [8] K. Pawlikowski, «Do not trust all simulation studies of telecommunication networks», en *Information Networking*, 2003, pp. 899–908.
- [9] A. Rachedi, S. Lohier, S. Cherrier, y I. Salhi, «Wireless network simulators relevance compared to a real testbed in outdoor and indoor environments», *International Journal of Autonomous and Adaptive Communications Systems*, vol. 5, n.º 1, pp. 88–101, ene. 2012.
- [10] S. Kristiansen, T. Plagemann, y V. Goebel, «Extending network simulators with communication software execution models», en *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*, 2013, pp. 1–10.
- [11] D. Melendi, R. García, X. G. Pañeda, y V. García, «Modelling and simulation of a real Internet radio service», *Journal of Simulation*, vol. 5, n.º 2, pp. 111–122, may 2011.
- [12] K. Salah, J. Hamodi, Z. A. Baig, y F. Al-Haidari, «Video-on-Demand (VoD) deployment over hospitality networks», *International Journal of Network Management*, vol. 22, n.º 1, pp. 65–80, 2012.
- [13] J. Liu, «A Primer for Real-Time Simulation of Large-Scale Networks», en *41st Annual Simulation Symposium (anss-41 2008)*, Ottawa, Canada, 2008, pp. 85–94.
- [14] D. Mahrenholz y S. Ivanov, «Real-Time Network Emulation with ns-2», en *Eighth IEEE International Symposium on Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004*, 2004, pp. 29–36.

-
- [15] L. Carter, J. Dyal, S. Doshi, y R. Bagrodia, «A hardware-in-the-loop (HWIL) network simulator for analysis and evaluation of large-scale military wireless communication systems», en *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pp. 1–8.
- [16] J. Mohorko, M. Fras, y Ž. Čučej, «Real video stream transmission over simulated wireless link», en *Advanced Technologies for Communications, 2008. ATC 2008. International Conference on*, 2009, pp. 231–234.
- [17] S. Kristiansen y T. Plagemann, «Accuracy and scalability of ns-2's distributed emulation extension», *SIMULATION*, vol. 87, n.º 1-2, pp. 45 -65, ene. 2011.
- [18] A. Alvarez, R. Orea, S. Cabrero, X. G. Pañeda, R. García, y D. Melendi, «Limitations of network emulation with single-machine and distributed ns-3», en *Proceedings of the 3rd International Conference on Simulation Tools and Techniques, Torremolinos, Spain*, 2010, pp. 1-9.
- [19] K. Yocum, E. Eade, J. Degeys, D. Becker, J. Chase, y A. Vahdat, «Toward scaling network emulation using topology partitioning», en *Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2003. MASCOTS 2003. 11th IEEE/ACM International Symposium on*, 2003, pp. 242–245.
- [20] C. Kiddle, R. Simmonds, C. Williamson, y B. Unger, «Hybrid packet/fluid flow network simulation», en *Parallel and Distributed Simulation, 2003.(PADS 2003). Proceedings. Seventeenth Workshop on*, 2003, pp. 143–152.
- [21] E. Weingärtner, F. Schmidt, T. Heer, y K. Wehrle, «Synchronized network emulation: matching prototypes with complex simulations», *SIGMETRICS Perform. Eval. Rev.*, vol. 36, pp. 58-63, ago. 2008.
- [22] A. Alvarez, R. Garcia, S. Cabrero, X. G. Paneda, D. Melendi, y R. Orea, «In pursuit of massive service emulation: a methodology for testbed building», *IEEE Communications Magazine*, vol. 49, n.º 9, pp. 162 -168, sep. 2011.
- [23] L. Peterson y T. Roscoe, «The design principles of PlanetLab», *ACM SIGOPS Operating Systems Review*, vol. 40, n.º 1, pp. 11–16, 2006.
- [24] J. S. Turner, M. Wilson, C. Wiseman, D. Zar, P. Crowley, J. DeHart, A. Freestone, B. Heller, F. Kuhns, S. Kumar, J. Lockwood, y J. Lu, «Supercharging planetlab», *SIGCOMM Comput. Commun. Rev.*, vol. 37, n.º 4, p. 85, oct. 2007.
- [25] C. Kiddle, B. Unger, y R. Simmonds, «Advances in Network Emulation», *Annual Review of Network Management and Security*, pp. 57–66, 2006.
- [26] S. Hemminger, «Netem-emulating real networks in the lab», en *Proc. Linux Conference Australia*, 2005.
- [27] M. Zee y M. Mikuc, «Real-time ip network simulation at gigabit data rates», en *Proceedings of the 7th International Conference on Telecommunications, Contel.*, 2003, vol. 1, pp. 235–242.
- [28] E. Weingärtner, H. Vom Lehn, y K. Wehrle, «A performance comparison of recent network simulators», en *Communications, 2009. ICC'09. IEEE International Conference on*, 2009, pp. 1–5.
- [29] N. I. Sarkar, K. Nisar, y L. Babbage, «Performance Studies on Campus-Wide Focus on FTP, Video and VoIP Ethernet Network»:., *International Journal of Advanced Pervasive and Ubiquitous Computing*, vol. 4, n.º 1, 2012.
- [30] B. Palaniappan, «Streaming Real-time Video Using Hybrid Transport Layer Protocols in Sim-real-sim Environment», *International Journal of Computer Science*, vol. 2, n.º 01, 2013.
- [31] S. C. Muthukumar, X. Li, C. Liu, J. B. Kopena, M. Oprea, R. Correa, B. T. Loo, y P. Basu, «RapidMesh: Declarative Toolkit for Rapid Experimentation of Wireless Mesh Networks», en *Proceedings of the 4th ACM International Workshop on*

- Experimental Evaluation and Characterization*, New York, NY, USA, 2009, pp. 1–10.
- [32] M. Varela, L. Skorin-Kapov, K. De Moor, y P. Reichl, «QoE – Defining Service Quality from the User’s Perspective», en *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, John Wiley & Sons, 2014.
- [33] H.-J. Lee, M.-S. Kim, J. W. Hong, y G.-H. Lee, «QoS parameters to network performance metrics mapping for SLA monitoring», *KNOM Rev*, vol. 5, n.º 2, 2002.
- [34] P. Pérez, J. Macías, J. J. Ruiz, y N. García, «Effect of packet loss in video quality of experience», *Bell Labs Technical Journal*, vol. 16, n.º 1, pp. 91–104, 2011.
- [35] A. Canovas, M. Garcia, J. Lloret, M. Atenas, y R. Rizo, «Improving IPTV QoE taking the suitable MPEG-2/MPEG-4 Quantizer based on Jitter, Delay and lost packets measurements», *International Journal On Advances in Telecommunications*, vol. 3, n.º 3 and 4, pp. 16-26, abr. 2011.
- [36] C. Perkins, *RTP: audio and video for the Internet*. Addison-Wesley, 2003.
- [37] T.-H. Truong, T.-H. Nguyen, y H.-T. Nguyen, «On Relationship between Quality of Experience and Quality of Service Metrics for IMS-Based IPTV Networks», en *2012 IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, 2012, pp. 1-6.
- [38] «Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison», *Broadcasting, IEEE Transactions on*, vol. 57, n.º 2, pp. 165-182, jun. 2011.
- [39] A. K. Moorthy, L. K. Choi, G. de Veciana, y A. C. Bovik, «Subjective analysis of video quality on mobile devices», *VPQM*, 2012.
- [40] Zhou Wang, A. C. Bovik, H. R. Sheikh, y E. P. Simoncelli, «Image quality assessment: from error visibility to structural similarity», *Image Processing, IEEE Transactions on*, vol. 13, n.º 4, pp. 600-612, 2004.
- [41] M. Rosenberg, «Eye, Brain, and Vision», *JAMA: The Journal of the American Medical Association*, vol. 260, n.º 24, p. 3677, 1988.
- [42] J. Klaue, B. Rathke, y A. Wolisz, «Evalvid—A framework for video transmission and quality evaluation», *Computer Performance*, pp. 255–272, 2003.
- [43] Zhou Wang y A. C. Bovik, «Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures», *Signal Processing Magazine, IEEE*, vol. 26, n.º 1, pp. 98-117, 2009.
- [44] J. Korhonen, U. Reiter, y Junyong You, «On the relationship between perceptual impact of source and channel distortions in video sequences», en *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 2933-2936.
- [45] Y. F. Ou, Z. Ma, y Y. Wang, «A novel quality metric for compressed video considering both frame rate and quantization artifacts», *Proc. of Intl. Workshop Video Processing and Quality Metrics for Consumer (VPQM)*, Scottsdale, AZ, ene. 2009.
- [46] E. Mendi, M. Milanova, Y. Zhou, y J. Talburt, «Objective video quality assessment for tracking moving objects from video sequences», en *Proceedings of the 9th WSEAS international conference on Signal processing, robotics and automation*, 2010, pp. 121–126.
- [47] Shu Tao, J. Apostolopoulos, y R. Guerin, «Real-Time Monitoring of Video Quality in IP Networks», *Networking, IEEE/ACM Transactions on*, vol. 16, n.º 5, pp. 1052-1065, 2008.

-
- [48] Z. Wang, L. Lu, y A. C. Bovik, «Video quality assessment based on structural distortion measurement», *Signal processing: Image communication*, vol. 19, n.º 2, pp. 121–132, 2004.
- [49] A. Hore y D. Ziou, «Image Quality Metrics: PSNR vs. SSIM», en *2010 20th International Conference on Pattern Recognition (ICPR)*, 2010, pp. 2366-2369.
- [50] M. Uhrina, J. Hlubik, y M. Vaculik, «Correlation between objective and subjective methods used for video quality evaluation», en *ELEKTRO, 2012*, 2012, pp. 103-108.
- [51] D. Jiménez Bermejo, «High definition video quality assessment metric built upon full reference ratios», Tesis Doctoral, E.T.S.I. Telecomunicación (UPM), 2012.
- [52] K. Seshadrinathan y A. C. Bovik, «Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos», *Image Processing, IEEE Transactions on*, vol. 19, n.º 2, pp. 335 -350, feb. 2010.
- [53] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, y L. K. Cormack, «Study of subjective and objective quality assessment of video», *Image Processing, IEEE Transactions on*, vol. 19, n.º 6, pp. 1427–1441, 2010.
- [54] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, y A. Perkis, «Perceptual-based quality assessment for audio-visual services: A survey», *Signal Processing: Image Communication*, vol. 25, n.º 7, pp. 482-501, ago. 2010.
- [55] T. Zinner, O. Abboud, O. Hohlfeld, T. Hossfeld, y P. Tran-Gia, «Towards QoE Management for Scalable Video Streaming», en *21th ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE*, 2010.
- [56] Hosik Sohn, Hana Yoo, W. De Neve, Cheon Seog Kim, y Yong Man Ro, «Full-Reference Video Quality Metric for Fully Scalable and Mobile SVC Content», *IEEE Transactions on Broadcasting*, vol. 56, n.º 3, pp. 269-280, sep. 2010.
- [57] R. Skupin, C. Hellge, T. Schierl, y T. Wiegand, «Packet level video quality evaluation of extensive H.264/AVC and SVC transmission simulation», *Journal of Internet Services and Applications*, vol. 2, n.º 2, pp. 129-138, jun. 2011.
- [58] Y. Pitrey, M. Barkowsky, P. Le Callet, y R. Pèpion, «Evaluation of MPEG4-SVC for QoE protection in the context of transmission errors», en *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2010, vol. 7798, p. 37.
- [59] Q. Ma, F. Wu, y M. T. Sun, «Error concealment for spatially Scalable Video Coding using hallucination», en *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, 2009, pp. 129–132.
- [60] M. Brandas, M. G. Martini, M. Uitto, y J. Vehkaperä, «Quality assessment and error concealment for SVC transmission over unreliable channels», en *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, 2011, pp. 1–6.
- [61] Tien Anh Le, Hang Nguyen, y Hongguang Zhang, «EvalSVC — An evaluation platform for scalable video coding transmission», en *2010 IEEE 14th International Symposium on Consumer Electronics (ISCE)*, 2010, pp. 1-6.
- [62] A. Detti, G. Bianchi, C. Pisa, F. S. Proto, P. Loretì, W. Kellerer, S. Thakolsri, y J. Widmer, «SVEF: an open-source experimental evaluation framework for H.264 scalable video streaming», en *IEEE Symposium on Computers and Communications, 2009. ISCC 2009*, 2009, pp. 36-41.
- [63] K. D. Singh, A. Ksentini, y B. Marienval, «Quality of Experience Measurement Tool for SVC Video Coding», en *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1-5.
- [64] Q. Huynh-Thu, M. N. Garcia, F. Speranza, P. Corriveau, y A. Raake, «Study of rating scales for subjective quality assessment of high-definition video», *Broadcasting, IEEE Transactions on*, n.º 99, pp. 1–1, 2011.

-
- [65] K. Brunnstrom, D. Hands, F. Speranza, y A. Webster, «VQeg validation and ITU standardization of objective perceptual video quality metrics [Standards in a Nutshell]», *IEEE Signal Processing Magazine*, vol. 26, n.º 3, pp. 96-101, may 2009.
- [66] Video Quality Experts Group (VQEG), «Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase I, (FRTV-I)». 2000.
- [67] Video Quality Experts Group (VQEG), «Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II, VQEG FRTV-II». .
- [68] T. Shanableh, «No-Reference PSNR Identification of MPEG Video Using Spectral Regression and Reduced Model Polynomial Networks», *Signal Processing Letters, IEEE*, vol. 17, n.º 8, pp. 735–738, 2010.
- [69] M. Naccari, M. Tagliasacchi, y S. Tubaro, «No-reference video quality monitoring for H. 264/AVC coded video», *Multimedia, IEEE Transactions on*, vol. 11, n.º 5, pp. 932–946, 2009.
- [70] Video Quality Experts Group (VQEG), «Final report from the Video Quality Experts Group on the validation of objective models Reduced Reference (RR) and No Reference (NR) that predict the quality of standard definition Television, VQEG RRNR-TV». .
- [71] Video Quality Experts Group (VQEG), «Final report from the Video Quality Experts Group on the validation of subjective models of multimedia quality assesment, phase I, VQEG MM-I». .
- [72] Video Quality Experts Group (VQEG), «Final report from the Video Quality Experts Group on the validation of objective models of full-reference video quality assesment of High Definition Television (HDTV), VQEG HDTV». .
- [73] R. Eg, C. Griwodz, P. Halvorsen, y D. Behne, «Audiovisual robustness: exploring perceptual tolerance to asynchrony and quality distortion», *Multimed Tools Appl*, vol. 74, n.º 2, pp. 345-365, ago. 2014.
- [74] M. A. Usman, M. R. Usman, y S. Y. Shin, «The Impact of Temporal Impairment on Quality of Experience (QoE) in Video Streaming: A No Reference (NR) Subjective and Objective Study».
- [75] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, y C. Lorentzen, «Initial delay vs. interruptions: Between the devil and the deep blue sea», en *2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2012, pp. 1-6.
- [76] R. K. P. Mok, E. W. W. Chan, y R. K. C. Chang, «Measuring the quality of experience of HTTP video streaming», en *2011 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2011, pp. 485-492.
- [77] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, y L. Martens, «Quantifying the Influence of Rebuffering Interruptions on the User's Quality of Experience During Mobile Video Watching», *IEEE Transactions on Broadcasting*, vol. 59, n.º 1, pp. 47-61, mar. 2013.
- [78] D. Ghadiyaram, A. C. Bovik, H. Yeganeh, R. Kordasiewicz, y M. Gallant, «Study of the effects of stalling events on the quality of experience of mobile streaming videos», en *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, 2014, pp. 989–993.
- [79] J. Liu, Y. Geng, D. Wang, W. Li, y X. Qiu, «An objective multi-factor QoE evaluation based on content classification for H.264/AVC encoded video», en *2013 IEEE Symposium on Computers and Communications (ISCC)*, 2013, pp. 000137-000142.

-
- [80] T. Zinner, O. Hohlfeld, O. Abboud, y T. Hossfeld, «Impact of frame rate and resolution on objective QoE metrics», en *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, 2010, pp. 29-34.
- [81] P. Kortum y M. Sullivan, «The effect of content desirability on subjective video quality ratings», *Human factors: the journal of the human factors and ergonomics society*, vol. 52, n.º 1, pp. 105–118, 2010.
- [82] T. Hosfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, y M. Fiedler, «The memory effect and its implications on Web QoE modeling», en *Teletraffic Congress (ITC), 2011 23rd International*, 2011, pp. 103-110.
- [83] J. S. Lee, F. De Simone, T. Ebrahimi, N. Ramzan, y E. Izquierdo, «Quality assessment of multidimensional video scalability», *Communications Magazine, IEEE*, vol. 50, n.º 4, pp. 38–46, 2012.
- [84] J. S. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, y T. Ebrahimi, «Subjective evaluation of scalable video coding for content distribution», en *Proceedings of the international conference on Multimedia*, 2010, pp. 65–72.
- [85] T. Oelbaum, H. Schwarz, M. Wien, y T. Wiegand, «Subjective performance evaluation of the SVC extension of H.264/AVC», en *15th IEEE International Conference on Image Processing, 2008. ICIP 2008*, 2008, pp. 2772 -2775.
- [86] F. Niedermeier, M. Niedermeier, y H. Kosch, «Quality assessment of the MPEG-4 scalable video CODEC», en *Image Analysis and Processing-ICIAP 2009*, 2009, pp. 297-306.
- [87] Y. Pitrey, M. Barkowsky, P. Le Callet, y R. Pepion, «Subjective quality assessment of MPEG-4 Scalable Video Coding in a mobile scenario», en *2010 2nd European Workshop on Visual Information Processing (EUVIP)*, 2010, pp. 86 -91.
- [88] A. Eichhorn y P. Ni, «Pick your layers wisely - a quality assessment of H.264 scalable video coding for mobile devices», en *Proceedings of the 2009 IEEE international conference on Communications*, Piscataway, NJ, USA, 2009, pp. 5446–5451.
- [89] Y. Pitrey, U. Engelke, M. Barkowsky, R. Pepion, y P. Le Callet, «Subjective quality of SVC-coded videos with different error-patterns concealed using spatial scalability», en *Visual Information Processing (EUVIP), 2011 3rd European Workshop on*, 2011, pp. 180 -185.
- [90] R. K. P. Mok, X. Luo, E. W. W. Chan, y R. K. C. Chang, «QDASH: A QoE-aware DASH System», en *Proceedings of the 3rd Multimedia Systems Conference*, New York, NY, USA, 2012, pp. 11–22.
- [91] J.-S. Lee, F. De Simone, y T. Ebrahimi, «Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding», *IEEE Transactions on Multimedia*, vol. 13, n.º 5, pp. 882 -893, oct. 2011.
- [92] N. Staelens, W. Van den Broeck, Y. Pitrey, B. Vermeulen, y P. Demeester, «Lessons learned during real-life QoE assessment», en *10th European Conference on Interactive TV and Video (Euro ITV-2012)*, 2012, pp. 1–4.
- [93] Q. Xu, J. Xiong, Q. Huang, y Y. Yao, «Robust evaluation for quality of experience in crowdsourcing», en *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 43–52.
- [94] B. Gardlo, M. Ries, y T. Hossfeld, «Impact of screening technique on crowdsourcing QoE assessments», en *Radioelektronika (RADIOELEKTRONIKA), 2012 22nd International Conference*, 2012, pp. 1–4.

-
- [95] R. Shannon y J. D. Johannes, «Systems Simulation: The Art and Science», *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-6, n.º 10, pp. 723-724, oct. 1976.
- [96] C. Kiddle, «Scalable network emulation», Tesis Doctoral, University of Calgary, 2004.
- [97] Beuran, Razvan, *Introduction to Network Emulation*, 1.^a ed. Singapore: Pan Stanford Publishing, 2012.
- [98] R. Sherwood, G. Gibb, K.-K. Yap, G. Appenzeller, M. Casado, N. McKeown, y G. M. Parulkar, «Can the production network be the testbed?», en *OSDI*, 2010, vol. 10, pp. 1–6.
- [99] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, y M. Bowman, «Planetlab: an overlay testbed for broad-coverage services», *ACM SIGCOMM Computer Communication Review*, vol. 33, n.º 3, pp. 3–12, 2003.
- [100] A. Nayyar y R. Singh, «A Comprehensive Review of Simulation Tools for Wireless Sensor Networks (WSNs)», *Journal of Wireless Networking and Communications*, vol. 5, n.º 1, pp. 19–47, 2015.
- [101] J. Lessmann, P. Janacik, L. Lachev, y D. Orfanus, «Comparative Study of Wireless Network Simulators», en *Seventh International Conference on Networking, 2008. ICN 2008*, 2008, pp. 517-523.
- [102] D. Mahrenholz y S. Ivanov, «Adjusting the ns-2 Emulation Mode to a Live Network», en *Kommunikation in Verteilten Systemen (KIVS)*, P. Müller, R. Gotzhein, y J. B. Schmitt, Eds. Springer Berlin Heidelberg, 2005, pp. 205-217.
- [103] «NEMAN: a network emulator for mobile ad-hoc networks», vol. 1, pp. 155- 161, jun. 2005.
- [104] «Software, Patches and Docs». [En línea]. Disponible en: <http://www.ssi.bg/~ja/#loop>. [Accedido: 23-ago-2015].
- [105] B. Hubert, T. Graf, G. Maxwell, R. van Mook, M. van Oosterhout, P. Schroeder, J. Spaans, y P. Larroy, «Linux advanced routing & traffic control», en *Ottawa Linux Symposium*, 2002, p. 213.
- [106] P. Arlos, *On the quality of computer network measurements*. Blekinge Institute of Technology, 2005.
- [107] G. Seguin, «Multi-core parallelism for ns-3 simulator», *INRIA Sophia-Antipolis, Tech. Rep*, vol. 106, p. 110, 2009.
- [108] A. Álvarez, L. Pozueco, X. G. Pañeda, R. García, D. Melendi, S. Cabrero, y G. D. Orueta, «Limits for the real-time simulation of video services over commodity hardware», *J of Sim*, jun. 2015.
- [109] D. Bullock y A. Catarella, «A real-time simulation environment for evaluating traffic signal systems», *Transportation Research Record: Journal of the Transportation Research Board*, n.º 1634, pp. 130–135, 1998.
- [110] M. Fras y J. Mohorko, «Real time “system-in-the-loop” simulation of tactical networks», en *Software, Telecommunications and Computer Networks. SoftCOM 2008. 16th International Conference on*, 2008, pp. 105–108.
- [111] R. Martinez, W. Wu, K. McNeill, J. Deal, T. Haynes, y D. Bradford, «Hardware and software-in-the-loop techniques using the OPNET modeling tool for JTRS developmental testing», en *2003 IEEE Military Communications Conference, 2003. MILCOM '03*, 2003, vol. 1, pp. 469-474 Vol.1.
- [112] Z. Lu y H. Yang, *Unlocking the Power of OPNET Modeler*. Cambridge University Press, 2012.

-
- [113] S. Bhatt, R. Fujimoto, A. Ogielski, y K. Perumalla, «Parallel simulation techniques for large-scale networks», *Communications Magazine, IEEE*, vol. 36, n.º 8, pp. 42–47, 1998.
- [114] S. Hemminger, «Network emulation with NetEm», en *Proceedings of the 6th Australia's National Linux Conference (LCA)*, Canberra (Australia), 2005.
- [115] García Gutiérrez Alberto Eloy, «Contribución al desarrollo de herramientas estratégicas para el diseño, dimensionado y evaluación de redes de telecomunicación de banda ancha», Tesis Doctoral, Universidad de Cantabria, 2009.
- [116] S. J. Vaughan-Nichols, «New Approach to Virtualization Is a Lightweight», *Computer*, vol. 39, n.º 11, pp. 12-14, nov. 2006.
- [117] M. S. Hasan, H. Yu, A. Carrington, y T. C. Yang, «Co-simulation of wireless networked control systems over mobile ad hoc network using SIMULINK and OPNET», *Communications, IET*, vol. 3, n.º 8, pp. 1297-1310, 2009.
- [118] G. F. Riley, M. H. Ammar, R. M. Fujimoto, A. Park, K. Perumalla, y D. Xu, «A federated approach to distributed network simulation», *ACM Trans. Model. Comput. Simul.*, vol. 14, n.º 2, pp. 116-148, abr. 2004.
- [119] R. Bradford, R. Simmonds, y B. Unger, «A parallel discrete event IP network emulator», en *Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2000. Proceedings. 8th International Symposium on*, 2000, pp. 315–322.
- [120] A. Fraga, L. Pozueco, X. Garcia Pañeda, R. García, D. Melendi, y S. Cabrero, «A non-intrusive estimation for high-quality Internet TV services», *Multimedia Tools and Applications*, pp. 1–20, 2009.
- [121] S. Winkler y P. Mohandas, «The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics», *Broadcasting, IEEE Transactions on*, vol. 54, n.º 3, pp. 660-668, 2008.
- [122] Tao Liu, Yao Wang, J. M. Boyce, Zhenyu Wu, y Hua Yang, «Subjective Quality Evaluation of Decoded Video in the Presence of Packet Losses», en *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, vol. 1, pp. I-1125-I-1128.
- [123] Nan Li, Bo Yan, y Guanling Chen, «Measurement study on wireless camera networks», en *Distributed Smart Cameras, 2008. ICDS-C 2008. Second ACM/IEEE International Conference on*, 2008, pp. 1-10.
- [124] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, y T. Ebrahimi, «Subjective assessment of H. 264/AVC video sequences transmitted over a noisy channel», en *Quality of Multimedia Experience, 2009. QoMEX 2009. International Workshop on*, 2009, pp. 204–209.
- [125] A. Lie y J. Klaue, «Evalvid-RA: trace driven simulation of rate adaptive MPEG-4 VBR video», *Multimedia Systems*, vol. 14, n.º 1, pp. 33–50, 2008.
- [126] P. Casas, P. Belzarena, y S. Vaton, «End-2-end evaluation of ip multimedia services, a user perceived quality of service approach», en *Proceedings of the 18th ITC Specialist Seminar on Quality of Experience*, 2008, pp. 13–23.
- [127] P. Romaniak, M. Mu, A. Mauthe, S. D'Antonio, y M. Leszczuk, «Framework for the Integrated Video Quality Assessment», *Multimedia Tools and Applications*, vol. 61, n.º 3, pp. 787-817, 2012.
- [128] M. Zink, «Scalable Internet Video-on-Demand Systems», Tesis Doctoral, TU Darmstadt, 2003.

-
- [129] L. Pozueco Álvarez, «Diseño y evaluación de sistemas de estimación de ancho de banda disponible para servicios adaptativos de vídeo streaming», Tesis Doctoral, 2015.
- [130] M. Azni, R. Pyndiah, K. Mokrani, y F. Guilloud, «Joint Source and Channel Coding for Real Time H.264 coded Video Transmission Using Mobile Hand-held Devices», *IETE Journal of Research*, vol. 55, n.º 4, pp. 154-161, jul. 2009.
- [131] L. Pozueco, X. G. Pañeda, R. García, D. Melendi, y S. Cabrero, «Adaptable system based on Scalable Video Coding for high-quality video service», *Computers & Electrical Engineering*, 2013.
- [132] R. Kuschnig, I. Kofler, M. Ransburg, y H. Hellwagner, «Design options and comparison of in-network H.264/SVC adaptation», *Journal of Visual Communication and Image Representation*, vol. 19, n.º 8, pp. 529-542, dic. 2008.
- [133] M. Nuutinen, O. Orenius, T. Saamanen, y P. Oittinen, «A framework for measuring sharpness in natural images captured by digital cameras based on reference image and local areas», *Eurasip Journal on Image and Video Processing*, 2011.
- [134] N. M. De Simone Francesca, T. Marco, D. Frederic, T. Stefano, y E. Touradj, «Subjective Quality Assessment of H. 264/AVC Video Streaming with Packet Losses», *EURASIP Journal on Image and Video Processing*, vol. 2011.
- [135] T. Zinner, O. Abboud, O. Hohlfeld, T. Hossfeld, y P. Tran-Gia, «Towards QoE Management for Scalable Video Streaming», en *21th ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE*.
- [136] Y.-H. Kim, J. Park, y J.-W. Kim, «Multithreading Architecture for Real-Time MPEG-4 AVC/H.264 SVC Decoder», en *Advances in Image and Video Technology*, vol. 7088, Y.-S. Ho, Ed. Springer Berlin / Heidelberg, 2012, pp. 13-24.
- [137] A. Eleftheriadis, S. Wenger, T. Schierl, y Y. K. Wang, «RTP Payload Format for Scalable Video Coding», IETF, 2011.
- [138] Y. Chen, K. Xie, F. Zhang, P. Pandit, y J. Boyce, «Frame loss error concealment for SVC», *Journal of Zhejiang University-Science A*, vol. 7, n.º 5, pp. 677–683, 2006.
- [139] P. Lambert, P. Debevere, J. De Cock, J. F. Macq, N. Degrande, D. De Vleeschauwer, y R. Van de Walle, «Real-time error concealing bitstream adaptation methods for SVC in IPTV systems», *Journal of Real-Time Image Processing*, vol. 4, n.º 1, pp. 79–90, 2009.
- [140] C. S. Park, T. S. Wang, y S. J. Ko, «Error concealment using inter-layer correlation for scalable video coding», *ETRI journal*, vol. 29, n.º 3, pp. 390–392, 2007.
- [141] M. Stoufs, A. Munteanu, J. Cornelis, y P. Schelkens, «Error concealment for the scalable extension of H. 264/MPEG-4 AVC», en *Proc. Picture Coding Symp.(PCS2007)*, 2007, p. 1.
- [142] M. Blestel y M. Raulet, «Open SVC decoder: a flexible SVC library», en *Proceedings of the international conference on Multimedia*, 2010, pp. 1463–1466.
- [143] C. Perkins y T. Schierl, «Rapid Synchronisation of RTP Flows», IETF.
- [144] J. Chen, D. Wu, H. Gao, y J. Y. Tham, «A novel scheme of generating streamable MP4 files for scalable video coding», en *Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on*, 2009, pp. 1–5.
- [145] H. Gao, J. Y. Tham, K. H. Goh, W. S. Lee, y K. S. Aw, «MP4 File Creator for SVC Adaptive Video Streaming», en *Internet Technology and Applications, 2010 International Conference on*, 2010, pp. 1–4.
- [146] Alberto Alvarez, Sergio Cabrero, Xabiel García Pañeda, Roberto García, David Melendi, Rafael Orea, «A Flexible QoE Framework for Video Streaming Services», en *Workshop on Ubiquitous Computing and Networks*, 2011.

-
- [147] U. Reiter, J. Korhonen, y J. You, «Comparing apples and oranges: assessment of the relative video quality in the presence of different types of distortions», *EURASIP Journal on Image and Video Processing*, vol. 2011, n.º 1, p. 8, sep. 2011.
- [148] J. Navarro Salmerón, «Contribución a los Modelos de Estimación de la Calidad Percibida en Servicios de Vídeo sobre Internet mediante Parámetros Objetivos», Tesis Doctoral, E.T.S.I. Telecomunicación (UPM), 2015.
- [149] P. de la Cruz Ramos, «Contribución a los Modelos y Metodologías para la Estimación de la Calidad Percibida por los Usuarios (QoE) a partir de Parámetros de Calidad de Red/Servicio (QoS) en Servicios Convergentes Multimedia (Triple-Play)», Tesis Doctoral, E.T.S.I. Telecomunicación (UPM), 2012.
- [150] F. De Simone, L. Goldmann, J.-S. Lee, y T. Ebrahimi, «Towards high efficiency video coding: Subjective evaluation of potential coding technologies», *Journal of Visual Communication and Image Representation*, vol. 22, n.º 8, pp. 734–748, 2011.
- [151] S. Winkler, «Analysis of public image and video databases for quality assessment», *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, n.º 6, pp. 616–625, 2012.
- [152] P. Paudyal, F. Battisti, y M. Carli, «A Study On The Effects Of Quality Of Service Parameters On Perceived Video Quality», presentado en Procs. of 5th European Workshop on Visual Information Processing, EUVIP 2014, 2014.
- [153] International Telecommunication Union, «Recommendation ITU-R BT.500-12: Methodology for the subjective assessment of the quality of television pictures», vol. 12, 2009.
- [154] M. J. Crawley, *The R book*. John Wiley & Sons, 2012.
- [155] Y. Qi y M. Dai, «The Effect of Frame Freezing and Frame Skipping on Video Quality», en *International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP '06*, 2006, pp. 423-426.
- [156] G. Muntean, P. Perry, y L. Murphy, «Subjective Assessment of the Quality-Oriented Adaptive Scheme», *IEEE Transactions on Broadcasting*, vol. 51, n.º 3, pp. 276-286, sep. 2005.
- [157] C. Kim, A. Kondoz, y X. Shi, «Investigation into spatial audio quality of experience in the presence of accompanying video cues with spatial mismatch», en *2013 IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 1192-1197.
- [158] T. N. Minhas y M. Fiedler, «Impact of disturbance locations on video quality of experience», *Quality of Experience for Multimedia Content Sharing, EuroITV*, 2011.
- [159] Christine Bagwell, «Multimedia on the Internet: Streaming Media West '99», *Library Hi Tech News*, vol. 17, n.º 4, abr. 2000.
- [160] L. Cruvinel y T. Vazão, «Improving performance for multimedia traffic with distributed dynamic QoS adaptation», *Computer Communications*, vol. 34, n.º 10, pp. 1222-1234, jul. 2011.
- [161] N. Cranley, P. Perry, y L. Murphy, «User perception of adapting video quality», *Int. J. Hum.-Comput. Stud.*, vol. 64, n.º 8, pp. 637–647, ago. 2006.
- [162] J. Rückert, O. Abboud, T. Zinner, R. Steinmetz, y D. Hausheer, «Quality Adaptation in P2P Video Streaming Based on Objective QoE Metrics», *Networking 2012*, vol. 7290, pp. 1-14, 2012.
- [163] L. Caviglione y C. Cervellera, «Design, optimization and performance evaluation of a content distribution overlay for streaming», *Computer Communications*, vol. 34, n.º 12, pp. 1497-1509, ago. 2011.

-
- [164] P. Juluri, «Measurement And Improvement of Quality-of-Experience For Online Video Streaming Services», Tesis Doctoral, Jawaharlal Nehru Technological University, 2015.
- [165] A. Alvarez, L. Pozueco, S. Cabrero, X. G. Pañeda, R. Garcia, D. Melendi, y G. D. Orueta, «Subjective evaluation of critical success factors for a QoE aware adaptive system», *Computer Communications*, vol. 36, n.º 15-16, pp. 1608-1620, sep. 2013.
- [166] J. Lloret, A. Canovas, J. Tomas, y M. Atenas, «A network management algorithm and protocol for improving QoE in mobile IPTV», *Computer Communications*, 2012.
- [167] I. Kofler, R. Kuschnig, y H. Hellwagner, «In-network adaptation of H. 264/SVC for HD video streaming over 802.11 g networks», en *Proceedings of the 21st international workshop on Network and operating systems support for digital audio and video*, 2011, pp. 9–14.
- [168] D. T. Nguyen y J. Ostermann, «Congestion control for scalable video streaming using the scalability extension of H. 264/AVC», *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, n.º 2, pp. 246–253, 2007.
- [169] B. Görkemli y A. M. Tekalp, «Adaptation strategies for MGS scalable video streaming», *Signal Processing: Image Communication*, 2012.
- [170] ITU-T Rec. P.800, «Methods for subjective determination of transmission quality». 1996.
- [171] M. Vaahteranoksa y T. Vuori, *Threshold value for acceptable video quality using signal-to-noise ratio*. International Society for Optics and Photonics, 2007.
- [172] S. Salsano, F. Ludovici, y A. Ordine, «Definition of a general and intuitive loss model for packet networks and its implementation in the Netem module in the Linux kernel», Technical report.-University of Rome «Tor Vergata», 2009.–60 p, 2009.
- [173] M. Fiedler, T. Hossfeld, y P. Tran-Gia, «A generic quantitative relationship between quality of experience and quality of service», *IEEE Network*, vol. 24, n.º 2, pp. 36-41, mar. 2010.
- [174] M. Vranješ, S. Rimac-Drlje, y D. Žagar, «Objective video quality metrics», en *49th International Symposium ELMAR-2007 focused on Mobile Multimedia*, 2007.
- [175] Q. Huynh-Thu y M. Ghanbari, «Scope of validity of PSNR in image/video quality assessment», *Electronics Letters*, vol. 44, n.º 13, pp. 800-801, 2008.
- [176] S. Wulf y U. Zolzer, «About the imperfection of objective quality metrics on high-definition video content», en *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, 2013, pp. 384-389.
- [177] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, y S. Wolf, «Objective video quality assessment system based on human perception», presentado en IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology, 1993, pp. 15-26.
- [178] L. Zhou, S. Wu, X. Shi, H. Jin, y J. Zhou, «Guaranteeing QoS of media-based applications in virtualized environment», *New Review of Hypermedia and Multimedia*, vol. 19, n.º 3-4, pp. 217-233, dic. 2013.

Glosario de términos

Símbolo	Término
ACR	Absolute Category Rating
ANOVA	Analysis of Variance
API	Application Programming Interface
AU	Access Unit
AVC	Advanced Video Codec
BER	Bit Error Rate
BPMN	Business Process Model and Notation
CCR	Comparison Category Rating
CGS	Coarse Grain Scale
CMOS	Comparison Mean Opinion Score
cPSNR	Continuous Peak Signal-to-Noise Ratio
CPU	Central Processing Unit
CSMA	Carrier Sense Multiple Access
DASH	Dynamic Adaptive Streaming over HTTP
DES	Discrete Event Simulation
DMOS	Diferential Mean Opinion Score
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
DVD	Digital Versatile Disc
FC	Frame Copy
FR	Full Reference
GOP	Group of pictures
HD	High definition

Símbolo	Término
HSD	Honest Significant Difference
HTTP	Hipertext Transfer Protocol
HVS	Human Vision System
IP	Internet Protocol
IPv4	Internet Protocol version 4
ITU	International Telecommunication Institute
JSVM	Joint Scalable Video Model
LAN	Local Access Network
LXC	Linux Contaniers
MANET	Mobile ad hoc network
MOS	Men Opinion Score
MOVIE	MOtion-tuned Video Integrity Evaluation Index
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
NAL	Network Abstraction Layer
PC	Personal Computer
PDCA	Plan-do-check-act
PDES	Parallel Discrete Event Simulation
PLR	Packet Loss Ratio
PSNR	Peak Signal-to-Noise Ratio
PTS	Presentation TimeStamp
QoE	Quality of Experience
QoS	Quality of Service
RAM	Random Access Memory
RMSE	Root Mean Square Error
RR	Reduced Reference
RTCP-APP	Application-defined RTCP packet
RTP	Real Time Transport Protocol

Símbolo	Término
RTSP	Real Time Streaming Protocol
SI	Spatial Information
SITL	System in the Loop
SLA	Service Level Agreement
SR	Sender Report
SS	Single Stimulus
SSACJ	Single Stimulus Adjectival Categorical Judgement
SSCQE	Single Stimulus Continuous Quality Evaluation
SSIM	Structural Similarity Index
SVC	Scalable Video Coding
TCP	Transmission Control Protocol
TI	Temporal Information
TV	Television
UDP	User Datagram Protocol
UML	Unified Modeling Language
UTP	Unshielded twisted pair
VoIP	Voz IP
VQ	Video Quality
VQEG	Video Quality Experts Group
VQM	Video Quality Metric
XML	Extensible Markup Language

ANEXO

Curriculum Vitae

Alberto Álvarez González

Resumen

En la actualidad, compagino mi labor como profesional Investigador en la empresa con el desarrollo de mi carrera investigadora como miembro del grupo de Investigación de Sistemas de Distribución Multimedia (DMMS) de la Universidad de Oviedo.

Desde el curso 2011/2012 colaboro, en el marco del convenio de colaboración vigente entre las dos entidades, en proyectos de Investigación en el grupo DMMS y el máster en Investigación en Ingeniería Eléctrica, Electrónica y Control Industrial de la UNED.

Cuento con varias publicaciones en revistas indexadas en el JCR así como varios congresos nacionales e internacionales. En mi etapa como investigador en la empresa, he podido establecer contacto con los últimos estándares de la industria de las comunicaciones a nivel de redes vehiculares.

Formación académica

2001-2012 **Máster en Comunicación, Redes y Gestión de Contenidos**, UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

2003-2009 **Ingeniero de Telecomunicación**, UNIVERSIDAD DE OVIEDO. *Doble Intensificación, Telemática y Comunicaciones.*

Experiencia profesional

2012-Act Investigador, ADN Mobile Solutions, Departamento de I+D, *Desarrollo proyectos de innovación en el área de la monitorización eficiente de flotas de vehículos.*

2009-2012 Investigador en Grupo de Investigación de Sistemas. de Distribución Multimedia, Área de Ingeniería Telemática de la Universidad de Oviedo.

Certificaciones oficiales

2012 ITL V3 Foundation

2011 Certified Associate in Project Management (CAPM), CAPM N° 1436609

Idiomas

Inglés, Nivel Avanzado. Certificación CAE (Certificate in Advance English, Cambridge University)

Francés, Nivel medio. (Curso Francés II CaSa de las Lenguas, Universidad de Oviedo)

Formación Complementaria

2012 Curso en Gestión de Tecnologías de la Información Basado en ITIL v3. *21 horas. Gijón*

2011 Curso en Gestión de Proyectos orientado a la certificación PMI. *Colegio Oficial de Ingenieros de Telecomunicación, 50 PDUs*

2010 Curso en Dirección de Proyectos para la gestión de I+D+I. *Colegio Oficial de Ingenieros de Telecomunicación, 56 horas*

2009 Curso presencial de peritaciones *Colegio Oficial de Ingenieros de Telecomunicación, 56 horas*

Publicaciones

Subjective video quality evaluation of different content types under different impairments, Pozueco, L., Álvarez, A., García, X., García, R., Melendi, D., & Díaz, G. *New Review of Hypermedia and Multimedia*, 1-28. 2016

Limits for the real-time simulation of video services over commodity hardware
A Álvarez, L Pozueco, XG Pañeda, R García, D Melendi, S Cabrero, Gabriel Díaz Orueta
Journal of Simulation. 2015

Service To Manage The Efficient Driving Of Combustion Vehicle Fleets To Support ISO 50001, A Rionda, D Martinez, XG Pañeda, A Álvarez, D Arbesú, GD Orueta, J Pérez, AG Pañeda.
Latin America Transactions, IEEE (Revista IEEE America Latina) 13 (4), 1198-1204. 2015

A Framework to Measure and Estimate Video Quality in SVC Real-Time Adaptive Systems
A Álvarez, L Pozueco, S Cabrero, XG Pañeda, R García, D Melendi, Gabriel Díaz
International Journal of Business Data Communications and Networking (IJBDCN) 2014

Subjective evaluation of critical success factors for a QoE aware adaptive system
Alberto Alvarez, Laura Pozueco, Sergio Cabrero, Xabiel G. Pañeda, Roberto García, David Melendi y Gabriel Díaz Orueta, *Computer Communications*. 2013.

Evaluaciones subjetivas de servicios streaming adaptativos vs no-adaptativos
A. Álvarez, L. Pozueco, S. Cabrero, X. G. Pañeda, R. García, D. Melendi y G. Díaz,
XI Jornadas de Ingeniería Telemática, Granada, España.2013.

Automatic UMTS system resource dimensioning based on service traffic analysis,
P. Alonso, A. Alvarez, A. Alonso, B. Carro, J. M. Aguiar, and A. Sánchez, *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 323, Oct. 2012..

A method for the detection of QoS degradation in UMTS Networks, P. Alonso, A. Alvarez, A. Alonso, B. Carro, J. M. Aguiar, and A. Sánchez, *2nd International Workshop on Context-aware QoS Provisioning and Management for Emerging Networks, Applications and Services, Germany*, 2012..

Sistema de distribución de vídeo streaming adaptativo basado en codificación SVC, Laura Pozueco, Xabiel G. Pañeda, Alberto Álvarez, Sergio Cabrero, David Melendi, Roberto García. *Jitel* 2011.

A flexible QoE Framework for video streaming services, Alberto Álvarez, Sergio Cabrero, Xabiel G. Pañeda, Roberto García Fernandez, David Melendi, Rafael Orea. *2nd Workshop on Ubiquitous Computing and Networks (UbiCoNet, 2011)*. Houston, Estados Unidos de América..

In pursuit of massive service emulation: a methodology for testbed building,
Alberto Álvarez, Rafael Orea, Sergio Cabrero, Xabiel G. Pañeda, Roberto García, David Melendi. *IEEE Communications Magazine*. 2010.

Limitations of network emulation with single-machine and distributed ns-3, Alberto Álvarez, Rafael Orea, Sergio Cabrero, Xabiel G. Pañeda, Roberto García, David Melendi. *International ICST Conference on Simulation Tools and Techniques (SIMUTOOLS)* . Torremolinos, España. 2010.

Sistema autónomo para monitorización de incendios utilizando cámaras de vídeo sobre redes inalámbricas, Alberto Álvarez, Sergio Cabrero, Roberto García, Xabiel G. Pañeda, David Melendi. *Conferencia Iberoamericana de Telemática (CITA)*. Gijón, Asturias. 2009.
