

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE INGENIERÍA

INSTITUTO DE INGENIERÍA



PROGRAMA DE MAESTRIA Y DOCTORADO EN CIENCIAS E INGENIERÍA

Sistema de teleasistencia médica en movimiento para casos de emergencias utilizando videoconferencia sobre redes inalámbricas existentes en México.

TESIS

que presenta para obtener el grado de
DOCTORADO EN CIENCIAS

Presenta:

RAÚL RIVERA RODRÍGUEZ

DIRECTOR DE TESIS

Dr. Arturo Serrano Santoyo

Mexicali, Baja California, México, diciembre de 2010

CONTENIDO

	Página
Resumen español.....	i
Resumen inglés.....	ii
Dedicatorias.....	iii
Agradecimientos.....	iv
Contenido.....	v
Lista de Figuras.....	vii
Lista de Tablas.....	ix
CAPÍTULO I.....	3
<i>Introducción.....</i>	<i>3</i>
I.1 Marco de Referencia.....	3
I.2 Antecedentes.....	9
I.3 Planteamiento del Problema.....	13
I.4 Objetivo.....	15
I.5 Metodología de Investigación.....	16
I.6 Infraestructura Empleada.....	17
I.7 Organización de la Tesis.....	17
CAPÍTULO II.....	20
<i>Cross Layer Design y Videoconferencia Móvil.....</i>	<i>20</i>
II.1 Cross Layer Design.....	20
II.1.1 Arquitectura de Cross Layer.....	21
II.1.2 Aproximaciones de Cross Layer.....	23
II.2 Videoconferencia Móvil.....	25
II.2.1 Codificación de video H.263 y H.264.....	29
II.2.2 PSNR (Peak Signal to Noise Ratio).....	31
II.3 Resumen del Capítulo II.....	33
CAPÍTULO III.....	35
<i>Materiales y Métodos.....</i>	<i>35</i>
III.1 Diseño de Cross Layer Propuesto.....	35
III.1.1 Abstracción de Parámetros de Capa de Radio.....	39
III.1.2 Abstracción de Parámetros de Capa de Aplicación.....	45
III.1.3 Propuesta de aproximación de Capas Intermedias.....	52
III.2 Modelado Matemático del Diseño de Cross Layer Propuesto.....	55
III.2.1 Abstracción de Parámetros.....	55
III.2.1.1 Parámetros de Capa de Radio.....	56
III.2.2 Optimizador de Cross Layer.....	63
III.2.3 Distribución de Decisiones.....	68
III.3 Validación del diseño de Cross Layer.....	69
III.4 Resumen del Capítulo III.....	73
CAPÍTULO IV.....	75
<i>Resultados y Análisis Numérico.....</i>	<i>75</i>
IV.1 Consideraciones de Simulación.....	75
IV.2 Resultados de Simulaciones.....	80
IV.2.1 Resultados de la Implementación del Diseño de Cross Layer.....	81
IV.2.2 Análisis de distorsión de la imagen de video.....	93
IV.3 Resumen del Capítulo IV.....	97
CAPÍTULO V.....	99
<i>Discusión y Conclusiones.....</i>	<i>99</i>

V.1 Discusión de Resultados	99
V.2 Conclusiones	104
V.3 Trabajo Futuro	108
LITERATURA CITADA	110
ANEXO A. CARACTERIZACIÓN DE FUENTES DE TRÁFICO	115
ANEXO B. MODELO EMPÍRICO DE PÉRDIDAS DE PAQUETES Y ESTADO DEL CANAL	122
ANEXO C. PARÁMETROS DE QOS	127
ANEXO D. REDES INALÁMBRICAS DE TERCERA GENERACIÓN.....	132

Capítulo I

Introducción

I.1 Marco de Referencia

En términos generales la Telemedicina consiste en la provisión de servicios médicos a distancia, usando procedimientos electrónicos y de telecomunicaciones. En este sentido, el telégrafo, el teléfono, la radio y la televisión han sido utilizados desde sus inicios para llevar asistencia sanitaria a barcos en alta mar, a plantas petrolíferas, a zonas despobladas y, en general, a aquellos escenarios o lugares remotos en donde existía escasez y/o dificultad para proporcionar asistencia sanitaria. El concepto inicial de la Telemedicina se asociaba al de la práctica médica realizada a distancia como única forma de vencer las barreras geográficas.

Los desarrollos tecnológicos actuales han establecido nuevos paradigmas de cómo el uso de la tecnología puede potenciar los servicios haciendo uso adecuado de la información. La convergencia de diferentes tecnologías de comunicación y la Internet, han dado lugar a las nuevas *Tecnologías de la Información y Comunicación* (TIC) . Su penetración a promovido en todos los sectores de la sociedad establecer términos de discusión como “ Sociedad de la Información” y el de la “ Sociedad del Conocimiento”, lo cual ha promovido que en los países desarrollados exista un fuerte impulso y determinación en el desarrollo de aplicaciones en la Telemedicina.

La Telemedicina establece los métodos y formas para establecer la comunicación e intercambiar información entre los actores que participan en una consulta médica o práctica médica en general. Por tal motivo la Telemedicina no puede ser considerada como un elemento tecnológico en sí mismo, sino como una nueva forma de organizar y llevar al cabo, a través de tecnologías, la provisión de servicios médicos a distancia.

Los avances tecnológicos por novedosos que puedan ser, no constituyen más que una herramienta, y como tal, deben valorarse ya que no todo lo que es técnicamente posible resulta necesario o añade valor a la organización de salud. Por eso es absolutamente necesario realizar una gestión eficiente de las TIC en el área de la salud para que su aplicación cubra las necesidades reales y facilite los procesos asociados a la salud.

Otro factor que favorece el término al término e-salud es sin duda, el acceso remoto de los pacientes a los servicios de salud a través de la Web. En la actualidad más del 70% de los usuarios de Internet solicitan información sobre temas médicos y se estima que el 40% de los contenidos Web están relacionados con la salud (Martínez R. Ignacio, 2006). El 35% de los pacientes con padecimientos crónicos de la unión americana, accede habitualmente a información en Internet, relacionada con su enfermedad (K.C. Lun, 2000). Este comportamiento y necesidad de tener acceso a información útil para la salud, establece oportunidad de incorporar las TICs en diferentes áreas de la salud en forma organizada. Para esto a continuación se trata de ubicar y definir la teleasistencia médica.

Los servicios de salud basados en telemedicina requieren ser clasificados por su tipo de actividad, esto puede variar dependiendo del país. Aquí se presentan tres grupos principales divididos según los siguientes criterios (Martínez, 2009):

Según el área de Aplicación:

- Servicios de información o educación, orientados tanto a los profesionales (programas de educación médica continua, educación en línea, gestión administrativa de datos, teleauditorías, proyectos de investigación y de evaluación)

como a los pacientes (ayuda en línea para pacientes crónicos, servicios de salud pública y discapacitados, centros de toxicología, informes de enfermedades).

- Servicios de tratamiento, en los que se aplican las tecnologías y el control a distancia sobre especialidades e intervenciones clínicas tradicionales, dando lugar a nuevos servicios como telecardiología, teleecografía, telecirugía, teledermatología, teleencefalografía, teleradiología, etc., que se describen técnicamente en el siguiente apartado.
- Servicios de atención al paciente o diagnóstico. A este grupo pertenecen diversos servicios:
 - Teleconsulta, entre médico y paciente, pero sin necesidad de estar físicamente presentes.
 - Teleconferencia como una videoconferencia entre profesionales médicos que comparten la responsabilidad del cuidado del paciente, pero sin el enfermo presente. Se suele realizar entre centros rurales de atención primaria y el hospital, o incluso dentro del área urbana.
 - Telemonitorización, en que la información del paciente es recopilada continuamente o a intervalos y analizada en un centro remoto. Este método puede reemplazar las caras visitas a domicilio, monitorizando al paciente en casa en vez de en el hospital.
 - Telediagnóstico, en el que un médico pide información sobre una prueba complementaria, recibiendo el informe del especialista que ha visto la imagen o el registro del paciente. Permite un diagnóstico rápido, barato y ubicuo para dilucidar posibles desacuerdos.
 - Teleasistencia, telecitación, selección de urgencias, procesos asistenciales y asistencia remota. En lo que se refiere a servicios y sistemas de soporte a tareas “asistenciales”, situamos los que se identifican genéricamente con el clásico concepto del término “Telemedicina”. Consiste en la utilización de sistemas de telecomunicación para proporcionar asistencia médica a distancia. Dentro de esta modalidad se puede distinguir entre:

- Teleconsulta-Telediagnóstico: la capacidad de realizar consultas remotas entre pacientes y facultativo o entre diferentes facultativos, para la elaboración de un diagnóstico común. Se vienen utilizando estos servicios en diferentes especialidades médicas (Radiodiagnóstico; Dermatología; Cardiología; Psiquiatría; Oftalmología; Anatomía Patológica, Otorrinolaringología, etc.)
- Telemonitorización-Teleasistencia: permite conocer y realizar un seguimiento a distancia de la situación de un paciente y de sus parámetros vitales y de esta manera permite la provisión de asistencia y cuidados de salud a los pacientes en su entorno habitual (domicilio). Estos servicios permiten apoyar la atención a determinados grupos de pacientes con necesidades especiales, situados fuera del entorno hospitalario: procesos crónicos, programas de cuidados paliativos, medicina de urgencias, etc. Los sistemas suelen ser interactivos e incluir alguna forma de telealarma (para poner en marcha un procedimiento de urgencia en caso de ser preciso) y si están apropiadamente conectados en una red de sistemas integrados, es también factible el acceso a la historia clínica del paciente y un sistema de intercambio de información entre los diferentes profesionales implicados.

La teleasistencia médica puede considerarse como una actividad en la cual se ofrece atención médica a distancia, en donde las tecnologías de la información y comunicaciones (TIC), en particular para este trabajo de investigación, los sistemas de comunicaciones inalámbricas de tercera generación (3G) y de espectro disperso, juegan un papel fundamental en la transmisión de información médica como signos vitales, expedientes clínicos, imágenes médicas, videoconferencia, etc. En este trabajo se estudia el escenario de la teleasistencia médica en movimiento aplicada a casos de emergencia en particular el uso de la videoconferencia como una herramienta de la teleasistencia. En este caso, personal especializado en la sala de urgencias del hospital o centro de control, monitorea y analiza la información transmitida desde una unidad médica en movimiento que le permite emitir

órdenes de acción al paramédico en la ambulancia durante el traslado del sitio de la emergencia hacia el hospital. Desde el punto de vista médico el video y audio en el lugar de la emergencia, es un elemento clave para determinar la condición preliminar del paciente, esto es nombrado hallazgos físicos.

De esta forma la aplicación de videoconferencia en entornos inalámbricos será motivo principal de estudio en este trabajo de investigación. Una característica importante que se considera es la aplicación de la videoconferencia en movimiento, en donde los aspectos de integridad, restablecimiento de las señales, permanencia de la sesión y las capacidades de ancho de banda del canal son fundamentales para lograr un sistema estable, es decir que la sesión de videoconferencia se mantenga en operación.

Las tendencias actuales de las comunicaciones apuntan a la utilización cada vez mayor de dispositivos inalámbricos y móviles, a la convergencia de servicios basados en IP y en general a la existencia de tráfico que es una mezcla de tráfico sensible al retardo como la voz, conferencias multimedia, “streaming video” (transmisión de video en flujo continuo) y, tráfico de datos que no son sensibles al retardo como la navegación por Web, y las transferencias de archivos (Shakkottai et al. 2003).

La implementación de las redes inalámbricas de tercera generación (3G) ha permitido a los usuarios de éstas utilizar satisfactoriamente aplicaciones sensibles al retardo que demandan grandes flujos de datos en el enlace descendente RL (Reverse Link, por sus siglas en inglés). Ejemplos de dichas aplicaciones son el Streaming Video y la Televisión Móvil¹.

Otras aplicaciones sensibles al retardo como la videoconferencia móvil, Voz sobre IP (VoIP, por sus siglas en inglés) ó la video telefonía se encuentran limitadas al utilizarlas sobre redes inalámbricas, debido a que en este tipo de redes los enlaces son generalmente

¹ La aplicación aquí mencionada no se refiere a las del tipo Forward Link Only (FLO, por sus siglas en inglés).

asimétricos y en estos casos se requieren los mismos parámetros de QoS en ambos enlaces RL y FL.

Los esfuerzos por mejorar el funcionamiento de aplicaciones y servicios multimedia en el ambiente inalámbrico se enfocan en optimizar el desempeño de aplicaciones de transmisión de Streaming video y datos en general en el enlace FL. Sin embargo, el uso de videoconferencia en el ambiente inalámbrico, vista como la transmisión de datos sensibles al tiempo, implica minimizar la tasa de pérdida, el retardo de paquetes transmitidos, y la variación del retardo en el inter-arribo de los paquetes, así como mantener constante la tasa de transferencia de datos.

Actualmente, las aplicaciones de videoconferencia utilizan la pila de protocolos TCP/IP y protocolos de establecimiento y control de sesión como H.323 (ITU-T) y SIP (Session Initiation Protocol, RFC 3261, Rosenberg, et al. 2002). El control de flujo multimedia es realizado de acuerdo a los estándares de transporte RTP (Real-Time Transport Protocol) y RTCP (Real-Time Transport Control Protocol) presentados en la (RFC 3550 Schulzrinne, et al, 2003). En este escenario, el flujo de video y audio es enviado en la forma tradicional de comunicación (Open Systems Interconnections, ISO-IEC 7498, ITU-T X.200) entre capas pares, siguiendo una comunicación ascendente y descendente entre capas adyacentes. Esta forma de envío de flujos de información multimedia no considera los problemas de transmisión en las capas no adyacentes del modelo OSI, como se presenta en la capa de aplicación donde no se toma en cuenta los problemas que se presentan en los canales inalámbricos (estos pueden ser principalmente, desvanecimientos de pequeña y gran escala que ocasionan retardos, variaciones en los retardos y pérdida de paquetes); el manejo de los errores provocados por la transmisión inalámbrica, recae en las capas de radio. Estas capas buscan implementar contramedidas que generalmente, implican la adaptación dinámica de los esquemas de modulación digital y las tasas de codificación de los códigos correctores de errores. Desde la perspectiva de la capa de red, las adaptaciones dinámicas mencionadas se perciben como variaciones en la tasa de datos que el radio puede ofrecer. Dado que los protocolos mencionados no contemplan la señalización directa de parámetros de capas inferiores hacia capas superiores, las aplicaciones de videoconferencia generan caudales de

datos multimedia de forma constante, que no pueden ser transmitidos en tiempo real por el radio; esto genera condiciones de encolamiento de los paquetes de información que el mismo radio debe atender. Sin embargo, si los paquetes que integran a los flujos multimedia permanecen mucho tiempo en colas de espera, los protocolos de capa de sesión (RTP, RTCP) los descartan² en la terminal receptora, independientemente de que las capas de radio, red y transporte los hayan podido entregar sin errores.

Debido a lo anterior, la transmisión de una videoconferencia en un canal inalámbrico, sugiere la búsqueda de una estrategia de comunicación entre capas no adyacentes que permita la señalización de parámetros comprensibles entre las capas involucradas. Tal estrategia ha sido propuesta en años anteriores (Shakkottai, 2003) y es conocida como Diseño de Cross-Layer (CLD, por sus siglas en inglés). Esta propuesta de diseño puede tener múltiples estrategias de implementación (Srivastava & Motani, 2005) y continua siendo un área de investigación abierta.

El presente trabajo propone y evalúa un diseño de Cross Layer para notificar a capas superiores el estado del canal en un momento determinado. De esta forma, aplicaciones de videoconferencia móvil pueden adaptar la cantidad de tráfico que generan al caudal eficaz T_h (Throughput, por sus siglas en inglés) ofrecido por el canal inalámbrico reduciendo así, el retardo, la variación del retardo y la tasa de pérdida de paquetes; todos ellos, parámetros de calidad de servicio de aplicaciones sensibles al retardo.

I.2 Antecedentes

En México se han hecho varios esfuerzos en el ámbito de la telemedicina, uno de ellos es el caso del Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE), el cual contempla la consulta médica a distancia utilizando enlaces satelitales y

² Los protocolos de capa de sesión mencionados, descartan los paquetes de datos cuyas marcas de tiempo indicadas en el encabezado RTP, superen intervalos determinados en el estándar RFC 3550; también pueden descartar los paquetes cuyas marcas de tiempo lleguen en desorden, pero esto depende de la capacidad de recuperación de errores del codificador de capa de presentación (audio o video, según sea el caso).

líneas dedicadas para el acceso a Internet (Gómez, 2002). Otros esfuerzos son los del Instituto Mexicano del Seguro Social (IMSS), que se han enfocado a sistemas para la implantación de infraestructura en hospitales digitales. Los proyectos del IMSS se han enfocado en la aplicación del expediente clínico electrónico (ECE) utilizando tecnologías para el despliegue de la información, movilidad de médicos y pacientes, entre otras (Hospitales Digitales, 2007).

Existen otros trabajos de investigación en el área de la telesalud en diferentes instancias académicas y hospitalarias. A la fecha, se han identificado los retos y beneficios que se tienen al implantar y operar una red de comunicaciones inalámbrica en un hospital (Cypher et al., 2006) para proveer un ambiente de conectividad ubicua. Por otro lado, en (Istepanian et al., 2004) se han descrito los conceptos de la llamada disciplina “salud móvil” (M-Health), en la cual se define el uso de las tecnologías inalámbricas en forma general, considerando el caso de movilidad en la provisión de servicios y cuidados médicos. Existen otros trabajos (Poon et al., 2006) que se enfocan en la aplicación de redes de sensores y métodos biométricos al ofrecer servicios de salud móvil.

El envío de tráfico multimedia en el enlace ascendente FL (Forward Link, por sus siglas en inglés) utilizado en comunicaciones inalámbricas ha sido estudiado de manera amplia en los años recientes (Shan y Zakhor, 2002; Yavuz et al. 2006; Stockhammer et al. 2003; Choi et al. 2004; Choi et al. 2005; Ksentini et al. 2006; Khan et al. 2006; Choi et al. 2006) tanto en el entorno de las WLAN (Redes Inalámbricas de Área Local, por sus siglas en inglés) como en el entorno de las WWAN (Redes Inalámbricas de Área Amplia, por sus siglas en inglés). Lo anterior ha sido motivado por la búsqueda de mecanismos que permitan establecer niveles de QoS para los diferentes flujos de información que atraviesan las redes de datos y en el caso de las WWAN, ha sido también porque el Th que las redes de tercera generación ofrecen, soporta grandes flujos de información, como los requeridos por las aplicaciones multimedia.

En este contexto, es posible encontrar trabajos sobre el desarrollo de mecanismos que permiten a las aplicaciones el envío de streaming video en el FL así como investigación que busca optimizar el comportamiento de capas aisladas. (Choi, et al 2005) propone un modelo

del impacto que tiene la pérdida de paquetes en la distorsión de una secuencia de video transmitida en un canal inalámbrico, demostrando como éste contribuye a la distorsión de la imagen de video. A diferencia del trabajo de (Choi, et al 2005) este trabajo propone el uso de un modelo semiempírico que correlaciona la potencia percibida por el radio con la tasa de pérdida de paquetes a nivel de capa de red. El uso de modelos semiempíricos o empíricos proporciona certidumbre a los resultados de la propuesta. Por su parte (Khan, et al, 2006) propone una estrategia de comunicación entre capas no adyacentes del modelo OSI con la intención de minimizar la distorsión y maximizar la percepción cualitativa del video transmitido sobre un canal inalámbrico. El presente trabajo difiere del mencionado en la estrategia de comunicación entre capas no adyacentes, como se presentará a detalle más adelante. (Haratcherev, et al. 2005) propone una adaptación y sincronización de las capas de aplicación y de radio con el objetivo de maximizar la tasa de transferencia de secuencias de video cuando éstas son transmitidas en un canal inalámbrico descendente. La adaptación propuesta por el artículo implica modificaciones a los parámetros del radio, mismos que no se pueden implementar con facilidad en la práctica. A diferencia del artículo de (Haratcherev, et al. 2005), el presente trabajo de investigación propone realizar adaptaciones dinámicas únicamente en la capa de aplicación, permitiendo que el radio adopte las contramedidas al canal radio ya establecidas por el estándar de comunicaciones en cuestión. Más adelante se explicará a detalle cuales son las estrategias, que este trabajo de investigación propone, para mitigar los efectos del canal sobre el flujo de video de una videoconferencia móvil.

A pesar de que ya existen múltiples aplicaciones que soportan el envío y recepción de flujos de audio y video en tiempo real,³ éstas no aprovechan la información del estado de la red o del canal. El diseño de Cross Layer es utilizado para tratar de mejorar el diseño de arquitectura de comunicaciones poniendo énfasis en la optimización y traspaso de parámetros de estados entre capas, y no solo entre capas pares, como es el diseño tradicional en la pila de comunicación OSI. En el ambiente inalámbrico como se mencionó anteriormente se pueden presentar fenómenos propios del medio físico de transmisión que

³ Cuando se habla de aplicaciones en tiempo real, se debe entender que son aplicaciones que envían datos al mismo tiempo que los generan y se espera que los receptores reciban dichos datos en el menor tiempo posible.

afectan directamente a las sesiones multimedia, como lo es la videoconferencia. En los trabajos presentados por parte de (Khan, et al, 2006) expone como la capa de aplicación es considerada para reconocer el efecto de distorsión provocado por la pérdida de paquetes en la capa de red provocado por congestión, por su parte la capa física y de enlace de datos son consideradas de igual forma debido a que estas estiman las capacidades del medio inalámbrico adaptándose rápidamente a las variaciones en el canal. La pérdida de paquetes es propiciada por las variaciones de potencia o desvanecimientos afectando a parámetros a nivel de red como la variación del retardo, así como se mencionó en el capítulo anterior las variaciones de retardo a lo largo del proceso de transmisión afecta a la capa de transporte RTP/RTCP descartando paquetes en esta capa, y en la capa de presentación al descartar paquetes de información tardíos en el proceso de decodificación. Es por ello necesario realizar un diseño donde las capas superiores sepan lo que acontece en las capas inferiores (aproximación de abajo hacia arriba) y de esta forma las fuentes de información se puedan ajustar a las prestaciones de las capas inferiores. También se puede utilizar una aproximación en el diseño de arriba hacia abajo, donde las capas inferiores se adapten a los requerimientos de las capas superiores o puede haber aproximaciones con combinaciones en el diseño, llamadas híbridas.

Este trabajo de tesis enfrenta los retos de utilizar videoconferencia punto a punto bidireccional en un enlace inalámbrico de 3G, donde la arquitectura de diseño normalmente toma en consideración fortalecer al enlace FL. Se consideran también la movilidad y el entorno como variables que afectan la transmisión de la videoconferencia y se analiza la ventaja de la arquitectura de Cross Layer utilizada en este punto.

De acuerdo a los trabajos que se mencionan en esta sección como antecedentes, existen esfuerzos por satisfacer el enlace descendente FL utilizando CLD con estrategias para optimizar parámetros en las capas de radio, como la potencia necesaria para la transmisión asociado a un esquema de modulación. Sin embargo este trabajo de tesis se enfoca en el enlace ascendente RL, el cual tiene características diferentes (el móvil está limitado en capacidad de procesamiento, en potencia de transmisión y en el consumo de energía; además hay restricción en la dimensión y geometría de las antenas) que impiden la

aplicación directa de los modelos del FL. Otra razón es el hecho de que esta investigación busca una solución que coexista con las tecnologías de comunicaciones inalámbricas estandarizadas ya existentes en el mercado. Otros trabajos como el de (Choi, et al 2006) propone de igual forma optimizar la transmisión de video en el FL modificando la estrategia en las capas física y de enlace de datos, al priorizar los paquetes de información de video que corresponden a los cuadros de imagen con mayor cantidad de información en escenarios de estrés en el sistema. Este trabajo de tesis propone una estrategia de CLD a capas intermedias donde las modificaciones de adaptación se hacen en la capa de aplicación. Para poder realizar esto es necesario una abstracción de parámetros de cada una de las capas involucradas en el proceso de CLD como se presenta a detalle en el capítulo II y capítulo III.

I.3 Planteamiento del Problema

La teleasistencia médica en movimiento requiere de diferentes herramientas para y sobretodo mantener con vida al paciente, vigilando y manteniendo sus signos vitales. Estas actividades son realizadas por parte de los paramédicos y médicos a bordo de una ambulancia. Actualmente la creciente demanda en servicios de salud que atienden accidentes o urgencias médicas requiere de una gran inversión en recursos humanos y de equipo médico así como de comunicaciones. La indagación inicial realizada con el Hospital Medica Sur, revela que la comunicación con una ambulancia se limita a conocer las condiciones del paciente en forma general, sin poder realizar un diagnóstico prehospitalario que redunde en la mejora de la condición del paciente o incluso evite su ingreso al hospital. Es aquí donde existe la necesidad de que el médico especialista pueda tener un contacto visual para realizar actividades de hallazgos físicos y poder tener un mayor conocimiento del estado del paciente y su gravedad de salud. De esta forma este trabajo de tesis considera como caso de estudio, la transmisión de videoconferencia sobre canales inalámbricos, en un escenario de atención médica en casos de urgencias, donde la movilidad es continua a velocidades vehiculares (40 a 80 km/hr) en entornos urbanos. Los trabajos previos

revisados en los antecedentes, están orientados a optimizar la transmisión de streaming video y otras aplicaciones multimedia en el enlace ascendente FL, lo anterior provocado por diversos factores, uno de ellos es la baja capacidad de procesamiento disponible en las unidades móviles así como la potencia disponible para transmitir y principalmente el comportamiento en las prácticas del uso de Internet o acceso a datos en general, el cual presenta un comportamiento asimétrico. Es decir, es más la demanda de información por el móvil y el comportamiento en datos representa menor costo (en bytes) al solicitar información, comparado con la cantidad de bytes que se reciben en respuesta a la solicitud. Debido a que los enlaces inalámbricos tienden a ser asimétricos respecto a la tasa de transferencia que ofrecen. Esto último afecta directamente a los flujos simétricos de envío de video y audio bidireccional utilizado en una sesión de videoconferencia. Las tecnologías de redes inalámbricas de 3G utilizan enlaces asimétricos siendo el enlace RL el de menor capacidad de transmisión de datos (ver la figura 1).

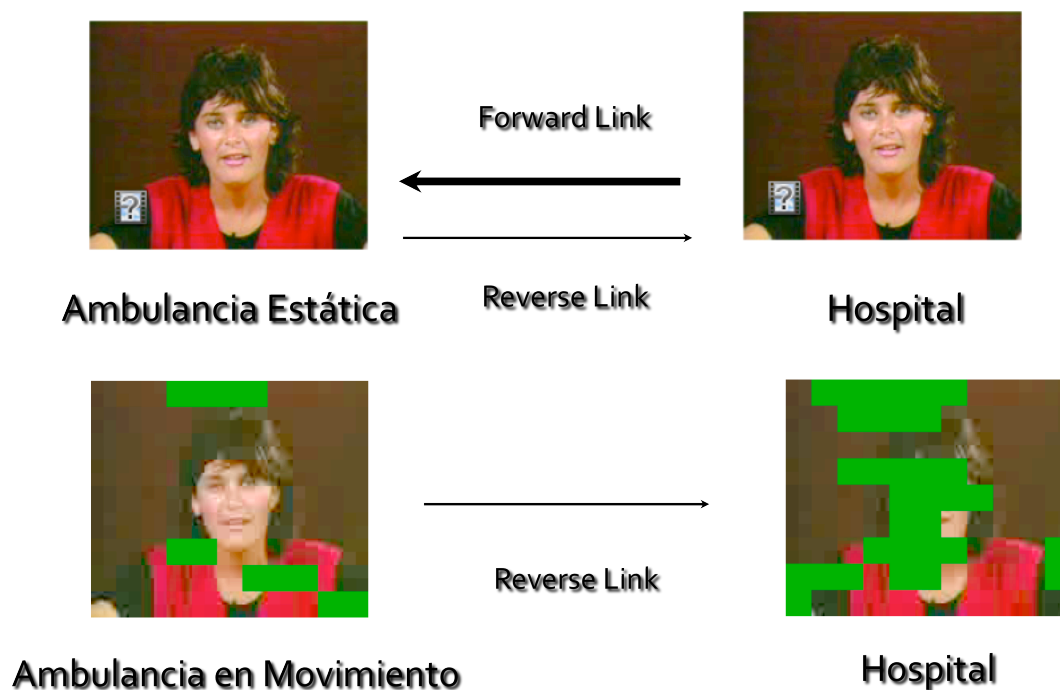


Figura 1. Degradación del flujo de video transmitido desde la ambulancia hasta el hospital, utilizando un enlace inalámbrico 3G asimétrico.

Las tecnologías 3G tales como 1x EV-DO rev. 0 y superiores, ofrecen mayores caudales eficaces en el enlace RL que cualquier tecnología celular previa, haciendo factible el uso de aplicaciones como videoconferencia, que demandan los mismos parámetros de calidad de servicio tanto en el enlace FL como en el mencionado enlace RL. Lo anterior ocasiona que el diseño e implementación de mecanismos que optimicen la transmisión de flujos sensibles al retardo y de gran Th sobre el enlace ascendente de redes WWAN, sea un tema de investigación abierto.

En virtud de que las aplicaciones no aprovechan la información del estado de las capas de Enlace de Datos y Física o viceversa, se requiere un esquema de señalización entre capas que permita que dichas aplicaciones tengan conocimiento del comportamiento variable del canal inalámbrico en el enlace RL. Sin embargo, no es suficiente que solo conozca el estado del canal, adicionalmente se necesita un proceso de control con retroalimentación que minimice el error entre parámetros válidos (descritos en el capítulo III), obtenidos de un proceso de abstracción de las capas en cuestión y que satisfagan las necesidades de la capa de aplicación.

I.4 Objetivo

Este trabajo de tesis tiene como objetivo general, diseñar un método de adaptación de parámetros de calidad de servicio para videoconferencia móvil aplicado a la teleasistencia médica prehospitalaria.

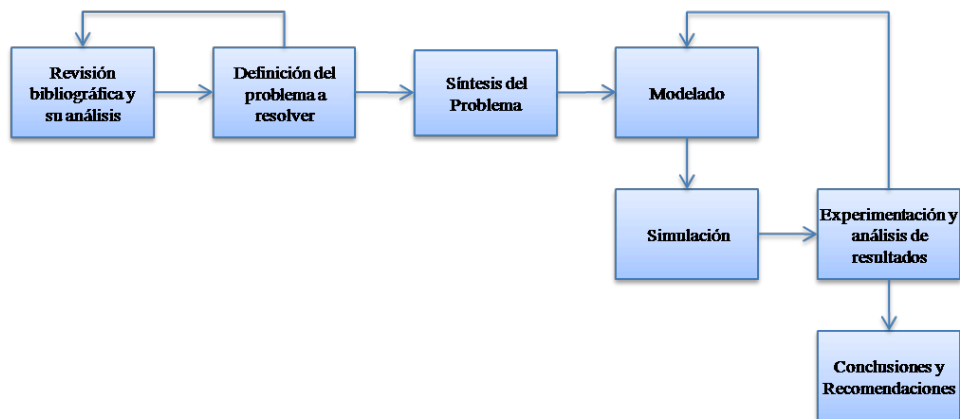
Los objetivos particulares que se han fijado son los siguientes:

- Proponer un método de control para ajuste dinámico del flujo de datos generado por una videoconferencia desde una terminal móvil.
- Realizar un estudio del comportamiento del método de control de la videoconferencia en entornos inalámbricos utilizando la red celular de tercera generación IS-856 (cdma2000 1xEV-DO Rev. 0).

I.5 Metodología de Investigación

La metodología y etapas del desarrollo del trabajo de investigación, utilizada para alcanzar los objetivos planteados se muestran a continuación.

(a) Metodología general del trabajo de investigación desarrollado



(b) Etapas desarrolladas para la solución del problema



Durante la fase de revisión bibliográfica y análisis se realizó una investigación del estado del arte del Diseño de Cross Layer con el objetivo de identificar la aproximación y la arquitectura adecuadas que ayudaran a la definición y eventual solución del problema a resolver. Una vez definido y sintetizado el problema a resolverse se procedió a modelar matemáticamente la solución propuesta para comenzar el proceso de simulación, el cual estuvo retroalimentado a través de la evaluación del modelado resultado de la experimentación y el análisis de los resultados parciales que se fueron obteniendo. Una vez determinado el modelo final y la validez de la simulación se analizaron los resultados finales y se desarrollaron las conclusiones del trabajo.

I.6 Infraestructura Empleada

Durante el proceso de desarrollo de la tesis fue necesario realizar mediciones de campo que permitan obtener un modelo empírico que relaciona las variaciones del canal y movilidad con la tasa de pérdida de paquetes en la red 1x EV-DO fueron utilizados los siguientes elementos:

- Laboratorio Móvil de Telemedicina equipado con sistema de videoconferencia.
- Tarjeta de red inalámbrica 1x EV-DO Rev. 0 adicional.

I.7 Organización de la Tesis

En el capítulo I se explica como la telemedicina presenta una amplia variedad de casos de uso en la salud ubicando a la “teleasistencia médica” con las tecnologías de información y comunicaciones. La videoconferencia en general, es una aplicación en la que la información de video y la de audio son transmitidas por canales lógicos separados, i. e., son dos flujos de información sensible al retardo diferentes. Este trabajo de tesis se centra únicamente en la optimización de la transmisión del video de una videoconferencia sobre el canal ascendente de la comunicación entre una Red de Acceso y una Terminal de Acceso

EV-DO, lo anterior considerando que el caso de uso de “teleasistencia médica” utiliza la videoconferencia móvil en un enlace inalámbrico asimétrico, en donde el enlace ascendente RL, posee menos capacidad de en la tasa de transmisión efectiva (ver anexo D) y a través de este se transmite el video desde una unidad móvil de teleasistencia médica hacia un hospital.

En el capítulo II se presenta el marco teórico que envuelve este trabajo de investigación destacando cuatro conceptos clave:

1. Las necesidades de la videoconferencia móvil como aplicación multimedia sensible al retardo,
2. Los parámetros de calidad de servicio que normalmente se utilizan para evaluar el desempeño de las redes que transportan las aplicaciones y el impacto que algunos de ellos tienen respecto a las aplicaciones multimedia,
3. El diseño de Cross Layer, la arquitectura utilizada y la diferentes aproximaciones existentes en la literatura y,
4. Características generales de las redes inalámbricas de tercera generación, en particular cdma2000 1xEV-DO en la que está basada la parte de radio de esta tesis.

En el capítulo III se presenta el Diseño de Cross Layer propuesto en este trabajo de investigación y éste es formalizado matemáticamente en concordancia con el estado del arte disponible en la literatura. Se destaca la importancia del proceso de abstracción de parámetros y se describe detalladamente el proceso de Aproximación a Capas Intermedias utilizado en este trabajo de investigación. En este capítulo se describe la función objetivo y se describe su regla de correspondencia, se hace énfasis también en el proceso de distribución de decisiones. En este capítulo se detallan los modelos analíticos, empíricos y de simulación utilizados en este trabajo y se valida la propuesta a través de simulación.

En el capítulo IV se presentan los resultados obtenidos y se hace un análisis numérico de los mismos. Finalmente en el capítulo V se discuten los resultados obtenidos a través de un

análisis profundo de los mismos y se presentan las conclusiones y los posibles trabajos futuros a realizarse a partir de esta tesis. La importancia de la modularidad de la arquitectura de Cross Layer utilizada es destacada y se explica su aplicabilidad para redes diferentes a la utilizada por este trabajo de investigación.

Finalmente en el capítulo V se presenta la discusión sobre los resultados obtenidos y su conclusión.

Cross Layer Design y Videoconferencia Móvil

En este capítulo se introducirá el concepto de Cross Layer Design, su arquitectura y diferentes aproximaciones. Posteriormente se establecerán las características de la videoconferencia móvil, enfatizando qué aspectos se utilizarán para la propuesta de optimización de Cross Layer cuando ésta aplicación es transmitida sobre un enlace de redes inalámbricas de tercera generación EV-DO.

II.1 Cross Layer Design

De acuerdo a Choi et al. (2006) el concepto Cross Layer Design introduce acoplamiento entre capas a través de la pila de protocolos y permite el intercambio de información entre capas diferentes con el objetivo de optimizar la comunicación entre dos o más capas. Aunque este enfoque puede ser utilizado en cualquier tipo de redes de comunicaciones, es de especial interés su utilización en las comunicaciones móviles debido a las particularidades del canal inalámbrico.

La naturaleza del canal radio, variante en tiempo y frecuencia, aunado a la movilidad de los usuarios, genera que el desempeño de las comunicaciones inalámbricas tenga una variación prácticamente aleatoria (Choi et al, 2004). La idea de ofrecer diferentes niveles de QoS a diferentes flujos de información bajo estas condiciones es un concepto difícil de lograr con la estructura de comunicación establecida en las pilas de protocolos correspondientes y es por ello que el concepto de Cross Layer Design ha recibido mucha atención en las comunicaciones móviles cuando se deben ofrecer servicios multimedia (Choi et al, 2004). De acuerdo con (Khan et al, 2006), Cross Layer Design debe ser visto como un

complemento a la estructura de comunicación por capas tradicional, dado que dicho esquema ha demostrado su eficiencia en las redes del pasado y en las actuales.

Más aún, es fundamental contar con una arquitectura que garantice la estabilidad del sistema de comunicaciones y considerar que a medida que se incrementan los procesos de señalización (procesos de Cross Layer con el propósito de mejorar las condiciones de la comunicación) se incrementan también los retardos en el proceso de envío y recepción de información. Kawadia y Kumar (2005) indica que la implementación de diferentes esquemas de Cross Layer en diferentes nodos de una red, podría volver caóticas las comunicaciones y hacen énfasis en que las nuevas propuestas de diseños de Cross Layer deben considerar la totalidad del diseño, incluyendo las interacciones con otras capas.

La implementación del diseño de Cross Layer había sido visto de manera positiva por (Shakkottai et al. 2003) y los trabajos más recientes comprueban dicha visión. Ya desde entonces era posible vislumbrar la necesidad de crear modelos que relacionaran las características del canal con parámetros manejables a niveles de capa de red y superiores (Shakkottai et al. 2003), propuesta que se retoma en este trabajo de tesis para efectos de una adecuada abstracción de parámetros de capas inferiores.

II.1.1 Arquitectura de Cross Layer

Kawadia y Kumar (2005) enfatizan la importancia de la arquitectura en el diseño de un sistema al mencionar que ésta “acelera el desarrollo del diseño y la implementación permitiendo esfuerzos paralelos”. Establecen también que la modularidad provee la abstracción necesaria para comprender la totalidad de los sistemas y ha sido mostrado a lo largo de la historia que el desarrollo de arquitecturas conlleva a la larga vida de las mismas.

El seguimiento de una arquitectura y su implícita modularidad permiten también el desarrollo individual de los módulos y por consiguiente su eventual actualización sin tener la necesidad de rediseñar el sistema completo. Quizá uno de los ejemplos más

sobresalientes del éxito en el uso de arquitecturas es el modelo de referencia OSI, arquitectura que descompone el proceso de comunicación entre dos equipos remotos en capas que tienen funciones específicas.

La literatura muestra que los esfuerzos por generar arquitecturas de Cross Layer sólidas iniciaron posteriormente a las primeras propuestas de implementación de dichas técnicas. En estas primeras propuestas (Shan y Zakhor, 2002) no es posible identificar módulos de una arquitectura y sin embargo, se puede vislumbrar el contorno de los procesos necesarios para llevar a cabo un diseño de Cross Layer formal. Existen posteriores propuestas de arquitecturas de Cross Layer de las que en este trabajo de investigación se tiene conocimiento (Choi et al. 2004; Khan et al. 2006) y en dónde es posible identificar, a pesar de sus respectivas diferencias, tres procesos o módulos de las arquitecturas en común:

- a) Proceso de abstracción,
- b) Proceso de optimización y,
- c) Proceso de reconfiguración o distribución de decisiones.

En todo diseño de Cross Layer es necesario un proceso de abstracción de parámetros dado que los parámetros propios de una capa son ininteligibles a otras capas (e. g., esquemas de modulación o tasas de corrección de errores son parámetros propios de la capa física que no pueden ser interpretados, ni siquiera, a nivel de capa de Red). Lo anterior implica la necesidad de una transformación de parámetros específicos de una capa en parámetros que sean comprensibles para otras capas o, como se verá, para el Optimizador de Cross Layer.

El segundo proceso, el de optimización, es aquél en dónde se toma una decisión que eventualmente, mejorará el desempeño del proceso de comunicación. El optimizador de Cross Layer tiene como entradas uno o varios conjuntos de parámetros abstraídos de las diferentes capas que participarán en el proceso de optimización y tiene como salida la optimización de dichos parámetros con respecto a una función objetivo particular.

El proceso de reconfiguración o distribución de decisiones es el proceso inverso al proceso de Abstracción de Parámetros. En este módulo se reciben los valores óptimos provenientes del Optimizador de Cross Layer y se transforman en parámetros específicos de las capas relacionadas al proceso de optimización. Las respectivas capas ejecutarán adaptaciones específicas en función de lo señalado por el Distribuidor de Decisiones.

II.1.2 Aproximaciones de Cross Layer

De lo establecido en párrafos anteriores, es evidente la posibilidad de generar múltiples enfoques en el Diseño de Cross Layer, aún para optimizar una misma aplicación. Es posible intentar señalar a la capa de Enlace de Datos (MAC)⁴ las prioridades de diferentes flujos de información en orden de importancia para que dicha capa pueda colocar los flujos en colas apropiadas según la importancia de los datos, pero bien podría hacerse lo mismo para la capa de Red, señalizando la importancia de los flujos para que se modifiquen los valores del campo DSCP (Differentiated Services Code Point) del encabezado IP; inclusive ambas al mismo tiempo.

Van Der Schaar y Shankar (2005) propusieron una clasificación de las aproximaciones de Cross Layer basándose en la forma en la que se lleva a cabo la optimización:

- Aproximación de arriba hacia abajo (Top-down approach). Los protocolos de capas superiores optimizan sus parámetros y estrategias en la capa inferior siguiente, i. e., las aplicaciones determinan parámetros y estrategias de capa MAC y ésta determina los parámetros y estrategias que la capa Física (PHY)⁵ debe seguir. Este tipo de aproximación es útil cuando las capas inferiores tienen los recursos suficientes como para responder a las demandas de las capas superiores.
- Aproximación de abajo hacia arriba (Bottom-up approach). Las capas inferiores tratan de aislar a las capas superiores de las pérdidas y las variaciones en el Th que

⁴ La capa de Enlace de Datos del modelo de referencia OSI es comúnmente citada en la literatura como la capa MAC (del inglés, Media Access Control).

⁵ La capa Física del modelo de referencia OSI es comúnmente citada en la literatura como la capa PHY.

se sufren en el canal. La estrategia consiste en señalar datos a las capas superiores para que tengan conocimiento de las restricciones que las capas bajas perciben. El objetivo es que las capas superiores se adapten a las prestaciones ofrecidas por las capas inferiores.

- Aproximación centrada en la aplicación (App-centric approach). La capa de Aplicación (APP) optimiza los parámetros de una capa inferior a la vez, la optimización puede tener origen en un enfoque Bottom-up o Top-down. Se distingue de la primera aproximación en que la capa de Aplicación determina los parámetros de todas y cada una de las capas inferiores. Esta estrategia puede utilizarse cuando existe abundancia de recursos de red a través de las diferentes capas del modelo OSI, computacionalmente es compleja dado el incremento de las primitivas que debe controlar la aplicación.
- Aproximación centrada en la capa de enlace de datos (MAC-centric approach). La capa de aplicación señala a la capa MAC la importancia de los flujos de datos que se generan con el objetivo de que ésta determine que paquetes o flujos serán enviados y con qué nivel de QoS lo harán. En forma general, esta aproximación es un caso particular de la estrategia Top-down, pero la señalización es estrictamente realizada a la capa MAC. La estrategia es adecuada cuando es posible modificar la forma en la que la capa MAC administra las colas o líneas de espera.
- Aproximación integrada (Integrated Approach). Todas las estrategias son determinadas de manera conjunta, vistas como un todo. Es una estrategia similar a App-centric, con la diferencia de que se pueden generar primitivas específicas para cada capa, no sólo para la capa de aplicación. La implementación práctica de esta estrategia es complicada debido a la gran cantidad de procesos de señalización requeridos, implica un desarrollo específico para cada una de las capas del modelo OSI.

En tiempos más recientes han surgido nuevas propuestas de aproximaciones híbridas de las ya mencionadas (Choi et al, 2006). En estas nuevas aproximaciones se busca la mejor adaptación al problema a resolver. En el Diseño de Cross Layer es fundamental una

adecuada abstracción de parámetros porque esta abstracción determinará las posibles formas que podrá tomar la función objetivo dentro del Optimizador de Cross Layer. La función objetivo per se, puede ser diferente (aún siendo alimentada con los mismos parámetros abstraídos) si se busca la optimización de diferentes parámetros.

La aproximación de Cross Layer elegida tendrá que estar en función de los parámetros que se desean optimizar y de la abstracción realizada. En este trabajo de investigación se propone el uso de una aproximación que no empata explícitamente en ninguna de las mencionadas. La propuesta de este trabajo es una estrategia híbrida⁶ entre un enfoque Top-down y uno Bottom-up centrado en la capa de red. Esta aproximación es adecuada para el caso de uso en cuestión debido a la necesidad de mantener la independencia de las capas inferiores. Esto es, lograr que el diseño sea utilizable aún cuando la tecnología de radio cambie. Lo anterior implica abstraer parámetros de la capa de radio a una capa intermedia en donde parámetros de la capa de aplicación abstraídos previamente, pueden alimentar a una función objetivo. La aproximación propuesta será descrita en el capítulo III.

II.2 Videoconferencia Móvil

La videoconferencia es una aplicación multimedia que involucra el envío y recepción de video y audio en tiempo real y requiere los mismos parámetros de QoS tanto en recepción, como en transmisión.

En la videoconferencia tradicional ambos extremos del enlace son fijos, i. e., no tienen movilidad y el tráfico que generan es manejado con diferente prioridad al resto del tráfico que circula por las redes que se interconectan a través de la videoconferencia. Esta condición se debe cumplir por los requerimientos especiales que estas aplicaciones tienen.

⁶ Se utiliza el término “híbrido” propuesto por (Choi et al, 2006) para describir el concepto de la estrategia, pero la estrategia per se, es diferente.

En general es posible establecer que una videoconferencia será de buena calidad si se aseguran y establecen como fijos (o con variación limitada) algunos valores típicos de las comunicaciones de datos (Ojanperä y Prasad, 1998). Dichos valores típicos son conocidos como parámetros de Calidad de Servicio (ver anexo C) y dentro de los más representativos se encuentran el caudal eficaz (T_h) requerido por la aplicación, el retardo extremo-a-extremo (End-to-end Delay), la Variación del Retardo (Jitter) y la Tasa de Pérdida de Paquetes (Packet Loss Rate, PLR).

Existen diferentes tipos de tráfico sensibles a diferentes umbrales de cada uno de los parámetros de QoS mencionados, siendo las aplicaciones en tiempo real (como la videoconferencia) las que más restricciones imponen al respecto. La videoconferencia es un caso extremo en cuanto a demanda de reservación de recursos de red para su óptimo funcionamiento, dado que ya ha sido establecido que no sólo requiere niveles elevados de calidad de servicio en el enlace de recepción sino que requiere los mismos niveles en el enlace de transmisión.

La videoconferencia móvil a velocidades vehiculares (40 a 80 km/hr) añade un nivel de desafío mayor dado que involucra comunicaciones móviles e inalámbricas en por lo menos, uno de los extremos de la videoconferencia y dada la asimetría convencional de los enlaces inalámbricos, mantener los mismos niveles de QoS en ambos enlaces (FL y RL) no es una tarea sencilla.

Como se presento en el capítulo anterior, existen otros trabajos (Khan, et al. 2006; Shan y Zakhor, 2002; Choi, et al. 2004; Choi, et al 2005; Choi, et al 2006) respecto a cómo mejorar la calidad del envío de streaming video en el enlace FL e inclusive se han propuesto diferentes mecanismos de QoS para hacer confiable la transmisión de este tipo de aplicaciones. Sin embargo, hay poco trabajo dirigido hacia la optimización del envío de video desde una terminal móvil, i. e., el uso del enlace RL para envío de tráfico sensible al retardo.

Lo anterior posiblemente obedezca a dos factores:

- Las investigaciones sobre transmisión de datos sensibles al retardo a través de enlaces inalámbricos se habían centrado en las tecnologías que permitían hacerlo, i. e., las tecnologías de redes de área local inalámbricas (WLAN). En las mencionadas tecnologías los enlaces inalámbricos no son asimétricos y los caudales eficaces ofrecidos son elevados, de forma tal que los mecanismos para transmitir hacia las estaciones inalámbricas son aplicables también a la transmisión desde las estaciones hacia el punto de acceso (AP, por sus siglas en inglés) y no hay restricciones prácticas en cuanto al T_h que demandan aplicaciones sensibles al retardo (aún si generan caudales de datos del orden de centenas de kilobits por segundo).
- Las tecnologías de transmisión de datos sobre redes celulares no ofrecían caudales eficaces lo suficientemente amplios como para hacer factible el envío de audio y video sobre el enlace RL.

Con la llegada de las comunicaciones inalámbricas de tercera generación (3G) el T_h ofrecido en los enlaces ascendentes se ha visto incrementado posibilitando así, el envío de tráfico sensible al retardo a través de dichos enlaces. La condición de asimetría respecto a los caudales eficaces sigue existiendo, dado que normalmente los usuarios descargaban más información de la que enviaban, sin embargo, la demanda de nuevas aplicaciones multimedia (e. g., la videoconferencia móvil) está cambiando este comportamiento.

Tal y como en la videoconferencia convencional, en la videoconferencia móvil se deben mantener los mismos parámetros de calidad de servicio tanto en el enlace FL como en el enlace RL y es en esta parte dónde se presenta el mayor desafío, dado que mantener parámetros de QoS en umbrales adecuados en el enlace RL está en función de las condiciones del canal.

Siendo el canal radio altamente variable y dependiente de muchos factores (entre ellos la velocidad del móvil), lograr mantener parámetros de QoS en niveles específicos tales que permitan el establecimiento, mantenimiento y término adecuado de una videoconferencia móvil, es una tarea que demanda gestionar los recursos de red para obtener un desempeño satisfactorio desde el punto de vista de la aplicación. Particularmente en este trabajo de investigación, como se mencionó en la sección anterior, se propone utilizar una aproximación de CLD híbrida, en donde se toma en cuenta la capa de aplicación, como la videoconferencia desde el punto de vista del móvil. En el capítulo III se presenta el proceso de abstracción de parámetros que determinan los parámetros válidos a utilizarse en la propuesta de CLD. Es importante mencionar que a diferencia de las aplicaciones de videoconferencia tradicionales donde la negociación de los recursos se realiza al inicio de la sesión, en este trabajo se propone un cambio en el paradigma al considerar un sistema de videoconferencia el cual pueda modificar sus parámetros (propios de la aplicación y definidos en el proceso de abstracción de CLD) en el momento que los recursos de red disponibles cambien o se modifiquen a favor o en contra. Estos recursos de red por su parte, atienden al comportamiento y prestaciones propias de las capas de Enlace de Datos y Física del modelo OSI, en este caso de estudio la tecnología de transmisión inalámbrica utilizada en particular es cdma2000 1xEV-DO (ver anexo D).

Las nuevas tecnologías de transmisión inalámbricas buscan la manera de evaluar el estado del canal y en la medida de lo posible, de predecir su comportamiento. En respuesta a dicha búsqueda, es posible encontrar actualmente esquemas de modulación adaptativos, tasas de codificación adaptativa y en general, mecanismos que se adaptan a las condiciones que el canal radio ofrece en un momento determinado. La forma en la que impactan las adversidades del canal a los flujos de información y las contramedidas que se emplean para evitarlas, son particulares de cada tecnología de transmisión.

El hecho de que las capas de Enlace de Datos y Física del modelo OSI se adapten a las condiciones del canal impacta en los recursos de comunicación que se encuentran disponibles y que eventualmente son ofrecidos a las aplicaciones. Por ejemplo, el Th

máximo que se puede ofrecer con un esquema de modulación QPSK (Quadrature Phase Shift Keying) es, obviamente, mayor que el que se puede ofrecer con un esquema de modulación BPSK (Binary Phase Shift Keying). Si se establece una sesión de videoconferencia mientras el canal permite utilizar QPSK, los parámetros de QoS se establecerán de acuerdo al Th que se ofrece en ese momento. Cuando las capas inferiores se adapten a un canal más agresivo y utilicen BPSK como esquema de modulación el Th máximo ofrecido tendrá, necesariamente, que ser menor. Esto modifica los niveles de QoS que la tecnología puede ofrecer en función de las variaciones del canal lo que impactará en el desempeño de las aplicaciones.

Como resumen, las aplicaciones sensibles al retardo como la videoconferencia, no tienen mecanismos implícitos que les permita adaptarse a situaciones cambiantes en los recursos de red, como la descrita en el párrafo anterior, afectando directamente a la calidad de la imagen. A continuación se presenta en forma general el proceso de codificación de video estándar para videoconferencia.

II.2.1 Codificación de video H.263 y H.264.

H.263 es un estándar de codificación de video diseñado originalmente por ITU-T en 1995/1996 como un formato de compresión de baja tasa de transmisión en bits utilizado para la transmisión de video en videoconferencia. Es un miembro de la familia de H.26x de los estándares de codificación de video de la ITU-T Grupo de expertos video de la codificación (VCEG).

El siguiente codificador de la ITU-T VCEG (en sociedad con MPEG) después de H.263 es el estándar H.264, también conocido como AVC y MPEG-4. Pues H.264 proporciona una mejora significativa en capacidad más allá de H.263.

La compresión de video es posible gracias a que en una secuencia de imágenes puede existir redundancia espacial (figura 2) tanto como temporal, así como mecanismos de predicción [Richardson, 2003]. La redundancia espacial se refiere a que en una imagen pueden existir regiones donde la textura en la imagen no tenga grandes variaciones (ej. fondos lisos, como puede ser una pared blanca). La redundancia temporal indica que a lo

largo del tiempo, un objeto se representa de la misma manera, es decir, no tuvo cambios (ej. no hubo movimiento en esa zona). En la figura 3 se muestran estas características.



Figura 2. Cuadro de un video (mostrando ejemplos de zonas homogéneas)

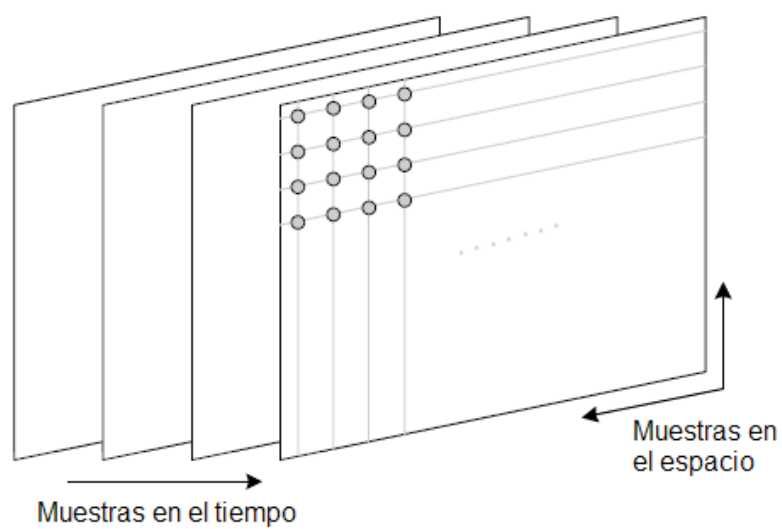


Figura 3. Muestras en el tiempo y en el espacio de una secuencia de video.

Una secuencia de video se encuentra definida por un grupo de imágenes (GOP). Un GOP especifica la manera en la que están ordenados los cuadros de video. Existen 5 tipos de cuadros, pero solo dos se utilizan en el perfil básico, el cual es el empleado en la videoconferencia [Richardson, 2003]. Estos cuadros son:

- Cuadro I. Es una imagen de referencia, la cual contiene la mayor parte de la información en un GOP y es independiente del resto.
- Cuadro P. Es un cuadro predictivo, el cual depende directamente de los cuadros I y sin estos no se puede decodificar.

Una vez que estos cuadros de información son transmitidos a través de una red de datos, estos pueden ser extraviados por procesos de congestión o retardos en la red y para poder medir cual fue el impacto de deterioro en la información de video en el extremo receptor, a continuación se introduce el concepto de PSNR y su uso como un indicador de la calidad del video recibido.

II.2.2 PSNR (Peak Signal to Noise Ratio).

La relación señal a ruido pico (PSNR) es una medida logarítmica, que depende del error cuadrático medio (MSE) entre la imagen original y la imagen reconstruida, relativo al cuadrado del valor más alto posible de la imagen (en bits)

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (1)$$

La PSNR puede ser calculada por medio de la expresión anterior (ecuación 1), esta es una medida de calidad de imagen muy popular [Richardson, 2003], ampliamente usada para comparar la calidad de las imágenes comprimidas contra las descomprimidas. Sin embargo, tiene algunas limitantes. Para hacer uso de esta medida, se requiere la imagen original y no

se garantiza que ésta tenga una fidelidad perfecta. La figura 4 muestra un ejemplo con la imagen original y diferentes grados de distorsión.

Para una imagen o secuencia de imágenes dada, una alta PSNR normalmente indica alta calidad, mientras que una baja PSNR indica baja calidad. Sin embargo, un valor en particular de PSNR no necesariamente significa en lo absoluto calidad subjetiva.



Figura 4. Ejemplos de PSNR en una imagen. (a) Original. (b) 30.6 dB. (c) 28.3 dB



Figura 5. Imagen con el fondo borroso (PSNR = 27.7 dB)

Por ejemplo, en la figura 5, se muestra la imagen original comprimida, donde solo el fondo ha sido degradado. Esta imagen tiene 27.7 dB de PSNR, este es resultado de la

comparación con la imagen original (figura 4a.), afectando los detalles de la imagen. Muchas personas podrán decir que se ve significativamente mejor debido a que el rostro es más claro, pero esto contradice la evaluación usando la PSNR, que indica que a mayor cantidad de PSNR, mejor será la imagen. Se considera que un valor entre 30 dB y 40 dB de PSNR es suficiente para decir que la imagen tiene buena calidad [Carlisle, 2007].

Evaluar un aspecto de apreciación numéricamente resulta complicado, ya que cada observador tiene diferente criterio, más aun en el área de la telesalud. Sin embargo, para este trabajo de investigación se propone el uso de la PSNR como medida de evaluación de la propuesta de CLD y así poder tener un parámetro de evaluación que permita evaluar un video a través de todas las secuencias de imágenes que lo conforman. Esto puede ser motivo de estudio, al evaluar un caso de uso de telemedicina con la opinión de los médicos y comparándolo con los valores de PSNR obtenidos en las pruebas, y así determinar cuales valores o intervalos de PSNR son validos y satisface el criterio del grupo de médicos en el estudio en cuestión. Sin embargo esto queda fuera del alcance de este trabajo de investigación y será motivo de otro trabajo.

A continuación se presenta una introducción de CLD como base para el capítulo III donde se profundiza en este tema.

II.3 Resumen del Capítulo II

En este capítulo fue descrito el marco teórico que envuelve al presente trabajo de investigación enfatizando sus tres componentes principales:

- Diseño de Cross Layer,
- Videoconferencia Móvil y,
- Parámetros de QoS,

todos ellos dentro del contexto de los trabajos relacionados y el estado del arte. Se definieron las características particulares de la videoconferencia móvil, las consideraciones hechas respecto a los parámetros de QoS que se consideran para las aplicaciones de tipo conversacional. También se introdujo el concepto del Diseño de Cross Layer y se enfatizó la importancia de una arquitectura robusta, todo lo anterior englobado dentro de una tecnología inalámbrica de tercera generación (ver anexo D) disponible actualmente en el país (cdma2000 1xEV-DO Rev. 0) y que soporta aplicaciones multimedia de tipo conversacional como la videoconferencia.

Materiales y Métodos

Ya se ha establecido previamente que la capa de Aplicación no tiene los mecanismos para conocer las condiciones del canal a las que se enfrentan las capas inferiores y que, en general, las aplicaciones generan tráfico sin importar si existen pérdidas u otra clase de problemas durante la transmisión. En particular, las aplicaciones multimedia de tipo conversacional (sensibles al retardo), percibirán los problemas que se presenten en el canal como pérdidas de paquetes. La estrategia de CLD propuesta facilita el que la aplicación pueda adaptarse a los cambios en las prestaciones de red, utilizando la estrategia basada en una aproximación híbrida enfocada a capas intermedias del modelo OSI, donde los parámetros formales de cada capa puedan ser evaluados para tomar una decisión en el ajuste de parámetros en la capa de Aplicación.

En este capítulo se expondrá el Diseño de Cross Layer propuesto que permite a la parte móvil de una videoconferencia conocer las condiciones del canal inalámbrico y adaptar en función de esto, los parámetros con los que genera los flujos de video a ser transmitidos en el enlace RL. Se disertará sobre la arquitectura utilizada y sus beneficios enfatizando la independencia del diseño propuesto a la tecnología de radio utilizada. Se presenta también la formalización matemática del Diseño de Cross Layer y la función objetivo propuestos, así como los métodos utilizados para los procesos de abstracción de parámetros.

III.1 Diseño de Cross Layer Propuesto

El Diseño de Cross Layer que se propone en este trabajo de tesis hace uso de la arquitectura mencionada en (Choi et al. 2004), arquitectura que se compone de tres procesos o módulos:

- Abstracción de Parámetros,
- Optimización de Cross Layer y,
- Distribución de Decisiones.

Partiendo del análisis de parámetros de QoS del capítulo II y considerando aplicaciones de tipo conversacional, es posible distinguir que el retardo, el jitter y la tasa de pérdida de paquetes están en función del Th ofrecido por el canal en un momento determinado. Lo anterior permite establecer que si las aplicaciones conocieran el Th que las capas inferiores pueden ofrecer y se adaptaran a este parámetro, los demás parámetros de QoS mejorarían como consecuencia directa del ajuste hecho.

El Diseño de Cross Layer que se propone, busca que las aplicaciones tengan conocimiento del estado del canal y por ende, del Th que pueden ofrecer, para adaptar el tráfico que generan a la capacidad del enlace. En otras palabras, el diseño busca adaptar el Th generado por la capa de aplicación, al Th ofrecido por la capa de Radio⁷. Pudiera pensarse en este punto que, cómo se pretende señalar parámetros propios de la capa de Radio a la capa de Aplicación, se está hablando de una aproximación de abajo hacia arriba. Esto no es estrictamente verdadero ya que la aproximación está definida por el proceso de abstracción de parámetros y la posterior distribución de decisiones; se mostrará en las secciones siguientes, que el diseño propuesto no cumple con las características necesarias para ser identificado como tal.

Ante el objetivo establecido en el párrafo anterior (i. e., adaptar el Th generado por la aplicación al Th ofrecido por el canal), y considerando que la estrategia de llevar el análisis de CLD a la capa intermedia, para este caso la capa de Transporte del modelo OSI, surgen cuestionamientos que deben ser respondidos por el Diseño de Cross Layer:

⁷ En la literatura se habla de la capa de Radio cuando se hace referencia a las capas PHY y MAC de manera simultánea, lo anterior en virtud de que son éstas capas las que forman la parte de radio de las comunicaciones y en general, una tecnología inalámbrica debe especificar el funcionamiento de la comunicación en lo que respecta a las capas mencionadas.

- ¿Cómo se relacionan y cuantifican las pérdidas en el canal inalámbrico con respecto al T_h ofrecido por la capa de Radio?
- ¿La variación de qué parámetros de la videoconferencia modifican, y cómo lo hacen, el T_h que se demanda al canal?

La respuesta a las preguntas planteadas no es trivial y forma la base del proceso de Abstracción de Parámetros de Cross Layer. Se debe establecer entonces, que la abstracción de parámetros de las diferentes capas que participan en el proceso, esté en función de la meta de diseño y que ésta pueda responder de forma convincente a los cuestionamientos anteriores.

El proceso de Optimización de Cross Layer involucra la toma de una decisión respecto a los parámetros que las diferentes capas pueden modificar. En la propuesta realizada, la única capa que modificará parámetros posteriormente a la optimización, será la capa de Aplicación. El módulo Optimizador de Cross Layer deberá decidir de entre un conjunto de parámetros de capa de Aplicación, cuál es el que cumple la meta de diseño. En otras palabras, el optimizador debe determinar cómo se debe comportar la aplicación para poder adaptarse a las prestaciones que ofrecen las capas inferiores en un instante de tiempo. Tanto los parámetros de capa de Aplicación como los parámetros de la capa de Radio se encuentran a disposición del Optimizador y en este proceso son identificados como parámetros de Cross Layer (i. e., cuando los parámetros abstraídos de las diferentes capas se encuentran dentro del optimizador, son conocidos como parámetros de Cross Layer). El optimizador deberá tomar una decisión en función de los parámetros que alimentan la función objetivo que se busca optimizar. La optimización de la función objetivo arroja un grupo de parámetros que garantizan que la meta de diseño se cumple y que se cumple de manera óptima.

Una vez que el módulo Optimizador de Cross Layer obtiene los parámetros que optimizan la función objetivo, los arroja como una salida y éstos deben ser señalizados a la capa de

Aplicación. El módulo que se encarga de notificar a la capa de Aplicación de los ajustes que debe realizar es el Distribuidor de Decisiones. El distribuidor debe realizar un proceso inverso al proceso de abstracción de parámetros dado que la salida del optimizador es un parámetro de Cross Layer, i. e., un parámetro abstraído. En el diseño propuesto, tal parámetro es un valor de T_h que la aplicación debe generar, lo cual implica que el codificador de video debe modificar algunos de los parámetros con los que codifica el video: tasa de cuadros por segundo o formato de resolución.

Cuando la aplicación recibe la decisión tomada, ajusta los parámetros del codificador, modificando de esta forma, el T_h que se genera. El T_h generado está en función del T_h ofrecido por las capas inferiores y como consecuencia los parámetros de QoS son ajustados. En otras palabras, se realiza un ajuste automático de los parámetros de calidad de servicio.

En la figura 6 se muestra la arquitectura del Diseño de Cross Layer propuesto, es posible observar los módulos o elementos de la arquitectura.

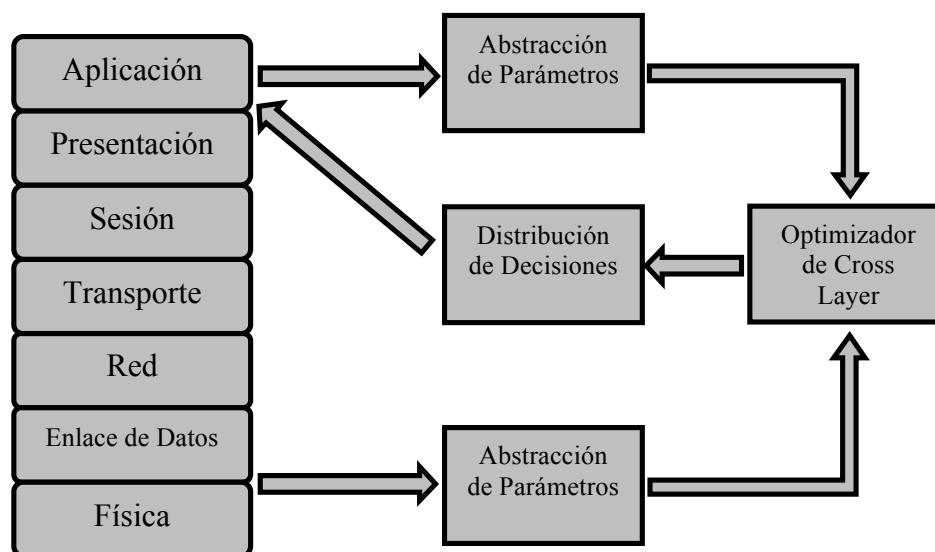


Figura 6. Arquitectura del Diseño de Cross Layer: Abstracción de Parámetros, Optimizador de Cross Layer y Distribución de Decisiones.

III.1.1 Abstracción de Parámetros de Capa de Radio

El proceso de abstracción de parámetros de las capas de Radio tiene como objetivo hacer inteligibles a capas superiores, parámetros propios de las capas de Enlace de Datos y Física.

En todo proceso de abstracción de parámetros de Cross Layer, se deben escoger tanto los parámetros a ser abstraídos como la abstracción per se, respecto a la meta de diseño planteada. Dado que la meta de diseño es una adaptación del Th generado por la aplicación al Th ofrecido por el radio, el resultado de la abstracción debe ser el Th ofrecido por las capas de Radio como tal, o bien, un parámetro que permita obtener dicho Th.

Todas las tecnologías de comunicación a nivel de capa física ofrecen un Th máximo en los diferentes enlaces (ascendente o descendente), no está de más enfatizar que dicho Th es teórico, desde el punto de vista de que en la práctica rara vez se llegan a alcanzar esos caudales eficaces. Suponiendo que sólo existiera un nodo en la red y que tenga información para transmitir suficiente tal que, desea y puede utilizar el Th máximo, el Th ofrecido por las capas inferiores puede ser, teóricamente, el Th máximo. Si la aplicación genera, por ejemplo, 144 kbps y el Th máximo que ofrece la tecnología es 144 kbps, es claro que el nodo transmitirá 144 kbits en un segundo y se espera que en la recepción, se reciban los 144 kbits en un segundo más el retardo que sea introducido por la propagación de la señal en los medios físicos que apliquen. Si existe más de un nodo en la red intentando transmitir, entrarán en función los mecanismos de acceso múltiple, mecanismos que en algunos casos generarán procesos de encolamiento; si la red se encuentra congestionada las colas crecerán y los retardos que experimentarán las aplicaciones serán mayores. Como ya se mencionó, en el caso de las aplicaciones conversacionales, esto se verá reflejado como una pérdida de paquetes.

Si por el contrario, la aplicación genera 144 kbps y el Th que puede ofrecer la tecnología en un momento dado es de 100 kbps, los 44 kbps restantes entrarán en procesos de encolamiento. Muy probablemente, después de pasada cierta cantidad de tiempo, llegará la

información completa al nodo destino, pero el retardo que experimentaron los 44 kbps que no pudieron ser enviados cuando fueron generados podría hacer que la capa de Sesión descartara esos paquetes con retardo. Por el contrario, si la aplicación tiene conocimiento de que las capas inferiores sólo pueden ofrecer 100 kbps y ajusta el Th generado a ese valor, no habrá información en colas de espera y ésta no se perderá por ese motivo.

Es evidente que el Th que las tecnologías de transmisión ofrecerán a las capas superiores depende de múltiples factores que están ubicados en dos diferentes capas: Física y Enlace de Datos. A nivel de capa Física los factores que impactan en el Th que se ofrece son (entre otros) los esquemas de modulación, las tasas de corrección de errores de los FEC y las condiciones del canal. A nivel de capa de Enlace de Datos están las técnicas de acceso múltiple y el manejo de las colas, calidad de servicio, entre otras. La forma en la que se manejan todos estos parámetros es dependiente de la tecnología en cuestión.

Sin embargo, desde el punto de vista de la capa de Red, lo único que puede percibirse como consecuencia de todos los parámetros que manejan las capas inferiores, es que algunos paquetes enviados llegarán a su destino y otros no llegarán, traducándose esto en una tasa de pérdida de paquetes:

$$PLR = 1 - \left(\frac{\text{Paquetes Recibidos}}{\text{Paquetes Enviados}} \right) \quad (2)$$

De la expresión anterior, es posible ver que cuando todos los paquetes que son enviados se reciben, la tasa de pérdida de paquetes es cero. Es deseable que los parámetros abstraídos no sean dependientes de la tecnología en la que subyacen (Choi et al. 2004 y Choi et al. 2006) debido a que esto afecta la modularidad de la arquitectura, por lo tanto, una buena abstracción de los problemas que enfrenta la capa de Radio es la PLR a nivel de capa de Red. La PLR ofrecida estará en función de los problemas que enfrenten las capas inferiores y de la forma en la que éstas los resuelvan. Es claro entonces que algunas tecnologías ofrecerán mejores tasas de pérdida de paquetes que otras ante las mismas circunstancias.

En este trabajo de tesis se optó por un enfoque experimental en los procesos de abstracción de parámetros (la abstracción de capa de Aplicación se trata en la sección siguiente, pero el enfoque es igualmente experimental) por lo que la capacidad de medir en entornos reales los parámetros abstraídos y antes de la abstracción fue fundamental. El enfoque práctico-experimental facilita enormemente la generación de valor del proyecto científico en el corto plazo, dado que la implementación es más susceptible de llevarse a cabo.

Los trabajos de campo para lograr la abstracción a nivel de capa de Red estuvieron basados en la tecnología que ofrecía el proveedor de servicios, esto implica que el módulo de Abstracción de capa de Radio es específico de esa tecnología (IS-856). Los parámetros medidos antes de la abstracción están relacionados con la velocidad del móvil, su entorno, y las características del medio; siendo éstos el máximo desplazamiento Doppler (el cual define el espectro Doppler de Jakes) y la potencia percibida por el móvil en el enlace FL.⁸

En resumen, se requiere un modelo que relacione la velocidad del móvil y su percepción del estado del canal inalámbrico con la tasa de pérdida de paquetes (PLR) a nivel de capa de Red. Dicho modelo es propiamente el módulo de abstracción de parámetros de capa de Radio dado que permite relacionar parámetros propios de las capas inferiores con parámetros que no pertenecen a ellas y que son más inteligibles por capas superiores; en este caso el parámetro abstraído es la PLR que se experimenta a nivel de capa de red en función de la velocidad del móvil, las condiciones del canal y las contramedidas que IS-856 utiliza para contrarrestar las dos primeras.

Es claro que la existencia previa de un modelo con las características requeridas simplificaría el proceso de abstracción de parámetros y evitaría el trabajo de campo para obtenerlo, sin embargo, dado que la PLR que las capas de Radio ofrecen a la capa de Red está en función de las contramedidas que la tecnología de radio implementa para lidiar con

⁸ El que se considere la potencia percibida por el móvil en el Forward Link para evaluar pérdidas en el Reverse Link obedece más bien a las limitaciones del hardware disponible. Ver anexo B.

el canal inalámbrico, se requiere un modelo específico que ofrezca la PLR que experimenta EV-DO Rev. 0 (IS-856) en función de la velocidad y las condiciones del canal.

Este equipo de investigación no tiene conocimiento de que tal modelo exista o haya sido descrito en la literatura del estado del arte. Inferimos al respecto que esta condición se genera porque dicho modelo es específico para el trabajo requerido, i. e., responde a un proceso de abstracción de parámetros de Cross Layer de una propuesta que no ha sido descrita con anterioridad; de aquí la necesidad de generarlo de manera experimental.

Conviene anotar tal como se indica en (Rappaport, 1996) que “los modelos de propagación se han enfocado tradicionalmente en la predicción de la intensidad de la señal recibida a una distancia determinada desde el transmisor, así como la variabilidad de la intensidad de la señal en la proximidad cercana espacial de un lugar particular”.

Es cierto que existen o pueden existir aproximaciones teóricas que relacionen la PLR a nivel de capa de Red ante variaciones de potencia, de la relación señal a ruido (SNR) o de otros parámetros, sin embargo, no se tiene conocimiento de que tales aproximaciones (modelos) teóricas hayan considerado a IS-856 como la tecnología de radio que enfrenta al canal.⁹

Otro aspecto importante a considerar es la capacidad que el hardware existente tendría para medir efectivamente los parámetros que el modelo implicara, e. g., para una terminal móvil resulta difícil evaluar la SNR; los dispositivos actuales no señalizan al software que los controla este tipo de parámetros que son utilizados comúnmente en los modelos teóricos, simplemente porque no tienen la capacidad para obtenerlos. Se debe recordar que los dispositivos móviles están grandemente limitados en sus capacidades.

⁹ Aunque existieran, la forma de modelar el canal inalámbrico y los parámetros con los que se habría hecho, influirían enormemente en los resultados ofrecidos.

La figura 7 muestra el modelo obtenido experimentalmente, que relaciona la potencia que el móvil percibe en el enlace FL (parámetro que puede ser medido en la terminal móvil) con la PLR medida a nivel de capa de Red, así como la expresión matemática que describe el comportamiento de dicho modelo.

El hecho de que la tasa de pérdida de paquetes que se experimenta a bajas velocidades sea prácticamente cero ante cualquier nivel de potencia, implica que no se requiere optimizar el envío de datos en el enlace RL cuando el móvil se desplaza a velocidades iguales o menores a 20 km/hr, no siendo así cuando el móvil se desplaza a velocidades vehiculares. Los detalles de la obtención experimental del proceso de abstracción de parámetros de capa de Radio están disponibles en el Anexo B.

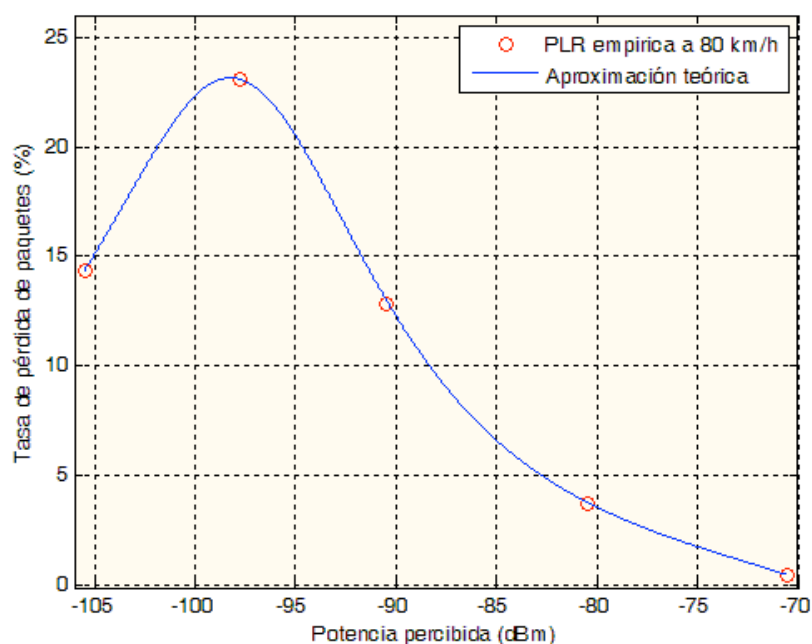


Figura 7. Modelo empírico obtenido que relaciona la potencia percibida por el móvil con la PLR en el enlace RL.

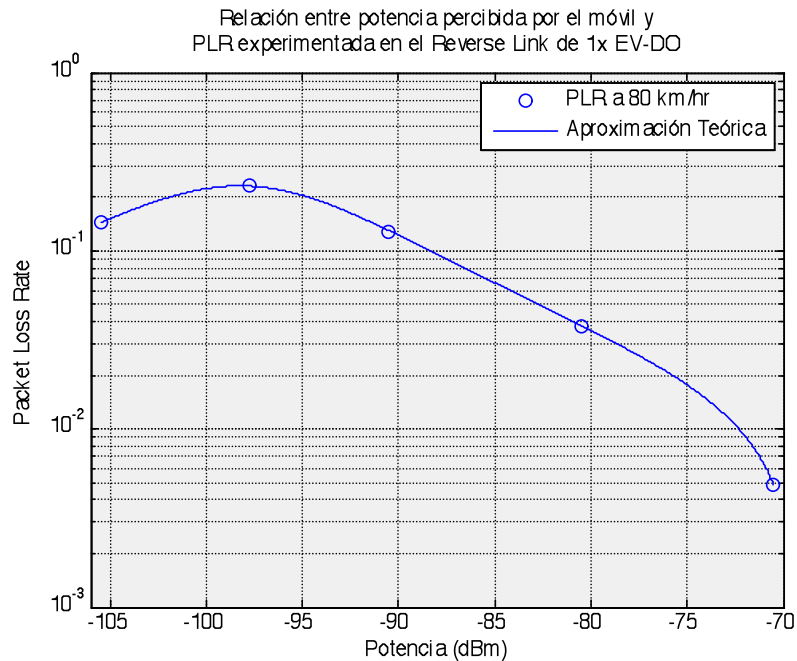


Figura 8. Modelo empírico obtenido que relaciona la potencia percibida por el móvil con la PLR en el enlace RL (Log).

Los círculos en la figura 7, representan los valores de la PLR promedio que fue medida en el enlace RL de una terminal móvil mientras se iniciaba, mantenía y terminaba una sesión de videoconferencia de 10 segundos de duración, aproximadamente, y la línea azul continua es el modelo matemático que corresponde a los puntos obtenidos a velocidades de 80 km/hr. Dicho modelo matemático nos permite estimar la PLR que se experimentará en el enlace RL a partir de la lectura de potencia que percibe el móvil. Dicho de otra manera, el modelo obtenido responde a las necesidades del proceso de abstracción de parámetros de capa de Radio requerido.

$$f(x) = \frac{2457x + 169580}{x^3 + 257.3x^2 + 21709x + 594487} \quad (3)$$

Llama la atención la caída en la pérdida de paquetes que se observa cuando la potencia medida es la menor. Aunque la discusión detallada al respecto se encuentra en el anexo B, se presenta en la figura 8 el mismo modelo expresado en escala logarítmica en el eje de las ordenadas, que es la forma convencionalmente utilizada en la literatura para representar los

modelos que relacionan variaciones de potencia o de otra índole con tasas de pérdida de paquetes. En esta figura es posible percibir que la curva presentada sigue un comportamiento similar a las curvas típicas que pueden ser encontradas en la literatura (Choi et al. 2006).

III.1.2 Abstracción de Parámetros de Capa de Aplicación

El proceso de abstracción de parámetros de capa de aplicación tiene como objetivo hacer inteligibles a las capas inferiores los parámetros propios de las aplicaciones.

En este trabajo, la abstracción de parámetros de la videoconferencia móvil (la aplicación de interés), sólo hará referencia a parámetros de video. Lo anterior debido a que, como se mencionó previamente, la optimización propuesta es sobre el flujo de información de video. Sin embargo, como también se verá, el Diseño de Cross Layer propuesto puede ser adaptado para optimizar el flujo de información de audio sin mayores problemas que el cambio del módulo de abstracción de parámetros.

Modificar el Th que genera la parte de video de una videoconferencia implica modificar parámetros propios del codificador. En este trabajo fue utilizada una aplicación de videoconferencia que permitía el uso del codificador de video H.263 (ITU-T Rec. H.263, 2005). En la literatura se establece que existen algunos parámetros clave en el comportamiento del tráfico que generan los codificadores de video, dichos parámetros son el formato de resolución (medido en pixeles por pixeles), la tasa de cuadros por segundo y el tipo de escena que codifica el codificador (Choi et al. 2006).

Para conocer la forma en la que la variación de los parámetros del codificador modifican el Th generado, se requiere una adecuada caracterización de las fuentes de tráfico, i. e., caracterizar el comportamiento del codificador ante un cambio en el formato de

resolución¹⁰ o una modificación en la tasa de cuadros por segundo (fps) que es considerada. Los detalles de la caracterización de fuentes de tráfico realizada durante este trabajo de tesis se presenta en el Anexo A.

Hacer inteligibles a las capas inferiores los parámetros de la aplicación, consiste en este caso, en interpretar el flujo de video generado como un Th que demanda la aplicación, i. e., un Th Generado. A nivel de capa de Red, los datos de capa de Sesión generados por RTP (Schulzrinne, et al. 2003), encapsulados en segmentos UDP (Postel, J., 1980), son encapsulados en paquetes IP (Postel, J., 1981) u otro protocolo de capa de Red. La capa inmediatamente inferior a IP percibirá el flujo de datos generado en la capa de aplicación, como una cierta cantidad de paquetes por unidad de tiempo, que la capa de Red genera.

Si la aplicación varía alguno de los parámetros con los que se está codificando el video, esta variación será percibida necesariamente, como una variación en el Th que se genera a nivel de capa de red. Esta abstracción es útil porque la capa de Radio será la encargada de enviar todos los datos que reciba de la capa de red, y se está buscando optimizar el envío de datos desde la terminal de acceso de DO Rev. 0 en función de lo que el radio perciba. Las pérdidas estarán en función de lo que tengan que transmitir las capas de Radio.

Para el codificador de video, modificar el formato de resolución implica una mayor o menor área de imagen a codificar. Es claro que a mayor tamaño de la imagen existe más cantidad de información a codificar, sencillamente por el hecho de que existen más macrobloques para ser codificados.¹¹ De la división de un cuadro (imagen fija) en macrobloques también se desprende el hecho de que diferentes tipos de escena generen diferentes tipos de tráfico a nivel de capa de Red (Choi et al. 2006). Algunos macrobloques

¹⁰ El formato de resolución es propiamente el tamaño del cuadro, se mide en pixeles por pixeles. Tamaños típicos de cuadro son CIF (Common Intermediate Format, 352x288 pixeles), QCIF (Quarter CIF, 176x144 pixeles), entre otros.

¹¹ Los macrobloques son bloques de 16x16 pixeles en los que es dividida una imagen para ser codificada. Cada macrobloque es codificado de manera individual (Stockhammer et al. 2003).

de la imagen presentaran variación temporal y otros no, esto implica que en algunas partes de la imagen existirá o no, movimiento temporal.

El hecho de que cierta cantidad de macrobloques no presente movilidad con respecto al tiempo, hace disminuir la cantidad de tráfico que genera la aplicación en un instante determinado porque los macrobloques que no presentan cambios con respecto a la fps (tasa de cuadros por segundo, por sus siglas en inglés), no son actualizados, i. e., no se codifican nuevamente (Stockhammer, et al. 2003). A mayor movimiento del video, más tráfico se genera debido a que más macrobloques presentan movimiento temporal y más información tiene que ser codificada y enviada. Cuando existe poco movimiento en el video, menos macrobloques presentan variación temporal, menos macrobloques necesitan ser codificados y se genera menos información.

En base a lo anterior, es posible encontrar en la literatura 3 tipos de escena que al ser codificados, generan tres tipos de tráfico diferentes (Choi, et al. 2006):

- Talking Heads,
- Talking Heads más movimiento de fondo y,
- Talking Heads con movimiento y movimiento de fondo.

Cabe señalar que en los artículos citados, se hace referencia a los anteriores tipos de escena por videos típicos utilizados por los grupos de investigación que trabajan en codificación de video. La secuencia que aquí ha sido mencionada como talking heads corresponde a la secuencia denominada Madre-Hija, la cual describe una conversación por videoconferencia entre una madre y su hija. Existe nulo movimiento del plano de fondo y muy poco movimiento del plano frontal (sólo se mueven los ojos y la boca de los interlocutores). La secuencia denominada Talking Heads más movimiento de fondo, corresponde a la secuencia conocida como Teléfono de Auto. Existe movimiento continuo del plano de fondo pero poco movimiento en el plano frontal. Finalmente, la secuencia que se ha nombrado como Talking Heads con movimiento y movimiento de fondo corresponde a la

secuencia conocida como Capataz. En esta secuencia un capataz de una obra en construcción toma una cámara y se auto graba mientras habla y muestra un edificio generando así una secuencia de video en la que existe un movimiento intenso del plano de fondo y un movimiento intenso del plano frontal también.

La fps es la medida del número de imágenes fijas que una cámara obtiene durante un segundo. La reproducción sucesiva de dichas imágenes es la que genera a la vista la sensación de movimiento. Cada imagen fija es descompuesta en macrobloques o grupos de éstos y cada macrobloque es posteriormente codificado. A medida que existen más imágenes fijas existen más macrobloques a ser codificados y por ende, más tráfico es generado.

De lo descrito en los párrafos anteriores es posible inferir que existen tres parámetros que impactan en la cantidad de tráfico que puede generar un codificador de video¹²: el formato de resolución, la fps y el tipo de escena. El módulo de abstracción de capa de Aplicación requerido debe especificar cuánta cantidad de tráfico es generado en función de los parámetros de codificación descritos. La cantidad de tráfico que genera la aplicación deberá estar dada en un parámetro que sea entendible por la capa de red y que sea compatible con el módulo de abstracción de capa de Radio mencionado en la sección previa. Más aún, la abstracción debe ser de tal forma que el proceso sea completamente reversible y entendible por la capa de Aplicación.

Bajo esta última condición, es claro que el tipo de escena es un parámetro que no puede ser considerado dentro del proceso de Distribución de Decisiones (el inverso a la Abstracción de Parámetros), debido a que no es posible que la aplicación modifique el tipo de escena que se está codificando. Es deseable, por lo tanto, que el tipo de escena no pertenezca a los parámetros a ser abstraídos en la capa de Aplicación.

¹² Cada codificador de video tiene características especiales que generan diferentes cantidades de tráfico, según las propiedades del codificador per se.

De la caracterización de fuentes de tráfico, descrita con detalle en el Anexo A, se pudo concluir que el tipo de escena impacta en la distribución de los datos generados, no así en los caudales eficaces máximos generados por la aplicación.

Esta conclusión reviste particular importancia debido a que en los procesos de codificación de video (en particular en H.263) no todos los paquetes generados son igualmente importantes. Específicamente, se debe dar máxima prioridad a los paquetes que contienen cuadros I (del inglés Intra-Frame) debido a que ellos contienen información de codificación clave sin la cual no es posible decodificar los paquetes subsecuentes (cuadros tipo P o tipo B). Los paquetes a nivel de capa de Red que contienen los cuadros I son los de mayor tamaño, son los que llevan la información más importante y, el tráfico que generan es el que debería ofrecerse en capas inferiores.

En otras palabras, es deseable que el Th Ofrecido por las capas de Radio esté ajustado al Th Generado por los paquetes que contienen a los cuadros I, Th que puede ser medido a nivel de capa de red. Para una aplicación que genera video, no es funcional el que las capas de Radio ofrezcan menos Th que lo requerido por sus cuadros I, debido a que la información que éstos contienen no llegará a tiempo (en el mejor de los casos) o simplemente no llegará y esto ocasionará que la información de video que llegue posteriormente (que sí podrá ser transmitida por el radio debido a que contiene menos información) no pueda ser decodificada por el receptor. El resultado obtenido al no llegar a tiempo los cuadros I o no llegar, es el mismo a que si se perdiera toda la información de codificación del cuadro completa. Es claro entonces, que se debe buscar satisfacer los caudales eficaces pico que generan las aplicaciones de video. Y dado que el tipo de escena no modifica el Th pico sino la distribución del tráfico generado, éste parámetro puede ser omitido del proceso de abstracción.

Para efectos del proceso de Abstracción de Parámetros de capa de Aplicación, esto implica que el Diseño de Cross Layer debe responder al Th Pico Generado por la aplicación. La

figura 9 muestra el resultado de la caracterización de fuentes de tráfico para el Th Pico Generado en función del formato de resolución y la fps.

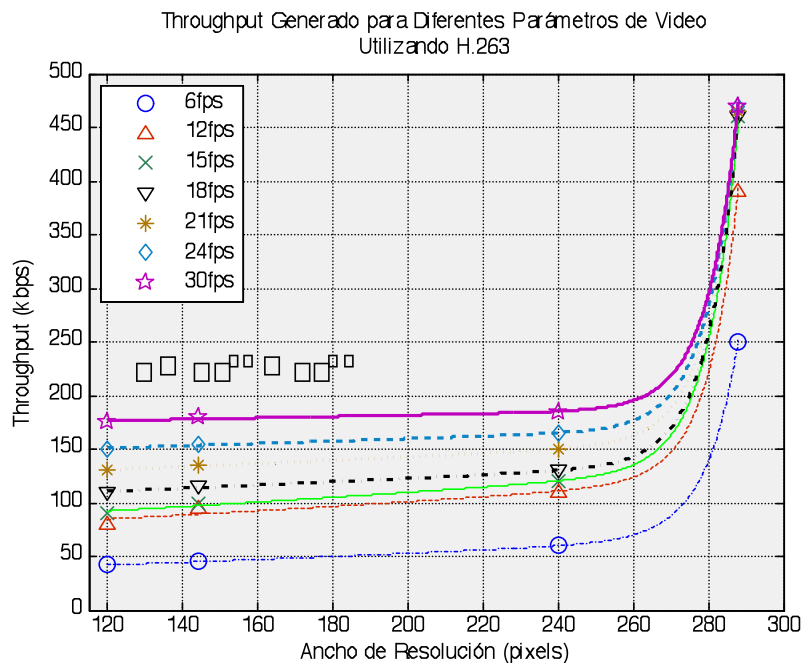


Figura 9. Th Pico Generado por H.263 en función de la fps y el ancho de resolución.

Los puntos mostrados en la figura corresponden al promedio de las mediciones realizadas y las líneas corresponden al modelo matemático también presente en la gráfica. A diferencia del proceso de abstracción anterior, en dónde se deseaba conocer o estimar la PLR para todo nivel de potencia, el modelo matemático que sigue al comportamiento de los puntos en esta abstracción es únicamente para mostrar el comportamiento exponencial que tiene el Th con respecto al tamaño de la imagen. En este caso no se desea conocer el Th Generado para cualquier formato de resolución con ancho entre 120 píxeles y 288 píxeles dado que los valores que toma el ancho del formato de resolución son discretos y están bien definidos; ahora se desea obtener un conjunto de valores de Th Pico Generados por la aplicación, en función de valores discretos de formato de resolución y de tasas de cuadros por segundo.

Este conjunto de valores de Th obtenido, es la Abstracción de Parámetros de capa de Aplicación per se. Existe una relación entre este conjunto y el conjunto de tasas de cuadro por segundo posibles y el conjunto de formatos de resolución posibles. La formalización de dichos conjuntos se presenta en la sección III.2.1.2.

Aunque su importancia será enfatizada posteriormente, conviene hacer notar aquí que el conjunto resultado de la abstracción deberá estar ordenado en forma ascendente. La regla de correspondencia entre el conjunto de parámetros a abstraerse y el conjunto de parámetros abstraídos deberá ser además una función biyectiva, entendiéndose por esto que no deben existir dos valores de Th Pico Generado iguales (en lo sucesivo Th Generado) que correspondan a diferentes valores de formato de resolución y fps y además, que todos los valores de Th Generado tengan una contraparte en el conjunto de parámetros a abstraerse.

Lo segundo, como se verá, estará dado por definición, mientras que lo primero deberá de asegurarse en todo proceso de Abstracción de Parámetros de capa de Aplicación a ser implementado en este Diseño de Cross Layer. También es conveniente destacar, aunque implícitamente ya se estableció, que la fuente de tráfico será considerada como una fuente que genera bits a una tasa constante (CBR, por sus siglas en inglés), el valor de la tasa constante es el mismo que el Th Generado en cuestión. Esta consideración asegura que todo el tráfico generado por la aplicación es tomado en cuenta al momento de realizar la optimización, porque todo el tráfico que no contenga cuadros tipo I es necesariamente menor en throughput, que el Th Generado (valor pico).

Lo anterior permite establecer que el tamaño de los paquetes a nivel de capa de Red, es también fijo y máximo, en correspondencia a los parámetros de codificación. Esta forma de visualizar el tráfico de capa de aplicación para un codificador de video H.263 no es del todo realista, puesto que como se muestra en el Anexo A, las distribuciones del tráfico generado corresponden más bien a un proceso markoviano ON-OFF. Sin embargo, esta aproximación demanda más recursos a las capas inferiores de lo que realmente requiere la aplicación; si satisfacemos la demanda de recursos en cuestión, se proporciona de manera implícita una

protección adicional a los datos no críticos del video y se garantiza la integridad del flujo de los mismos.

III.1.3 Propuesta de aproximación de Capas Intermedias

Ya ha sido establecido que la Abstracción de Parámetros de capa de Radio ofrece como resultado una PLR a nivel de capa de Red y que la Abstracción de Parámetros de capa de Aplicación ofrece como resultado un Th Generado, también a nivel de capa de Red.

El Optimizador de Cross Layer tomará una decisión en base a los parámetros abstraídos mencionados y buscará que el Th Generado por la aplicación sea adaptado en función de la PLR que se percibe. Como tal, la PLR es un valor adimensional. Sin embargo, desde el punto de vista de la capa de Red, es una relación entre los paquetes que se envían y los paquetes que se reciben, según lo descrito en la ecuación (2). Es conveniente, por lo tanto, traducir la PLR en un valor que no esté relacionado necesariamente a la PDU de capa de Red, sino más bien al Th Generado por la Aplicación.

El Th ofrecido por el canal (o por las capas de Radio, desde el punto de vista de la capa de Red) es siempre menor o igual al Th que genera la aplicación:

$$Th_{of} \leq Th_{gen} \quad (4)$$

En el mejor de los casos, cuando las condiciones del canal son óptimas, el Th Ofrecido es igual al Th Generado, nunca mayor, lo que implica que la tecnología no puede generar más datos de los que genera la aplicación¹³. Adicionalmente el Th Máximo Ofrecido está en función de las capacidades de la tecnología de transmisión. En ningún momento la

¹³ Es conveniente notar que cuando no se ha considerado la carga de los encabezados, la tecnología sí genera mayores caudales eficaces que la aplicación. En el caso de interés el caudal eficaz máximo que maneja la tecnología es efectivo, DO Rev. 0 soporta 200 kbps por sector como carga máxima, al usuario se le ofrecen únicamente 153.6 kbps, la diferencia en Th es lo que la aplicación consume por encabezados y señalización (Qualcomm, 2001).

tecnología puede ofrecer mayores caudales eficaces que su máximo, cuando las aplicaciones generan más tráfico del que la tecnología puede manejar se crean procesos de encolamiento, que es precisamente lo que se desea evitar.

Dado que se asume que las fuentes de tráfico utilizadas generan datos de manera constante, se puede establecer que los paquetes generados a nivel de capa de red son de longitud constante para el Th actual. Es posible entonces, interpretar un conjunto de paquetes de datos del mismo tamaño como un Th :

$$\frac{n \cdot P}{seg} = m \frac{bits}{seg} \quad (5)$$

Donde n es el número de paquetes de longitud P bits y m es la cantidad de bits resultantes. Si se generan nP paquetes en un segundo, se generan m bits por segundo. La ecuación (5) nos muestra que es posible entonces, visualizar un Th generado por la aplicación medido en kbps, como una cierta cantidad de paquetes por segundo si garantizamos que la tasa de bits es constante y los paquetes son de longitud fija, como es el caso.

Así, el Th ofrecido puede ser definido como:

$$Th_{of} = \begin{cases} Th_{gen} (1 - PLR) & \text{si } Th_{gen} \leq Th_{of} \\ Th_{max} (1 - PLR) & \text{si } Th_{gen} > Th_{of} \end{cases} \quad (6)$$

Dónde Th_{gen} es el Th generado por la aplicación en un momento determinado, Th_{max} es el Th máximo que la tecnología puede ofrecer y PLR es la tasa de pérdida de paquetes que experimenta la capa de Red en el mismo instante. Cuando la PLR es cero, el Th ofrecido es exactamente igual al Th generado y cuando el Th generado excede la capacidad de la tecnología, el Th ofrecido es simplemente igual al máximo que soporta la tecnología multiplicado por el complemento a uno de la PLR .

En base a la ecuación (6) también es posible establecer que la diferencia entre el Th ofrecido y el Th generado define la cantidad de información que se pierde en la capa de Radio:

$$Th_{lost} = Th_{gen} - Th_{of} \quad (7)$$

El objetivo del Diseño de Cross Layer es minimizar el valor que toma Th_{lost} en todo momento. Es decir es necesario minimizar el error tomando en cuenta los parámetros de aplicación donde la PLR se considera una variable independiente que afecta Th_{gen} por la aplicación.

En la figura 10 se muestra nuevamente la arquitectura de Cross Layer utilizada y se han agregado los parámetros que conforman las entradas y las salidas de los diferentes módulos del proceso.

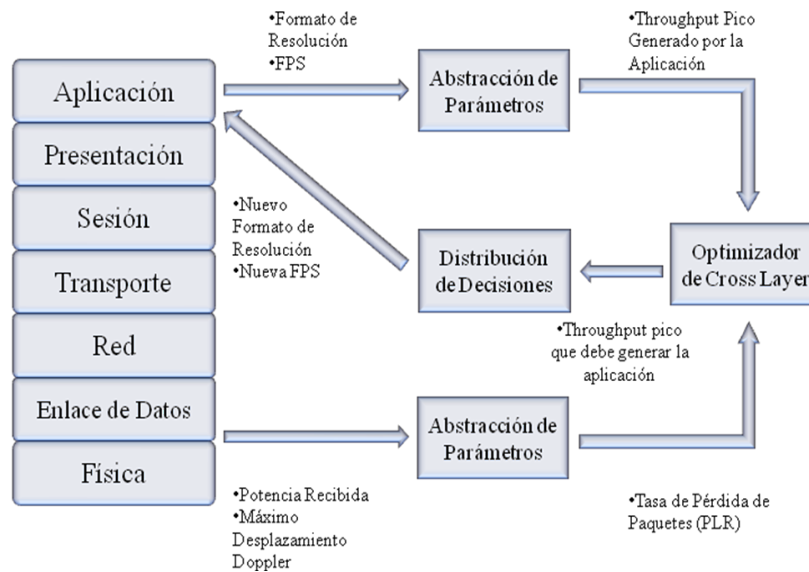


Figura 10. Arquitectura de Cross Layer y parámetros de entrada y salida

A continuación se presenta la abstracción de parámetros necesaria para poder llevar al cabo la formalización de Cross Layer de capas intermedias.

III.2 Modelado Matemático del Diseño de Cross Layer Propuesto

En esta sección se presenta la formalización matemática del diseño de Cross Layer que ha sido descrito de manera conceptual en las secciones previas. La primera parte de esta sección describe la necesidad de llevar a cabo el proceso de abstracción de parámetros en los diseños de Cross Layer seguida de la definición de los parámetros de capa de Radio y los de capa de Aplicación.

Posteriormente se definen los parámetros de Cross Layer y la función que realiza el optimizador de Cross Layer. La función objetivo también es definida así como el proceso de Distribución de Decisiones

III.2.1 Abstracción de Parámetros

El proceso de abstracción de parámetros es necesario para poder desarrollar e implementar cualquier de Diseño de Cross Layer en virtud de que se tomarán parámetros de las capas que tienen sentido únicamente dentro del contexto de la capa en cuestión. Como ejemplo se puede mencionar que la variación en la potencia percibida por el móvil, sólo tiene sentido a nivel de la capa Física, si es necesario considerar este valor en alguna otra capa diferente se debe realizar un proceso de abstracción que permita que la capa destino pueda interpretar de manera adecuada un valor de potencia determinado.

En función de lo anterior, es claro que el proceso de abstracción de parámetros debe estar en completa relación con el objetivo principal del Diseño de Cross Layer, dado que la capacidad de interpretar adecuadamente la información que permite optimizar alguna parte del proceso de comunicación, dependerá de una correcta abstracción. Los resultados del proceso de abstracción conforman las entradas que tiene el Optimizador de Cross Layer.

De no llevarse a cabo el proceso de abstracción, se enfrentarían dificultades en la comunicación de información específica de capas debido principalmente a la compatibilidad, el exceso de señalización y la estabilidad del sistema (Choi et al. 2006).

III.2.1.1 Parámetros de Capa de Radio

Tal y como ha sido establecido en las secciones previas, la abstracción de parámetros de capa de Radio evaluará parámetros que están relacionados con la velocidad de desplazamiento del móvil y el estado del canal inalámbrico que la capa de Radio percibe y los abstraerá en parámetros más comprensibles por las capas superiores. Se debe recordar que la intención es optimizar la transmisión de información sobre el enlace RL de un enlace IS-856.

Los parámetros a ser abstraídos de la capa de Radio son el máximo desplazamiento Doppler y la potencia que el móvil percibe en el enlace FL, ésta última como una medida de estimación de las condiciones del canal en general y debido a que es posible medir ambos parámetros con el hardware disponible.

La velocidad del móvil y el entorno que lo rodea determinan, junto a otros parámetros, la calidad del canal que perciben tanto la estación base como el móvil. Debido a la movilidad, dicha calidad del canal depende del tiempo, y en general éste es modelado como un proceso de desvanecimientos aleatorios. El estado del canal inalámbrico es considerado constante durante el llamado tiempo de decorrelación T_{dec} y se considera que pasado este tiempo las condiciones del canal cambian de manera abrupta a un valor completamente aleatorio (Choi et al, 2005).

El tiempo de decorrelación del canal depende de la velocidad del móvil y define el espectro Doppler. La forma de este espectro está en función de la distribución espacial de los interferentes que se encuentran en las cercanías de la terminal móvil y de la relativa

movilidad de éste con respecto a los interferentes. Una simplificación útil empleada de manera frecuente en el análisis teórico de la propagación de señales de radio, consiste en asumir que los interferentes están distribuidos uniformemente de manera circular alrededor del móvil de forma tal, que las fronteras de dichos interferentes se comportan como fuentes de difracción de la señal, generando multi-trayectos. La frecuencia con la que cada uno de estos multi-trayectos es percibida por el móvil, está en función del máximo desplazamiento Doppler que define, bajo las condiciones mencionadas, lo que se conoce como el espectro Doppler de Jakes. La variación en frecuencia del n -ésimo multi-trayecto percibido por el móvil está definido como

$$f_n = f_d \cos \alpha_n \quad (8)$$

donde f_d es el máximo desplazamiento Doppler y α_n es el ángulo de la circunferencia de interferentes del que proviene el n -ésimo multi-trayecto. El máximo desplazamiento Doppler está definido por

$$f_d = \frac{v \cdot f_c}{c} \quad (9)$$

donde v es la velocidad del móvil, f_c es la frecuencia de portadora y c es la velocidad de propagación de las ondas electromagnéticas en el aire. De las ecuaciones (8) y (9) es posible observar que la velocidad del móvil impacta directamente en el efecto de las multi-trayectorias que se presentan en el canal radio.

En la formalización matemática del Diseño de Cross Layer se seguirá el enfoque propuesto por Choi et al, (2006) y definiremos el conjunto

$$R = \{r_1, r_2, \dots\} \quad (10)$$

de tuplas $r_i = (P_i, f_{d_i})$ de parámetros específicos de capa de Radio r_i , donde P es la potencia percibida por el móvil y f_d es el máximo desplazamiento Doppler. En virtud de que estos parámetros pueden ser variables, el conjunto R contiene todas las posibles combinaciones de ellos y cada tupla r_i representa una posible combinación.

Para poder formalizar la abstracción de parámetros definimos el conjunto

$$\tilde{R} = \{\tilde{r}_1, \tilde{r}_2, \dots\} \quad (11)$$

de tuplas $\tilde{r}_i = (PLR_i)$ de parámetros abstraídos, donde PLR_i es la i -ésima tasa de pérdida de paquetes que ofrece el canal correspondiente a la tupla r_i de potencia y máximo desplazamiento Doppler. La relación entre el conjunto R de todas las posibles tuplas de parámetros de capa de Radio y el conjunto \tilde{R} de todas las posibles tuplas de parámetros abstraídos de capa de Radio está definida por la relación

$$B \subseteq R \times \tilde{R} \quad (12)$$

con dominio en R y codominio \tilde{R} y donde el operador \times denota el producto cartesiano. Se establece que B es una relación en la que *sólo* están incluidos *todos* los pares válidos (r_i, \tilde{r}_i) , i. e., se descartan pares ordenados como (r_1, \tilde{r}_2) o (r_2, \tilde{r}_1) y, B es un subconjunto de $R \times \tilde{R}$. Llamamos al conjunto B la abstracción de parámetros de capa de Radio.

Los pares ordenados que no son válidos, corresponden a las combinaciones en las que la potencia y el máximo desplazamiento Doppler, no corresponden a ese valor de PLR. En otras palabras, se define que para una determinada combinación de potencia y máximo desplazamiento Doppler, corresponde uno y sólo un valor de PLR.

La cardinalidad de los conjuntos R y \tilde{R} está determinada por la granularidad de los valores de potencia que pueden ser medidos en el móvil. Aunque la definición lo permite, el valor

de potencia del conjunto R no toma cualquier valor. En la práctica, el valor de la potencia corresponderá a números enteros que representarán valores de potencia en dBms medidos en el móvil. Los posibles valores de potencia que puede señalar el móvil¹⁴ determina la cardinalidad del conjunto R , ya que como se describe en el Anexo B, sólo se consideró una velocidad del móvil. De esta forma sólo existe un posible valor de máximo desplazamiento Doppler que corresponde a la frecuencia de portadora de DO Rev. 0 en México:

$$f_d = \frac{80\text{km/hr} \cdot 1.9 \times 10^9 \text{ Hz}}{3 \times 10^8 \text{ m/s}} = 140.74\text{Hz} \quad (13)$$

Tal y como se detalló en las secciones anteriores, las tuplas abstraídas son más fácilmente interpretadas por las capas superiores a la capa de Radio, en particular, por la capa de Red.

III.2.1.2 Parámetros de Capa de Aplicación

De acuerdo a lo establecido en la sección III.1.2 y de forma similar a la abstracción descrita en la sección anterior, definimos de manera formal el conjunto

$$A = \{a_1, a_2, \dots\} \quad (14)$$

de tuplas $a_i = (f_{res_i}, fps_i)$ de parámetros específicos de capa de Aplicación, donde f_{res} es el formato de resolución del video y fps es la tasa de cuadros por segundo con la que se codifica el video¹⁵. En virtud de que estos parámetros pueden ser variables, el conjunto A contiene todas las posibles combinaciones de ellos y cada tupla a_i representa una posible combinación.

¹⁴ A nivel de hardware, el móvil es una terminal de acceso DO Rev.0 que funciona de la misma manera que lo hace un módem. El módem puede señalar la potencia que está percibiendo como respuesta a una petición explícita. Las terminales de acceso utilizadas para el desarrollo de esta tesis pueden señalar de manera explícita hasta 35 valores del RSSI.

¹⁵ Existe una diferencia entre la fps con la que se capturan las imágenes y la fps con la que se codifica el video.

Definimos además, el conjunto

$$\tilde{A} = \{\tilde{a}_1, \tilde{a}_2, \dots\} \quad (15)$$

de tuplas $\tilde{a}_i = (\text{Th}_{\text{gen}_i})$ de parámetros abstraídos, donde Th_{gen_i} es el i -ésimo Th generado que la aplicación puede ofrecer, correspondiente a la tupla a_j de formato de resolución y tasa de cuadros por segundo. La relación entre el conjunto A de todas las posibles tuplas de parámetros de capa de Aplicación y el conjunto \tilde{A} de todas las posibles tuplas de parámetros abstraídos de capa de Aplicación está definida por la relación \tilde{a}_i

$$\Delta \subseteq A \times \tilde{A} \quad (16)$$

con dominio en A y codominio \tilde{A} y donde el operador \times denota el producto cartesiano. Se establece que Δ es una relación en la que *sólo* están incluidos *todos* los pares válidos (a_i, \tilde{a}_i) y que además, Δ es una función biyectiva. Llamamos al conjunto Δ la abstracción de parámetros de capa de Aplicación.

La cardinalidad del conjunto A se puede definir de manera explícita debido a que todos los elementos de las tuplas a_i toman valores discretos. Si se consideran siete tasas de cuadros por segundo y cuatro formatos de resolución, de acuerdo al anexo A, las tuplas del conjunto A son

$$\begin{aligned} a_1 &= (160 \times 120, 6 \text{ fps}) \\ a_2 &= (160 \times 120, 12 \text{ fps}) \\ a_3 &= (160 \times 120, 15 \text{ fps}) \\ &\vdots \\ a_{27} &= (352 \times 288, 24 \text{ fps}) \\ a_{28} &= (352 \times 288, 30 \text{ fps}) \end{aligned} \quad (17)$$

de dónde es posible determinar que la cardinalidad del conjunto es

$$|A| = 28 \quad (18)$$

y por ende, la cardinalidad del conjunto Δ es igualmente 28.

Las combinaciones de formatos de resolución y tasas de cuadro por segundo, corresponden a uno, y sólo un, Th generado específico. Esta condición (que Δ sea una función y además sea biyectiva) es necesaria debido a que durante el proceso de Distribución de Decisiones se invierte el proceso de abstracción y si algún valor de Th generado correspondiera a más de una combinación de formato de resolución y fps, existiría una ambigüedad en el ajuste que se debe realizar y el proceso completo se volvería inestable.

III.2.1.3 Parámetros de Cross Layer

Los conjuntos de parámetros abstraídos \tilde{R} y \tilde{A} de capa de Radio y Aplicación respectivamente, forman la entrada del Optimizador de Cross Layer y como cualquier combinación de las tuplas de parámetros abstraídos es válida, es conveniente definir el conjunto

$$\tilde{X} = \tilde{R} \times \tilde{A} \quad (19)$$

que integra los dos conjuntos de entrada del optimizador en uno solo y que está compuesto de las tuplas $\tilde{x}_n = (\tilde{r}_i, \tilde{a}_i)$. En este punto es importante destacar que la cardinalidad del conjunto \tilde{X} crece de forma exponencial con respecto al número de parámetros de Cross Layer, e. g., si todos los parámetros n de Cross Layer tienen un número fijo de valores posibles m , la cardinalidad de \tilde{X} es q^n . Debido a esto, el número de parámetros abstraídos debe mantenerse tan bajo como sea posible, la función objetivo tendrá que tomar una

decisión con respecto a las tuplas que conforman el conjunto \tilde{X} y puede darse el caso de que el valor que optimiza la función tenga que ser encontrado mediante búsqueda exhaustiva. Los tiempos de respuesta del Diseño de Cross Layer pudieran llegar a ser un problema en aplicaciones sensibles al retardo si éstos se vuelven muy grandes, debido a que el Optimizador tardaría mucho en encontrar el valor óptimo dentro de un conjunto \tilde{X} de cardinalidad grande.

Se debe notar también, la diferencia en la granularidad de los conjuntos de parámetros abstraídos \tilde{R} y \tilde{A} . Mientras que el conjunto \tilde{R} puede tomar teóricamente cualquier valor real (eso está en función de las combinaciones de los diferentes niveles de potencia y máximos desplazamientos Doppler), el conjunto \tilde{A} toma valores discretos. Más aún, el conjunto de parámetros específicos de capa de Aplicación A , sólo está definido para ciertos valores determinados por la aplicación.

Esta situación tiene repercusiones al momento de simular e implementar el comportamiento del Diseño de Cross Layer, debido a que los tiempos de respuesta de la aplicación influyen de manera directa en el desempeño de la optimización. La diferente granularidad de las capas de Aplicación y Radio pudiera llevar a la inestabilidad de las aplicaciones de video si éstas son forzadas a responder a las rápidas variaciones del canal inalámbrico. Los tiempos de respuesta y operación de la capa de Aplicación son considerablemente mayores a los tiempos de respuesta y operación de las capas de Radio.

En el primer caso podemos hablar de granularidad del orden de segundos, mientras que el segundo caso la granularidad depende, inclusive del tiempo de decorrelación del canal, que para velocidades de 80 km/hr a una frecuencia de portadora de 1.9 GHz es de apenas 2 ms. La granularidad es más fina en las capas de Radio que en la capa de Aplicación.

III.2.2 Optimizador de Cross Layer

Redefinimos el conjunto \tilde{A} como

$$\tilde{A} = \{\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{j-1}, \tilde{a}_j, \tilde{a}_{j+1}, \dots, \tilde{a}_{k-1}, \tilde{a}_k\} \quad (20)$$

de forma tal que

$$\tilde{a}_k > \tilde{a}_{k-1} \quad (21)$$

para todo valor de k y reescribimos el conjunto \tilde{R} como

$$\tilde{R} = \{\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{i-1}, \tilde{r}_i, \tilde{r}_{i+1}, \dots, \tilde{r}_{m-1}, \tilde{r}_m\} \quad (22)$$

notando que

$$|\tilde{R}| \neq |\tilde{A}| \quad (23)$$

y que por lo tanto, $i \neq j$. Definimos entonces la operación Ω realizada por el Optimizador de Cross Layer como

$$\Omega : \tilde{X} \rightarrow \tilde{Y} \subset \tilde{X}. \quad (24)$$

El Optimizador tiene como entradas todas las posibles tuplas de parámetros abstraídos de Cross Layer y tiene como salida el subconjunto no vacío \tilde{Y} . En lo sucesivo se asume que

$$|\tilde{Y}| = 1 \quad (25)$$

y que

$$\tilde{Y} = \{\tilde{x}_{opt}\} \in \tilde{X} \quad (26)$$

III.2.2.1 Función Objetivo

La salida \tilde{x}_{opt} del Optimizador de Cross Layer está definida respecto a una función objetivo particular

$$O: \tilde{X} \rightarrow \mathfrak{R} \quad (27)$$

donde \mathfrak{R} es el conjunto de los números reales y con regla de correspondencia

$$O(\tilde{x}) = \begin{cases} (\tilde{a}_{i+1}) & \text{if } \tilde{a}_i(1 - \tilde{r}_{\max}) \leq \tilde{a}_i(1 - \tilde{r}_i) < \tilde{a}_i \left(1 - \frac{2}{3} \tilde{r}_{\max}\right) \\ (\tilde{a}_i) & \text{if } \tilde{a}_i \left(1 - \frac{2}{3} \tilde{r}_{\max}\right) \leq \tilde{a}_i(1 - \tilde{r}_i) < \tilde{a}_i \left(1 - \frac{1}{3} \tilde{r}_{\max}\right) \\ (\tilde{a}_{i-n}) & \text{if } \tilde{a}_i \left(1 - \frac{1}{3} \tilde{r}_{\max}\right) \leq \tilde{a}_i(1 - \tilde{r}_i) < \tilde{a}_i(1 - \tilde{r}_{\min}) \end{cases} \quad (28)$$

donde \tilde{a}_i es el parámetro abstraído de capa de Aplicación que representa al Th generado por la aplicación en el estado actual, \tilde{r}_{\max} es el parámetro abstraído de capa de Radio que representa la máxima PLR que puede presentarse a nivel de capa de Red, \tilde{r}_i es el parámetro abstraído de capa de Radio que representa la PLR en el estado actual, \tilde{r}_{\min} es el parámetro abstraído de capa de Radio que representa la mínima PLR que puede presentarse a nivel de capa de Red y \tilde{a}_{i-n} es el parámetro abstraído de capa de Aplicación que garantiza que

$$\tilde{a}_i(1 - \tilde{r}_i) - \tilde{a}_{i-n} > 0. \quad (29)$$

Es necesario que la tupla \tilde{a}_{i-n} esté presente como una opción dentro de la función objetivo debido a que puede presentarse el caso de que el Th ofrecido $\tilde{a}_i(1-\tilde{r}_i)$ ¹⁶ sea menor que el estado inmediatamente inferior al actual \tilde{a}_i , i. e., que exista más de un posible Th generado dentro de las Regiones de Decisión.

Es conveniente recordar en este punto que las Regiones de Decisión (representadas en la ecuación (29) por las desigualdades) son determinadas a partir de los valores máximo y mínimo que la PLR puede presentar. Sin embargo y tomando en cuenta la disposición no lineal de los posibles valores de Th generado mostrados en la figura 9, pudiera darse el caso de que el Th ofrecido fuera menor que el Th generado actual. En tal caso, la función objetivo debería optar por indicar que se genere menos tráfico que el indicado en el estado anterior a \tilde{a}_i . Como pudiera presentarse el caso de que los estados $\tilde{a}_{i-1}, \tilde{a}_{i-2}, \tilde{a}_{i-3}, \dots$ también fueran mayores que el Th Ofrecido, se debe buscar y encontrar el valor \tilde{a}_{i-n} que es *menor* que el Th Ofrecido. La ecuación (29) define el momento en el que \tilde{a}_{i-n} es encontrado. Una vez localizado este valor, la función objetivo señala que ésa tupla es la óptima en ese instante¹⁷ y es enviada al Distribuidor de Decisiones para su eventual envío a la capa de Aplicación.

Los casos en los que se puede presentar que el Th Ofrecido $\tilde{a}_i(1-\tilde{r}_i)$ sea menor que el Th Generado en el estado inferior al actual \tilde{a}_{i-1} se presentan cuando existen altas pérdidas en el canal. Un valor elevado de PLR genera que el Th Ofrecido se aleje bastante del Th Generado, sobre todo si el Th Generado en ese instante es elevado, e. g., si se tiene un Th Generado de 180 kbps (\tilde{a}_i) y una PLR de 25% (\tilde{r}_i), de acuerdo a la expresión (6), el Th Ofrecido $\tilde{a}_i(1-\tilde{r}_i)$ sería de 135 kbps. Este valor de Th Ofrecido es numéricamente menor que varios estados previos al estado \tilde{a}_i actual, según la figura 9.

¹⁶ Ver la ecuación (4)

¹⁷ Es óptima en ese instante porque es la más cercana al Th Ofrecido.

Es conveniente enfatizar que el codominio de la función objetivo Ómicron es el conjunto de los números reales debido a que la función opera sobre los valores numéricos de los elementos de los conjuntos de parámetros abstraídos que tienen una representación dentro de los conjuntos \tilde{A} y \tilde{R} , la función objetivo no opera con tuplas, sino con los valores numéricos correspondientes a las tuplas.

La salida \tilde{x}_{opt} que arroja el Optimizador de Cross Layer corresponde a una tupla $\tilde{a}_i \in \tilde{A}$, misma que fue encontrada en respuesta a la regla de correspondencia de la función Ómicron descrita en la ecuación (28). Al observar esta ecuación, se puede notar que los posibles valores de $O(\tilde{x})$ son valores numéricos que tienen una representación relativa al estado actual \tilde{a}_i del Th Generado.

También debe hacerse énfasis en la importancia de las expresiones (20) y (21) para llevar a cabo una decisión adecuada respecto al Th que se deberá generar. El ordenamiento especificado en la ecuación (21) garantiza que el estado inferior a \tilde{a}_i , i. e., \tilde{a}_{i-1} , es numéricamente menor al valor de \tilde{a}_i . Debido a que la relación Δ de la expresión (16) es una función biyectiva¹⁸, se garantiza que el proceso inverso a la Abstracción de Parámetros (Distribución de Decisiones) relacionará de manera adecuada los caudales eficaces \tilde{a}_j arrojados por el Optimizador de Cross Layer, con las tuplas correspondientes $a_i = (f_{res_i}, f_{ps_i})$, sin importar el orden establecido en las ecuaciones (20) y (21).

Lo anterior significa que el ordenamiento del conjunto \tilde{A} carece de sentido fuera de la operación Ω de la ecuación (24) que define al proceso de Optimización de Cross Layer, sin que esto afecte la señalización que realiza la Distribución de Decisiones a la capa de Aplicación.

¹⁸ I. e., es una función inyectiva y también suprayectiva.

De esta forma se propone el uso de un Optimizador de Cross Layer que tiene como entradas la PLR percibida en ese instante, la cantidad de tráfico que está generando la aplicación y el conjunto de posibles valores que puede tomar el Th Generado. La parte central del Optimizador es una función objetivo que se desea optimizar para generar la menor cantidad de pérdidas posibles. Dicha función objetivo lee los valores actuales del Th Generado y el Th Ofrecido y evalúa la posición del Th Ofrecido con respecto a ciertas regiones de decisión para determinar si la aplicación debe generar menos tráfico, si debe seguir generando el mismo tráfico o, si puede generar más tráfico. Las regiones de decisión están definidas con respecto a la mínima y máxima PLR que se puede presentar en el modelo que relaciona la tasa de pérdida de paquetes con la potencia percibida por el móvil.

Si el Th Ofrecido se encuentra en la región de decisión inferior, la función objetivo indica que la aplicación debe generar menos tráfico, si el Th Ofrecido se encuentra en la región de decisión intermedia se le indica a la aplicación que mantenga el mismo tráfico y si el Th Ofrecido se encuentra en la región de decisión superior, la función objetivo le señala a la aplicación que puede generar más tráfico.

El módulo Optimizador de Cross Layer tiene entonces como salida el Th que debe generar la aplicación. Este valor es la entrada del módulo de Distribución de Decisiones, el cual se encargará de realizar el proceso inverso a la Abstracción de Parámetros y convertirá el valor del Th que debe generarse en una fps y un formato de resolución, que serán finalmente señalizados a la aplicación para que realice los cambios.

De lo descrito hasta el momento en relación al Diseño de Cross Layer propuesto, es posible visualizar que se está realizando una abstracción de parámetros de capa de Radio con enfoque bottom-up y una abstracción de parámetros de capa de Aplicación con enfoque up-bottom. Sin embargo, ninguna de las abstracciones llega a las capas relacionadas en el proceso de optimización, i. e., la abstracción de capa de Radio no genera señalización hasta la capa de Aplicación y tampoco es cierto el sentido inverso. Ambas abstracciones están siendo llevadas a un nivel intermedio, que es la capa de Red. El Optimizador de Cross

Layer tomará decisiones en base a parámetros abstraídos que son propios de la capa de Red y el proceso de Distribución de Decisiones tomará parámetros abstraídos en capa de Red y los convertirá en parámetros propios de la capa de Aplicación.

Esta aproximación de Cross Layer no puede ser clasificada según las definiciones propuestas por Van Der Schaar y Shankar (2005), quizá podría mencionarse como una propuesta centrada en la capa de Red o simplemente una aproximación a capas intermedias. Si se observa el proceso como un todo, la intención es señalar parámetros de la capa de Radio a la capa de aplicación, por lo que podría considerarse esta aproximación como Bottom-Up en el largo plazo (Choi et al. 2006).

III.2.3 Distribución de Decisiones

El proceso de Distribución de Decisiones es el encargado de señalar a la capa de Aplicación la decisión tomada (óptima) por el Optimizador de Cross Layer. Siguiendo el formalismo presentado en las secciones anteriores, la salida del Optimizador

$$\tilde{x}_{opt} = (\tilde{r}_{opt}, \tilde{a}_{opt}) \quad (30)$$

debe ser señalizada a las capas correspondientes. En el caso de interés, el valor de \tilde{r}_{opt} es exactamente igual al de \tilde{r}_i porque la función objetivo sólo arroja como resultado una adaptación del Th \tilde{a}_i que debe generar la aplicación. La función objetivo no modifica la PLR que presenta el canal, que es formalmente el valor de \tilde{r}_i en un momento determinado.

De cualquier forma, la tupla $\tilde{a}_{opt} \in \tilde{A}$ debe ser comunicada a la capa de Aplicación y para ello la Abstracción de Parámetros debe ser revertida de forma tal que \tilde{a}_{opt} se convierta en $a_{opt} \in A$. Esta transformación está dada por la expresión

$$a_{opt} \in \{a \mid (a, \tilde{a}_{opt}) \in \Delta\} \quad (31)$$

y debido a que Δ es una función biyectiva, el valor de a que corresponde a \tilde{a}_{opt} es único y está definido.

En la práctica se deben generar mecanismos que permitan a la aplicación leer la señalización hecha por el Distribuidor de Decisiones y reconfigurar la codificación de video en función de los valores pertenecientes a la tupla a , adaptando así, el Th Generado por la Aplicación.

Hasta este punto ha sido descrito el Diseño de Cross Layer que se propone para optimizar la transmisión de video sobre el enlace RL de 1xEV-DO Rev.0, la validez de la propuesta se llevó a cabo a través de simulación, según lo descrito en la sección siguiente.

III.3 Validación del diseño de Cross Layer

La validación del Diseño de Cross Layer propuesto fue realizada a través de simulación. La caracterización de fuentes de tráfico descrita en el Anexo A y la obtención del modelo empírico de PLR en función de la potencia, descrito en el Anexo B, garantizan la fiabilidad de los datos que están siendo manejados dentro de la simulación. En otras palabras, no se asumen fuentes ideales de tráfico ni se consideran modelos teóricos para la estimación de la PLR que experimenta la capa de Red como consecuencia de los problemas que enfrentan las capas de Radio.

La confiabilidad de que las fuentes de tráfico siguen un comportamiento real, práctico; es un proceso fundamental para dar validez a cualquier simulación. En la investigación y caracterización de las fuentes de tráfico realizadas, pudimos observar que el tipo de escena no modifica los caudales eficaces pico, generados por la aplicación. Esta observación reviste fundamental importancia debido a que no se encuentra publicada en las referencias

consultadas durante el proceso de revisión bibliográfica¹⁹ y es crítica (la observación *per se*) al momento de visualizar la posible distorsión que sufriría la aplicación si se perdiera la información contenida en los cuadros I (Choi et al. 2005), que a su vez está contenida en los paquetes de capa de Red de mayor tamaño, que son los que generan mayor Th, i. e., donde se encuentran los valores pico del Th Generado.

El comportamiento exponencial del Th Generado en función del formato de resolución (más que de la fps) por el codificador de video H.263, tampoco había sido reportado de manera explícita en la literatura consultada y es fundamental para el proceso de optimización del Th Generado. Su comportamiento define la forma de operación de la función objetivo Ómicron. A través de observar detenidamente la figura 5, se puede concluir que el impacto más grande en el Th que genera la aplicación está dado por el formato de resolución y no por la tasa de cuadros por segundo. Más aún, la fps modifica en mucha menor medida el Th Generado.

Estas consideraciones obtenidas a raíz de la caracterización de las fuentes de tráfico, nos permiten establecer con fundamento, la razón por la que se asume que el codificador de video genera tráfico como si fuera una fuente CBR, según lo descrito en la sección III.1.2. Esta consideración ofrece una protección adicional al tráfico no crítico del video (enviado en los cuadros tipo B y P por H.263) por parte del Diseño propuesto, y aunque su efecto no fue medido durante este trabajo de investigación, aseguraría un mejor rendimiento del decodificador de video.

Por otra parte, la obtención de un modelo empírico que relacione la PLR del enlace RL con la potencia percibida en el móvil y evaluada a través del RSSI (Indicador de Intensidad de la Señal Recibida, por sus siglas en inglés), garantiza que la simulación estará basada en el comportamiento *real* de la tecnología evaluada, en este caso, de cdma2000 1xEV-DO Rev. 0.

¹⁹ Sin pretender asumir que la revisión bibliográfica realizada es completa y total, sin embargo, la literatura citada habla de las propiedades de codificación más que del Th que se demanda a las capas inferiores y no se especifica la distribución del mismo.

Asumir modelos teóricos de tasas de pérdidas de paquetes en función de valores de potencia, siempre conlleva la incertidumbre de que el canal y la tecnología realmente se comporten de acuerdo a las estimaciones deterministas, sin mencionar que modelar el comportamiento del estándar IS-856 reviste en sí una dificultad grande.

Más aún, la abstracción requerida debía ser visualizada a nivel de capa de Red y esta capa es completamente independiente de los fenómenos propios de las capas inferiores a ella. Las contramedidas al canal empleadas por la tecnología en cuestión son fundamentales para conocer el estado de la red que *realmente* ofrecen las capas de Radio a las aplicaciones a partir de la capa de Red.

Los resultados de las mediciones realizadas mostraron un comportamiento atípico²⁰ de la PLR en el enlace RL que sin embargo, es continuamente asumido en la literatura que trata sobre la cantidad de información que se pierde en el canal inalámbrico. Comúnmente se asume que mientras menos información sea enviada al canal radio, menos probabilidad hay de que se pierda, y si ésta llega a perderse, la información perdida es mínima. En el caso de DO Rev. 0, cuando las condiciones del canal del enlace RL son severas, la tecnología envía menos bits por ranura (Qualcomm, 2001), limitando el Th que se ofrece a las capas superiores pero garantizando que se pierde menos información por ranura perdida.

Si adicionalmente, se toma en cuenta que el tamaño de los paquetes IP puede ser considerablemente mayor que el tamaño de las tramas a nivel de capa MAC, enviar menos datos por ranura (a nivel de capa Física) le permite al radio no perder todo el paquete completo e implementar mecanismos de tipo ARQ para la recuperación de errores a nivel

²⁰ En las figuras 3 y 4 es posible observar que a menor potencia se presenta una menor tasa de pérdida de paquetes, lo cual es teóricamente inesperado. En teoría, a menor potencia debería de existir una PLR mayor, pero esta consideración no toma en cuenta las contramedidas que la tecnología de radio en cuestión toma al respecto. Si la cantidad de información que se envía a nivel de capa PHY se mantiene constante a pesar de percibir un canal agresivo, es más probable que la PLR disminuya con respecto a la potencia.

de capa física, adicionalmente a los códigos correctores de errores; mismos que eventualmente, impactarán de manera positiva en la PLR percibida a nivel de capa de Red.

El comportamiento atípico de la PLR mostrado en las mediciones está plenamente justificado por lo mencionado anteriormente y sin embargo, los modelos teóricos rara vez toman en cuenta esta consideración. Adicionalmente fue posible concluir que DO Rev. 0 proporciona canales muy suaves a las capas superiores cuando no existe movilidad, condición que nos permitió descartar una optimización cuando las velocidades del móvil son bajas, e. g., cuando el móvil se encuentra en condiciones de tráfico denso.

Lo descrito en los párrafos anteriores le proporciona validez a la simulación, que considera una PLR más aproximada a la realidad que lo estimado por las predicciones teóricas.

Las simulaciones realizadas también consideran los tiempos de respuesta de la aplicación y evalúan el comportamiento en función de esta respuesta. También se ha considerado la validez temporal de la lectura de potencia obtenida cada vez que se toma una muestra de las condiciones del canal. Esta validez temporal está en función del tiempo de decorrelación del canal, que como ya se ha mencionado, está en función de la velocidad del móvil y considera los efectos de las multi-trayectorias generadas por el entorno en el que se encuentra la terminal móvil. Su impacto en la optimización también es considerado.

El hecho de que la simulación esté basada en una caracterización real de fuentes de tráfico y en un modelo que presenta la PLR para niveles de potencia que pueden ser medidos en la práctica en cualquier terminal de acceso DO Rev. 0, incrementa la validez de los resultados de ésta. De manera adicional, el tomar en cuenta parámetros reales, facilita la generación de valor del proyecto científico dado que su aplicación inmediata se vuelve factible. El diseño de Cross Layer propuesto y evaluado en simulación, puede ser implementado tal y como ha sido propuesto en virtud de que opera con valores que son medibles de manera inmediata en los entornos reales. Tanto el valor de la potencia como el de la velocidad del móvil (que define el máximo desplazamiento Doppler y éste el espectro Doppler de Jakes) son valores

que pueden ser medidos directamente, el primero en la terminal de acceso EV-DO Rev. 0 (a través del RSSI) y el segundo en el móvil. La caracterización de las fuentes de tráfico fue realizada tomando en cuenta las capacidades del codificador en cuestión (H.263) en secuencias de video reales.

Más aún, la modularidad de la arquitectura empleada asegura que el Diseño propuesto puede ser llevado fácilmente hacia otras tecnologías de radio, no sólo celulares, sino implementaciones con IEEE 802.16 o tecnologías de WLAN inclusive. El único requerimiento es proporcionar el módulo de Abstracción de Parámetros de capa de Radio de la tecnología en la que se pretenda evaluar el desempeño pudiendo ser éste, inclusive, un módulo obtenido teóricamente.

Lo descrito en el párrafo anterior es aplicable también a la abstracción de capa de Aplicación, evaluar el comportamiento de la simulación para otros codificadores de video como H.264 depende únicamente de que se provea al Diseño de Cross Layer propuesto con la abstracción adecuada, pudiendo haber sido obtenida ésta de manera experimental como en este caso, o teórica.

III.4 Resumen del Capítulo III

En este capítulo fueron descritos los materiales y métodos empleados para alcanzar el objetivo general planteado para este proyecto de investigación. En particular, se describió el Diseño de Cross Layer propuesto para lograr un ajuste automático de parámetros de QoS en videoconferencia móvil a velocidades vehiculares, considerando el enlace RL de una tecnología IS-856 (cdma2000 1xEV-DO Rev. 0).

Fue descrita la arquitectura empleada en el diseño y las ventajas de su modularidad así como los procesos de Abstracción de Parámetros de capa de Aplicación y de capa de Radio. Se formalizó también el diseño propuesto a través de un modelado matemático correspondiente a lo encontrado en la literatura del estado del arte, definiéndose

matemáticamente los procesos de Abstracción de Parámetros, Optimizador de Cross Layer y Distribuidor de Decisiones.

Finalmente se estableció la forma en la que son validados los resultados de la propuesta hecha a través de la simulación y se estableció la facilidad de generación de valor de este proyecto científico en función de la posibilidad de su inmediata aplicación.

Capítulo IV

Resultados y Análisis Numérico

En este capítulo se presentan los resultados de las simulaciones realizadas para evaluar y validar la propuesta de Diseño de Cross Layer; adicionalmente se expone un análisis numérico de los resultados obtenidos y se presentan gráficas que permiten evaluar y medir el desempeño del Diseño de Cross Layer propuesto.

Se evalúan las pérdidas que se generan a nivel de capa de Red a partir del modelo empírico que relaciona la PLR con las condiciones del canal y la movilidad cuando no se implementa el Diseño y cuando éste es aplicado, esto último aunado a las consideraciones de simulación y el análisis numérico, permiten validar la propuesta y los resultados de la simulación.

IV.1 Consideraciones de Simulación

Aunque ya fueron mencionadas en el capítulo anterior, es conveniente destacar las consideraciones que se tuvieron al realizar las simulaciones ya que de ellas depende la validación de la propuesta realizada.

Ciertamente, una de la partes más complejas en la simulación de tecnologías de comunicación inalámbricas es la simulación del canal radio (Rappaport, 1996). En entornos urbanos donde normalmente no existe trayectoria de visibilidad directa (LOS, del inglés Line of Sight) se presentan la propagación multi-trayectoria que implica los desvanecimientos de pequeña escala y los desvanecimientos de gran escala. Ya se ha establecido anteriormente que los modelos de propagación han centrado sus esfuerzos en

predecir la potencia de la señal a una cierta distancia de la radio base, dado que dichos modelos se utilizan comúnmente en la planificación celular.

En la simulación desarrollada no son utilizados modelos teóricos de propagación, debido principalmente, a que los modelos existentes no satisfacen las necesidades del Diseño propuesto. El Diseño de Cross Layer que se ha planteado requiere establecer una relación entre la PLR percibida a nivel de capa de Red en el enlace RL de un enlace IS-856 con las variaciones del canal y la velocidad de desplazamiento del móvil. En la simulación se ha considerado un modelo empírico que satisface los requerimientos planteados y los detalles de la obtención de dicho modelo se encuentran en el Anexo B.

La utilización de este modelo implica que en la simulación se está considerando un canal radio que es enfrentado con las contramedidas que utiliza cdma2000 1xEV-DO Rev. 0 y que al final de cuentas, arroja una PLR determinada a nivel de capa de Red. Todos los mecanismos propios de la propagación y la forma en la que estos son tratados por IS-856 están implícitos en el modelo empírico, reforzando así, la validez de los resultados obtenidos de la simulación.

Otra consideración que es conveniente destacar es la relativa a la caracterización de las fuentes de tráfico. Asumir fuentes de tráfico ideales, caracterizadas por su comportamiento estadístico o en base a la distribución que presentan, es una práctica común encontrada en la literatura consultada. Su utilización no es crítica si los parámetros a evaluar en la simulación no están relacionados directamente al desempeño de las fuentes. Sin embargo, en este trabajo de investigación los resultados de la simulación están estrechamente ligados al comportamiento de las fuentes. Las fuentes per se, definen el comportamiento de la simulación.

Del párrafo anterior se desprende la importancia de que las fuentes de tráfico consideradas sean lo más reales posibles. Esta conclusión nos llevó a caracterizar las fuentes de tráfico que serían empleadas en la simulación. Se caracterizó el comportamiento del Th a nivel de

capa de Red que genera un codificador de video H.263 en función del formato de resolución de los cuadros que codifica y, de la tasa de cuadros por segundo que debe codificar. Los detalles de la caracterización de las fuentes de tráfico se encuentra en el Anexo A.

Como resultado del proceso de caracterización se observó que el tipo de escena que está siendo codificada no impacta en el Th Pico Generado por la aplicación, i. e., los recursos máximos que la aplicación demanda a las capas inferiores no están en función de las condiciones capturadas por los cuadros que componen el video. La consecuencia inmediata de esta observación es fundamental para justificar el tratamiento que se da al tráfico generado por la aplicación. Debido a que las secuencias de video del tipo Talking Head generan distribuciones de tráfico caracterizadas por un proceso Markoviano ON-OFF, pudiera establecerse que éstas deben ser tratadas por la red de transporte²¹ de forma diferente que las secuencias que presentan alta movilidad del plano de fondo y del plano frontal (caracterizadas por una distribución prácticamente aleatoria y consideradas comúnmente como fuentes de tasa de bits variable (VBR, por sus siglas en inglés)).

Sin embargo, el hecho de que el tipo de escena no modifique el Th Pico Generado por la aplicación (que es dónde se encuentra la información de codificación más importante), implica que todas las secuencias de video deban ser manejadas de igual forma, i. e., se debe satisfacer la máxima demanda de Th para mantener al mínimo la distorsión que pueda presentar el video en la recepción.

La caracterización de las fuentes de tráfico arrojó además los valores pico promedio del Th que genera un codificador de video H.263 a nivel de capa de Red y que son los que alimentan al simulador, validando así, el tratamiento que se da a la información generada por la aplicación por parte del Diseño de Cross Layer simulado.

²¹ Por red de transporte debe entenderse todo el conjunto de dispositivos de comunicación y enlaces que están presentes en la transmisión de información de extremo a extremo.

Hasta este punto han sido analizadas las consideraciones de simulación relativas a las entradas del Diseño a ser evaluado. Las entradas del sistema (PLR a nivel de capa de Red y Th Generado por la aplicación) están basadas en mediciones reales, y los resultados de la simulación son por tanto, susceptibles de implementarse en el corto plazo.

Las simulaciones realizadas consideran un escenario en el que un dispositivo móvil transmite información de video durante 10 segundos a través del enlace RL de un enlace IS-856, el móvil se desplaza a una velocidad constante de 80 km/hr y las condiciones del entorno son tales que la potencia que percibe a través del valor del RSSI, van desde la mínima potencia promedio que puede detectar (-105.5 dBm), hasta la máxima potencia promedio que puede detectar (-70.5 dBm). La simulación considera que el móvil puede medir cualquier valor de potencia posible entre estos extremos, aunque en la realidad no es así. Esta última condición garantiza que existirá un valor de PLR a nivel de capa de Red para cualquier valor de potencia que el móvil pudiera percibir.

Asumiendo un tiempo de decorrelación del canal definido por

$$T_{dec} = 0.4 \cdot \left(\frac{\lambda}{v} \right) \quad (32)$$

donde λ es la longitud de onda de la señal transmitida y v es la velocidad del móvil²²; sustituyendo los valores se tiene

$$T_{dec} = 0.4 \cdot \left(\frac{c/f}{v} \right) = 0.002s \quad (33)$$

para una frecuencia de portadora de 1.9 GHz y una velocidad de desplazamiento del móvil de 80 km/hr. Considerando 10 segundos de duración de la secuencia de video, la variación de potencia mencionada anteriormente y el tiempo de decorrelación del canal, se establece

²² Ver la sección 2.1.1 del capítulo III en relación al tiempo de decorrelación del canal.

que el número de veces que debe ser muestreado el canal para asumir que la lectura es válida durante el tiempo de decorrelación está dada por

$$\text{Muestras} = \frac{\text{Duración de la Prueba}}{T_{dec}} = \frac{10s}{.002s} = 5000 \quad (34)$$

La ecuación (34) nos dice que deben existir 5000 lecturas de potencia y cada una de estas lecturas debe corresponder a un valor de PLR. Esta consideración es imprescindible para validar que la potencia que se está *midiendo* en la simulación es constante²³, i. e., si la simulación toma una lectura de potencia y en base a ella determina un valor de PLR, dicho valor de PLR se mantiene cierto, únicamente, durante el tiempo de decorrelación del canal.

El hecho de contar con un modelo matemático que describe el comportamiento de la PLR con respecto a la potencia, nos permite tener la cantidad de muestras que se requieran para cualquier valor de potencia comprendido entre -105.5 dBm y -70.5 dBm, pudiendo satisfacer lo requerido por la ecuación (34). El modelo matemático que describe la tasa de pérdida de paquetes que se experimenta en el enlace RL de DO Rev.0 con respecto a la potencia que percibe el móvil a través del valor del RSSI está indicado en la figura 3 y se reproduce aquí a través de la ecuación

$$f(x) = \frac{2457x + 169580}{x^3 + 257.3x^2 + 21709x + 594487}, \quad x \in [-105.5, -70.5]. \quad (35)$$

Condiciones adicionales de simulación incluyen las ya mencionadas en la sección 1.3 del capítulo III respecto a que el tamaño de los paquetes que genera la aplicación es constante. Este punto permite estimar de manera precisa la cantidad de bits que se pierden ante las variaciones del canal considerando también, el tiempo de respuesta de la aplicación a los

²³ Y en verdad lo es durante el tiempo de decorrelación del canal. Pasado este tiempo, el valor de la potencia cambia a otro valor completamente aleatorio.

cambios señalados por el Distribuidor de Decisiones y el período de muestreo de la potencia por parte del módulo de Abstracción de Parámetros de capa de Radio²⁴.

Es necesario señalar que la caracterización de fuentes de tráfico descrita en el Anexo A arrojó valores de Th Generado mayores a los que soporta el enlace RL de IS-856. Valores superiores al máximo Th que pueden ofrecer las capas de Radio no son considerados al momento de realizar las simulaciones. Considerarlos implicaría generar procesos de encolamiento a nivel de capa MAC que no fueron tomados en cuenta al momento de obtener el modelo empírico de PLR.

Tampoco se considera tráfico agregado en el nodo que genera el tráfico de video sobre el enlace RL, i. e., se asume que únicamente se están transmitiendo datos como consecuencia de la transmisión de video.

IV.2 Resultados de Simulaciones

Los resultados de las simulaciones y su análisis numérico se presentan en las dos secciones siguientes. En la primera de ellas se presentan los resultados de la simulación evaluando diferentes caudales eficaces generados por la aplicación y las pérdidas que enfrentan cuando no se implementa el Diseño de Cross Layer propuesto.

En la sección posterior se presentan los resultados de la simulación cuando se implementa la optimización y se evalúa el comportamiento que presenta el Th Generado respecto al Th Ofrecido para diferentes tiempos de respuesta de la aplicación a la señalización realizada por el Distribuidor de Decisiones. El mismo comportamiento es evaluado para diferentes tiempos de muestreo de la potencia por parte del proceso de Abstracción de Parámetros de capa de Radio. En este punto se analiza el estado del canal (evaluado como una PLR) que

²⁴ Las terminales móviles de DO Rev.0 tienen mecanismos para estimar las condiciones del canal en cualquiera de los enlaces, sin embargo estos mecanismos no son señalizados hacia las capas superiores por parte de las AT. El período de muestreo al que se hace referencia aquí es propiamente al intervalo de tiempo que transcurre entre cada vez que se le solicita a la terminal de acceso que señalice el valor que tiene del RSSI en ese instante.

percibe o asume la optimización y se compara con el comportamiento real que presentan, tanto el Th Generado como el Th Ofrecido.

IV.2.1 Resultados de la Implementación del Diseño de Cross Layer

Los resultados de la implementación del Diseño de Cross Layer fueron evaluados en función de la pérdida de información a nivel de capa de Red. Este parámetro de evaluación fue escogido dado que nos permite conocer qué tan cerca se encuentra el Th Generado del Th Ofrecido; el cual a su vez, refleja la meta de diseño propuesta desde un principio, misma que está implícita en la optimización que realiza la función objetivo como parte central del Optimizador de Cross Layer. Mientras el Th Generado se siga adaptando al Th Ofrecido se realizará un ajuste automático de los parámetros de calidad de servicio que están en función del Throughput, según lo establecido en anexo C.

El Th que realmente está siendo ofrecido por las capas de Radio es variante con respecto a la potencia y se puede observar que las pérdidas que se presentan siguen el comportamiento de la PLR mostrada en el modelo empírico. En este caso, el módulo de abstracción de parámetros de Capa de Radio toma muestras de la potencia cada 2 ms, i. e., cada vez que el tiempo de decorrelación cambia. A una tasa constante de 150 kbps, durante 10 segundos se generan 1500.000 kbits.

En este punto es conveniente hacer notar que las unidades en las que están presentados los valores del Th Ofrecido, Th Generado y Pérdidas es algo que se ha denominado kbps equivalentes o Th Generado Equivalente en kbps. Esto significa que el valor que puede leerse en la ordenada de las gráficas mostradas, es un valor de kbits que se generarían *si al menos* se transmitiera durante un segundo, i. e., si el valor indicado se mantuviera por un segundo, la aplicación habría generado la cantidad de kbits señalados por la ordenada. Como esto no es necesariamente cierto para todos los casos (no siempre se mantiene constante durante un segundo el valor representado), se hace referencia a un valor de kbps que es multiplicado por el tiempo de decorrelación del canal T_{dec} , para determinar cuánta

cantidad de tráfico realmente se generó en el tiempo de decorrelación, que es el tiempo en el que se mantiene constante el valor medido de potencia y por ende, el de la PLR.

Así, el Th Generado en la muestra k se define a partir del Th Generado Equivalente en la expresión

$$\text{Th}_{\text{gen}_k} = \text{Th}_{\text{gen-equiv}_k} \cdot T_{\text{dec}} \quad (36)$$

Para poder calcular las pérdidas, se debe obtener el valor del Th Ofrecido para cada Th_{gen_k} . Esto nos lleva a reescribir la ecuación (6) de la siguiente manera

$$\text{Th}_{\text{of}_k} = \text{Th}_{\text{gen}_k} \cdot (1 - \text{PLR}_k), \quad (37)$$

asumiendo que el Th Generado es menor que el Th máximo ofrecido por la tecnología de radio y donde PLR_k es el valor de la tasa de pérdida de paquetes en la muestra k .

De manera similar, reescribimos la ecuación (7) para obtener el Th perdido en la muestra k como

$$\text{Th}_{\text{lost}_k} = \text{Th}_{\text{gen}_k} - \text{Th}_{\text{of}_k} \quad (38)$$

La expresión anterior implica que las pérdidas se calculan en cada muestra que se toma del canal, cuando es posible establecer un valor de la PLR y por ende, del Th Ofrecido. El valor obtenido de PLR en cada muestra se mantiene constante únicamente durante el tiempo de decorrelación, por lo que el Th Generado es válido también durante ese instante de tiempo.

Finalmente, la cantidad de datos que se pierden es calculada a partir de la expresión

$$\text{Pérdidas} = \sum_{k=1}^n \text{Th}_{\text{loss } k} \quad (39)$$

donde n es igual al número de muestras que se toman durante el tiempo que se transmiten datos, que de acuerdo a lo establecido en la ecuación (34) es igual a 5000 para una secuencia de video de 10 segundos. Las pérdidas totales, calculadas a través de la expresión (39), suman así 165.112 kbits.

Tal y como había sido establecido previamente, es fundamental para este análisis el que se considere que los paquetes generados por la aplicación sean de longitud constante y que la fuente genere información de manera constante también. La validez de asumir estas condiciones radica en que las fuentes de tráfico están debidamente caracterizadas y se sabe con certeza cuánto Th medido en kbps generan las aplicaciones para ciertos valores de formato de resolución y fps, lo cual permite evaluar las pérdidas según lo establecido en las ecuaciones previas.

Como se mencionó previamente, en las simulaciones que involucran la implementación del Diseño de Cross Layer propuesto, se varían dos parámetros que impactan en el nivel de pérdidas que se presentan durante la transmisión. El primero de estos parámetros es el tiempo de respuesta de la aplicación a las señalizaciones realizadas por el Distribuidor de Decisiones, que está relacionado con la granularidad de la capa de Aplicación. Como se dijo, la granularidad de las capas de Radio, es diferente a la de la capa de Aplicación, y su impacto será evaluado en este punto. Dependiendo del tiempo de respuesta de la aplicación, puede llegar a presentarse inestabilidad en el sistema y además, se impacta directamente en la cantidad de pérdidas que se presentan.

En los resultados siguientes se consideró que el Optimizador de Cross Layer toma lecturas del nivel de potencia a intervalos de tiempo iguales al tiempo de decorrelación del canal, lo que implica

$$T_s = T_{dec} \quad (40)$$

donde T_s es el período de muestreo de la potencia por parte del Optimizador y T_{dec} es el tiempo de decorrelación del canal, establecido a 2 ms, que corresponden según lo descrito en secciones anteriores, a una velocidad de 80 km/hr para una frecuencia de portadora de 1.9 GHz.

En la figura 11 se muestran los resultados de la simulación cuando el tiempo de respuesta de la aplicación, T_{ans} es igual al tiempo de muestreo de la potencia T_s . Al principio de la simulación, la aplicación está generando el mayor Th posible, valor que es ajustado de manera inmediata debido al tiempo de muestreo de la potencia, que es igual al tiempo de decorrelación del canal y a que la aplicación responde de manera prácticamente inmediata a una señalización del Distribuidor de Decisiones. En este caso el Th Ofrecido Virtual y Real, son iguales debido a que los tiempos de muestreo y decorrelación del canal son iguales. Posteriormente se profundizará en este último aspecto.

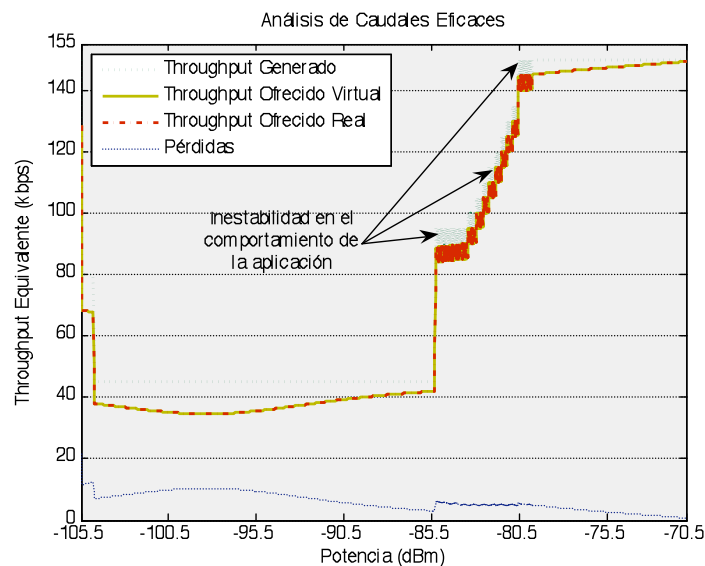


Figura 11. Análisis de Caudales eficaces con tiempo de respuesta de aplicación igual al tiempo de muestreo de potencia.

Puede verse que el hecho de que la aplicación responda de manera inmediata a las señalizaciones del Distribuidor de Decisiones genera inestabilidad en el comportamiento del Th Generado por la aplicación a partir de una potencia medida de -85.5 dBm, aproximadamente. En el momento en que inicia la simulación, la función objetivo determina que ante el nivel de pérdidas que existe, se debe generar menos tráfico y la aplicación se adapta de manera instantánea a esta situación. Conforme las pérdidas continúan incrementándose, se observa que la función objetivo indica que la aplicación debe generar cada vez menos tráfico, hasta llegar al mínimo posible. Cuando la potencia llega a los -85.5 dBm, la PLR ha mejorado considerablemente y el Optimizador a través de la función objetivo determina que se puede comenzar a generar más tráfico. Una vez superados los -80.5 dBm de potencia percibida, la aplicación vuelve a generar el máximo Th posible pasando previamente, por momentos de inestabilidad.

En la figura 12 se presenta la normalización de la PLR determinada por el modelo empírico y la normalización del Th Generado, donde es posible observar claramente que el Diseño de Cross Layer indica a la aplicación que genere el menor tráfico posible mientras la PLR es elevada y le permite generar cada vez más tráfico conforme la PLR disminuye.

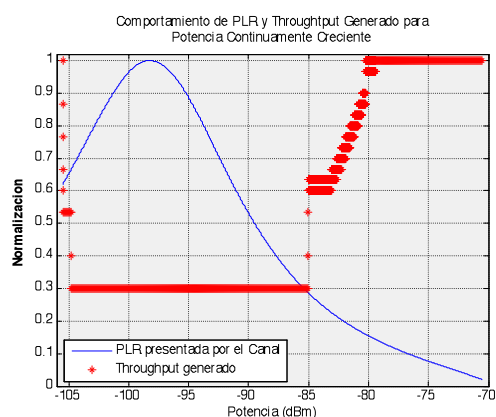


Figura 12. Normalización de PLR y Th Generado.

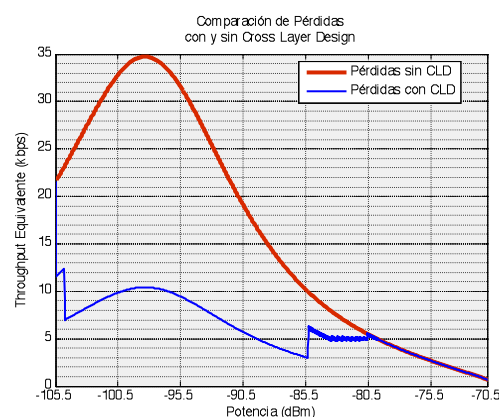


Figura 13. Comparación de Pérdidas con y sin Cross Layer Design.

La figura 13 muestra la comparación entre las pérdidas que se generan cuando se implementa el diseño de Cross Layer y cuando éste no es implementado. Es posible observar que las curvas de pérdidas siguen el comportamiento de la PLR del modelo empírico, pero cuando está presente la optimización, la curva de pérdidas es considerablemente menor.

El comportamiento de la curva de pérdidas con el Diseño de Cross Layer presenta algunos cambios abruptos para ciertos valores de potencia. Estos cambios abruptos y la forma similar a un diente de sierra que se observan son ocasionados por los ajustes del Th Generado. Una vez que el Optimizador de Cross Layer determina un cambio en el comportamiento de la aplicación, se crea un ajuste en el Th Generado y en consecuencia, en el Th Ofrecido. Mientras no se señalice un nuevo ajuste del Th que genera la aplicación, la curva de pérdidas se comportará de acuerdo al modelo de PLR y potencia y, si ahí se mantuviera constante, observaríamos una curva semejante a la de las pérdidas sin el Diseño de Cross Layer. Como el Th Generado no se mantiene constante, sino que es incrementado o disminuido, las pérdidas generadas se desplazan entre una y otra curva de pérdidas característica a ese Th en particular.

Una vez que se han realizado ajustes en el Th Generado por la aplicación es necesario calcular la cantidad de datos que son enviados hacia las capas de Radio dado que ahora, el Th Generado se está adaptando al Th Ofrecido que es variante según las condiciones del canal. Siguiendo un razonamiento similar al utilizado para calcular las pérdidas en la sección anterior, se tiene que el Th Generado está definido por

$$Th_{gen} = \sum_{k=1}^n (Th_{gen-equiv_k} \cdot T_{dec}) \quad (41)$$

donde el k-ésimo $Th_{gen-equiv}$ (kbps) corresponde al Th Generado que generaría la aplicación si por lo menos transmitiera un segundo, T_{dec} es el tiempo de decorrelación del canal y n el

número de muestras señalado por la ecuación (36). El análisis dimensional de esta expresión demuestra que las unidades del Th_{gen} son kbits, lo cual es correcto. La ecuación (41) puede reescribirse como

$$Th_{gen} = T_{dec} \sum_{k=1}^n Th_{gen-equiv_k} \cdot \quad (42)$$

Para esta simulación, el tiempo de respuesta de la aplicación es igual al tiempo de muestreo de la potencia, por lo que haciendo uso de la ecuación (41), se obtiene que en total fueron generados 831.390 kbits. Para calcular el Th Ofrecido y las pérdidas se hace uso de las ecuaciones (36) a (39) obteniéndose que se perdieron en total 60.363 kbits; valor que contrasta con los 165.112 kbits que se pierden cuando no se implementa el Diseño de Cross Layer.

A continuación se presentan los resultados cuando el tiempo de respuesta de la aplicación es mucho mayor al tiempo de muestreo de la potencia. En este caso se considera que la aplicación tarda 200 ms en responder a una señalización del Distribuidor de Decisiones. En la figura 14 se muestran los resultados de esta simulación. Es posible observar que las inestabilidades en el comportamiento de la aplicación han desaparecido pero también se observa una respuesta más lenta ante condiciones de altas pérdidas. Esta respuesta más lenta ocasiona que las pérdidas sean mayores que en el caso anterior. En la figura 16 se muestra la normalización de la PLR y el Th Generado, se observa que cuando la potencia alcanza los -85 dBm aproximadamente, la aplicación comienza a generar más tráfico, al igual que en el caso anterior.

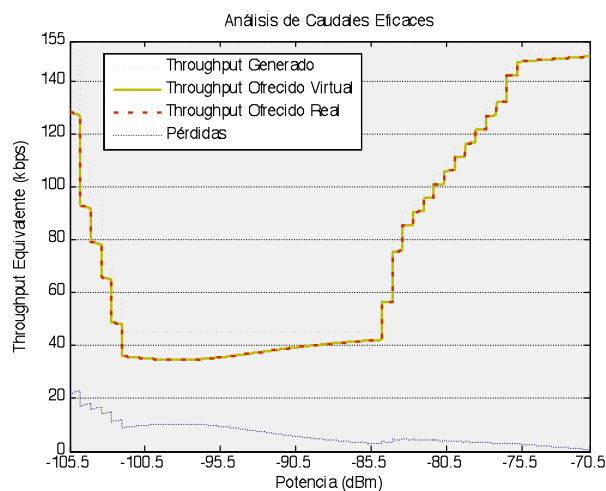


Figura 14. Análisis de Caudales Eficaces con tiempo de respuesta de aplicación de 200 ms.

La figura 16 ilustra el comportamiento de las pérdidas con el tiempo de respuesta de la aplicación indicado. Es posible observar los cambios abruptos en la forma de la curva de pérdidas cuando se implementa el Diseño de Cross Layer a medida que se van realizando los ajustes en el Th Generado.

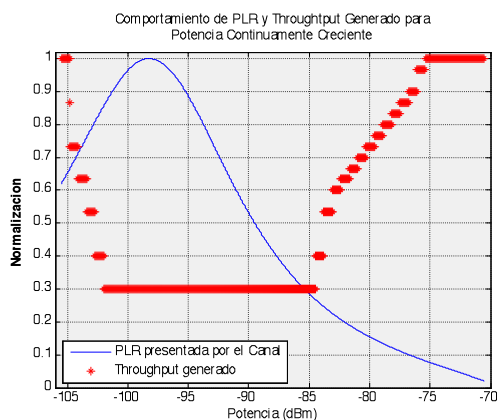


Figura 15. Normalización de PLR y Th Generado con tiempo de respuesta de la aplicación igual a 200 ms.

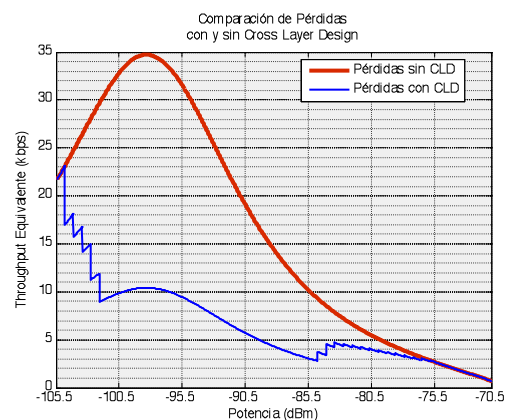


Figura 16. Comparación de pérdidas con y sin CLD para un tiempo de respuesta de la aplicación de 200 ms.

Haciendo uso de las ecuaciones (36) a (39) y (42), se tiene que se generaron 815.960 kbits de los cuales se perdieron en total 65.502 kbits. Las pérdidas son algo mayores que el ejemplo anterior.

El Distribuidor de Decisiones notifica a la aplicación que debe ajustar el Th Generado pasados apenas 2 ms después de iniciada la simulación, sin embargo, debido a que el tiempo de respuesta de la aplicación es de 200 ms, se puede observar en la figura 15 que al principio de la simulación no hay un ajuste en el Th Generado; el ajuste se presenta pasados los 200 ms que ha sido establecido, tarda la aplicación en responder.

Esta misma condición genera que durante los primeros 200 ms las pérdidas sean iguales con y sin Cross Layer Design, debido a que en ambos casos se están transmitiendo 150 kbps. Al final de la simulación la aplicación vuelve a generar el máximo Th posible y por ello su curva de pérdidas empata perfectamente con la curva de pérdidas sin CLD.

Un tiempo de respuesta de la aplicación igual a 1 segundo, produciría los resultados mostrados en las figuras 17 y 18, con un total de 975.040 kbits de datos generados y pérdidas totales de 109.375 kbits.

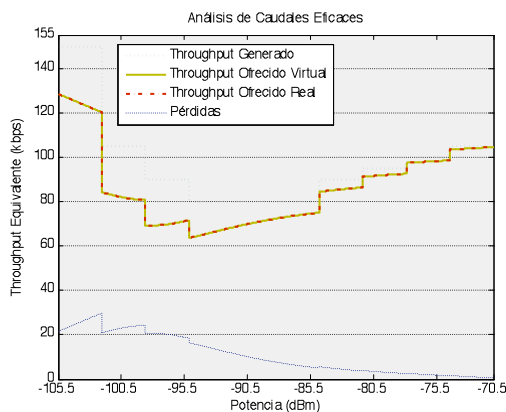


Figura 17. Análisis de Caudales Eficaces con tiempo de respuesta de aplicación de 1000 ms.

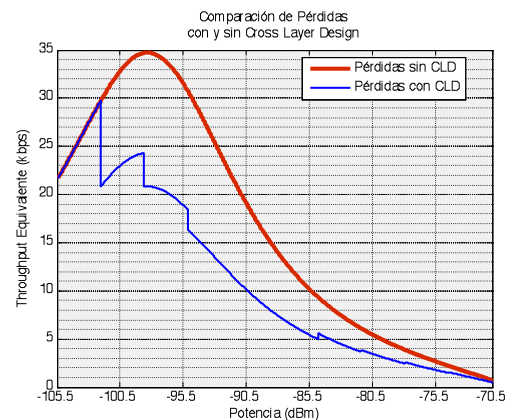


Figura 18. Comparación de pérdidas con y sin CLD para un tiempo de respuesta de la aplicación de 1000 ms.

En los siguientes resultados se ha mantenido fijo el tiempo de respuesta de la aplicación y éste ha sido establecido a .1 ms, i. e., se asume que la aplicación responde de manera prácticamente inmediata a una señalización del Distribuidor de Decisiones; el parámetro a ser evaluado es el tiempo de muestreo del estado del canal T_s , que ahora tomará valores de 0.01 s, 0.1 s y 1 s²⁵. En la figura 19 se observa que con este tiempo de muestreo se presenta inestabilidad en la aplicación que, sin embargo, es menor que la presentada cuando $T_{ans} = .001$ s. Aunque a simple vista no es posible visualizarlo, en esta gráfica el Th Ofrecido Virtual y el Real no son iguales. En la figura 20 se muestra un detalle de la gráfica 19 correspondiente a un valor de potencia de aproximadamente -104 dBm.

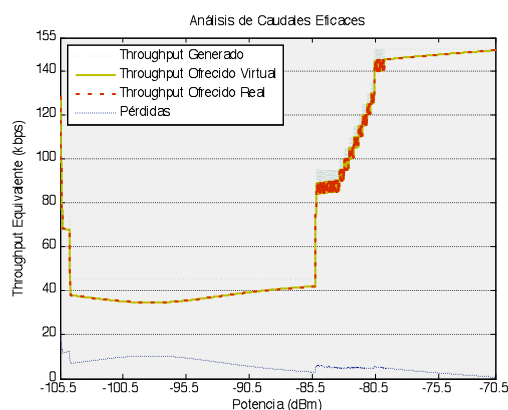


Figura 19. Análisis de Caudales Eficaces con tiempo de muestreo de las condiciones del canal de .01 s.

Ahí, es posible observar que el Th Ofrecido Virtual (TOV) se presenta con forma escalonada mientras que el Th Ofrecida Real (TOR) es más bien una línea continua. En realidad, El TOV es el Th que el Diseño de Cross Layer asume que el canal ofrece a las capas superiores mientras que el TOR, es el Th que *realmente* es ofrecido por el radio. Esta diferencia se presenta porque el módulo Optimizador de Cross Layer toma un valor de PLR (parámetro abstraído de capa de Radio) correspondiente a un nivel de potencia y lo asume constante hasta que vuelve a tomar la siguiente muestra, generando una concepción errónea

²⁵ Los resultados para $T_s = T_{dec}$ son los presentados en las figuras 12, 13 y 14.

del estado del canal. Este fenómeno es más visible mientras menor es la frecuencia con la que se muestrea el estado del canal, i. e., mientras mayor es el valor de T_s .

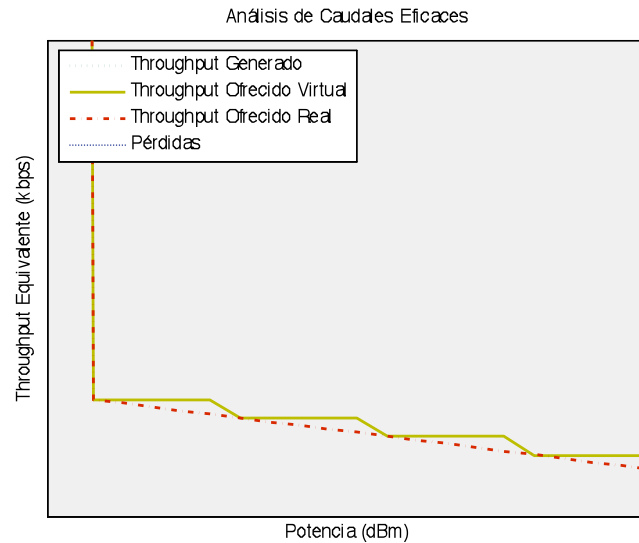


Figura 20. Detalle de una sección de la figura 20 que muestra una concepción errónea del estado del canal debido a un período de muestreo superior al tiempo de decorrelación del canal.

Esta concepción poco precisa del estado del canal no tiene impacto numérico en el entorno de la simulación dado que los cálculos se realizan en concordancia a los modelos matemáticos presentados. Sin embargo, tanto en simulación como en la eventual implementación de este trabajo de investigación, el hecho de que el tiempo de muestreo sea diferente del tiempo de decorrelación del canal sí tiene efectos secundarios sobre el desempeño del Optimizador debido a que éste toma decisiones en base a la posición relativa del Th Ofrecido dentro de las regiones de decisión. Dado que los valores de PLR que el Optimizador posee definen el Th Ofrecido, un tiempo de muestreo grande generará respuestas por parte del Optimizador que no corresponden al comportamiento real del canal en un momento determinado.

Utilizando las ecuaciones (36) a (39) y (42), se obtiene un total de datos de 832.350 kbits y pérdidas totales por 60.603 kbits. En la figura 21 se muestran los resultados de la simulación cuando el tiempo de muestreo T_s es igual a 100 ms.

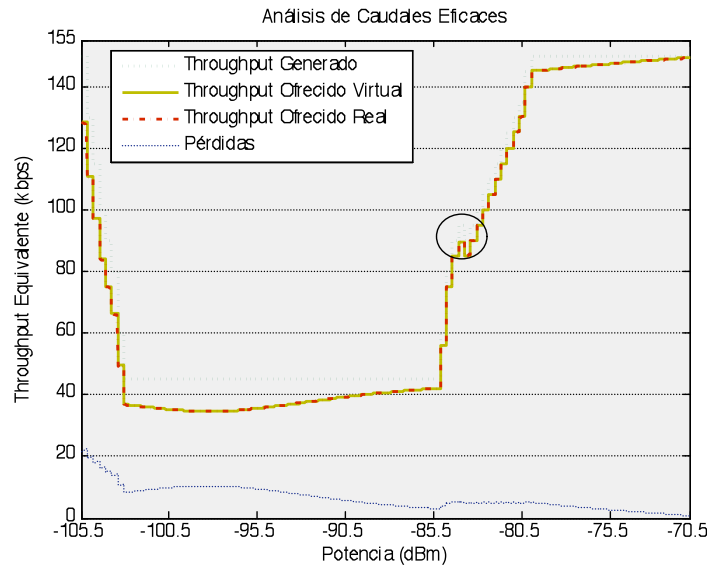


Figura 21. Análisis de Caudales Eficaces con tiempo de muestreo de las condiciones del canal de .1 s.

Se debe notar que a pesar de que la inestabilidad en el Th Generado ha disminuido, ésta no desaparece por completo tal y como se señala en la figura 21. En total se generaron 856 kbits y se perdieron 65.156 kbits, valores muy similares a los obtenidos cuando el tiempo de respuesta de la aplicación era de 200 ms y el de muestreo era de 2 ms. Adicionalmente, el fenómeno de la variación del TOV frente al TOR comienza a hacerse más presente.

La figura 22 muestra los resultados de la simulación cuando el tiempo de muestreo T_s es de 1 segundo. La similitud del comportamiento mostrado en esta simulación con la mostrada en la figura 18 es grande y es posible distinguir claramente que el fenómeno que rige adecuadamente el comportamiento del Th Generado es el tiempo muestreo del estado del canal. Obsérvese que al Optimizador le toma más tiempo lograr que el Th Generado llegue a un nivel mínimo, comparando con los resultados de la figura 17. Esta condición es generada debido a la mala estimación que se tiene del Th Ofrecido ocasionada a su vez, por

el tiempo de muestreo mucho mayor que el tiempo de decorrelación del canal (500 veces más grande).

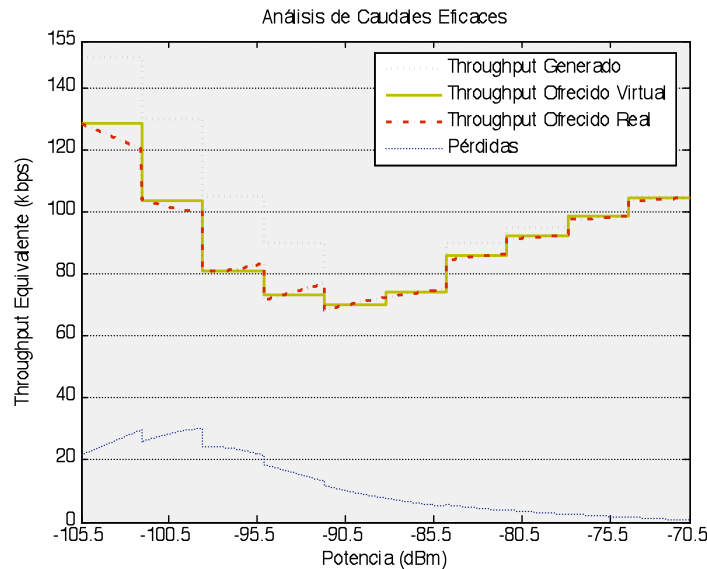


Figura 22. Análisis de Caudales Eficaces con tiempo de muestro de las condiciones del canal de 1 s.

En esta figura es clara la diferencia entre el TOV y el TOR y su impacto puede ser evaluado visualmente en el primer ajuste del Th que señala el Optimizador. Nótese que el hecho de que se asuma un Th Equivalente Ofrecido de casi 130 kbps le hace determinar a la función objetivo que el nuevo Th Generado sea precisamente, 130 kbps. Mientras que si comparamos la respuesta de la función objetivo cuando el Th Ofrecido está bien estimado (ver figura 17) para el mismo nivel de potencia, podemos observar que la función determina un nuevo Th Generado de 105 kbps que es inclusive, menor que el Th Ofrecido en ese instante. La función objetivo ha buscado y determinado, en este caso, el estado a_{j-n} que es menor que el Th Ofrecido, ver el cuarto caso de la ecuación (28). Para esta simulación y de acuerdo a las ecuaciones correspondientes descritas anteriormente, se tiene que fueron generados 1025.000 kbits de información, de los cuales se perdieron en total 119.934 kbits.

En la práctica, el tiempo de respuesta de la aplicación no puede llegar a ser tan pequeño, de hecho, ajustar un formato de resolución y una fps en una videoconferencia real es un proceso del orden de varias centenas de milisegundos. Un mal muestreo del canal, aunado a la condición de respuesta mencionada, generaría resultados que son inclusive perjudiciales para la aplicación. En las figuras 23 y 24 se muestran algunos resultados de la simulación para un $T_{ans} = 2s$ y un $T_s = 1s$.

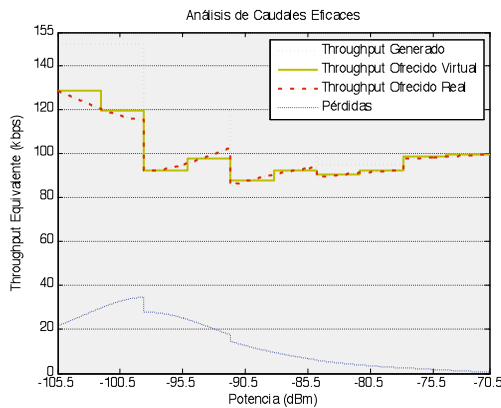


Figura 23. Análisis de Caudales Eficaces con $T_s = 1s$ y $T_{ans} = 2s$.

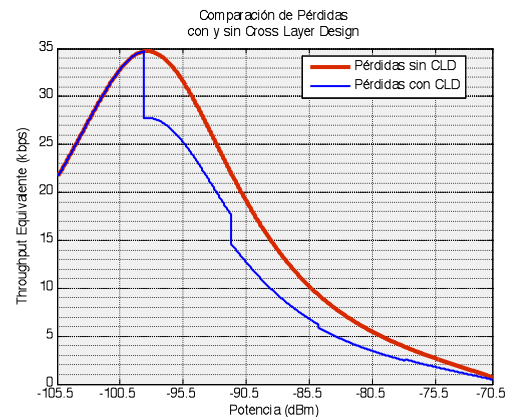


Figura 24. Comparación de pérdidas con $T_s = 1s$ y $T_{ans} = 2s$.

De la figura 24 pudiera concluirse que el comportamiento de las pérdidas es algo mejor cuando se implementa el Diseño de Cross Layer que cuando no se hace, sin embargo, este razonamiento es engañoso. Si bien es cierto que la curva azul (línea delgada) de la figura 24 representa menores pérdidas que la curva roja, también es cierto que el Th Generado fue diferente, de hecho, se generó menos información que cuando no se realiza la optimización.

Este es un resultado esperado, dado que se asume que en malas condiciones del canal, la función objetivo indicará a la aplicación generar menos tráfico. Para esta simulación, las pérdidas, de acuerdo a las ecuaciones (36) a (39) y (42) sumaron un total de 137.114 kbits para un monto de datos generados de 1130 kbits.

IV.2.2 Análisis de distorsión de la imagen de video.

Para evaluar el desempeño del diseño propuesto se propone analizarlo de manera cualitativa. El uso de la PSNR permitirá evaluar y obtener un indicador, no la calidad la imagen por si misma, si no la comparación de la imagen original enviada con la imagen decodificada recibida. Si bien, en algún momento determinado el codificador se ve en la necesidad de emplear una calidad baja, la idea es que por lo menos no disminuya aun más la calidad en el extremo receptor.

Se ha establecido en el capítulo II, que la PSNR se puede obtener por medio de la ecuación (1). Sin embargo, un parámetro que no se definió fue el de MSE, que es el error cuadrático medio.

El error cuadrático medio entre la imagen original y la recibida se define como [Choi, 2006]

$$MSE = MSE_s + MSE_l \quad (43)$$

Donde MSE_s es el error cuadrático medio producido por la fuente (codificador) y MSE_l es el error cuadrático medio originado por pérdidas en la transmisión, el cual se describe como:

$$MSE_l = \sum_{i=1}^n D_i P_i \quad (44)$$

donde P_i es la probabilidad de que el i -ésimo cuadro sea el que se pierda durante la transmisión de un GOP, y D_i es el error cuadrático medio que es inducido por la pérdida de ese cuadro. Simular las pérdidas producidas por la codificación en la compresión de video resulta altamente complejo, ya que no solo depende de la imagen que se esté mostrando en un momento determinado, sino que también depende de los recursos de la computadora, por lo que estas pérdidas no se tomarán en cuenta para este estudio, considerando que si existiesen, afectaría de igual manera con la implementación de CLD y sin ella.

Para obtener la distorsión generada por la pérdida de algún cuadro dentro del GOP, se utilizaron los resultados obtenidos en el trabajo de [Choi et al., 2006], los cuales se muestran en la figura 25, donde el índice i representa el número de cuadro dentro de un GOP, y la distorsión representa la cantidad de MSE que se genera con la pérdida de ese cuadro.

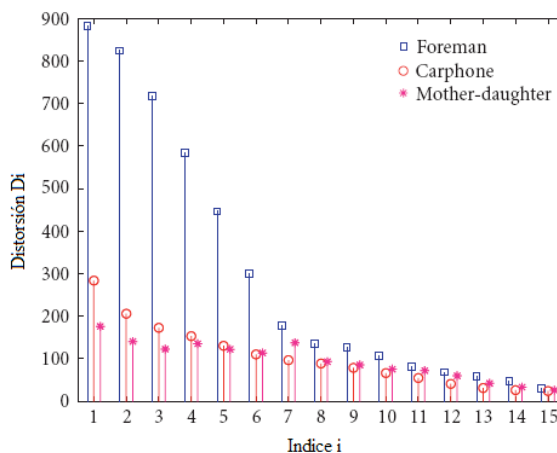


Figura 25. Distorsión generada en la imagen dentro de un GOP según el cuadro perdido.

EL perfil de distorsión D_i de la figura 25 es calculado considerando considera 3 tipos de escenas: Mother and Daughter, Carphone y Foreman.

En la escena Mother and Daughter, el área de movimiento corresponde al de la mano de la madre, la cual se encuentra con su hija (ver figura 26a). En el tipo de escena Carphone (ver figura 26b)., el interlocutor se encuentra dentro de un vehiculo en movimiento, donde se perciben movimientos faciales y de fondo. La escena Foreman (ver figura 26c)., se observa a un trabajador portando una cámara, en la cual se va filmando a él mismo mientras camina, por lo que hay movimiento en toda la escena.



Figura 26. Tipos de escena. a) Mother and Daughter, b) Carphone, c) Foreman

Estos tipos de escena, además de poder considerarse como escenas comunes, exponen que tanto tráfico genera el codificador. Es de esperarse que el tipo de escena Foreman fuese el que más le exigiera al codificador, lo cual resulto ser así.

La razón a lo anterior, resulta de la naturaleza misma de la imagen, donde un interlocutor se encuentra hablando, generando movimientos faciales, y además lleva consigo la cámara, lo que ocasiona movimiento en el fondo. El codificador H.263 y H.264 utiliza codificación espacial y temporal. Esta codificación saca provecho de escenas donde hay poca actividad, lo cual no sucede en este tipo de escena.

Para evitar un funcionamiento incorrecto de la optimización implementada en este trabajo de tesis, se opto por considerar el tipo de escena que demandara mayor tasa de bit para su envío, la cual es el tipo de escena Foreman.

La figura 27 muestra los resultados obtenidos al utilizar CLD en la transmisión de video, se aprecia como las pérdidas crecen considerablemente comparado con la curva inferior que representa las pérdidas obtenidas utilizando el mecanismo de CLD propuesto.

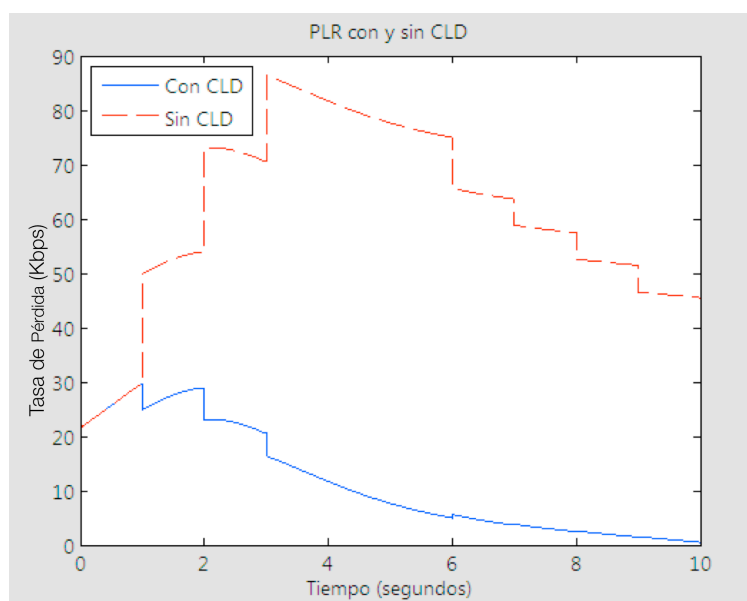


Figura 27. Comparación de la tasa de pérdida con y sin mecanismo de CLD.

La figura 28 muestra el resultado al aplicar la ecuación (44) los GOP generados en 10 segundos. Se aprecia que en la curva que representa el comportamiento sin CLD, el nivel de PSNR se mantiene bajo porque el canal en un principio no satisface las condiciones de transmisión a la tasa de bit de la aplicación, por lo que la distorsión se presenta desde el inicio. Los valores de PSNR promedio sin utilizar mecanismo de CLD son de aproximadamente de 17 dB, representando una imagen distorsionada y de mala calidad.

Por otro lado al utilizar el mecanismo de CLD se puede apreciar como la calidad de la imagen es afectada de igual forma por el cambio abrupto de niveles de potencia afectando la tasa efectiva de transmisión, generando errores en la imagen del video. También es de destacar que el mecanismo de retroalimentación de Cross Layer en aproximación de capas

intermedias empieza a funcionar tratando de ajustarse a las condiciones del canal en un lapso de 10 segundos de simulación. Numéricamente el mecanismo de CLD se adapta y sobrepasa el umbral de los 30 dB de PSNR considerado como un valor límite para imágenes de buena calidad o calidad aceptable.

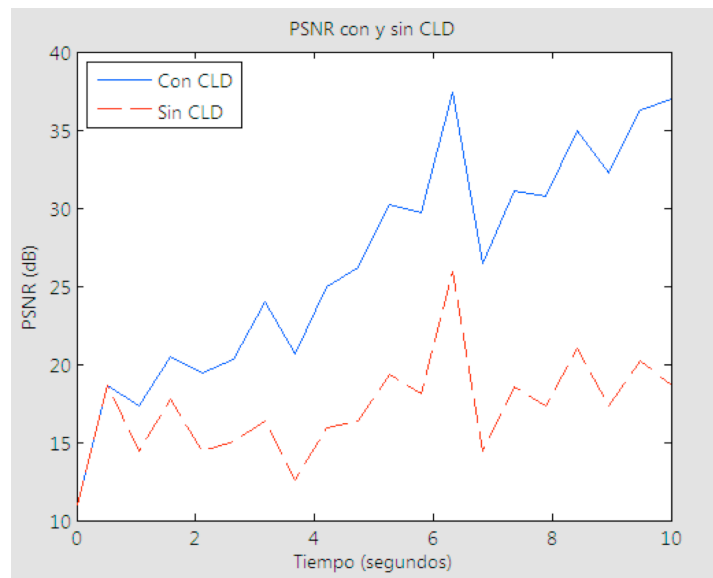


Figura 28. Distorsión generada en la imagen dentro de un GOP según el cuadro perdido.

Otra observación es el hecho de que al no utilizar el mecanismo de CLD el flujo de las imágenes del video no logra sobrepasar el límite de los 30dB.

IV.3 Resumen del Capítulo IV

En este capítulo fueron presentados los resultados de las simulaciones realizadas así como el análisis numérico de los mismos. Se presentaron las consideraciones de simulación y los resultados fueron divididos en dos secciones, una sección fue enfocada a evaluar y caracterizar el comportamiento de la aplicación bajo las consideraciones de simulación hechas, con el objetivo de establecer un punto de referencia para validación de resultados. La segunda sección se centró en los resultados y el análisis numérico de la simulación del Diseño de Cross Layer propuesto con énfasis en la variación de dos parámetros que caracterizan el comportamiento del mismo. Por otro lado, el análisis de distorsión de la

imagen del video permite conocer que tanto se degrado la imagen recibida respecto a la original debido a las perdidas, sin importar la cantidad de pérdidas, si no que tanto impactan en el video.

Discusión y Conclusiones

En este capítulo se presentan la discusión final de los resultados obtenidos, así como las conclusiones del trabajo de investigación realizado y las propuestas de trabajo futuro. La discusión se centra en los parámetros de la simulación evaluados, la factibilidad de llevar a la práctica los resultados obtenidos y, las respuestas al planteamiento del problema original. Finalmente se propone el posible trabajo futuro a realizarse tomando como punto de partida la investigación realizada.

V.1 Discusión de Resultados

La evaluación del comportamiento del Diseño de Cross Layer Propuesto para variaciones del tiempo de respuesta de la aplicación y variaciones del tiempo de muestreo de las condiciones del canal, permite comprender el impacto que ocasiona la diferente granularidad de las capas que participan en el proceso de optimización propuesto. La presencia de inestabilidad en el desempeño de la aplicación es prueba de ello.

La inestabilidad del Th Generado surge como consecuencia de pretender que la aplicación responda con la granularidad de los tiempos propios de las capas de Radio. Aunque este enfoque logra minimizar las pérdidas de información a nivel de capa de Red, la eventual aplicación de este tiempo de respuesta no es práctica por dos razones principalmente:

- La modificación del Th Generado implica la variación del formato de resolución, o la fps, o ambas a la vez; a menos que un cambio de Th implicara únicamente un cambio de fps, el impacto de variar el tamaño del cuadro a codificar, varias veces en

un segundo, es negativo desde la perspectiva del usuario final ya que esto sería observado como una vibración en la imagen, en el mejor de los casos.

- Cualquier cambio en los parámetros de codificación del video implica un reprocesamiento completo de las secuencias de imágenes que se estén codificando en ese momento. Normalmente dicho reprocesamiento conlleva la congelación de la imagen, parpadeo de ésta ó pérdida de nitidez durante unas centenas de milisegundos. La aplicación no podría responder a una señalización de cambios que fuera más rápida que el valor mencionado.

La granularidad temporal de la capa de aplicación es por tanto, del orden de segundos, y la respuesta real que puede ofrecer a las señalizaciones del Distribuidor de Decisiones no es menor de este valor.

Por otra parte, quedó demostrada la importancia de que la frecuencia con la que se muestrean las condiciones del canal, sea elevada. La naturaleza agresiva del canal inalámbrico y la condición de desplazamiento del móvil a velocidades vehiculares, ocasionan que el tiempo que se mantienen constantes los valores que se miden del canal en sí, sean muy cortos. En particular se establece que un valor de potencia medida en un canal inalámbrico se mantiene constante únicamente en la duración del llamado tiempo de decorrelación. Con la intención de que los valores de PLR con los que se calcula el Th Ofrecido en la propuesta puedan ser considerados como ciertos, se debe muestrear el canal con un período igual o menor al tiempo de decorrelación. Un valor así medido, se mantendrá constante durante dicho tiempo y posteriormente cambiará a otro nivel diferente cuyo valor es aleatorio. Asumir que un valor de PLR se mantendrá por más tiempo que el de decorrelación del canal es inherentemente incierto.

Sin embargo, las simulaciones realizadas consideran el valor del tiempo de decorrelación del canal como el correcto para efectos de valorar las pérdidas y se mostró que inclusive para tiempos de muestreo de 100 ms (50 veces el tiempo de decorrelación), el desempeño del Optimizador de Cross Layer no se ve comprometido.

Este resultado no implica que se deba asumir que un período de muestreo de 100 ms es adecuado en entornos reales. Lo que este resultado implica directamente es que en condiciones ideales del canal, el tiempo de respuesta de la aplicación puede ser regulado a través del tiempo de muestreo de las condiciones del entorno. Más aún, este resultado muestra el impacto que el tiempo de muestreo tiene sobre el tiempo de respuesta de la aplicación. Los tiempos de muestreo y de respuesta deben mantener su granularidad de acuerdo a la capa a la que pertenecen.

El período de muestreo debe ser tan, o más pequeño que el tiempo de decorrelación del canal para asegurar la fiabilidad de las decisiones tomadas por el Optimizador de Cross Layer. Adicionalmente a esta condición, el tiempo de respuesta de la aplicación debe ser considerado en función del tiempo que al codificador le tome ajustarse a los parámetros indicados por el distribuidor de decisiones, valor que será, necesariamente y por lo menos, tres órdenes de magnitud mayor que el tiempo de decorrelación del canal. Esto último con el objetivo de garantizar la estabilidad del comportamiento de la aplicación.

Una parte fundamental del funcionamiento del Optimizador de Cross Layer son las Regiones de Decisión. Éstas se encuentran implícitamente declaradas dentro de la función objetivo Ómicron²⁶, en esta tesis no se evaluó el comportamiento del Diseño de Cross Layer en función de dichas regiones puesto que no era parte de los objetivos planteados originalmente. Sin embargo, durante el desarrollo de la fase de simulación fue posible encontrar que cuando el canal presenta altas pérdidas, reducir el tamaño de las Regiones de Decisión mejora el desempeño del Diseño, mientras que cuando el canal presenta bajas pérdidas, Regiones de Decisión más grandes mejoran el desempeño.

En las últimas gráficas presentadas se puede comprobar que cuando más datos se transmiten más datos se pierden, aparentemente, de forma independiente al canal. En las simulaciones esto obedece al comportamiento que presenta el modelo que relaciona las

²⁶ Ver ecuación (26).

pérdidas con la potencia; siempre existe una región de altas pérdidas cuando la potencia está en los -98 dBm aproximadamente. Esta característica genera que siempre se pierdan en esa zona alrededor del 25% de lo que se envíe, sin importar si es mucha o poca información²⁷. En la práctica es aceptado que mientras más datos se transmitan, si se mantienen las mismas condiciones del canal, se perderá más información. Sin embargo en entornos reales se deben considerar otros parámetros aparte del estado del canal, como el número de usuarios activos en los sectores.

Debe notarse también la forma escalonada en las gráficas del Th Generado, este comportamiento también está presente en el Th Ofrecido, sin embargo, cuando se muestrea el canal a intervalos menores o iguales que el tiempo de decorrelación esto ya no puede ser apreciado en el Th Generado. Dicho comportamiento resulta del carácter discreto del conjunto de valores que puede tomar la aplicación. Esto significa que los valores de Th Equivalente que la aplicación puede tener están definidos de acuerdo al Proceso de Abstracción de Parámetros de capa de Aplicación, y éstos a su vez relacionados al proceso de caracterización de fuentes de tráfico realizado.

Un incremento en el número de usuarios transmitiendo datos por el enlace RL de un sector de DO Rev.0 limita la capacidad de transporte de datos de la tecnología celular debido al incremento del valor de la relación I_0/N_0 (Bhushan et al. 2006). Una condición de este tipo no sería percibida a través del valor del RSSI de una terminal de acceso puesto que es una circunstancia ajena a las condiciones propias del enlace FL, lo que seguramente generaría valores de PLR diferentes a los registrados en el modelo empírico. También es cierto que al final de cuentas, un número elevado de usuarios generaría más interferencia en CDMA y que en teoría, el control de potencia de lazo cerrado del enlace RL controlado por la radio base aliviaría esta condición. Sin embargo, si el móvil se encuentra en situaciones extremadamente complejas, e. g., en la frontera entre tres sectores con carga densa en el enlace RL, en solicitud de un traspaso y a velocidades vehiculares, poco se podrá hacer

²⁷ Recordando que consideramos que nunca sobrepasamos el límite de la tecnología, i. e., nunca se supera el caudal eficaz máximo que ofrece 1xEV-DO REv.0

para que los datos que son enviados no experimenten altas pérdidas, sobre todo si los datos son tráfico de tipo conversacional.

Finalmente, conviene hacer énfasis en que la optimización propuesta, i. e., el Diseño de Cross Layer, modifica el Th que genera un codificador de video H.263 cuando la información es transmitida sobre el enlace RL de un enlace cdma2000 1xEV-DO Rev.0, en función de la tasa de pérdida de paquetes a nivel de capa de Red sobre el mencionado canal ascendente. En otras palabras, se buscó lograr que el Th Generado siguiera el comportamiento del Th Ofrecido por las capas de Radio. Para ello se diseñó y evaluó en simulación un nuevo Diseño de Cross Layer que ha sido formalmente presentado.

La validez de los resultados está fundamentada en el análisis numérico realizado que toma en cuenta parámetros aceptados en la literatura del estado del arte respecto a las consideraciones de los modelos que estiman la calidad del canal radio (Choi et al, 2005; Rappaport, 1996; Ojanperä y Prasad, 1998).

Las fuentes de tráfico caracterizadas utilizando el codificador H.263 y los parámetros que son considerados en la literatura, fueron tomados en cuenta para el desarrollo de la simulación. Esto implica que las fuentes de tráfico utilizadas en las simulaciones corresponden a parámetros reales que fueron medidos. Todas las consideraciones hechas al respecto, están fundamentadas en comportamientos reales del codificador de video H.263.

Lo anterior, aunado a la obtención de un modelo empírico que ofrece la PLR para cualquier valor de potencia que comercialmente pueden detectar las terminales de acceso 1xEV-DO Rev.0, soporta y proporciona validez a los resultados obtenidos por las simulaciones debido a que estos yacen sobre modelos provenientes de mediciones reales, no son comportamientos teóricos.

Adicionalmente, el Diseño de Cross Layer presentado es acorde a lo disponible actualmente en la literatura especializada y la formalización empleada en este trabajo de tesis, ha sido

aplicada en otros trabajos relacionados y publicada en revistas arbitradas internacionalmente.

V.2 Conclusiones

El Diseño de Cross Layer presentado en este trabajo de investigación realiza mejoras en el envío de tráfico de video correspondiente a una videoconferencia móvil que se transmite en el enlace RL de un enlace cdma2000 1xEV-DO Rev. 0 mientras el móvil se desplaza a velocidades de hasta 80 km/hr. La optimización se basa en la adaptación del Th Generado por un codificador de video H.263, al Th Ofrecido por la tecnología de radio en función de las pérdidas que presenta el canal en un momento determinado.

Esta adaptación del Th Generado conlleva un ajuste automático de parámetros de calidad de servicio (QoS) debido a que el retardo, la variación del retardo y la tasa de pérdida de paquetes, dependen en gran medida de que el Th que generan las aplicaciones no exceda los niveles de Th que la red de transporte pueda ofrecer.

La propuesta hecha fue validada a través de simulación. En ella están incluidas, la caracterización real de las fuentes de tráfico que alimenta a la simulación per se, la obtención de un modelo empírico de PLR en función de la potencia que percibe una terminal de acceso DO Rev.0 que se desplaza a velocidades vehiculares y, una arquitectura de Cross Layer modular y escalable.

Respecto al proceso de caracterización de fuentes de tráfico, fue posible concluir que los parámetros que dominan el comportamiento del Th Generado por el codificador de video H.263 son el formato de resolución y la tasa de cuadros por segundo con la que se realiza la codificación, siendo el primero de ellos, el que más impacto tiene en la cantidad de tráfico que se genera. El tipo de escena que se codifica no tiene impacto en el Th que genera el codificador H.263 pero sí afecta la distribución estadística del mismo.

También a manera de conclusión sobre la caracterización de fuentes de tráfico, se debe mencionar que se analizó la importancia de la información que genera el codificador en función del impacto que ésta tiene durante el proceso de decodificación en el receptor y se evaluó su impacto sobre el Th Generado, encontrando que, la información más importante (los cuadros 'Intra-Picture' de H.263) genera caudales eficaces pico. En base a esto fue posible establecer que, una de las metas de diseño de Cross Layer debía ser ofrecer el mejor rendimiento posible a nivel de capa de Red para los niveles pico de Th que genera el codificador. Lo anterior nos llevó a considerar a las fuentes de tráfico como fuentes de tasa de bit constantes (CBR) ajustadas al valor pico del Th Generado para una y solo una, combinación de formato de resolución y tasa de cuadros por segundo.

Adicionalmente, como resultado del proceso de caracterización de fuentes de tráfico se encontró que la cantidad de tráfico que genera el codificador H.263 se comporta como una suma de dos exponenciales, lo cual tiene consecuencias a considerarse cuidadosamente en el desempeño del Optimizador de Cross Layer perteneciente al Diseño propuesto, debido a que el valor del Th Generado no puede ser tratado de manera lineal.

Respecto al modelo empírico que relaciona la PLR percibida en el enlace RL de un enlace cdma2000 1xEV-DO Rev. 0 (IS-856) con la potencia que el móvil percibe a través de su valor de RSSI, se concluye que a pesar de que dicho valor mide la potencia que el móvil percibe desde la radio base, i. e., la estimación del estado del enlace FL, existe una correlación entre las variaciones de potencia percibidas por el móvil y la PLR que se presenta en el enlace RL. Lo anterior fue corroborado mediante extensas mediciones de campo en las que se tomaba lectura de la cantidad de información enviada desde el móvil desplazándose a velocidades vehiculares, de la cantidad de información recibida por otro dispositivo de datos activo dentro de la red celular IS-856 en cuestión y, del valor que presentaba el RSSI en el dispositivo móvil.

Fue evidenciado que existe una relación entre la potencia que el móvil percibe en el enlace FL y la PLR asociada a la transmisión en el enlace RL como una *forma indirecta* de

establecer la calidad de éste último. Un ejemplo que ayuda a comprender este fenómeno es el que se presenta cuando el móvil atraviesa un túnel largo, dentro del cual no existen estaciones repetidoras de la portador de servicio celular. En el momento en el que el móvil entra al túnel, el valor del RSSI decae abruptamente indicando un canal radio muy agresivo en el enlace FL. Desde un punto de vista práctico, el valor de RSSI presenta una estimación del estado del canal y en este caso, es claro que el enlace RL tendrá un comportamiento muy similar al enlace FL.

Otra conclusión fundamental respecto al modelo empírico es la relacionada con los resultados de las mediciones de campo bajo condiciones similares a las mencionadas en el párrafo anterior, pero cuando el móvil se desplazaba a baja velocidad, de forma similar a las condiciones de tráfico urbano. Las mediciones hechas en el contexto ahora descrito, ofrecieron como resultado que la tecnología DO. Rev.0 no presenta pérdidas significativas en el envío de datos sobre el enlace RL cuando la velocidad de desplazamiento del móvil es baja. Esto último permitió retirar esta variable del problema planteado.

Como resultado de las simulaciones de la implementación del Diseño de Cross Layer desarrollado en este trabajo de tesis, se puede concluir que existen dos parámetros que afectan el comportamiento del diseño propuesto, a decir, el tiempo de muestreo del estado del canal T_s y, el tiempo de respuesta de la aplicación a un cambio señalado por el Distribuidor de Decisiones de Cross Layer, T_{ans} . La cuadros Intra-Picture de H.263 generan las tasas de datos pico. y la pérdida del frame “i” corresponde al 30 % de la información de un GoP para el caso de estudio La propuesta de CLD mejora significativamente dando prioridad a los frames “i” con una ganancia de 20 dB de PSNR.

Al respecto concluimos que el tiempo de muestreo del estado del canal debe ser, al menos, igual al tiempo de decorrelación del canal debido a que, pasado este último, todo parámetro que haya sido medido en relación a este medio físico de propagación, es inherentemente impreciso debido a la variabilidad del mismo. El no respetar este parámetro redundaría en que el Optimizador de Cross Layer pueda tomar como óptimos valores de T_h Generado (que

serán señalizados a la aplicación) que están basados en información errónea del estado del canal.

En relación al tiempo de respuesta de la aplicación, concluimos que éste debe respetar la granularidad propia de la capa de Aplicación y que deberá estar en función de las capacidades de respuesta del software que codifique el video. La recomendación para un buen desempeño es que T_{ans} sea del orden de segundos. No seguir esta recomendación pudiera ocasionar comportamientos inestables de la aplicación aunque, se encontró, que el mínimo de pérdidas se logra cuando el tiempo de respuesta de la aplicación responde a la granularidad de las capas inferiores.

Respecto a la arquitectura empleada para definir, modelar y simular el Diseño de Cross Layer propuesto, concluimos que ésta permite alta flexibilidad para adaptar el trabajo desarrollado a cualquier otra tecnología de comunicación inalámbrica e inclusive, cableada, así como permite que la optimización pueda implementarse sobre cualquier codificador de video. Lo anterior es posible gracias a la modularidad de la arquitectura. El módulo central, que es el Optimizador de Cross Layer, sólo requiere los bloques de Abstracción de Parámetros adecuados para poder operar sobre ellos. De esta manera se puede aplicar el Diseño para optimizar transmisiones sobre tecnologías nuevas como Wi-Max. El único requerimiento es contar con un modelo que relacione las pérdidas de paquetes con los valores que se desee y estén relacionados a nivel de capa de Radio. De la misma forma, es posible intercambiar el módulo de Abstracción de Parámetros de capa de Aplicación para que el codificador de video sea H.264, por ejemplo. Lo necesario para ello es la caracterización de las fuentes de tráfico correspondientes.

En relación a las Regiones de Decisión podemos concluir que hace falta desarrollar más trabajo de investigación relativo al impacto que éstas tienen en el desempeño de la función objetivo Ómicron. De forma breve sólo es posible mencionar que se observó que cuando el canal presenta grandes pérdidas es conveniente que las Regiones de Decisión sean pequeñas ya que aseguran una rápida convergencia del Th Generado al Th Ofrecido. De

forma inversa, cuando las pérdidas del canal son bajas, es conveniente que las Regiones se mantengan pequeñas, dado que esto elimina problemas de inestabilidad en la aplicación.]

Consideramos conveniente hacer énfasis en que el carácter experimental del trabajo desarrollado facilita su inmediata aplicación. Con los resultados obtenidos y los productos que se derivaron de este trabajo de investigación (Diseño de Cross Layer, caracterización de fuentes de tráfico, modelo empírico de pérdidas, software de simulación desarrollado, entre otros) es posible proponer de manera inmediata la implementación del diseño desarrollado.

A modo de conclusión general, podemos decir que se han cumplido satisfactoriamente los objetivos generales y particulares planteados para resolver el problema original. Se ha propuesto, modelado, simulado y evaluado en simulación un nuevo esquema de ajuste automático de parámetros de calidad de servicio para videoconferencia móvil utilizando Cross Layer Design en el enlace RL de una red celular IS-856.

V.3 Trabajo Futuro

Se recomienda la evaluación del impacto que las Regiones de Decisión tienen sobre el desempeño de la función objetivo Ómicron. De manera complementaria al modelo empírico de pérdidas, un excelente trabajo de investigación sería obtener un modelo ajustado cuando en la red celular existe tráfico denso en el enlace RL. Este modelo puede ser obtenido de manera muy similar al presentado en este trabajo. Se requerirían más terminales de acceso DO Rev. 0 que pudieran estar en el mismo sector y que generaran grandes caudales eficaces de datos mientras se evalúan las pérdidas en el nodo de interés. Esto complementaría adecuadamente el modelo empírico utilizado aquí.

Será motivo de estudio futuro el tiempo de respuesta de la aplicación óptimo tal que, minimice las pérdidas y mantenga los parámetros de video en valores estables durante una sesión de videoconferencia. Este valor es crítico al momento de evaluar las pérdidas que se

generan cuando se implementa el diseño y el desempeño del sistema está en estrecha relación con este valor.

También es motivo de estudio la optimización del tráfico de audio que se genera en una videoconferencia a través de mecanismos como el propuesto. Es posible utilizar el mismo diseño sin mayores modificaciones, siempre y cuando se obtenga el módulo de Abstracción de Parámetros que caracteriza el tráfico de los codecs de audio de interés.

Literatura Citada

- 3GPP2. (2000). cdma2000 High Rate Data Packet Air Interface Specification.
- 3GPP2. (2006). cdma2000 High Rate Data Packet Air Interface Specification.
- Agrawal Rakesh, Tyrone Grandison, Christopher Johnson, Jerry Kiernan. (2007). Enabling the 21st Century health care information technology revolution. *Communications of the ACM*, Vol 50, No 2, pp 35-42.
- Bender, P., Black, P., Grob, M., Padovani, R., Sindhushayana, N., & Viterbi, A. (2000, July). CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users. *IEEE Communications Magazine*, 70-77.
- Bhushan, N., Loth, C., Black, P., Attar, R., Jou, Y.-C., Fan, M., et al. (2006, February). CDMA2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview. *IEEE Communications Magazine*, 75-87.
- Choi, L., Ivrlac, M., Steinbach, E., & Nossek, J. (2005). Analysis of Distortion Due to Packet Loss in Streaming Video Transmission over Wireless Communications Links. *Proceedings of the International Conference on Imaging Processing (ICIP'05)*, (pp. 189-192). Génova, Italia.
- Choi, L., Ivrlac, M., Steinbach, E., & Nossek, J. (2005). Bottom-Up Approach to Cross-layer Design for Video Transmission over Wireless Channels. *Proceedings of the IEEE Vehicular Technology Conference (VTC'05)*, (pp. 3019-3023). Estocolmo, Suecia.
- Choi, L., Kellerer, W., & Steinbach, E. (2004). Cross Layer Optimization for Wireless Multi-User Video Streaming. *Proceedings of the International Conference on Image Processing*, (pp. 2047-2050). Singapur, República de Singapur.

- Choi, L.-U., Kellerer, W., & Steinbach, E. (2006). On Cross-Layer Design for Streaming Video Delivery in Multiuser Wireless Environments. EURASIP Journal on Wireless Communications and Networking, 1-10.
- Cypher, David, Nicolas Chevrollier, Nicolas Montavont, and Nada Golmie. (2006) Prevailing over Wires in Healthcare Environments: Benefits and Challenges. IEEE Communications, Vol 4, No 4, pp 56-63.
- Gómez González, Amanda. Programa Nacional de Telesalud [en línea]. (2001) [fecha de consulta 15 de enero 2010]. Disponible en:
<http://ciberhabitat.gob.mx/hospital/telesalud/>
- Haratcherev Ivaylo, Koen Langendoen, Reginald Lagendijk, and Henk Sips, Delft University of Technology, (2005). Link Adaptation and Cross-Layer Signaling for Wireless Video-Streaming in a Shared Medium. IEEE International Conference on Wireless Networks, IEEE Communications and Mobile Computing, pp. 1522-1526
- Hospitales Digitales. Suplemento especial: Política Digital, Nexos [en línea]. (2007), No 35, pp 39-40. [fecha de consulta 15 de enero 2010] Disponible en:
http://www.politicadigital.com.mx/pics/edito/multimedia/476/num_35_multimedia.pdf
- ISO/IEC;. (1994). Open System Interconnection Basic Reference Model: The Basic Model,. International Organization for Standardization.
- Istepanian, R. S. H., E. Jovanov, and Y. T. Zhang. (2004). M-health: Beyond Seamless Mobility for Global Wireless Healthcare Connectivity. IEEE Transactions Information Technology in Biomedicine, Vol 8, No 4, pp. 405-414
- Jou, Y. (2000). Developments in Third Generation (3G) CDMA Technology. Proceedings of IEEE 6th Symposium Spread-Spectrum Technologies and Applications, (pp. 460-464). Newark, New Jersey.
- Kawadia, V., & Kumar, P. (2005). A Cautionary Perspective on Cross-Layer Design. IEEE Wireless Communications, 12 (1), 3-11.

- Khan, S., Steinbach, E., Sgroi, M., & Kellerer, W. (2006, January). Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks. *IEEE Communications Magazine*, 122-130.
- Ksentini, A., Naimi, M., & Guéroui, A. (2006). Toward an improvement of H.264 Video Transmission over IEEE 802.11e through a Cross-Layer Architecture. *IEEE Communications Magazine*, 44 (1), 107-114.
- Lun K.C., (2000) “On-line healthcare,” *Internat Conf on Medical Aspects of Telemedicine*, 1(1):15-9.
- Martínez R. Ignacio. (2006) *Contribuciones a Modelos de Tráfico y Control de QoS en los Nuevos Servicios Sanitarios Basados en Telemedicina*, Tesis Doctoral, Universidad de Zaragoza, España, pp. 17-18.
- Martínez Ramos C. (2009) *Telemedicina. Aspectos Generales*, Reduca (Recursos Educativos). *Serie Medicina* 1 (1): 61-79.
- Poon, Carmen C. Y. and Yuan-Ting Zhang. (2006). A Novel Biometrics Method to Secure Wireless Body Area Sensor Networks for Telemedicine and M-Health. *IEEE Communications*, Vol 4, No 4, pp. 73-81.
- Postel, J. (1981). *Transmission Control Protocol*.
- Postel, J. (1980). *User Datagram Protocol*.
- Qualcomm. (2001). *1xEv Airlink Overview*.
- Qi Bi, Ronald R. Brown, Dongzhe Cui, Asif D. Gandhi, Ching-Yao Huang, and Stan Vitebsky. (2003) *Performance of 1xEV-DO Third-Generation Wireless High-Speed Data Systems*. *Bell Labs Technical Journal*, Vol. 7, No. 3, pp. 97–107.
- Rosenberg J., et al. (2002). *Session Initiation Protocol*, RFC 3261, IETF, [fecha de consulta 15 de noviembre 2010] Disponible en: <http://www.ietf.org/rfc/rfc3261.txt>
- Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (1996). *RFC 1889 RTP: A Transport Protocol for Real-Time Applications*. [fecha de consulta 15 de noviembre 2010] Disponible en: <http://www.ietf.org/rfc/rfc1889.txt>

- Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). RFC 3550 RTP: A Transport Protocol for Real-Time Applications. [fecha de consulta 15 de noviembre 2010] Disponible en: <http://www.ietf.org/rfc/rfc3550.txt>
- Shakkottai, S., Rappaport, T., & Karlsson, P. (2003). Cross-Layer Design for Wireless Networks. *IEEE Communications Magazine*, 41 (10), 74-80.
- Shan, Y., & Zakhor, A. (2002). Cross Layer Techniques for Adaptive Video Streaming Over Wireless Networks. *Proceedings of the International Conference on Multimedia*, pp. 277-280.
- Srivastava Vineet, & Motani Mehul (2005), Cross-Layer Design: A Survey and the Road Ahead, *IEEE Communication Magazine*, pp. 112-119
- Stockhammer, T., Hannuksela, M. M., & Wiegand, T. (2003). H.264/AVC in Wireless Environments. *IEEE transactions On Circuits and Systems For Video Technology* , 13 (7), 657- 673.
- Theodore S. Rappaport, *Wireless Communications Principles and Practice*. 2^a edición. New Jersey. Prentice-Hall PTR. pp. 117-248. 2002.
- Tokgoz, Y., Fan, M., & Smee, J. E. (2005). On the Reverse Link Performance of the cdma2000 1xEV-DO Revision A System with Antenna Array Receivers. *IEEE Globecom 2005*, (pp. 3373-3377).
- Van Der Schaar, M., & Shankar, S. (2005). Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms. *IEEE Wireless Communications*, 12 (4), 50-58.
- Wenger, S. (2003). H.264/AVC Over IP. *IEEE Transactions On Circuits And Systems For Video Technology*, 13 (7), 645-656.
- Wenger, S., Hannuksela, M., Stockhammer, T., Westerlund, M., & Singer, D. (2005). RTP Payload Format for H.264 Video. RFC 3984 .
- Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of teh H.264/AVC Video Coding Standard. *IEEE TransactionsOn Circuits And Systems For Video Technology*, 13 (7), 560-576.

- Yavuz, M., Diaz, S., Kapoor, R., Grob, M., Black, P., Tokgoz, Y., et al. (2006, February). VoIP over cdma2000 1xEV-DO Revision A. IEEE Communications Magazine, 88-95.

Anexo A. Caracterización de Fuentes de Tráfico

Con el objetivo de proveer al Diseño de Cross Layer Propuesto con un proceso de Abstracción de Parámetros de capa de Aplicación adecuado, se realizó la caracterización del tráfico que genera un codificador de video H.263 a nivel de capa de Red.

El proceso de abstracción de parámetros de capa de aplicación para este diseño de Cross Layer requiere identificar el Th a nivel de capa de Red que genera una fuente de tráfico de video codificado con H.263 en función de tres parámetros principalmente:

- Tasa de cuadros por segundo (fps, por sus siglas en inglés),
- Formato de Resolución y,
- Tipo de escena.

Según Choi et al (2006), el tipo de escena conlleva diferentes características de codificación en función del movimiento del plano frontal y del plano de fondo de la imagen. En general, una secuencia de video no es más que un conjunto de imágenes fijas reproducidas en sucesión. Los codificadores de video, generalmente dividen cada imagen fija en bloques denominados macro bloques. Los macrobloques son secciones de imágenes de 16 por 16 pixeles de resolución y son escogidos de ese tamaño debido a que así se pueden procesar mejor a través de la transformada del coseno discreto. Los codificadores de video normalmente, codifican macrobloques, i. e., cada imagen es dividida en un determinado número de macrobloques y posteriormente cada uno de ellos será procesado y codificado. Dependiendo del formato de resolución es la cantidad de macrobloques que se tienen que

codificar por imagen y la tasa de cuadros por segundo es un indicador de cuantas imágenes fijas por segundo debe codificar el codificador.

En resumen, una secuencia de video es procesada como una sucesión de imágenes fijas, cada imagen fija es dividida en macrobloques y cada macrobloque es codificado en función de las características particulares que posea.

Cuando existe mucho movimiento en una secuencia de video, los macrobloques cambian sus parámetros de manera muy rápida, de forma tal que cada vez que llega una imagen nueva, sus macrobloques traen información diferente y estos deben ser procesados conforme van llegando. Si una escena no contiene movimiento en alguna parte de la imagen, ese macrobloque no habrá cambiado mucho con respecto a macrobloques previos que corresponden a la misma zona de la imagen, por lo que no es estrictamente necesario codificarlos de nuevo dado que se enviaría información repetida. La capacidad de un codificador de identificar qué información es redundante y no la codifica es un indicador de su capacidad de compresión. El que el decodificador en la recepción pueda reconstruir lo más fielmente posible la imagen codificada es un indicador de la de la calidad del codificador de video.

Es sencillo intuir entonces, que cuando una escena a ser codificada presenta mucho movimiento en cualquiera de sus partes, el codificador generará más datos a nivel de capa de Aplicación. Esta información es clasificada y empaquetada en unidades de datos de capa de Sesión, típicamente basados en el protocolo RTP (Schulzrinne et al. 2003). La clasificación de la información incluye identificadores de tipos de cuadro. En H.263 se manejan 3 tipos de cuadro distintos: I, P y B. Los cuadros I son los que contienen la información de codificación más importante. Cuando este tipo de cuadros se pierden durante una transmisión, se presenta el peor perfil de distorsión del video. Los cuadros tipo P y B contienen información que no es crucial para la correcta decodificación de los subsecuentes datos que arriben, por lo que normalmente la distorsión que sufre una imagen como consecuencia de perder algunos de sus cuadros P o B es mucho menor que cuando se

pierde un cuadro I. Los cuadros I son enviados normalmente, al principio de una transmisión. Esto es debido a que tienen la información más importante para el decodificador, información sin la cual los subsecuentes datos que arriben no podrán ser decodificados hasta que llegue un nuevo cuadro I. Cada vez que existe un movimiento fuerte en una imagen o se actualiza la imagen, se generan nuevos flujos de datos que contienen cuadros tipo I.

Los tipos de escena son clasificados generalmente en 3 grupos, el primero de ellos son las secuencias que presentan poco o nulo movimiento del plano de fondo y movimiento ligero del plano frontal, el segundo tipo de escena es aquel que tiene mucho movimiento del plano de fondo, pero movimiento relativamente bajo de plano frontal y, finalmente, escenas que tienen movilidad en ambos planos.

La caracterización de las fuentes de tráfico debe ser capaz de especificar qué parámetros de la codificación de video impactan y de qué forma lo hacen en el Th que genera la aplicación a nivel de capa de red. Para lograr esta caracterización se utilizó una herramienta software analizador de protocolos (Wireshark). Esta herramienta permite capturar las tramas que son transmitidas y recibidas en una tarjeta de interfaz de red (NIC, por sus siglas en inglés) y permite hacer análisis relativamente profundos sobre los datos capturados. Lo que interesa en este caso es el Th que se va a demandar a las capas de Radio, por lo que un análisis a nivel de capa de Red es adecuado y la herramienta permite hacerlo.

El software utilizado para llevar a cabo la codificación del video es un software de licencia libre denominado JM Studio de Java. Este software es un conjunto de aplicaciones que permiten capturar, codificar y enviar datos multimedia a través de enlaces basados en IP. Los parámetros que permite controlar son la tasa de cuadros por segundo, el formato de resolución y otros parámetros más específicos del codificador H.263. Cuando se realizó la caracterización de las fuentes de tráfico, se buscó manipular la menor cantidad de parámetros posibles, dado que como fue establecido en el capítulo III de esta tesis, mientras

más parámetros abstraídos se manejen, el crecimiento del conjunto sobre el cual el Optimizador de Cross Layer debe tomar una decisión, crece exponencialmente.

De esta forma se limitaron los parámetros a ser variados a un número de tres: fps, formato de resolución y envío de cuadros clave, más el tipo de escena. Con un total de 7 tasas de cuadros por segundo disponibles, 5 formatos de resolución, 4 valores diferentes de envío de cuadros clave y 3 tipos de escena; fueron evaluadas 420 fuentes de tráfico diferentes.

Limitaciones del hardware disponible nos hicieron reducir el número de formatos de resolución a sólo 4 y se optó por no considerar el período de envío de cuadros clave en H.263 debido a que los primeros resultados mostraron que no impactaba fuertemente el comportamiento del Th Generado. Así, únicamente fueron considerados la fps, el formato de resolución y el tipo de escena. Para cada combinación diferente (de las 84 posibles) se realizaron un promedio de 5 mediciones para dar validez estadística a los resultados, por lo que aproximadamente se realizaron 420 mediciones.

Una vez obtenidos los datos de la caracterización se procedió a su análisis, encontrando que el Th pico generado por la aplicación no estaba en función del tipo de escena sino de los otros dos parámetros, la fps y el formato de resolución. El tipo de escena sí definía por el contrario, la distribución de los datos que representaban el Th medido.

El tipo de escena Talking Head (poco o nulo movimiento del plano de fondo, poco movimiento de plano frontal) genera datos que siguen una distribución caracterizada por un proceso markoviano ON-OFF (García Mendoza, 2007). Cuando se tiene un tipo de escena Talking Head + Movimiento de Fondo se aprecia una distribución que por momentos se asemeja a una fuente de bits variable (VBR, por sus siglas en inglés) y cuando se tiene una escena Talking Head + Movimiento del plano frontal y de fondo, se tiene para bajos formatos de resolución un comportamiento CBR y para altos formatos de resolución un comportamiento poco uniforme y difícil de describir estadísticamente.

La característica común entre los tres tipos de escena es que no modifican el Th pico que genera la aplicación. Lo que define el Th pico alcanzado es en primer lugar, el formato de resolución y en menor grado la fps.

Las secuencias de video utilizadas para la caracterización de fuentes de tráfico tuvieron una duración promedio de 10 segundos y fueron probadas tres cámaras diferentes, Creative PC Cam 600, Creative Instant y, Creative ULTRA. Estas cámaras también fueron caracterizadas para determinar si podían cumplir con las demandas de la aplicación de video, siendo la tercera cámara la que superó todas las pruebas y por ello fue utilizada para la caracterización de las fuentes y para la transmisión de videoconferencia durante la obtención del modelo empírico de PLR.

Al analizar el tráfico generado por las fuentes con el software wireshark se detectó que en los momentos en los que se generaba mayor tráfico era cuando la aplicación enviaba los cuadros I y se procedió a verificar en todas las pruebas realizadas con anterioridad. Pudimos comprobar que cuando la aplicación envía la información de video más importante es cuando genera más tráfico. Cuando hay más movimiento en la imagen, más cuadros I se generan y genera mayor tráfico la aplicación, pero no rebasa el umbral determinado por la fps y el formato de resolución.

El hecho de que la aplicación se comporte como se ha descrito en el párrafo anterior, nos permite considerar las fuentes de tráfico como si fueran una fuente CBR. Considerar la fuente de tráfico como un CBR es adecuado porque garantizamos ofrecer el mejor Th disponible para la aplicación, en la realidad, se pudo comprobar que la distribución de la fuente de tráfico no es normal ni constante, que esto depende más bien, del tipo de escena. Sin embargo, intentar satisfacer el Th Pico Generado a nivel de capa de Red e inferiores, garantiza la integridad del video porque no siempre es requerido ese nivel de throughput. Con esta aproximación se ofrece una protección extra al envío de información de video, ya que si se llega a perder algo por condiciones del canal, existe una probabilidad de que lo que se haya perdido no haya sido información crítica o que inclusive, no se pierda

información debido a que en ese instante no se estaba generando el Th Pico sino un valor mucho menor.

La figura 28 es una reproducción de la figura 5 mostrada previamente y que es el resultado gráfico del proceso de caracterización de fuentes de tráfico que ha sido descrito.

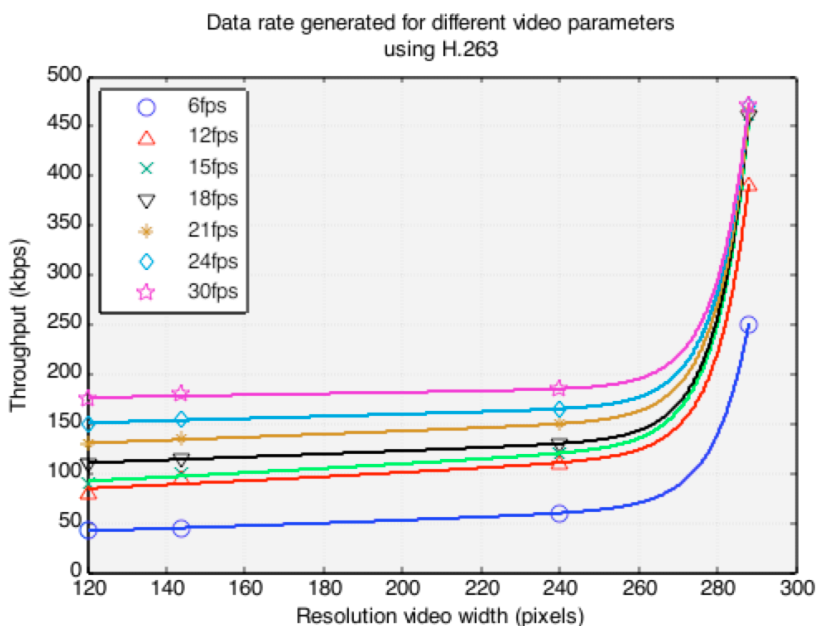


Figura 2. Resultado del proceso de caracterización de fuentes de tráfico.

El resultado de la caracterización ofrece valores discretos y bien determinados de Th Pico. Las líneas que unen los puntos de la figura 28 sólo son representativas del comportamiento exponencial que sigue la generación de tráfico con H.263, el modelo que representa este comportamiento está mostrado en la gráfica. Cada punto en la figura 28 representa un valor de Th Pico Generado que se asume constante y cuyas unidades son Kbps. Esto, al igual que en las gráficas anteriores presentes en la tesis, representa un Th Equivalente, i. e., si la aplicación se mantuviera constante por lo menos un segundo, generaría el tráfico señalado por la ordenada. La figura 28 representa la caracterización de 28 fuentes de tráfico, con 7 tasas de cuadros por segundo diferentes y con 4 formatos de resolución diferentes. Las unidades de la abscisa son el Ancho de Resolución en pixeles debido a que es común hacer referencia a los formatos de resolución únicamente por el ancho del cuadro.

Es de notar que 21 fuentes de tráfico están localizadas entre los 50 y 200 kbps de Th equivalente y que 5 fuentes de tráfico con ancho de resolución de 288 (el formato de resolución es de 352x288 pixeles) generan prácticamente el mismo Throughput. Asumimos este comportamiento a deficiencias del hardware utilizado, ya que al momento de estar generando las secuencias de video para caracterizarlas, los recursos del equipo de cómputo utilizado eran explotados al máximo. Aún así, se debe recordar que el valor mostrado es el promedio de las 25 mediciones realizadas en esos puntos.

Anexo B. Modelo Empírico de Pérdidas de Paquetes y Estado del Canal

En los capítulos previos de la tesis ya ha abordado con suficiente énfasis las ventajas que presentó el haber obtenido un modelo empírico de pérdida de paquetes y el estado del canal inalámbrico. Más importante es quizá, que el hecho de que este modelo haya sido obtenido de manera empírica le proporciona validez a los resultados de la simulación para las consideraciones que ya se han hecho.

El objetivo de este anexo no es repetir o concentrar los aspectos ya mencionados tanto de las implicaciones o las consideraciones que se deben tomar en cuenta al ofrecer resultados apoyándose en el modelo, como de la utilidad que tuvo para el desarrollo del proyecto.

El objetivo de este anexo es dejar establecido de manera clara la forma en la que fueron llevadas a cabo las mediciones con el afán de clarificar cualquier posible duda que surja respecto a la metodología empleada.

Las terminales de acceso IS-856 utilizadas son tarjetas de datos que pueden ser conectadas a equipos de cómputo convencionales a través de puertos USB o ranuras PCMCIA o ExpressCard. Una computadora portátil fue utilizada como extremo móvil de la videoconferencia con una de las tarjetas conectada a ella, el equipo de cómputo cuenta con software de captura de tráfico y con software que permite establecer la video conferencia.

Para evitar variaciones en el proceso de codificación y captura de video, se grabó una secuencia de video que la computadora portátil podía leer y enviar a través del software de videoconferencia como si el video grabado fuera una sesión de videoconferencia en vivo. Se tomó esta opción principalmente por la complejidad de conducir por las calles con el Laboratorio Móvil y aparte estar controlando manualmente la videoconferencia. El video grabado y transmitido tiene una duración aproximada de poco más de 10 segundos.

Una vez que se ubicaba al laboratorio móvil en una zona de cobertura dónde la lectura de potencia fuera la deseada, se alcanzaba la velocidad requerida y se iniciaba la sesión de videoconferencia de 10 segundos. Un extremo de la videoconferencia era el laboratorio móvil y el otro extremo de la videoconferencia fue otro equipo portátil con el mismo software que el equipo usado como terminal móvil, con una tarjeta EV-DO conectada a ella y posicionada en una zona donde existiera trayectoria de visibilidad directa entre la tarjeta EV-DO y la radio base y la potencia detectada por el móvil fuera máxima.

Este arreglo asegura que los datos enviados por la terminal móvil experimentarán los mayores problemas de transmisión durante su paso por el canal inalámbrico, que es el canal que se desea evaluar. Al llegar los datos desde el laboratorio móvil hasta la radio base, la información es transportada a través de la red interna que conecta las celdas del proveedor de servicio, prácticamente sin pérdidas y se asume que el canal inalámbrico que enlaza a la terminal no móvil de la video conferencia con la radio base no presenta problemas debido a las condiciones mencionadas.

Lo anterior nos permite establecer que el comportamiento de pérdida de paquetes que estamos observando en la comunicación, corresponde efectivamente a los problemas que enfrenta el flujo de información cuando atraviesa el canal radio desde la terminal móvil hasta la radio base.

Ya se discutió también en el último capítulo de la tesis los detalles respecto a la relación que existe entre los valores que señala el RSSI de la terminal de acceso y, las pérdidas

sufridas en el enlace RL. A través de numerosas mediciones realizadas fue posible obtener los valores promedio de PLR que presenta el enlace RL de IS-856 para velocidades vehiculares (80 km/hr) en función de la potencia que percibe el móvil desde la radio base y que es evaluada a través del RSSI. El valor RSSI ofrece una medición de la intensidad de la potencia de la señal que percibe el móvil desde la radio base. Propiamente este valor mide la potencia percibida por el móvil en el enlace FL, sin embargo, se encontró que existe una relación estrecha entre las variaciones de potencia del RSSI y la PLR que experimenta el móvil en el enlace RL, valores representados en la figura 29.

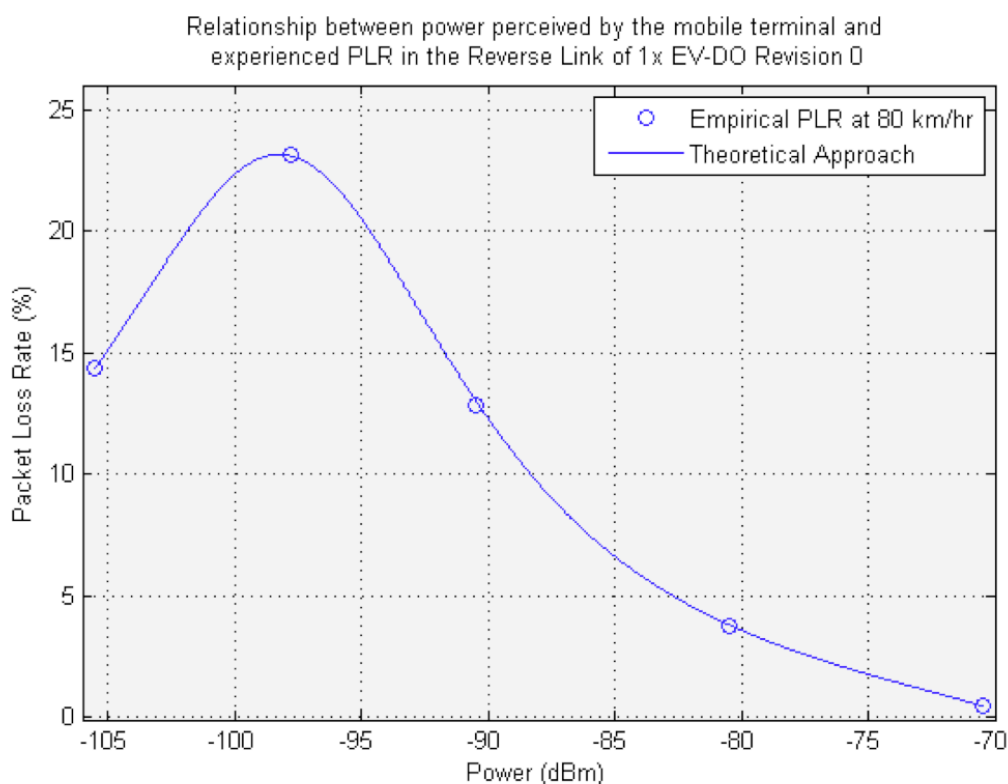


Figura 3. Modelo empírico de pérdidas y estado del canal.

La forma en la que fue obtenido el modelo anterior fue midiendo los valores de potencia indicados por el RSSI de la terminal de acceso EV-DO Rev.0 mientras un equipo de cómputo (al que estaba conectado la terminal EV-DO) establecía mantenía y terminaba una sesión de videoconferencia de parámetros tales que no superaban el T_h Máximo permitido por la tecnología, i. e., 153.6 kbps; mientras la unidad móvil se desplazaba a velocidades

vehiculares. Los puntos en la gráfica son la PLR promedio observada en ese valor de potencia promedio, ambos promedios obtenidos de las múltiples mediciones realizadas.

Se realizaron dos series de experimentos para dos tipos de velocidades diferentes, alta velocidad (80 km/hr) y baja velocidad (20 km/hr). Los límites de potencia de las terminales de acceso EV-DO que fueron utilizadas (Audiovox PC5740 y Franklin CDV-550) van desde los -66 dBm hasta los -110 dBm. En particular las mediciones del canal fueron realizadas con la tarjeta Audiovox que para el software de videoconferencia se comporta como si fuera un módem y a petición explícita de comandos del sistema operativo la terminal Audiovox despliega los valores del RSSI que percibe. Los resultados mostraron que a bajas velocidades, las pérdidas que sufre la información son prácticamente nulas y por lo tanto ya no fueron consideradas en los análisis subsecuentes. Se encontró evidencia de resultados reportados en donde se muestra como a velocidades bajas las pérdidas son bajas y solo se maximizan a velocidades de 80 km/hr, como muestra la siguiente figura.

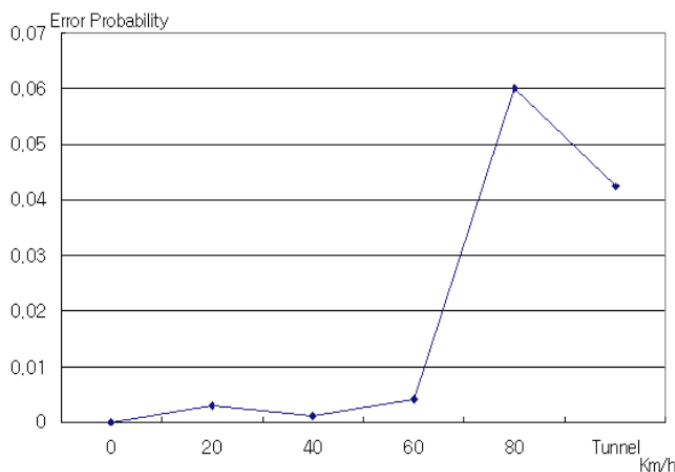


Fig. 6. Error probability in terms of vehicle speed

El modelo matemático mostrado en la figura 29 fue obtenido con ayuda de la herramienta MatLab y es posible ver que existe alta correlación entre el modelo presentado y los valores empíricos obtenidos de PLR. Este modelo es el resultado del proceso de Abstracción de

Parámetros de capa de Radio dado que nos permite predecir un valor de PLR para una potencia dada.

Anexo C. Parámetros de QoS

Tanto en las videoconferencias tradicionales como en las móviles, los parámetros de QoS deben mantenerse constantes no sólo dentro de las redes en dónde se genera el tráfico de las aplicaciones, sino a través de todas las redes de transporte que son requeridas para establecer el enlace. Durante todo el transporte es fundamental reservar un nivel de Th adecuado para la transmisión de la videoconferencia, en ambos sentidos, debido a que la limitación en este parámetro impacta negativamente en los demás debido a los efectos de la congestión.

El congestionamiento es la condición en la que una red recibe más paquetes de los que puede despachar en un momento determinado. Esta condición genera procesos de encolamiento y con esto último se incrementa el retardo extremo a extremo.

Existe un retardo inherente dentro de los procesos de comunicación, retardo generado por el tiempo que toma en ser procesada la información por cada uno de los dispositivos de red que tienen que manejar la información hasta que esta sea entregada en su destino final. Mientras más dispositivos existan entre el origen y destino de los datos, mayor es el retardo en la comunicación. Estos retardos son predecibles siempre que los dispositivos de red no estén congestionados.

En entornos reales los dispositivos de red rara vez son capaces de despachar todo el tráfico que reciben sin tener que generar procesos de encolamiento. Es por ello que se han desarrollado múltiples mecanismos para identificar la prioridad que tienen cada uno de los

flujos que los equipos de red reciben. Esas prioridades asignadas a los flujos de datos, son lo que genera los diferentes niveles de calidad de servicio y en los procesos de colas implica la capacidad de seleccionar adecuadamente a qué cola se le da preferencia de servicio. Los flujos de datos que no son sensibles al retardo tienen la más baja calidad de servicio puesto que existen mecanismos en la capa de Transporte que enfrentan los problemas de congestión y que, en caso de que cierto dispositivo de red decida descartar algunos paquetes de ese flujo, pueden solicitar retransmisiones de dichos paquetes (Postel, J., 1981).

Las aplicaciones sensibles al retardo son encapsuladas en segmentos²⁸ UDP a nivel de capa de Transporte (Postel, J., 1980) porque es más importante que los segmentos sean despachados con la mayor prontitud posible (para minimizar el retardo) a que lleguen exactamente todos al destino y en orden. Esto último funciona bien para aplicaciones en las que la integridad de la información es más importante que el retardo que experimentan.

Las aplicaciones sensibles al retardo descartan los segmentos UDP que llegan fuera de tiempo e inclusive fuera de secuencia, de ahí que mantener en niveles bajos el retardo y la variación del retardo sea fundamental. De esta situación también se desprende la importancia de que los canales de comunicación presenten bajas tasas de pérdidas de paquetes.

Por lo anteriormente expuesto, la forma en la que se garantizan los mejores niveles de QoS en las redes de transporte es asegurando niveles de Th exclusivos para las aplicaciones que así lo demanden. Si se hiciera una analogía entre la capacidad de transporte de una red y una tubería de agua, reservar niveles de Th equivaldría a introducir un tubo más pequeño dentro de la tubería de agua para que por ahí fluyera de manera exclusiva un determinado flujo. Mientras más Th se desee reservar, más grande debe ser el tubo interno.

²⁸ Las PDUs (del inglés, Protocol Data Unit) de las diferentes capas del modelo OSI son nombradas de diferente manera para poder hacer una referencia adecuada al tipo de datos del que se está hablando. La PDU de capa de Transporte son los segmentos. La PDU de capa de Red son los paquetes.

Si el Th se mantiene constante y reservado durante todo el enlace, el retardo se mantiene en niveles mínimos así como su variación. Por el contrario, si el Th que demanda la aplicación no está reservado, habrá momentos en los que los paquetes de datos que integran al flujo de información sensible al retardo, pasen más tiempo en colas de espera generando más retardo y con ello generando mayores pérdidas de paquetes desde el punto de vista de la aplicación.

Es claro entonces que uno de los parámetros de QoS que más impactan en el nivel de servicio de los datos es el Th que se le ofrece a la aplicación.

Esto último ha mostrado su veracidad en la transmisión de tráfico sensible al retardo en las redes cableadas, en donde los mecanismos comunes de QoS buscan identificar los flujos de información que deben ser tratados de manera especial y se les reserva un Th específico.

Lo anterior no siempre se puede llevar a cabo en las redes inalámbricas, debido a que los desvanecimientos de pequeña y gran escala y el fenómeno de múltiples trayectorias del canal radio generan errores durante la transmisión (Rappaport, 1996) que, eventualmente, se percibirán como pérdidas de paquetes a nivel de capa de Red, o pérdidas de Tramas a nivel de capa de enlace de Datos o pérdidas de bits a nivel de capa Física. Es bien sabido que además estos problemas que presenta el canal radio no pueden ser predichos de manera determinista, sino más bien mediante procesos estadísticos, dado que los efectos del canal son variantes en tiempo y en frecuencia.

Esta variabilidad del canal hace prácticamente imposible la reservación de Th para las aplicaciones que así lo demanden, debido a que el Th máximo que soporta el canal no siempre será el mismo y aunque las capas inferiores del modelo OSI (que son las que lidian con el canal inalámbrico) reservaran toda su capacidad de transmitir información para un cierto tráfico sensible al retardo, dicha capacidad puede ser simplemente, inferior a lo demandado, lo que genera un proceso de encolamiento.

Más aún, la misma variabilidad del canal hace que las colas no puedan ser despachadas en intervalos regulares de tiempo, éstas son servidas, nuevamente, en función de las condiciones del canal. Este fenómeno hace variar los tiempos de espera de las Tramas²⁹ en las colas, lo que provoca que el retardo extremo a extremo también varíe, i. e., se genera más Jitter.

Por otra parte, las pérdidas de paquetes en flujos de datos sensibles al retardo, suelen impactar de forma diferente a la aplicación dependiendo de qué paquete es el que se pierde. Esto debido a que en ocasiones se envía información de redundancia o información con detalles que no necesariamente hacen imposible la decodificación del flujo total de datos. Esto es, necesariamente, dependiente del codificador que se utilice para la aplicación en cuestión. Diferentes codificadores son susceptibles a diferentes niveles de pérdida de información.

En resumen, existen dos procesos por los cuáles se pueden generar niveles bajos de parámetros de QoS:

- Congestionamientos y,
- Pérdidas durante la transmisión.

Ante el primer proceso, el Th que es ofrecido a las aplicaciones disminuye y el retardo se incrementa dado que los paquetes de información o las tramas, deben esperar a ser despachados. Algunos de los mecanismos desarrollados en (Postel, 1981) fueron diseñados para mitigar los efectos de la congestión, sin embargo, éstos introducen retardo por los procesos de envío de acuses de recibo y retransmisiones, de ahí que las aplicaciones sensibles al retardo utilicen en capa de transporte los mecanismos propuestos en (Postel, 1980). Ante la congestión, los segmentos que arriben tarde o fuera de secuencia, deben ser descartados.

²⁹ La Trama (del inglés, Frame) es la PDU de capa de enlace de datos.

Lo anterior no corrige el problema del congestionamiento, simplemente mitiga sus efectos en las aplicaciones sensibles al retardo. Esto debido a que los orígenes de un proceso de congestionamiento no se encuentran (normalmente) en la capa de transporte, sino en las capas inferiores donde se forman las colas y se enfrenta al canal³⁰.

Es por ello que otra de las causas por las que se generan bajos niveles de QoS son las pérdidas que se presentan durante la transmisión. El Th disminuye si los errores que se presentan en la capa física son elevados. Esta disminución del Th generará procesos de encolamiento, se incrementarán los retardos y si el canal tiene un comportamiento variable (e. g., el canal radio) también se incrementará el jitter.

Es claro que si existen niveles elevados de pérdidas en el canal, se generará un proceso de congestionamiento en los accesos a dicho canal y los parámetros de QoS se verán afectados; las aplicaciones sensibles al retardo observarán estos efectos como pérdidas de paquetes.

³⁰ Los orígenes de un proceso de congestionamiento también pueden encontrarse en un número elevado de tráfico que pretende circular por un mismo canal.

Anexo D. Redes Inalámbricas de Tercera Generación

Los principales requerimientos que fueron establecidos para las redes inalámbricas de tercera generación (3G) fueron las necesidades de servicios de datos con tasas de transferencia mayores y una mejor eficiencia espectral (Ojanperä y Prasad, 1998), esto último relacionado con las limitaciones que la FCC (Comisión de Comunicaciones Federales, por sus siglas en inglés) impuso en el espectro radio eléctrico para las comunicaciones telefónicas móviles.

Los objetivos que se marcaron para las interfaces aéreas de las redes 3G eran la cobertura y movilidad a una tasa de 144 kbps y cobertura y movilidad limitada para 2 Mbps, la ya mencionada eficiencia espectral y además, alta flexibilidad para la introducción de nuevos servicios. Esto es, las redes inalámbricas de 3G debían soportar los servicios tradicionales de voz para los que originalmente fueron diseñados y ofrecer servicios de comunicación de datos con caudales eficaces elevados.

Sin embargo, el concepto de ofrecer servicios de datos a los usuarios móviles a principios de este siglo, estaba basado en la idea de los enlaces asimétricos que eran utilizados en las líneas de datos cableadas en las cuales, aún era típico utilizar módems con enlaces de 64 kbps. En los países más desarrollados ya se utilizaban enlaces xDSL con los cuáles era comparado el desempeño que debían ofrecer los servicios de datos de las redes inalámbricas 3G (Bender et al. 2000).

Bajo este contexto, era común visualizar las comunicaciones de datos con un énfasis mayor en la diferencia que existía entre los requerimientos de la mayoría de las aplicaciones de datos y los requerimientos de las conversaciones que ya soportaban las redes (i. e., las conversaciones telefónicas en las que los parámetros de QoS son simétricos), lo que se resume en la asimetría del tráfico y la tolerancia al retardo (Bender et al. 2000).

Nuevas aplicaciones de datos demandadas son aquellas de tipo conversacional también, en las que se incluyen servicios de VoIP, streaming video, juegos en red móviles, videoconferencias o video telefonía. Estas aplicaciones de datos difieren de las demás en la necesidad de simetría en la comunicación y la susceptibilidad al retardo; las nuevas redes de tercera generación han sido diseñadas para satisfacer tales demandas (Bhushan et al. 2006). Lo anterior implica más Th y el soporte de diferentes niveles de QoS en el enlace RL.

D.1 cdma2000 1xEV-DO (IS-856)

Hasta el momento, sólo existe en México un proveedor de servicios que ofrece comunicaciones inalámbricas de tercera generación EV-DO, la tecnología que emplea dicho proveedor es cdma2000 1xEV-DO (3GPP2, 2000) en todo el país y en ciertos lugares ofrece también cobertura para cdma2000 1xEV-DO Rev. A (3GPP2, 2006).

La particularidad de estas tecnologías es que están diseñadas para soportar caudales eficaces elevados en el enlace RL (en EV-DO Rev. 0 el máximo Th es de 153.6 kbps mientras que en EV-DO Rev. A el máximo Th es de 1.8 Mbps³¹) lo que implica que

³¹ A la primera versión de EV-DO se le conoce como revisión 0, estandarizada por el 3GPP2 en el año 2000. TIA también estandarizó esta versión con el nombre de IS-856. La siguiente revisión de EV-DO se conoce como EV-DO Rev. A y también fue estandarizada por la TIA bajo el nombre de IS-856A. El nombre que el 3GPP2 otorga a ambas revisiones es el mismo: cdma2000 High Rate Data Packet Air Interface Specification (Bhushan et al. 2006).

aplicaciones de datos de tipo conversacional de alta demanda en throughput, e. g., videoconferencia, pueden ser soportadas.

El estándar cdma2000 1xEV-DO Revision 0 (conocido como DO Rev 0) fue diseñado con la visión de un “Ethernet inalámbrico móvil de área amplia” desde el punto de vista de las capacidades que debía ofrecer (Bender et al. 2000). Como resultado de esta visión, se generó un sistema de paquetes de datos inalámbricos de alta capacidad con mejoras considerables en el enlace FL y en la cobertura respecto a los sistemas tradicionales de cdma2000 como IS-95 e IS-2000 (Bhushan et al, 2006 y Jou, 2000).

Adicionalmente al incremento en el Th del enlace FL, DO Rev 0 provee soporte para diferentes niveles de QoS, el uso de paquetes de mayor tamaño codificados con turbo códigos de baja tasa, transmitidos con esquemas de modulación y tasas de codificación adaptativas, ARQ híbrido en el enlace FL (H-ARQ) y diversidad de antenas en el receptor. DO Rev 0 soporta calidad de servicio por flujo en el enlace FL y calidad de servicio por terminal móvil en el enlace RL.

D.1.1 Forward Link

Con el objetivo de ofrecer compatibilidad hacia atrás, los sistemas EV-DO utilizan un ancho de banda de 1.25 MHz y ensanchan (o esparcen) el espectro mediante secuencia directa (DSSS, Espectro Esparcido por Secuencia Directa, por sus siglas en inglés) a una tasa de 1.2288 Mchips/s. El tamaño de la trama en DO Rev 0 es de 26.66... ms con 16 ranuras, a diferencia de los sistemas IS-2000. El canal piloto se transmite a máxima potencia durante 96 chips cada media ranura proveyendo así, una referencia para la demodulación coherente del tráfico y los canales de la capa MAC y además un muestreo de las condiciones del canal cada 1200 Hz. Estas muestras le ayudan a la terminal de acceso

(Access Terminal en inglés, AT)³² a determinar el estado del canal en el futuro cercano y por ende, a determinar el T_h que soportará en el enlace FL. Esto es un mecanismo para poder proveer una rápida adaptación de los esquemas de modulación y tasas de codificación en diferentes situaciones del canal radio.

El máximo T_h que soporta DO Rev 0 en el enlace FL es 2457.6 kbps y el mínimo es 38.4 kbps. Los esquemas de modulación utilizados van desde QPSK hasta 16QAM y utiliza 2 diferentes tasas de codificación (Qualcomm, 2001).

D.1.2 Reverse Link

DO Rev 0 tiene las mismas características en el enlace RL que su predecesor IS-2000, las que incluyen el control de potencia de lazo cerrado. Este control de potencia es necesario debido a la existencia del efecto cercano-lejano (near-far effect, en inglés). Dicho efecto surge de la interferencia por acceso múltiple presente en CDMA. Desde el punto de vista de la Red de Acceso (Access Network, AN) y debido a los mecanismos de propagación, una AT que se encuentre más cerca de la AN será percibida con mayor potencia que una AT que se encuentre en la frontera del sector o de la celda. Esto ocasionará que las terminales más cercanas dominen en interferencia a las más lejanas. (Ojanperä y Prasad, 1998). Para resolver este problema, la estación base controla la potencia con la que los móviles transmiten intentando así, percibir a todas las terminales con la misma potencia. Los móviles que se encuentre más cerca de la radio base transmitirán con menor potencia que las que se encuentran a mayor distancia.

Además de lo anterior, DO Rev 0 introduce la capacidad de la AN de medir la relación I_o/N_o , donde I_o es la potencia total recibida y N_o es el piso de ruido térmico, lo anterior es posible debido a que existe un periodo de silencio para todas las terminales, permitiendo a la radio base medir el N_o . En DO Rev 0 las AT reciben un bit de actividad en el enlace RL de cada sector en el que se encuentren activas indicándoles cuando la relación I_o/N_o supera

³² En EV-DO una terminal de acceso es la terminal móvil, la red de acceso (Access Network, AN) es la radio base.

un umbral permitiéndoles determinar la tasa de transferencia a la que deben transmitir. Cabe señalar que las AT pueden estar activas en más de un sector para efectos de traspasos suaves (Bhushan et al. 2006 y Ojanperä y Prasad, 1998).

Este estándar soporta un par de tasas de codificación adaptativas y utiliza BPSK como esquema de modulación en el enlace RL, siendo el máximo T_h de 153.6 kbps y el mínimo de 9.6 kbps (Qualcomm, 2001).