



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

CALIDAD DE EXPERIENCIA EN EL SERVICIO BANDA ANCHA MÓVIL

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN INGENIERÍA
DE REDES DE COMUNICACIONES

JORGE IGNACIO SANDOVAL ARENAS

SANTIAGO DE CHILE

JUNIO 2011



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

CALIDAD DE EXPERIENCIA EN EL SERVICIO BANDA ANCHA MÓVIL

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN INGENIERÍA
DE REDES DE COMUNICACIONES

JORGE IGNACIO SANDOVAL ARENAS

PROFESOR GUÍA:
ALFONSO EHIJO BENBOW

MIEMBROS COMISIÓN:
PATRICIO PARADA SALGADO
PATRICIO VALENZUELA CANO
PATRICIO RODRÍGUEZ PIZARRO

SANTIAGO DE CHILE

JUNIO 2011

RESUMEN INFORME FINAL
PARA OPTAR AL GRADO DE
MAGÍSTER EN INGENIERÍA DE
REDES DE COMUNICACIONES
POR: JORGE SANDOVAL ARENAS
PROF GUÍA: SR. ALFONSO EHJO

CALIDAD DE EXPERIENCIA EN EL SERVICIO BANDA ANCHA MÓVIL

El servicio de banda ancha móvil es uno de los de más alto crecimiento tanto en telefonía móvil como en telecomunicaciones en general. Acceso inalámbrico móvil a Internet con alta disponibilidad y presente en todas las zonas urbanas e incluso con una amplia cobertura rural, sumado a su bajo precio, ha sido la clave para este crecimiento.

Esto ha significado que los operadores móviles han consolidado su ingreso al mundo de los ISP, pero con los exigentes usuarios de telefonía móvil y el competitivo mercado es necesario entregar un servicio de alta calidad.

De este modo, se introduce el concepto de calidad de experiencia (QoE, Quality of Experience), que muestra la percepción extremo a extremo que tiene el usuario del servicio, y surge como necesidad establecer los parámetros de medición apropiados (KQI, Key Quality Indicator) que sirven de puntos de control para que ésta se cumpla y además se puedan incorporar tareas que mejoren la experiencia de servicio de los abonados.

Se estudia la tecnología y arquitectura del operador móvil, su evolución y el comportamiento de los abonados, lo que permite modelar el servicio para luego introducir puntos de control y mejoras en el sistema, con el objetivo de entregar una calidad de experiencia satisfactoria y altamente diferenciadora.

Adicionalmente se analiza el comportamiento de los abonados, determinando el tipo de contenido que éstos prefieren, los protocolos que utilizan y cómo éstos afectan a las redes. Se describe la tecnología utilizada por los operadores móviles para brindar el servicio e indican los puntos críticos para el diseño y planificación de la red.

Finalmente, se define un indicador de control para la calidad de experiencia del servicio y se estudian diversas tareas en la red que determinan niveles de mejora del servicio.

Este trabajo explora las diferentes alternativas que tienen los operadores para incorporar y mejorar la calidad de experiencia que ofrecen a sus usuarios, entregando una metodología de modelamiento que sirve de punto de partida para esta tarea.

Del mismo modo, es posible a partir de este trabajo profundizar en los mecanismos y facilidades que puede proporcionar una tecnología específica, como es el caso específico de LTE o LTE advance con mira a ofrecer servicios de nuevas generaciones.

Dedicatoria

A mi familia que siempre ha estado junto a mí. Loreto mi compañera de vida que camina a mi lado y me llena de felicidad, Francisca que ilumina mis mañanas y me muestra que el futuro siempre será mejor y al que viene en camino que me llena de esperanza y me invita a comenzar de nuevo.

Agradecimientos

Mis sinceros agradecimientos a quienes me apoyaron en la elaboración de este trabajo. Alfonso, que compartiste opiniones, confiaste en mí y me guiaste desde el inicio. A mi Universidad, que abrió mi mente plantando en mí la semilla del conocimiento y la inquietud en la docencia.

También agradezco a quienes sin razón aparente compartieron sus trabajos, sus emociones y alimentan sus almas viviendo la vida a plenitud.

Tabla de Contenidos

Índice de Figuras y Tablas	4
Abreviaturas	8
Capítulo 1. Introducción.....	11
1.1 Motivación	12
1.2 Aportes.....	12
1.3 Objetivos	13
1.3.1 Objetivo General.....	13
1.3.2 Objetivos Específicos	13
1.4 Descripción	13
Capítulo 2. Antecedentes	14
2.1 Evolución Internet	14
2.2 Descripción de la Tecnología y de los Servicio de Datos.....	14
2.2.1 Evolución de Tecnologías de Acceso	15
2.2.2 Packet Scheduling.....	19
2.3 Situación Local.....	21
2.4 Desempeño Protocolos de Transporte.....	23
2.4.1 TCP	23
2.4.2 Control de Flujo y Throughput de TCP.	24
2.4.3 UDP.....	25
2.4.4 Comparación de rendimiento entre los protocolos TCP y UDP	25
2.5 Descripción Servicios de Internet.....	26
2.6 Descripción de Servicio BAM	27
2.6.1 Movilidad	27
2.6.2 Velocidad de Datos.....	28
2.6.3 Caracterización de Servicios	29
2.6.4 Servicios Intensivos	30
2.7 Variables de Desempeño de una Red Móvil	31
2.7.1 Latencia	31
2.7.2 Throughput	32
2.7.3 Pérdida de Paquetes	33

2.7.4	Rendimiento de Aplicaciones	34
2.8	Costos de Red	35
2.8.1	Modelo de Costos en BAM	35
2.8.2	El dilema Intensidad v/s Volumen.....	36
2.8.3	Usuarios Ofensores	37
2.8.4	Fair Use	37
2.8.5	Disminución de Costos v/s Calidad de Experiencia.....	38
2.9	Calidad de Experiencia	39
2.9.1	Definición KQI.....	39
2.9.2	Modelo de QoE.....	41
Capítulo 3.	Metodología.....	43
3.1	Introducción.....	43
3.2	Impacto de Protocolos.....	43
3.3	Análisis de Tráfico.....	43
3.4	Simulación de Servicios.....	44
3.5	Medición.....	44
3.6	Indicador de Calidad de Experiencia.....	45
Capítulo 4.	Resultados	47
4.1	Modelamiento del Servicio BAM	47
4.1.1	Perfil de Navegación.....	47
4.1.2	Movilidad	49
4.1.3	Intensidad de uso.....	51
4.1.4	Usuarios Intensivos	52
4.1.5	Fair Use	55
4.2	Desempeño de Aplicaciones.....	60
4.2.1	Throughput	60
4.2.2	Control de flujo en TCP	65
4.2.3	Aplicaciones Interactivas	66
4.2.4	Aplicaciones Intensivas	67
4.3	Modelo Calidad de Experiencia.....	67
4.4	Modelo de Costos	70
Capítulo 5.	Discusión.....	75
5.1	Caracterización de usuarios BAM	75
5.1.1	Movilidad	75

5.1.2	Concentración horaria	75
5.1.3	Usuarios Intensivos.	76
5.1.4	Concentración geográfica.....	76
5.1.5	Aplicaciones y servicios.....	77
5.1.6	Políticas de Fair Use.....	77
5.2	Desempeño de Aplicaciones.....	78
5.3	Calidad de Experiencia	79
5.4	Análisis de Costos.....	80
5.5	Mejoras	81
Capítulo 6.	Conclusiones.....	84
Capítulo 7.	Referencias Bibliográficas	86
Apéndice A.	Arquitectura de Tecnologías Móviles.....	89
A.1	GSM/GPRS	89
A.2	Arquitectura UMTS.....	90
A.3	Long Term Evolution	91
A.4	WiMax	95
Apéndice B.	Desempeño de Protocolos de Transporte.....	97
B.1	Descripción de TCP	97
B.2	Funcionamiento.....	98
B.3	Control de Flujo y Throughput de TCP.....	100
B.4	Pérdida de Paquetes.....	102
B.5	Descripción de UDP.....	106
B.6	Comparación Rendimiento TCP/UDP	107
Apéndice C.	Descripción Servicios de Internet	110
C.1	Web Browsing.....	110
C.2	Streaming y Flash Video	111
C.3	P2P	114
C.4	Network Storage	115
C.5	Redes Sociales	115
C.6	Otros Servicios.....	116
Apéndice D.	Caracterización de Servicios	117

Índice de Figuras y Tablas

Figura 2.1. Evolución Redes Móviles	15
Figura 2.2. Constelación de señales para QPSK, 16QAM y 64QAM	15
Figura 2.3. Modulación TDMA y WCDMA.	16
Figura 2.4. Modulación OFDM y FDMA.	17
Figura 2.5. Modulación OFDMA y SC-FDMA.	17
Figura 2.6. Incremento de las velocidades móviles	18
Figura 2.7. Tiempo de descarga de archivos según tecnología de acceso	18
Figura 2.8. Arquitectura EPS y Legacy	19
Figura 2.9. Distribución banda 850 y 1900 MHz utilizada por actuales operadores móviles y banda 1700/2100 MHz de nuevos operadores.....	21
Figura 2.10. Crecimiento banda ancha móvil (BAM) y Fija (BAF) en Chile	22
Figura 2.11. Market share BAM en Chile	22
Figura 2.12. Adaptación de TCP luego de pérdida de paquetes con y sin Fast Recovery	24
Figura 2.13. Variación en flujos de tráfico TCP v/s UDP al pasar a un canal de menor capacidad.	25
Figura 2.14. Overhead protocolos de transporte	26
Figura 2.15. Movilidad en tecnologías móviles	28
Figura 2.16. Clasificación de Servicios de Banda Ancha	30
Figura 2.17. La latencia es la suma de los retardos parciales.....	31
Figura 2.18. La información enviada corresponde a la definida por la ventana de recepción.....	32
Figura 2.19. La pérdida de paquetes puede ocurrir en cualquier punto de la comunicación.	33
Figura 2.20. El servicio es afectado por todos los nodos a lo largo de la comunicación.	34
Figura 2.21. Modelo de Costos.	36
Figura 2.22. Diferentes perfiles de tráfico aportan con distinta intensidad en la hora cargada.	36
Figura 2.23. Comparación de efecto de cambio de MBR y prioridad como política de Fair Use.....	38
Figura 2.24. Cuatro polos de la calidad de servicio.....	40
Figura 2.25. Proceso para obtener la calidad de funcionamiento extremo a extremo....	41

Figura 3.26. Algoritmos de tomas de muestras.....	45
Figura 3.27. Frecuencia correcta para la toma de muestras.....	45
Figura 3.28. Mediciones en la navegación web.....	46
Figura 4.29. Uso de servicios por categoría para usuarios BAM.....	47
Tabla 4.30. Principales servicios utilizados en BAM.	48
Figura 4.31. Principales protocolos utilizados por los usuarios de BAM.	48
Tabla 4.32. Características servicios más utilizados por los usuarios de BAM.	49
Tabla 4.33. Categorización por movilidad.	50
Figura 4.34. Movilidad de Usuarios Banda Ancha Móvil	51
Figura 4.35. Throughput medio en la hora cargada y volumen medio	52
Figura 4.36. Líneas de tendencia de throughput medio en la hora cargada y volumen medio	52
Figura 4.37. Relación throughput medio en hora cargada v/s consumo total.....	53
Figura 4.38. Aporte de los heavy users al volumen total traficado en un mes y al throughput en la HC.	54
Figura 4.39. Sólo el 25% de los usuarios cumplen ambos criterios.	54
Figura 4.40. Tráfico de dos usuarios compartiendo la misma celda.....	56
Figura 4.41. Al caer en FU el usuario 1 ve degradado su servicio limitándose su throughput máximo.....	56
Figura 4.42. El usuario 1 tiene limitado su throughput independientemente si existe o no capacidad en el sistema.	57
Figura 4.43. Al caer en FU el usuario 1 ve degradado su servicio sólo cuando no existe capacidad en la celda.....	57
Figura 4.44. El usuario 1 utiliza la capacidad remanente del sistema.	58
Figura 4.45. La política de cambio de prioridad (THP) del <i>intelligent fair use</i> logra una mejor percepción del servicio ya que su tráfico es restringido sólo en la medida que no hay capacidad disponible.	58
Figura 4.46. Efecto en el throughput total tras aplicar políticas de fair use.	59
Figura 4.47. Aplicación de <i>fair use</i> con múltiples ciclos de facturación.	59
Figura 4.48. El uso de la política de <i>intelligent fair use</i> no mostraría una variación importante del tráfico total.	60
Figura 4.49. Throughput en función del RTT de una conexión TCP	61
Figura 4.50. Throughput Real en función del RTT para los sitios más comunes.	61
Figura 4.51. Throughput en función del RTT de una conexión TCP con y sin escalamiento	62

Figura 4.52. Throughput y tiempos de descarga con y sin escalamiento.....	63
Tabla 4.53. Comparación de throughput con y sin usar RFC1323.....	63
Figura 4.54. Throughput en función del RTT para distintos sitios	64
Figura 4.55. Mejora del throughput al mejorar el retado.....	65
Figura 4.56. diagrama de flujo para el envío de datos sobre TCP	66
Figura 4.57. Secuencia de ejecución de la aplicación de medición de calidad de experiencia.....	69
Tabla 4.58. Tabla de KPI generados por la aplicación de medición de QoE.....	69
Figura 4.59. Indicador Calidad de Experiencia del servicio BAM	70
Figura 4.60. Modelo de costos.	70
Figura 4.61. Costos de Red por gigabyte de datos.	71
Figura 4.62. El usuario de la línea verde trafica más volumen de datos pero tiene el mismo costo de red que el usuario de la línea roja.	72
Figura 4.63. Distribución de costos de red.	72
Figura 4.64. Distribución de usuarios entre las celdas de un sitio.....	73
Figura 4.65. Costos de servicios de un nuevo sitio.	74
Figura 5.66. Indicador Calidad de Experiencia del servicio BAM	76
Figura A.67. Arquitectura GSM/GPRS	90
Figura A.68. Arquitectura 2G y 3G	91
Figura A.69. Arquitectura EPS y Legacy	92
Figura A.70. Tabla velocidades de LTE	93
Figura A.71. Tabla de especificaciones de las categoría de los UE de LTE	94
Figura A.72. Arquitectura WiMax.....	95
Figura B.73. Distribución de protocolos de transporte en Internet.....	97
Figura B.74. Encabezado TCP.....	98
Figura B.75. Mecanismo three way handshake.....	99
Figura B.76. Mecanismo de finalización de la conexión.....	100
Figura B.77. Evolución de Ventana de Congestión durante sesión TCP	101
Figura B.78. Contenido enviado versus throughput alcanzado	101
Figura B.79. Flujo de pérdida y recuperación de segmento usando <i>Time-Out</i>	103
Figura B.80. Flujo de pérdida y recuperación de segmento usando <i>Fast Retransmit</i> .	104
Figura B.81. Adaptación de TCP luego de pérdida de paquetes.....	105

Figura B.82. Adaptación de TCP luego de pérdida de paquetes con y sin Fast Recovery	106
Figura B.83. Encabezado UDP	106
Figura B.84. Variación en flujos de tráfico TCP v/s UDP al pasar a un canal de menor capacidad.....	107
Figura B.85. Control de Flujo TCP y UDP y pérdida de paquetes.....	108
Figura B.86. Overhead protocolos de transporte.....	108
Figura C.87. Diagrama de flujo de la descarga de un contenido web.....	110
Figura C.88. Throughput de un usuario descargando contenido web.....	111
Figura C.89. Funcionamiento de streaming tradicional.....	112
Figura C.90. Funcionamiento Streaming.....	112
Figura C.91. Funcionamiento RTMP.....	113
Figura C.92. Streaming de Video	113
Figura C.93. Constelación P2P	114
Figura C.94. Flujo de tráfico P2P	114
Figura D.95. Clasificación de Servicios de Banda Ancha.....	117
Tabla D.96. Características de Servicios de Banda Ancha	118
Tabla D.97. Clasificación de Servicios y Aplicaciones en términos de categorías y tipos de información.....	119
Figura D.98. Demanda de ancho de banda por aplicación	120
Figura D.99. Clasificación de Servicios de Banda Ancha.....	120

Abreviaturas

2G	2nd Generation
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	4rd Generation
ADSL	Asymmetric Digital Subscriber Line
ARP	Allocation and Retention Priority
ARPU	Average Revenue Per User
ASN	Access Service Network
AWS	Advanced Wireless Services
BAF	Banda Ancha Fija
BAM	Banda Ancha Móvil
BH	Busy Hour
B-ISDN	Broadband Integrated Services Digital Network
BPSK	Binary Phase-Shift Keying
BS	Base Station
BSC	Base Station Controller
BSS	Base Station Subsystems
BTS	Base Transceiver Station
C/I	Carrier to Interference Ratio
CPC	Continuous Packet Connectivity
CPE	Consumer Premise Equipment
CSD	Circuit-Switched Data
CSN	Connectivity Service Network
DB-HSPA	Dual Band High-Speed Packet Access
DC-HSPA	Dual Carrier High-Speed Packet Access
DNS	Domain Name Service
DOCSIS	Data Over Cable Service Interface Specification
DPI	Deep Packet Inspection
E2E	End to End
EDGE	Enhanced Data rates for GSM of Evolution
EMM	EPS Mobility Management
eNB	Evolved Node B
EPC	Evolved Packet Core
EPS	Evolved Packet System
ESM	EPS Session Management
E-UTRA	Evolved UMTS Radio Access
E-UTRAN	Evolved UMTS Radio Access Network
FBB	Fixed broadband
FTP	File Transfer Protocol

FU	Fair Use
GBR	Guaranteed Bit Rate
GERAN	GSM/GPRS/EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile communication
HC	Hora Cargada
HLR	Home Location Registers
HSCSD	High-Speed Circuit-Switched Data
HSDPA	High-Speed Downlink Packet Access
HSPA	High-Speed Packet Access
HSUPA	High-Speed Uplink Packet Access
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
ITU	International Telecommunication Union
KBO	Key Business Objectives
KPI	Key Performance Indicator
KPO	Key Performance Objectives
KQI	Key Quality Indicator
LTE	Long Term Evolution
MBB	Mobile Broadband
MBR	Maximun Bit Rate
MBS	Mobile Broadband Systems
MME	Mobility Management Entity
MoU	Minutes of Use
MoU	Measurement of Use
MSS	Mobile Switching Subsystem
MSS	Maximum Segment Size
NAP	Network Access Provider
NSN	Nokia Siemens Networks
OFDMA	Orthogonal Frequency-Division Multiple Access
P2P	Peer to Peer
PCRF	Policy Charging and Rules Function
PCS	Personal Communications Service
PCU	Packet Control Unit
PDN	Packet Data Network

QAM	Quadrature amplitude modulation
QoE	Quality of Experience
QoS	Quality of Service
QoSE	Quality of Service Experimented
QPSK	Quadrature Phase-Shift Keying
RAN	Radio Access Network
RFC	Request for comments
RNC	Radio Network Controller
RT	Real Time
RTMP	Real Time Messaging Protocol
RTO	Retransmission Time-Out
RTP	Real Time Protocol
RTT	Round-Trip Time
RWND	Received Window
SAE	System Architecture Evolution
SAES	System Architecture Evolution Specification
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SG	Serving Gateway
SON	Self Organizing Networks
SRTT	Smoothed Round-Trip Time
SUBTEL	Subsecretaría de Telecomunicaciones (Regulador de Chile)
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
THP	Traffic Handling Priority
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over IP
VoLTE	Voice over LTE
WCDMA	Wideband Code Division Multiple Access
WiFi	Wide Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WS	Window Scale

Capítulo 1. Introducción

El servicio de banda ancha móvil es uno de los con más alto crecimiento en telefonía móvil en el mundo. En Chile, tuvo un crecimiento de 4,4 veces el 2009 , 2,2 veces el 2010, producto de las contingencias nacionales, y se proyecta que para el 2011 éste sea de al menos 3 veces¹. Acceso inalámbrico móvil a Internet con alta disponibilidad y presencia en todas las zonas urbanas de baja y alta densidad y una amplia cobertura rural, sumado a su bajo costo han sido la clave para este crecimiento. El potencial de estar siempre conectado, en cualquier lugar y las continuas innovaciones en materia de terminales y aplicaciones integradas en forma natural a Internet ha sido crucial en la penetración del servicio lo que se ha sumado la creciente penetración del uso de notebooks, netbooks y tablets en los últimos años.

Esto ha significado que los operadores móviles han consolidado su ingreso al mundo de los ISP, pero con los exigentes usuarios y el competitivo mercado de la telefonía móvil es necesario entregar un servicio de alta calidad.

Como un elemento diferenciador, se introduce el concepto de calidad de experiencia (QoE, Quality of Experience) y surge como necesidad establecer los parámetros de medición apropiados (KQI, Key Quality Indicator) que sirvan de puntos de control para que ésta se cumpla y además se puedan incorporar tareas que mejoren la experiencia de servicio de los abonados.

Para esto es necesario conocer la tecnología utilizada por los operadores móviles, su evolución y el comportamiento de los abonados, pudiendo así modelar el servicio y luego introducir puntos de control y mejoras en el sistema que permitan entregar una calidad de experiencia satisfactoria y altamente diferenciadoras.

En este trabajo se analiza el comportamiento de los abonados, determinando el tipo de contenido que éstos prefieren, los protocolos que utilizan y cómo esto afecta las redes. Se describe la tecnología utilizada por los operadores móviles para brindar el servicio indicando los puntos críticos para el diseño y planificación de la red.

Finalmente, considerando los puntos anteriores se define un indicador de control para la calidad de experiencia y se estudian diversas tareas en la red que determinarán niveles de mejora del servicio.

¹ de acuerdo a datos de la Subtel y estimaciones de crecimiento de Cisco.

1.1 Motivación

El servicio Banda Ancha Móvil (BAM) es complejo, los operadores móviles mundiales lo ofrecen mostrando el valor agregado que les entrega la movilidad, pero en Chile ha sido distinto y los usuarios no son capaces de distinguir si lo que están usando es un servicio de banda ancha móvil o sólo una extensión inalámbrica de la banda ancha tradicional. En pocas palabras, los operadores venden banda ancha móvil y los usuarios creen comprar ADSL, debido a que Internet es Internet y los abonados esperan la mejor calidad que le puedan dar los Proveedores de Servicio.

Esto significa que los usuarios no están conscientes del servicio que están comprando y luego le exigen un rendimiento similar al de Internet de banda ancha fijo, considerando además escenarios de movilidad.

Esta diferencia genera una notable desviación y una mala percepción del servicio móvil y aunque la naturaleza de la tecnología disponible tiene muchas limitaciones, si es posible modificar y mejorar la percepción del servicio introduciendo algunos cambios en la red.

1.2 Aportes

La Banda Ancha Móvil es un servicio relativamente reciente. Con su bajo nivel de penetración, incluso para mercados maduros en telefonía móvil, y su constante evolución, se presenta como un gran desafío para los operadores de red que aún no comienzan su incursión en la calidad de experiencia, elemento que los diferenciará en los siguientes años.

Los principales aportes de este trabajo pueden ser resumidos en los siguientes puntos:

1. Caracterización de la situación actual del servicio y escenario local en Chile.
2. Caracterización de los diferentes servicios que utilizan los usuarios de BAM.
3. Caracterización y comportamiento de los usuarios del servicio BAM.
4. Modelamiento de costos de red para el servicio BAM.
5. Desarrollo de una metodología para la definición de un indicador de calidad de experiencia.

1.3 Objetivos

1.3.1 Objetivo General

El objetivo general de esta tesis de grado es determinar los factores que influyen en la calidad de experiencia percibida por el usuario del servicio de banda ancha móvil.

1.3.2 Objetivos Específicos

Los objetivos específicos de este trabajo son los siguientes:

1. Caracterizar la tecnología que soporta el servicio de banda ancha móvil actual y su evolución.
2. Caracterizar los servicios que utiliza el usuario y los protocolos que le dan soporte.
3. Generar indicadores de calidad de experiencia que permitan reflejar el comportamiento de los usuarios.
4. Simular el comportamiento de los usuarios y servicios, y medir su experiencia determinando así el nivel de servicio ofrecido por el operador.
5. Introducir mejoras en el sistema móvil, que permitan potenciar la percepción del servicio que tiene el abonado móvil.

1.4 Descripción

En el siguiente capítulo se presentan los antecedentes que permiten conocer cómo ha evolucionado la tecnología móvil y cuáles son las principales características de ésta. También se puede encontrar la descripción de los servicios que utilizan los abonados y las principales variables que definen desempeño, eficiencia y calidad de experiencia.

En el capítulo tercero, se realiza una descripción de la metodología utilizada para el estudio de los temas que involucra este trabajo.

A continuación se describen los principales resultados obtenidos en el modelamiento del servicio y de las principales aplicaciones que los abonados usan, luego el modelo de calidad de experiencia utilizado y como afectan los costos al servicio BAM.

Finalmente en los capítulos 5 y 6 se puede encontrar una discusión de los principales tópicos estudiados y las principales conclusiones que se alcanzan en este trabajo.

Capítulo 2. Antecedentes

En este capítulo se presenta una revisión de los antecedentes necesarios para el desarrollo de este trabajo. En él se analiza la evolución que ha tenido Internet, se hace una revisión de las tecnologías más importantes utilizadas para brindar el servicio de banda ancha móvil y además se incluye una breve visión del estado actual del servicio en Chile hasta finales del año 2010.

A continuación se muestra una descripción y análisis de los principales protocolos y aplicaciones que influyen en el servicio y finalmente una descripción de los factores que influyen tanto en los costos del servicio como en el modelamiento de calidad de experiencia.

2.1 Evolución Internet

Internet ha tenido un constante desarrollo y evolución en los últimos años, incluso rompiendo varias veces las predicciones de los expertos y desviando su camino hasta tener lo que hoy se conoce.

Desde sus orígenes, hace 40 años, se ha visto como ha cambiado desde ser una pequeña red universitaria de baja velocidad a una red mundial de alta velocidad y rica en contenidos.

Se ha podido apreciar cómo ha evolucionado el tipo de contenido ofrecido en la Red y cómo ésta ha debido adaptarse para dar soporte a las demandantes aplicaciones y las altas exigencias de los usuarios.

2.2 Descripción de la Tecnología y de los Servicio de Datos

Tras el fuerte incremento en el uso de Internet de los últimos años, se ha visto cómo las distintas formas de ofrecer el servicio han ido mejorando. Primero fueron las distintas evoluciones de las conexiones conmutadas, luego la Internet ilimitada que llegó primero con llamadas ilimitadas al servidor de acceso conmutado y luego las evoluciones hacia ADSL y DOCSIS que permite ofrecer altas velocidades utilizando el cableado ya desplegado para otros servicios.

A comienzos de 2000 son lanzadas las primeras redes GPRS en el mundo, naciendo así el servicio de banda ancha móvil para más tarde evolucionar a EDGE, luego, la incorporación de redes 3G introducen las tecnologías WCDMA y HSPA y en paralelo WiMAX, para finalmente aparecer LTE como se ha podido apreciar en los últimos lanzamientos de redes móviles.

En los siguientes puntos se analiza la evolución de la tecnología de banda ancha móvil deteniéndose en las principales tecnologías de acceso utilizadas hoy día, pieza clave para ofrecer un servicio de calidad.

2.2.1 Evolución de Tecnologías de Acceso

Las redes móviles han tenido una fuerte evolución y un gran incremento en los anchos de bandas disponibles para el servicio. Es así como se puede ver en la Figura 2.1 como se va incrementando la velocidad de acceso y se logra tener un servicio de banda ancha que es capaz de competir con los servicios fijos.

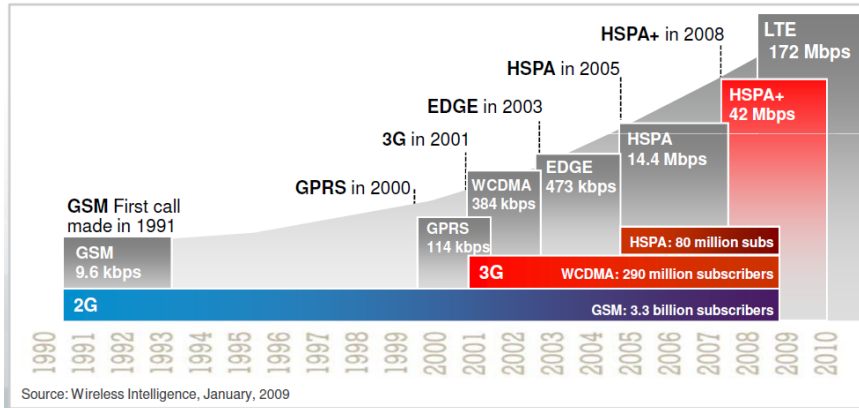


Figura 2.1. Evolución Redes Móviles

Para incrementar las velocidades en el acceso ha sido necesario realizar importantes cambios y mejoras en las tecnologías. Se han incorporado técnicas de modulación digital y mejoras en las tecnologías de acceso que permiten que el servicio ofrezca tasas de transmisión más altas y mejores niveles de retardo.

Modulación Digital

En términos de modulación digital, los principales tipos utilizados por las tecnologías UMTS y LTE son QPSK, 16QAM y 64QAM. Estas modulaciones ofrecen una constelación más densa a medida que se mejora la codificación, tal como se muestra en la figura 2.2.:

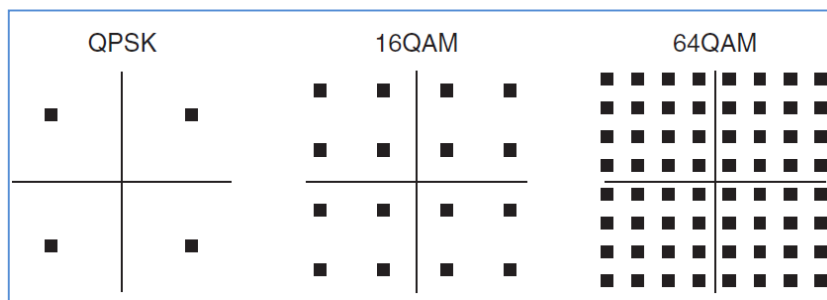


Figura 2.2. Constelación de señales para QPSK, 16QAM y 64QAM

Los distintos tipos de modulación utilizados actualmente han permitido mejorar las velocidades de acceso sin tener que realizar cambios importantes en las redes y soportando los usuarios existentes. Las modulaciones más utilizadas por las redes modernas son las siguientes:

- Quadrature Phase-Shift Keying (QPSK), como su nombre lo indica utiliza modulación por desplazamiento de fase de 4 símbolos desplazados entre sí en 90°. QPSK permite el envío de dos bit de información por cada símbolo.
- Quadrature amplitude modulation (QAM), consiste en la modulación de amplitud en cuadratura y puede tener diferentes números de estados (nQAM). 16QAM permite 4 bits y 64QAM por su parte 6 bits. Esto significa que una eficiencia espectral (bits/s/Hz) de dos y tres veces mejor que QPSK.

El uso de modulaciones de más alto orden, permite alcanzar mayores velocidades para un ancho de banda fijo. Sin embargo, ello requiere canales de mayor calidad, inmunes al ruido y la interferencia. Las modulaciones de orden superior como 16QAM o 64QAM requieren una mayor relación señal a ruido E_b/N_0 en el receptor para una probabilidad de error de bits dada, en comparación con QPSK.

Métodos de Acceso

El método de acceso utilizado permite obtener distintas tasas de transmisión y dependiendo de las características del espectro utilizado, lograr mejores eficiencias. En lo que se refiere a los métodos de acceso en tecnologías inalámbricas, la evolución de estos sólo es posible tras cambios importantes en la tecnología que utiliza el operador y terminales utilizados por los abonados. Es por ello que el cambio es lento y requiere inversiones considerables. Los principales métodos de acceso utilizados en la actualidad son los siguientes:

- Wideband Code Division Multiple Access (WCDMA), consiste en la ortogonalización del espectro radioeléctrico mediante el uso de códigos que además utilizan como identificador único de cada usuario y permite que ellos tengan acceso a todo el recurso de frecuencia y en todo instante [24] . Ello permite lograr enviar una mayor cantidad de información. Ésta técnica es utilizada en UMTS.

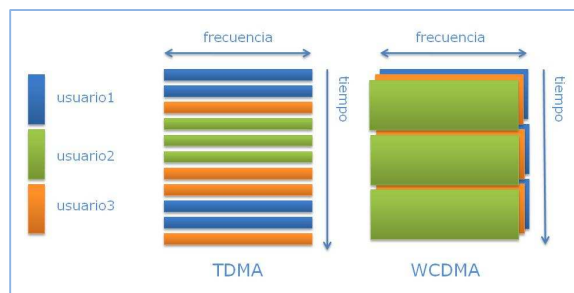


Figura 2.3. Modulación TDMA y WCDMA.

- Orthogonal Frequency-Division Multiple Access (OFDMA), consiste en una técnica de acceso en la cual se divide el ancho de banda disponible en

canales los cuales son utilizados por los usuarios en forma separada. El esquema considera la división del canal en un conjunto de subportadoras que se reparten en el conjunto de usuarios para el uso eficiente del espectro donde estos pueden usar un subcarrier o un conjunto de éstos y además se puede codificar utilizando diferentes métodos dependiendo de la calidad del canal [24] . OFDMA utiliza elementos de Time Division Multiple Access (TDMA) para que cada subcarrier pueda ser asignado dinámicamente entre los diferentes usuarios del canal a lo largo del tiempo, en función de la demanda de capacidad que requieran los usuarios, tal como se muestra en la figura 2.4.

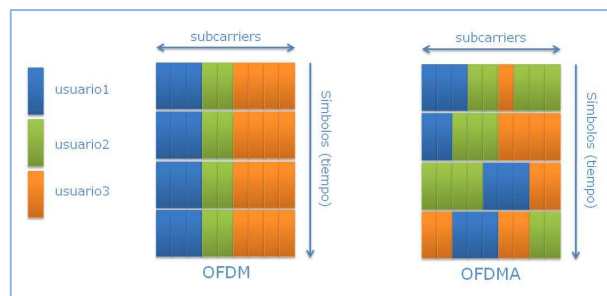


Figura 2.4. Modulación OFDM y FDMA.

Cada grupo de subcarriers es transmitido en paralelo y a continuación se deja un intervalo de tiempo para evitar el solapamiento de la información.

- Single-Carrier Frequency Division Multiple Access (SC-FDMA), consiste en método de acceso basado en la ecualización del dominio de la frecuencia. Al igual que en el método anterior, el canal es dividido en subportadoras y su asignación se realiza en forma secuencial, es decir, se asigna todo el canal a uno de los usuarios el que dispondrá de todos los subcanales para el envío de la información, donde cada subcanal puede ser codificado en forma distinta dependiendo del nivel de interferencia que pueda tener cada canal.

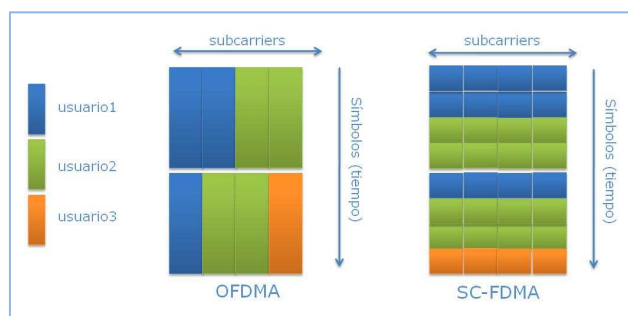


Figura 2.5. Modulación OFDMA y SC-FDMA.

El uso conjunto de los métodos de acceso y codificación ha permitido que las principales tecnologías utilizadas actualmente, como son WCDMA y LTE, evolucionen y aumenten sus velocidades, como se muestra en la figura 2.6.

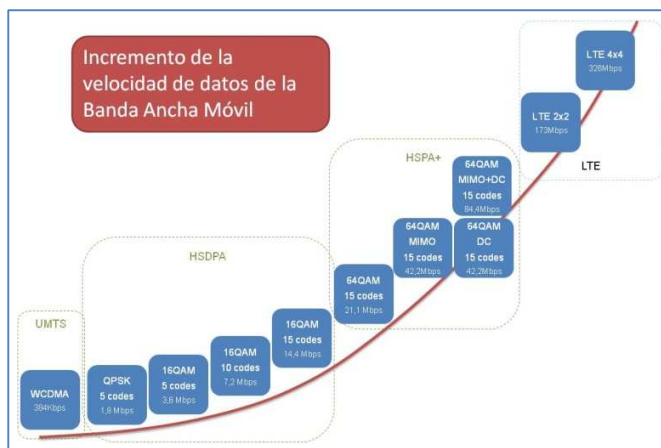


Figura 2.6. Incremento de las velocidades móviles

Los incrementos continuos de velocidad de acceso en las redes móviles generan además una disminución en los tiempos de descarga. Como se puede ver en la figura 2.7, para una transmisión usando HSDPA con codificación 16QAM y 10 códigos se logran tiempos de 16 segundos para la descarga de un archivo de música y en el caso de LTE se puede obtenerlo en 1 segundo, o sea, se puede observar una disminución en el tiempo de 16 veces.

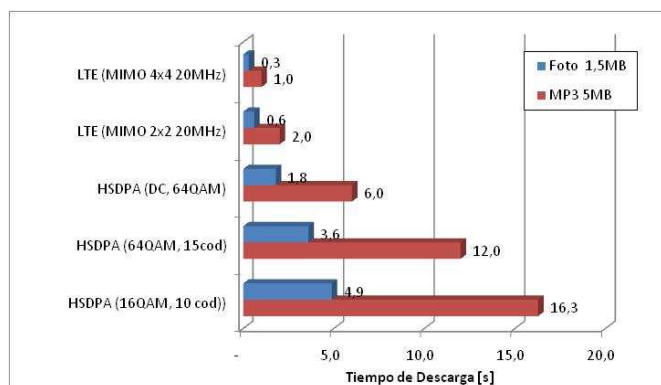


Figura 2.7. Tiempo de descarga de archivos según tecnología de acceso

Esta evolución no solo se restringe al ámbito del aumento de velocidad, sino también a mejoras en *delay*, permitiendo la incorporación de nuevos servicios.

En el Apéndice A se realiza una descripción de la arquitectura de las diferentes tecnologías de acceso utilizadas hoy en día para brindar el servicio BAM como son GSM/GPRS/EDGE, WCDMA, LTE, LTE advance y WiMax. Cabe destacar que las

cuatro primeras tecnología convergen en un núcleo común que permite la interoperación entre ellas, definiendo así la evolución de las redes móviles hacia una cuarta generación liderada por LTE que permite la continuidad del servicio sobre las redes *legacy* más maduras.

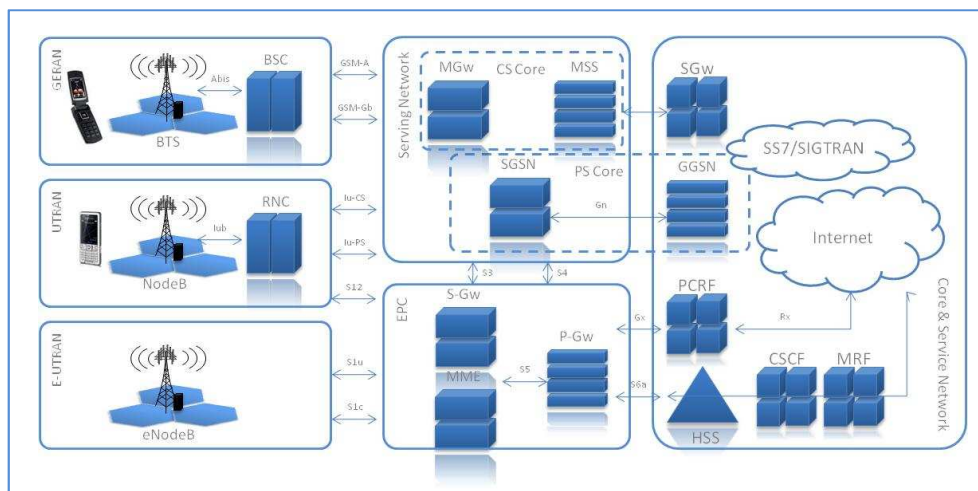


Figura 2.8. Arquitectura EPS y Legacy

2.2.2 Packet Scheduling

Una característica que comparten todas las tecnologías revisadas en la sección 2.2.1 es que todas utilizan un medio compartido para transmitir datos. Esto implica que es necesario desarrollar diferentes algoritmos que permitan administrar el ancho de banda existente.

Cuando existe más de un usuario en una celda esperando por recursos para usar el medio de transmisión es necesario implementar un protocolo que permita administrar el recurso compartido eficientemente entre todos los usuarios.

En este punto se hablará de tres tipos de esquemas utilizados para administrar el recurso de radio [4] :

C/I based. Esta estrategia de asignación de recursos basada en políticas de C/I (Carrier to Interference Ratio) favorece a usuarios con mejores condiciones de radio, maximizando la capacidad del sistema, pero tiene la desventaja de proporcionar una falta de equidad debido a que los usuarios con mala calidad de radio sólo serán atendidos una vez que aquellos con mejores calidades de radio finalicen el uso del recurso. Es importante tener en cuenta que los usuarios con una pobre calidad de radio, como aquellos en el borde de una celda, recibirán una calidad excesivamente deficiente en comparación con el resto de los usuarios y, en el caso de encontrarse en una celda de alta carga, podrían quedarse sin servicio ya que nunca tendrán la posibilidad de que se les asignen recursos.

Fair Resources. Esta estrategia asigna equitativamente los recursos a los usuarios de una celda. Si bien este mecanismo intenta distribuir homogéneamente los recursos, las condiciones de radio en la que se encuentran los abonados genera un menor throughput ofrecido a los que se encuentran en el borde de la celda, comparados con los que se encuentran más cerca de ésta. Esto se debe a que, aunque a los usuarios se les reparte equitativamente los recursos, se tiene que la codificación que pueden optar los usuarios producto de su condición de calidad de radio generará una diferencia de rendimiento y por lo tanto una distribución no homogénea en términos de throughput percibido.

Fair Throughput. El objetivo de este tipo de asignación es ofrecer a los usuarios el mismo throughput independientemente de las condiciones de radio en que éstos se encuentren, es decir, un usuario con condiciones de radio desfavorables, como los del borde de la celda, tendrán el mismo throughput que los con buenas condiciones de radio. Este método da una distribución más justa a los usuarios pero con el costo de un menor rendimiento de la celda.

Los diferentes esquemas de asignación de recursos de red, permiten ofrecer a los usuarios diferentes experiencias de uso a diferentes costos de red. Una adecuada elección permitirá ofrecer a los abonados la calidad deseada a un costo de red definido por la eficiencia que tendrá la celda sobre el rendimiento total del sistema.

A los diferentes esquemas mencionados se debe sumar los parámetros de calidad de servicio que permite la tecnología asignar al flujo de los usuarios. En el caso de HSPA se puede distinguir los siguientes:

- a) *Allocation and Retention Priority (ARP)*. Este parámetro determina la prioridad relativa que tiene un flujo de datos respecto a los flujos de otros usuarios pudiendo asignar dichos flujos a colas de prioridad distintas que generará que el flujo de algunos usuarios sea priorizado por sobre otros, generando diferentes experiencias de throughput entre los usuarios.
- b) *Traffic Handling Priority (THP)*. Opera en forma similar al ARP pero sólo para tráficos pertenecientes a la clase interactiva, de acuerdo a la QoS asignada para el punto de acceso utilizado.
- c) *Maximum Bit Rate (MBR)*. Este parámetro determina la tasa de conexión máxima que tendrán los abonados. Este parámetro permite que se ajusten los recursos asignados limitando el uso excesivo a la tasa máxima MBR cuando el usuario demande altos niveles de tráfico, dejando recursos disponibles para otros abonados.
- d) *Guaranteed Bit Rate (GBR)*. Este parámetro determina la tasa de conexión asegurada que podrá tener un abonado cuando se encuentre transmitiendo datos. El uso de este parámetro permite que tráficos sensibles a variaciones de ancho de banda puedan ofrecerse sin degradar su experiencia.

2.3 Situación Local

En Chile, al año 2010, existían tres operadores móviles ofreciendo servicios en las bandas 850MHz y 1900MHz (Claro, Entel PCS y Movistar) todos con redes GSM y HSPA. Tras una reciente licitación de nuevas banda, VTR y Nextel se adjudican bandas de tercera generación en las frecuencias AWS (Advanced Wireless Services) que opera en los 1700/2100 MHz y se espera entren en operación comercial en el segundo semestre de 2011.

Las asignaciones de bandas que se tienen hoy día en Chile se muestra en la figura 2.9:

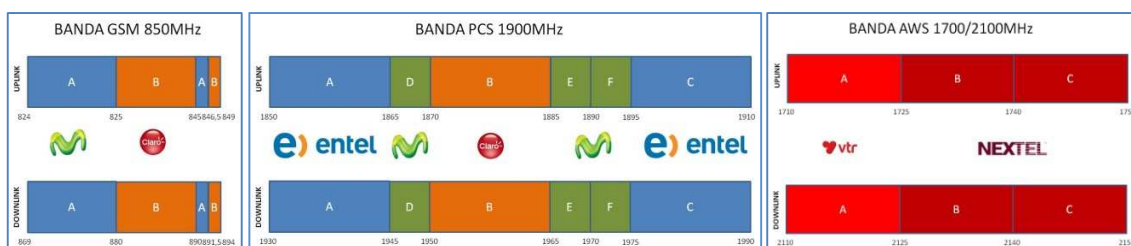


Figura 2.9. Distribución banda 850 y 1900 MHz utilizada por actuales operadores móviles y banda 1700/2100 MHz de nuevos operadores

En el caso de los operadores antiguos, se puede ver que los que tienen asignada banda baja (850 MHz) han preferido ésta para el despliegue de su red de tercera generación debido a que por su característica logra mejor cobertura indoor. Y en el caso del tercer operador ha realizado su despliegue usando sus bandas PCS.

El servicio de banda ancha móvil ha tenido un fuerte aumento en Chile a diferencia del servicio de banda ancha fijo el que se ha visto frenado en el último tiempo. Este comportamiento muestra que aún existen grandes oportunidades en el mercado y una gran tarea para que este crecimiento se mantenga y se consolide.

De acuerdo a los datos de la Subtel² se puede ver cómo ha crecido el mercado chileno de la banda ancha móvil y fija.

² Estadísticas servicio de acceso a Internet disponibles en www.subtel.cl



Figura 2.10. Crecimiento banda ancha móvil (BAM) y Fija (BAF) en Chile

Durante el 2010 se aprecia un crecimiento de la BAF llega a un 8,6% y en cambio BAM un 103,2% lo que muestra el rápido crecimiento de banda ancha móvil y el estancamiento de la BAF. En lo que respecta a la penetración del servicio BAM se puede ver que a fines del 2009 ésta era menor al 4% y que para fines del 2010 ésta alcanza el valor de 7,3%. Inclusive se espera que durante la primera mitad del 2011 la cantidad de conexiones BAM supere a las de BAF³.

Por otro lado, se ve que el market share se mueve continuamente en los últimos 18 meses disputándose el primer lugar entre los dos principales operadores de la industria como se muestra en la figura 2.11, y dejando Entel como líder absoluto al finalizar el segundo semestre del 2010 con casi un 50% del mercado.

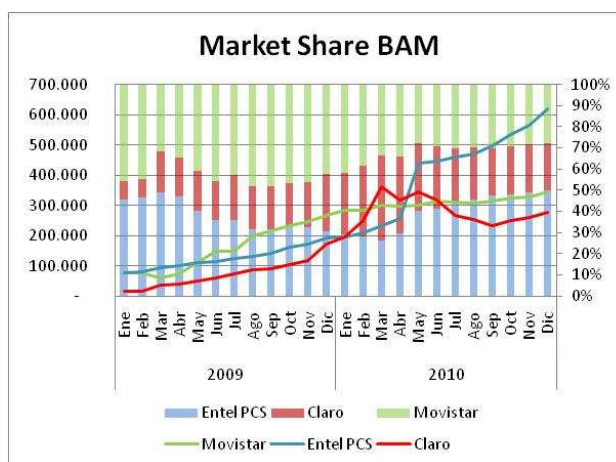


Figura 2.11. Market share BAM en Chile

En lo que respecta a la tecnología de acceso, se puede apreciar que la situación local es bastante similar a lo que ocurre en el resto del mundo. En Chile, se tiene que el

³ Estadísticas actualizadas se encuentran en www.subtel.cl

operador Claro poseen tecnología HSPA pudiendo ofrecer hasta 14,4 Mbps teóricos. En cambio, Entel y luego Movistar ofrecen tecnología HSPA+⁴, que permite lograr hasta 21 Mbps teóricos. Durante el primer trimestre de 2011, Entel lanza la tecnología DC-HSPA+ que permite alcanzar 42 Mbps teóricos, posicionándose como el primer operador en Latinoamérica en tener dicha tecnología.

Se puede agregar que Entel presentó a finales de 2009 pruebas de LTE siendo el cuarto operador a nivel mundial en probar esta tecnología y primeros en Latinoamérica[22] , en la misma línea los operadores Movistar y Claro realizaron pruebas de la tecnología durante 2010. Este hito muestra el alto interés de los operadores y fabricantes por empujar la nueva tecnología, la que podrá materializarse una vez que el regulador (Subtel) licite las bandas de frecuencias apropiadas para su implementación, lo que se prevé hará durante el mes de marzo de 2011 y específicamente licitando la banda 2.600 MHz, la misma que hoy está implementada comercialmente en Europa con sus primeras operaciones en Suecia y Noruega por el operador TeliaSonera. Esto ayudará a una rápida adopción de la tecnología aprovechando las economías de escalas impulsadas por el viejo mundo.

Por último, la empresa china ZTE y la Universidad de Chile inauguran el pasado 31 de marzo 2011 el primer laboratorio docente de LTE en Latinoamérica, dejando a Chile como líder indiscutido de la región en esta tecnología.

2.4 Desempeño Protocolos de Transporte

Es de gran importancia para el análisis conocer cómo operan realmente los protocolos de transporte que son usados en Internet, ya que su operación afecta directamente en el rendimiento de las aplicaciones. Hoy en día casi el 90% del volumen de tráfico en Internet está basado en el uso de TCP [1] y aunque el tráfico UDP sea minoritario es importante destacar que la cantidad de flujos son del orden del 70% por lo que no debe ser descuidado.

2.4.1 TCP

El protocolo TCP (*Transmission Control Protocol*) es uno de los principales protocolos del stack TCP/IP y se encuentra definido en RFC 793 de la IETF [5] .

Es un protocolo orientado a la conexión diseñado para brindar una comunicación bidireccional, confiable de extremo a extremo, sobre una red no confiable. Para ello, implementa una serie de mecanismos como el control de flujo, detección y corrección de errores y control de congestión que le dan un alto nivel de confiabilidad al protocolo, asegurando a las aplicaciones la entrega íntegra de la información transmitida.

En términos generales se puede decir que una conexión TCP posee 3 fases. La primera conocida como sincronización, donde se inicia el primer contacto entre las

⁴ Entel es el primer operador en Latinoamérica en lanzar HSPA+ en diciembre 2009. [22]

partes e intercambian parámetros que permiten asegurar el correcto funcionamiento de la conexión. Además, se informan las características que ésta tendrá.

La segunda, donde la conexión ya se encuentra establecida y se realiza el intercambio de la información. Junto con la información de las aplicaciones se realiza el envío de mensajes de control que permiten que ésta se realice de manera fluida. En particular, se realiza el control de flujo de conexión y para ello se utiliza el mecanismo de *slow start* y *congestion avoidance*, los que son descritos en detalle en Apéndice A.

Finalmente la tercera fase donde se negocia y realiza el cierre de la conexión.

2.4.2 Control de Flujo y Throughput de TCP.

TCP posee un mecanismo de control de flujo para regular la velocidad de transferencia de datos en función de las variaciones de capacidad y congestión del medio de transmisión, adaptándose automáticamente a las condiciones. Este mecanismo está basado en un lento inicio e incrementa la velocidad paulatinamente hasta lograr la máxima velocidad que le permite el medio. Esto significa que para alcanzar un throughput de 1Mbps es necesario transmitir al menos 400kbytes⁵, lo que significa que para transmitir datos de menor tamaño no será posible alcanzar velocidades altas usando una sólo conexión simultánea.

Cuando un segmento TCP se pierde se activa un algoritmo de recuperación de información y además se produce un fuerte descenso en la tasa de transmisión y una recuperación más lenta, que con el uso de *Fast Recovery* [9] se logran mejorar, tal como se muestra en la figura 2.12.

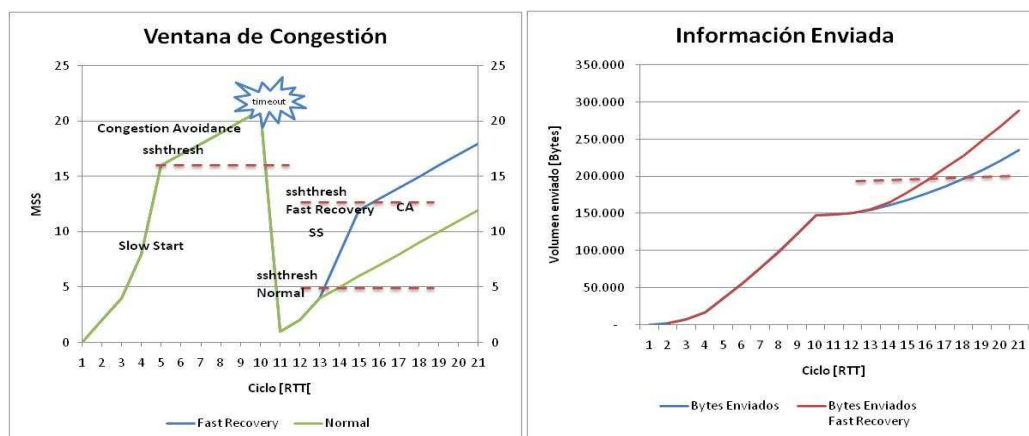


Figura 2.12. Adaptación de TCP luego de pérdida de paquetes con y sin Fast Recovery

La mejora de *Fast Recovery* permite alcanzar con mayor rapidez el throughput máximo de la conexión y lograr que la información llegue a destino varios ciclos antes.

⁵ Este valor se obtiene al simular una transmisión sobre un canal sin pérdida de paquetes y operación por defecto de slow start y congestion avoidance, más detalles se pueden apreciar en sección B.3

2.4.3 UDP

El protocolo UDP (*User Datagram Protocol*) es un protocolo de transporte no orientado a la conexión parte del stack TCP/IP definido en el RFC 768 [5] . Permite el envío de datagramas a través de la red sin que exista una conexión previa. No tiene confirmación de envío ni control de flujo. Tiene un bajo overhead haciéndolo ideal para el transporte de audio y video en tiempo real.

Debido a que UDP es un protocolo no orientado a la conexión, éste no requiere de mecanismos de inicio, fin y mantención de la conexión como en el caso de TCP, haciendo más simple el protocolo de comunicación y minimizando el overhead. Por otro lado, no posee un mecanismo de control de flujo lo que puede generar un gran problema cuando las aplicaciones son altamente demandantes de ancho de banda, ya que pueden saturar el canal de comunicación.

2.4.4 Comparación de rendimiento entre los protocolos TCP y UDP

Una diferencia importante entre TCP y UDP es que el primero tiene mecanismos de control de flujo y congestión que permite una adaptación extremo a extremo a las variaciones de las capacidades del camino adaptándose al canal más estrecho. En cambio UDP puede inundar el medio hasta encontrarse con un segmento más estrecho que descartará los datagramas, tal como se ilustra en la figura 2.13.

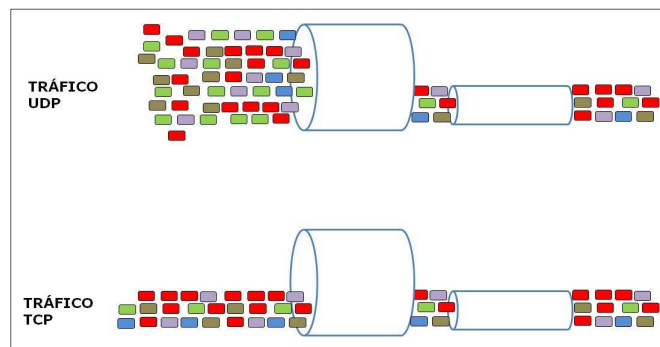


Figura 2.13. Variación en flujos de tráfico TCP v/s UDP al pasar a un canal de menor capacidad.

Otro punto importante que se debe considerar es el overhead que aportan los protocolos de transporte a las aplicaciones ya que éste incide en el throughput efectivo que aprecia el cliente.

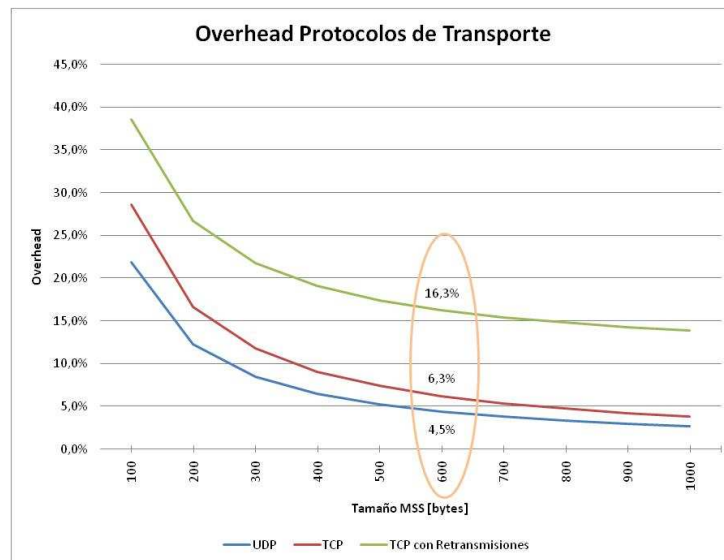


Figura 2.14. Overhead protocolos de transporte

Como muestra la figura 2.14, para una transmisión de paquetes promedio de 600 bytes, UDP aporta un 4,5% de overhead y TCP un 6,3% y pudiendo llegar a más de un 16,3% si se considera una tasa normal de 10% de retransmisiones.

2.5 Descripción Servicios de Internet

Lo que busca un usuario de banda ancha móvil es usar los servicios disponibles en Internet y, es por ello, que es importante conocer sus características técnicas principales, de modo tal que se pueda comprender su funcionamiento y los factores que pueden alterar su rendimiento y cómo se puede mejorar su percepción. La descripción de los principales servicios utilizados por los usuarios de Internet se puede apreciar en el Apéndice C.

Dentro de los principales servicios se puede encontrar:

Web Browsing. Es uno de los más masivos en Internet, corresponde a lo que el usuario conoce como navegación propiamente tal.

Streaming y Flash Video. Servicio de alto uso permite la reproducción en línea de contenido multimedia.

P2P. Servicio de compartición de archivos que utiliza como medio de transporte los protocolos UDP y TCP, generando múltiples conexiones entre todos los nodos que comparten la información.

Network Storage. Servicio de compartición de archivos basado en web que permite el almacenamiento y distribución de contenido en la red.

Redes Sociales. Corresponde a un conjunto de servicios que permite la interacción y autogeneración de contenido por parte de los usuarios que comparten la red.

2.6 Descripción de Servicio BAM

El servicio de banda ancha móvil no es un servicio de banda ancha tradicional, no sólo porque es esencialmente móvil sino también porque los requerimientos de servicios de los abonados no necesariamente son iguales que los del servicio de banda ancha tradicional. Es por eso que es importante estudiar en detalle el comportamiento de estos usuarios agrupándolos por sus características comunes, de modo tal, que ayude a comprender su funcionamiento, las necesidades de los abonados y se pueda definir indicadores de calidad de experiencia.

Los primeros usuarios en utilizar este servicio son definidos como usuarios “nativos digitales” refiriéndose a las personas que “nacieron conectadas” y que “necesitan” estar 100% conectados 7x24 365 días del año. Un segundo grupo son los “inmigrantes digitales”, usuarios que han tenido que adaptarse a la tecnología y entienden que estar “on line” es cada vez más necesario. Luego se puede ver que se está frente a usuarios más especializados que los de la banda ancha tradicional o que simplemente han evolucionado de ésta con el objetivo de buscar nuevos servicios.

También existe un grupo de usuarios que no diferencian entre el servicio inalámbrico brindado por un hot-spot WiFi y el de un servicio móvil de tercera o cuarta generación, ya que en ambos casos el usuario usará su laptop o Smartphone para conectarse en forma inalámbrica a Internet.

Los usuarios “always on” usan servicios de Internet que no necesariamente son los mismos que los de un usuario promedio, tanto en cantidad, frecuencia e intensidad por lo que es imprescindible determinar cuál es su perfil de navegación en términos de movilidad, tipo de servicios, intensidad y frecuencia, lo que definirá qué es lo que se debe monitorear y cómo se debe crear el vector de calidad de experiencia asociado al servicio.

2.6.1 Movilidad

Existe una disminución en la movilidad de los usuarios a medida que la tecnología de banda ancha móvil ofrece mayores velocidades [3] , tal como se muestra en la Figura 2.15

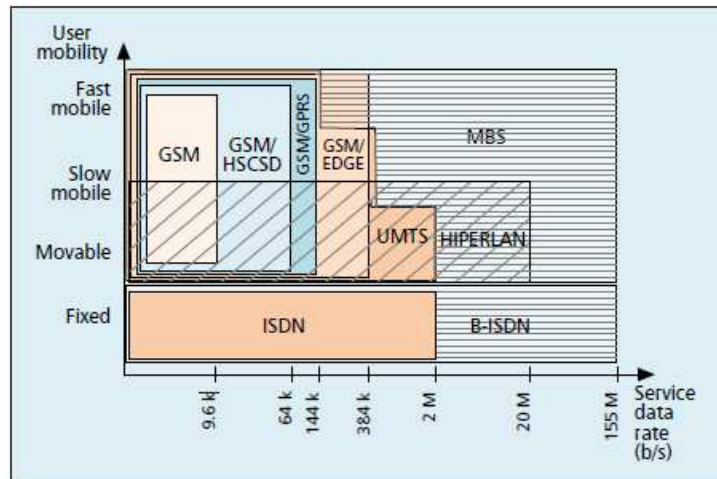


Figura 2.15. Movilidad en tecnologías móviles

Esta disminución de movilidad viene dada, principalmente, porque los servicios que demandan altos niveles de tráfico requieren grandes niveles de procesamiento que habitualmente sólo computadores personales fijos o laptops pueden realizar, obligando al usuario a tener una movilidad reducida vinculándolo sólo a un perfil nómada. Por otro lado, los servicios altamente móviles son soportados en terminales tradicionales con procesadores de capacidades reducidas y pantallas de menor tamaño y corresponden a servicios de Correos Push o navegación en el Móvil con micro-browsers. La introducción de tablets y Smartphones con procesadores más potentes y pantallas más amplias permitirían que muchas de las funcionalidades asociadas a computadores personales estén siendo absorbidas por ellos, exigiendo mayores velocidades a servicios que normalmente eran considerados exclusivamente fijos. De este modo, es posible ver que aplicaciones como descargas multimedia (youtube, grooveshark, podcast, etc), compartición de archivos (rapidshare, megaupload, dropbox, application stores, etc) son realizadas en estos dispositivos.

2.6.2 Velocidad de Datos

Cuando se requiere determinar cuál es el throughput total de la red en la hora cargada, se puede suponer que corresponde a la suma de las velocidades de transmisión que tienen los usuarios simultáneos que están traficando en dicha hora.

$$Throughput = \sum Velocidad_i Pr(U_i) \quad (1)$$

Luego, sólo sería necesario determinar cuál es efectivamente la velocidad instantánea de los usuarios y cuál es la probabilidad que éstos se encuentren traficando en un instante dado.

Este cálculo no es trivial por lo que se debe recurrir a una simplificación del problema utilizando una velocidad de diseño equivalente que representa el aporte que

realiza todo el parque del servicio a éste throughput, tal como se muestra en la ecuación (2).

$$\text{Throughput} = \text{VelocidadDiseño} * \text{NumeroUsuarios} \quad (2)$$

Ésta velocidad de diseño es directamente proporcional al volumen mensual que trafica un usuario promedio y al uso del servicio en términos de conexiones diarias y días al mes que utiliza el servicio.

$$\text{Throughput} = \text{VolDatosMes} * \text{HorasDía} * \text{DiasMes} * \text{NumeroUsuarios} \quad (3)$$

Si bien esta forma de cálculo es bastante simple, se puede considerar como una buena aproximación para el cálculo de capacidades de una red móvil y fija. Pero si se quiere mejorar la exactitud es necesario considerar que efectivamente no todos los usuarios tienen el mismo comportamiento, como se ve en el siguiente punto, y se puede utilizar valores indexados al tipo de plan comercial que tienen estos usuarios, obteniendo:

$$\text{Throughput} = \sum_{i=\text{planes}} \text{VolDatosMes}_i * \text{HorasDía}_i * \text{DiasMes}_i * \text{NumeroUsuarios}_i \quad (4)$$

2.6.3 Caracterización de Servicios

De acuerdo a la recomendación I.221 de la ITU-T [15] los servicios y aplicaciones pueden clasificarse de acuerdo a las características del servicio, parámetros de tráfico y comunicaciones y ambientes de operación, como se puede apreciar en detalle en Apéndice C.

Esta caracterización permite asociar parámetros de calidad a las diferentes aplicaciones y servicios que puede ser medidos y monitoreados en distintos puntos de la red, determinando así el efecto que estos tienen en los servicios de los abonados.

Además, se pueden utilizar indicadores específicos para cada tipo de aplicación, como por ejemplo, cuando se habla de servicios dependientes del tiempo o *Real Time*(RT) o interactivos, donde los efectos del retardo y jitter son relevantes. Luego si se hace una caracterización utilizando ancho de banda que necesitan, dependencia al retardo y el jitter se pueden obtener interesantes resultados que permiten comprender más el servicio.

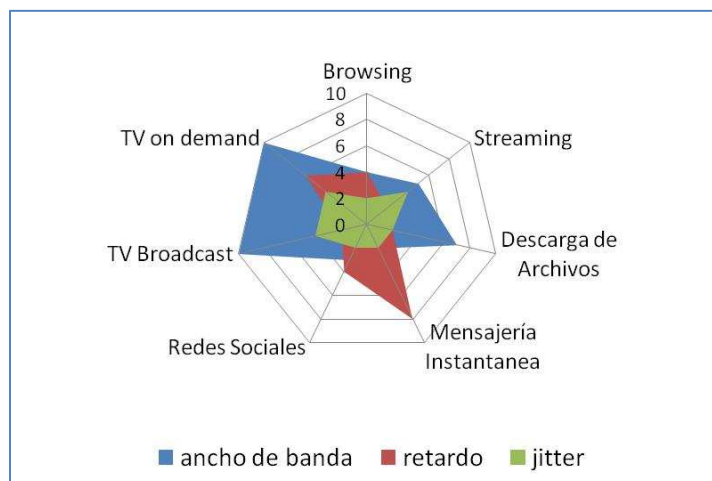


Figura 2.16. Clasificación de Servicios de Banda Ancha

Una categorización adecuada permitirá al operador optimizar su red atendiendo los factores que afectan los servicios que realmente utilizan los abonados, enfocando sus esfuerzos en la mejora sostenida de la calidad de experiencia del abonado.

2.6.4 Servicios Intensivos

Uno de los grandes problemas para los ISP son los protocolos P2P debido que a su característica de operación tiende a utilizar en forma agresiva todo el ancho de banda disponible. En un modelo de servicio fijo esto se traduce en que los protocolos agresivos utilizan todo el ancho de banda de la línea de acceso, incluso al punto de canibalizar los servicios interactivos del mismo usuario haciendo que su experiencia de navegación decaiga.

En el caso de los servicios móviles el efecto es mucho mayor. Dado que el acceso móvil es compartido, el tráfico interactivo no sólo afectará al mismo usuario que lo utiliza si no también a los usuarios que comparten el servicio en la misma celda o incluso a los usuarios de la misma zona de cobertura del RNC.

Claramente el uso del servicio depende del modelo comercial del servicio. Si un usuario tiene un plan con un precio por volumen, se preocupará de usar sólo lo que realmente necesita y, por otro lado, cuando se encuentra con planes ilimitados se ve que el usuario usa todo lo que necesita o más aún todo lo que puede. En este último caso, se tienen usuarios intensivos que distorsionan las estadísticas, sobre usan los recursos y que pueden incluso afectar a los demás clientes demandando un ancho de banda excesivo que sobrepasa los parámetros técnicos de diseño y de factibilidad comercial.

Si este número de usuarios es excesivamente alto se puede tener un servicio que no es rentable y que además no entrega un servicio de calidad. Luego, es necesario determinar las características de estos usuarios en términos de tipos de servicios que utiliza y de movilidad para buscar una estrategia que permita que éstos no afecten al resto de los usuarios.

2.7 Variables de Desempeño de una Red Móvil

A continuación se analiza una serie de variables de desempeño que permiten describir el estado de la red y de los diferentes servicios que utilizan los abonados. Estas variables, utilizadas adecuadamente y tomando las consideraciones pertinentes permitirán definir la variable de calidad de experiencia de una red.

2.7.1 Latencia

La latencia consiste en la suma de los retardos parciales del camino que debe recorrer la información entre un punto y otro de la red, propios de la propagación de la señal en el medio de transmisión y del tiempo de procesamiento y conmutación en su paso por los diferentes nodos de red.

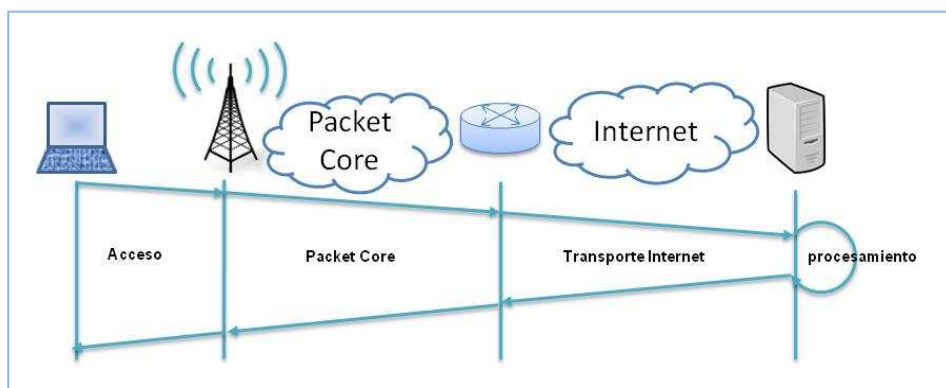


Figura 2.17. La latencia es la suma de los retardos parciales

La latencia es uno de los indicadores más utilizados para mostrar la calidad de una conexión. Esto se realiza habitualmente midiendo el RTT (*Round-Trip Time*, que equivale al doble de la latencia) hacia un host específico utilizando la aplicación ping.

Ping envía un datagrama ICMP *echo* hacia el servidor destino y mide el tiempo entre el envío y el arribo del datagrama. La aplicación permite, entre otras opciones, la configuración del tamaño del datagrama lo que permite simular el envío de un paquete promedio y así obtener una muestra más fidedigna. Esta operación es repetida por la aplicación varias veces y obteniendo el RTT promedio, mínimo y máximo en el período de medición, permitiendo conocer el estado de la conexión hacia un host específico.

Es importante recalcar que lo que se mide es la latencia del canal de comunicación y no incluye los tiempos de procesamiento de las aplicaciones que podría estar accediendo el usuario, es decir, se podría obtener una latencia baja hasta el servidor aún teniendo un gran retardo producto de una aplicación lenta (este último retardo no sería detectado en las mediciones). Luego, este indicador por sí solo no es capaz de entregar la percepción del servicio que experimenta el usuario.

2.7.2 Throughput

El throughput, o tasa de transferencia, es la razón de datos entregados a un destino, medido en términos de bit por segundo y considera no sólo los bytes transmitidos sino también los erróneos, retransmitidos e información de control propia de la conexión. En telecomunicaciones es importante determinar cuál es la máxima tasa de transmisión que alcanzan los dispositivos, pero es importante destacar que el máximo throughput posible de alcanzar es diferente al máximo throughput teórico, debido a diferentes consideraciones que se ven a continuación.

El máximo throughput teórico se puede determinar midiendo la cantidad de datos enviados en un tiempo específico.

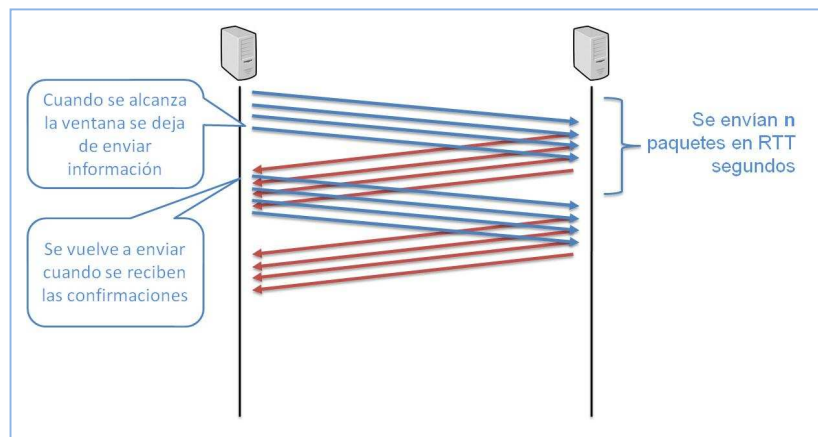


Figura 2.18. La información enviada corresponde a la definida por la ventana de recepción

En régimen permanente se puede apreciar que la cantidad de datos enviados en un periodo de ida y vuelta corresponde a la ventana de recepción, de este modo, se tiene que la tasa de transferencia es directamente proporcional a la ventana de congestión e inversamente proporcional al RTT.

$$\text{Throughput} = \frac{RWND}{RTT} \quad (5)$$

Esta razón es válida por cada conexión TCP y está limitada a la capacidad del canal de comunicaciones. Por ello, los desarrolladores de aplicaciones han optado por la utilización de múltiples conexiones de modo tal que se pueda optimizar el uso del canal y aumentar la tasa de transmisión percibida por el cliente.

Pero en una red móvil el throughput máximo que puede percibir el cliente depende no sólo de las aplicaciones que está accedendo sino también de las condiciones de radio que el abonado tenga y de la utilización que están realizando el

resto de los clientes que comparten su celda, viéndose directamente afectado su rendimiento cuando el resto de los abonados se vuelve más intensivo.

2.7.3 Pérdida de Paquetes

La pérdida de paquetes en TCP genera que el flujo de datos decrezca abruptamente y disminuya el throughput de una conexión y producto de la naturaleza del protocolo, la velocidad original se restablezca lentamente⁶.

Por ello, resulta relevante considerar el nivel de pérdida de paquetes que puede existir en una conexión, medirlo y controlarlo. En la práctica se prefiere mantener la tasa de pérdidas en un nivel bajo de modo tal que los usuarios no vean afectado su servicio por la mala calidad del canal.

Las pérdidas de paquetes pueden ocurrir en diferentes puntos de una conexión, lo habitual es que ellas sucedan en las colas de acceso a los routers. Por ello, es habitual contabilizar este efecto mediante la probabilidad de pérdida de paquetes.

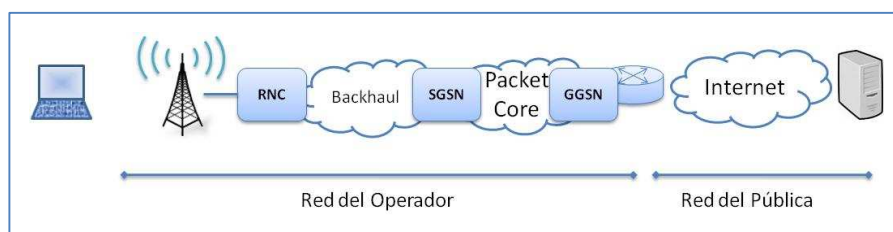


Figura 2.19. La pérdida de paquetes puede ocurrir en cualquier punto de la comunicación.

Si bien es posible que existan pérdidas a lo largo de todo el camino que recorren los paquetes, cada entidad es responsable de minimizar esta pérdida en función de los compromisos de calidad de servicios que ellos tengan comprometidos. En el caso del operador móvil, existe un amplio segmento de red de su responsabilidad administrativa con diversos puntos donde puede actuar y realizar acciones de mejora que se traducirán en una mejor experiencia del servicio.

La pérdida de paquetes genera un cambio de estado en la conexión TCP haciendo que esta vuelva a *slow start* y además a que el umbral *slow start threshold* disminuya su valor lo que se traduce que la tasa con que TCP retoma su velocidad antes de la pérdida sea menor luego de cada evento. Es decir, cada vez tarda más en recuperarse, o sea, la pérdida de un paquete no solo genera la retransmisión de la información perdida sino también que la conexión se haga más lenta.

⁶ Tal como se ve en el Apéndice B.4,

De este modo, es importante que el operador móvil monitoree y controle su tasa de pérdidas o tasas de retransmisiones de modo tal que esta no afecte la calidad de experiencia del servicio que tienen los clientes.

2.7.4 Rendimiento de Aplicaciones

El análisis de las aplicaciones es quizás el punto más importante cuando se analiza la calidad de experiencia del servicio de banda ancha móvil.

Los operadores están acostumbrados a monitorear sus parámetros de red y realizar mejoras en función de estos, lo que muchas veces está desalineado con la percepción del servicio que tiene el usuario.

La percepción que tiene el cliente corresponde al servicio extremo a extremo y no a una parte. Los factores que influyen en la calidad de experiencia se encuentran a lo largo de toda la comunicación, tal como se muestra en la Figura 2.20.

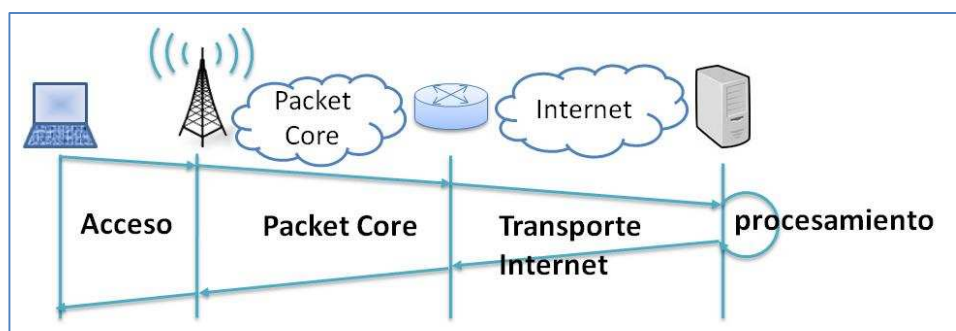


Figura 2.20. El servicio es afectado por todos los nodos a lo largo de la comunicación.

Cuando un servidor web se encuentra saturado y genera retardos en sus respuestas el cliente verá afectado su servicio y será el Servicio al Cliente del operador quien recibirá las quejas por mal funcionamiento, ya que es él quien le brinda el servicio de conectividad y asumirá que cualquier falla que ocurra será de su responsabilidad.

De este modo, es importante considerar que los indicadores de calidad de experiencia deben estar alineados con esta variable y cuando exista una falla en el proveedor de contenido se debe ver que este indicador debe caer y reflejar la falla, aunque esta no sea responsabilidad del operador.

2.8 Costos de Red

2.8.1 Modelo de Costos en BAM

El análisis de costos en banda ancha móvil es una pieza clave para comprender el servicio. Diversos trabajos realizados por proveedores de red como Ericsson [18] [19] y Nokia Siemens Network[20] [26] explican en diferentes niveles de profundidad los aspectos que determinan los costos a los que deben incurrir los operadores para suministrar el servicio de banda ancha móvil.

El análisis radica en el estudio de los gastos e inversiones que debe realizar el operador en la red móvil para brindar el servicio BAM y como afectan éstos en los costos del servicio brindado.

En primer lugar se tiene los costos asociados al acceso, estos consideran las inversiones en hardware y software de radio bases, infraestructuras, arriendo de sitios, costos de energía, mantenimientos del sitio, caminos, etc.

El número de radio bases que despliega el operador depende principalmente de la demanda de datos, la cobertura brindada y de la distribución geográfica de los abonados. Dentro de la cobertura es importante destacar las obligaciones que el regulador le haya impuesto al operador en el momento de asignar el espectro ya que éstas pueden obligarlo a realizar una fuerte inversión en cobertura aun cuando la demanda en esos puntos sea limitada.

A continuación están los costos de transmisión, tanto de backhaul como transmisión entre elementos del núcleo de paquetes.

Las capacidades del backhaul dependen del uso máximo que tenga la celda y de la velocidad que se le permitirá alcanzar al usuario, por lo tanto, es necesario considerar las holguras necesarias para cubrir no sólo los peak de tráfico sino también las estacionalidades y los crecimientos a corto plazo que dependen además de los tiempos de respuesta que pueda ofrecer el proveedor de transmisión para los upgrade de velocidad de los enlaces.

Luego están los costos de controladores, RNC, BSC y equipos de core de paquetes, y finalmente, se tienen los costos asociados al acceso a Internet. En estos casos, el tráfico de las celdas es agregado y concentrado por los nodos lo que permite ofrecer capacidades consolidadas.

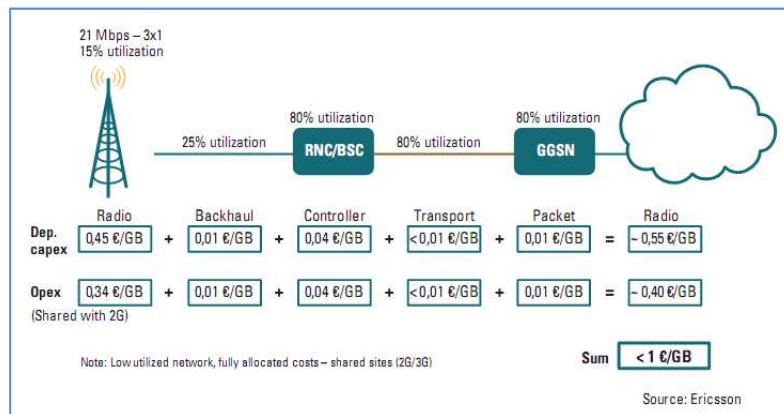


Figura 2.21. Modelo de Costos.

Como se ve en la Figura 2.21, los costos se encuentran fuertemente concentrados en radio ocupando el 60% de los costos de red involucrados en el servicio, por lo cual los operadores deben concentrar sus esfuerzos en generar eficiencias principalmente en este ítem.

En ambos estudios mencionados, los proveedores indican que los costos en una red madura tiende a ser menor a 1€/GByte, es decir, considerando que un usuario promedio trafica 3 GBytes en un mes, el costo de red de este usuario para el operador debiera ser menor a los 3€, pero este valor depende de otros factores que hace que el monto sea considerablemente superior como se indicará más adelante.

2.8.2 El dilema Intensidad v/s Volumen

Como se mencionó en el punto anterior, el costo de un usuario depende exclusivamente del tráfico que este tenga en la hora cargada independiente si su nivel de consumo mensual sea significativo o no. Esta velocidad se determina de la razón entre el throughput máximo y el número de usuarios del servicio, obteniéndose así la velocidad media por suscriptor durante la otra cargada.

En la Figura 2.22 se ve que el perfil de uso de diferentes tipos de usuarios y su aporte en la hora cargada. Aunque el volumen total descargado para los tres perfiles es el mismo, el aporte al throughput total en la hora cargada, en rojo, es claramente diferente y se puede apreciar que $V_{m2} > V_{m3} > V_{m1}$

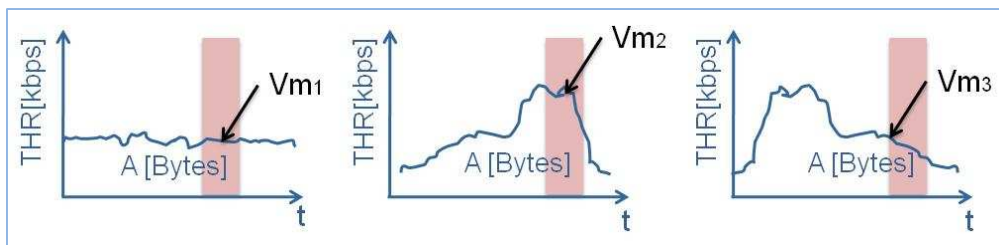


Figura 2.22. Diferentes perfiles de tráfico aportan con distinta intensidad en la hora cargada.

La forma tradicional de medir el tráfico de un usuario corresponde al volumen consumido durante un mes. De este modo, se tiene que un usuario que tiene un alto MOU, o volumen consumido mensual, será un usuario más intensivo. Pero como se ha visto en el ejemplo anterior, esta forma de medir el consumo no representa la intensidad de tráfico que aporta éste en la hora cargada y, por consiguiente, el costo que representa para el operador, lo que induce a un importante error.

Ante esto existen dos posibles enfoques: el primero consiste en caracterizar a un usuario por el throughput medio que aporta en la hora cargada, y el segundo en agrupar a los usuarios según su comportamiento y establecer la relación entre intensidad y volumen que permitirá utilizar la metodología tradicional.

2.8.3 Usuarios Ofensores

Existen distintos tipos de usuarios del servicio BAM, y ellos pueden ser catalogados en función de la utilización de recursos que realizan, en cuatro grupos: usuarios de bajo uso, usuarios normales, usuarios intensivos y usuarios ofensores. Estos últimos son una preocupación permanente para los proveedores de servicios de banda ancha debido a que se alejan del promedio de uso y pese a ser un número bajo de usuarios, representan una gran proporción del tráfico cursado por la red.

El 5% de los usuarios que más trafican en la red pueden llegar a sobrepasar el 40% del tráfico cursado en la red (como se puede ver en el punto 4.1.4), lo que es un gran problema ya que unos pocos abonados usan intensivamente la red sin dejar espacio a que el resto de los usuarios pueda navegar libremente por ella. Esto obliga a los operadores a realizar fuertes inversiones en la red haciendo cada vez menos rentable el negocio y llevándolo al borde del colapso.

2.8.4 Fair Use

Los planes comerciales del servicio de BAM son caracterizados, en algunos operadores, por la velocidad máxima que pueden alcanzar los usuarios. Para ello se utiliza el campo MBR (Maximum Bit Rate) que permite que la red controle la velocidad máxima que puede alcanzar el móvil.

Como se ha visto en el punto anterior, existe un número de usuarios que contribuyen con un importante nivel de tráfico, congestionando las redes y degradando la calidad de experiencia del resto de los usuarios del servicio.

Con la finalidad de controlar a estos usuarios, los operadores han optado por la estrategia del uso justo o *fair use*, que consiste en definir un límite de volumen de tráfico que, una vez alcanzado, modifica su calidad de servicio reduciendo su velocidad de acceso en forma considerable (del orden del 10~20% % del valor original). En la mayoría de los casos este límite sólo es alcanzado por unos pocos usuarios.

Este mecanismo puede ser inefectivo en la práctica si el usuario no cambia su comportamiento. En ocasiones, el usuario sigue siendo intensivo y afecta el rendimiento

de los demás suscriptores. Además esta estrategia puede generar un uso ineficiente de los recursos, pues restringe la velocidad máxima del usuario aún teniendo disponibilidad de recursos.

Existe una alternativa que consiste en manipular la prioridad del cliente, esto es, una vez que el usuario supera el umbral de *fair use*, al usuario se le modifica su prioridad por una menor, lo que se traduce en que el usuario percibirá una baja en su velocidad en la hora cargada del sitio, pero podrá navegar libremente a máxima velocidad del plan cuando el sitio se encuentre con una holgura de capacidad.

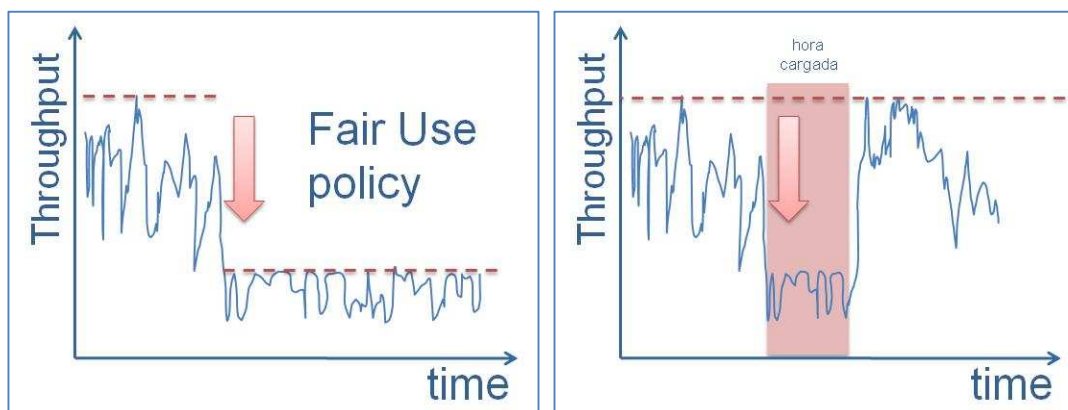


Figura 2.23. Comparación de efecto de cambio de MBR y prioridad como política de Fair Use.

Esta nueva política, también conocida como *Intelligent Fair Use*, o *fair use* inteligente, ofrece al abonado intensivo la opción de tener una mejor experiencia en las horas valles los que incentiva a un cambio de comportamiento y al uso del servicio cuando más existen recursos disponibles, permitiendo un uso más eficiente de los recursos de red disponibles.

2.8.5 Disminución de Costos v/s Calidad de Experiencia

En los puntos anteriores se describen diversos fenómenos que afectan la calidad de experiencia de los usuarios y los costos del servicio. Pero cuando se quiere ofrecer un servicio de una alta calidad de experiencia a costos bajos se debe tener en cuenta todos estos fenómenos y lograr un equilibrio, entre la calidad de servicio que se quiere ofrecer y el costo que tendrá el servicio.

En este sentido se deben tener claros los objetivos del servicio, cuál será la característica de éste como diferenciador frente al resto de los operadores y en un mercado activo hay ser capaces de ofrecer un servicio atractivo para los abonados y para los accionistas del operador.

2.9 Calidad de Experiencia

2.9.1 Definición KQI

De acuerdo a la recomendación E.419 de la ITU-T [13] , el objetivo de describir en detalle indicadores clave de desempeño o Key Performance Indicator (KPI) es establecer una correlación sólida entre los factores tradicionales que servían de indicadores para la gestión de red y el "enfoque comercial" agresivo que impera actualmente en la industria de las telecomunicaciones. La percepción de la calidad de servicio por los usuarios depende del funcionamiento extremo a extremo del servicio y, por lo tanto, se debe encontrar una métrica, o conjunto de ellas, que considere este hecho y que además permita reflejar la percepción del servicio aun cuando esta se vea degradada por factores externos al operador.

Los objetivos comerciales esenciales (KBO, Key Business Objectives) se deducen de los asuntos que cada empresa considera importantes, tales como:

- Ingresos.
- Reducción de costos.
- Mejora de la satisfacción del cliente.
- Cumplimiento de la reglamentación estatal.
- Reducción del uso de recursos.

Los objetivos clave de desempeño (KPO, key performance objectives) son los factores cuantificables que contribuyen a realizar los objetivos de la empresa. Por ejemplo: en relación con el aumento de ingresos, la disponibilidad de servicios o de la red garantizaría ingresos. Ahora bien, la disponibilidad de la red podría ser un requisito impuesto por la reglamentación estatal con el fin de garantizar el servicio, como ocurre en los servicios básicos que proveen los operadores.

Los KPO son los factores cuantificables que directamente permiten realizar objetivos comerciales esenciales (KBO) y cuyo valor puede medirse en tiempo real. Los KPO son los siguientes:

- Disponibilidad.
- Reducción de los efectos de las interrupciones de red.
- Integridad.
- QoS (calidad de servicio).
- Confiabilidad

Los indicadores clave de desempeño (KPI, Key Performance Indicator) son indicadores concretos y cuantificables que se miden en tiempo real y que facilitan

directamente el cumplimiento de los objetivos principales de la empresa al permitir que se alcancen los objetivos clave de desempeño. Estos indicadores se relacionan entre sí para determinar un modelo de calidad.

En esta misma línea la recomendación E.800 [14] muestra que los componentes que contribuyen a la calidad de servicio extremo a extremo se encuentran a lo largo de toda la red y para determinar la QoS extremo a extremo es necesario determinar las condiciones operativas en que se proporcionará el servicio.

Finalmente la Calidad de servicio experimentada o percibida por el usuario QoSE corresponderá a la QoS proporcionada por el proveedor del servicio e influenciada por los factores psicológicos que influyen en la percepción del usuario es por ello que es fundamental conocer la QoSE para optimizar los ingresos y recursos del proveedor del servicio.

La calidad de servicio se puede desglosar en cuatro polos como se muestra en la Figura 2.24:

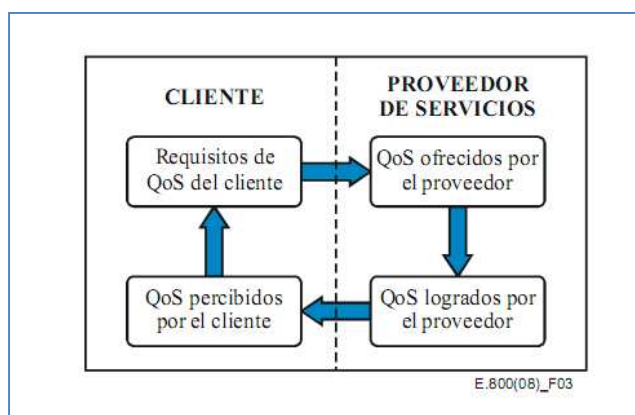


Figura 2.24. Cuatro polos de la calidad de servicio.

Los Requisitos de QoS del cliente, corresponden a la declaración de requerimientos de QoS de un usuario o necesidades de calidad de funcionamiento de un servicio. El QoS ofrecido por el proveedor corresponde a la declaración del nivel de calidad planificada por el proveedor de servicios. Los QoS logrados por el proveedor corresponden al nivel de calidad logrado o alcanzado por el proveedor. Finalmente el QoS percibidos por el cliente, corresponde al nivel de calidad que el cliente declara haber experimentado por el servicio brindado.

Estos cuatro polos muestran lo distinto que pueden ser los indicadores que se puede estar midiendo o considerando a la hora de evaluar un servicio y lo indispensable que es que estos se encuentren fuertemente correlacionados y sea posible medir de modo tal que cuando se mida la calidad de servicio que se está ofreciendo a un cliente sea en realidad la QoS percibida, de modo tal que se tenga claridad en lo que está recibiendo el usuario.

Pero de esta misma forma, los indicadores de calidad de servicio no necesariamente son capaces de mostrar la experiencia que está recibiendo un usuario ya que los indicadores por sí mismos, sólo son capaces de mostrar niveles de servicio de los diferentes factores que influyen en la experiencia del usuario y no la experiencia propiamente tal.

Este nuevo concepto se conoce como calidad de experiencia o QoE (Quality of Experience) que recoge la totalidad de los efectos del sistema a lo largo de la comunicación extremo a extremo y se ve influenciado por las expectativas del cliente y el contexto en que recibe el servicio. Este enfoque permite conocer la calidad percibida por el cliente y de este modo poder modelar el comportamiento que tendrá la red y así determinar cuando el usuario está percibiendo una mala calidad del servicio, aunque los indicadores de QoS tradicionales no se vean alterados y más aun adelantarse a posibles problemas y corregir las falencias antes que se transformen en un incidente.

2.9.2 Modelo de QoE

Considerando los puntos anteriores, es indispensable adoptar una metodología de trabajo que permita modelar el servicio y de ese modo establecer un indicador de desempeño de la calidad de experiencia percibida por el usuario.

La ITU-T [12] recomienda el siguiente proceso para estimar la calidad de funcionamiento de un servicio extremo a extremo:

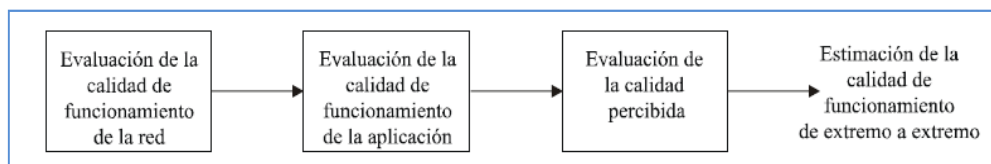


Figura 2.25. Proceso para obtener la calidad de funcionamiento extremo a extremo

El primer paso consiste en la evaluación de la calidad de funcionamiento de la red. Existen dos fuentes principales de información sobre la calidad de funcionamiento de la red, las mediciones y el modelamiento.

Las **mediciones** permiten que la red sea tratada como una caja negra y generan información útil para el resto del modelamiento, es su realización es importante tener en cuenta las siguientes consideraciones:

1. La prueba no debe alterar el medio de la aplicación considerada. Por ejemplo, el control de flujo de TCP responde a las condiciones de red y tiende a llenar las colas en los cuellos de botella, incrementando el retardo por sobre lo que podría medirse en caso que no realizara la prueba.
2. La carga sin mediciones debe ser similar en tamaño y características a las condiciones en las que han de aplicarse las pruebas. Las mediciones en una red no cargada no son particularmente útiles.

El **modelamiento** de la red puede proporcionar la caracterización de la calidad de funcionamiento necesaria cuando la red no esté aún plenamente construida o cuando no puedan alcanzarse las consideraciones esenciales para las mediciones. Las herramientas de modelado exigen bastante pericia e información sobre la red considerada para conseguir los resultados previstos. No obstante, la precisión de este método dependerá en gran medida de la precisión de la información disponible.

El segundo paso consiste en la evaluación de la calidad de funcionamiento de la aplicación. Los modelos de aplicación utilizan como entradas las estimaciones de la calidad de funcionamiento de la red y producen como resultado una o varias métricas de calidad de funcionamiento de la aplicación.

El tercer paso consiste en la evaluación de la calidad percibida como un valor de percepción del usuario. Esto sumado a la información proporcionada por los pasos anteriores entrega como resultado final la estimación de la calidad de funcionamiento extremo a extremo de la aplicación en cuestión.

Esta metodología considera la información de la red, aplicación y percepción del cliente que al ser modelada y correlacionada apropiadamente entregará como resultado el o los indicadores de calidad de experiencia que perseguimos, pudiendo así describir el comportamiento de los usuarios y su percepción del servicio. De este modo, será posible determinar los factores que están afectando la experiencia del usuario y realizar rápida y eficazmente las correcciones necesarias para mejorarla.

La recomendación además incluye un modelo de opinión para las aplicaciones de servicios web estableciendo una correspondencia entre los tiempos de respuesta y descarga con la calidad percibida en una sesión de navegación web que sirve de ejemplo para la implementación del modelo de calidad de experiencia del servicio.

Esta metodología permite la elaboración del modelo de calidad de experiencia y entrega como resultado un indicador que permita establecer el nivel de satisfacción que está recibiendo un abonado del servicio de BAM, el cual se adapta a las particularidades de cada operador y permite la flexibilidad necesaria para comparar resultados entre distintos operadores y así establecer además ranking de operadores.

Capítulo 3. Metodología.

Este capítulo muestra la metodología de trabajo utilizada para abordar los distintos temas que se contemplan en este trabajo. Cómo se aborda el estudio del impacto de los diferentes protocolos que utilizan los servicios, el análisis de tráfico y servicios utilizados por los clientes del servicio y la forma en que se simulan los servicios del usuario para generar el indicador de calidad de experiencia.

3.1 Introducción.

Como se ve a lo largo de este trabajo los usuarios sólo se concentran en usar servicios que sean atractivos para ellos sin tener conocimiento de cómo fueron construidos y ni de la naturaleza del transporte que están utilizando para acceder a éstos.

Ellos desconocen qué protocolos usan o si fueron optimizados para su uso en BAM, es más, no esperan distinguir diferencias de calidad entre redes fijas y móviles.

De este modo, este estudio se concentra en determinar cuáles son los servicios que utiliza el usuario, cuáles son sus características principales y cómo están construidos, para así determinar cómo se ven impactados cuando el servicio de transporte registra alteraciones.

3.2 Impacto de Protocolos.

La elección de los protocolos de transporte elegidos por los desarrolladores para sus aplicaciones no es casual, pero en la mayoría de los casos su elección no contempla todos los casos de uso, dado que fueron concebidos para su uso en la Internet tradicional, es decir, utilizando un servicio de banda ancha fija.

Es por ello que en este trabajo se realiza un análisis de desempeño de las aplicaciones mirado desde el punto de vista de los protocolos utilizados y como se ven afectadas en un ambiente de banda ancha móvil.

Adicionalmente se analiza cómo estos mismos protocolos pueden ver afectado su rendimiento con pequeños cambios de configuración en los sistemas operativos.

3.3 Análisis de Tráfico.

Se realiza un análisis del tráfico que cursan los usuarios de un operador móvil para determinar cuáles son los servicios que utilizan y en qué proporción.

Esta información muestra cual es el perfil de tráfico de un abonado promedio y permite definir cuáles son los principales servicios que se debe simular para el estudio de calidad de experiencia.

Esto ayuda además a tener claridad de cómo el tráfico de los usuarios afectan su experiencia y la del resto de los abonados.

3.4 Simulación de Servicios.

Es importante recordar que para realizar una buena medición de la calidad de experiencia es necesario elaborar una herramienta que sea capaz de simular los servicios que utiliza un usuario. Para ello se utiliza la plataforma de monitoreo del servicio BAM desarrollada [11] . Esta plataforma permite, en primer lugar, la simulación del comportamiento del usuario y la medición de una serie de indicadores de calidad que permite caracterizarlo. Luego, la información recolectada es transmitida, almacenada y procesada en la plataforma que facilita la visualización de estos indicadores y además de la construcción del indicador de calidad de experiencia que se anda buscando.

3.5 Medición.

A continuación se describe la metodología utilizada para la recolección de datos.

En primer lugar se establece que el método de recolección utilizado consiste en la simulación automática del comportamiento de un usuario de acuerdo al punto anterior. Ésta consiste en la ejecución de una serie de tareas en forma serial que da como resultado una muestra característica del comportamiento de un usuario.

Existen diferentes mecanismos para repetir estas muestras en el intervalo de medición. El primero consiste en una medición aleatoria simple, donde se realiza la medida n veces en forma aleatoria durante el intervalo de medición. La segunda consiste en una medición aleatoria estratificada, que divide el intervalo de medición en n intervalos iguales y dentro de cada uno de los intervalos se realiza la medición eligiendo el instante de inicio en forma aleatoria. Finalmente se tiene la medición sistemática donde al igual que la anterior se decide el intervalo de medición en n intervalos iguales pero en este caso el inicio de la muestra coincide con el inicio de cada uno de los intervalos, luego las muestras son iniciadas con una frecuencia constante. En la Figura 3.26 se puede apreciar los diferentes algoritmos que se explicaron.

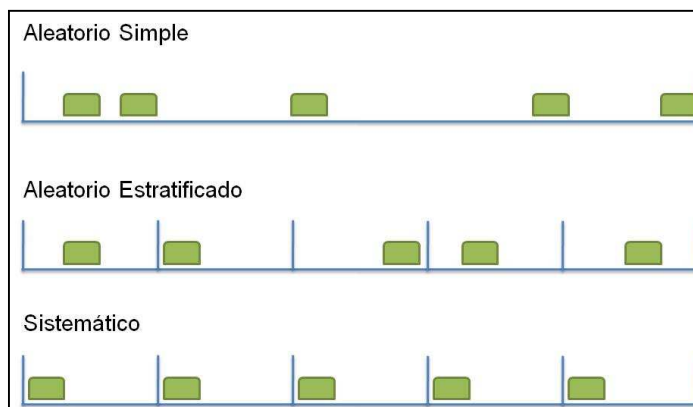


Figura 3.26. Algoritmos de tomas de muestras.

Para este trabajo se eligió el algoritmo sistemático ya que al tener muestras periódicas es posible correlacionar errores sistemáticos o eventos de la red que puedan alterar las muestras.

En cuanto a la frecuencia de la toma de muestra se realiza una serie de toma de muestras para determinar la duración de éstas y así determinar el intervalo de tiempo que permita realizar el experimento con la suficiente distancia entre uno y otro de modo, tal que no se traslapen pero procurando tomar la mayor cantidad de ellas en el período de muestreo, tal como se muestra en la Figura 3.27.

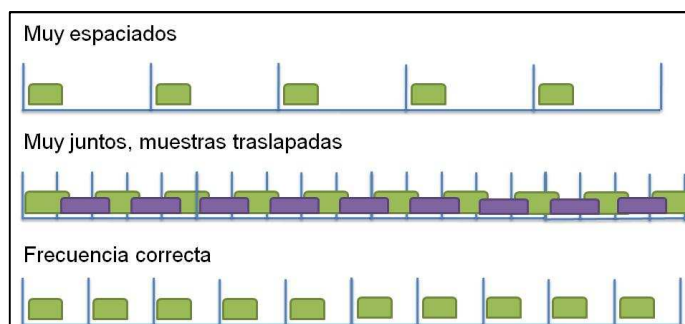


Figura 3.27. Frecuencia correcta para la toma de muestras.

Para ajustar la frecuencia se considera además la desviación y peores casos de modo tal que si la duración de la toma de muestra varía en función del tiempo, debido a la carga de la red o del servidor, no se cometan errores.

3.6 Indicador de Calidad de Experiencia.

Para realizar una medición adecuada de los diferentes indicadores que afectan la calidad de experiencia se desarrollan una serie de rutinas que permiten la extracción de datos esenciales de las aplicaciones más utilizadas por los abonados.

Como primer paso se realiza el estudio detallado de las principales aplicaciones que utilizan los abonados del servicio BAM, luego se estudia su funcionamiento y se

determina cuales son indicadores de calidad que se utilizan para caracterizar el servicio, para luego determinar los puntos de medición adecuados para cada uno de los servicios. Por ejemplo, para el caso de la navegación por sitios web se desarrolla una aplicación que simula la interacción de un usuario frente a un browser:

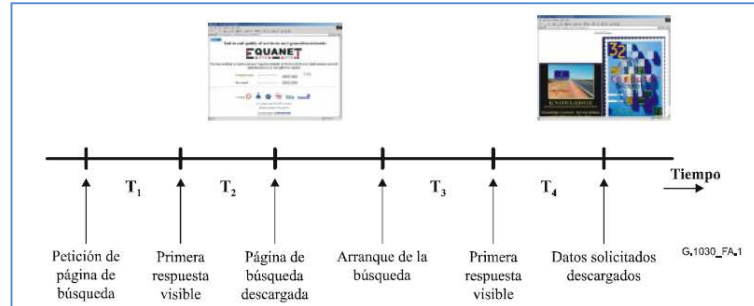


Figura 3.28. Mediciones en la navegación web.

Finalmente, se determina cual es la función óptima que considera todos estos indicadores antes generados y los procesa para determinar el indicador de calidad de experiencia que refleja el estado del servicio.

$$KQI = f(x_1, x_2, x_3, y_1, y_2, z_1, z_2, z_3) \quad (6)$$

La elaboración de dicha función considera además la percepción del servicio de modo tal que se pueda ponderar correctamente los distintos indicadores medidos y se obtenga un indicador de QoE que represente el estado del servicio.

Capítulo 4. Resultados

En este capítulo se analiza el comportamiento de los usuarios del servicio de banda ancha móvil en Chile caracterizando sus hábitos y el desempeño de las aplicaciones más utilizadas, su implicancia en el modelo de costos del servicio y la definición de un indicador de calidad de experiencia.

Durante los meses de julio y septiembre de 2009 y abril a mayo de 2010 se tomaron una serie de muestras del servicio de banda ancha móvil para los distintos operadores principalmente en la ciudad de Santiago, las que sirvieron de base para los análisis que se muestran más adelante. Las mediciones fueron tomadas periódicamente cada 20 minutos durante las 24 horas del día en los meses indicados. Éstas arrojaron diversos indicadores de desempeño del servicio los cuales fueron analizados y se exponen a continuación.

4.1 Modelamiento del Servicio BAM

4.1.1 Perfil de Navegación

¿Qué tipo de tráfico generan los usuarios?, es una interrogante que se debe responder cuando se quiere determinar cuál es el nivel de tráfico que alcanzan los usuarios, y es de vital importancia cuando se quiere que tengan una buena experiencia del servicio.

El perfil de navegación varía continuamente ya que los hábitos de los usuarios de Internet se van modificando a medida que ésta evoluciona o que se ponen de moda nuevos servicios como Twitter, Facebook o Youtube. Y por otro lado, estos hábitos son propios de cada país o región, como se puede ver más adelante.

En la figura 4.29 se ve un ejemplo de perfil de navegación donde se muestra la distribución en diferentes categorías de protocolos para el tráfico visto en forma macro.

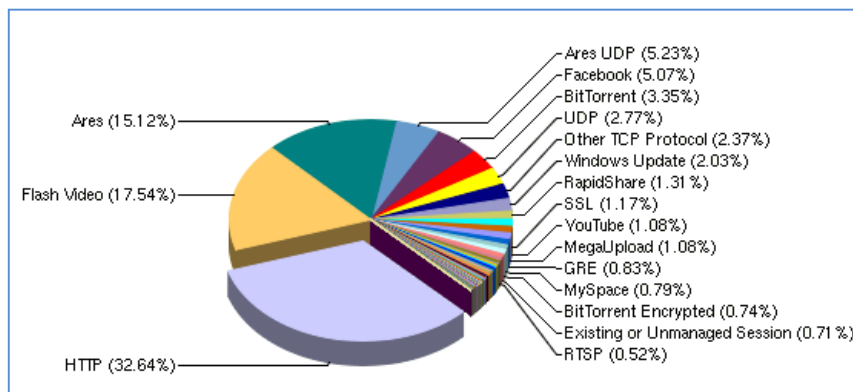


Figura 4.29. Uso de servicios por categoría para usuarios BAM.

Esta figura se obtiene analizando el tráfico de los usuarios por el periodo de una semana y determinando el volumen total transmitido por ellos en ese periodo. Ella muestra una fuerte concentración en las categorías Web Browsing, Peer to Peer, y Streaming Protocols, lo que indica claramente cuáles son las preferencias de los usuarios por los distintos tipos de servicios ofrecidos en Internet.

Esta distribución es característica de los usuarios del servicio BAM y sólo cambia cuando el servicio es utilizado por Smartphone, donde la distribución de tráfico es principalmente browsing y una limitada descarga de archivos. Ésto se debe a las restricciones de los smartphones, que poco a poco se parecen más a un computador por lo que se espera que los tráficos sean similares.

Finalmente se puede decir que en términos generales el 85% del tráfico de los usuarios de BAM es Browsing, Streaming y Peer to Peer como se muestra en la siguiente tabla.

Servicio	Distribución
Browsing	35%
Streaming	20%
Peer to Peer	30%

Tabla 4.30. Principales servicios utilizados en BAM.

Es por ello que es necesario centrar el estudio en el funcionamiento de dichos protocolos para así modelar el servicio.

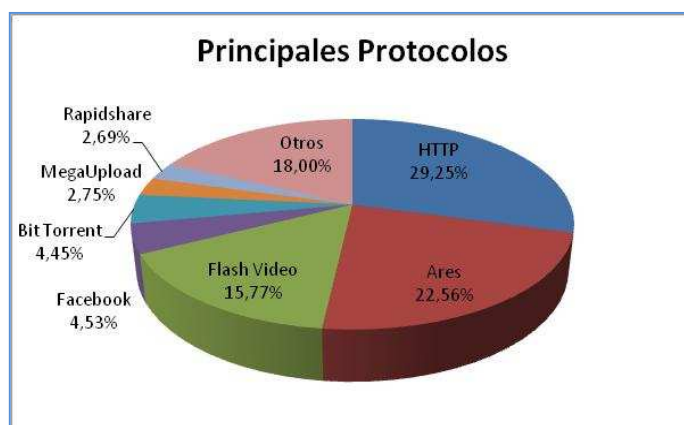


Figura 4.31. Principales protocolos utilizados por los usuarios de BAM.

En la Figura 4.31 se puede ver los principales protocolos utilizados por los usuarios: HTTP, Flash Video y P2P.

El protocolo HTTP como se ha visto es un protocolo cliente-servidor y su implementación en los navegadores (Mozilla Firefox, Google Chrome, Safari, Internet Explorer, etc) se caracteriza por utilizar múltiples conexiones simultáneas lo que ayuda

a utilizar al máximo el ancho de banda disponible, pero no es un protocolo esencialmente intensivo y se caracteriza por ser utilizado en ráfagas, ya que los usuarios visitan una página, la leen y luego visitan otra, donde cada una de las descargas bordea los 450 Kbytes por sitio.

Por otro lado Flash Video utiliza el protocolo RTMP basado en TCP y dada la característica del protocolo se establece una conexión con el servidor del cual se descarga una gran cantidad de tráfico que habitualmente bordea los 40 kbytes por segundo de video, luego un video de 3 minutos tiene un volumen de unos 6,56 Mbytes (15 veces más que una página web).

El caso de Peer to Peer es un poco diferente, primero que todo debido a que el contenido que descarga el usuario es virtualmente infinito ya que habitualmente el usuario está continuamente descargando contenido en modo background y, por otro lado, a diferencia de los anteriores tiene un alto volumen de subida (relación 2:1 y hasta 4:3 bajada:subida en cambio los anteriores mantienen una relación 4:1 bajada:subida)

Para el caso de los servicios de redes sociales, se puede ver que el caso del tráfico hacia Facebook es del orden del 4,5% y a este hay que sumarle el acceso otros tipos de portales como Hyves, LinkedIn o Twitter, que en volumen puede ser menor pero como bien se sabe se tiene un alta penetración en los usuarios.

Finalmente a modo de resumen, se puede ver que la naturaleza de estos permite estudiar el servicio y determinar la experiencia que está teniendo el usuario. En la siguiente tabla se puede apreciar las principales características de los protocolos las que se utilizan para la construcción del modelo de calidad de experiencia.

Servicio	protocolos	Conexiones Simultáneas	RT / NRT	Interactividad	Volumen
Web Browsing	HTTP, TCP	Hasta 20	RT	Interactivo	Bajo
Network Storage	HTTP, TCP	Normalmente 1 Hasta 20	NRT	No interactivo	Alto
P2P	TCP, UDP	cientos	NRT	No interactivo	Alto
Flash Video	HTTP, TCP	Normalmente 1	RT	Interactivo	Medio
Social Network	HTTP, TCP	Normalmente 1	RT	Interactivo	Bajo

Tabla 4.32. Características servicios más utilizados por los usuarios de BAM.

4.1.2 Movilidad

El dimensionamiento tradicional en una red móvil considera que los usuarios se encuentran en movimiento, o sea, que se están desplazando constantemente mientras usan el servicio o que pueden utilizarlo en forma semi estática cambiando de lugar a lo largo del día.

Ésta consideración tiene una alta validez cuando se analiza el comportamiento de usuario de los servicios de voz y de servicios de datos poco intensivos, como es el caso de correo *push* o navegación en el móvil, pero esta premisa se rompe con facilidad en los usuarios del servicio de Banda Ancha Móvil (BAM), debido a que el servicio y los usuarios poseen una naturaleza distinta.

Si bien el servicio que ofrecen los operadores de telefonía móvil es esencialmente un servicio de datos móvil, esta movilidad es interpretada como la posibilidad de disponer del servicio en cualquier lugar y en cualquier momento y a velocidades similares que las de la banda ancha fija.

Por otro lado, se debe recordar que la banda ancha fija hoy día se entrega con un dispositivo WiFi luego es natural que las diferencias entre banda ancha fija y banda ancha móvil sean muy pocas para el usuario, tendiendo a ser considerado como bienes sustitutos y no como complementarios. Esto significa que los abonados exigen los anchos de banda del Internet tradicional que utiliza una red de acceso dedicado, pero en una red con medio de acceso compartido, convirtiéndose en un enorme desafío para los operadores móviles.

Inicialmente BAM se entiende como un segundo servicio de acceso a Internet que aporta la movilidad como valor agregado. Estos usuarios utilizan la banda ancha móvil principalmente para conectar sus laptops o netbook a Internet y acceder a los mismos servicios que cuando se encuentran en su acceso principal a Internet (en su trabajo, un café, la Universidad e incluso desde su casa). Por otro lado, se tiene a otro tipo de usuarios que han reemplazado su acceso a Internet fijo por este servicio.

Estos usuarios, intensivos, esencialmente nómades e incluso estacionarios distorsionan el modelo de dimensionamiento y hacen obligatorio realizar los ajustes necesarios que permitan determinar sus efectos dentro de la red.

Estudiando el comportamiento de los usuarios de BAM durante 24 horas se puede determinar cuántas celdas distintas visitan. Con esta información se puede categorizar a los usuarios en usuarios móviles, nómades o estacionarios de acuerdo a la siguiente tabla:

Movilidad	Número de celdas visitadas
Estacionario	1-2
Nómade	3
Móvil	3+

Tabla 4.33. Categorización por movilidad.

Esto lleva al siguiente resultado

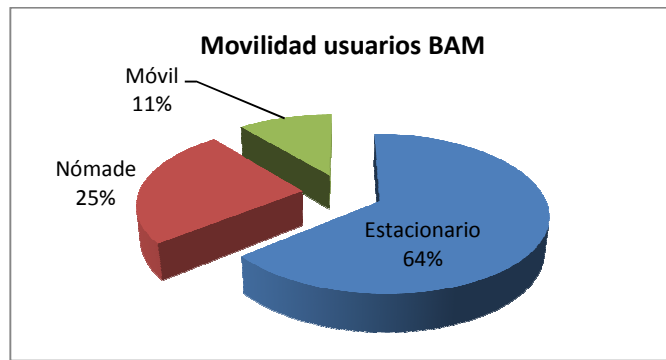


Figura 4.34. Movilidad de Usuarios Banda Ancha Móvil

Donde se puede ver que los usuarios son principalmente estacionarios o nómades y sólo una pequeña proporción son esencialmente móviles. Luego se puede ver que se está en un escenario distinto al inicialmente esperado lo que claramente puede distorsionar los parámetros de diseño y hace necesario un análisis más fino cuando se realiza la planificación y la ingeniería de detalles de la red móvil. Por otro lado, esta característica de los usuarios del servicio muestra que su comportamiento es muy similar al del servicio de banda ancha fija lo que invita a estudiar en detalle su comportamiento en lo que respecta a los servicios que utiliza, lo que es analizado en los siguientes puntos.

Otro punto que es importante de analizar es la distribución geográfica que tienen los usuarios, la que habitualmente está altamente correlacionada con la distribución de la población, pero además existen otros factores que modifican esta distribución, como son la distribución de socioeconómica, la cultura digital que tengan las regiones y además la presencia de servicios de banda ancha fija que puedan existir en las zonas.

4.1.3 Intensidad de uso.

Con la finalidad de establecer una relación entre throughput medio en la hora cargada y volumen mensual se tomó una muestra aleatoria de usuarios con el mismo plan de banda ancha móvil, estableciéndose una clara tendencia, tal como se muestra en la Figura 4.35.

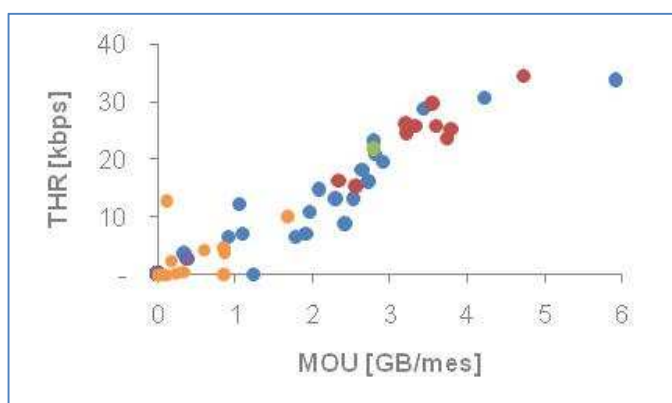


Figura 4.35. Throughput medio en la hora cargada y volumen medio

En la Figura 4.35 se puede ver el comportamiento de los diferentes usuarios y que claramente tienen un comportamiento lineal, pero tras una segunda mirada se puede ver que el comportamiento es lineal pero por tramos, o más bien que el segmento superior de la curva está acotado. En la Figura 4.36 se puede ver el comportamiento al agregarse la línea de tendencia de los puntos obtenidos.

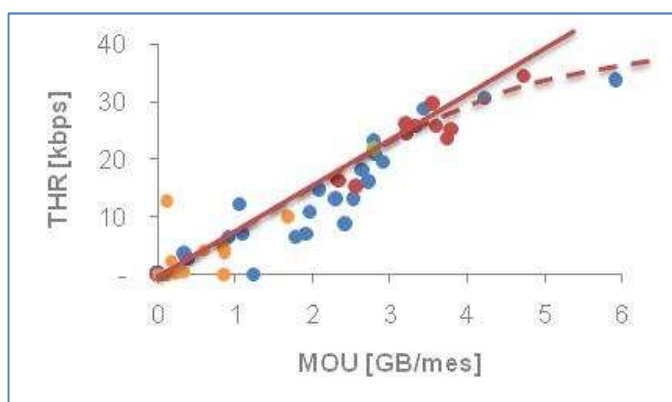


Figura 4.36. Líneas de tendencia de throughput medio en la hora cargada y volumen medio

En ella se puede ver cómo, a medida que aumenta el consumo, el throughput medio también aumenta pero luego de un punto la curva comienza a cambiar su pendiente y a aumentar con mayor velocidad el MOU que el throughput, esto se debe principalmente a la limitación de velocidad del medio de transmisión, es decir, aunque el usuario requiera transmitir un alto volumen de datos, su velocidad no puede superar la ofrecida por el plan o el medio de transmisión, pero si puede aumentar el número de horas que transmite datos al día lo que hace incrementar el volumen transmitido.

4.1.4 Usuarios Intensivos

Otra variable interesante que se analiza es el volumen de datos que trafica el usuario considerando las diversas familias de planes comerciales del servicio de BAM.

Cuando se estudia el nivel de volumen que trafican los usuarios se descubre que existe un porcentaje de usuarios intensivos y más aún una proporción de usuarios que pueden ser categorizados como usuarios ofensores. Estos últimos, ya no sólo están en el borde de la rentabilidad para el operador móvil sino más bien generan un alto costo y es importante analizar si existen beneficios cruzados que justifiquen la rentabilidad y permanencia de este abonado.

Si se considera que los usuarios son altamente estacionarios se tiene que ellos pueden ser atraídos a un servicio fijo de alta velocidad entregándole un servicio de mayor calidad y acorde a sus necesidades, convirtiéndose en una importante oportunidad de negocio que además disminuiría los altos costos estos usuarios representan para la red móvil.

Cuando se quiere determinar qué usuarios son ofensores, lo primero que se debe determinar es la métrica que se utilizará para determinar qué usuario es ofensor. La forma tradicional para determinarlo es utilizando el nivel de consumo que realiza un usuario mensualmente, es decir, midiendo la cantidad de Bytes que trafica en un mes, también llamado MoU por su analogía a la voz (MoU, minutes of use) aunque para generalizar su significado debiera entenderse como Measurement of Use.

Otra forma de medir el uso que hacen los usuarios corresponde a determinar la intensidad de uso que tiene un abonado en la hora cargada, que corresponde efectivamente al impacto que tiene en la red su uso.

Tras hacer un análisis del consumo en Gbytes mensuales que consumen los usuarios y su aporte en la hora cargada se puede ver que no existe una relación directa entre ambos factores, tal como muestra la figura 4.37.

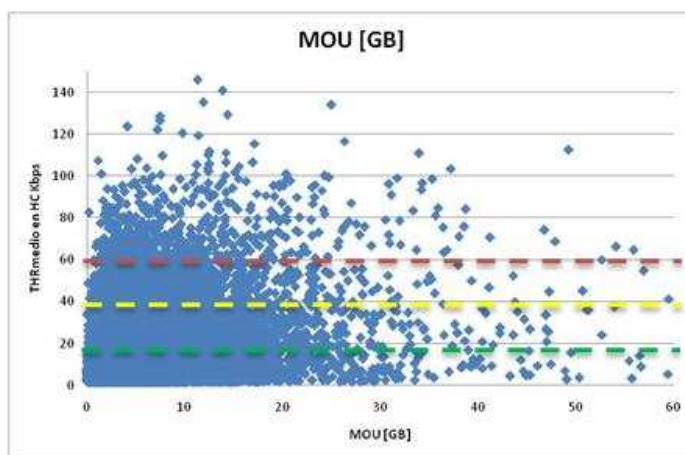


Figura 4.37. Relación throughput medio en hora cargada v/s consumo total

Esto significa que un usuario de alto MoU no necesariamente genera tráfico importante en la hora cargada, debido a que su tráfico puede estar concentrado en la

hora valle, lo que significa que el usuario tiene menos costo que otro que realiza su tráfico en la hora cargada.

Si se considera al 5% de los usuarios que más trafica de acuerdo a los criterios de throughput en la hora cargada y MoU se puede ver que son responsables de diferentes niveles de tráfico, tal como lo muestra la Figura 4.38:

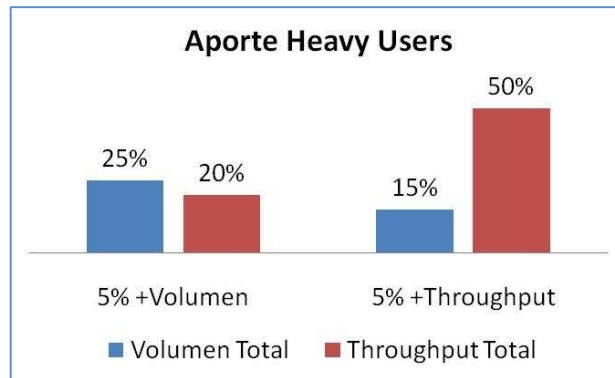


Figura 4.38. Aporte de los heavy users al volumen total traficado en un mes y al throughput en la HC.

Esto quiere decir que el 5% que más volumen consume en la red sólo es responsable del 20% del tráfico en la HC; en cambio, el 5% con más alto throughput en la HC pueden aportar hasta el 50% del tráfico en ese periodo, o dicho en otras palabras, son responsables del 20% y 50% del costo del servicio respectivamente.

Dicho esto, queda la duda de cuántos de los *heavy users* por volumen, son además *heavy users* por throughput en HC, o sea, cumplen ambos criterios. Tras analizar los datos se obtiene que el 25% de los usuarios son *heavy users* por throughput y volumen. Esto quiere decir que el 75% de los usuarios que más trafican no cumplen el criterio de heavy user por throughput, lo cual indica que no consumen excesivos recursos de la red tal como muestra la figura 4.39.

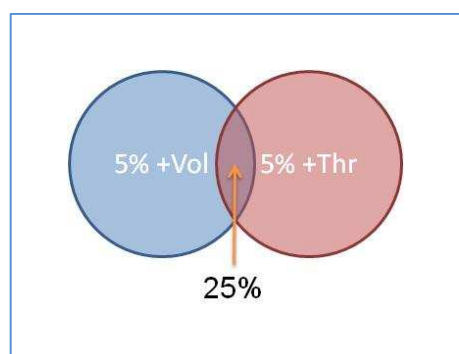


Figura 4.39. Sólo el 25% de los usuarios cumplen ambos criterios.

Independiente de la metodología que se utilice para indicar si un usuario es o no *heavy user*, ellos son responsables de una parte importante del tráfico de la red y su nivel de consumo afecta al resto de los usuarios, degradando la calidad de experiencia que pueden alcanzar los usuarios. Por otro lado, es importante definir el criterio con el cual se determinará si un usuario es o no *heavy user* de modo tal que se pueda determinar a ciencia cierta cuánto afectan a los costos y al resto de los usuarios, si no se estaría aplicando políticas de control de congestión a usuarios que no degradan la red.

El problema de la metodología de throughput es que es difícil de implementar. La forma más natural es medir el consumo de tráfico que tiene un usuario en un periodo específico, es decir, contar sus bytes. Pero un punto intermedio, consiste en medir la cantidad de bytes que consumen los usuarios en un periodo corto de tiempo como lo puede ser en la última hora, lo que se acerca mucho a la posibilidad de medir el throughput instantáneo que cursan los clientes.

Utilizando esta última metodología se puede ver que esta forma de identificar a los *heavy users* muestra que éstos concentran del orden del 35% del throughput, mostrando además lo eficiente que resulta ser el método.

En el caso de uno de los proveedores más importante de equipos de inspección de paquetes, como es Sandvine⁷, ellos incluyen dentro de sus posibilidades la funcionalidad de identificar a los heavy user o como ellos llaman a los Top Talker utilizando este criterio de identificación, o sea, el nodo marca a los usuarios como top talker cuando trafican más de un cierto volumen en las últimas horas.

Finalmente es importante destacar que la linealidad que se mostraba en el punto anterior entre volumen e intensidad de uso se rompe cuando se quiere analizar el caso de los heavy users ya que, como se muestra, su comportamiento no está correlacionado, adicionalmente es importante repetir que un usuario que consume un gran volumen de datos en un mes, no necesariamente es un usuario que realiza su tráfico en la hora cargada y por lo tanto el costo para el operador podría ser incluso menor que el de un usuario normal.

4.1.5 Fair Use

Para analizar el efecto de las políticas de *fair use* se modela el sistema considerando una celda con capacidad de 2Mbps, política de asignación de recursos Round Robin y dos usuarios simultáneos traficando distintos niveles de datos tal como se muestra en la figura 4.40.

⁷ www.sandvine.com

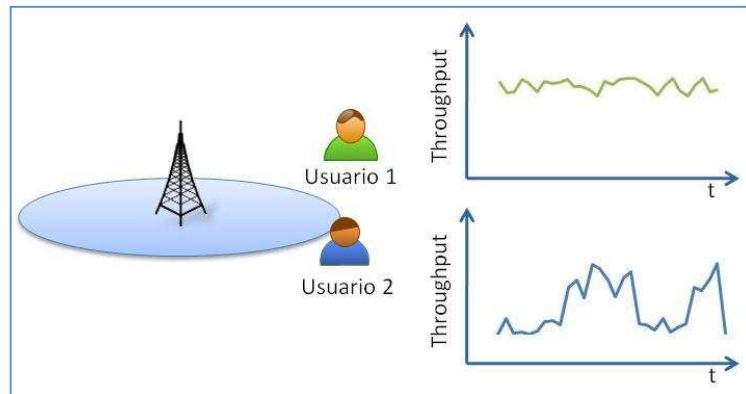


Figura 4.40. Tráfico de dos usuarios compartiendo la misma celda.

Como se puede observar, el primer usuario presenta un alto nivel de tráfico a lo largo del tiempo y el usuario 2 se observa un nivel de tráfico bajo con altos niveles a ráfagas. Para el análisis se consideró que al usuario 1 se le aplica una política de *fair use* en medio de su transmisión.

Cuando se encuentran los dos usuarios en la misma celda compartiendo el ancho de banda disponible, se puede observar que las condiciones de throughput para el usuario que cae en la política de *fair use* son claramente afectadas, viéndose y su throughput máximo se limita al límite máximo de la política aplicada (MBR=200kbps), tal como se muestra en la figura 4.41.

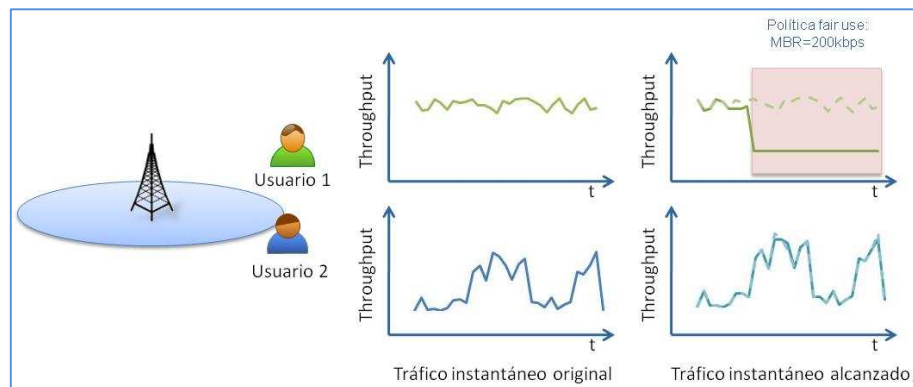


Figura 4.41. Al caer en FU el usuario 1 ve degradado su servicio limitándose su throughput máximo.

Si se analiza el tráfico total del sistema se puede ver que sólo en algunos puntos se alcanza la capacidad máxima de la celda y el usuario 1, que tiene aplicada la política de *fair use*, ve degradado su flujo de datos aun cuando existe capacidad disponible en la celda la mayor parte del tiempo.

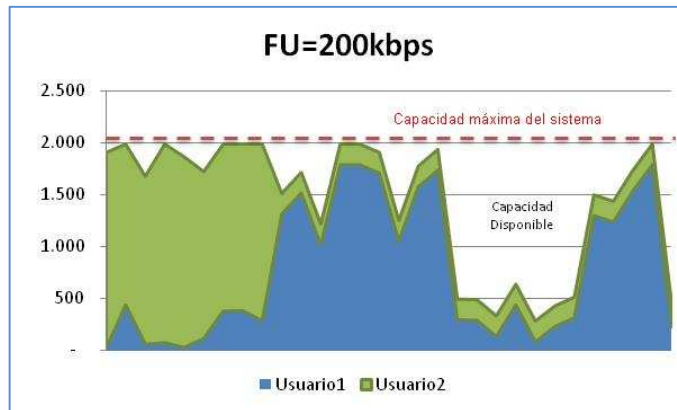


Figura 4.42. El usuario 1 tiene limitado su throughput independientemente si existe o no capacidad en el sistema.

En el caso del Fair Use inteligente se puede ver que a nivel de celda el escenario es muy distinto. En este caso al usuario afectado por la política sólo se le cambia la prioridad y por lo tanto sólo se ve degradado cuando hay poca capacidad en el sistema.

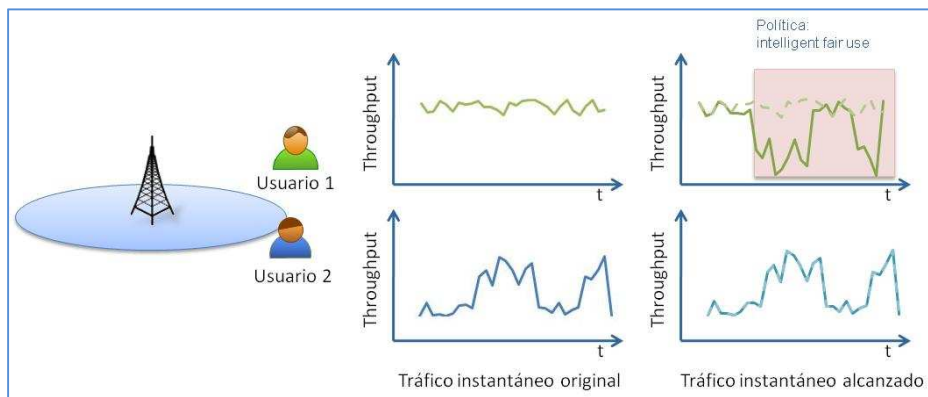


Figura 4.43. Al caer en FU el usuario 1 ve degradado su servicio sólo cuando no existe capacidad en la celda.

Este mecanismo mejora notablemente la experiencia del usuario ya que cuando existe capacidad en el sistema el usuario puede navegar a la máxima velocidad que le permita el sistema sin degradar al resto de los usuarios, consiguiéndose una utilización más eficiente del sistema.

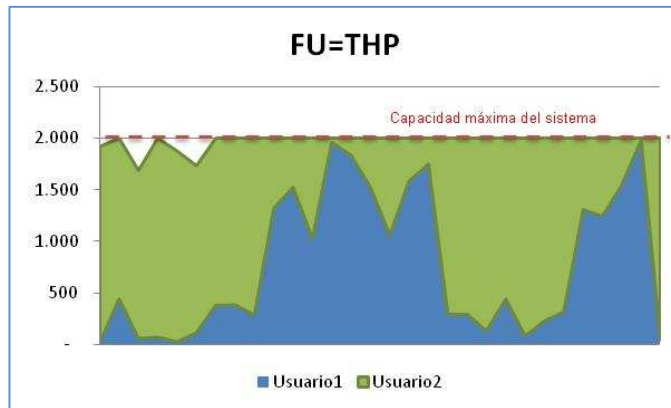


Figura 4.44. El usuario 1 utiliza la capacidad remanente del sistema.

El problema de este mecanismo es que el tráfico agregado es mucho mayor, por lo tanto, es necesario contemplar un mayor uso de recursos en los RNC y en el núcleo de paquetes, pero con un mejor resultado para el cliente final.

Aunque en ambos casos se aplica una política de restricción, estos tienen efectos diferentes en el usuario y por consiguiente en la experiencia que logra del servicio.

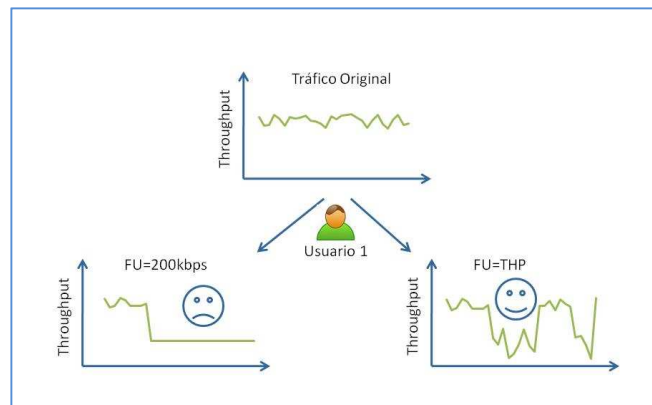


Figura 4.45. La política de cambio de prioridad (THP) del *intelligent fair use* logra una mejor percepción del servicio ya que su tráfico es restringido sólo en la medida que no hay capacidad disponible.

Por otro lado, si se analiza el tráfico global resultante de las distintas políticas de *fair use* se puede ver resultados muy distintos.

En el caso de la política de fair use, se puede ver el throughput de los abonados que poseen el mismo ciclo de facturación comienza a decaer a medida que avanza el periodo de facturación y una vez que se inicia un nuevo ciclo se restablece al nivel que debiera haber tenido sin aplicar el criterio de FU, tal como muestra la figura 4.46.

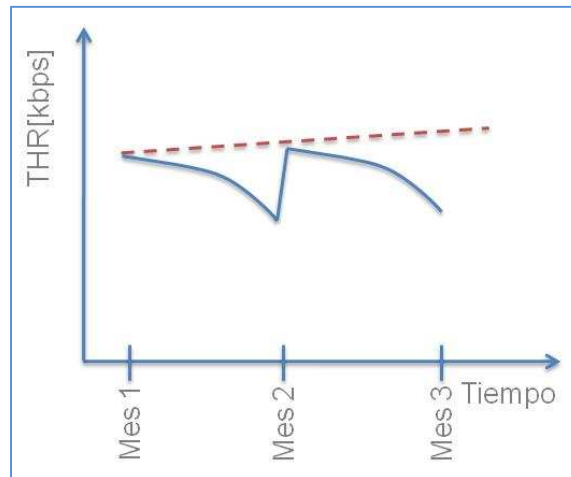


Figura 4.46. Efecto en el throughput total tras aplicar políticas de fair use.

Esta política genera una disminución temporal en el throughput total de los abonados pero en términos de capacidad de red, es necesario planificar el crecimiento de la red en función del primer día lo que muestra que no tiene ningún efecto en la red, o sea, la penalización del usuario no genera mejoras en la red.

Pero si este mecanismo se utiliza combinado con el uso de múltiples ciclos de facturación se puede lograr una importante disminución el throughput total de la red que en el caso de 3 ciclos bordea el 10% tal como lo muestra la figura 4.47.

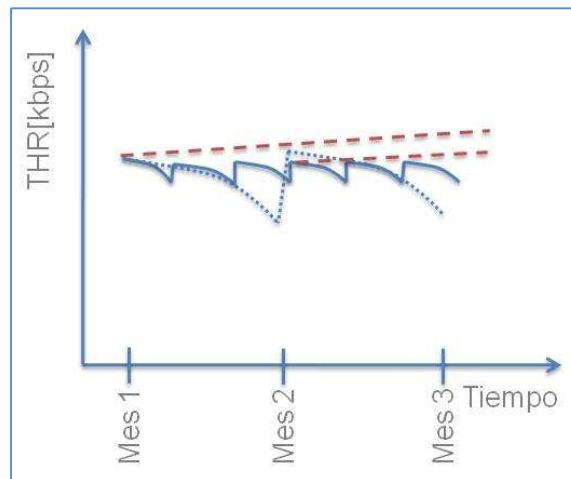


Figura 4.47. Aplicación de *fair use* con múltiples ciclos de facturación.

Finalmente, en el caso de intelligent fair use, se tiene que el tráfico agregado final no debiera variar en demasía debido a que sólo es restringido el tráfico de los usuarios cuando la capacidad del sistema es alcanzada, optimizando el uso de recursos.

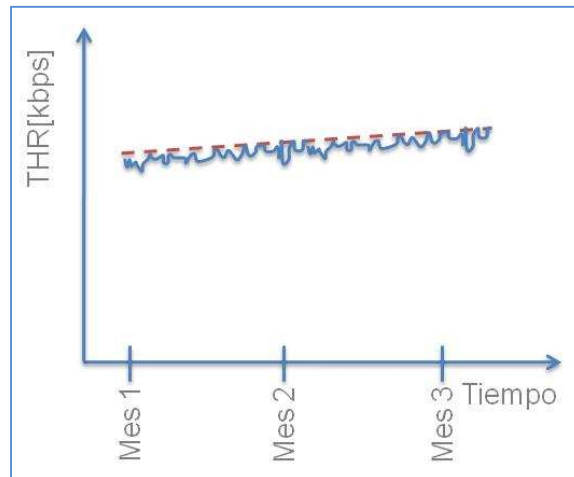


Figura 4.48. El uso de la política de *intelligent fair use* no mostraría una variación importante del tráfico total.

Esta estrategia permite un uso más eficiente de los recursos de radio, no así en el lado del core, ya que esta política aumenta el uso de recursos debido a que usuarios en fair use que antes estaban limitados aunque hubiese capacidad disponible ahora podrán navegar a máxima capacidad.

Por otro lado, los usuarios que se encuentran con política de *intelligent fair use* no distinguen una gran degradación al traspasar su límite salvo al encontrarse en la hora cargada, este nuevo estado podría generar que los usuario no distinguan su estado y naveguen libremente incentivando el uso desmedido del servicio lo que jugaría en contra de la política.

4.2 Desempeño de Aplicaciones

4.2.1 Throughput

Como se ve en el capítulo anterior el throughput máximo que puede alcanzar una conexión TCP viene dado por la siguiente fórmula.

$$\textit{Throughput} = \frac{RWND}{RTT} \quad (7)$$

Esta relación es de gran importancia cuando se quiere saber cual el *throughput* máximo que puede alcanzar una conexión. Para determinarlo basta con saber cuál es el retardo entre el cliente y el servidor (RTT) que para una conexión de banda ancha típica es del orden de los 300 [ms], luego se puede calcular que en este caso el throughput máximo alcanzable por una conexión TCP es de:

$$\text{Throughput} = \frac{RWND}{RTT} = \frac{2^{16} \text{ Bytes}}{300 \text{ ms}} \approx 1,6 \text{ Mbps} \quad (8)$$

Y la gráfica del throughput en función de RTT es la siguiente:

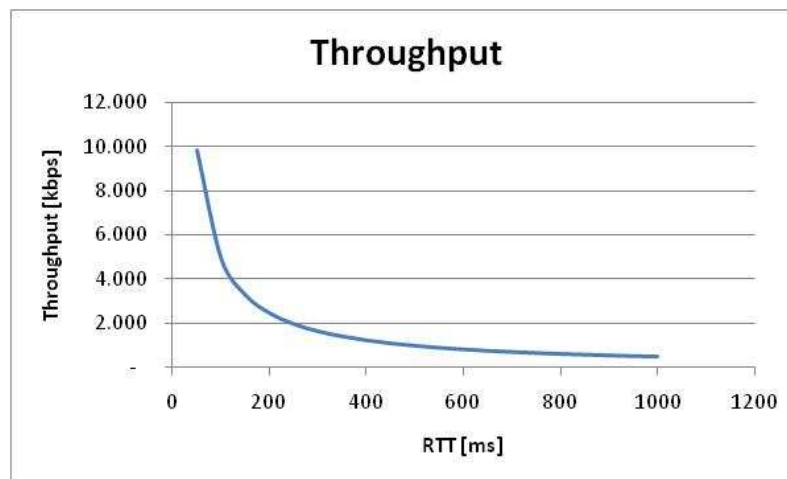


Figura 4.49. Throughput en función del RTT de una conexión TCP

Al contrastar esta curva con los resultados obtenidos durante diferentes pruebas del servicio BAM se obtienen los siguientes resultados:

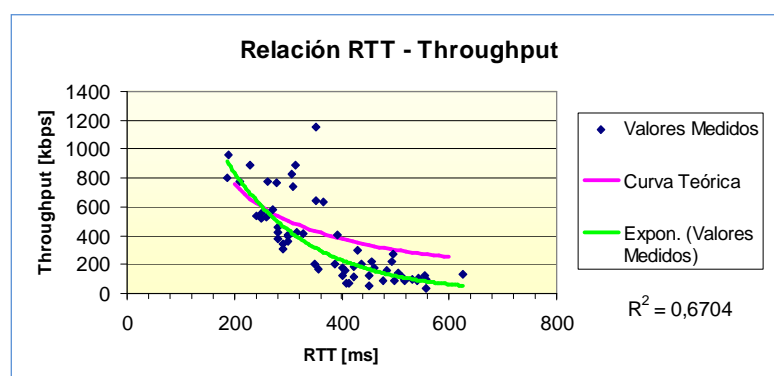


Figura 4.50. Throughput Real en función del RTT para los sitios más comunes.

Estos resultados confirman que el throughput de una conexión disminuye a medida que aumenta el RTT al servidor, luego para aumentar el throughput de una conexión se debe o disminuir el RTT o aumentar la ventana. El primer caso no siempre se puede hacer mucho ya que normalmente el RTT depende directamente del servidor, como se ve más adelante. Luego, la única posibilidad que queda es aumentar la ventana.

En ese sentido el RFC1323 de la IETF [7] define un factor de escalamiento para el campo window que permite definir valores mayores, de este modo se se tiene que:

$$window = WIN * WS \quad (9)$$

donde WIN es el valor del campo window original y WS (window scale) es el factor de escalamiento definido por:

$$WS = 2^S \quad \text{con } S \leq 14 \quad (10)$$

De este modo se tiene que el valor máximo será:

$$window = 2^{16} * 2^{14} \quad (11)$$

Es decir, se pasa de una ventana de 64KBytes a una de 1 GBytes

Se debe destacar que el valor normalmente utilizado para el factor de escalamiento es 4, luego el efecto de este aumento puede verse en la figura 4.51:

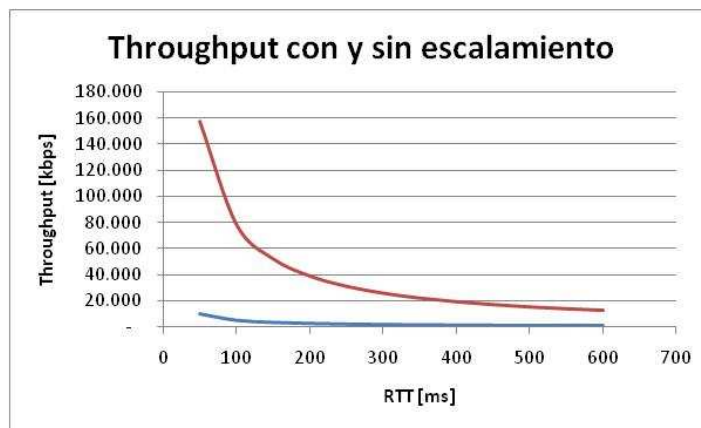


Figura 4.51. Throughput en función del RTT de una conexión TCP con y sin escalamiento

Es decir, 16 veces mayor para un flujo TCP sólo con utilizar el RFC1323. Esto significa que se pueden lograr importantes mejoras en los tiempos de descargas al utilizar la opción de escalamiento. En la figura 4.52 se puede apreciar esta ganancia para diferentes valores de RTT para el caso de máximo escalamiento.

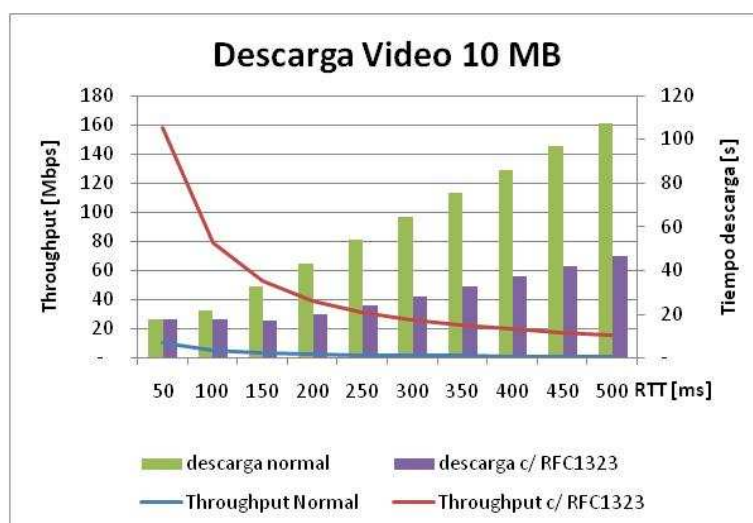


Figura 4.52. Throughput y tiempos de descarga con y sin escalamiento

El valor de escalamiento que utiliza TCP es definido automáticamente en función de las capacidades del medio de transmisión eligiendo un valor óptimo.

Por otro lado, si se considera que el medio de transmisión no es ideal y se incluye la tasa de pérdida del canal y sólo utilizando la mejora de escalamiento que introduce el RFC1323 se puede apreciar una mejora en la tasa de descarga de más del 20%, como se puede observar en la siguiente tabla.

Ítem	Sin RFC1323	Con RFC1323	Mejora
Tamaño Ventana [Bytes]	65.535	223.040	3,4 veces
Tasa media de descarga de un sitio Web [kbps]	323,41	438,54	26,25%
Tiempo medio de descarga de un sitio Web [s]	9,66	8,15	18,46%
Tasa media de descarga de un sitio FTP [kbps]	1.232,36	1.628,87	23,88%

Tabla 4.53. Comparación de throughput con y sin usar RFC1323

Aunque la ventana se aumenta 3,4 veces sólo se puede observar una mejora en torno al 25% en la tasa de transferencia, debido a las restricciones que impone el medio de transmisión y la naturaleza de las aplicaciones utilizadas que no se encuentran diseñadas para sacar el máximo provecho al medio.

Es interesante recordar que en Chile existe una latencia importante comparada con EEUU o Europa debido a la distancia que existe hacia los principales servidores de

contenido, lo que se traduce en que muchos sitios bordeen los 500ms de retardo como se muestra en la figura 4.54:

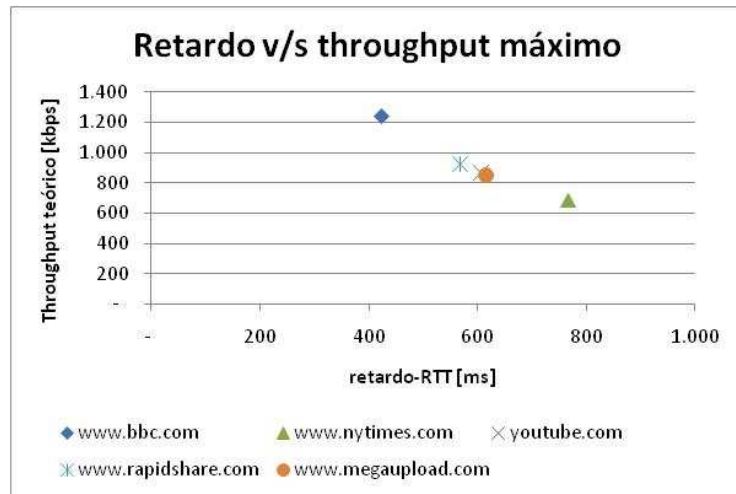


Figura 4.54. Throughput en función del RTT para distintos sitios

Esto muestra que existen grandes desafíos para realizar optimizaciones en el retardo, lo que puede mejorar considerablemente el throughput a un sitio. Si se recuerda que el throughput máximo de TCP es la razón entre la ventana y el RTT, se puede calcular cual sería la mejora en throughput si se disminuye el retardo del enlace, teniendo de este modo la siguiente tasa de mejora:

$$mejora(\alpha) = \frac{\frac{RWND}{RTT'}}{\frac{RWND}{RTT}} = \frac{RTT}{RTT'} = \frac{RTT}{RTT \cdot (1-\alpha)} = \frac{1}{1-\alpha} \quad (12)$$

Esto quiere decir que si se disminuye un 10% el retardo, se podría mejorar un 11% el throughput, en cambio si se hace hasta la mitad de su valor, el throughput podría aumentar en un 100% como se puede apreciar en la siguiente gráfica

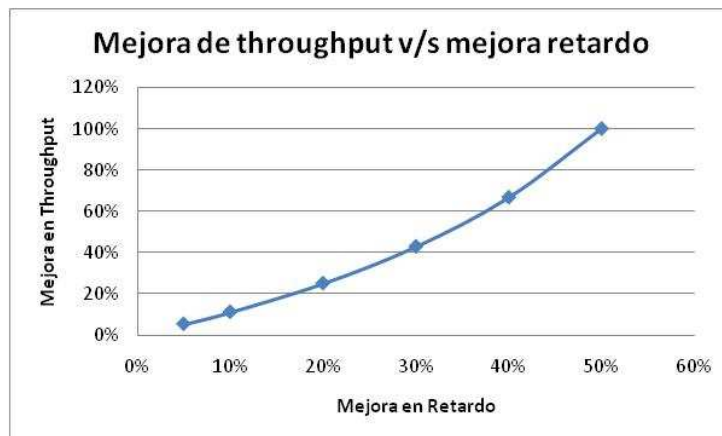


Figura 4.55. Mejora del throughput al mejorar el retado

Esto no sólo permite hacer importantes mejoras en el throughput sino además explica el motivo por el cual se pueden lograr grandes tasas de descargas cuando se encuentra cerca de los proveedores de contenido.

Finalmente, se puede ver que si el objetivo es maximizar la tasa de transmisión, se tiene que este se puede lograr considerando los siguientes métodos.

- Minimizar la latencia de la red.
- Minimizar la tasa de errores.
- Maximizar la ventana de congestión

4.2.2 Control de flujo en TCP

El throughput instantáneo se ve profundamente afectado por las condiciones del entorno.

En primer lugar una correcta configuración en los parámetros window y window scaling permiten importantes mejoras en el rendimientos de las aplicaciones, y por otro lado, agentes externos como tasa de errores son capaces de disminuir significativamente el throughput debido a la dificultad de adaptación que tiene TCP cuando se encuentra en un medio con altas pérdidas.

Por otro lado, las principales aplicaciones son basadas en TCP y debido al lento inicio, este no siempre es capaz de alcanzar las máximas tasas permitidas por el medio de transporte debido a que al transmitir niveles bajos de información slow start regula el throughput. Como se vio anteriormente para lograr una tasa de 1 [Mbps] es necesario transmitir al menos 400kBytes.

Debido a esta desventaja es que las aplicaciones han debido adaptarse teniendo que levantar múltiples conexiones TCP para compensar este problema y así alcanzar

las tasas necesarias que logren que el contenido llegue al usuario en el tiempo estimado.

4.2.3 Aplicaciones Interactivas

Las aplicaciones interactivas son una de las más importantes al momento de analizar la percepción del usuario. Esto es debido a que el usuario está continuamente interactuando ya sea con una máquina o una persona a través de la aplicación que está utilizando. Esto significa que el usuario toma conciencia del retardo en la respuesta y cualquier mejora que se pueda realizar en este sentido es percibida inmediatamente por el usuario.

Dentro de esta familia de aplicaciones se encuentran las aplicaciones de mensajería instantánea, VoIP y algunas aplicaciones web, entre otras.

Tradicionalmente el operador busca mejorar la tasa de transferencia realizando mejoras en su infraestructura que permitan asegurar este indicador. El problema es que no necesariamente esta es la forma adecuada de realizar mejoras.

Muchas de estas aplicaciones interactivas se destacan por transferir una baja cantidad de datos y considerando que las aplicaciones son mayoritariamente basadas en TCP, se puede ver que el mecanismo *slow start* juega en contra, ya que el flujo de datos nunca logra altas tasas de transferencia independientemente si el canal de comunicación lo permite o no.

Por ejemplo, para enviar 160 caracteres (bytes) se tiene lo siguiente:

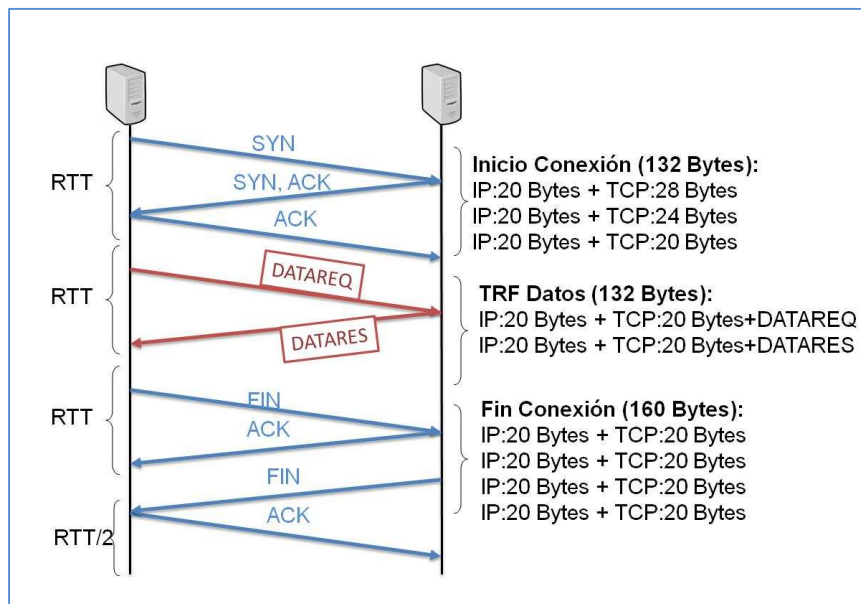


Figura 4.56. diagrama de flujo para el envío de datos sobre TCP

Es decir el throughput alcanzado corresponde a:

$$\text{Throughput} = \frac{132+80+DATA+160}{3,5*RTT} \text{ (13)}$$

Si se considera un retardo medio de 200 [ms] se obtiene que el throughput alcanza los 16,7 kbps, cifra que está por debajo de las capacidades ofrecidas a los usuarios.

Del análisis se puede desprender que el tiempo de respuesta es de $3,5 * RTT$, o sea, de 500 [ms] y más aun, se tiene que es directamente proporcional al delay, dejando en manos del retardo la experiencia que pueda lograr el usuario.

4.2.4 Aplicaciones Intensivas

Como se ha visto, las principales aplicaciones intensivas que usan los usuarios en la red de un operador de banda ancha son las relacionadas con compartición de archivos. Esto es, aplicaciones de P2P y Network storage⁸ (descargas http).

En el primer caso, se tiene que P2P es una aplicación muy agresiva que inunda la red mediante conexiones TCP y UDP de modo tal que se logre una alta tasa de transferencia independientemente de la pérdida de paquetes. Esto quiere decir que dichas aplicaciones son capaces de generar cuellos de botella para el caso de las conexiones UDP, debido a que este protocolo de transporte carece de control de flujo. Y en el caso de las conexiones TCP se ve que la adaptación del protocolo regula el uso de los canales.

Finalmente se analiza el tipo de contenido que es descargado por los usuarios, se ve que éste es principalmente contenido como imágenes de software, películas y música, la cual es altamente demandada en mercados como el latinoamericano, pero que no está muy presente en mercados como el norteamericano o el europeo. Lo que significa que dicho contenido es parte de la realidad del operador y debe tomar medidas para controlarlo sin degradar la calidad de experiencia ofrecida al abonado.

4.3 Modelo Calidad de Experiencia

En primer lugar, basándose en la información obtenida del DPI y de las caracterizaciones ya descritas en los puntos anteriores, se definen una serie de aplicaciones que se simulan y a partir de ellas una serie de KPI que son medidos durante cada simulación tal como se pasa a describir a continuación:

Disponibilidad de servicio. Corresponde a la razón de conexiones exitosas sobre la base total, donde se entiende por una conexión exitosa cuando se logra establecer

⁸ Corresponde a descargas de archivos desde sitios especializados como rapidshare, megaupload o similares.

una conexión de datos y se marcan como fallidas por ejemplo cuando un usuario se encuentra sin cobertura del servicio o la red no es capaz de entregarle recursos.

Tiempo de Inicio de Sesión. Este indicador corresponde al tiempo que demora en asignar su dirección IP desde que se solicita crear el contexto de conexión.

Tasa de Transferencia. Corresponde a la cantidad de bits medios que son enviados/recibidos por unidad de tiempo (Kbps), es decir corresponde al throughput medio de una conexión ya sea en la descarga de datos de una página web, sitio ftp o servidor de video flash.

Retardo. Este indicador corresponde al RTT medido entre el cliente y el servidor destino.

Tiempo de Respuesta DNS. Corresponde al tiempo que tarda en recibirse la respuesta de los DNS.

Número de Interrupciones Video Flash. Corresponde al número de veces que la reproducción de un video flash se ve interrumpida debido a la falta de información en el buffer de la aplicación produciéndose la interrupción de la reproducción del video.

Las aplicaciones desarrolladas para la simulación fueron las siguientes:

DNS test. Corresponde a la resolución de diferentes nombres donde se encuentran los 25 nombres más resueltos por los DNS del operador.

PING test. Corresponde a la medición de retardo a los servidores más accedidos por los usuarios para los servicios más relevantes para los usuarios.

HTTP ping. Corresponde a la simulación de descarga de un sitio web simulando la conexión utilizando el o los browsers más populares en la red del operador.

FTP DL/UP. Corresponde a la descarga y subida de archivos de tamaño conocido desde un servidor dispuesto por el operador.

Flash Video. Corresponde a la simulación de descarga y reproducción de un video en formato flash video.

La secuencia de ejecución de la aplicación es la que se muestra en la siguiente figura.



Figura 4.57. Secuencia de ejecución de la aplicación de medición de calidad de experiencia

Finalmente cada una de las pruebas anteriores generó una serie de indicadores los que se pasan a mencionar en la siguiente tabla:

Aplicación	KPI	unidades
Inicio de Sesión	Disponibilidad del servicio	% disponibilidad
	Tiempo de inicio de sesión	milisegundos
DNS Test	Tiempo de Descarga	milisegundos
	Tasa de Descarga	kbps
HTTP Ping	Tasa de descarga	kbps
FTP DL/UL	Tasa de descarga/subida	kbps
Flash Video	Número de interrupciones	Veces
	Tiempo de interrupciones	Milisegundos
	Tasa de Descarga	kbps

Tabla 4.58. Tabla de KPI generados por la aplicación de medición de QoE

Utilizando los indicadores antes descritos, normalizándolos y ponderándolos en función de su importancia se logra un indicador de calidad de experiencia del servicio de banda ancha móvil.

$$QoE = \sum \alpha_i * KPI_i \quad (14)$$

Donde α_i corresponde a la constante de ponderación del indicador KPI_i . El valor que tiene esta constante se obtiene de acuerdo a la importancia que tenga el KPI en cuestión, el cual viene dado por el uso que tiene la aplicación aludida y la caracterización de los servicios de acuerdo a la Tabla 4.32.

De este modo, es posible obtener la medición del indicador de calidad de experiencia que refleja el estado actual del servicio, tal como lo percibiría el usuario, como se puede ver en la Figura 4.59

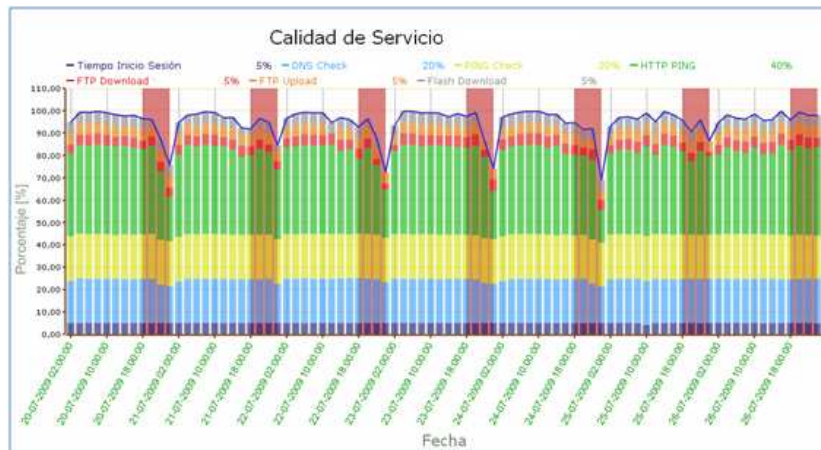


Figura 4.59. Indicador Calidad de Experiencia del servicio BAM

El proceso de definición de las constantes de ponderación debe permitir que el indicador refleje la experiencia del usuario. Por ejemplo en la Figura 4.59 se puede ver que éste decae en la zona roja, la que corresponde al rango horario de más tráfico en la red, donde efectivamente la experiencia del usuario se ve disminuida.

4.4 Modelo de Costos

Tanto el modelo de costos de Ericsson[18] [19] como el de Nokia Siemens Networks (NSN) [20] ofrecen al operador la oportunidad de racionalizar la estrategia de inversión y gastos del servicio. Si bien ambos concuerdan con que el costo del servicio tiende a 1€/GB es importante revisar en detalle la metodología y los supuestos de modo tal que se pueda contextualizar adecuadamente al mercado local. Por otro lado, lo importante no es sólo tener un costo bajo y en descenso, sino un costo por usuario bajo y acotado que permita tener los márgenes de utilidades esperados por el operador.

Los costos de red además dependen de las particularidades de cada región, como el área de cobertura que es necesario cubrir, la geografía, las facilidades tecnológicas, los costos de terrenos, etc.

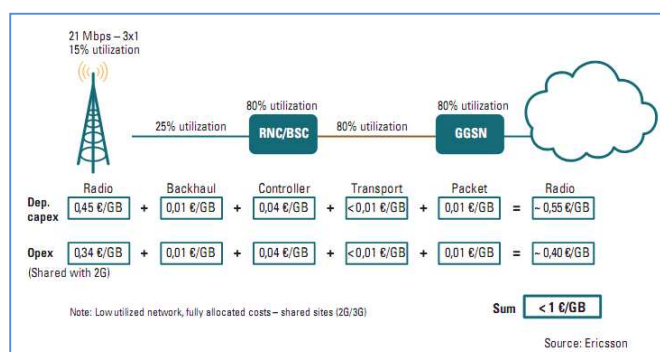


Figura 4.60. Modelo de costos.

Al analizar los costos que los operadores móviles tienen en Chile para brindar el servicio, se puede ver que este valor está bastante alejado del valor esperado en otros mercados. Esto se debe principalmente a que los estudios hablan de una red madura, y cuando se refieren a este concepto se están refiriendo específicamente a una red con una penetración del orden del 25%, o como se muestra en el caso del estudio de NSN, un mercado con 40% y un uso del servicio de 2 GB/mes por suscriptor. Efectivamente en este caso si existe un costo de red menor al 1€/GB.

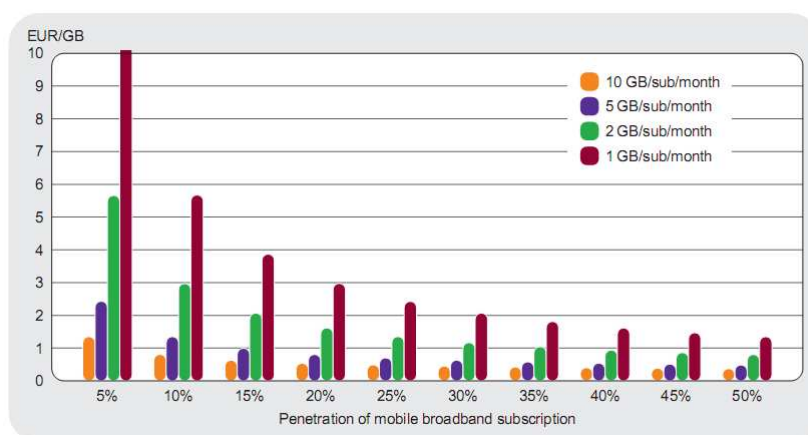


Figura 4.61. Costos de Red por gigabyte de datos.

En el caso de Latinoamérica, se tiene que la penetración del servicio es menor al 5% en promedio y el consumo es del orden de 3-4GB/mes lo que indica que el costo de red no es menor a 6€/GB, es decir, el costo por usuario es de 18 €/mes lo que es muy cercano al ARPU de un cliente BAM, dejando márgenes de rentabilidad muy estrechos para el operador del servicio de datos.

Otro punto interesante de analizar al momento de comparar los costos, es la metodología utilizada para asignar los montos de inversiones, esto es cuando un operador construye un nuevo sitio con tecnologías 2G y 3G para los servicios básicos de Voz y Datos, debe decidir cómo asigna los costos a cada uno de los servicios básicos en cada una de las tecnologías. Si bien un método corresponde a asignar proporcionalmente las inversiones de acuerdo al nivel de capacidad disponible para cada servicio, también es válido utilizar un criterio de marginalidad, es decir, si la inversión se realiza si o si para un servicio específico los otros tendrán un costo marginal cero para ese ítem. Esta consideración es importante ya que si no se unifican y se informan claramente los criterios utilizados, se puede llegar a conclusiones erróneas al momento de comparar eficiencias en la inversión.

Por otro lado, aunque normalmente los estudios muestran los costos de red en función de los bytes traficados, lo correcto es que estos dependen directamente de la intensidad de uso, es decir al throughput que cursan los usuarios en la hora cargada y no a los bytes estos puedan traficar en el mes, de este modo se podrían tener dos usuarios, uno con el doble de volumen transmitido, pero a ambos con el mismo

throughput en la hora carga, luego el costo en ambos casos es el mismo, pese a que uno traficó el doble que el otro.

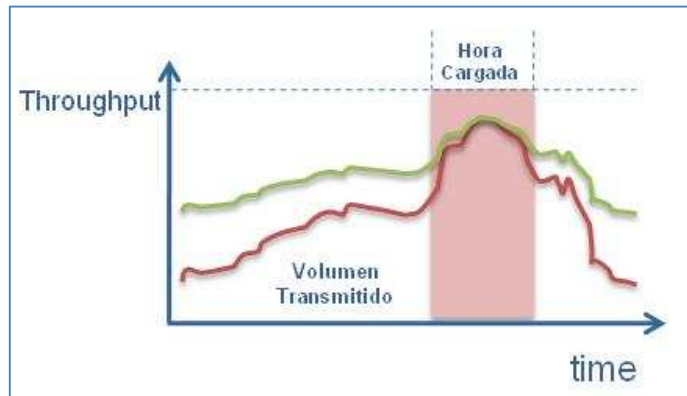


Figura 4.62. El usuario de la línea verde trafica más volumen de datos pero tiene el mismo costo de red que el usuario de la línea roja.

Como una analogía al modelo de voz se acostumbra a estimar el costo de red en función del Byte transmitido, pero es recomendable sólo en forma paralela al análisis del costo como función del flujo de datos (Kbps) que aportan los usuarios en la hora cargada, que corresponde a una mirada real al problema.

A continuación es importante realizar un análisis de tallado de cuáles son los factores que influyen en cada uno de los puntos descritos anteriormente y cómo es posible atacar cada uno de ellos.

En el estudio de Ericsson, se puede ver que ellos distribuyen los costos de red de la siguiente forma:

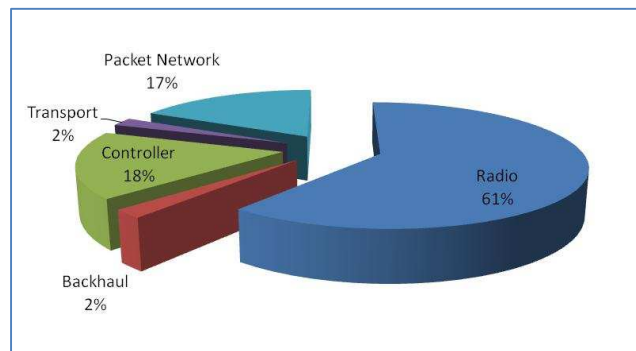


Figura 4.63. Distribución de costos de red.

Cada uno de estos ítems puede variar dependiendo de la realidad de cada país. En el caso de Chile, los costos de *backhaul* y de transporte son claramente mayores debido a la geografía y a la distribución poblacional que éste presenta.

Por otro lado, costos como los de radio pueden aumentar notablemente dependiendo de las exigencias que el regulador imponga, en términos de cobertura geográfica exigida y en función de los niveles de penetración que el país tenga. Esto puede generar que una red tenga una gran cantidad de nodos con baja utilización y sólo unos poco utilizados al máximo. Lo que no ocurre cuando se despliega una red estrictamente en base a criterios de la demanda de tráfico.

Esta capacidad ociosa se utiliza a medida que aumenta la penetración del servicio haciendo la inversión cada vez más eficiente.

Otro punto importante que hay que destacar de la particularidad de cada despliegue de red, es el hecho que los usuarios no se distribuyen uniformemente dentro de las diferentes celdas de un sitio, lo que afecta su desempeño, ya sea por la forma en que se distribuyen los clientes o por la presencia de edificios o instituciones que tienden a densificar una zona, tal como se puede ver en la siguiente figura.

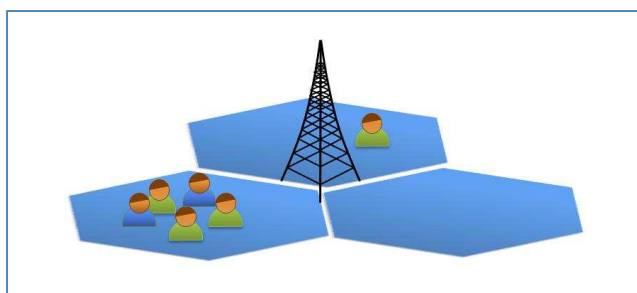


Figura 4.64. Distribución de usuarios entre las celdas de un sitio.

En resumen, los costos de radio dependerán del factor de utilización que poseen las celdas, como función de la densidad poblacional que podrán atender y de la distribución que estos tengan entre los diferentes sectores.

Finalmente los niveles de volumen que se puedan alcanzar también afectan en los costos, ya que claramente las economías de escalas son aplicables y no se puede comparar una red para unos cientos de miles de abonados con una red de varias decenas de millones.

Una vez que se tienen identificados los costos que influyen en la red, se debe proseguir al siguiente paso que consiste en comenzar con la identificación de los componentes de segundo orden que afectan cada uno de los ítems mencionados anteriormente.

Por ejemplo, como mencionaba Silva [21] los costos de servicios asociados al despliegue de un nuevo sitio se dividen en trabajos de construcción de sitios, costos de adquisición, implementación y otros ítems, como se muestra a continuación.

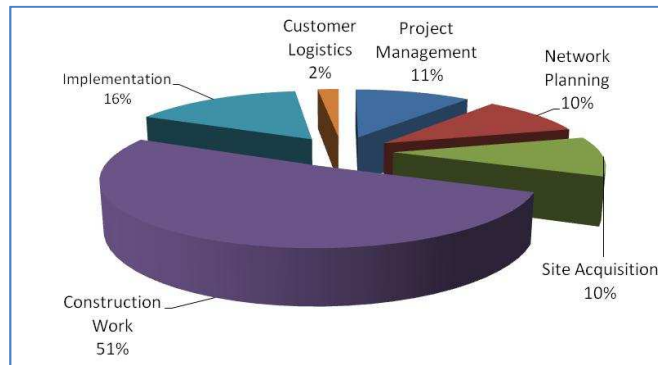


Figura 4.65. Costos de servicios de un nuevo sitio.

En este caso, se presenta una alternativa para disminuir los costos de planificación de red a través de la utilización de SON (Self Organizing Networks)⁹ reduciendo el CAPEX y luego también disminuyendo el OPEX de la operación de la red.

Tal como muestra esta iniciativa, existen diversas otras que tienen como objetivo reducir los costos, las cuales deben ser estudiadas detalladamente e ir adoptándolas rápidamente de modo tal que se pueda disminuir el costo del servicios haciendo que éste sea más rentable para el operador móvil y luego sea más fácil su masificación lo que claramente favorece al cliente con la disminución de los precios del servicio.

El tamaño de la red influye en la calidad de experiencia que se puede brindar a los usuarios, pero es importante concentrarse en los distintos factores y elementos que influyen en los costos de red, de este modo, es posible determinar donde es preferible invertir maximizando la experiencia del usuario.

⁹ SON es parte de la especificación 3GPP release 8 incluido en el TS 36.902 [17]

Capítulo 5. Discusión

En este capítulo se discuten los resultados obtenidos en el capítulo anterior, son interpretados y puestos en el contexto analizando el servicio de banda ancha móvil. En él se revisa la caracterización de los usuarios del servicio BAM en lo que respecta a su movilidad, concentración horaria y geográfica, la intensidad de uso y las aplicaciones y servicios que estos utilizan.

5.1 Caracterización de usuarios BAM

El primer paso para iniciar la discusión de la calidad de servicio en la banda ancha o en cualquier servicio, es el conocer en detalle el comportamiento que los usuarios tienen. Para ello fue vital el uso de plataforma de DPI, ya que ella proporciona una imagen de los servicios más comunes utilizados por los usuarios de BAM. En conjunto con el análisis de protocolos que utiliza cada aplicación, permite determinar cuál es el comportamiento de cada uno de los servicios y su vulnerabilidad a fluctuaciones en el ancho de banda disponible.

5.1.1 Movilidad

Los usuarios BAM son principalmente fijos o nómades, luego es importante considerar que la movilidad es considerablemente menor en comparación de los servicios de navegación en los móviles. Los principales usuarios ven al servicio como un sustituto de la banda ancha fija y no necesariamente aprecian el atributo de movilidad. Considerando esto, es importante que los operadores analicen las concentraciones geográficas de los usuarios para implementar soluciones enfocadas a solucionar los problemas de densidad de uso en sitios particulares y además deben considerar factores de densificación, como edificios residenciales y oficinas, al momento de diseñar las redes.

5.1.2 Concentración horaria

El costo de un abonado depende directamente del uso que éste realice en la hora cargada, es por ello que es importante definir políticas de incentivo de uso en las horas valle de modo tal que el tráfico total se distribuya homogéneamente durante el día y se utilice la infraestructura durante los momento que hay más holgura haciendo más eficiente la inversión.



Figura 5.66. Indicador Calidad de Experiencia del servicio BAM

Esta práctica de aplanamiento del tráfico permite que los usuarios puedan traficar volúmenes altos de datos a un precio menor, lo que genera una mejor percepción del servicio a bajo costo.

5.1.3 Usuarios Intensivos.

El primer paradigma del servicio BAM es que los usuarios intensivos son los usuarios que más descargan. Este paradigma es roto cuando se demuestra que los usuarios más intensivos son quienes trafican en forma más intensiva en hora cargada y no necesariamente quien descargar mayor volumen en un mes.

Cuando se determina que sólo el 25% de los usuarios que trafica más volumen también son parte del grupo de los más intensivos en la hora cargada, se puede ver que un gran número de usuarios que descargan grandes volúmenes de contenido en la hora valle. Estos abonados tienen una alta tasa de satisfacción usando los recursos en la hora en que están disponibles, es decir tienen un bajo costo para el operador, incluso menor que el de un usuario normal.

Este resultado obliga a cambiar la mirada y enfocarla en la intensidad de uso. El estudio además muestra que usando un criterio mixto que consiste en clasificar de *heavy user* a quienes consuman más de un cierto volumen en la última hora, se puede acceder a una categorización más factible técnicamente y así lograr aplicar políticas de control a quienes realmente afecten el servicio.

5.1.4 Concentración geográfica

Los usuarios se encuentran distribuidos geográficamente y la densidad de uso tiene directa relación con la densidad poblacional del territorio. Debido a las obligaciones regulatorias existe un amplio despliegue de las redes que no obedece estrictamente a criterios económicos locales, por lo tanto existe una gran capacidad ociosa en las redes que sólo son ocupadas cuando comienza a crecer la penetración del servicio, haciendo que éste sea más rentable.

5.1.5 Aplicaciones y servicios.

Las principales aplicaciones y servicios que usan los abonados, corresponde principalmente a navegación web, streaming y peer to peer. Esto muestra la clara preferencia de los usuarios por el contenido interactivo de la red. Pese a esto, aún existe una gran proporción de contenido no interactivo y ante una situación de congestión, donde normalmente se ve afectado todo el tráfico del abonado, es posible priorizar el contenido interactivo por sobre el no interactivo, de modo tal, que la degradación del servicio se enfatice en aplicaciones que el usuario no perciba o que afecten menos su experiencia y así lograr una degradación que maximice la experiencia.

Es importante considerar que los equipos de inspección de paquetes requieren una constante actualización, ya que continuamente se desarrollan nuevas aplicaciones y, por otro lado, la preferencia por los servicios varía en el tiempo, teniendo importantes explosiones como la de los videos en línea y de los servicios de redes sociales.

La evolución de las aplicaciones genera un usuario cada vez más demandante, con un perfil de tráfico que va en aumento debido a la aparición de nuevas aplicaciones que consumen mayor cantidad de recursos. Este aumento pone en riesgo la rentabilidad del servicio. Si bien los costos van a la baja, lo que hay que procurar es el costo por abonado, que ante este aumento de perfil podría enviar a los proveedores de servicio fuera del umbral de rentabilidad.

Finalmente, los últimos estudios [19] muestran que los nuevos usuarios que ingresan son menos demandantes que los que ya se encuentran en el operador, debido a que han encontrado los beneficios del servicio en forma tardía al no necesitar tan intensamente el servicio.

5.1.6 Políticas de Fair Use

La aplicación de políticas de *fair use* puede controlar el uso desmedido que realizan los heavy users y que afectan la calidad de experiencia del resto de los abonados del servicio. Estas medidas pueden ser muy eficaces como se pudo ver en los capítulos anteriores, pero si no se realiza una correcta implementación genera una deficiente percepción del servicio y estas medidas resultan ser inútiles. En este sentido es importante contemplar al menos los siguientes puntos [23] :

Límites o umbrales. Los umbrales deben ser correctamente definidos, asignados y administrados, su uso debe poder ser monitoreado sin que este proceso interfiera la experiencia del usuario.

Seguimiento de uso. Es esencial la transparencia para la aplicación y aceptación de las políticas de *fair use* por parte del cliente. Permitir a los usuarios determinar cuánto han consumido genera confianza, proporcionar información actualizada sobre el uso de datos, mediante el uso de alertas de notificación automática o consulta de saldo en línea es esencial para el buen funcionamiento del *fair use*.

Notificaciones. Un componente importante es la capacidad de enviar notificaciones de acuerdo con el uso del plan adquirido. Con el fin de comunicarse eficazmente y hacer cumplir los límites de uso, los operadores deben ser capaces de enviar notificaciones personalizadas a sus abonados, indicando cuando se alcanza un umbral o que se está a punto de alcanzarlo, opciones para incrementarlo temporalmente o realizando una modificación en el plan y finalmente informando cuando se están aplicando cargos extras.

Ejecución. Es importante aplicar las políticas definidas en las condiciones de *fair use* una vez que los límites son alcanzados, por el periodo estipulado, ya sea horas, días o hasta el fin del ciclo de facturación. Una vez que se termina el periodo de restricción, el usuario debe ser vuelto a su condición inicial

En las primeras aplicaciones de políticas de *fair use* se pudo ver que los operadores no siempre cumplían con una correcta notificación y no siempre hacían una efectiva ejecución de la política. Esto generó que los abonados no comprendieran las reglas del servicio y lo percibieran como un servicio de poca calidad. Esta situación ha sido corregida en algunos operadores sin que exista una notificación y ejecución 100% eficiente de parte de todos los operadores.

Si un usuario no es notificado de la ejecución de la política de *fair use*, nota una degradación del servicio y lo asocia a las condiciones de la red, interpretando que su operador ofrece una mala calidad y no a que se encuentra en esa condición debido a que superó los límites de su plan. Por otro lado, si el abonado es notificado y no se le aplica la política creará que la degradación no es tal y será un incentivo para aumentar su uso promedio, transformando al usuario en un usuario potencialmente intensivo.

5.2 Desempeño de Aplicaciones

Uno de los puntos interesantes que se estudia, corresponde a la relación que existe entre el throughput, RWND y el RTT, dada en la ecuación 7.

Se puede obtener mejoras tan sólo habilitando *window scale* en el computador del usuario que superan el 20% de aumento de throughput. Pero esta solución, si bien es importante, comienza a disminuir su relevancia ya que hoy se encuentra habilitada en sistemas operativos como Windows 7.

La relación inversamente proporcional entre el RTT y el throughput tiene un alto potencial, ya que como mostró en el capítulo anterior, se tiene que una mejora del 50% en el retardo se traduce en un aumento del 100% en el throughput esperado, esto significa que el retardo es una de las variables que más afectan al desempeño de las aplicaciones y por ende a la percepción que el usuario puede tener del servicio. Producto de ésto, técnicas como *cache* y *peering* tienen un alto impacto y es importante incorporarlas dentro de las tareas de mejoras que puede realizar el operador de red.

Otro punto importante de destacar es la importancia que tienen las aplicaciones, si bien éstas son de responsabilidad del proveedor de contenido, afectan directamente

la percepción que el usuario tiene del servicio global y normalmente culpará a su proveedor de red por todos los problemas que pudiera tener, sin importar si estos son efectivamente del operador o del proveedor de contenido.

5.3 Calidad de Experiencia

El indicador de calidad de experiencia definido en este trabajo se encuentra altamente correlacionado con la percepción del servicio de banda ancha móvil, es importante destacar que la metodología elegida y las consideraciones adoptadas permite que el operador pueda realizar correcciones en los ponderadores o pesos asociadas a los indicadores unitarios, de modo tal que pueda adaptarse a los cambios de comportamiento y aparición de nuevas aplicaciones que puedan modificar el comportamiento de los usuarios.

La calidad de experiencia entregada es producto de la calidad de la red y de los servicios brindados, es por ello que los niveles de congestión elevados afectarán negativamente a la percepción del servicio. En ese sentido es importante contener la congestión priorizando los servicios interactivos por sobre los no interactivos de modo tal que la degradación de la conexión se traduzca en la degradación de servicios que el usuario no perciba y que no afecten mayor mente su experiencia.

En la misma línea es importante tomar acciones de mitigación sobre los *heavy user*, que siendo menos de un 5% pueden traficar más del 40%. Estas acciones pueden ser: la adopción de políticas de *fair use* efectivas o la oferta comercial acordes a sus necesidades, evitando así que degraden el servicio de la gran mayoría de los abonados.

Otro punto que no debe dejarse de lado es el factor de costo del servicio, ya que claramente una mejora en la calidad también pasa por invertir más en la red, lo que debe ajustarse a los niveles de ARPU que tenga el servicio. Si el operador sobre invierte en el servicio y no es eficiente tendrá un servicio con un alto costo y se verá imposibilitado de realizar mejoras para aumentar la calidad de experiencia del servicio.

La diferenciación por calidad de servicio aporta claros beneficios y los operadores están comenzando a incorporarlo. En ese sentido es importante considerar una serie de puntos al momento de planificar la red.

- *Garantizar QoS elevada a los servicios clave.* Priorizando servicios sensibles.
- *Aplicar diferenciación a los usuarios.* Permitiendo diferenciar planes comerciales.
- *Limitación a usuarios y aplicaciones no deseadas.* Control de Heavy Users y aplicaciones intensivas en ambientes de congestión.
- *Considerar el tratamiento especial a servicios de emergencia.*

- *Buscar el uso eficiente de los recursos.* Considerando inversiones costo eficientes haciendo más rentable el servicio.

Finalmente, el modelo de calidad de experiencia realizado permite tener un indicador de la calidad de experiencia en tiempo real. De este modo, se pueden tomar acciones correctivas ante la degradación del servicio. Adicionalmente, el modelo permite predecir la calidad de experiencia percibida ante variaciones de los factores que influyen en ella. Por ejemplo, se puede estimar cuanto mejora la calidad al disminuir retardos, ya sea utilizando técnicas de *cache* o de *peering*

5.4 Análisis de Costos

El análisis de costos es uno de los puntos más importantes para el operador móvil, primero porque lo ayudará a identificar cuáles son los principales factores que influyen en él, dónde debe enfocarse al momento de buscar eficiencias y finalmente permite compararse con otros operadores para así definir su nivel objetivo para los siguientes años.

Al realizar el análisis de costos, se puede ver que uno de los puntos más relevantes es la inversión de acceso, debido al despliegue de infraestructura necesario para dar cobertura ya sea por demanda comercial o por obligaciones con el regulador. Esta inversión en cobertura será desplazada por inversión en capacidad a medida que madura el mercado y los tráficos aumentan. Teniendo así una inversión más eficiente que permite mejorar el nivel de costos.

Los niveles de inversión y costos dependen de cada operador y de las características propias de cada país por lo que es importante contextualizar los análisis al momento de pretender compararse con otros operadores. Ésto permite hacer un correcto benchmark y tomar decisiones acertadas. En esta línea, es necesario considerar cuál es el nivel de penetración que tiene cada país, ya que el aumento de usuarios permite un uso más eficiente de las redes desplegadas y, por lo tanto, un mejor uso de la inversión. En este caso Chile pasa de una penetración cercana al 4% para finales del 2009, a una mayor al 8% para fines del 2010, esto indica que los costos podrían verse disminuidos en un 50% si se consideran los resultados del estudio de NSN mencionado anteriormente.

Otro de los puntos importantes es el desafío que tienen los operadores de fomentar el uso del servicio en las horas valle, esto permitirá utilizar infraestructura cuando exista capacidad disponible con mejor satisfacción lo que a su vez permitirá al usuario priorizar ese horario para traficar descongestionando la hora cargada.

No hay que dejar de lado el dilema volumen-intensidad cuando se realiza el análisis de costos, esto es, lo que realmente le cuesta al operador es la intensidad de uso o el throughput en la hora cargada y no el volumen que descarguen los abonados, este punto llevado al extremos de los *heavy users* muestra que usuarios que consumen altos volúmenes de contenido, no necesariamente representan un alto costo para el operador, siempre y cuando su tráfico lo realicen en las horas valles y no en la hora cargada.

El fenómeno de los *heavy user* no sólo significa un aumento en los costos para el operador, sino también generan una degradación en la experiencia de los usuarios ya que al usar excesivamente los recursos disponibles no dejan capacidad para los usuarios normales y lo que es peor para los usuarios de alto ARPU.

Hay que recordar la importancia de realizar un análisis de uso geográfico del servicio, ya que, éste dejó de ser un servicio móvil y hoy se destaca por tener características estáticas o máximo nomádicas, luego es necesario tomar las consideraciones del caso al realizar la planificación de la red y más aun proponer soluciones de *off-loading* que pretendan descongestionar la red utilizando alternativas de bajo costo.

Finalmente, el análisis de costo no solo pretende disminuir las inversiones y gastos de los operadores, sino también busca que éstas sean eficientes y se puedan concentrar donde mejoren la experiencia.

5.5 Mejoras

El servicio se puede ver afectado por factores que el usuario o el operador puede manejar, mejorando su experiencia. En esta línea, y a modo de resumen, se propone una serie de medidas que permite mejorar la experiencia del usuario.

- a) Usuarios
 - i. RCF 1323. La utilización de esta recomendación en los sistemas operativos de los computadores de los usuarios, permite mejorar el orden de un 20% el desempeño de las conexiones TCP, mejorando considerablemente la experiencia. Para llevar a cabo esto es necesario habilitar la opción para los usuarios que utilizan Windows XP y ya que para el caso de Windows 7 y Linux, ésta viene habilitada por defecto.
 - ii. Modem más modernos. La incorporación de módems más modernos que permitan un mejor desempeño de acuerdo a la evolución y requerimiento de los usuarios. En este caso, se recomienda el uso de módems con Rx *diversity* y con soporte de HSPA+ o DC-HSPA+, dependiendo de la disponibilidades del operador.
- b) Evolución de Red. La evolución de los *release* de la 3GPP para la tecnología WCDMA y para LTE no solo permite incorporar mejoras en el throughput, sino también mejoras en el retardo. Es por ello que es importante incorporar mejoras como:
 - i. One Tunnel. Permite una comunicación directa entre NodoB y SGSN solo pasando por RNC la señalización.
 - ii. ARP y THP. Permite incorporar distintos niveles de priorización y calidad de servicio de cara al cliente, pudiendo diferenciar los distintos planes comerciales en función de ARPU.

- iii. 64QAM. Esta modulación permite lograr tasas reales de 16Mbps en los usuarios (21 Mbps teóricos).
 - iv. CPC¹⁰. Permite mejoras en la señalización, disminuyendo el retardo, junto con la carga de los RNC y el uso de batería en los dispositivos móviles.
 - v. DC-HSPA. Permite un uso más eficiente del espectro usando dos carrier simultáneamente permite llegar a tasas reales de 32Mbps (42 Mbps teóricos).
 - vi. DB-HSPA. Permite el uso de bandas distintas para hacer *dual carrier* o *multiple carrier*, habilitando el uso de bandas discontinuas.
- c) Mejoras en la Infraestructura de Red. Estas permitirán tener una red más robusta y con menos latencia.
- i. Menos satélite. Disminuir el uso de satélite buscando alternativas rentables que además permita mejorar retardos.
 - ii. Menos Micro ondas, más fibra. Permite alcanzar mejores tasas de transmisión necesarias para redes de cuarta generación.
- d) Aplicaciones. Incorporación de Aplicaciones que permita mejorar desempeños.
- i. *Network Optimisation*. Permite la optimización de protocolos y compresión de tráfico donde es permitido, mejorando la experiencia del usuario donde los desarrolladores de contenido no han procurado optimizarlo.
 - ii. *Cache Server*. Incorporando este tipo de tecnología permite “acercar” el contenido a los usuarios ahorrando los saltos internacionales y reduciendo el retardo en las transmisiones.
 - iii. *Peering*. Consiste en la interconexión directa a los proveedores de contenido, ésta técnica permite disminuir los tiempos de transmisión al contenido ahorrándose parte de los retardos que agrega Internet.
 - iv. *Housing/Hosting*. Al alojar localmente los contenidos, se disminuyen los retardos que agrega el segmento de red hasta el proveedor de contenido, habitualmente internacional, permitiendo además utilizar conexiones locales de alta velocidad que aceleran la entrega del contenido y que además tienen un costo mucho menor.

¹⁰ CPC (continuous packet connectivity) es parte del 3GPP Rel 7 TR 25.903

Capítulo 6. Conclusiones

El servicio de banda ancha móvil es un servicio en constante maduración, éste se encuentra continuamente cambiando sus características invitando al operador a mantenerse investigando sobre su funcionamiento e innovando para mejorar el servicio que ofrece a los exigentes clientes.

La calidad de experiencia del servicio de banda ancha móvil es una característica muy apreciada y que los usuarios están aprendiendo a distinguir, es por ello que los operadores no deben dejarla de lado y para desarrollarla y planificarla deben tener en cuenta una serie de puntos que se describen a continuación:

Planes de datos con QoE diferenciados. Permite que los usuarios que lo requieran puedan optar a pagar más por planes con mejor calidad de experiencia y mejorar el ARPU del servicio, junto con la experiencia global de éste.

Incentivar uso del servicio en horas valle. De este modo se logra que los usuarios utilicen el servicio en los horarios en que existe capacidad ociosa, maximizando la eficiencia de los recursos disponibles.

Políticas de *fair use* robustas. Con límites correctamente definidos, control de uso que permita al operador y abonado saber en cada momento su consumo, notificación de umbrales que permita saber cuándo se alcanzan los límites de su plan y políticas bien ejecutadas, que reflejen las políticas definidas y que diferencien claramente los estados en que se encuentre el abonado. Todas estas características permiten a los usuarios tener claridad de las condiciones de su servicio teniendo así la claridad de que si permanentemente alcanza los límites de *fair use* es muy probable que necesite mejorar su plan comercial.

Intelligent fair use. Una política interesante, que mejora bastante la experiencia de los usuarios y permite un uso más eficiente de los recursos de radio, pero que puede ser contraproducente, ya que bajo esta política, los usuarios podrían no distinguir la degradación del servicio incentivándolo a traficar más de lo normal, causando el efecto contrario.

Control de heavy users. Utilizando una correcta categorización de los usuarios es importante determinar quienes utilizan el servicio en forma abusiva degradando la calidad del resto de los usuarios, de esta forma y aplicando políticas claramente informadas se puede hacer control de estos usuarios y mejorar la percepción de los abonados, tanto heavy user, como usuarios afectados por el uso descontrolado de estos abonados.

Análisis y control de costos. Realizando un análisis de costos del servicio se puede lograr determinar qué ítems afectan mayormente al servicio y mejor aún comparar resultados con otros operadores de modo tal que sea posible aplicar políticas de control de costos coherentes y siempre con la visión puesta en mejorar la calidad de

experiencia del usuario, disminuyendo gastos innecesarios e incentivando los que la mejoren.

Calidad de Experiencia. Es importante utilizar un mecanismo robusto que permita disponer de un indicador de calidad de experiencia que refleje los requerimientos de los usuarios y que muestre los estados de la red, considerando el servicio en toda su magnitud, es decir, como un servicio extremo a extremo donde cada uno de las partes puede afectar la calidad de experiencia.

Este último punto es el que define el tipo de servicio que brindará el operador, el que puede variar en cada región mostrando diferentes niveles, debido al uso que le dan los abonados.

El desafío consiste en conocer a los suscriptores y poder ofrecerles el servicio que ellos buscan.

En esa línea toma importancia la necesidad de implementar un indicador de calidad de experiencia, como el que se muestra en este trabajo. Ello permite no sólo determinar cual es el comportamiento del usuario y el nivel de calidad que se les entrega, sino también es una herramienta de predicción al momento de definir qué tipo de medidas es necesario tomar para mejorar el servicio. Por ejemplo, se puede modelar cuál es el impacto en la calidad de experiencia al aplicar mejoras del tipo *cache* o *peering* para un contenido específico o el impacto de priorizar un contenido durante episodios de congestión. También se pueden hacer correlaciones entre el indicador de calidad de experiencia y los números de reclamos que realizan los usuarios del servicio o ir más allá y correlacionar el indicador con la tasa de Churn

Este trabajo deja las puertas abiertas a nuevas investigaciones que permitirán conocer más cerca del servicio de banda ancha móvil. En particular se sugieren los siguientes temas:

Calidad de experiencia en redes de cuarta generación. Tecnologías como LTE ya se encuentran implementadas a lo largo del globo y es importante conocer cuales son las limitaciones que ésta tiene al momento de necesitar implementar calidades diferenciadas para distintos servicios como voz y video, esto permitirá definir los productos que se pueden comercializar y definir claramente qué es lo que se puede esperar de esta tecnología cuando se comience a implementar en Latinoamérica.

Calidad de experiencia en redes multiacceso convergentes. Cómo se pueden aplicar distintos tipos de calidades de servicios independientes al medio de acceso, cuales con las restricciones que imponen las tecnologías actuales y cómo la enfrentarán las nuevas tecnología. Responder estas interrogantes permitirá definir la evolución tecnológica que deben seguir los operadores, considerando sus diferentes realidades y los diferentes segmentos que quieran abordar.

Capítulo 7. Referencias Bibliográficas

- [1] Rysavy Research. "Transition to 4G. 3GPP Broadband Evolution to IMT-Advanced". [online]. 3G Americas. Septiembre 2010. Disponible en: http://www.4gamericas.org/documents/3G_Americas_RysavyResearch_HSPA-LTE_Advanced_FINALv1.pdf
- [2] CAIDA. "Analyzing UDP usage in Internet traffic". Mayo 2009. Disponible en: <http://www.caida.org/research/traffic-analysis/tcpudpratio/index.xml>
- [3] Velez, F.J y Correia, L.M. "Mobile broadband services: classification, characterization, and deployment scenarios.". Communications Magazine, IEEE. Volume 40, Issue 4, April 2002 Page(s):142 – 150
- [4] Ameigeiras, Pablo. "Packet Scheduling And Quality of Service in HSDPA ". PhD Thesis, Aalborg University. Aalborg Øst, Denmark, 2003.
- [5] Postel, Jon. "User Datagram Protocol". RFC 768, agosto, 1980. Disponible en: <http://www.faqs.org/rfcs/rfc768.html>
- [6] Postel, Jon. "Transmission Control Protocol". RFC 793, Septiembre, 1981. Disponible en: <http://www.faqs.org/rfcs/rfc793.html>.
- [7] Jacobson V., et al. "TCP Extensions for High Performance". RFC 1323, Mayo, 1992. Disponible en: <http://www.faqs.org/rfcs/rfc1323.html>.
- [8] Allman M., et al. Increasing TCP's Initial Window. Experimental Protocol RFC 2414, Septiembre, 1998. Disponible en: <http://www.faqs.org/rfcs/rfc2414.html>.
- [9] Allman M., et al. "TCP Congestion Control". Standard Track RFC 2414, Abril, 1999. <http://www.faqs.org/rfcs/rfc2581.html>.
- [10] Karn, Phil; Craig Partridge (11-13 Agosto), "Improving Round-Trip Time Estimates in Reliable Transport Protocols" (PS), *ACM SIGCOMM '87*, pp. 2-7, <http://www.ka9q.net/papers/rtt.ps.gz>, visitado el 15 Julio de 2009.
- [11] Casals, Andrés. "Desarrollo de Plataforma para Monitoreo de Servicio de BAM". Memoria de Título Ingeniero Civil Electricista/Profesor Guía Jorge Sandoval Arenas, Universidad de Chile. Santiago 2009.
- [12] ITU-T G.1030 Recommendation. "Quality of service and performance – Generic and userrelated aspects. Estimating end-to-end performance in IP networks for data applications". International Telecommunication Union, Telecommunication Standardization Sector. Noviembre 2005.

- [13] ITU-T E.419 Recommendation. "Business oriented key performance indicators for management of networks and services". International Telecommunication Union, Telecommunication Standardization Sector. Febrero 2006.
- [14] ITU-T E.800 Recommendation. "Definitions of terms related to quality of service". International Telecommunication Union, Telecommunication Standardization Sector. Septiembre 2008.
- [15] ITU-T I.211 Recommendation. "B-ISDN Service Aspects". International Telecommunication Union, Telecommunication Standardization Sector. Marzo 1993.
- [16] Hågstran, Karlsson y Silverplats. "Proof of Concept Implementation of UMTS Long Term Evolution". MSc Programmes in Engineering Thesis. Luleå University of Technology. Sweden 2007.
- [17] 3GPP. TS 36 Series. [online]. Disponible en: http://www.3gpp.org/ftp/Specs/2009-09/Rel-8/36_series/, visitado 6 enero 2011.
- [18] Blennerud, Greger. "Don't worry – Mobile broadband is profitable". [online], Ericsson Business Review N°2 2009, pp.54-48. Disponible en: http://www.ericsson.com/ericsson/corpinfo/publications/ericsson_business_review/pdf/209/209_BUSINESS_CASE_mobile_broadband.pdf, visitado el 11 de Mayo de 2010.
- [19] Blennerud, Greger. "Mobile broadband – busting the myth of the scissor effect". [online], Ericsson Business Review N°2 2010, pp. 50 -55, Disponible en: http://www.ericsson.com/ericsson/corpinfo/publications/ericsson_business_review/pdf/210/210_strategy_mobile_broadband.pdf, visitado el 1 de Noviembre de 2010.
- [20] Nokia Siemens Network. "Mobile broadband with HSPA and LTE – capacity and cost aspects". [online], White Paper Nokia Siemens Network, Disponible en: http://www.nokiasiemensnetworks.com/sites/default/files/document/Mobile_broadband_A4_26041.pdf, visitado el 22 de Mayo de 2010.
- [21] Silva, Ricardo. (2010, 27-28 abril) "Reducing the Cost Per Bit of LTE with Self Organizing Networks (SON)", Congreso LTE LATAM 2010, Rio de Janeiro, Brasil. Abril 2010.
- [22] Sandoval, Jorge (2010, 27-28 abril) "Conducting LTE Trials in Latin America. What should operators ensure that an LTE Trial Covers?", Congreso LTE LATAM 2010, Rio de Janeiro, Brasil. Abril 2010.
- [23] OPENET Telecom. "How To Guide Fair Usage", Disponible en: http://img.en25.com/Web/Openet/HTG_Convergent_Fair_Usage_A4.pdf, visitado el 4 de Octubre de 2010.

- [24] Rumney, Moray ed. "LTE and the Evolution to 4G wireless". Agilent Technologies publication. China 2009.
- [25] Omerovic, Sanida. "WiMax overview". Faculty of Electrical Engineering, University of Ljubljana, Slovenia.
- [26] Nokia Siemens Network. "¿Entregar servicios universales de datos móviles a bajo costo? ¡Es posible!". [online]. Unite Latinoamérica, Ejemplar 04 2010, pp 4-8. Disponible en: http://lat.nokiasiemensnetworks.com/sites/default/files/document/UniteLat_04_S PANISH_FINAL.pdf, visitado 10 diciembre 2010.
- [27] Nokia Siemens Network. "La diferenciación en la Calidad del Servicio". Unite Latinoamérica, Ejemplar 04 2010, pp 9-13. Disponible en: http://lat.nokiasiemensnetworks.com/sites/default/files/document/UniteLat_04_S PANISH_FINAL.pdf, visitado 10 diciembre 2010.

Apéndice A. Arquitectura de Tecnologías Móviles

En este capítulo se describen las arquitecturas de las distintas tecnologías móviles que permiten ofrecer el servicio BAM.

A.1 GSM/GPRS

La infraestructura de una red móvil de segunda generación como la red GSM se divide en tres partes: la red de radio, la red de conmutación de circuitos y la red de conmutación de paquetes, identificándose los siguientes elementos esenciales en cada una de ellas.

Base Station Subsystems

La red de radio o Base Station Subsystems (BSS) comprende el Base Station Controller (BSC) y la Base Transceiver Station/Base Station (BTS/BS).

La BTS corresponde a la radio base propiamente tal, provee los canales de radio para señalización y tráfico de datos de usuarios en las celdas, además de algunas pocas funciones de señalización y procesamiento de protocolos.

Por otro lado, los BSCs se encargan de la administración de frecuencias, control de las BTSs y de funciones de movilidad.

Mobile Switching Network

El Mobile Switching Subsystem (MSS) incluye Mobile Switching Centres (MSCs) y bases de datos de abonados que almacenan información requerida para el ruteo y provisión de servicios.

El nodo encargado de la conmutación de es llamado Mobile Switching Centre (MSC). Una red puede tener varios MSCs los cuales gestionan las llamadas de un grupo de BSCs.

El Home Location Registers (HLR) almacena la identidad y los datos de usuario de todos los abonados, las claves de autenticación, los servicios suplementarios permitidos e informaciones temporales necesarias para el funcionamiento de la red.

Packet Switching Network

Los servicios GPRS (General Packet Radio Service) y EDGE (Enhanced Data Rate for GSM Evolution) son entregados a través del uso de una red paralela compuesta por los siguientes nodos: PCU, SGSN y GGSN.

LA PCU (Packet Control Unit) corresponde a un dispositivo instalado en la BSC que permite separar el flujo de paquetes y enviarlo al SGSN (Serving GPRS Support Node) que se encarga de administrar la señalización y la movilidad del abonado y envía el flujo de datos al GGSN (Gateway GPRS Support Node) que a su vez lo dispondrá en la red de datos que le brinda el servicio que el cliente ha especificado en su conexión.

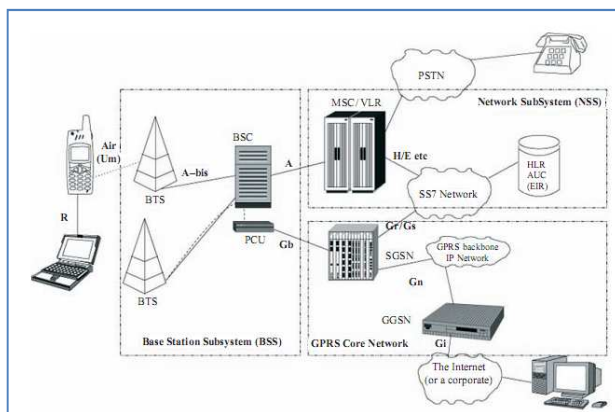


Figura A.67. Arquitectura GSM/GPRS

A.2 Arquitectura UMTS

La Arquitectura UMTS es muy similar a la red 2G distinguiéndose la red de acceso de radio, red de voz y red de paquetes:

Radio Access Network

Los componentes de la Radio Access Network (RAN) comprende el Radio Network Controller (RNC) y el Nodo B.

El nodo B, al igual que la BTS en GSM, corresponde a la radio base propiamente tal, provee los canales de radio para señalización y tráfico de datos de usuarios en las celdas, además de funciones de señalización, control de potencia, codificación de canales y procesamiento de protocolos.

Por otro lado los RNC se encargan de la administración y control de recursos de radio, control de potencia, asignación de canales, cifrado, etc.

Circuit Switch and Packet Switch Network

El núcleo de circuitos y paquetes es compartido con la red GSM como puede verse en la siguiente figura, manteniendo las funcionalidades descritas anteriormente.

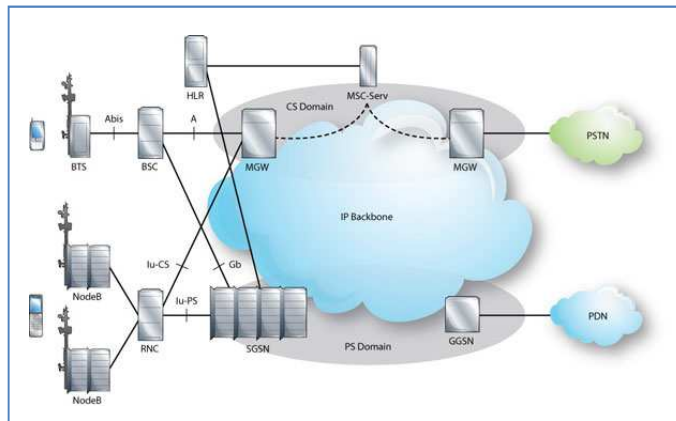


Figura A.68. Arquitectura 2G y 3G

Esta arquitectura, permite evolucionar hacia una red de mayor capacidad manteniendo el núcleo de la red y realizando cambios en la red de acceso. Estos cambios permiten además la compatibilidad y operación entre ambas redes, de manera que un usuario inicie una llamada en la red UMTS y la pueda terminar en la red GSM, permitiendo así desplegarla sólo donde sea necesaria asegurando cobertura a los nuevos usuarios pero con las limitaciones de capacidad y servicios de la red 2G.

A.3 Long Term Evolution

La evolución de las tecnologías móviles está clara hacia la adopción de la tecnología LTE (Long Term Evolution) [24], esta tecnología es especificada en la serie 36 de las especificaciones técnicas del release 8 de la 3GPP (cerrado en Diciembre 2008) [17] y nuevas mejoras y beneficios introducidos en el release 9 (cerrado en Diciembre 2009). El foco de esta tecnología es el soporte de servicios de conmutación de paquetes con altas velocidades de datos y bajas latencias. Lo que permite entregar al usuario servicios con una alta calidad de experiencia.

En el caso actual de UMTS/HSPA, la tecnología soporta la voz utilizando una red de circuitos paralela, pero con una baja latencia y un rendimiento general mucho más alto ya no es necesaria la conmutación de circuitos y en su lugar se puede utilizar confiablemente VoIP.

LTE introduce una nueva red de acceso (eUTRAN) con nuevos conceptos para el plano de control y de usuario.

El método de acceso utilizado para el downlink es OFDMA (Orthogonal Frequency Division Multiple Access) que consiste en dividir un flujo de alta velocidad en varios de menor velocidad, transmitidos en simultáneo, donde cada flujo no necesariamente es asignado al mismo usuario.

El esquema de modulación multiportadora OFDM divide el espectro disponible en múltiples subportadoras ortogonales. Cada subportadora puede tener asignado su propio esquema de modulación, ya sea QPSK, 16QAM o 64QAM para el flujo de bajada.

OFDM además permite traslapar los subcarrier sin generar interferencia ahorrando hasta un 50% de ancho de banda debido a la ortogonalidad del código [16], además de otorgar soporte para MIMO y robustez frente a desvanecimiento por multitrayecto.

El desarrollo de las interfaces de aire para LTE está estrechamente ligado al simultáneo proyecto System Architecture Evolution (SAE) que ha definido la nueva arquitectura. Esta arquitectura es llamada Evolved Packet System (EPS) y abarca el LTE Evolved UMTS Radio Access (E-UTRA) y Evolved UMTS Radio Access Network (E-UTRAN) así como el nuevo Evolved Packet Core (EPC). El EPS es a veces llamado como arquitectura LTE/SAE y la 3GPP entrega las especificaciones con el nombre de System Architecture Evolution Specification (SAES).

Los nuevos elementos e interfaces definidos por la 3GPP para LTE junto con los elementos legacy de las redes de paquetes y circuitos son los que se muestran en la siguiente figura.

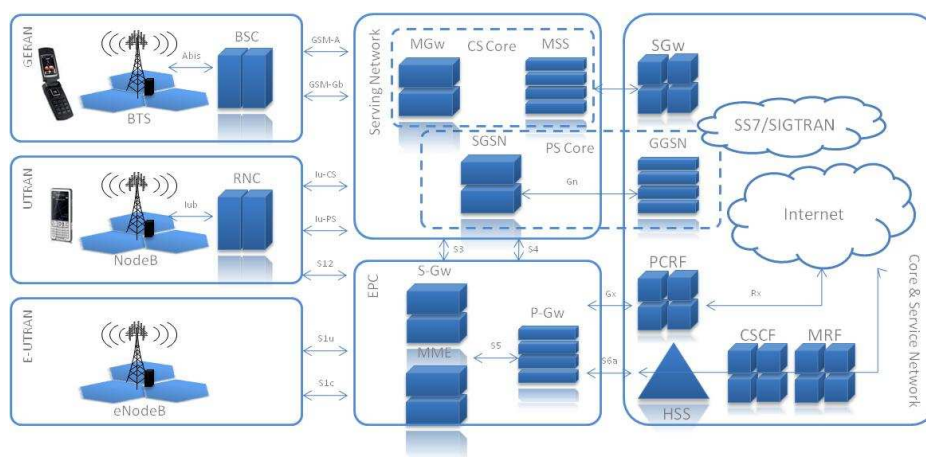


Figura A.69. Arquitectura EPS y Legacy

Mobility Management Entity

La entidad de gestión de movilidad, Mobility Management Entity (MME), es la entidad del plano de control que implementa los procedimientos usados por EPS Mobility Management (EMM) y EPS Session Management (ESM) en el EPS. El MME se comunica con el HSS para recibir la información de suscripción y con el Serving Gateway para establecimiento y liberación de los EPS bearers. La información de autenticación recibida desde el HSS es usada para la generación de llaves de integridad y cifrado, que son usadas para la protección de integridad y cifrado de los mensajes del plano de control del NAS. El MME se comunica con el eNB sobre la interfaz S1 y también con el SGSN usando la interfaz S3 para brindar soporte de movilidad entre redes de acceso 3GPP (2G y 3G).

Serving Gateway

El Serving Gateway (SG) termina la interfaz hacia la red de radio. Éste maneja el tráfico del plano de usuario ruteando y reenviando los paquetes de uplink y downlink entre el PDN Gateway y la red de radio (eNBs). Cada User Equipment (UE) se comunica con sólo un SG y el SG puede comunicarse con diferentes PDN gateways para diferentes conexiones que genere el UE realizando esta comunicación sobre la interfaz S5. El SG actúa como ancla para la movilidad inter-eNB e inter-3GPP. Éste se comunica con la entidad PCRF para las funciones de interceptación legal y control de tasación.

PDN Gateway

El PDN Gateway provee la conectividad a los UE a la red de paquetes de los diferentes servicios IP que proveen los operadores. El PDN Gateway configurado en la información de suscripción asigna una dirección IP al UE durante el attachment a la red. Un UE en estado conectado puede solicitar conexión a otro PDN Gateway para diferentes servicios IP. El PDN Gateway es responsable por la ejecución de la tasa de uplink y downlink dependiendo de la tasa máxima (Maximum Bit Rate) asignado al UE. El PDN Gateway nunca cambia durante una sesión sin tener en cuenta la movilidad del usuario.

Las velocidades que se pueden lograr con LTE dependen del ancho de banda utilizado en la portadora y del tipo de antenas que se utilice, tal como se muestra en la siguiente tabla.

Antenna Technology		Bandwidth		
TYPE	QTY	5 MHZ	10 MHZ	20 MHZ
MIMO	2x2	43	86	173
MIMO	4x4	82	163	326

Table 1a. LTE Peak Data Rates (Mbps) - No Error Rate Coding

Antenna Technology		Bandwidth		
TYPE	QTY	5 MHZ	10 MHZ	20 MHZ
MIMO	2x2	29	59	117
MIMO	4x4	55	113	226

Table 1b. LTE Peak Data Rates (Mbps) - 5/6 Error Rate Coding

Figura A.70. Tabla velocidades de LTE

Para el caso de los terminales, se puede apreciar las diferentes especificaciones para las categorías definidas en los estándares:

Category		1	2	3	4	5
Peak rate Mbps	DL	10	50	100	150	300
	UL	5	25	50	50	75
Capability for physical functionalities						
RF bandwidth		20MHz				
Modulation	DL	QPSK, 16QAM, 64QAM				
	UL	QPSK, 16QAM				QPSK, 16QAM, 64QAM
Multi-antenna						
2 Rx diversity		Assumed in performance requirements.				
2x2 MIMO		Not supported	Mandatory			
4x4 MIMO		Not supported				Mandatory

Figura A.71. Tabla de especificaciones de las categorías de los UE de LTE

LTE tiene definida su evolución hacia LTE Advance que es especificada en el release 10 de la 3GPP la cual aún no ha cerrado sus definiciones¹¹ pero que prevé lograr velocidades de 1Gbps.

¹¹ Para conocer el estado actual se puede visitar <http://www.3gpp.org/LTE-Advanced>

A.4 WiMax

WiMax (Worldwide Interoperability for Microwave Access) es el nombre con que se conoce al estándar IEEE 802.16. Éste se muestra como una alternativa altamente viable para satisfacer la demanda del servicio de banda ancha y como una alternativa de conectividad de última milla.

WiMax permite brindar conectividad en forma inalámbrica pero no brinda las facilidades de movilidad que ofrece, por ejemplo, LTE.

La tecnología de acceso utilizada es OFDM con modulaciones que van desde BPSK a 64QAM pasando por QPSK, 16QAM, dependiendo de las condiciones del canal.

El sistema lo componen diversos dispositivos que se muestran a continuación [25] :

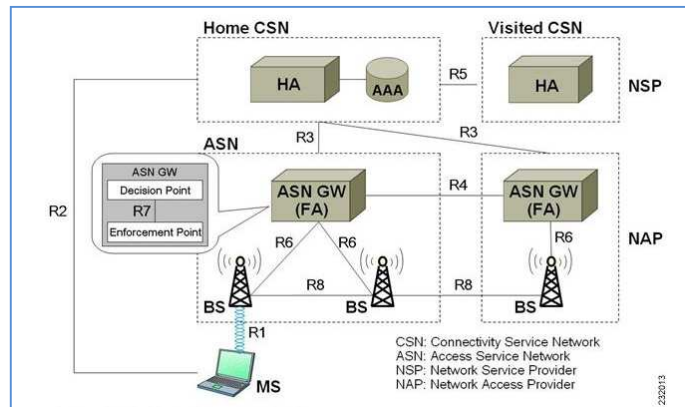


Figura A.72. Arquitectura WiMax

Mobile Subscriber Station

El Mobile Subscriber Station (MSS), típicamente llamado *Consumer Premise Equipment* (CPE) consiste en un equipo receptor y una antena que puede ser una caja independiente o una tarjeta que se conecta directamente a un computador y que brindan la conectividad al usuario hacia la estación base, entregándole acceso inalámbrico.

Access Service Network

La Access Service Network (ASN) consiste en un conjunto de nodos que proveen el acceso de radio al suscriptor WiMax. Este acceso es bridado a través de una estación base que la compone la electrónica *indoor* y una torre WiMAX. Por lo general, una estación de base puede cubrir un radio de hasta 10 km (En teoría, una estación de base puede cubrir hasta el radio de 50 kilómetros, sin embargo, las consideraciones

prácticas se limitan a unos 10 km). Cada una de estas estaciones bases se encuentra conectada a un ASN Gateway que sirve de agregador para luego enrutar los tráficos de usuario hacia el CSN.

Connectivity Service Network

La Connectivity Service Network (CSN) corresponde al conjunto de nodos que proveen el servicio de conectividad a los suscriptores hacia el Network Access Provider, típicamente un ISP hacia Internet. Entre los diferentes nodos se encuentran AAA, Routers, Home Agent, Base de Clientes, DNS y routers de acceso.

La tecnología incluye diversos niveles de calidad de servicio, seguridad y ciertos niveles de movilidad, permitiendo ofrecer servicios versátiles, principalmente de banda ancha.

Apéndice B. Desempeño de Protocolos de Transporte

Es de gran importancia para el análisis conocer cómo operan realmente los protocolos de transporte que son usados en Internet, ya que ellos afectan directamente en el rendimiento de las aplicaciones.

Hoy en día casi el 90% del volumen de tráfico en Internet está basado en el uso de TCP [1] y aunque el tráfico UDP sea minoritario es importante destacar que la cantidad de flujos son del orden del 70% por lo que no se debe descuidar. Si se mira la siguiente figura se ve la evolución y su distribución en los últimos 7 años donde el volumen de TCP decae cerca de 8 puntos porcentuales en el periodo y por otro lado la proporción de flujos UDP aumenta casi 70 puntos porcentuales en igual periodo.

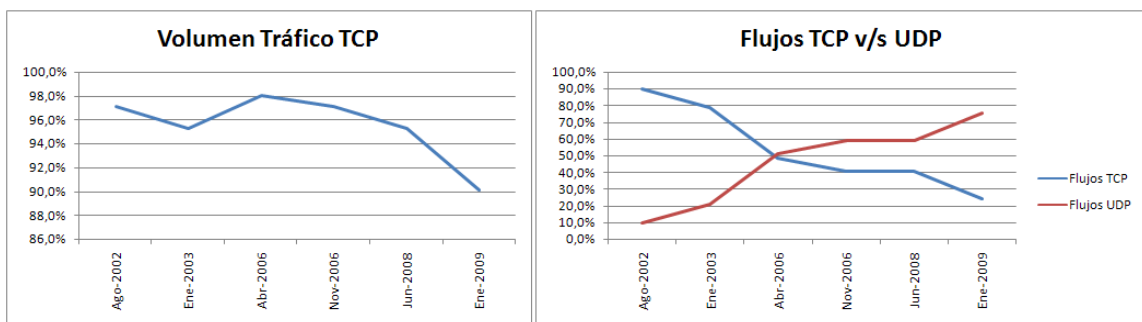


Figura B.73. Distribución de protocolos de transporte en Internet

TCP es un protocolo maduro pero ha sido optimizado para su operación en redes fijas y no para su operación en redes móviles, por lo que es recomendable tomar ciertas consideraciones al momento de buscar mejoras, tal como se ve en este capítulo.

Por otro lado, se puede ver que UDP genera una gran cantidad de flujos que luego deben ser evaluados al momento de realizar el diseño de los equipos de Internet, es decir, no sólo incorporar las variables de capacidad de throughput si no también la capacidad de conmutación y manejo de flujos.

B.1 Descripción de TCP

El protocolo TCP (Transmission Control Protocol) es uno de los principales protocolos del stack TCP/IP y se encuentra definido en RFC 793 de la IETF [5].

Es un protocolo orientado a la conexión diseñado para brindar una comunicación bidireccional, confiable de extremo a extremo, sobre una red no confiable. Para ello, implementa una serie de mecanismos como el control de flujo, detección y corrección de errores y control de congestión que le da un alto nivel de confiabilidad al protocolo, asegurando a las aplicaciones la entrega íntegra de la información transmitida.

Para lograr esto, es necesario definir una robusta cabecera tal como se ve en la siguiente figura:

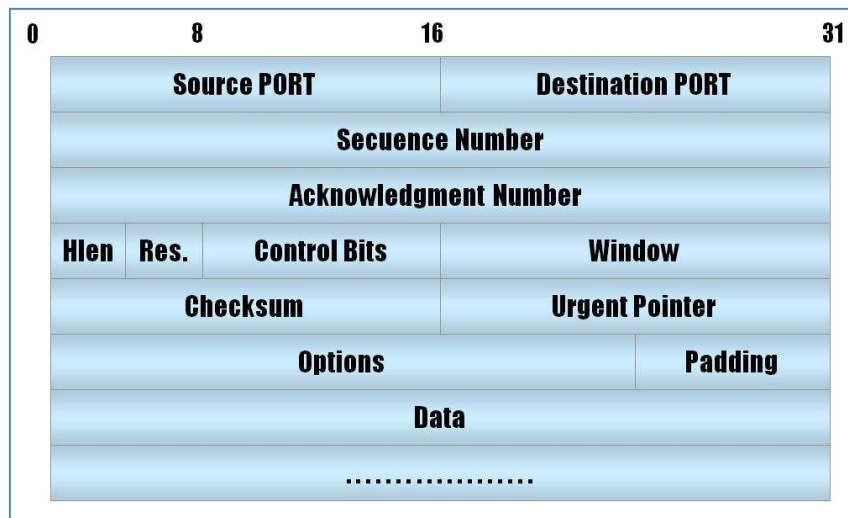


Figura B.74. Encabezado TCP

A continuación se describen los principales campos que componen la cabecera de TCP.

Los campos *Source Port* y *Destination Port* son utilizados para el direccionamiento de capa 4 en este protocolo.

Sequence Number y *Acknowledgment Number* son utilizados para enumerar los bytes enviados y confirmar su recepción lo que es utilizado para establecer el mecanismo de confiabilidad de asegura la recepción de la información enviada.

Window indica los bytes que pueden ser transmitidos sin haber recibido aún su confirmación de recepción. Este mecanismo permite aumentar la tasa de transmisión y hacer más eficiente el protocolo.

Control Bits incluye una serie de flags de control que indican, entre otras cosas, si se encuentra en el inicio o fin de la comunicación, información de acknowledgment relacionada con el campo *Acknowledgment Number* para la confirmación de recepción de datos.

Finalmente el campo *Options* incluye distintas informaciones, tales como la utilización del RFC1323 y el largo del MSS entre otros.

B.2 Funcionamiento

En términos generales se puede decir que una conexión TCP posee 3 fases.

La primera conocida como sincronización, donde se inicia el primer contacto entre las partes y se intercambian parámetros que permiten asegurar el correcto funcionamiento de la conexión y se informan las características que ésta tendrá. Esta fase se desarrolla con el intercambio de tres mensajes y es por ello que se conoce como *three way handshake*, en ella el extremo requiere iniciar una conexión envía un

SYN el que debe ser confirmado con un SYN-ACK y finalmente debe ser vuelto a confirmar por quien inicia la conexión con un ACK, quedando en condiciones de poder intercambiar información, tal como se muestra a continuación.

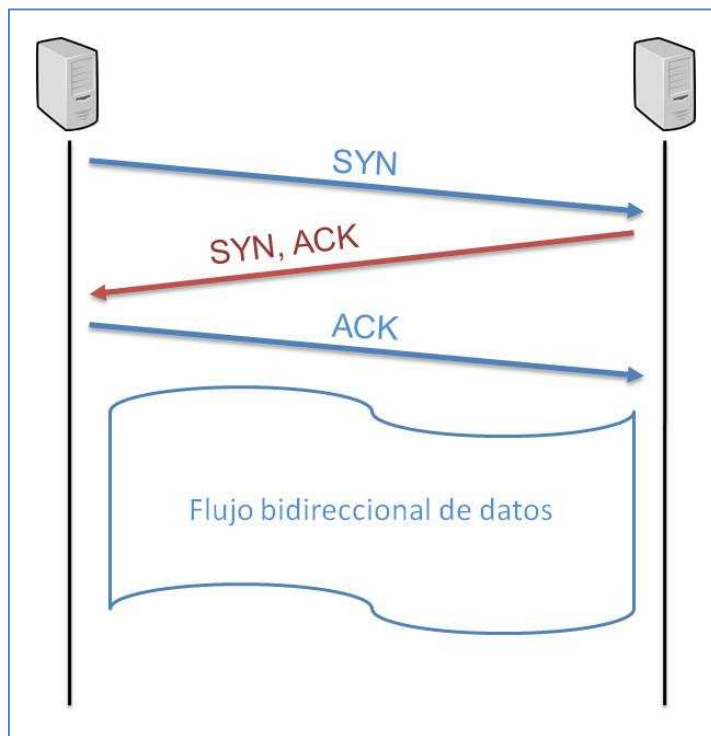


Figura B.75. Mecanismo three way handshake

Este mecanismo permite la negociación de parámetros de la conexión de red antes de iniciarse la comunicación. Por ejemplo, se definen los valores iniciales de los *sequence number* que se utilizarán para confirmación de recepción de la información independiente para cada sentido de la conexión lo que permite evitar la suplantación de conexiones conocido como *spoofing* y la detección de pérdida y duplicación de paquetes.

También se define el tamaño de la ventana de TCP, que corresponde a la cantidad de bytes que pueden ser enviados sin esperar confirmación de recepción, para ello se utiliza el campo *window* que posee 16 bit, lo que permitiría tener una ventana de 65.535 bytes, pero con el uso del escalamiento, que se ve con más detalle más adelante, se puede modificar este valor hasta 1 Gigabyte.

La segunda, donde la conexión ya se encuentra establecida y se realiza el intercambio de la información. Junto con la información de las aplicaciones se realiza el envío de mensajes de control que permiten que ésta se realice de manera fluida. En particular, se realiza el control de flujo de conexión y para ello se utiliza el mecanismo de *slow start* y *congestion avoidance*.

Y finalmente la tercera fase donde se negocia y realiza el cierre de la conexión. Ésta se realiza desde el extremo que requiere finalizarla, enviando un *flag* FIN el cual debe ser confirmado por el receptor mediante un ACK. A continuación este último debe enviar todos los datos pendientes para luego finalizar también la conexión mediante el envío de otro FIN el que debe ser confirmado con un SYN, tal como se muestra en la siguiente figura.

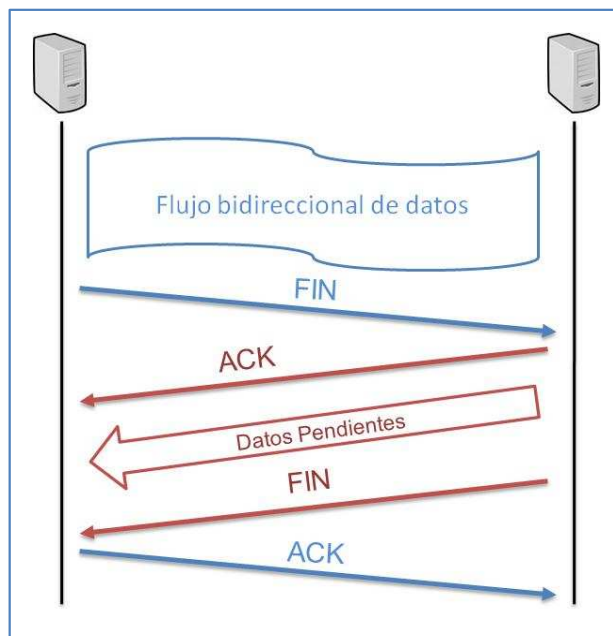


Figura B.76. Mecanismo de finalización de la conexión

B.3 Control de Flujo y Throughput de TCP.

Con la finalidad de regular el flujo de información y evitar la congestión, después del establecimiento una conexión TCP se activa un algoritmo de llamado *slow start*. Este algoritmo aumenta el volumen de datos en el aire (sin confirmar su recepción) exponencialmente y tras superar el umbral *slow start threshold* se activa un algoritmo de control de congestión llamado *congestion avoidance* [9] que regula el volumen de datos con el objetivo de no sobrepasar el límite físico de la conexión en curso, durante el cual se incrementa el número de bytes en forma lineal y con una baja pendiente, ya que se asume que la conexión se encuentra cercana a alcanzar el máximo de capacidad, evitando así saturar el canal.

En la siguiente figura se ver como evoluciona la ventana de congestión en términos de MSS (Maximum Segment Size) a medida que se reciben las confirmaciones de recepción de los segmentos enviados.

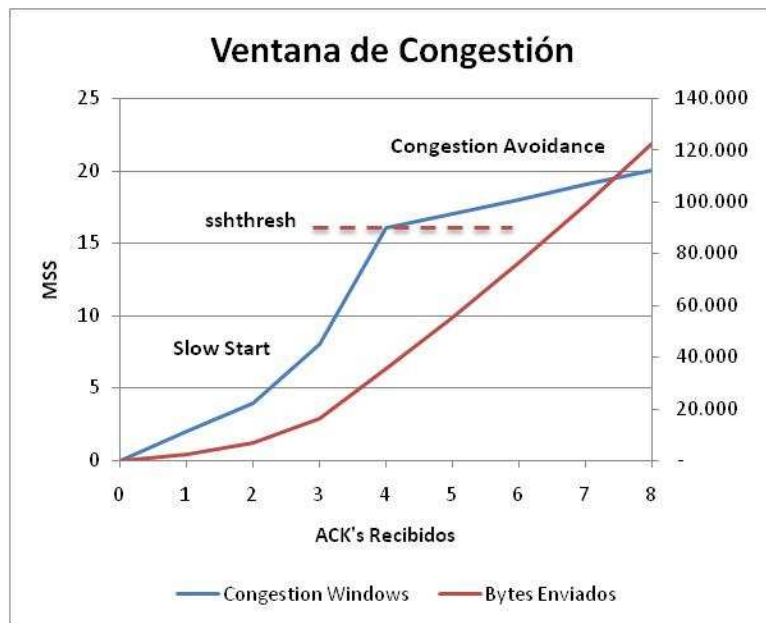


Figura B.77. Evolución de Ventana de Congestión durante sesión TCP

El estado de congestión se logra tras pasar *slow start threshold* y con la linealidad en el incremento de la ventana de congestión se tiene que el throughput de la conexión se incrementa lentamente. Es importante destacar que para alcanzar un throughput de 1Mbps es necesario transmitir al menor 400Kbytes, lo que significa que para transmitir datos de menor tamaño no es posible alcanzar velocidades altas usando una sola conexión.

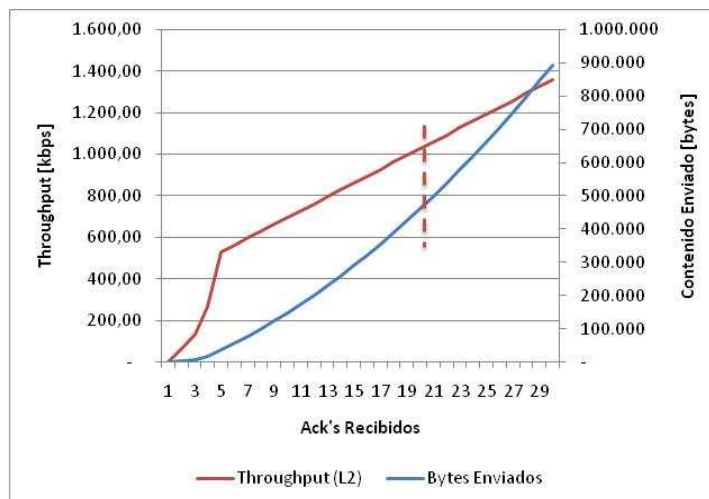


Figura B.78. Contenido enviado versus throughput alcanzado

Quando se inicia la conexión se define un valor inicial para la ventana de control de congestión dado por la siguiente fórmula [8] :

$$IW = \min(4 * MSS; \max(2 * MSS; 4380)) \quad (15)$$

Y mientras se encuentre en *slow start* la ventana de congestión se va incrementando a medida que van llegando las confirmaciones de recepción de la siguiente forma:

$$CWND_i = CWND_{i-1} + 2 * MSS \quad (16)$$

El umbral *slow start threshold* se define de acuerdo a la siguiente fórmula:

$$SSTHRESH_{inicial} = window \quad (17)$$

Cuando es superado este umbral y la conexión se encuentra en *congestion avoidance* se modifica la ventana de congestión de modo tal que se disminuya la pendiente como se muestra:

$$CWND_i = CWND_{i-1} + MSS * \frac{MSS}{CWND_{i-1}} \quad (18)$$

B.4 Pérdida de Paquetes.

La pérdida de paquetes pertenecientes a una comunicación TCP se debe, principalmente, al descarte de paquetes en algún tramo del camino, como por ejemplo en la cola de algún router congestionado, o a que simplemente el segmento siguió un camino más largo superando el tiempo de espera máximo y luego es considerado como perdido.

Cuando un segmento TCP es enviado, se activa un contador de tiempo de espera de la confirmación de recepción del segmento respectivo, pero mientras llega esta confirmación se sigue enviando información cumpliendo los criterios de control de flujo y ventana deslizante. En caso de superar el tiempo de espera se asume pérdida del segmento y se realiza el reenvío del segmento tal como se puede ver en la siguiente figura.

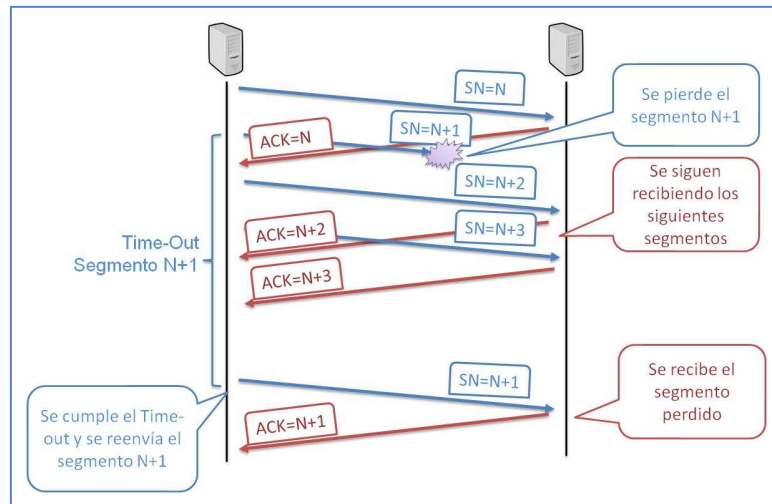


Figura B.79. Flujo de pérdida y recuperación de segmento usando Time-Out

El tiempo de espera debe definirse lo suficientemente alto para evitar reenvío de segmentos con ACK retrasados y los suficientemente bajo para que no existan periodos de inactividad de datos. Para ello se utiliza el algoritmo de Karn[10] que define el tiempo de espera de retransmisión, RTO (*Retransmission Time-Out*) de la siguiente forma:

$$RTO_i = \beta * SRTT_i \quad (19)$$

Donde SRTT (*smoothed round-trip time*) corresponde a una estimación del tiempo de viaje que toma un segmento en ir y volver, y que se calcula de la siguiente forma:

$$SRTT_{i+1} = \alpha * SRTT_i + (1 + \alpha)S_i \quad (20)$$

Donde S_i es el RTT medido en la iteración i , α es una constante definida entre 0 y 1 que controla cuán rápido se adapta SRTT al cambio y habitualmente se define en $7/8$ y β es otra constante mayor que 1 definida como la probabilidad que el RTT del paquete sea mayor que el RTO, en este caso el valor recomendado es $\beta = 2$.

Este mecanismo permite definir el RTO en forma dinámica dependiendo de las condiciones del medio y ajustándose a éste.

Cuando existe una pérdida de segmentos, se tiene que la tasa de transmisión de la conexión decrece debido a que el extremo transmisor debe esperar que se cumpla el RTO para poder retransmitir, luego después de llenar la ventana de transmisión este se queda sin enviar información hasta que pueda retransmitir.

Para mejorar esta ineficiencia se implementa el mecanismo *Fast Retransmit*, que consiste en el envío de un ACK triplicado cuando se detecta una pérdida de secuencia en los segmentos recibidos, es decir, se asume que el segmento se ha perdido y no se espera que se cumpla el time-out asociado y el extremo receptor reenvía el ACK anterior tres veces seguidas, lo que se interpreta como la pérdida de la información enviada y el extremo transmisor vuelve a enviarla, de esta forma se disminuye el tiempo de espera mejorándose la eficiencia del protocolo tal como se muestra en la figura.

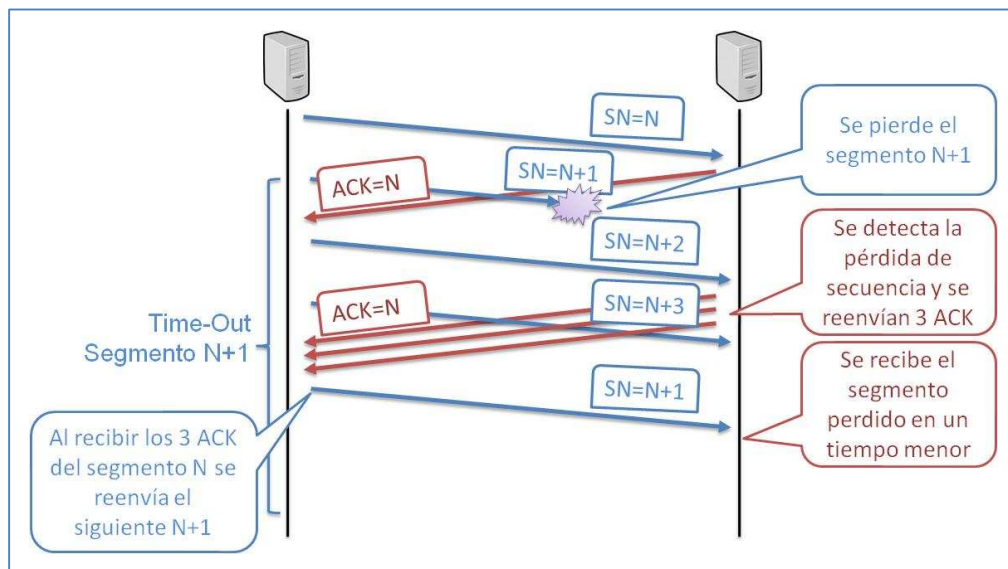


Figura B.80. Flujo de pérdida y recuperación de segmento usando *Fast Retransmit*

La pérdida de segmentos es interpretada por el algoritmo de control de flujo como congestión y luego éste debe ajustar el flujo y adaptarse a la nueva condición.

Cuando se detecta congestión debido a la pérdida de paquetes, TCP ajusta el flujo de paquetes modificando la ventana de congestión y *slow start threshold* de la siguiente forma:

$$CWND_i = MSS \quad (21)$$

$$ssthresh = \max\left(\frac{flightSize}{2}; 2 * MSS\right) \quad (22)$$

Donde *flightSize* corresponde al número de bytes que hay en el aire.

Esta modificación de parámetros genera un fuerte descenso en la tasa de transmisión y una recuperación más lenta, tal como se puede ver en la siguiente figura:

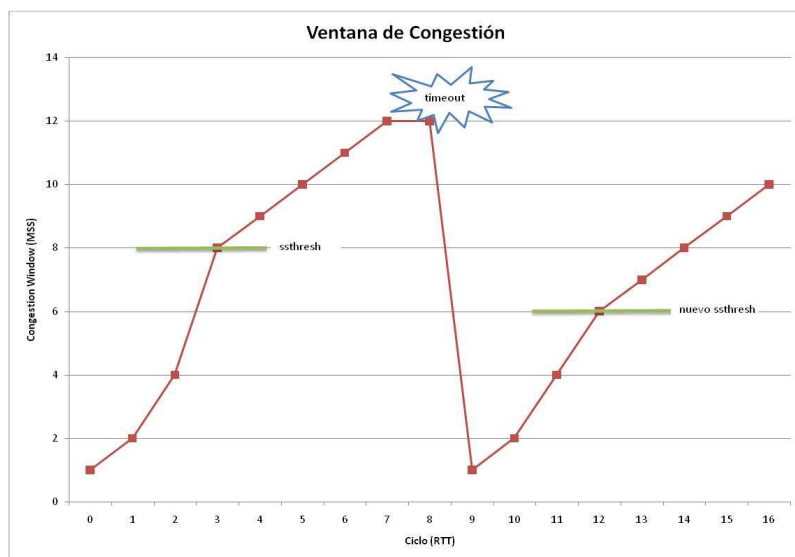


Figura B.81. Adaptación de TCP luego de pérdida de paquetes

Este esquema es ineficiente ya que si se lleva a la práctica, cuando existe una pérdida claramente es necesario disminuir la tasa de transmisión de datos pero no necesariamente comenzar como si se iniciara la conexión.

Para corregir esto se implementa el mecanismo *Fast Recovery* que considera los siguientes valores la ventana de congestión y manteniendo el mismo en el caso del umbral:

$$CWND_i = ssthresh + 3 * MSS \quad (23)$$

$$ssthresh = \max\left(\frac{flightSize}{2}; 2 * MSS\right) \quad (24)$$

La mejora se puede apreciar en las figuras:

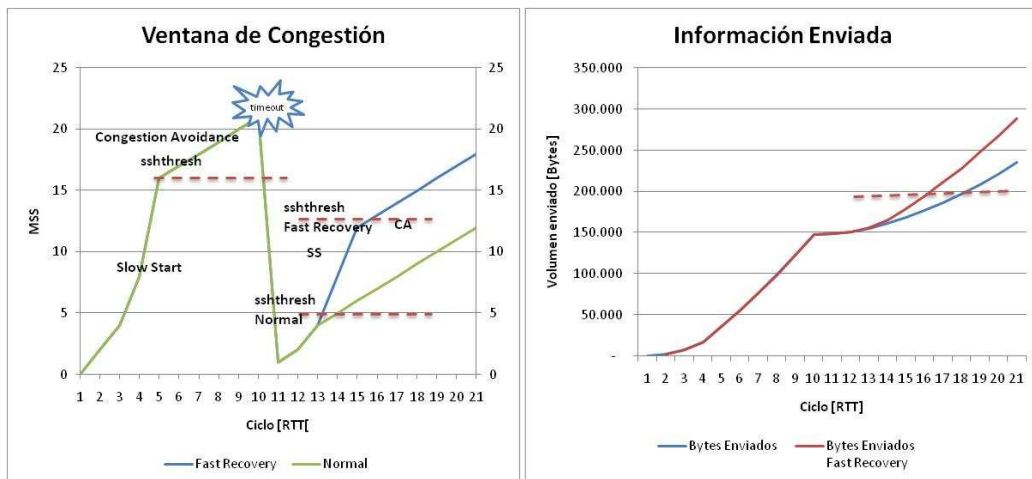


Figura B.82. Adaptación de TCP luego de pérdida de paquetes con y sin Fast Recovery

La mejora de *Fast Recovery* permite alcanzar con mayor rapidez el throughput máximo de la conexión y lograr que la información llegue a destino varios ciclos antes.

B.5 Descripción de UDP

El protocolo UDP (User Datagram Protocol) es un protocolo de transporte no orientado a la conexión parte del stack TCP/IP definido en el RFC 768. Permite el envío de datagramas a través de la red sin que exista una conexión previa. No tiene confirmación de envío ni control de flujo. Tiene un bajo overhead haciéndolo ideal para el transporte de audio y video en tiempo real.

El encabezado de UDP ha sido optimizado y reducido manteniendo los campos básicos como se puede ver a continuación:

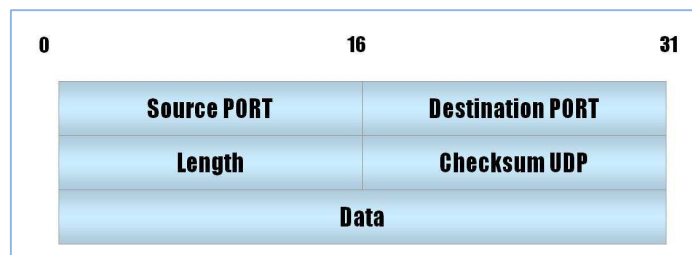


Figura B.83. Encabezado UDP

Los campos *Source Port* y *Destination Port* al igual que en TCP son utilizados para el direccionamiento de capa 4 en este protocolo.

Los campos *Length* indican el largo del paquete, el campo *Checksum* es el campo de chequeo de errores y a continuación viene la data encapsulada.

Debido a que UDP es un protocolo no orientado a la conexión, éste no requiere de mecanismos de inicio, fin y mantención de la conexión como en el caso de TCP, haciendo más simple y minimizando el overhead.

Es importante recordar que debido a que UDP no posee control de flujo puede generar un gran problema cuando las aplicaciones son altamente demandantes de ancho de banda, esto se debe a que pueden saturar el canal de comunicación o parte de éste generando pérdida de paquetes propios y de otras conexiones.

B.6 Comparación Rendimiento TCP/UDP

En términos de control de flujo y congestión existe una diferencia importante entre UDP y TCP, el primero no posee control de flujo en cambio TCP si lo tiene. Esto se traduce en que en UDP no existe una adaptación extremo a extremo a las variaciones de las capacidades del camino completo y en cambio en TCP si la hay. El control de flujo en TCP se realiza una adaptación al canal más estrecho, en cambio en UDP lo que ocurre es que utiliza el máximo el canal inicial y cuando se está con un canal estrecho existe pérdida de paquetes tal como se muestra en la figura.

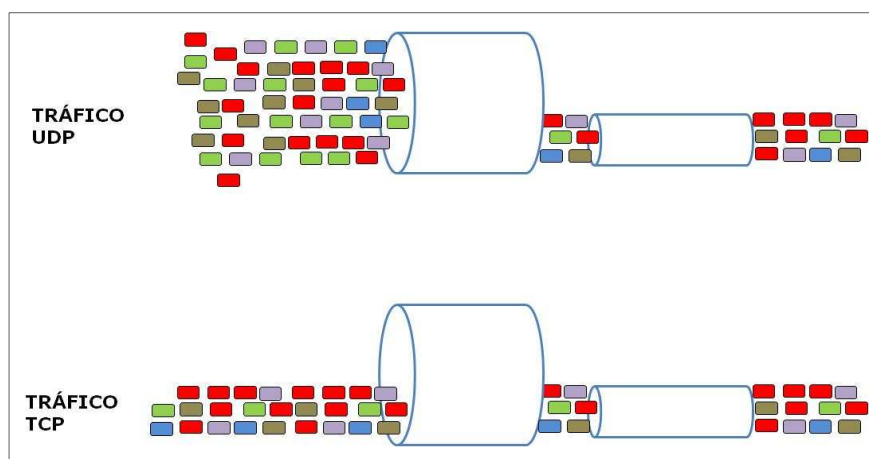


Figura B.84. Variación en flujos de tráfico TCP v/s UDP al pasar a un canal de menor capacidad.

Cuando convive tráfico UDP con tráfico TCP se puede ver que rápidamente el canal de comunicaciones se satura de paquetes UDP que en su gran mayoría serán descartados y nunca llegarán a destino y por otro lado, TCP se adapta al ancho de banda disponible dejándolo en una mínima expresión producto de la inundación de datagramas UDP.

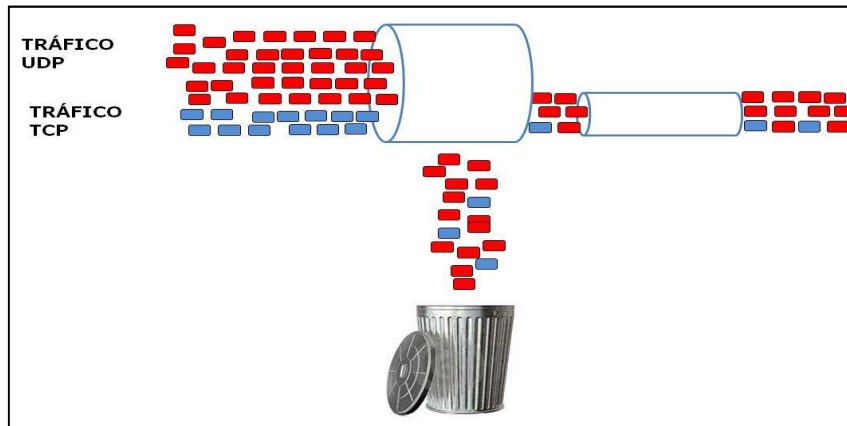


Figura B.85. Control de Flujo TCP y UDP y pérdida de paquetes

Esto significa que el consumo de ancho de banda es principalmente para aplicaciones intensivas basadas en UDP y queda muy poco disponible para compartir entre aplicaciones TCP, lo que genera una mala percepción del servicio.

Otro punto importante que se debe considerar es el overhead que aportan los protocolos de transporte a las aplicaciones ya que éste incide en el throughput efectivo que aprecia el cliente.

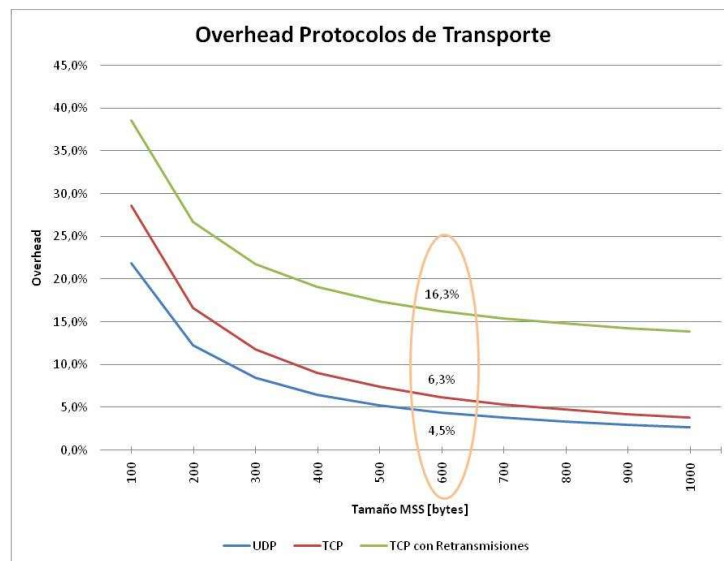


Figura B.86. Overhead protocolos de transporte

Como muestra la figura anterior, para una transmisión de paquetes promedio de 600 bytes, UDP aporta un 4,5% de overhead y TCP un 6,3% y pudiendo llegar a más de un 16,3% si se considera una tasa normal de 10% de retransmisiones.

Este nivel de overhead genera una disminución en el ancho de banda efectivo disponible para el flujo de datos de aplicación afectando directamente el tiempo de

transmisión de la información pudiendo variar si la aplicación utiliza UDP o TCP y dependiendo además de la tasa de errores end-to-end que afecte el flujo de datos.

Apéndice C. Descripción Servicios de Internet

Lo que busca un usuario de banda ancha móvil es usar los servicios disponibles en Internet y es por ello que es importante conocer las características técnicas principales de ellos, de modo tal que se pueda comprender su funcionamiento y los factores que pueden alterar su rendimiento y cómo se puede mejorar la percepción de ellos.

C.1 Web Browsing

El servicio de web browsing es uno de los más masivos en Internet, corresponde a lo que el usuario conoce como navegación propiamente tal.

Es un servicio interactivo donde el usuario solicita información a un servidor de contenidos, como es el caso de la página web de un banco, un periódico o web mail (gmail, yahoo mail, hotmail, etc). A continuación, el usuario lee (ve o escucha) la información y luego vuelve a solicitar una nueva.

Las páginas web están compuestas por una serie de meta-archivos y el navegador realiza una serie de descargar simultáneas y semi-simultáneas para obtener y desplegar la información. Debido a que el usuario se encuentra esperando que la información sea desplegada en el mínimo de tiempo posible, este servicio es altamente sensible, pero a su vez depende no solo del estado de la red sino también del nivel de congestión del servidor que brinda el contenido. En la siguiente figura se ve el flujo de la información producido durante una navegación normal.

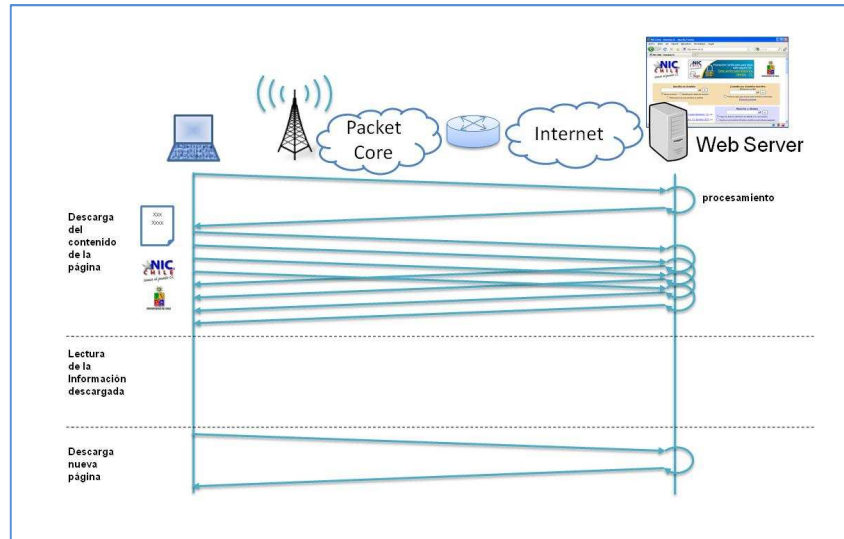


Figura C.87. Diagrama de flujo de la descarga de un contenido web.

El contenido descargado genera un flujo intensivo de paquetes de múltiples conexiones de datos en forma de ráfagas tal como se muestra en las figuras y comienza a utilizar la información descargada una vez que se despliega la primera información visible o audible, mientras tanto se sigue descargando el resto de la página.

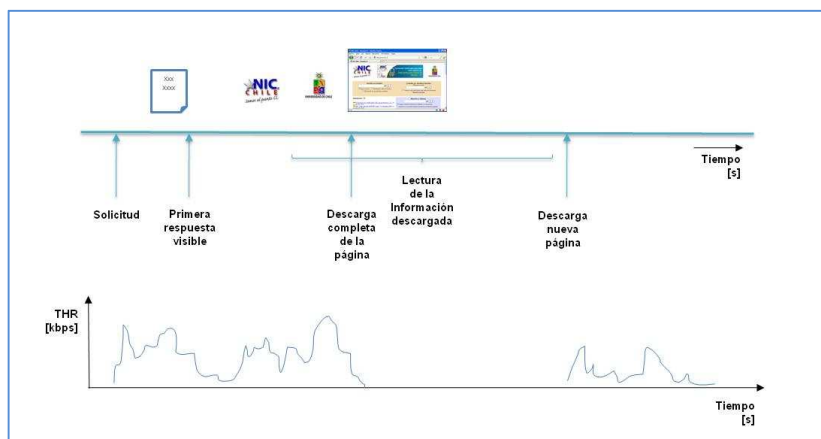


Figura C.88. Throughput de un usuario descargando contenido web.

El tráfico en forma de ráfagas requiere una rápida respuesta al servicio solicitado sin ser altamente demandante para la red, debido a que el throughput multiplexado de todos los abandonados no genera un alto nivel de tráfico para la red.

C.2 Streaming y Flash Video

El servicio de video streaming ha evolucionado considerablemente en los últimos años. Inicialmente se conceptualizaba considerando el uso de protocolos ad-hoc como RTP que opera sobre UDP y permite enviar un flujo de información con bajo overhead pero que tolera la pérdida de información. Hoy en día este mecanismo ha cambiado y el protocolo predominante para dar soporte al streaming de video es RTMP (Real Time Messaging Protocol) que opera sobre TCP y ofrece un transporte confiable para el streaming de video, incluso permitiendo el envío de video de alta calidad por la red. Esto sólo es posible gracias a los considerables aumentos de ancho de banda de Internet, pero todavía existen problemas de calidad los cuales aún no son abordados por los proveedores de acceso a internet.

En streaming de video es muy importante tener un flujo de datos que permita mantener una continuidad en la reproducción del contenido. Esto se logra principalmente utilizando la estrategia de buffering donde el cliente multimedia almacena la información a medida que esta llega a una tasa mayor de la de despliegue y de este modo permite que no existan interrupciones cuando el flujo de datos disminuye. Durante la primera parte de la transmisión el cliente sólo almacena datos y luego de completar un nivel aceptable el cliente comienza a reproducir la información. Si bien esta técnica reduce la pérdida de la continuidad en la reproducción, no se asegura que ésta no exista, luego, el algoritmo debe ser capaz de calcular el tamaño del buffer inicial en función de la calidad del canal de transmisión, tal que, extrapolando las condiciones asegure una reproducción libre de interrupciones.

El streaming tradicional de video o audio opera utilizando el protocolo RTSP especificado en el RFC 2326, el cual utiliza conexiones TCP y UDP para su

funcionamiento, la primera para establecer la sección y controlar el flujo de datos y la segunda para el envío de los datos propiamente tal, tal como se muestra en la figura.

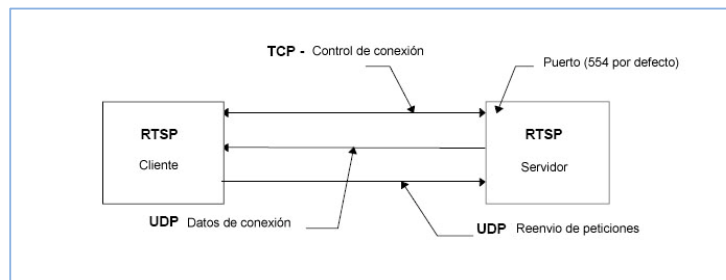


Figura C.89. Funcionamiento de streaming tradicional.

Primero se realiza una solicitud http con el descriptor de la información y datos del media server y luego se realiza un requerimiento hacia el media server que es el encargado de iniciar una sesión UDP enviando la información de video y audio que luego el cliente recibe y despliega.

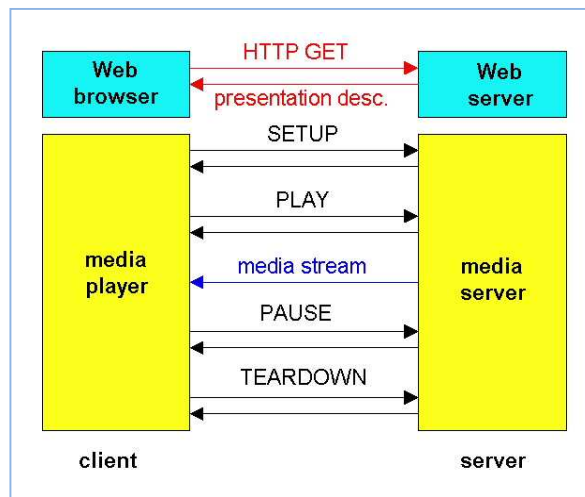


Figura C.90. Funcionamiento Streaming

Debido a que la información de audio y video es enviado usando transporte udp, se tiene que es susceptible a pérdida de datos, lo que puede degradar la información enviada.

La evolución de las redes de datos ha permitido disponer de conexiones con anchos de banda mucho mayores, es por ello que se comienza a utilizar RTMP como protocolo por defecto para el envío de video.

Su funcionamiento es similar al del streaming tradicional, pero se utiliza TCP como protocolo de transporte, esto genera que sea más robusto y permite se pueda emitir video en alta calidad pero demandando una mayor capacidad del enlace de

comunicaciones y adicionalmente requiriendo una capacidad mínima par su funcionamiento sin interrupciones.

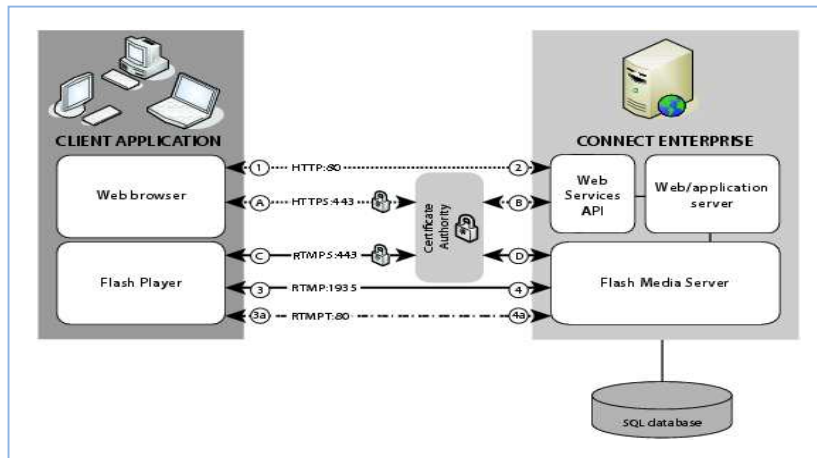


Figura C.91. Funcionamiento RTMP

En la siguiente figura se puede ver una simulación de una reproducción de flash video donde se puede ver la pérdida paquetes que afecta el tamaño del buffer al punto de generar interrupciones (línea verde) en la reproducción del video.

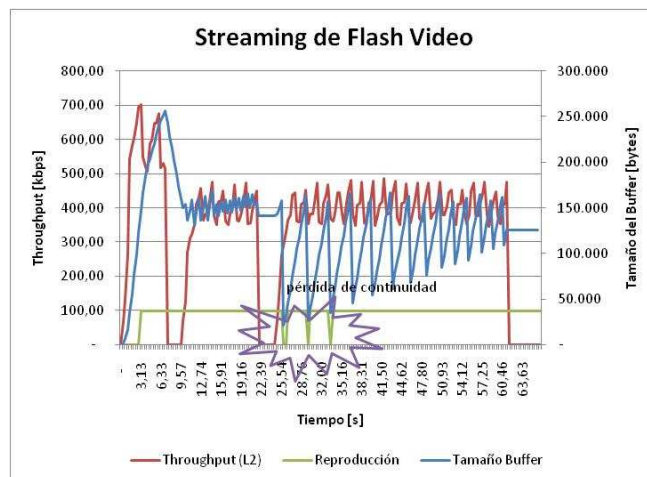


Figura C.92. Streaming de Video

Nuevamente si se considera que existen elevadas tasas de retransmisiones producto principalmente del escaso ancho de banda se ve que las aplicaciones son afectadas y en este caso comienza a decaer el tamaño del buffer a niveles críticos donde se detiene la reproducción del video hasta volver a tener datos.

C.3 P2P

El servicio de compartición de archivos vía peer-to-peer utiliza como medio de transporte los protocolos UDP y TCP, generando múltiples conexiones entre todos los nodos que comparten la información y operando como cliente y servidor a la vez, formando un gran enmallado interconectado. Para aumentar la eficiencia del protocolo, todos los clientes que descargan una parte del archivo se transforman en servidores y comienzan a compartirlo, aumentando así la disponibilidad del contenido en forma exponencial. Además el mecanismo contempla el uso de múltiples conexiones para aumentar la eficiencia y el uso de UDP para aumentar la velocidad generando un alto nivel de tráfico entre los nodos.

Cada nodo puede estar conectado a varias decenas de servidores a la vez que le entregarán parte de los archivos compartidos, una vez que el cliente termina de descargar una parte, inmediatamente lo pone a disposición del resto de los nodos y actúa como servidor compartiendo dicha información con los demás dispositivos. En la siguiente figura se puede ver la constelación formada.

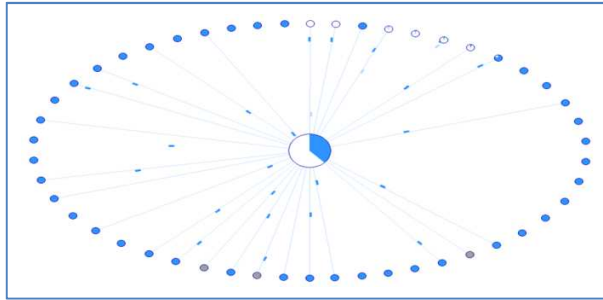


Figura C.93. Constelación P2P

Este mecanismo resulta ser altamente agresivo llegando a generar importantes niveles de congestión en los canales de acceso para la subida y el core o Internet en caso de bajada. Esto se debe a que el alto tráfico generado es contenido en los puntos de estrechamiento, descartándose la proporción que no se puede conmutar, pero debido a que UDP no posee control de flujo se tiene que la fuente sigue generando contenido a grandes tasa independiente si éste es recibido o no, tal como se muestra en la siguiente figura:

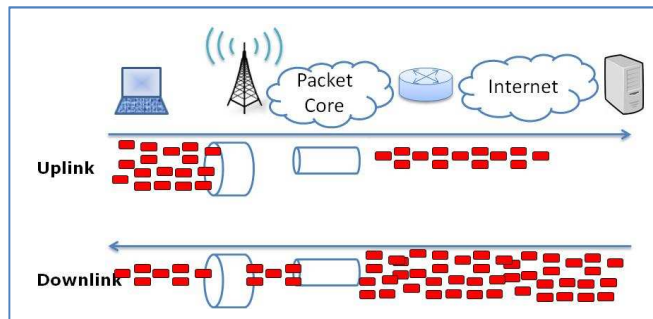


Figura C.94. Flujo de tráfico P2P

C.4 Network Storage

Uno de los servicios que está reemplazando a los tradicionales P2P para la compartición de archivos son los servicios de almacenamiento en línea, o network storage. Estos corresponden simplemente a servicios web capaces de almacenar grandes cantidades de información y ponerlas en línea para que los usuarios realicen su descarga en forma directa, tal como lo hacen cuando bajan un nuevo software.

Dentro de esta gama de proveedores se encuentran servicios como los de rapidshare, megaupload, depositfile, etc. que ofrecen el servicio en dos modalidades: gratis y Premium. El primero consiste en un servicio que me permite descargar sólo un archivo a la vez a bajas velocidades y el segundo permite múltiples conexiones simultáneas y el uso de servidores que permiten descargar el contenido a altas tasas de transferencias.

Estos servicios están siendo utilizados hoy en día para la compartición no sólo de música, sino principalmente de películas, es decir, para la descarga de archivos del orden de los GigaBytes de información, donde los usuarios descargan continuamente durante todo el día y la noche hasta lograr descargar todas las partes que componen la información que andan buscando, convirtiéndose en uno de los principales servicios utilizados por los clientes y uno de los más agresivos al momento de analizar el uso de recursos de red.

C.5 Redes Sociales

El servicio de redes sociales es el con mayor crecimiento en los últimos años. Si bien en estricto rigor corresponde a un subconjunto de los servicios existentes, el modo de ofrecer el servicio a marcado una gran diferencia que ha permitido que día a día se sumen más adeptos.

Uno de ellos es Facebook¹² que consiste en un portal web que permite compartir información entre los usuarios del sistema, primando la compartición de fotos y videos. También se encuentra un amplio surtido de pequeñas aplicaciones y juegos, basadas en flash que complementan el portal, enriqueciéndolo y haciéndolo más complejo.

En términos generales, el contenido no es muy pesado, pero si exige una alta interacción, exigiendo además una baja tasa de retardo para así tener una mejor experiencia.

Otro servicio altamente demandado es la aplicación Twitter¹³, en su origen permite la compartir información basada en pequeños mensajes de 160 caracteres que son recibidos por el resto de los usuarios que ha comparten la información. Este servicio comenzó a evolucionar, incluyendo imágenes y hasta video. Dado el potencial

¹² www.facebook.com

¹³ www.twitter.com

que este tiene, fue rápidamente adoptado como un canal de información utilizado por las empresas o medios informativos. Aunque el usuario espera una alta interacción, este servicio tiene además una alta tolerancia al retardo, permitiendo que éste se actualice periódicamente pero sin exigir instantaneidad.

C.6 Otros Servicios

La parrilla de servicios disponibles es amplia y crece cada día más, es por ello que no hay que limitarse a las aplicaciones típicas y es necesario ir más allá investigando las aplicaciones emergentes de modo tal que el operador pueda responder rápidamente a la llegada de nuevas aplicaciones y ofrecer una buena calidad considerando los requerimientos de los clientes.

Apéndice D. Caracterización de Servicios

De acuerdo a la ITU-T [15] se puede clasificar los servicios y aplicaciones utilizando características del servicio, parámetros de tráfico y comunicaciones y ambientes de operación.

Esta clasificación ayudan a comprender los diferentes requerimientos para la red que exigen indirectamente los usuarios considerando los servicios y aplicaciones que usan. Esto define las características que se deben mejorar al momento de determinar los niveles de calidad de la red.

La recomendación considera un árbol de características que se pueden ver en la siguiente figura:

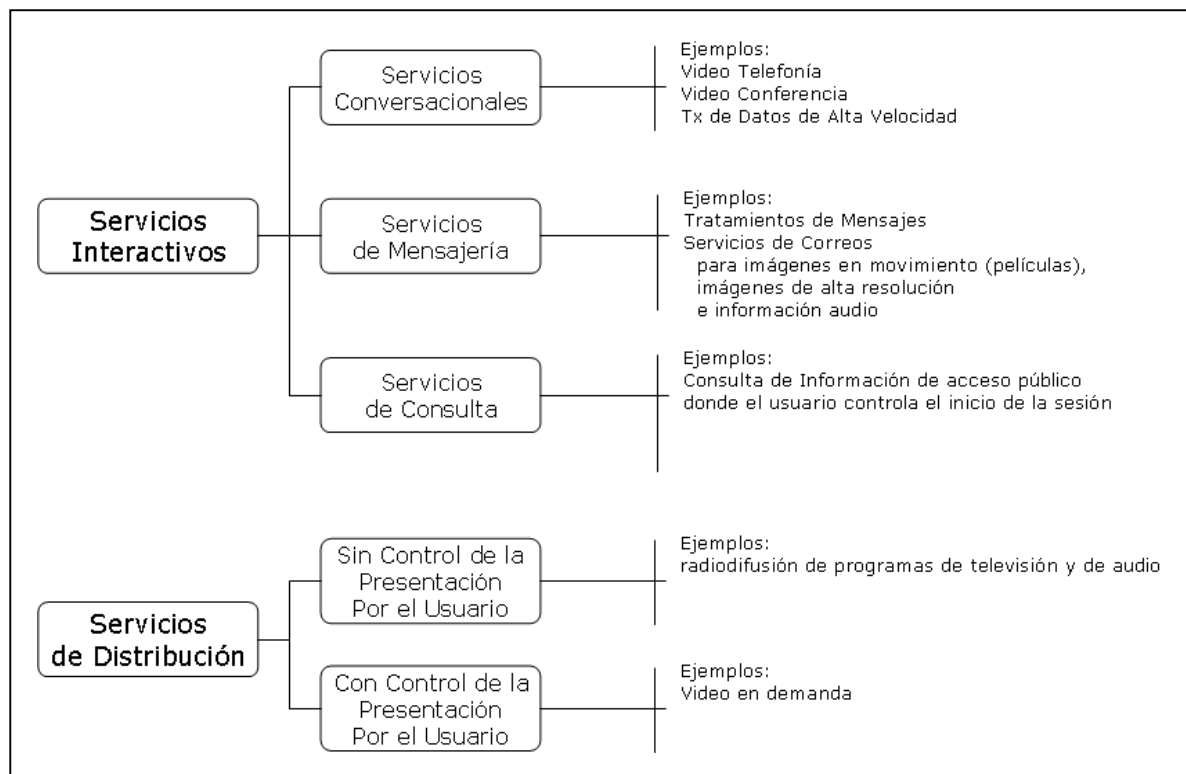


Figura D.95. Clasificación de Servicios de Banda Ancha

Se consideran dos grandes clasificaciones: Servicios Interactivos y Servicios de Distribución. La primera categoría agrupa servicios solicitados o accedidos interactivamente por el usuario y se subdivide en Servicios Conversacionales, Servicios de Mensajería y Servicios de Consulta. Y la segunda agrupa a servicios de distribución sin y con control de la presentación por el usuario.

Así mismo los servicios se pueden clasificar de acuerdo a las siguientes características de servicio[3] :

Característica	Iniciales	Significado
Dependencia del Tiempo	TB o NTB	Time Based: basado en el tiempo Not Time Based: no basado en el tiempo.
Requerimiento de Entrega	RT o NRT	Real Time: en tiempo real Not Real Time: no en tiempo real.
Direccionalidad (de la comunicación)	UND o BIND	Unidirectional: comunicación unidireccional Bidirectional: comunicación bidireccional
Simetría (de la comunicación)	SYM o ASY	Symmetric: comunicación simétrica. Asymmetric: comunicación asimétrica.
Interactividad		Comunicación Interactiva o no.
Número de participantes		Número de participantes en la comunicación.

Tabla D.96. Características de Servicios de Banda Ancha

De este modo, se puede definir las características del servicio de acuerdo a la clasificación anterior lo que puede entregar una idea clara de los requerimientos del sistema para su correcto funcionamiento.

La siguiente tabla muestra la caracterización de los distintos tipos de servicios clasificados según la recomendación ITU-T I.221

Service hierarchies	Type of information	Examples of broadband services	Examples of applications	Service characteristics			
				Intrinsic time dep.	Delivery reqs.	Directionality	Symm./asymm.
Interactive, conversational	Moving pictures and sound	HD videotelephony	1. Tele-education	TB	RT		Sym/Asy
		ISDN videoconference	2. Tele-advertising				
		Video surveillance	3. Mobile video surveillance				
		Video/audio information transmission service	4. HDTV outside broadcast				
	Data	High-speed unrestricted information Tx. service	5. Wireless LAN interconnection		NRT/RT	Bid	Asy
		High volume file transfer service – FTP	6. Data file transfer (FTP)	NTB	RT		
	Document (multimedia)	High resolution image communications service	7. Professional images	TB/NTB	RT		
		Mixed document communications service	8. Desktop multimedia	TB	RT		
			9. Mobile emergency services				
			10. Mobile repair assistance				
	11. Mobile teleworking						
			12. Freight and fleet management			Sym	
Interactive, messaging	Mixed documents	Multimedia mail	13. Electronic mailbox service for multimedia	TB	NRT	Bid	Asy
Interactive retrieval	Text, data, graphics, sound, still images, moving pictures	Broadband videotex	14. E-commerce	TB	RT	Bid	Asy
			15. Multimedia library				
			16. Tourist information	NTB			
	Data retrieval service	17. Remote procedure call		NRT/RT			
	Multimedia retrieval service	18. Urban guidance	TB				
		19. Assistance in travel					
Distributed broadcast	Video	MPEG2-4	20. TV programs distribution	TB	RT	Bid	Asy
Distributed cyclical	Text, graphics, sound, and still images	Full channel broadcast videography	21. E-newspaper	TB	NRT	Bid	Asy

Tabla D.97. Clasificación de Servicios y Aplicaciones en términos de categorías y tipos de información.

Cada una de estas características de servicios se asocia a un parámetro de calidad de servicio que puede ser medido y monitoreado en distintos puntos de la red. Por ejemplo, para asegurar el correcto funcionamiento de servicios en tiempo real (RT) se debe tener un nivel tolerable de retardo y de jitter. Por otro lado, existen servicios que demanda un alto ancho de banda y este puede ser unidireccional o bidireccional, por lo que se debe medir los niveles de ancho de banda disponible en ambos sentidos. Esta asociación es importante que sea realizada para así comprender que parámetros de la red se debe monitorear y como afectan los servicios de los abonados.

Cuando se quiere caracterizar a los servicios que utilizan los abonados de BAM, la metodología tradicional indica que se agrupan de acuerdo al ancho de banda que requiere cada una de las aplicaciones, tal como se ve en la siguiente figura que muestra el ancho de banda necesario para distintos servicios móviles.

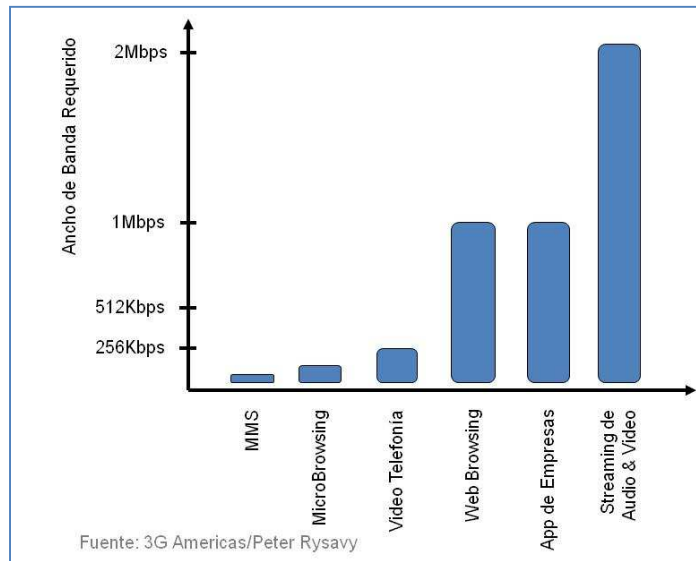


Figura D.98. Demanda de ancho de banda por aplicación

Pero como se ha visto también existen una serie de indicadores que pueden categorizar más fielmente las distintas aplicaciones, como por ejemplo, cuando se habla de servicios dependientes del tiempo (RT) o interactivos se debe considerar otros factores como son el retardo y el jitter. Luego si se hace una caracterización utilizando ancho de banda que necesitan, dependencia al retardo y el jitter se puede obtener interesantes resultados que permiten comprender más el servicio.

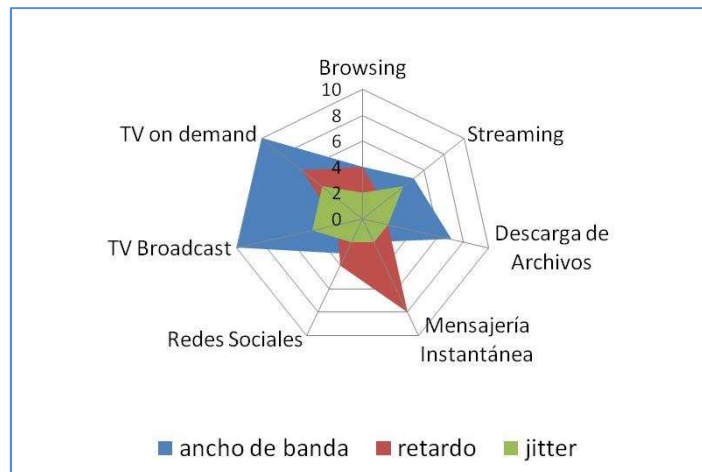


Figura D.99. Clasificación de Servicios de Banda Ancha

Una categorización adecuada permite al operador optimizar su red atendiendo los factores que afectan los servicios que realmente utilizan los abonados, enfocando sus esfuerzos en la mejora sostenida de la calidad de experiencia del abonado.