

Algoritmos de Planificación de recursos para redes inalámbricas 4G

Jaume Ramis, Loren Carrasco, Guillem Femenias y Felip Riera

Universitat de les Illes Balears

Cra. Valldemossa, km. 7,5. 07122-Palma. SPAIN

tel: (34) 971.171.390 fax: (34) 971.173.003

Correo Electrónico: {jaume.ramis, loren.carrasco, guillem.femenias, felip.riera}@uib.es

Resumen

Los algoritmos de planificación son los componentes fundamentales en el proceso de asignación de recursos para conseguir el cumplimiento de los requisitos de calidad de servicio. Cuando el diseño corresponde a redes inalámbricas éste se complica considerablemente debido a las características del canal radio: interferencias, desvanecimientos, variaciones de la capacidad, etc. Es por ello que no resulta una tarea fácil conseguir una planificación eficiente y equitativa de los recursos en este tipo de redes. En este artículo se describen los criterios de planificación en redes inalámbricas de cuarta generación y se lleva a cabo una extensa revisión de los mecanismos de planificación propuestos para las redes TDMA y CDMA.

1. Introducción

El desarrollo de las primeras redes cableadas de alta velocidad supuso grandes posibilidades para nuevas aplicaciones como la videoconferencia, la telemedicina, etc. Estas aplicaciones presentan unos requisitos rigurosos en cuanto a retardo, *jitter* del retardo, tasa de pérdida de paquetes y *throughput*. Por ello se invirtieron grandes esfuerzos con el objetivo de incorporar en las redes cableadas de conmutación de paquetes los mecanismos necesarios para garantizar los requisitos de calidad de servicio de las nuevas aplicaciones emergentes. Los planificadores (en inglés *schedulers*) son elementos fundamentales a la hora de asegurar el cumplimiento de estos requisitos, dado que son los responsables de repartir la capacidad del sistema llevando a cabo la asignación de recursos entre las distintas conexiones.

Cuando extendemos el campo de aplicación de los planificadores a un entorno inalámbrico 4G, caracterizado por la heterogeneidad del tráfico multimedia que transporta, la dificultad de esta tarea se incrementa considerablemente dada la escasez de los recursos radio y la hostilidad y naturaleza cambiante de este canal. Existe una gran número de planificadores para las redes inalámbricas, la mayoría de los cuales se derivan de los algoritmos diseñados para redes cableadas. En este estudio se lleva a cabo una clasificación de los principales tipos de planificadores inalámbricos propuestos en la literatura, y se analizan sus características distintivas. Estos planificadores son evaluados respecto a

un conjunto de criterios que se deben tener en cuenta en las redes inalámbricas 4G, como son la eficiencia, el mantenimiento de la QoS (del inglés *Quality of Service*) de las distintas aplicaciones, la utilización de información de la capa física en el proceso de planificación (diseño intercapa), etc.

Existen diferentes contribuciones en las que se revisan las diferentes propuestas de algoritmos de planificación de recursos. En [1] se realiza una revisión de los principales algoritmos de planificación para redes cableadas. Los análisis realizados en [2] y [3] se centran en redes inalámbricas. Akyildiz et al. en [3], analiza los protocolos MAC inalámbricos sin centrarse específicamente en los algoritmos de planificación, mientras que Fattah y Leung en [2] llevan a cabo un análisis de los algoritmos derivados del GPS (del inglés *Generalized Processor Sharing*), principalmente para redes TDMA (del inglés *Time Division Multiple Access*). A diferencia de las contribuciones anteriores en este estudio se realiza una revisión más exhaustiva de las propuestas existentes para redes inalámbricas, clasificándolas en diferentes familias y analizando sus ventajas y limitaciones. El objetivo final de este estudio es sentar las bases que nos permitan seleccionar y/o diseñar un planificador que sea capaz de explotar todo el potencial de una capa física MIMO-GO-MC-CDMA (MIMO Group Orthogonal MC-CDMA) como la definida en el proyecto MARIMBA (MIMO-MC/OFDM-CDMA Adaptativo para Redes Integradas Móviles de Banda Ancha) [4].

El documento se organiza de la siguiente forma: en primer lugar se establecen aquellos aspectos

que deben considerarse en el diseño de un planificador para redes inalámbricas 4G; a continuación se lleva a cabo una clasificación de los protocolos de planificación inalámbricos, evaluando cualitativamente sus prestaciones; por último se exponen las conclusiones resultantes de este estudio.

2. Criterios de planificación en redes inalámbricas 4G

Los enlaces radio poseen unas características que los diferencian enormemente de los enlaces por cable. Con el objetivo de llevar a cabo un análisis preciso de las técnicas de planificación existentes es necesario definir los aspectos de los entornos inalámbricos que determinan su comportamiento.

- Los canales radio móviles se caracterizan por la presencia de ráfagas de errores provocadas por desvanecimientos, interferencias y ruido, que dependen de la localización del terminal móvil. Los primeros planificadores diseñados para estos entornos modelaban el canal mediante una cadena de Markov discreta con dos estados: canal bueno y canal malo [5, 6]. El canal cambiaba de estado de acuerdo con una determinada matriz de probabilidades de transición. Este modelo resulta del todo inapropiado para ser utilizado en los sistemas inalámbricos actuales 3G-4G, dado que estos presentan la posibilidad de utilizar múltiples modos de transmisión (formados por diferentes esquemas de modulación, codificación, etc.), que permiten adaptarse al estado del canal. Por tanto los planificadores de nueva generación deben ser capaces de manejar un modelo del canal adaptativo con múltiples estados posibles. Es decir, el diseño del planificador no puede ser independiente de la capa física que se sitúa por debajo, se necesita un tipo de diseño intercapa (en inglés *cross-layer design*) que permita un cierto nivel de integración entre ambas capas.
- Por otra parte debe tenerse en cuenta la heterogeneidad del tráfico de las redes multimedia actuales y futuras, lo que representa diferencias considerables en los requisitos generales de QoS para los distintos tipos de servicios. Por tanto, el planificador debe gestionar diferentes tipos de tráfico, garantizando sus requisitos de QoS.
- No debemos olvidar el aspecto de la *justicia* en la asignación de los recursos del sistema. El planificador debe garantizar una repartición justa de los recursos entre las diferentes conexiones.

- La asignación del medio de transmisión debe ser eficiente. Este grado de eficiencia se mide habitualmente en términos de *throughput*.
- A ello se debe añadir, para el caso particular de redes CDMA (del inglés *Code Division Multiple Access*), el fenómeno de la *soft-capacity*, es decir, la variabilidad de la capacidad del sistema en función del nivel de interferencia experimentado. Por tanto el planificador debe ser capaz de trabajar con una capacidad variable que depende entre otros factores de la combinación de servicios elegidos para su transmisión simultánea. En definitiva el planificador debe programar las transmisiones de manera que se incremente la capacidad del sistema y debe asegurar que el número de transmisiones simultáneas previstas no provoque un exceso de interferencia.
- Finalmente debe tenerse en cuenta el grado de complejidad del algoritmo de planificación.

Desafortunadamente, existe un compromiso entre todos estos requisitos. Por ejemplo, debido a la hostilidad y variabilidad del canal radio móvil no podemos conseguir a la vez: una justa distribución de los recursos, la máxima eficiencia del algoritmo en términos de *throughput* y el cumplimiento de los requisitos de QoS de las diferentes conexiones.

3. Clasificación de protocolos de planificación

Los protocolos de planificación para redes de comunicaciones móviles definidos en la literatura pueden agruparse en diferentes familias. La fig. 1 muestra la clasificación realizada que aunque no es exhaustiva incluye una muestra representativa de las principales familias de algoritmos de planificación existentes.

3.1. La familia de protocolos de planificación GPS

3.1.1. Descripción

GPS (*Generalized Processor Sharing*)[7], también conocido como FFQ (del inglés *Fluid-flow Fair Queueing*) [8], es un protocolo eficiente, flexible y justo, propuesto originariamente para un entorno libre de errores.

En un servidor GPS cada sesión tiene asociado un número real, denominado *compartición relativa de servicio*, o también *peso*, ϕ_i . La tasa de transmisión a la que opera el servidor, C , se distribuye entre las distintas sesiones activas del

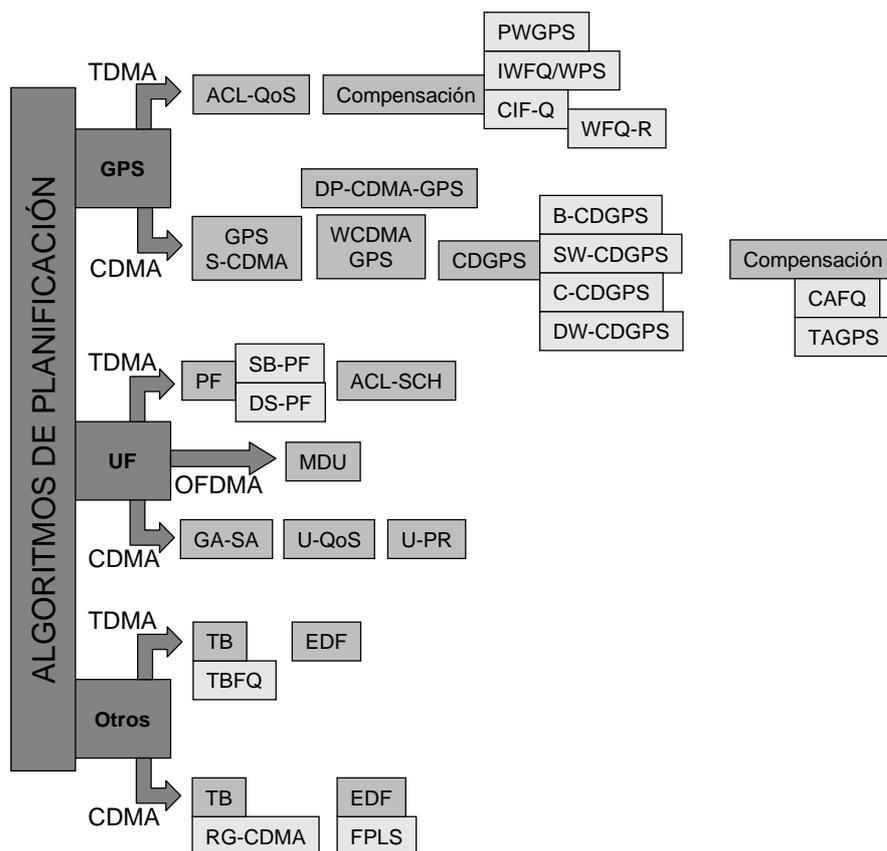


Figura 1: Clasificación de los planificadores para sistemas inalámbricos.

sistema, proporcionalmente a los valores de las ϕ_i .

Se trata de un planificador conservativo (en inglés *work-conserving*), es decir, que estará ocupado siempre que exista información esperando para ser transmitida en el sistema. Además, la diferencia máxima entre el servicio normalizado recibido por dos sesiones activas cualesquiera es cero, de manera que puede afirmarse que el reparto de recursos llevado a cabo por estos planificadores se caracteriza por su absoluta justicia.

Siempre que el tráfico generado por las fuentes sea restringido mediante un *leaky-bucket*, se puede acotar el retardo máximo de encolado. Es por ello que esta familia de algoritmos puede utilizarse para aplicaciones con requerimientos de retardo.

Es importante notar que determinar el valor que deben adoptar los coeficientes ϕ_i no es un problema de fácil solución. Para fijar su valor deben tenerse en cuenta la tasa de transmisión requerida y el retardo máximo que puede tolerarse.

GPS considera a las secuencias de datos como flujos fluidos, de manera que no permite una transmisión orientada a paquetes. Existe, pues, una adaptación que permite su utilización en

esquemas de transmisión de paquetes: se trata del PGPS (del inglés *Packet-by-packet GPS*) [7], conocido también con el nombre de WFQ (del inglés *Weighted Fair Queueing*) [8]. Este algoritmo se basa en la definición del *tiempo virtual del sistema*, concepto que permite seguir el progreso del sistema GPS y llevar a cabo una implementación práctica del PGPS. Cuando un nuevo paquete llega al servidor PGPS se marca con dos tiempos virtuales: el *tiempo virtual de inicio*, t_{vi} , definido como el instante de tiempo en que un paquete iniciaría su transmisión en el sistema GPS, y el *tiempo virtual de finalización*, t_{vf} , que se corresponde con el tiempo en que el paquete finalizaría su transmisión. Estos dos tiempos se calculan a la llegada del nuevo paquete. A continuación se ordenan los paquetes por orden creciente del tiempo virtual de finalización, estableciéndose de esta manera el orden en que los paquetes serán transmitidos. En la figura 2 se representa un ejemplo explicativo de estos algoritmos.

Obviamente esta adaptación del sistema GPS ideal a la transmisión de paquetes, conlleva un importante incremento de la complejidad del sistema, pues debe mantenerse actualizado en todo momento el tiempo virtual del sistema, que debe recalcularse cada vez que una conexión finaliza la transmisión de todos sus paquetes o

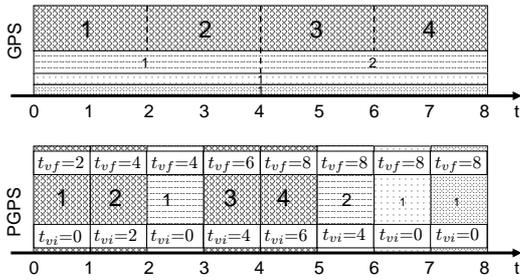


Figura 2: GPS/PGPS y Tiempo Virtual.

bien cuando se incorpora una nueva conexión al sistema.

Un análisis más exhaustivo del comportamiento de los sistemas GPS y PGPS permite apreciar diferencias de comportamiento que pueden llegar a ser considerables en situaciones con características heterogéneas de tráfico y de elevada carga. Por ello se desarrolló el algoritmo WF2Q [9]. En este algoritmo sólo son elegibles para ser transmitidos aquellos paquetes que en el instante de tiempo actual ya habrían empezado a recibir servicio en el sistema GPS correspondiente; de entre todos ellos se elige para ser transmitido aquél que terminaría de ser servido en primer lugar en este sistema GPS de referencia. De esta manera se reducen las diferencias de comportamiento que presentaban los algoritmos GPS y PGPS. Con el objetivo de cuantificar las discrepancias de comportamiento que presenta cualquier algoritmo con respecto al algoritmo GPS se define el *worst-case fair index*. Obviamente, su valor para el algoritmo WF2Q se reduce considerablemente con respecto al algoritmo PGPS.

Existen multitud de variantes derivadas de los algoritmos WFQ y WF2Q. El algoritmo WF2Q-M (del inglés *WF2Q with Maximum Rate Control*) [10] incorpora una modificación que permite limitar la tasa de transmisión máxima para las sesiones que así lo requieran, y lo hace distribuyendo el ancho de banda sobrante entre el resto de sesiones de acuerdo con sus respectivos pesos. Una variante de WFQ, que recibe el nombre DP-WFQ (del inglés *Dynamic Predictive WFQ*) [11], lleva a cabo una distribución del ancho de banda sobrante entre todas las conexiones teniendo en cuenta no tan sólo los pesos GPS sino que además considera lo cerca que está el buffer transmisor de desbordarse.

Existen varias modificaciones de estos algoritmos que persiguen reducir su complejidad. En [12] se modifica la definición de los tiempos virtuales de inicio y finalización, de manera que tan sólo es necesario mantener actualizados sus valores para los paquetes situados a la cabeza

de la cola de cada una de las conexiones, lo que supone una simplificación considerable de los cálculos que deben llevarse a cabo para la implementación del algoritmo, conocido como WF2Q+. Otra variante del algoritmo PGPS que simplifica su ejecución es el SCFQ (del inglés *Self-Clocked Fair Queueing*) [13]. En este sistema, a la llegada de cada paquete, éste es marcado con un *service finish tag*. Para determinarlo se necesita conocer el tiempo virtual, cuya definición se simplifica: en lugar de calcularlo respecto del sistema ideal GPS, se define como el *service tag* del paquete que está siendo servido en el instante de llegada del nuevo paquete. El algoritmo SFQ (del inglés *Start-time Fair Queueing*) [14] lleva a cabo un proceso prácticamente idéntico, con la diferencia de que utiliza el *service start tag*, es decir, se fija en el instante de inicio del servicio en lugar de en su finalización. Los algoritmos SCFQ y SFQ presentan un idéntico nivel de simplificación de la complejidad del algoritmo PGPS.

La extensión de la disciplina de planificación GPS a un entorno inalámbrico supone un verdadero reto. Téngase en cuenta que este tipo de canales se caracteriza, tal como se ha detallado en el Apartado 2, por el hecho de que la capacidad depende de la localización del terminal móvil y del instante temporal. Encontraremos estaciones con un canal nefasto, mientras que otras estaciones pueden tener un buen canal y transmitir. En un entorno de estas características es inevitable plantearse la posibilidad de que cuando una conexión experimente un canal con errores pueda ceder su porción de servicio a otro usuario con un buen canal; es más, ¿debe ser compensado por ello a posteriori?, ¿afectará el esquema de compensaciones a la justicia original del GPS?.

En los siguientes apartados se describirán los protocolos derivados de GPS para redes TDMA y CDMA. Los planificadores GPS requieren un control centralizado. En el enlace descendente la estación de base distribuye la capacidad disponible entre los usuarios móviles. Para el enlace ascendente este proceso se lleva a cabo en conjunción con un esquema de acceso al medio bajo demanda que consiste en que en cada slot de tiempo la estación móvil, mediante el canal de acceso aleatorio o bien a través de su propio canal dedicado, notificará a la estación de base las demandas de transmisión. Entonces el planificador asigna la tasa de transmisión para el siguiente slot, teniendo en cuenta los requisitos de QoS y la capacidad disponible. Los recursos asignados son comunicados a cada estación móvil al inicio del siguiente slot, mediante un canal de control o bien realimentando esta información en un canal dedicado de bajada.

3.1.2. Protocolos derivados de GPS para redes TDMA

Los primeros protocolos derivados de GPS para redes TDMA incorporan esquemas de compensación para gestionar la calidad variable del canal, como por ejemplo IWFQ (del inglés *Idealized Wireless Fair Queueing*) y WPS (del inglés *Wireless Packet Scheduler*) [5]. Para llevar a cabo la compensación se distinguen dos tipos de conexiones: las conexiones que han perdido parte de su servicio debido a las malas condiciones de su canal (etiquetadas como *lagging*) y aquellas que han visto incrementado el servicio que han recibido gracias a que han experimentado un mejor canal (etiquetadas como *leading*). Para ello se mantiene un sistema ideal libre de errores como referencia, que permite determinar en todo momento la diferencia de servicio recibida entre el sistema real y el sistema ideal. La compensación por el defecto de servicio experimentado por las sesiones con un mal canal puede realizarse cuando estas mismas conexiones posean un canal con calidad adecuada, teniendo en cuenta los requisitos de QoS, y que se llevará a cabo a costa de las conexiones *leading*. Se imponen ciertos límites a estas compensaciones, para evitar situaciones de monopolización de recursos por parte de un usuario.

Dado que en los sistemas IWFQ y WPS el paquete elegido para ser transmitido corresponde a aquél que presenta un menor tiempo virtual de finalización de entre las sesiones que experimentan un buen canal, se corre el riesgo de no servir a las sesiones *lagging* con la tasa de transmisión garantizada que les corresponde de acuerdo con la disciplina GPS. Es por ello que presentan deficiencias en cuanto a justicia a corto plazo y retardo. Para solventar este problema se desarrolló el algoritmo CIF-Q (del inglés *Channel Condition Independent Fair Queueing*) [15]. Este esquema toma las decisiones de planificación de acuerdo con el menor tiempo virtual de inicio del sistema libre de errores. Define una variable *lag* que mantiene la información acerca de la diferencia de servicio recibido en el sistema real y en el sistema de referencia ideal. Si el flujo elegido para ser transmitido es *leading* y además ya ha recibido la porción de servicio garantizado (de acuerdo con su peso) entonces se asignará el slot a los flujos *lagging* de manera proporcional a su coeficiente de compartición de servicio, evitando así que se produzcan situaciones con un reparto injusto de recursos.

Con el mismo objetivo surgió el algoritmo PWGPS (del inglés *Packet Wireless GPS*) [16], que incrementa los pesos de las sesiones que han experimentado un canal con errores hasta que consiguen recibir la proporción de servicio

que les corresponde. El planificador redefine el tiempo virtual de finalización de los paquetes de manera que no se requiere el mantenimiento de un modelo libre de errores, lo que permite reducir la complejidad con respecto a CIF-Q manteniendo las mismas propiedades de justicia. Se ordenan los paquetes según su tiempo virtual de finalización, pero debe tenerse en cuenta que su valor no puede calcularse a la llegada de los paquetes, debido a que los pesos son variables y dependen del estado del canal, lo que implica que su valor deberá recalcularse.

Cuando se trata de sistemas que implementan mecanismos de detección de errores y retransmisiones se plantea la necesidad de contabilizar el consumo de recursos que éstas representan. El algoritmo WFQ-R (del inglés *WFQ with link level Retransmission*) [17] lleva a cabo una repartición de los recursos de manera proporcional a los pesos de cada conexión, pero contabilizando las retransmisiones, es decir, no asume una predicción perfecta del canal. Se trata de combinar el algoritmo CIF-Q con un sistema de compensación de los recursos extra correspondientes a las retransmisiones, compensación que puede llevarse a cabo a costa de la sesión que las ha padecido, o bien distribuyéndola entre todas las sesiones de manera proporcional a sus pesos.

El algoritmo ACL-QoS (en inglés *Adaptive Cross-Layer scheduler with prescribed QoS guarantees*) [18] constituye una propuesta reciente que presenta un planificador para redes TDMA con un diseño intercapa. A diferencia de las propuestas anteriores se asume un sistema en el que cada usuario utiliza un esquema de modulación y codificación adaptativo (del inglés *Adaptive Modulation and Coding, AMC*) en la capa física. Se modela el canal mediante una cadena de Markov finita, y de acuerdo con el estado del canal se determinará el esquema de modulación y codificación a utilizar. El planificador distingue dos tipos de tráfico: tráfico con calidad de servicio garantizada y *best-effort*. El reparto de recursos (slots por trama) para el primero, no utiliza la política GPS sino que se lleva a cabo teniendo en cuenta la disponibilidad de datos en el buffer de transmisión y la capacidad de transporte del canal; es decir, el planificador utiliza información de las capas de enlace y física. Los recursos sobrantes se destinan al tráfico *best-effort*, siguiendo la disciplina GPS.

Como hemos indicado, el hecho de que la técnica GPS se base en un modelo de un flujo fluido que asume la posibilidad de que múltiples sesiones sean servidas simultáneamente y además con tasas distintas, representa un problema a la hora de aplicarla a los sistemas TDMA de transmisión de paquetes, en los que no son

posibles las transmisiones paralelas. Ésta es la causa de que en los sistemas reales sea necesario mantener el cómputo de un tiempo virtual, que permite realizar la implementación práctica del sistema GPS fluido ideal, lo que representa un aumento de la complejidad de estos sistemas. Ahora bien, en los sistemas CDMA sí que la transmisión paralela de múltiples servicios utilizando diferentes códigos es algo "natural". Además es posible realizar una compartición de los recursos que puede ajustarse de manera variable a partir del factor de ensanchado, de la potencia asignada, etc.

3.1.3. Protocolos derivados de GPS para redes CDMA

En este subapartado se describen los principales algoritmos de planificación derivados de GPS que se han desarrollado para su implementación en sistemas CDMA. En general se trata de variantes del algoritmo GPS original o de algunos de sus derivados, y llevan a cabo la asignación de recursos trama a trama (sistemas CDMA ranurados, en inglés *Slotted CDMA*).

El B-CDGPS (del inglés *Basic CDGPS*) [19], es un planificador muy simple donde se considera un canal ideal y además no se tienen en cuenta las variaciones de capacidad propias de los sistemas CDMA. Utiliza un umbral inferior de la *soft-capacity* del sistema, lo cual provoca una infrautilización de los recursos y una reducción de la eficiencia. A excepción del B-CDGPS, dada la importancia del fenómeno de la *soft-capacity*, todos los demás algoritmos considerados tratan de incorporar esta característica. Por ejemplo los esquemas GPS S-CDMA (del inglés *GPS in Slotted CDMA*) [20], SW-CDGPS (del inglés *Static Weighted CDGPS*) [19], C-CDGPS (del inglés *Credit based CDGPS*) [19] y el WCDMA GPS, propuesto en [21], consideran una capacidad variable aunque trabajan con un canal ideal. El esquema S-CDMA [20] derivado del PGPS, distribuye los recursos del sistema (potencia y ancho de banda) según los coeficientes de compartición relativa de servicio ϕ_i , en un entorno uncelular. Los requisitos de BER y de retardo de cada conexión permiten ajustar de manera adecuada los valores de potencia asignada y los pesos GPS, suponiendo que cualquier tasa de transmisión es posible. En el caso del SW-CDGPS (del inglés *Static Weighted CDGPS*) [19] se define la *capacidad nominal* de una célula como la cantidad de recursos disponibles en esa célula y que puede determinarse a partir de la medida de la interferencia intercelular. Teniendo en cuenta las demandas de los usuarios y el valor de la capacidad nominal se determinarán los valores de SIR y tasa de transmisión de cada

una de las estaciones móviles, de forma que el *throughput* agregado de todos los usuarios de la celda sea el máximo posible, siempre y cuando se cumpla el criterio de justicia GPS. Para mejorar el *throughput* del SW-CDGPS el algoritmo C-CDGPS (del inglés *Credit based CDGPS*) [19] tolera para las sesiones que no son en tiempo real una cierta pérdida de justicia a corto plazo. Se utiliza un esquema de compensación que asocia a cada flujo un contador de créditos que contabiliza el servicio recibido en comparación con el que correspondería según la justicia GPS. Para evitar situaciones de monopolización de los recursos sobrantes por parte de un usuario, el número de créditos máximo estará limitado, de forma que se asegura la justicia a largo plazo. El número de créditos asociado a un flujo, junto con la disciplina GPS, determinarán la asignación de los recursos del sistema. Otro algoritmo que utiliza el mismo concepto de capacidad nominal es el WCDMA GPS [21] aunque en este caso la distribución de recursos no está basada en el algoritmo GPS básico sino en su variante PGPS y se considera un conjunto de tasas de transmisión discreto.

Como hemos mencionado, los planificadores anteriores consideran un canal ideal. La introducción de un canal no ideal implica, al igual que en el caso TDMA, la incorporación de mecanismos de compensación para garantizar la justicia a largo plazo de las conexiones con un mal canal. Algunos ejemplos de este tipo de mecanismos son DW-CDGPS (del inglés *Dinamic Weighted CDGPS*) [19], CAFQ (del inglés *Channel Adaptive Fair Queueing*) [22] y TAGPS (del inglés *Traffic Aided GPS*) [6]. La propuesta de DW-CDGPS [19] primero planifica las transmisiones de todos los usuarios en tiempo real utilizando SW-CDGPS. A continuación la capacidad sobrante será repartida entre el resto de usuarios. Para mejorar el *throughput* del tráfico en tiempo no real el algoritmo ajusta en cada slot temporal las ϕ_i s de este tipo de tráfico de manera dinámica favoreciendo los flujos afectados por un buen canal. Se debe llegar a un compromiso entre la eficiencia y el cumplimiento de los requisitos de justicia a corto plazo. En el algoritmo CAFQ [22] se utiliza un sistema libre de errores como referencia, en el que el proceso de planificación se realiza de acuerdo con el algoritmo SFQ [14]. En el sistema real se ordenan las sesiones de acuerdo con su número de créditos (primero las que han recibido menos servicio que el que les correspondería en el sistema de referencia), y se servirán las conexiones con un canal que permita las transmisiones. Existe una excepción: en el caso en que el sistema de referencia escoja para su transmisión una conexión *lagging* con un canal libre de errores, se respetará esa elección. En esta misma línea está el algoritmo TAGPS [6]. Del

mismo modo que en el algoritmo GPS S-CDMA [20], se utiliza el concepto de *índice de potencia*, g_i , para representar los recursos del sistema. Se distinguen dos tipos de usuarios, los de voz, con una tasa de transmisión constante, y los de datos, con tasa de transmisión variable. La asignación de recursos consiste en determinar los valores de las g_i s de cada usuario de datos, lo cual nos permitirá determinar su tasa de transmisión (se considera que cualquier tasa es posible). Cuando un usuario *lagging* perciba un buen canal, el algoritmo asignará a esa conexión un incremento de su peso GPS (directamente relacionado con un incremento de su índice de potencia), de forma que recibirá una compensación de servicio, y lo hará a expensas de los usuarios *leading*.

Podemos concluir que, aunque los algoritmos de compensación incorporan información de la capa física en el proceso de planificación (diseño intercapa), utilizan modelos muy simples del canal (en general modelos on-off). Además debe tenerse en cuenta que los mecanismos de compensación de servicio implican un verdadero problema a la hora de asegurar el cumplimiento de los requisitos de retardo máximo y suponen un compromiso para el cumplimiento de la justicia GPS. Es importante mencionar también que en los planificadores basados en GPS las garantías de servicio de una sesión se controlan mediante un solo parámetro: el peso ϕ_i . Se produce, pues, un acoplamiento entre retardo y ancho de banda, ya que para garantizar retardos pequeños es necesario reservar una porción elevada del ancho de banda disponible en el sistema. Este comportamiento conlleva una infrautilización y desperdicio de los recursos, ya que en las redes multimedia nos encontramos con servicios multitasa con especificaciones de retardo muy diversas y distintos requisitos de ancho de banda. El algoritmo DP-CDMA-GPS (del inglés *Dynamic Programming CDMA-GPS*) [23] trata de resolver esta limitación, asignando unos pesos variables en el tiempo, y para ello utiliza una función de coste que representa el retardo de encolado experimentado por las estaciones móviles y que se debe minimizar. Esta función de coste tiene en cuenta los requisitos de QoS de las distintas conexiones y también la calidad del canal físico para cada una de ellas (aunque no considera la interferencia intercelular). Otra de las ventajas de este algoritmo es que considera un conjunto discreto de posibles tasas de transmisión.

3.2. Planificadores basados en funciones de utilidad

La teoría de utilidad, proveniente de economía, puede usarse en sistemas de comunicaciones para evaluar el grado de satisfacción que proporciona la red a los requerimientos de las aplicaciones del usuario, en lugar de considerar parámetros

del sistema como por ejemplo el *throughput* o la tasa de pérdida de paquetes [24]. Diferentes aplicaciones tendrán diferentes funciones de utilidad o simplemente diferentes parámetros en esas funciones. Por ejemplo las funciones de utilidad del tráfico *best-effort* serán en función del *throughput*, mientras que las de las aplicaciones sensibles al retardo deberán ser en función de este parámetro. Cómo encontrar las funciones de utilidad más adecuadas para los diferentes tipos de aplicaciones es uno de los elementos clave para el buen funcionamiento de este tipo de técnicas. Ejemplos de diferentes funciones de utilidad para diferentes tipos de tráfico pueden encontrarse en [25, 26, 27].

Una vez definidas las funciones de utilidad de las diferentes aplicaciones existentes en el sistema, la optimización de la planificación consiste en maximizar la utilidad agregada de todo el sistema teniendo en cuenta las limitaciones de la capa física subyacente, es decir, consiste en encontrar la combinación de tasas, retardos, etc., que hace que la utilidad sea máxima dentro de los límites fijados por la capa física. Dado que en general las funciones de coste son no lineales, el problema se convierte en un problema de maximización combinatoria no lineal de una función de coste (que es la utilidad agregada del sistema). Llevar a cabo esta optimización es un problema complejo y los algoritmos utilizados constituyen una parte esencial de la planificación basada en funciones de utilidad y en general son computacionalmente costosos. Además estos algoritmos deben tener en cuenta que no todas las soluciones son posibles ya que habitualmente las capas físicas trabajan con un conjunto discreto de tasas de transmisión o retardos posibles. Por ejemplo en [28] se definen un conjunto de algoritmos válidos para redes OFDM aunque algunos de ellos serían fácilmente adaptables también para redes CDMA .

Dado que este tipo de planificación requiere un control centralizado, este tipo de soluciones se ha propuesto en muchos casos en el enlace descendente. Al igual que en el caso de los algoritmos derivados del GPS su utilización en el enlace ascendente implica el uso de un protocolo MAC de asignación previa demanda. A continuación se presenta una clasificación de algoritmos basados en funciones de utilidad.

3.2.1. Planificadores proporcionalmente justos (Proportional Fair)

Se trata de un algoritmo ideado para tráfico de datos *best-effort*. De acuerdo con el algoritmo PF (del inglés *Basic Proportional Fair*), para cada estación móvil se monitorizan dos parámetros: en primer lugar la DRC_i , que es la tasa máxima a la

que puede transmitirse en el siguiente intervalo de planificación según la SINR medida, y en segundo lugar la R_i , que es la tasa promedio a la que se ha estado transmitiendo. Entonces el scheduler selecciona para transmitir el móvil con un cociente DRC_i/R_i mayor, que transmitirá con una tasa DRC_i . De hecho puede demostrarse que este sistema es equivalente a utilizar una función de utilidad logarítmica $U_i(R_i) = \ln(R_i)$ y que, para este tipo de función de utilidad, la optimización de la utilidad agregada del sistema puede obtenerse asignando a cada momento el recurso al usuario con un DRC_i/R_i mayor. Se trata por tanto de un ejemplo donde gracias al tipo de función de utilidad utilizada la optimización de la función de coste es muy simple. Gracias a su simplicidad y eficiencia este tipo de algoritmo se utiliza en sistemas reales, como es el caso de los canales HDR de CDMA2000.

Como puntos fuertes de este algoritmo debemos señalar que utiliza información de la capa física para la toma de decisiones de scheduling, se trata por tanto de un *sistema intercapa* ya que tiene en cuenta el estado del canal priorizando los usuarios con un mejor canal. Además este algoritmo asegura el cumplimiento del criterio de justicia proporcional dado que la selección de un usuario incrementará su *throughput* y, por lo tanto, reducirá su prioridad. Puede trabajar con un conjunto discreto de tasas de transmisión como se muestra en el ejemplo descrito en [29] para un canal TDMA dentro de un sistema CDMA.

Su principal inconveniente proviene del uso de una función de utilidad que sólo depende del *throughput*, lo que hace que este sistema sea incapaz de cumplir con los requisitos de retardo máximo aceptable para las aplicaciones sensibles a este parámetro. Para tratar de resolver esta limitación existen propuestas [30, 31] que mantienen la función de utilidad logarítmica para el tráfico *best-effort* i definen nuevas funciones de utilidad para datos sensibles al retardo. En el algoritmo DS-PF (del inglés *Delay Sensitivity PF*) [30] las sesiones en tiempo real incrementan su prioridad cuando se ha superado un retardo máximo, mientras que en la propuesta SB-PF (del inglés *Sender Buffer PF*) [31] se intenta evitar que el buffer del receptor se vacíe para las aplicaciones en tiempo real. En estos casos mixtos los algoritmos de optimización de la utilidad agregada dejan de ser tan simples como en el caso del algoritmo PF básico. Por este motivo, y para preservar la simplicidad, tanto Barriac en [30] como Koto en [31] proponen algoritmos de optimización sub-óptimos.

3.2.2. Planificadores basados en funciones de utilidad multimedia

Existen diferentes propuestas (GA-SA [25], MDU [26], U-PR [27], U-QoS [32], ACL-SCH [33]) en las que se definen diferentes funciones de utilidad para cada tipo de tráfico y algoritmos que permitan maximizar de forma eficiente la utilidad agregada del sistema, tanto en el caso de acceso TDMA [33], como CDMA [25, 27, 32] o OFDMA[26]. En [26] además se proporciona un marco teórico donde se discute la eficiencia, justicia y estabilidad de los mecanismos de gestión de recursos basados en funciones de utilidad.

3.3. Otros algoritmos de planificación

A continuación se describen algunos planificadores que no pueden ser incluidos en ninguna de las dos familias anteriores.

3.3.1. Algoritmos basados en Token Bucket

El concepto de *Token Bucket* constituye un mecanismo de control de tráfico que se basa en el símil de un *cubo de testigos*. Por cada testigo se permitirá transmitir una determinada unidad de tráfico, como puede ser un paquete. Entonces, un flujo podrá ser transmitido siempre que dispongamos de testigos suficientes dentro del cubo, que se va rellenando con nuevos testigos de acuerdo con la tasa media de transmisión deseada. El máximo número de testigos almacenables en el cubo determinará la tasa de pico permitida para un flujo dado. Este concepto es utilizado en ciertos algoritmos de planificación, tanto en sistemas TDMA como CDMA.

El algoritmo TBFQ (del inglés *Token Bucket Fair Queuing*) [34] fue diseñado para sistemas TDMA. Para poder hacer frente al comportamiento en ciertos casos impredecible del tráfico multimedia, permite violaciones del perfil de tráfico siempre que se disponga de suficiente ancho de banda y no se exceda el contrato a largo plazo. Para ello cada estación móvil dispondrá de un *token bucket* y de un crédito máximo de *tokens* que puede tomar prestados. Se dispone de un banco de *tokens*, al que cada estación contribuirá cuando disponga de *tokens* sobrantes, o del que podrá tomar *tokens* prestados cuando los necesite. Cada estación tiene asociado un contador que contabiliza las aportaciones/préstamos que ha realizado. Este contador determinará, junto con la tasa de generación de *tokens* de cada conexión, las prioridades para el préstamo de *tokens*.

El protocolo RG CDMA (del inglés *Rate Guarantee in CDMA*) [35] multiplexa el tráfico en el

enlace descendente de un sistema W-CDMA utilizando códigos OVSF (Orthogonal Variable Spreading Factor). Cada conexión tiene un ancho de banda garantizado y un número de créditos asociado. En cada slot temporal se incrementa el número de créditos (existe un valor máximo) según el ancho de banda garantizado y se decrementa por cada paquete recibido. De esta manera el número de créditos representa la diferencia entre las transmisiones garantizadas y las realmente realizadas. Se ordenan las conexiones según su número de créditos, de manera que los códigos OVSF asignados a las conexiones con un mayor número de créditos presentan una mayor tasa de transmisión. Este algoritmo no tiene en cuenta el fenómeno de la *Soft Capacity* y considera el canal como ideal.

3.3.2. Algoritmos basados en EDF

En el apartado 3.1.1 se ha descrito el algoritmo DP-WFQ [11], que en el proceso de planificación considera lo cerca que está el buffer transmisor de desbordarse. Siguiendo en esta misma línea encontramos, para las redes TDMA, el algoritmo EDF (del inglés *Earliest Deadline First*) [36], en que cada paquete se marca con su tiempo de expiración, cuyo cálculo vendrá determinado, entre otros factores, por los requisitos de retardo del paquete según la clase de tráfico de que se trate.

Para el caso de redes mixtas TD-CDMA el algoritmo FPLS (del inglés *Fair Packet Loss Sharing*) [37] planifica la transmisión de paquetes multimedia de tal forma que la pérdida de paquetes se distribuya entre todos los usuarios de manera justa y de acuerdo con sus requisitos de QoS. Para ello se ordenan los paquetes según el tiempo restante para su expiración. Si la capacidad disponible en el sistema no permite la transmisión de todos los paquetes que así lo requieren, se procede a distribuir las pérdidas (paquetes expirados) entre las distintas conexiones. Ello se lleva a cabo teniendo en cuenta los requisitos de BER (Bit Error Rate) de cada conexión, que determinarán cuáles serán los paquetes descartados. Las principales deficiencias de este algoritmo es que no considera la interferencia intercelular y supone un canal ideal.

4. Conclusiones

La primera conclusión que puede extraerse de este estudio es que resulta imposible conseguir un planificador para redes inalámbricas que sea a la vez: justo, eficiente, que cumpla con los requisitos de retardo de los servicios en tiempo real y que no sea demasiado complejo. Los compromisos existentes entre estos requisitos deben tenerse

muy en cuenta en el momento de diseñar un nuevo planificador. Por otra parte es importante mencionar que todos los planificadores descritos en este documento forman parte de un protocolo de acceso al medio (MAC) que requiere un nodo central donde se realiza el proceso de planificación. Existen otras alternativas de protocolos MAC distribuidos más simples donde no existe un proceso de planificación; sin embargo estas alternativas no alcanzan los niveles de eficiencia, justicia y cumplimiento de la QoS que se obtienen en estos esquemas centralizados. En el enlace descendente la estación de base actúa de nodo central distribuyendo la capacidad disponible entre los usuarios móviles. Para el enlace ascendente este proceso también se lleva a cabo en la estación de base, pero requiere un esquema de acceso al medio bajo demanda que incrementa la complejidad y nivel de señalización del proceso.

Otro aspecto general a tener en cuenta es que los nuevos sistemas inalámbricos cuentan con capas físicas adaptativas con múltiples modos de transmisión. Por lo tanto es necesario que los planificadores sean capaces de gestionar e interactuar con estas capas físicas más flexibles. Además en el caso de los sistemas CDMA debe tenerse en cuenta el fenómeno de la *soft capacity*: es decir, el planificador debe tener en cuenta que sus decisiones de planificación afectan a la capacidad del sistema.

Con respecto a la familia de protocolos derivados del GPS pueden resaltarse los siguientes aspectos:

- Los recursos son distribuidos proporcionalmente a unos pesos asignados a cada conexión. Sin embargo, en el caso de un entorno radio, la existencia de un canal no fiable hace que estos algoritmos tengan que incorporar mecanismos de compensación para conseguir mejores niveles de justicia y eficiencia aunque eso supone un aumento de la complejidad.
- En el caso en que se utilice un conformador de tráfico tipo *leaky bucket* se puede acotar el retardo máximo de los paquetes. Sin embargo dado que en estos planificadores las garantías de servicio de una sesión se controlan mediante el peso, ϕ_i , se produce un acoplamiento entre retardo y ancho de banda, ya que para garantizar retardos pequeños es necesario reservar una porción elevada del ancho de banda disponible en el sistema. Este comportamiento conlleva una infrautilización de los recursos, sobretudo en redes multimedia. Por este motivo no resulta adecuado utilizar un planificador de tipo GPS para gestionar de manera conjunta flujos de tráfico heterogéneos. De hecho en muchas de las

propuestas analizadas el planificador proporciona un tratamiento diferenciado al tráfico en tiempo real priorizándolo por encima del tráfico de datos.

- Uno de los problemas más significativos de este tipo de algoritmos es la determinación de los pesos de las conexiones, ya que se requiere una correcta asignación de estos valores para conseguir un buen funcionamiento del algoritmo. Sorprendentemente en muchas de las propuestas analizadas no se aborda este problema.
- Por otra, parte en el caso de que quiera planificarse la transmisión a nivel de paquetes, es necesario introducir el concepto de tiempo virtual, cuyo cómputo incrementa la complejidad del sistema. Dada la imposibilidad de realizar transmisiones simultáneas en TDMA los algoritmos GPS para estos sistemas requerirán siempre el uso de tiempos virtuales. Algunos de los algoritmos analizados utilizan simplificaciones en el cálculo de estos tiempos a costa de una relajación en su concepto de justicia.
- En los sistemas CDMA existe un conjunto discreto de tasas de transmisión posible. Este hecho no es tenido en cuenta en muchas de las soluciones analizadas, en las que se trabaja con el modelo de fluidos del GPS básico que supone un valor continuo en las tasas de transmisión.

En cuanto a los planificadores que utilizan el concepto económico de funciones de utilidad, éstos tratan de evaluar el grado de satisfacción que proporciona la red a los requerimientos de las aplicaciones del usuario, en lugar de considerar parámetros del sistema como el *throughput* o la tasa de pérdida de paquetes. Por ejemplo, si un pequeño incremento en la tasa de transmisión de una aplicación multimedia implica que ésta pueda empezar a llevarse a cabo, la utilidad de ese incremento es muy elevada, mientras que si se trata de una aplicación de datos *best-effort* con una tasa de transmisión elevada, un pequeño incremento no se traduce en un beneficio significativo. Cómo encontrar las funciones de utilidad más adecuadas para los diferentes tipos de aplicaciones es uno de los elementos clave para el buen funcionamiento de este tipo de técnicas. Su principal inconveniente es que la maximización de la utilidad agregada del sistema implica la resolución de sistemas no lineales, proceso computacionalmente muy costoso. Por tanto la pieza clave de estas propuestas es el diseño de un algoritmo de cálculo eficiente. Existen algunas excepciones en las que debido al tipo de función de utilidad empleada se genera una utilidad agregada que puede resolverse con un algoritmo muy simple. Éste es

el caso de la familia de algoritmos PF. Por otra parte es importante mencionar que estos planificadores permiten la introducción de información de la capa física (diseño intercapa). Lo más adecuado es introducir esta información a través de restricciones impuestas a la utilidad agregada del sistema [26]. La inclusión de esta información directamente dentro de la función de coste implica recalcular el orden de transmisión cada vez que un usuario experimenta un cambio en su canal [33], lo que hace esta opción computacionalmente inviable.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia y el FEDER dentro del proyecto MARIMBA (TEC2005-0997), y una beca Ramón y Cajal.

Referencias

- [1] H. Zhang, "Service disciplines for guaranteed performance service in packet switching networks," *Proceedings of the IEEE*, vol. 83, no. 10, pp. 1374–1396, October 1995.
- [2] H. Fattah and C. Leung, "An Overview of Scheduling Algorithms in Wireless Multimedia Networks," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 76–83, October 2002.
- [3] I. Akyildiz, J. McNair, L. Carrasco, R. Puigjaner, and Y. Yesha, "Medium access control protocols for multimedia traffic in wireless networks," *IEEE Network Magazine*, vol. 13, no. 4, pp. 39–47, July-August 1999.
- [4] F. Riera-Palou, G. Femenias, and J. Ramis, "Performance of downlink group-orthogonal multicarrier systems," in *Personal Wireless Communications, LNCS4217*, IFIP. Springer, September 2006, pp. 389–400.
- [5] S. Lu, V. Bharghavan, and R. Srikant, "Fair Scheduling in Wireless Packet Networks," *IEEE/ACM Transactions on Networking*, vol. 7, no. 4, pp. 473–489, August 1999.
- [6] D. Liao and L. Li, "Traffic aided fair scheduling using compensation scheme in CDMA cellular networks," *International Conference on Communications, Circuits and Systems 2005*, vol. 1, pp. 363–367, May 2005.
- [7] A. Parekh, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344–357, June 1993.

- [8] A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm," *Proceedings of ACM SIGCOMM'89*, pp. 1–12, 1989.
- [9] J. Bennet and H. Zhang, "WF2Q: Worst-case Fair Weighted Fair Queueing," *IEEE INFOCOM 1996*, vol. 1, pp. 120–128, March 1996.
- [10] J. Lee, M. Chen, and Y. Sun, "WF2Q-M : a worst-case fair weighted fair queueing with maximum rate control," *IEEE GLOBECOM 2002*, vol. 2, pp. 1576–1580, November 2002.
- [11] J. Gallardo and D. Makrakis, "Dynamic predictive weighted fair queueing for differentiated services," *IEEE ICC 2001*, vol. 8, pp. 2380–2384, June 2001.
- [12] J. Bennett and H. Zhang, "Hierarchical packet fair queueing algorithms," *IEEE/ACM Transactions on Networking*, vol. 5, no. 5, pp. 675–689, October 1997.
- [13] S. Golestani, "A self-clocked fair queueing scheme for broadband applications," *IEEE INFOCOM 1994*, vol. 2, pp. 636–646, June 1994.
- [14] P. Goyal, H. Vin, and H. Cheng, "Start-time fair queueing: a scheduling algorithm for integrated services packet switching networks," *IEEE/ACM Transactions on Networking*, vol. 5, no. 5, pp. 690–704, October 1997.
- [15] T. Ng, I. Stoica, and H. Zhang, "Packet fair queueing algorithms for wireless networks with location-dependent errors," *IEEE INFOCOM 1998*, vol. 3, pp. 1103–1111, March-April 1998.
- [16] M. Jeong, H. Morikawa, and T. Aoyama, "Wireless packet scheduler for fair service allocation," *IEEE APCC/OECC 1999*, vol. 1, pp. 794–797, October 1999.
- [17] N. Kim and H. Yoon, "Packet fair queueing algorithms for wireless networks with link level retransmission," *IEEE CCNC 2004*, pp. 122–127, January 2004.
- [18] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer scheduling with prescribed qos guarantees in adaptive wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 5, pp. 1056–1066, May 2005.
- [19] L. Xu, X. Shen, and J. Mark, "Dynamic Fair Scheduling With QoS Constraints in Multimedia Wideband CDMA Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 60–73, January 2004.
- [20] M. Arad and A. Leon-Garcia, "A generalized processor sharing approach time to scheduling in hybrid CDMA/TDMA," *IEEE INFOCOM 1998*, vol. 3, pp. 1164–1171, March 1998.
- [21] X. Wang, "An FDD Wideband CDMA MAC Protocol with Minimum-Power Allocation and GPS-Scheduling for Wireless Wide Area Multimedia Networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 1, pp. 16–28, January-February 2005.
- [22] L. Wang, Y. Kwok, W. Lau, and V. Lau, "Channel Adaptive Fair Queueing for Scheduling Integrated Voice and Data Services in Multicode CDMA Systems," *IEEE WCNC 2003*, vol. 3, pp. 1651–1656, March 2003.
- [23] A. Stamoulis, N. Sidiropoulos, and G. Giannakis, "Time-varying fair queueing scheduling for multicode CDMA based on dynamic programming," *IEEE Transactions on Wireless Communications*, vol. 3, no. 2, pp. 512–523, March 2004.
- [24] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks. Part I: Theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, March 2005.
- [25] W. Zhao and M. Lu, "CDMA downlink rate allocation for heterogeneous traffic based on utility function: GA-SA approach," *Proc. CNSR*, pp. 156 – 162, May 2004.
- [26] G. Song, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, pp. 127–134, December 2005.
- [27] X. Duan, Z. Niu, D. Huang, and D. Lee, "A dynamic power and rate joint allocation algorithm for mobile multimedia DS-CDMA networks based on utility functions," *Proc. PIMRC*, vol. 3, pp. 1107–1111, September 2002.
- [28] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks. Part II: algorithm development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, March 2005.
- [29] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *IEEE VTC*, vol. 3, pp. 1854–1858, May 2000.
- [30] G. Barriac and J. Holtzman, "Introducing Delay Sensitivity into the Proportional Fair

- Algorithm for CDMA Downlink Scheduling,” *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications*, vol. 3, pp. 652–656, September 2002.
- [31] H. Koto, M. Fukushima, S. Nomoto, and F. Takahata, “Scheduling Algorithm based on Sender Backlog for Real-Time Application in Mobile Packet Networks,” *IEEE WCNC 2005*, vol. 1, pp. 151–157, March 2005.
- [32] S. Shen and C. Chang, “A utility-based scheduling algorithm with differentiated QoS provisioning for multimedia CDMA cellular networks,” in *Proc. IEEE VTC’04 Spring*, vol. 3, May 2004, pp. 1421–1425.
- [33] K. Johansson and D. Cox, “An adaptive cross-layer scheduler for improved QoS support of multiclass data services on wireless systems,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 334–343, February 2005.
- [34] W. Wong, H. Zhu, and V. Leung, “Soft QoS provisioning using the token bank fair queuing scheduling algorithm,” *IEEE Wireless Communications*, vol. 10, no. 3, pp. 8–16, June 2003.
- [35] A. Kam, T. Minn, and K. Siu, “Supporting Rate Guarantee and Fair Access for Bursty Data Traffic in W-CDMA,” *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 11, pp. 2121–2130, November 2001.
- [36] Q. Pang, A. Bigloo, V. Leung, and C. Scholfield, “Service scheduling for general packet radio service classes,” *IEEE WCNC 1999*, vol. 3, pp. 1229–1233, September 1999.
- [37] V. Huang and W. Zhuang, “QoS-Oriented Packet Scheduling for Wireless Multimedia CDMA Communications,” *IEEE Transactions on mobile computing*, vol. 3, no. 1, pp. 73–85, January-March 2004.