

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 376 942**

51 Int. Cl.:
G01R 31/08 (2006.01)
G06F 11/00 (2006.01)
G08C 15/00 (2006.01)
H04J 1/16 (2006.01)
H04J 3/14 (2006.01)
H04L 1/00 (2006.01)
H04L 12/26 (2006.01)
H04L 12/28 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06785133 .7**
96 Fecha de presentación: **20.06.2006**
97 Número de publicación de la solicitud: **1913744**
97 Fecha de publicación de la solicitud: **23.04.2008**

54 Título: **PLANIFICACIÓN DE CALIDAD DE SERVICIO (QoS) SENSIBLE A LATENCIA DE PAQUETES INTEGRADA QUE USA EQUIDAD PROPORCIONAL Y PUESTA EN COLA EQUITATIVA PONDERADA PARA SERVICIOS DE PAQUETES MULTIMEDIA INTEGRADOS INALÁMBRICOS.**

30 Prioridad:
12.08.2005 US 707558 P
16.11.2005 US 274297

45 Fecha de publicación de la mención BOPI:
21.03.2012

45 Fecha de la publicación del folleto de la patente:
21.03.2012

73 Titular/es:
**CELLCO PARTNERSHIP D/B/A VERIZON
WIRELESS
ONE VERIZON WAY
BASKING RIDGE, NJ 07920-1097, US**

72 Inventor/es:
KAKADIA, Deepak, Kumar

74 Agente/Representante:
Pons Ariño, Ángel

ES 2 376 942 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Planificación de calidad de servicio (QoS) sensible a latencia de paquetes integrada que usa equidad proporcional y puesta en cola equitativa ponderada para servicios de paquetes multimedia integrados inalámbricos.

5

Solicitudes relacionadas

Esta solicitud reivindica el beneficio de la Solicitud Provisional de Estados Unidos No. 60/707.558 presentada el 12 de agosto de 2005 titulada "INTEGRATED PACKET LATENCY AWARE QOS SCHEDULING ALGORITHM USING PROPORTIONAL FAIRNESS AND WEIGHTED FAIR QUEUING FOR IMS SERVICES IN 3 G CDMA2000 1 X EV-DO AND IP NETWORKS", cuya descripción también se incorpora en su totalidad en este documento como referencia.

10

Campo técnico

15

La presente invención se refiere a técnicas y equipo para mejorar la asignación de recursos y, por lo tanto, la calidad de servicio para comunicaciones urgentes, tales como servicios multimedia integrados, a través de combinaciones de redes inalámbricas y de paquetes que implementan múltiples algoritmos de planificación.

20

Antecedentes

Los Servicios Multimedia Integrados (IMS) están siendo desplegados en redes inalámbricas públicas de tipo 3G cdma2000 conectadas a redes de Protocolos de Internet (IP). Sin embargo, las redes inalámbricas e IP utilizan ciertas estrategias de planificación de calidad de servicio (QoS) incongruentes, que dan como resultado decisiones de planificación de paquetes priorizada sub-óptimas. Por ejemplo, en la planificación en enlace descendente de paquetes en redes celulares cdma2000 1xEV-DO, existe un claro alineamiento erróneo debido a los dos dominios de QoS distintos de las redes inalámbricas e IP por cable. Las redes celulares inalámbricas públicas típicamente usan algoritmos de QoS basados en Equidad Proporcional (PF). La PF se ocupa de decidir qué paquete transmitir en un intervalo de tiempo particular en un único canal de banda ancha compartido, en base a la asignación equitativa de ancho de banda y de maximizar el rendimiento del sistema. Por el contrario, las redes IP que forman Internet y diversas Intranets por cable típicamente usan algoritmos de QoS basados en Intercambio de Procesador Generalizado (GPS), en particular Puesta en Cola Equitativa Ponderada (WFQ), que decide qué paquete transmitir en un intervalo de tiempo particular de un puerto de salida en base a la asignación equitativa de ancho de banda y minimizar el retardo de flujo promedio para tráfico de buen comportamiento o vigilado. La integración a ciegas de estas dos redes, con sus mecanismos de planificación asociados, puede dar como resultado asignaciones de recursos sub-óptimas que conducen a retardos excesivos e innecesarios para algunos usuarios.

25

30

35

40

45

50

La figura 1 representa una implementación típica de una red 3GPP2 1xEV-DO. El dibujo muestra elementos de la red en forma de diagrama de bloques funcionales de alto nivel, y muestra algunos aspectos del procesamiento implicado en las comunicaciones a través de los elementos ilustrados. El dispositivo móvil, algunas veces denominado como un Agente de Usuario (UA) móvil, se comunica a través de una Estación Transmisora Base (BTS), seleccionada entre aquellas que puede detectar el dispositivo móvil (aproximadamente dentro de alcance) por el aire, terminando con la BTS con la mejor relación Canal con respecto a Interferencia. La Interfaz de Datos de Paquete de Alta Velocidad (HPRD) en este segmento o dominio de red inalámbrica es la conexión de red de capacidad más cara y estrecha entre todos los segmentos (representados de forma temática mediante los tubos de diversos tamaños/anchos de banda) que llevarán la comunicación del UA. El planificador de paquetes en este segmento puede residir en un módulo DOM (Módulo de Datos Optimizado) como ejemplo de una implementación. El DOM, típicamente en forma de una tarjeta que encaja en una Estación Transmisora Base (BTS), comunica con el dispositivo móvil del UA con un enlace por aire usando un espectro de frecuencia específico. El planificador en el DOM optimiza el rendimiento del sistema en base a un algoritmo de equidad proporcional (PF), representado por el bloque funcional en el diagrama.

55

60

El siguiente segmento conecta la BTS al Controlador de Red de Radio (RNC), situado en el Centro de Conmutación de Servicios Móviles (MSC). Este segmento inalámbrico tradicionalmente utilizaba circuitos T1 basados en TDMA, pero el segmento está evolucionando actualmente a utilizar conexiones Metro Ethernet proporcionadas por proveedores de Servicio Regional. Estos enlaces son típicamente de 10/100 Mbs de capacidad. La conexión metro Ethernet en el lado del MSC es típicamente del orden de un gigabit de capacidad de enlace. El planificador de paquetes por cable en el RNC, en el enlace directo típicamente usa una variante de Puesta en Cola Equitativa Ponderada (WFQ), representada por el bloque funcional en el diagrama. La Puesta en Cola de Baja Latencia (LLQ) se implementa a menudo, dado que combina una cola de prioridad estricta con WFQ para soportar tráfico en tiempo real. El resto de los segmentos de red por cable usan todos alguna forma de planificador de paquetes de WFQ, como se muestra mediante los bloques del planificador de WFQ en el diagrama. La WFQ está diseñada para minimizar la latencia promedio para todos los flujos. Aunque se muestran por separado con fines ilustrativos, las funciones de PF y planificador de WFQ típicamente son aspectos funcionales de los elementos de enrutamiento

relevantes, tales como aquellos en o asociados con el DOM y el RNC.

El segmento de red por cable entre la red de retroceso IP y la HPRD, en el enlace directo (flecha que representa la comunicación de tráfico desde la red central IP que va al UA de la estación móvil), tiene un claro alineamiento erróneo de optimizaciones. Este segmento se extiende fuera de e incluye una parte del Módulo de Datos Optimizado (DOM), en la BTS. El DOM implementa un algoritmo especial para transmitir/(recibir) datos a/(de) los dispositivos móviles mediante el dominio de la red inalámbrica, incluyendo el enlace o enlaces por aire a y desde los dispositivos móviles de UA. Se ha descrito que el DOM implementa un servicio de comunicación de paquetes inalámbrico de tipo 1xEV-DO - un tipo de protocolo CDMA para la transmisión de datos de paquetes de alta velocidad para redes móviles. El planificador de paquetes de tipo PF inalámbrico implementado en el módulo DOM intenta maximizar el rendimiento del sistema, mientras que los planificadores de paquetes por cable de tipo WFQ usados para el envío de paquetes al DOM minimizan la latencia promedio.

La parte central de la red puede implementarse de diversas maneras diferentes, que proporcionarán una capacidad de transporte adecuada para el tráfico de paquetes IP. Para los fines de mostrar un ejemplo completo, el centro se muestra usando el envío rápido y eficaz de paquetes por Conmutación Multi-Protocolo mediante Etiquetas (MPLS) en un transporte de tipo célula de modo de transferencia asíncrona (ATM). La parte inferior del dibujo muestra las pilas de protocolos para una implementación ejemplar de la red ilustrada. Los expertos en la materia reconocerán que diversas redes pueden utilizar éstas u otras combinaciones de protocolos de comunicaciones.

Las figuras 2A a 2C representan con más detalle, problemas ejemplares que pueden surgir cuando se conecta una red por cable a una red inalámbrica. La figura 2A muestra el caso general en el que dos abonados de igual prioridad experimentan diferente congestión y diferentes cantidades de tráfico en los enlaces directos. Los paquetes enlazados para cada abonado se ponen en cola en un enrutador en la parte por cable de la red, que implementa planificación WFQ. El enrutador selecciona paquetes de las diversas colas y pasa los paquetes, según lo planificado, a la parte inalámbrica de la red. En la parte inalámbrica de la red, los paquetes son puestos en cola mediante un elemento tal como el DOM que utiliza el algoritmo de planificación de PF. La BTS transmite paquetes desde las colas, según lo planificado mediante la aplicación del algoritmo de PF, en el enlace por aire a los dispositivos cliente de la estación móvil de los respectivos abonados.

En el ejemplo, el dispositivo cliente del abonado 1 tiene una conexión a distancia con un servidor a través de un enlace muy descongestionado y, por lo tanto, está recibiendo una gran ráfaga de tráfico (representada por los anchos tubos de líneas de puntos en la figura 2A). El dispositivo cliente del abonado 2, por otro lado, tiene una conexión con el servidor a distancia mediante un enlace muy congestionado (representado por los tubos estrechos de línea de puntos en la figura 2A) y está recibiendo una cantidad muy pequeña de tráfico de paquetes. En este caso surgen dos problemas.

Las figuras 2B y 2C muestran las colas relevantes en los respectivos enrutadores, que manejarán los paquetes para los dos abonados a medida que pasan fuera del dominio por cable y a través de la red inalámbrica a los dispositivos cliente.

Como se muestra en la figura 2B, dado que son de la misma prioridad, si el dispositivo de red por cable está usando puesta en cola basada en la clase, no colas por flujo implementadas en la planificación WFQ, el planificador de WFQ simplemente coloca cualesquiera nuevos datos para el abonado 2 en la cola saliente detrás de cualesquiera datos ya planificados para la transmisión al abonado 1. Dado que hay muchos más datos para el abonado 1, a menudo habrá un número sustancial de paquetes por delante de cualesquiera paquetes para el abonado 2 en el momento de la planificación WFQ.

Dado de la red por cable está habitualmente sobre-abastecida, es fácil ver que en el elemento de dominio inalámbrico, el abonado 1 tendrá muchos más paquetes en su cola que el abonado 2 (el abonado 1 tiene una cola llena, mientras que para el abonado 2 está vacía en el ejemplo). Las transmisiones del abonado 1 son capaces de llenar la cola del abonado 1 a una velocidad relativamente más alta que el abonado 2, debido al enrutamiento y la planificación a través de la parte por cable de la red.

El algoritmo de PF usado por el planificador para el enlace inalámbrico básicamente dará una mayor preferencia a un abonado particular en base a la cantidad de paquetes en la cola para ese abonado y/o la cantidad de ancho de banda que el enlace por aire puede manejar entre un dispositivo del abonado móvil particular y la Red de Acceso. En el ejemplo, el planificador de PF continuamente da una mayor prioridad al abonado 1, dado que ese abonado tiene más paquetes en su cola. Por lo tanto, es muy fácil para el abonado 1 acaparar todo el ancho de banda, y virtualmente privar de éste al abonado 2. La condición puede ser particularmente problemática, si las condiciones de radio son iguales, y el abonado 1 es de menor prioridad. El planificador de PF dará de forma no equitativa un mayor número de intervalos de tiempo al abonado 1, incluso aunque los dos abonados sean de la misma prioridad.

El segundo problema se muestra en la figura 2C. Ahora, hay colas por flujo implementadas en el elemento por cable

que realiza la planificación WFQ. En el ejemplo, se supone que el tráfico del abonado 1 tiene la mayor prioridad. La WFQ planifica el tráfico al menos en una parte significativa en base a la prioridad, por lo tanto cuando el elemento que realiza la planificación WFQ tenga paquetes que enviar para ambos abonados, enviará aquellos para el abonado 1 en primer lugar, en base a la mayor prioridad del tráfico de ese abonado. Incluso aunque al abonado 1 se le de mayor prioridad en la transmisión, los paquetes para el abonado 1 mantendrán a la cola para ese abonado relativamente llena, debido al número de paquetes suministrados a través de su sesión de banda ancha desde el servidor (véase también la figura 2A).

Debido a las diferencias en la congestión de la red, el tráfico de menor prioridad se retrasa de forma significativa en la red por cable. Este retardo de la red por cable puede alcanzar un punto en el cual el tráfico de menor prioridad será abandonado si no es enviado inmediatamente al dispositivo cliente del abonado 2. Sin embargo, dado que la planificación WFQ da prioridad al abonado 1 y el abonado 1 tiene una cola relativamente llena, el enrutador por cable sigue enviando paquetes para el abonado 1 antes de transmitir paquetes para el abonado 2. Esto tiende a mantener a los paquetes para el abonado 1 en la cola en el dominio inalámbrico, de modo que el planificador de PF da continuamente una mayor prioridad al abonado 1. En el ejemplo, la cola que contiene paquetes del abonado 1, en el elemento de enrutamiento inalámbrico que usa planificación PF, está relativamente llena. Al mismo tiempo, la cola para el abonado 2 en ese elemento contiene pocos, si es que contiene alguno, paquetes de datos. Como resultado, se le da prioridad al tráfico del abonado 1 y es transmitido al dispositivo cliente significativamente por debajo de la adjudicación de retardo. En este caso, se ve claramente que la planificación WFQ se ocupa de minimizar el retardo promedio local de los flujos en cada cola y que el algoritmo de PF se ocupa de maximizar el rendimiento del sistema en el enlace por aire. Lo que se echa en falta es algo que también tenga en consideración el objetivo global de cumplir las adjudicaciones de tiempo de todos los flujos pero asegurando que el tráfico de mayor prioridad no está retardado más allá de una cantidad perceptible para los abonados.

Otro problema obvio se produce cuando la función de PF en la red inalámbrica planifica paquetes en base al estado del canal. Si el algoritmo de PF no considera la noción de prioridad de la cola, entonces puede verse claramente que a usuarios de baja prioridad, con buenos estados del canal de forma coherente se les asignarán más recursos de red y se les dará un tratamiento preferente respecto a abonados de mayor prioridad.

Como se muestra mediante la descripción anterior, existe un problema con los parámetros de planificación de Paquetes de QoS de Extremo a Extremo, cuando se despliega tráfico de red sensible al tiempo en combinaciones de redes 3G cdma2000 1xEV-DO y redes IP. Los algoritmos de planificación de paquetes no se diseñaron para despliegues híbridos de sesiones de paquetes inalámbricas a por cable, dando como resultado asignaciones de recursos sub-óptimas. Existe una necesidad de mejorar los algoritmos de planificación para servicios sensibles al tiempo, tales como servicios multimedia integrados, a través de combinaciones de redes de paquetes (por ejemplo IP) inalámbricas y por cable que evitan problemas de asignación de recursos y/o mejoran el rendimiento.

Resumen

Las enseñanzas en este documento proponen una solución, cuyos ejemplos incluyen añadir la noción de tiempo de holgura en el algoritmo o algoritmos de planificación de paquetes para la Puesta en Cola Equitativa Ponderada y/o la Puesta en Cola con Equidad Proporcional, en despliegues de red híbrida inalámbrica-por cable. Al modificar uno o más de estos algoritmos, por ejemplo para reordenar o reorganizar paquetes en base a tiempos de holgura, son posibles asignaciones de recursos óptimas globales, al menos en algunos casos.

Por ejemplo, la descripción en este documento describe un procedimiento de planificación de transmisiones de paquetes, para su uso para proporcionar un servicio de comunicación de paquetes a dispositivos cliente de un abonado inalámbricos a través de una red híbrida que tiene una parte por cable y una parte inalámbrica. Este procedimiento implica determinar una adjudicación de tiempo para el suministro de cada paquete respectivo a través de la combinación de las partes por cable e inalámbrica de la red a cada uno de una serie de los dispositivos cliente de abonado inalámbricos. Los paquetes son enrutados a través de la parte por cable de la red a la parte inalámbrica de la red, usando un primer algoritmo de planificación. En base al tránsito de cada paquete respectivo a través de la parte por cable de la red, un elemento computa un tiempo de holgura, que representa la cantidad restante de la adjudicación de tiempo para el suministro del paquete respectivo. El procedimiento también implica enrutar los paquetes a través de la parte inalámbrica de la red usando un segundo algoritmo de planificación diferente del primer algoritmo de planificación. En un punto en la red por cable o un punto en la red inalámbrica, antes de la transmisión de paquetes mediante un enlace inalámbrico a respectivos dispositivos cliente de abonado inalámbricos, al menos dos de los paquetes destinados a diferentes dispositivos cliente de abonado inalámbricos se reordenan, en base a los tiempos de holgura computados para esos paquetes. Los paquetes reordenados son enrutados de acuerdo con al menos uno de los algoritmos de planificación. La reordenación y posterior planificación de enrutamiento asegura que los paquetes puedan ser suministrados a los dispositivos cliente de abonado inalámbricos apropiados antes de que expiren las respectivas adjudicaciones de temporización.

En una implementación híbrida típica, la parte por cable de la red usa un algoritmo de Puesta en Cola Equitativa

Ponderada (WFQ) para planificar transmisiones de paquetes desde cada elemento de enrutamiento de la red o una combinación de WFQ con otro algoritmo, tal como de prioridad estricta. La parte inalámbrica de la red usa un algoritmo de Equidad Proporcional (PF) para planificar transmisiones de paquetes mediante el enlace o enlaces inalámbricos a los dispositivos cliente. La monitorización del tiempo de holgura y la reordenación de paquetes pueden realizarse en cualquiera o en ambas partes de la red, es decir con respecto a la planificación WFQ y/o la planificación PF.

En una implementación ejemplar, un monitor detecta la entrada de cada paquete en la red y genera un cronomarcador que indica ese tiempo de entrada. El tiempo de holgura para el paquete, después de la llegada a un nodo cadena abajo, iguala la diferencia entre el momento de llegada en el nodo cadena abajo y el momento de la entrada para el paquete respectivo, sustraído de la adjudicación de tiempo para el paquete respectivo.

La red proporcionará típicamente una serie de diferentes servicios de comunicación en apoyo de diferentes aplicaciones, que tendrán diferentes tolerancias a la latencia de la red. La red puede implementar diferentes adjudicaciones de tiempo para el suministro de paquetes con respecto a paquetes de abonado relacionados con esas diferentes aplicaciones o realizados mediante los diferentes servicios de red. En tal caso, la determinación de la adjudicación para cada paquete implica detectar la aplicación o servicio para cada paquete (por ejemplo en base a un flujo de tráfico respectivo), de entre los servicios o aplicaciones soportados mediante la red. En base al servicio o aplicación, se asigna una adjudicación de tiempo entre las posibles adjudicaciones de tiempo asociadas con los diversos servicios o aplicaciones diferentes soportadas mediante la red.

La metodología puede implementarse en diversos tipos diferentes de redes utilizando componentes de red por cable e inalámbricos para comunicar paquetes a dispositivos cliente inalámbricos. El ejemplo mostrado en los dibujos y descrito en detalle utiliza una implementación híbrida de una red 3GPP2 1xEV-DO. La comunicación de paquetes a través de dicha red soporta telefonía de voz usando VoIP, navegación por páginas Web, descarga de software y/o video, diversos servicios de e-mail y envíos de mensajes de texto, y diversas aplicaciones/servicios más.

Aspectos de la tecnología de enrutamiento de paquetes mejorada pueden realizarse en una red de comunicación o en un sistema de manejo de paquetes para su uso en una red.

Una red, por ejemplo, podría proporcionar un servicio inalámbrico para dispositivos cliente de abonado inalámbricos. Dicha red incluiría una parte por cable y una parte inalámbrica. La parte de red por cable incluye uno o más elementos de enrutamiento de paquetes. El elemento de enrutamiento de paquetes tiene una primera función planificadora de la transmisión de paquetes asociada, que utiliza un primer algoritmo de planificación. La parte inalámbrica recibe paquetes de la parte por cable y transmite los paquetes recibidos mediante uno o más enlaces por aire a los dispositivos cliente de abonado inalámbricos. La parte inalámbrica incluye al menos un elemento de transmisión de paquetes que tiene una segunda función planificadora de paquetes asociada. La segunda función planificadora utiliza un segundo algoritmo de planificación diferente del primer algoritmo de planificación. La red también incluye un monitor de paquetes, que monitoriza flujos de paquetes a través de la red, para determinar un tiempo de holgura restante para el suministro de cada paquete. El monitor también ordena a al menos uno de los planificadores que reordene los paquetes, para evitar que expire el tiempo de holgura para el suministro de uno o más de los paquetes reordenados.

El sistema de manejo de paquetes controla las comunicaciones de paquetes a través de una red que proporciona servicio inalámbrico para dispositivos cliente de abonado inalámbricos. Este sistema incluye un planificador de paquetes y un monitor de paquetes. El planificador de paquetes controla la planificación de transmisiones de paquetes, destinados a diferentes dispositivos cliente de abonado inalámbricos, a través de un salto de red hacia un enlace de red inalámbrico que suministra a los dispositivos cliente. El monitor de paquetes monitoriza el tráfico de paquetes para determinar un tiempo de holgura restante para el suministro de cada paquete. El monitor ordena al planificador de paquetes que planifique un primer paquete, que tiene un tiempo de holgura restante demasiado pequeño para permitir el suministro a un primer dispositivo cliente de abonado inalámbrico, antes de que expire el tiempo de holgura restante para el primer paquete, pero delante de un segundo paquete que tiene suficiente tiempo de holgura restante para permitir el suministro a un segundo dispositivo cliente de abonado inalámbrico después de la transmisión del primer paquete.

Ventajas adicionales y nuevas características se mostrarán en parte en la siguiente descripción, y en parte se volverán evidentes para los expertos en la materia después del examen de lo siguiente y los dibujos adjuntos o pueden aprenderse mediante la producción u operación de los ejemplos. Las ventajas de las presentes enseñanzas pueden advertirse y obtenerse mediante la práctica o el uso de las metodologías, instrumentos y combinaciones particularmente señaladas en las reivindicaciones adjuntas.

Breve descripción de los dibujos

Las figuras de los dibujos representan una o más implementaciones de acuerdo con las presentes enseñanzas, a

modo de ejemplo solamente, no a modo de limitación. En las figuras, números de referencia similares se refieren a elementos iguales o similares.

5 La figura 1 representa una arquitectura de red 3GPP2 1xEV-DO, capacidades de enlace, protocolos y planificadores de paquetes para QoS.

Las figuras 2A-2C muestra cómo objetivos del planificador de paquetes no alineados dan como resultado la utilización sub-óptima de recursos.

10 Las figuras 3A y 3B ilustran cómo un algoritmo de equidad proporcional (PF) considera el estado del canal y la cantidad de ancho de banda solicitada, para maximizar el rendimiento del sistema.

Las figuras 4A y 4B ilustran decisiones de planificación sub-óptimas, en base a políticas de planificación locales de puesta en cola equitativa ponderada (WFQ) y equidad proporcional (PF).

15 La figura 5 representa una arquitectura de red inalámbrica y por cable, que tiene un monitor de latencia de paquetes integrado distribuido y un mecanismo de planificación de paquetes que puede ajustarse para tiempos de holgura.

20 Las figuras 6A y 6B muestran el beneficio de reorganizar un algoritmo de planificación de paquetes, que da como resultado asignaciones de recursos de red óptimas.

La figura 7 ilustra una modificación de WFQ.

25 La figura 8 es una vista visual proyectada de paquetes en una cola particular y cuánto retardo puede absorberse debido a la reorganización de un posible paquete sin afectar a ningún otro paquete que ya está en la cola.

La figura 9 representa paquetes en una cola particular, mostrando el impacto sobre todos los demás paquetes al insertar un paquete P0.

30 Descripción detallada

En la siguiente descripción detallada, se muestran numerosos detalles específicos a modo de ejemplos para proporcionar una exhaustiva comprensión de las enseñanzas relevantes. Sin embargo, debe ser evidente para los expertos en la materia que las presentes enseñanzas pueden ponerse en práctica sin dichos detalles. En otros ejemplos, se han descrito procesos, procedimientos, componentes y circuitos bien conocidos a un nivel relativamente alto, sin detalle, para evitar oscurecer innecesariamente aspectos de las presentes enseñanzas.

40 La solución en los ejemplos descritos en este documento proporciona un algoritmo de planificación sensible a la latencia de paquetes integrada, para despliegues de red usando combinaciones de planificación de Equidad Proporcional y Puesta en Cola Equitativa Ponderada para paquetes destinados a dispositivos cliente inalámbricos. A continuación se hará referencia en detalle a los ejemplos ilustrados en los dibujos adjuntos y descritos a continuación.

45 Inicialmente, se presenta un análisis formal para soportar datos en tiempo real en redes 3GPP2 1xEV-DO, y asignar recursos en la optimización del cumplimiento de todas las adjudicaciones de latencia de tiempo, tal como para tráfico VoIP, la noción de tiempo de holgura debe considerarse en el borde de la red entre la red inalámbrica y por cable, que está habitualmente en los planificadores de BTS y/o RNC,

50 Como se ha indicado anteriormente, la red de radio usa una estrategia de planificación PF. El algoritmo de equidad proporcional (PF) intenta maximizar el rendimiento del sistema con la limitación añadida de evitar la falta de suministro a cualquiera de los flujos. El algoritmo de PF asigna recursos de radio a una cola de paquetes de datos, destinada a la transmisión a un dispositivo móvil, que tiene la máxima relación entre el caudal o ancho de banda solicitado para una próxima transferencia de datos pendiente y el valor real del caudal o ancho de banda para la transferencia de datos al dispositivo móvil en el intervalo de tiempo previo. El algoritmo de PF se indica de la siguiente manera:

$$PF = \arg_{\max_{j \in \{1, \dots, N\}}} (r_i / R_i)$$

donde r_i es el caudal o ancho de banda solicitado en el intervalo de tiempo actual, y R_i era el ancho de banda o caudal real que atravesó el enlace por aire en el intervalo temporal previo.

El elemento de la red que controla las transmisiones de paquetes pone en cola paquetes para cada dispositivo móvil. A continuación, el elemento asigna recursos de radio a las colas de paquetes de datos para los dispositivos móviles, en base a los valores relativos dados por el algoritmo de PF. Específicamente, a la cola para el dispositivo que tiene el valor máximo producido por el algoritmo se le dan recursos para permitir la transmisión de un paquete desde la cola, a la cola para el dispositivo con el siguiente valor más elevado se le dan recursos para permitir la transmisión y así sucesivamente.

Puede observarse fácilmente que esto impide la privación de recursos dado que si el marco temporal anterior asignaba un ancho de banda de cero $R_i = 0$, y hay un ancho de banda solicitado extraordinario positivo no igual a cero en el marco temporal actual $r_i > 0$, la relación r_i/R_i da como resultado un número muy grande, asegurando la asignación de ancho de banda por parte del planificador a este abonado. Esta característica del algoritmo de planificación de PF da como resultado una propiedad auto-adaptativa, donde los flujos que están empezando a agotarse, gradualmente ganan más peso en posteriores intervalos de tiempo.

Esto se ilustra adicionalmente con un ejemplo. Las figuras 3A y 3B ilustran la principal idea detrás del algoritmo de equidad proporcional. En estos dibujos, las colas para los paquetes de datos del abonado se muestran a la derecha del abonado de PF. Los tubos cilíndricos representan esquemáticamente el ancho de banda relativo proporcionado mediante el enlace por aire a los dispositivos cliente inalámbricos respectivos. Los dibujos representan el tráfico planificado por PF a través de dos intervalos de tiempo sucesivos, $T = 1$ (figura 3A) y $T = 2$ (figura 3B).

En el primer intervalo de tiempo $T = 1$ (figura 3A), se observa que el primer abonado tiene un estado de canal que puede manejar 1 kbs, mientras que el segundo abonado tiene un estado de canal que puede absorber 100 kbs. En el ejemplo, ambos están recibiendo flujos de la red por cable a la misma velocidad de 100 kbs. Suponiendo el peor caso, al abonado 1 no se le dio ningún ancho de banda en el primer intervalo de tiempo, por lo tanto $R_1 = 0$, y $R_2 = 100$. Dejando $r_1 = 1$ y $r_2 = 100$. La cola del abonado 1 obviamente comenzará a llenarse. Ahora, en el segundo intervalo de tiempo $T = 2$ (figura 3B), de acuerdo con el algoritmo de equidad proporcional, el primer abonado recibirá un peso de $1/0$ y el segundo abonado recibe un peso de $100/100$. Claramente el primer abonado recibirá casi toda su cantidad solicitada, aunque hay un límite de lo que puede manejar el canal por aire, de modo que intervalos de tiempo no usados pueden asignarse al segundo flujo, dando como resultado que no se malgasten recursos.

La mayor parte de proveedores de equipos de conmutación de paquetes de red por cable IP usan alguna forma del enfoque de Intercambio de Procesador Generalizado (GPS) para planificar paquetes. La puesta en Cola Equitativa Ponderada se aproxima estrechamente a GPS emulando a un planificador bit a bit pero usando paquetes de tamaño variable reales. La Puesta en Cola Equitativa Ponderada (WFQ) asigna a cada cola un factor de peso. El algoritmo usa el factor de peso para determinar una cantidad de recursos adecuada a asignar a la cola para transmisión de paquetes desde la cola a través de las redes.

La WFQ intenta minimizar el retardo promedio del sistema. La WFQ es indicada mediante tres componentes.

1. Retardo de sesión promedio:

$$D = \left(\sum_{i=1}^N d_i \right) / N$$

donde:

D = Retardo de flujo o Sesión del Sistema promedio

d_i = retardo para un flujo o sesión particular ;

N = Número de flujos o sesiones

El algoritmo de WFQ intenta conseguir el mínimo retardo de sesión del sistema promedio con la limitación añadida de asegurar la equidad por todos los flujos individuales.

2. Algoritmo de planificación de paquetes

$$\frac{W_i(\tau, t)}{W_j(\tau, t)} \geq \frac{\phi_i}{\phi_j}, j = 1, \dots, N$$

donde:

N = número de flujos activos

5 ϕ_1, \dots, ϕ_N = números naturales positivos, no iguales a cero, que representan un peso relativo del ancho de banda disponible global.

$W_i(\tau, t)$ = Ancho de banda asignado a un flujo particular i durante un intervalo de tiempo (τ, t)

10 Ésta es una característica significativa del algoritmo de WFQ, donde a cada flujo se le dará cierta parte del ancho de banda disponible global. En casos en los que otros flujos tienen colas vacías, un ancho de banda no usado puede asignarse a flujos con colas no vacías.

3. Asignación de ancho de banda mínimo garantizada

$$r_i = \frac{\phi_i}{\sum_{j=1}^N \phi_j} C$$

15 donde:

N = número de flujos activos

20 **C = Capacidad de ancho de banda del enlace de salida**

Esta ecuación simplemente indica que a un flujo particular i se le asignará al menos su proporción del ancho de banda disponible global en el enlace de salida en curso.

25 Las técnicas descritas en este documento implican determinar una adjudicación de tiempo para el suministro de cada paquete respectivo a través de la red híbrida por cable - inalámbrica a los respectivos dispositivos cliente de abonado inalámbricos. La adjudicación de tiempo para un paquete, en los ejemplos, es la máxima latencia permisible que la aplicación o servicio de comunicación puede aceptar sin alteración. En algunos casos, una aplicación en una estación receptora (por ejemplo una aplicación VoIP ejecutada en un terminal móvil) puede desechar un paquete que está llegando indebidamente tarde a la estación, lo que establecería esencialmente el límite superior en la adjudicación de tiempo para la comunicación en espera para esa aplicación ejecutada en el dispositivo cliente. Por ejemplo, un servicio o aplicación VoIP puede tener una adjudicación inicial de 150 ms en varios de nuestros ejemplos de VoIP. Otros servicios o aplicaciones, tales como envío de mensajes de texto o descargas de software/video tendrán típicamente otros valores para la adjudicación inicial.

35 La planificación sub-óptima de recursos puede producirse, como se ha descrito de forma breve anteriormente, cuando los paquetes se planifican a ciegas en base solamente a la prioridad. En redes 3GPP2 cdma2000 1xEV-DO, por ejemplo, paquetes VoIP de abonados con alta prioridad y abonados con baja prioridad pueden planificarse de modo que se superen las adjudicaciones de tiempo de retardo, incluso cuando dicho problema podría haberse evitado. Las figuras 4A y 4B ilustran dos casos en los que las adjudicaciones de retardo temporal se superan pero podrían haberse suministrado dentro de las adjudicaciones. De nuevo, las colas para los paquetes de datos del abonado se muestran a la derecha del planificador. En este caso, el abonado 1 tiene el tráfico de baja prioridad, mientras que el abonado 2 tiene el tráfico de mayor prioridad.

45 La figura 4A representa el Caso 1, en el que paquetes de alta prioridad se planifican por delante de paquetes de baja prioridad. El siguiente escenario describe cómo puede ocurrir esto.

50 En primer lugar, consideramos WFQ (como se muestra realmente en la figura 4A). En base a la diferencia de prioridad, los paquetes para el abonado 1 se transmiten de forma menos frecuente que los del abonado 2. Como resultado, muchos de los paquetes destinados al abonado 1 están sujetos a tiempos de retardo más largos. En las redes por cable, este primer caso es un escenario muy probable, en el que los paquetes de baja prioridad están retardados en el borde de entrada, el núcleo y el borde del enlace directo final. En el momento en el que estos paquetes alcanzan el borde entre las partes por cable e inalámbrica de la red, los paquetes de baja prioridad ya están obsoletos y es necesario planificarlos inmediatamente para la transmisión al dispositivo del abonado objetivo. 55 Mientras tanto, paquetes de alta prioridad no se retrasan significativamente en el borde de entrada, el núcleo y el borde del enlace directo final, dando como resultado un gran tiempo de holgura antes de alcanzar el límite de latencia de adjudicación de tiempo, es decir 150 ms para una aplicación de servicio VoIP típica. No supone una diferencia notable para el usuario final incurrir en un retardo de puesta en cola de cierta cantidad menor. Al reorganizar las prioridades, y permitiendo que los paquetes del abonado 1 sean planificados antes que paquetes de alta prioridad, podría resultar una asignación de recursos óptima. 60

En redes inalámbricas que usan planificación PF, este escenario puede surgir si las condiciones del canal del abonado de alta prioridad son muy buenas y las condiciones del canal del abonado de baja prioridad son malas, además de la situación anterior en la que los paquetes de baja prioridad se retardan en el borde de entrada, el núcleo y el borde del enlace directo final. Los paquetes de baja prioridad ya están obsoletos y es necesario planificarlos para la transmisión inmediatamente al abonado objetivo. Mientras tanto los paquetes de alta prioridad no sufren un retardo significativo en el borde de entrada, el núcleo y el borde del enlace directo final, dando como resultado un gran tiempo de holgura antes de alcanzar el límite de latencia de adjudicación de temporización de 150 ms. No supone una diferencia notable para el usuario final incurrir un retardo de puesta en cola de alguna cantidad secundaria para permitir que los últimos paquetes de baja prioridad sean planificados en primer lugar para la transmisión al dispositivo cliente objetivo.

En el Caso 2 (figura 4B), paquetes de baja prioridad se planifican por delante de paquetes de alta prioridad. Debido a la característica de no agotamiento de ambos algoritmos, puede ser un caso en el que paquetes de VoIP de baja prioridad se planifican por delante de paquetes retardados de alta prioridad. El siguiente escenario describe cómo puede ocurrir esto:

En redes por cable que usan planificación WFQ, este segundo caso es un escenario menos probable, en el que los paquetes de alta prioridad están sobresuscritos y retardados en el borde de entrada, el núcleo y el borde del enlace directo final. Estos paquetes ya están obsoletos y es necesario planificarlos inmediatamente para la transmisión al abonado objetivo. Mientras tanto paquetes de baja prioridad no están retardados significativamente en el borde de entrada, el núcleo y el borde del enlace directo final, por ejemplo, debido a que toman una ruta diferente enteramente. El abonado de baja prioridad puede tener realmente paquetes en el límite de latencia de adjudicación de temporización de 150 ms. Supone poca diferencia notable para ese usuario final incurrir en un retardo de puesta en cola de cierta cantidad secundaria. La reorganización de prioridades, y permitir que más de los paquetes del abonado 2 sean planificados antes de los paquetes de baja prioridad, dará como resultado más probabilidades de que todos los paquetes lleguen dentro de la adjudicación de latencia respectiva.

En redes inalámbricas que usan planificación PF (como se muestra en 4B), este escenario de segundo caso puede surgir si las condiciones del canal del abonado de baja prioridad son muy buenas y las condiciones del canal del abonado de alta prioridad son malas, además de la situación anterior en la que los paquetes de alta prioridad se retardan en el borde de entrada, el núcleo y el borde del enlace directo final debido a la sobresuscripción y una ruta diferente del tráfico del abonado de baja prioridad. Los paquetes de alta prioridad ya están obsoletos y es necesario planificarlos inmediatamente para la transmisión al abonado objetivo (es decir más rápido de lo que sería el caso si el planificador de PF basara solamente su decisión en condiciones del canal respectivas, como se muestra). Mientras tanto, paquetes de baja prioridad no están retardados significativamente en el borde de entrada, el núcleo y el borde del enlace directo final, dando como resultado un gran tiempo de holgura bien en el límite de latencia de adjudicación de temporización de 150 ms. Supone una diferencia poco notable para el usuario final de baja prioridad incurrir en un retardo de puesta en cola de alguna cantidad secundaria para permitir que los paquetes de alta prioridad que llegan tarde sean planificados para una transmisión más temprana al objetivo. Será necesario que el planificador PF aumente el peso del tráfico de alta prioridad para permitir que lleguen más paquetes dentro de las adjudicaciones de temporización, dando como resultado una optimización global.

Con esa visión general, puede ser útil considerar el sistema ejemplar ilustrado en la figura 5, con algo más de detalle. La figura 5 representa una arquitectura de red 3GPP2 1xEV-DO, superpuesta con un monitor de latencia de paquetes integrados distribuidos y un sistema de planificación de paquetes. El diagrama de bloques funcionales muestra elementos de una red de comunicación móvil inalámbrica ejemplar 11, para proporcionar servicios basados en paquetes, para aplicaciones de datos multimedia tales como servicios de teléfono de voz móvil en base a comunicaciones de paquetes de tipo VoIP. El dibujo muestra elementos de red en forma de diagrama de bloques funcionales de alto nivel, y muestra ciertos aspectos del procesamiento implicado en comunicaciones a través de los elementos ilustrados.

La red de comunicación 11 proporciona servicios de comunicación de paquetes para numerosas estaciones móviles, aunque para fines de descripción, el dibujo muestra un único dispositivo 13. El dispositivo móvil 13, algunas veces denominado como un Agente de Usuario (UA) móvil, típicamente ejecuta uno o más programas "cliente" para implementar la funcionalidad del agente con respecto a uno o más servicios de comunicación que el usuario obtiene o a los que se abona a través de la red 11. El dispositivo móvil 13, por ejemplo, puede asumir la forma de una estación de telefonía móvil, con pantalla y capacidades de entrada de información por parte del usuario para soportar comunicaciones multimedia. Actualmente, dichas implementaciones de telefonía móvil del dispositivo 13 típicamente asumen la forma de terminales portátiles, aunque pueden implementarse en factores de otra forma. Como otra clase de ejemplos de estación, el dispositivo móvil 13 puede asumir la forma de un asistente digital personal (PDA) o un ordenador personal portátil, que incorpora un transceptor inalámbrico compatible con el tipo particular de servicio de paquetes de datos inalámbrico ofrecido por la red 11. Por supuesto, las estaciones móviles pueden asumir otras formas o conectar a diversos dispositivos de datos más, que pueden permitir el uso de los servicios de comunicación de la red.

La red 11 incluye o implementa una o más redes de acceso por radio (RAN), para comunicación inalámbrica con los dispositivos móviles que reciben servicios a través de la red 11. Los elementos físicos de una red de acceso por radio incluyen una serie de estaciones base (BS) 15. Cada estación base 15 incluye un sistema de antena 17 y un sistema transceptor base (BTS) 19. Uno o más enrutadores 21, 23 proporcionan enrutamiento de paquetes al y desde el BTS 19 y un controlador de red de radio (RNC) 25 en un centro de conmutación de servicios móviles (MSC) 27.

El sistema transceptor base (BTS) 19 comunica mediante el sistema de antena 17 y el enlace por aire con una o más de las estaciones móviles 13, cuando las estaciones móviles están dentro del alcance. El BTS 19 es la parte de la red de radio que envía y recibe señales de radiofrecuencia que portan paquetes a/desde las estaciones móviles a las que la estación base 15 sirve actualmente. Las comunicaciones del BTS por el enlace por aire con el dispositivo cliente inalámbrico del UA 13 proporcionan la Interfaz de Datos de Paquete de Alta Velocidad (HPRD) para servicios de datos del cliente. El BTS 19 incluye o está asociado con un módulo DOM que controla las comunicaciones de paquetes inalámbricos a través de la BS y el dominio inalámbrico de la HPRD, incluyendo planificación de transmisiones de paquetes en el enlace o enlaces directos, en este ejemplo de la parte inalámbrica de la red híbrida.

La red 11 también incluye una serie de Nodos Servidores de Datos en Paquetes o "PDSN". En el MSC 27 que sirve al dispositivo móvil 13, el PDSN 31 sirve como un agente externo (FA). El PDSN está en comunicaciones de paquetes con el RNC 25, por ejemplo mediante otro enrutador 33. El PDSN de agente externo 31 establece, mantiene y termina enlaces lógicos con la parte asociada de la red de acceso por radio. El PDSN también soporta sesiones de protocolo punto a punto (PPP) con las estaciones móviles 13. El PDSN proporciona la función de enrutamiento de paquetes desde la red de radio a/desde otras redes conmutadas con paquetes, en este caso mediante la red IP 35 a un agente local (HA) 37, que a su vez proporciona enrutamiento de paquetes a/desde una red central IP 39, por ejemplo para acceso a Internet o Intranet.

Los agentes 31 y 37 están acoplados a un sistema de Autenticación, Autorización y Estadísticas (AAA) 41. En uno o más puntos en el procesamiento de una llamada u otra sesión de comunicación, el FA 31 o el HA 37 del PDSN accede al servidor AAA 41 para obtener autorización de acceso de llamadas, y el FA y el HA proporcionan información respecto a la duración/volumen de uso durante la sesión, al servidor AAA para fines de recuento de estadísticas.

Desde la perspectiva de la estación móvil, a menudo habrá dos o más estaciones base dentro del alcance. El dispositivo móvil 13 comunica a través del BTS 19 de la estación base 15, seleccionada entre aquellas estaciones base que puede detectar (aproximadamente dentro de alcance) por el aire, finalizando con el BTS 15 que proporciona la mejor relación de Canal con respecto a Interferencia. La Interfaz de Datos de Paquete de Alta Velocidad (HPRD) en este segmento inalámbrico es la conexión de red de capacidad más costosa y estrecha entre todos los segmentos (representados temáticamente mediante los tubos de diversos tamaños/ancho de banda) que portarán la comunicación del UA. Aunque se muestran por separado por motivos de conveniencia, el planificador de paquetes 43 en este segmento típicamente reside en el módulo DOM implementado en o asociado con el BTS 19. Típicamente, el planificador es una función programada del elemento de enrutamiento. El planificador de paquetes de la red inalámbrica 43 utiliza un algoritmo de planificación, que optimiza el rendimiento del sistema en base a un algoritmo de equidad proporcional (PF).

El siguiente segmento de red conecta el BTS 19 al controlador de la red de radio (RNC) 25, situado en el Centro de Conmutación de Servicios Móviles (MSC) 27. Aunque este segmento de red de retroceso IP podría usar circuitos TDMA, en el ejemplo, este segmento inalámbrico usa conexiones Metro Ethernet entre enrutadores/conmutadores tales como los mostrados en 21 y 23 para proporcionar transporte a y desde el MSC 27. Aunque mostrados por separado por motivos de conveniencia, el RNC 25 implementa un planificador de paquetes 47, típicamente como un aspecto programado de esta función de enrutamiento de paquetes. En el enlace directo, el planificador de paquetes 47 en el RNC 25 típicamente usa una variante de Puesta en Cola Equitativa Ponderada (WFQ). La puesta en cola de baja latencia (LLQ) es una implementación típica, que combina Puesta en cola de prioridad estricta con WFQ. El resto de los segmentos de la red por cable usan todos alguna forma de planificador de paquetes de WFQ, como se muestra mediante los bloques del planificador de WFQ 47 y 49 en el diagrama, que pueden implementarse en el software de control del enrutamiento, por ejemplo en el PDSN 31 y el agente local 37, respectivamente. La WFQ está diseñada para minimizar la latencia promedio para todos los flujos.

La solución ejemplar a los problemas de políticas de planificación sub-óptima implica una serie de dispositivos externos, que generan cronomarcadores sincronizados y son sensibles a las latencias de los paquetes de un flujo particular, y que introducen cambios en base a esta información en una o ambas de las decisiones de planificación de paquetes de WFQ y PF. Esto permite añadir nociones de tiempo de holgura y reordenación de paquetes en espera, en los algoritmos de planificación de paquetes, para la Puesta en Cola Equitativa Ponderada y/o la Puesta en Cola con Equidad Proporcional, que en efecto compensa o ajusta las incongruencias de los algoritmos de planificación de paquetes para optimizar la planificación de paquetes en base a las adjudicaciones de temporización de paquetes críticos en cuanto al tiempo, tales como VoIP. Los monitores de latencia de paquetes 51-57 están

integrados funcionalmente con los sub-sistemas de planificación de paquetes en el equipo de la red.

5 Un ejemplo de arquitectura de una solución propuesta se muestra en la figura 5, que incluye una serie de dispositivos basados en flujo distribuido que monitorizan, registran y comparten cronomarcadores de paquetes e información de latencia y computan tiempos de holgura para cada paquete y suministran esta información al planificador o planificadores de paquetes para decisiones de planificación de paquetes mejoradas. Los dispositivos de monitorización 51 y 53 comprenden un hardware de computación acoplado a o en comunicación con los elementos de la red, tal como el BTS/DOM en 19, el RNC 25 o a uno o más de los enrutadores asociados 21, 33. Dispositivos de monitorización similares 55 y 59 pueden proporcionarse para el FA 31 y/o el HA 37. El ordenador u ordenadores que realizan las funciones de los dispositivos de monitorización 51-57 están programados a su vez para proporcionar las funciones de monitorización, cronomarcado, computación de la holgura y control del planificador asociado descritas en este documento. Como alternativa, otros elementos de hardware en la red, tales como los BTS, RNC, FA y HA pueden estar programados para implementar las funciones de los dispositivos de forma interna. Para fines de descripción adicional, los dispositivos se denominan como monitores de latencia de paquetes, mostrados en 51-57 en el ejemplo de la figura 5.

20 En el ejemplo, los monitores se implementan en cada fase o sección de la red, para proporcionar datos para su uso en la optimización de las operaciones de planificación tanto de WFQ como de PF. Los expertos en la materia reconocerán que puede ser posible implementar los ajustes de tiempo de holgura de los algoritmos de planificación con respecto a un menor número de planificadores, por ejemplo solamente en asociación con el planificador de PF 43 en el dominio inalámbrico y/o solamente en asociación con el planificador de WFQ 45 en el borde del dominio por cable.

25 Existen muchas implementaciones posibles. Un posible ejemplo implica el marcado de paquetes con cronomarcadores sincronizados globalmente asociados que permiten que el monitor de paquetes deduzca latencias en aumento y la adjudicación de temporización o el tiempo de holgura restantes, que pueden usarse en la decisión de reorganización. Los monitores de latencia de paquetes 51-57 pueden escuchar en puertos SPAN o espejo de puerto de enlaces de agregación en los enrutadores de la red en la red portadora, así como recibir señales de sincronización temporal de GPS para cronomarcadores de alta precisión. Si una temporización de alta precisión no está disponible, otro enfoque incluiría monitores que crean una tabla de dispersión en base a la dirección IP de la fuente y el destino, puertos, números de secuencia y números de reconocimiento (ACK), que identifican únicamente cada flujo punto a punto particular, vigilando en busca de puertos reutilizados. Los monitores de latencia de paquetes 51-57 pueden compartir esta información entre cada uno de ellos (mediante enlaces de comunicación de datos representados mediante flechas en línea de puntos en el dibujo). Las adjudicaciones de tiempo se determinan detectando el servicio o aplicación para cada flujo, por ejemplo VoIP que tiene una adjudicación de 150 ms. Hay bastante información para identificar el tiempo de holgura de paquetes identificando en primer lugar el punto en el tiempo en el que el paquete entró por primera vez en la red. Por ejemplo, el monitor 57 generará un cronomarcador de entrada inicial para cada paquete respectivo que entra en la red mediante el elemento enrutador en el HA 37. Los monitores cadena abajo pueden determinar a continuación el tiempo de tránsito, restando el tiempo de entrada del tiempo actual desde la llegada al nodo particular. El tiempo de holgura restante es igual entonces a la adjudicación de tiempo original menos el tiempo de tránsito.

45 Típicamente, el tiempo de holgura inicial o la adjudicación en el punto de entrada es un valor fijo para un tipo dado de comunicación de paquetes. Por ejemplo, un servicio o aplicación VoIP puede tener una adjudicación inicial de 150 ms en el ejemplo de VoIP. La red proporcionará típicamente una serie de diferentes servicios de comunicación en apoyo de diferentes aplicaciones, tales como telefonía de voz VoIP, mensajería de texto, descargas de software, descargas de video, servicio de páginas Web y similares. Diferentes servicios o aplicaciones tienen diferentes tolerancias para latencia de la red. En apoyo, la red puede implementar diferentes adjudicaciones de tiempo para el suministro de paquetes con respecto a paquetes de abonado relacionados con aquellas diferentes aplicaciones o transportados mediante los diferentes servicios de red. Para cada servicio o aplicación, la adjudicación puede estar fijada de antemano. Sin embargo, los monitores detectan la aplicación o servicio para un flujo de tráfico de abonado y, por lo tanto, para cada paquete respectivo del flujo, de entre los servicios o aplicaciones soportadas a través de la red. En base al servicio o aplicación detectada para un flujo particular, los monitores asignan una adjudicación de tiempo para los paquetes en ese flujo, de entre las posibles adjudicaciones de tiempo asociadas con diversos servicios o aplicaciones diferentes soportadas a través de la red.

60 A medida que cada paquete de un flujo atraviesa la red 11, cada monitor de paquetes 57 a 51 puede computar el tiempo de holgura restante tomando la diferencia entre el tiempo en el que el paquete entró en la red y el tiempo transcurrido (más algún desfase para contar el enlace por aire) restado de la adjudicación de tiempo global. Este valor puede introducirse a continuación en el planificador de paquetes apropiado, por ejemplo 43 y/o 45, para ejecutar la política de planificación.

Las figuras 6A y 6B representan dos casos que ilustran un problema que puede ser causado por la planificación y un ejemplo del algoritmo de PF modificado propuesto que resuelve ese problema y, de este modo, muestran el

beneficio de la modificación del algoritmo de PF en base al tiempo de holgura.

En la figura 6A, caso 1, se observa el efecto de los actuales mecanismos de planificación de paquetes limitados, donde no hay reorganización de prioridades en base a los tiempos de holgura. En la figura 6B, caso 2, se observa que con la reorganización en base a los tiempos de holgura, todos los paquetes cumplen las adjudicaciones de tiempo, con un impacto mínimo o ningún impacto sobre los usuarios existentes. Esta reorganización debe ocurrir lo más cerca posible del objetivo, dado que las latencias exactas no se conocen en redes intermedias, lo que puede dar como resultado que se cumplan las adjudicaciones de temporización de paquetes de baja prioridad pero a expensas de retardar paquetes de alta prioridad, dado que estos pueden tener un grado inesperadamente alto de congestión, dando como resultado una latencia relativamente alta en la ruta de alta prioridad. Los ejemplos de las figuras 6A y 6B realizan la reorganización en el enrutador de la red inalámbrica que realiza la planificación PF, por ejemplo en el DOM en el BTS 19 en el ejemplo de la figura 5.

En el lado izquierdo de cada una de las figuras 6A y 6B, se observa un modelo de un enrutador típico que soporta QoS ofreciendo servicios diferenciados para el tráfico de la red de entrada, por ejemplo como podría implementarse en el FA PDSN 31. Para fines de descripción, en enrutador en 31 implementa tres colas de paquetes para sus puertos de salida, una cola de alta prioridad EF, una cola de prioridad media AF y una cola de baja prioridad BE. En realidad, servicios diferenciados tendrían al menos 6 colas, correspondiendo cada cola a una clase particular, que varía entre la clase de reenvío expedito (EF), clases de reenvío asegurado (AF1, AF2, AF3, AF4) y clases de mayor esfuerzo (BE). El diagrama solamente muestra 3 clases por motivos de sencillez. Colas similares se implementan en los enrutadores tardíos.

Cada cola es drenada por el planificador de paquetes del enrutador particular, en este caso, el planificador 47. Se muestra un segundo enrutador por cable, tal como uno que podría implementarse en el RNC 25. El enrutador en 25 implementa colas y un planificador 45, análogo a los del enrutador en la PDSN 31. Los planificadores en los enrutadores por cable (a la izquierda y en el centro en estos dibujos), implementan el algoritmo de Puesta en Cola Equitativa Ponderada (WFQ).

Muy a la derecha, se observa un elemento de enrutamiento de red para la red de radio, que controla la transmisión y la recepción de paquetes entre el móvil y la Red de Acceso. En la red de la figura 5, éste sería el enrutador en el módulo DOM en el BTS 19 (o un enrutador asociado 21), aunque otras redes de acceso pueden implementar el enrutador y la planificación en espera en otros nodos de red inalámbrica. Se supone para la descripción que la reordenación en base al tiempo de holgura se implementará en la parte inalámbrica de la red. El componente clave para fines de la descripción de estos ejemplos es, por lo tanto, el Planificador de Paquetes 43 en la parte inalámbrica de la red, es decir, el elemento que implementa el algoritmo de Planificación de Equidad Proporcional (PF), que estaría en o controlaría de otro modo el elemento que implementa las transmisiones del enrutador de red inalámbrica. El enrutador en 19 tiene las mismas 3 colas (EF, AF, BE) que los elementos de la red por cable, y transmitirá paquetes de una manera similar, aunque la planificación del algoritmo de PF tenderá a seleccionar paquetes de las colas en base a la asignación equitativa de ancho de banda y la maximización del rendimiento del equipo.

El elemento de red 43 que controla las transmisiones de paquetes en el nodo 19 pone en cola paquetes para cada dispositivo móvil. A continuación, el elemento 43 asigna recursos de radio a las colas de paquetes de datos para los dispositivos móviles, en base a los valores relativos dados por el algoritmo de PF. Específicamente, a la cola para el dispositivo que tiene el valor máximo producido por el algoritmo se le dan recursos para permitir la transmisión de un paquete desde la cola, a la cola para el dispositivo con el siguiente valor más elevado se le dan recursos para permitir la transmisión y así sucesivamente.

Los diagramas ilustran un escenario con 3 flujos, donde se muestra cada latencia de salto de red. Cada uno de los tiempos δt mostrados en cada una de las figuras 6A y 6B representa el retardo incurrido por el paquete respectivo como resultado de atravesar un conmutador/enrutador particular o un salto en una nube de red al siguiente elemento de enrutamiento. El flujo 1, incluye el paquete P1, con una latencia asociada de 15 ms para atravesar la primera nube de red, 5 ms de latencia de conmutación y puesta en cola a través de cada enrutador, y una latencia de 5 ms para atravesar la segunda nube de red. Por motivos de sencillez, se combinó la latencia de conmutación y puesta en cola a 5 ms para todos los conmutadores, que es acumulativa desde arriba hacia abajo. La cola de mayor prioridad tendrá solamente una única latencia de conmutación y puesta en cola de 5 ms a través de cada enrutador. La segunda cola (AF) tendrá una latencia de conmutación y puesta en cola de 5 ms + 5 ms = 10 ms a través de cada enrutador. Finalmente la tercera cola (BE) tendrá una latencia de conmutación y puesta en cola de 5 ms + 5 ms + 5 ms = 15 ms. El flujo 2 que contiene el paquete P2 tiene una latencia de 55 ms para atravesar la primera nube de red y una latencia de 25 ms para atravesar la segunda nube de red. Cada flujo comienza y termina en los mismos elementos de red, pero puede atravesar diferentes rutas de red o incurrir en diferentes tiempos de retardo de tránsito, debido a diferentes congestiones de red.

Estos diagramas mostrarán que en la figura 6A, no se cumplen las adjudicaciones de temporización. Este primer

ejemplo muestra que existe una adjudicación de temporización total de 150 ms. Éste es el tiempo máximo que puede transcurrir para que un paquete alcance su destino. Un ejemplo sería el caso de tráfico de voz, donde un paquete IP tardío que contiene información de voz codificada es inútil a menos que alcance su destino en cierto marco temporal. El primer diagrama mostrará que, sin ninguna modificación, se puede tener un caso en el que algunos paquetes no cumplirán la adjudicación de temporización, y algunos paquetes estarán en la adjudicación de temporización. El segundo diagrama (figura 6B) muestra que, si se aplica una sencilla modificación de reorganización del algoritmo de planificación PF, puede conseguirse una solución óptima, donde todos los paquetes alcanzarán el destino dentro de sus adjudicaciones de temporización. La principal idea es explotar el tiempo en el que los paquetes están bajo la adjudicación de tiempo, y ofrecer una planificación más temprana a paquetes en peligro de superar su adjudicación de temporización. La siguiente descripción proporciona más detalle.

La idea básica es computar los tiempos de holgura de cada paquete en cada cola. Trabajando desde la cola de prioridad más baja, arriba hacia las colas de prioridades más altas, si existe una oportunidad de reinsertar un paquete tardío sin afectar a los paquetes existentes, entonces ese paquete puede reorganizarse, o no conseguir reorganizarse y el paquete puede desecharse pronto para dejar espacio para otros paquetes que pueden ser capaces de usar los recursos de red liberados que pueden ser capaces ahora de cumplir su adjudicación de temporización, como resultado.

Los planificadores de WFQ seleccionan la cola EF en primer lugar, a continuación la cola AF y a continuación la cola BE en base a los pesos dados al tráfico de diferente prioridad. Para fines del simple ejemplo, se supondrá que los números de paquetes en las colas y las condiciones del canal de enlace por aire tienden a hacer que el planificador PF seleccione paquetes entre las colas en un orden similar.

En los ejemplos, cada uno de los enrutadores (rectángulos de línea de puntos en el dibujo) muestra una latencia combinada de planificación y conmutación de paquetes de 5 ms. La latencia promedio para el enlace por aire es de 10 ms. Entre el primer y el segundo enrutador por cable, los paquetes de la cola EF de alta prioridad experimentan un retardo de 15 ms. Entre el segundo enrutador por cable y el enrutador del dominio inalámbrico, los paquetes de la cola EF de alta prioridad experimentan un retardo de 5 ms.

Los valores de T_s en los dibujos representan tiempos de holgura detectados por los dispositivos monitores asociados (véase la figura 5). En el ejemplo, los tiempos de holgura iniciales igualan la adjudicación o adjudicaciones para cada paquete (mostradas en las colas de la izquierda), que son todas de 150 ms, un valor típico de latencia permisible para comunicación por VoIP. Otras aplicaciones pueden dictar diferentes adjudicaciones de tiempo de holgura. Como se muestra en el dibujo, los valores de T_s que representan la adjudicación de retardo restante decrecen a medida que cada paquete atraviesa la red y se somete a diversos retardos.

Por ejemplo, el paquete P1 de la cola de alta prioridad EF inicialmente tiene un tiempo de holgura adjudicado $T_s = 150$ ms. Incurre en un retardo de 5 ms para planificación y conmutación de paquetes combinadas a través de el primer enrutador por cable en 31/47 y un retardo de 15 ms en el transporte por el salto hasta el segundo enrutador en 25. Como resultado, en la cola EF en el segundo enrutador, ese mismo paquete P1 tiene un tiempo de holgura restante $T_s = 130$ ms. A su vez, el paquete P1 incurre en otro retardo de 5 ms para la planificación y la conmutación de paquetes a través del segundo enrutador por cable en 25/45 y un retardo de 5 ms en transporte por el salto hasta el enrutador en el dominio inalámbrico en 19. En el momento de la planificación PF (en 43) en el enrutador de la red inalámbrica en 19, el paquete P1 tiene un tiempo de holgura restante $T_s = 120$ ms. La planificación y la conmutación a través de ese enrutador en 19/43 requieren 5 ms adicionales, y el transporte por el salto a través del enlace inalámbrico a la estación móvil del abonado implica 10 ms adicionales. Debe observarse que el paquete P1 llega mucho antes de que expire el tiempo de holgura (aproximadamente 105 ms).

Los paquetes en la cola AF que tienen la siguiente prioridad más alta se someten a un retardo combinado de planificación/conmutación y retardos de transporte, y esos paquetes también son retardados el tiempo necesario para el servicio a la cola de mayor prioridad EF. En el primer enrutador por cable en la ubicación 31, el paquete P1 es planificado y conmutado a su través primero (5 ms), y el paquete P2 de la cola AF es planificado a continuación. Dado que el paquete P2 de la cola AF incurre en un retardo de planificación y conmutación de 5 ms, éste es retardado en realidad un total de 10 ms antes de la transmisión. En el ejemplo, dado que usa instalaciones de menor prioridad sujetas a más congestión, por lo tanto el transporte del paquete P2 desde la cola AF incurre en un retardo de 55 ms en el salto entre los dos enrutadores por cable.

Suponiendo que el paquete P2 de la cola AF inicialmente tiene una adjudicación de tiempo de holgura $T_s = 150$ ms, cuando alcanza el segundo enrutador por cable, ese paquete tiene un tiempo de holgura restante de 85 ms ($150 - 5 - 5 - 55$). A su vez, el paquete P2 incurre en un retardo de 5 ms mientras un paquete es enviado desde la cola EF, y otro retardo de 5 ms para la planificación y conmutación de paquete combinada del propio paquete P2 a través del segundo enrutador por cable en 25/45. El transporte por el salto desde el segundo enrutador por cable en la ubicación 19 hasta el enrutador en el dominio inalámbrico tarda 25 ms adicionales. En el momento de la planificación PF en el enrutador de la red inalámbrica en 43, el paquete P2 tiene un tiempo de holgura restante $T_s = 50$ ms ($85 - 5$).

- 5 - 25). La planificación para la transmisión después de una transmisión de paquetes desde la cola EF tarda 5 ms, la planificación y conmutación combinada del propio paquete P2 a través de ese enrutador del dominio inalámbrico requiere 5 ms adicionales, y el transporte a través del enlace inalámbrico a la estación móvil del abonado apropiado implica 10 ms adicionales. Cabe destacar que el paquete P2 llega antes de que expire el tiempo de holgura (aproximadamente 30 ms restantes).

Consideremos ahora la comunicación del paquete P3 desde la cola de menor prioridad BE. Los paquetes en la cola BE que tienen la siguiente prioridad más alta están sujetos a retardos de conmutación y transporte, y aquellos paquetes también son retardados el tiempo necesario para el servicio a las colas de mayor prioridad EF y AF.

En el primer enrutador por cable en el nodo 31, los paquetes P1 y P2 se planifican y se conmutan a su través en primer lugar (5 ms + 5 ms), y el paquete P3 de la cola BE se planifica a continuación. Dado que el paquete de la cola BE incurre en un retardo de planificación y conmutación de 5 ms, actualmente está retardado un total de 15 ms. En el paquete P3 procedente de la cola BE incurre un retardo de 95 ms en tránsito en el salto entre los dos enrutadores por cable. Suponiendo que el paquete P3 de la cola BE inicialmente tiene una adjudicación de tiempo de holgura $T_s = 150$ ms, cuando alcanza el segundo enrutador por cable en el nodo 25, ese paquete tiene un tiempo de holgura restante de 40 ms ($150 - 5 - 5 - 5 - 95$).

A su vez, el paquete P3 incurre en dos retardos de 5 ms (10 ms en total) mientras los paquetes son enviados desde las colas EF y AF, y otro retardo de 5 ms para la planificación y conmutación de paquetes a través del segundo enrutador por cable 25/45 (total 15 ms). El transporte por el salto desde el segundo enrutador por cable en 25 hasta el enrutador en el nodo 19 en el dominio inalámbrico tarda 5 ms adicionales, suponiendo que no hay congestión para este simple ejemplo. En el momento de la planificación PF en el enrutador de la red inalámbrica por parte del planificador 43, el paquete P3 tiene un tiempo de holgura restante $T_s = 20$ ms ($40 - 5 - 5 - 5 - 5$).

En el enrutador del dominio inalámbrico en el nodo 19, la planificación para la transmisión después de las transmisiones de paquetes desde las colas EF y AF incurre en un retardo de 10 ms, y la planificación y conmutación del propio paquete P3 a través de ese enrutador del dominio inalámbrico requiere 5 ms adicionales. El transporte real a través del enlace inalámbrico a la estación móvil del abonado apropiado implica 10 ms adicionales. Por lo tanto, el suministro del paquete P3 requiere 25 ms adicionales. Sin embargo, el tiempo de holgura para P3 es solamente de 20 ms, lo que significa que la red no será capaz de suministrar el paquete P3 a la estación del abonado dentro de la adjudicación de tiempo.

Es interesante observar que cada uno de los paquetes P1 y P2 están bien en su respectiva adjudicación de tiempo de holgura, en el ejemplo de la figura 6A. El paquete P1 tenía un tiempo de holgura no usado de aproximadamente 105 ms, y el paquete P2 tenía un tiempo de holgura no usado de aproximadamente 30 ms restantes en su adjudicación de temporización. En la figura 6B, se ha movido al paquete P3 delante del paquete P2 en la cola AF en el enrutador del dominio inalámbrico para fines de planificación PF, dado que no existe ningún impacto material sobre el suministro del paquete P1 o el paquete P2, pero existe un beneficio significativo para el paquete P3. Ésta es la idea central detrás de la técnica de reorganización de paquetes.

El paquete P1 fluye a su través como en el ejemplo de la figura 6A; y los paquetes P2 y P3 fluyen a través de los enrutadores por cable hasta el enrutador del dominio inalámbrico en el nodo 19, como en ese ejemplo anterior. En el nodo 19, el paquete P2 tiene un tiempo de holgura $T_s = 50$ ms; y el paquete P3 tiene un tiempo de holgura $T_s = 20$ ms, como en el ejemplo de la figura 6A. Por lo tanto, el monitor (por ejemplo 51 mostrado en la figura 5) controla el enrutador y el planificador 43 para reordenar los paquetes P3 y P2, para facilitar la transmisión de P3 por delante de P2 y el suministro de P3 dentro de su tiempo de holgura restante. La reordenación podría implicar un ajuste del algoritmo de planificación PF, por ejemplo para aumentar el peso del valor del algoritmo de "equidad" para la cola que contiene el paquete P3 con el tiempo de holgura bajo. Si los paquetes P2 y P3 ya están en la misma cola, el procesamiento podría reordenar eficazmente los paquetes para colocar al P3 por delante del P2. En el ejemplo, el planificador 43 controla al enrutador para mover al paquete P3 desde la cola BE en el nodo 19 hasta la cola AF en ese nodo y para colocar al paquete P3 por delante del paquete P2 en la cola AF.

Dado que P3 se coloca por delante de P2 en la cola AF en el enrutador del dominio inalámbrico, los 20 ms restantes en su adjudicación de tiempo de holgura son suficientes para permitir que la red suministre ese paquete antes de que su adjudicación expire. La planificación para esperar mientras un paquete es transportado desde la cola EF incurre en un retardo de 5 ms. La planificación y conmutación combinadas del propio paquete P3 en 19/43 tarda 5 ms. El transporte del paquete P3 en el enlace de aire tarda 10 ms.

El paquete P2 es retardado, digamos durante 5 ms extra en el ejemplo de la figura 6B. Sin embargo, en el momento de la planificación PF en el enrutador de la red inalámbrica en el nodo 19, el paquete P2 tiene un tiempo de holgura restante $T_s = 50$ ms. La planificación para la transmisión después de una transmisión de paquete de P1 desde la cola EF tarda 5 ms, el retardo para la transmisión de P3 era de 5 ms, la planificación y la conmutación a través de ese enrutador del dominio inalámbrico para el propio paquete P2 requiere 5 ms adicionales, y el transporte a través

del enlace inalámbrico a la estación móvil del abonado apropiado implica 10 ms adicionales. Por lo tanto, el suministro del paquete P2 a la estación móvil del abonado requiere 25 ms, lo que deja aún 25 ms por debajo de la adjudicación del tiempo de holgura para ese paquete.

5 Este segundo ejemplo (figura 6B) muestra que, mediante la reorganización, más paquetes cumplirán sus respectivas adjudicaciones de temporización de lo que, en caso contrario, habría sido posible. El paquete P2 fue retardado ligeramente, como resultado, pero todos los paquetes siguen cumpliendo sus adjudicaciones de temporización.

10 **Algoritmo de planificación de qos sensible a latencia de paquetes integrada que usa puesta en cola equitativa ponderada**

La figura 7 ilustra un modelo de cola simplificado de un enrutador de red típico, que describe con más detalle, cómo esta técnica de reorganización puede aplicarse a la puesta en cola equitativa ponderada, por ejemplo, controlando uno o más de los planificadores por cable 45, 47, 49 en respuesta a la monitorización del tiempo de holgura por parte de uno o más de los dispositivos 53, 55, 57. En cada puerto de salida del enrutador respectivo, los paquetes se ponen en cola en una cola apropiada, en base a la marca QoS asignada a cada paquete, y a continuación son planificados por el planificador de paquetes WFQ fuera del puerto de salida. Se sabe, que si la longitud del buffer para cada cola tiene el tamaño correcto, y que si el tráfico de llegada está vigilado, y hay un flujo por cola, entonces puede afirmarse que cada flujo recibirá cierta velocidad garantizada de servicio ϕ_i , como se ilustra a continuación.

20 Dado que cada cola i está garantizada para procesar paquetes una velocidad mínima de:

$$r_i = \frac{\phi_i}{\sum_{j=1}^N \phi_j} C$$

donde:

25 N = número de flujos activos

C = Capacidad de ancho de banda del enlace de salida

30 ϕ_1, \dots, ϕ_N = números naturales positivos, no iguales a cero, que representan un peso relativo del ancho de banda disponible global.

El tiempo para drenar una cola i es:

$$T_i = \frac{\text{Número de Bytes} * 8}{r_i}$$

35 En otras palabras, se computan todos los bytes de todos los paquetes en una cola particular. Este valor nos dice el tiempo total, que se tardará en drenar una cola particular - enteramente.

Ahora, debemos asegurarnos de que cada paquete actual en la cola será servido dentro del tiempo en que el planificador será capaz de servir ese paquete. Esto puede hacerse de muchas maneras. Una manera es suponer que los paquetes en las colas están ordenados por tiempo de holgura en aumento. Entonces se puede ir paquete por paquete y asegurarse de que el paquete es servido dentro del tiempo de holgura, manteniendo un registro del recorrido del paquete con la menor holgura después de haberlos servido. Este valor es t_d . Ahora, se puede determinar, de manera rápida y aproximada, si una cola particular es capaz o no de alojar a algún paquete, sin superar los tiempos de holgura del actual paquete, dentro de una cola particular, verificando el tiempo de procesamiento del nuevo paquete y t_d . Esto se muestra de forma visual en la figura 8 y la figura 9.

45 Las figuras 8 y 9 representan el efecto de reorganizar proyectando paquetes en un gráfico bidimensional con un eje de bytes y de tiempo. La velocidad a la cual los paquetes salen de la cola se muestra mediante la derivada de la pendiente de la línea de puntos. La línea de puntos describe la velocidad, que es el movimiento de bytes dividido por el intervalo de tiempo.

50 La figura 8 es una vista visual proyectada de paquetes en una cola particular y cuánto retardo puede absorberse debido a la reorganización de un posible paquete sin afectar a ningún otro de los paquetes que ya están en la cola. La velocidad de drenaje de la cola modificada se desplaza en el eje de tiempo en 1 segundo y sigue siendo capaz de cumplir todas las adjudicaciones de temporización. Nótese que la punta del paquete P3 tiene todos los 3 bytes

transmitidos por su plazo de $t = 11$ segundos. Todos los demás paquetes pueden tolerar retardos adicionales y seguir cumpliendo sus adjudicaciones de temporización, como se muestra en el diagrama.

5 En la figura 8, el gráfico muestra el impacto de retardar la planificación de paquetes sobre el cumplimiento de las adjudicaciones de temporización. El rectángulo del paquete representa el retardo de serialización aproximado para recibir y almacenar un paquete en una cola particular. El rectángulo de paquete adyacente a la derecha, representa la cantidad de tiempo de holgura que un paquete particular puede tolerar, y seguir consiguiendo la limitación de adjudicación de tiempo. La línea de puntos a 45 grados representa la velocidad a la cual una cola puede drenar paquetes. La línea de puntos a 45 grados inferior representa la misma línea, pero retardada, la cantidad de retardo es reflejada por un desplazamiento horizontal de la línea de puntos a la derecha, a lo largo del eje de tiempo. En esta figura, se observa que se puede soportar alojar un retardo, desplazando la línea de velocidad de drenaje a la derecha, y seguir cumpliendo todos los plazos de paquetes, hasta el punto, donde los rectángulos del tiempo de holgura intersectan con la línea desplazada. En la figura 8 se observa que se puede alojar un retardo de 1 segundo, que es donde el tiempo de holgura del paquete P3 intersecta con la línea de velocidad de drenaje desplazada.

15 En la figura 9, se observa el efecto de insertar un paquete P0, de modo que todos los paquetes, incluyendo el paquete recién insertado pueden planificarse todos y seguir cumpliendo las adjudicaciones de temporización. Esto puede indicarse simplemente de tal manera que si es necesario reorganizar un paquete recién llegado dado que tiene un tiempo de holgura que corre el peligro de ser superado, entonces su tiempo de procesamiento es menor que t_d , entonces es seguro insertarlo en esa cola.

20 El beneficio de este algoritmo se vuelve evidente al observar un gran dispositivo de red con miles de colas, por ejemplo seis o más colas por puerto y muchos puertos de salida. En lugar de ir por cada cola, paquete por paquete, se puede identificar inmediatamente qué cola es capaz de alojar a un paquete insertado, sin superar los tiempos de holgura para todos los demás paquetes en esa cola particular.

Algoritmo de planificación de QoS sensible a latencia de paquetes integrada que usa puesta en cola equitativa proporcional

30 El algoritmo de puesta en cola de equidad proporcional no tiene garantías de tiempo deterministas que puedan deducirse para cada flujo en una cola particular, debido al hecho de que las condiciones del canal cambian de ventana a ventana y, por lo tanto, las asignaciones de ancho de banda cambian. A partir de las secciones previas, en el segmento final del enlace directo, se sabe en este punto, cuáles son los tiempos de holgura finales de los paquetes y, en base a eso, puede decidirse qué paquetes están en peligro de superar adjudicaciones de temporización, así como saber qué paquetes tienen alta y baja prioridad. La modificación del algoritmo de PF puede describirse de la siguiente manera

```

        If new arriving packet is in danger of exceeding time budget then {
    For each queuei, Start at lowest priority queuei {
    For each packet in queueido{
        If arriving packet can be inserted without exceeding other packets
        slack time then {
            then insert into queue
            done
        } else {
            go to next queue
        }
    } /* continue to next queue */
} /* no queues found */
drop packet
} /* arriving packet in danger of exceeding time budget */
Else {
    Insert packet in assigned queue
}

```

La modificación puede implementarse mediante diversos procedimientos:

40 i. a través de la red - esto permitirá que toda la red realice ajustes finos según sea necesario para asegurar que los paquetes cumplen sus adjudicaciones de temporización.

ii. En el borde de la red de acceso por radio en la ruta del enlace directo - esto permitirá a la red realizar un único ajuste para asegurar que los paquetes cumplen su adjudicación de temporización.

45 Los monitores 51-57 mantienen el rastro de los cronomarcadores de paquetes necesarios para determinar los

tiempos de holgura de los paquetes.

5 Aunque lo anterior ha descrito lo que se considera que es el mejor modo y/o otros ejemplos, se entiende que pueden realizarse diversas modificaciones en ellos y que el asunto descrito en este documento puede implementarse en diversas formas y ejemplos, y que las enseñanzas pueden aplicarse en numerosas aplicaciones, de las cuales solamente algunas se han descrito en este documento. Las siguientes reivindicaciones pretenden reivindicar todas y cada una de las aplicaciones, modificaciones y variaciones que estén dentro del verdadero alcance de las presentes enseñanzas.

10 Apéndice: lista de acrónimos

15 La descripción anterior ha usado gran número de acrónimos para referirse a diversos servicios, mensajes y componentes del sistema. Aunque conocidos en general, el uso de algunos de estos acrónimos no está estrictamente estandarizado en la técnica. Para comodidad del lector, la siguiente lista correlaciona términos con acrónimos, como se usan en la descripción detallada anterior.

Reconocimiento (ACK)

20 Modo de Transferencia Asíncrona (ATM)

Autenticación, Autorización y Estadísticas (AAA)

Estación Base (BS)

25 Estación Transmisora Base (BTS)

Acceso Múltiple por División de Código (CDMA)

30 Módulo de Datos Optimizado (DOM)

Solamente Datos de Evolución de Primera Generación (1xEV-DO)

Agente Externo (FA)

35 Intercambio de Procesador Generalizado (GPS)

Interfaz de Datos de Paquete de Alta Velocidad (HPRD)

40 Agente Local (HA)

Servicios Multimedia Integrados (IMS)

Protocolo de Internet (IP)

45 Kilo-bits por segundo (kbs)

Puesta en Cola de Baja Latencia (LLQ)

50 Mega-bits por segundo (Mbs)

Centro de Conmutación de Servicios Móviles (MSC)

Conmutación Multi-Protocolo Mediante Etiquetas (MPLS)

55 Función de Control de Paquetes (PCF)

Asistente Digital Personal (PDA)

60 Equidad Proporcional (PF)

Calidad del Servicio (QoS)

Red de Acceso por Radio (RAN)

Controlador de Red de Radio (RNC)

Red Inalámbrica de Tercera Generación (3G)

5 Proyecto Asociación 2 de Tercera Generación (3G PP2)

Acceso Múltiple por División de Tiempo (TDMA)

Agente de Usuario (UA)

10

Voz sobre Protocolo de Internet (VoIP)

Puesta en Cola Equitativa Ponderada (WFQ)

REIVINDICACIONES

1. Un procedimiento de planificación de transmisiones de paquetes, para su uso en proporcionar servicios de comunicación de paquetes a dispositivos cliente de abonado inalámbricos a través de una red híbrida que tiene una parte por cable y una parte inalámbrica, comprendiendo el procedimiento:
- determinar una adjudicación de tiempo para el suministro de cada paquete respectivo a través de una combinación de las partes por cable e inalámbrica de la red a cada uno de una pluralidad de los dispositivos cliente de abonado inalámbricos;
- registrar un cronomarcador respectivo que indica el tiempo de entrada en la red para cada paquete;
- enrutar los paquetes a través de la parte por cable de la red a la parte inalámbrica de la red, usando un primer algoritmo de planificación;
- enrutar los paquetes a través de la parte inalámbrica de la red usando un segundo algoritmo de planificación diferente del primer algoritmo de planificación;
- con respecto a un punto en la red por cable o un punto en la red inalámbrica antes de la transmisión de paquetes por enlace inalámbrico a respectivos dispositivos cliente de abonado inalámbricos, restar una diferencia entre el tiempo de llegada de cada paquete en el punto antes de la transmisión por enlace inalámbrico y el tiempo de entrada indicado por el cronomarcador respectivo, de la adjudicación de tiempo para el paquete, para computar un tiempo de holgura que representa una cantidad restante de la adjudicación de tiempo para el suministro de cada paquete respectivo desde dicho punto a través de la red a uno de los dispositivos cliente de abonado inalámbricos; y
- en dicho punto, reordenar al menos dos de los paquetes destinados a diferentes dispositivos cliente de abonado inalámbricos para enrutamiento de acuerdo con al menos uno de los algoritmos de planificación, en base a los tiempos de holgura computados para dichos al menos dos paquetes de tal manera que permitirá el suministro de los paquetes destinados a diferentes dispositivos cliente de abonado inalámbricos antes de que expiren las respectivas adjudicaciones de temporización.
2. El procedimiento de la reivindicación 1, en el que:
- el segundo algoritmo de planificación es un algoritmo de Equidad Proporcional (PF); y
- la reordenación reordena dichos al menos dos paquetes para planificación de transmisión por el enlace inalámbrico de acuerdo con el algoritmo de PF.
3. El procedimiento de la reivindicación 1, en el que:
- el primer algoritmo de planificación es un algoritmo de Puesta en Cola Equitativa Ponderada (WFQ); y
- la reordenación reordena dichos al menos dos paquetes para planificación de transmisión a la parte inalámbrica de la red de acuerdo con el algoritmo de WFQ.
4. El procedimiento de la reivindicación 1, en el que:
- el primer algoritmo de planificación es un algoritmo de Puesta en Cola Equitativa Ponderada (WFQ); y
- el segundo algoritmo de planificación es un algoritmo de Equidad Proporcional (PF).
5. El procedimiento de la reivindicación 4, en el que:
- la reordenación reordena dichos al menos dos paquetes para planificación de transmisión a la parte inalámbrica de la red de acuerdo con el algoritmo de WFQ; y
- el procedimiento comprende además reordenar al menos dos paquetes destinados a diferentes dispositivos cliente de abonado inalámbricos para enrutamiento de acuerdo con el algoritmo de PF, en base a los tiempos de holgura computados.
6. El procedimiento de la reivindicación 1, que comprende además desechar un paquete que tiene un tiempo de holgura demasiado corto para permitir el suministro a un dispositivo cliente de abonado inalámbrico de destino pretendido, para avanzar la transmisión de otro paquete.

7. El procedimiento de la reivindicación 1, en el que la parte inalámbrica de la red es una red 3GPP2 1xEV-DO.
8. El procedimiento de la reivindicación 1, en el que la determinación de la adjudicación de tiempo para cada paquete respectivo comprende:
- 5 determinar un servicio o aplicación de comunicación para el paquete respectivo de entre una pluralidad de servicios o aplicaciones soportados a través de la red; y
- 10 asignar una adjudicación de tiempo asociada con el servicio o aplicación determinado de entre una pluralidad de posibles adjudicaciones de tiempo asociadas con los respectivos servicios o aplicaciones soportados a través de la red.
9. Una red para proporcionar servicio inalámbrico para dispositivos cliente de abonado inalámbricos, que comprende:
- 15 una parte por cable, que incluye al menos un elemento de enrutamiento de paquetes por cable que tiene una primera función planificadora de la transmisión de paquetes asociada para planificar transmisiones de paquetes desde el elemento de enrutamiento de paquetes por cable, utilizando la primera función planificadora un primer algoritmo de planificación;
- 20 una parte inalámbrica para recibir paquetes de la parte por cable y transmitir los paquetes recibidos por uno o más enlaces por aire a los dispositivos cliente de abonado inalámbricos, incluyendo la parte inalámbrica un elemento de transmisión de paquetes inalámbrico que tiene una segunda función planificadora de paquetes asociada para planificar los paquetes recibidos para transmisiones por el uno o más enlaces por aire, utilizando la segunda función planificadora un segundo algoritmo de planificación diferente del primer algoritmo de planificación; y
- 25 un sistema monitor de paquetes para monitorizar flujos de paquetes a través de la red, en el que el sistema monitor de paquetes está dispuesto para:
- 30 registrar un cronomarcador respectivo que indica el tiempo de entrada en la red para cada paquete;
- 35 determinar una adjudicación de tiempo para el suministro de cada paquete respectivo a través de una combinación de las partes por cable e inalámbrica de la red a cada uno de una pluralidad de dispositivos cliente de abonado inalámbricos;
- 40 con respecto a uno de los elementos de transmisión, determinar un tiempo de holgura restante para el suministro de cada paquete respectivo, que representa una cantidad restante de la adjudicación de tiempo para el suministro del paquete, restando una diferencia entre el tiempo de llegada de cada paquete en el un elemento de transmisión y el tiempo de entrada indicado por el cronomarcador respectivo, de la adjudicación de tiempo para el paquete; y
- 45 ordenar a la función planificadora asociada con el un elemento de transmisión que reordene al menos dos de los paquetes destinados para el enrutamiento a través del un elemento de transmisión y el suministro a diferentes dispositivos cliente de abonado inalámbricos, para evitar que expire el tiempo de holgura para el suministro de uno de los al menos dos paquetes, en base a los tiempos de holgura computados para dichos al menos dos paquetes.
- 50 10. La red de la reivindicación 9, en la que:
- el segundo algoritmo de planificación es un algoritmo de Equidad Proporcional (PF); y
- 55 el monitor de paquetes está dispuesto para ordenar a la segunda función planificadora que reordene paquetes para evitar que expire el tiempo de holgura para el suministro de uno de los paquetes.
11. La red de la reivindicación 9, en la que:
- 60 el primer algoritmo de planificación es un algoritmo de Puesta en Cola Equitativa Ponderada (WFQ); y
- el monitor de paquetes está dispuesto para ordenar a la primera función planificadora que reordene paquetes para evitar que expire el tiempo de holgura para el suministro de uno de los paquetes.
12. La red de la reivindicación 9, en la que:

el primer algoritmo de planificación es un algoritmo de Puesta en Cola Equitativa Ponderada (WFQ); y

el segundo algoritmo de planificación es un algoritmo de Equidad Proporcional (PF).

5 13. La red de la reivindicación 9, en la que la parte inalámbrica de la red comprende una red 3GPP2 1xEV-DO.

14. La red de la reivindicación 13, en la que:

10 la parte inalámbrica comprende una Estación Transmisora Base (BTS); y

el al menos un elemento de transmisión de paquetes comprende un Módulo de Datos Optimizado (DOM) en la BTS.

15 15. La red de la reivindicación 14, en la que:

un controlador de red de radio (RNC) sirve como interfaz de la parte por cable y la parte inalámbrica; y

20 el al menos un elemento de enrutamiento de paquetes está implementado en o asociado con el RNC.

16. La red de la reivindicación 9, en la que el sistema monitor de paquetes comprende:

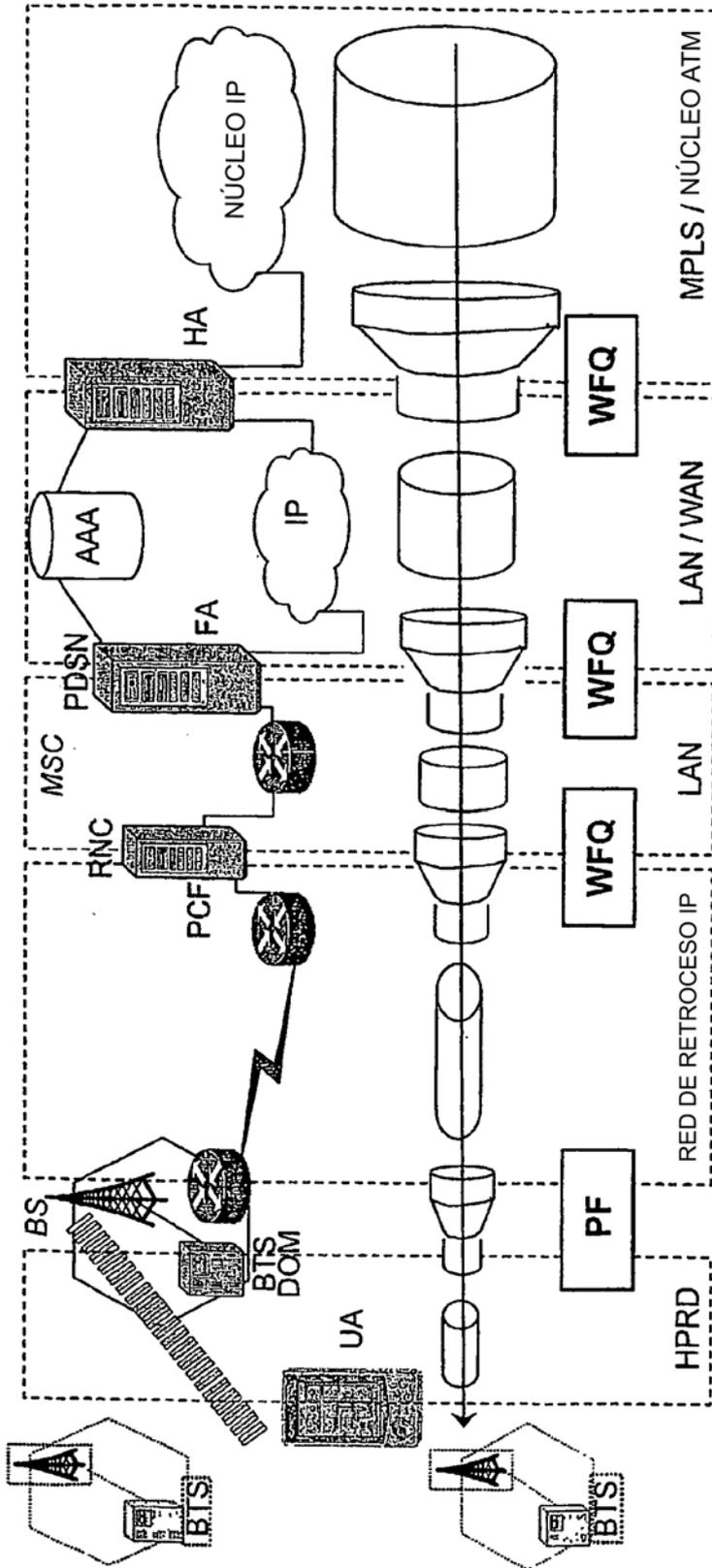
un primer dispositivo monitor asociado con el elemento de enrutamiento de paquetes por cable; y

25 un segundo dispositivo monitor asociado con el elemento de transmisión de paquetes inalámbrico.

17. La red de la reivindicación 16, en la que:

30 el primer dispositivo monitor está dispuesto para generar el cronomarcador para cada paquete respectivo que fluye a través del elemento de enrutamiento de paquetes por cable, que indica el tiempo de entrada en la red para el paquete respectivo; y

el segundo dispositivo monitor está dispuesto para computar el tiempo de holgura restante para cada paquete respectivo durante la llegada del paquete respectivo al elemento de transmisión de paquetes inalámbrico.



MIP	IP
	MPLS
	ATM
	PHY

MIP	MIP				
MIP	PPP	GRE	IP	MAC	PHY

RLP	RMI	TCP	GRE	IP	MAC	PHY

		RMI	TCP	IP	MAC	PHY

MIP	PPP	RLP	MAC	PHY

FIG. 1 -Técnica Anterior

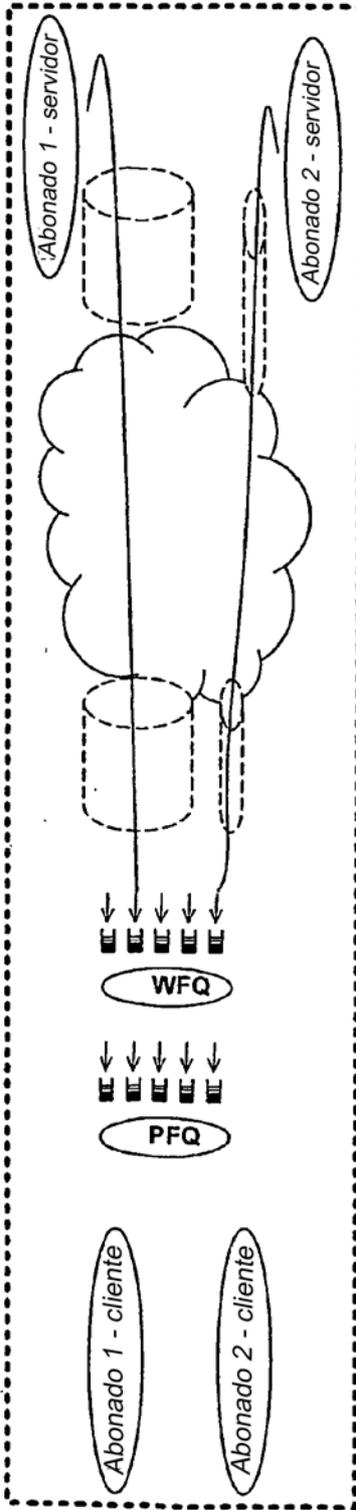


FIG. 2A - Técnica Anterior

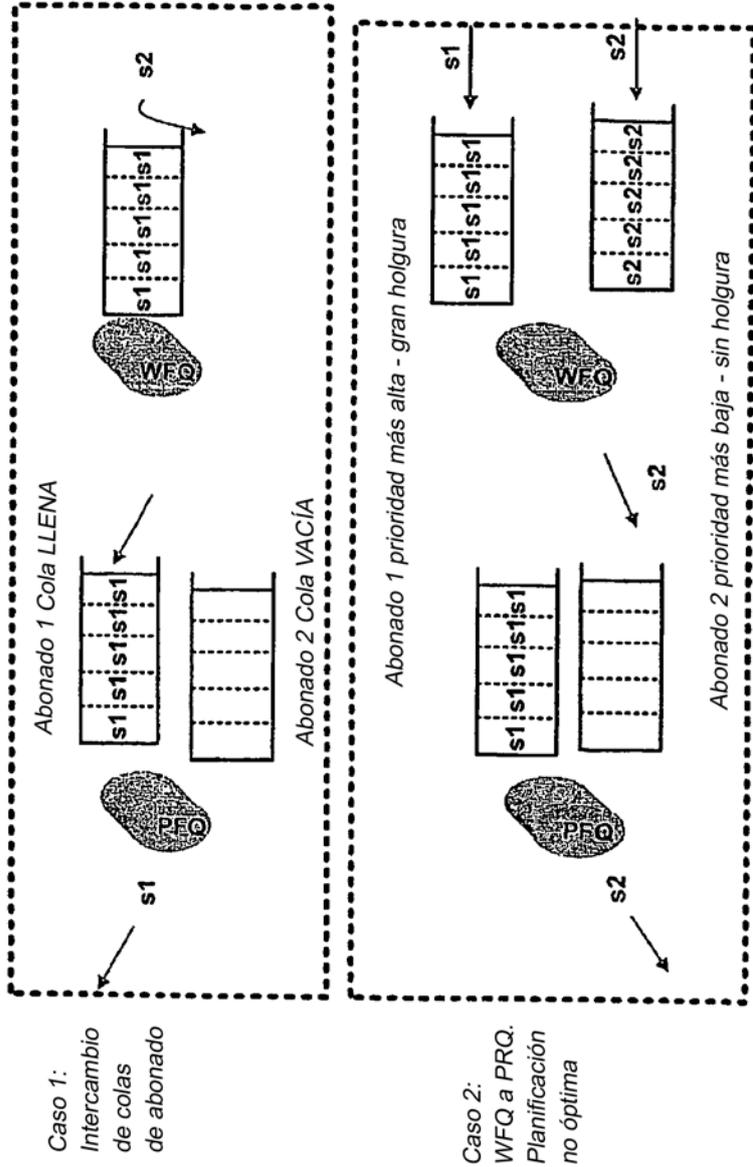


FIG. 2B - Técnica Anterior

FIG. 2C - Técnica Anterior

FIG. 3A

T=1 - al abonado 1 se le asignó un ancho de banda cero, esto se recordará en el siguiente intervalo de tiempo

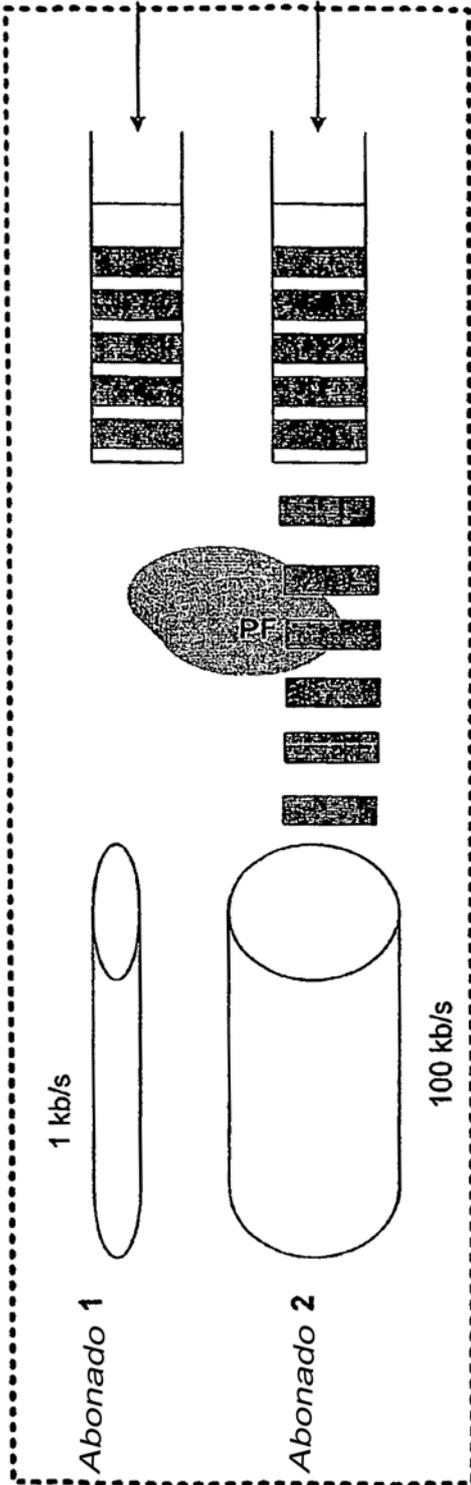
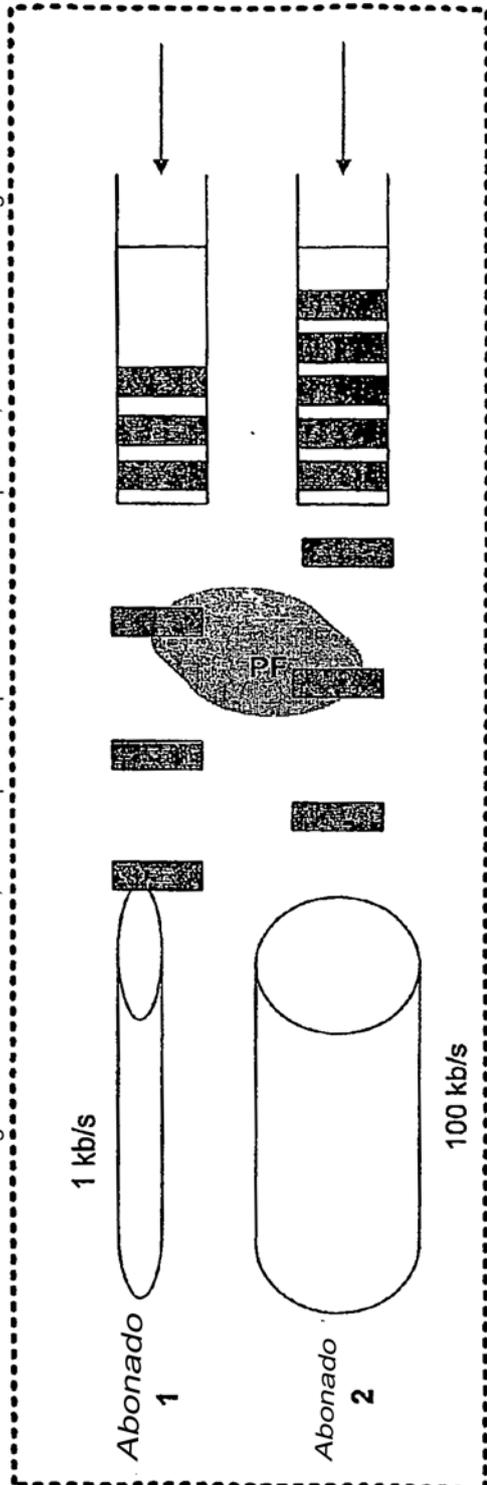


FIG. 3B

T=2 - al abonado 1 se le asignó tanto como solicitó, limitado por la capacidad del enlace por aire, el no usado se asignó al abon. 2



Caso 1 Los paquetes de alta prioridad tienen mucha holgura, pero se siguen planificando por delante de paquetes de baja prioridad sin holgura

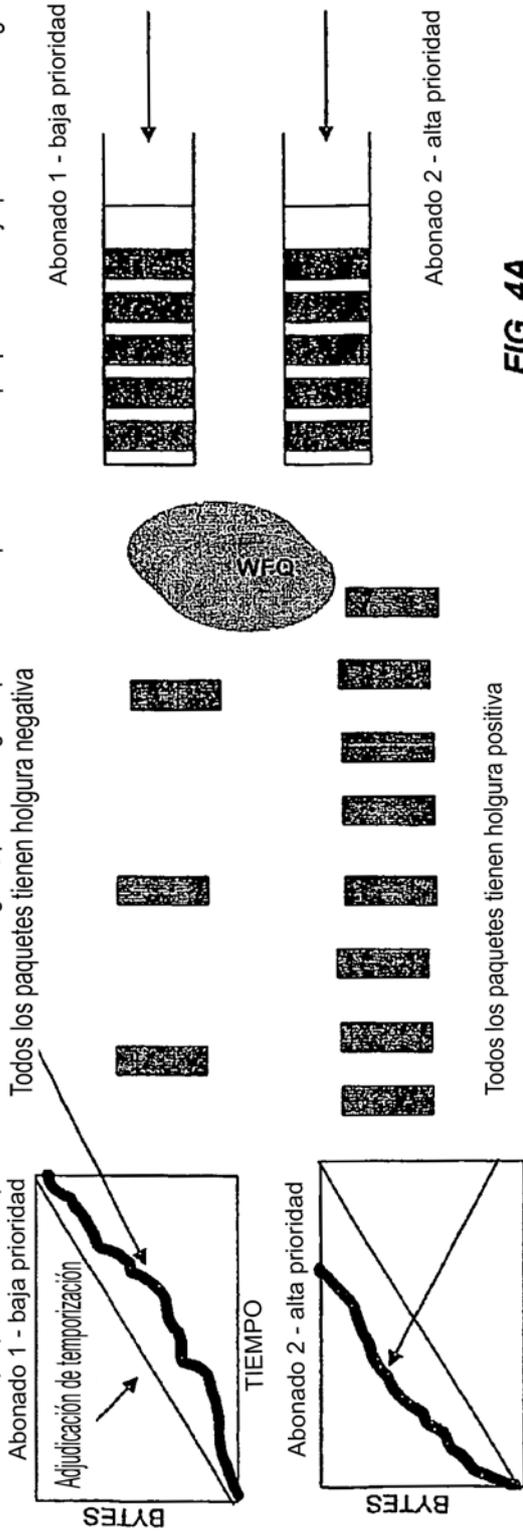


FIG. 4A

Caso 2 = Los paquetes de alta prioridad tienen holgura negativa, pero se siguen planificando detrás de paquetes de baja prioridad

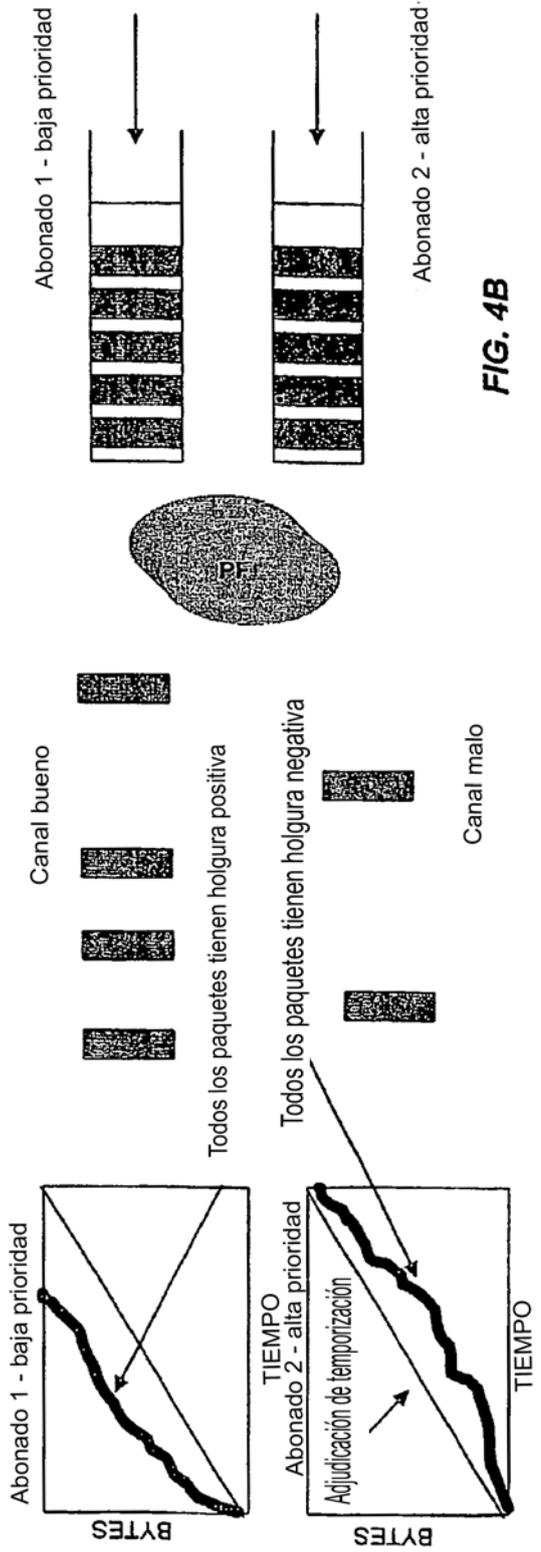


FIG. 4B

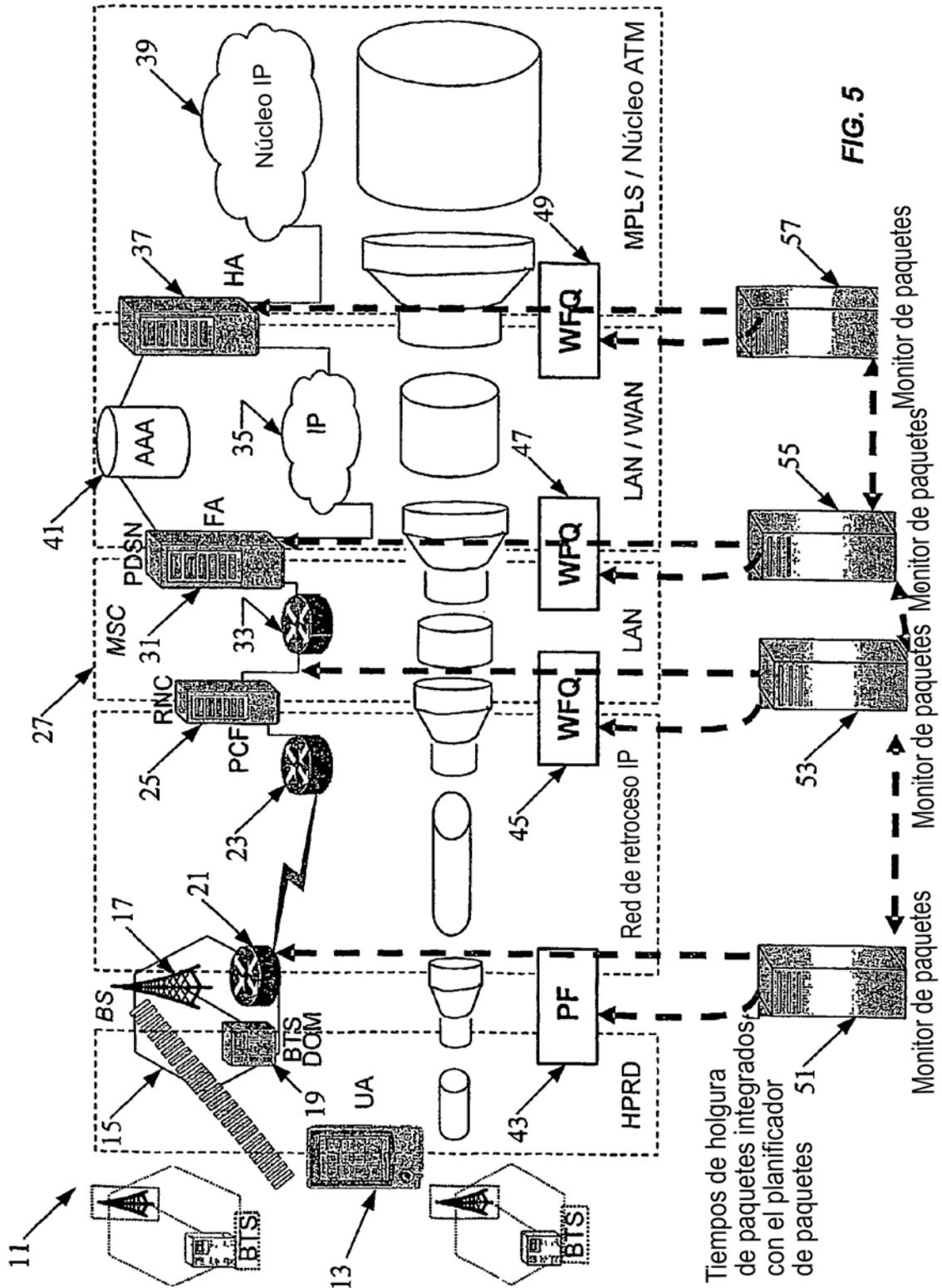
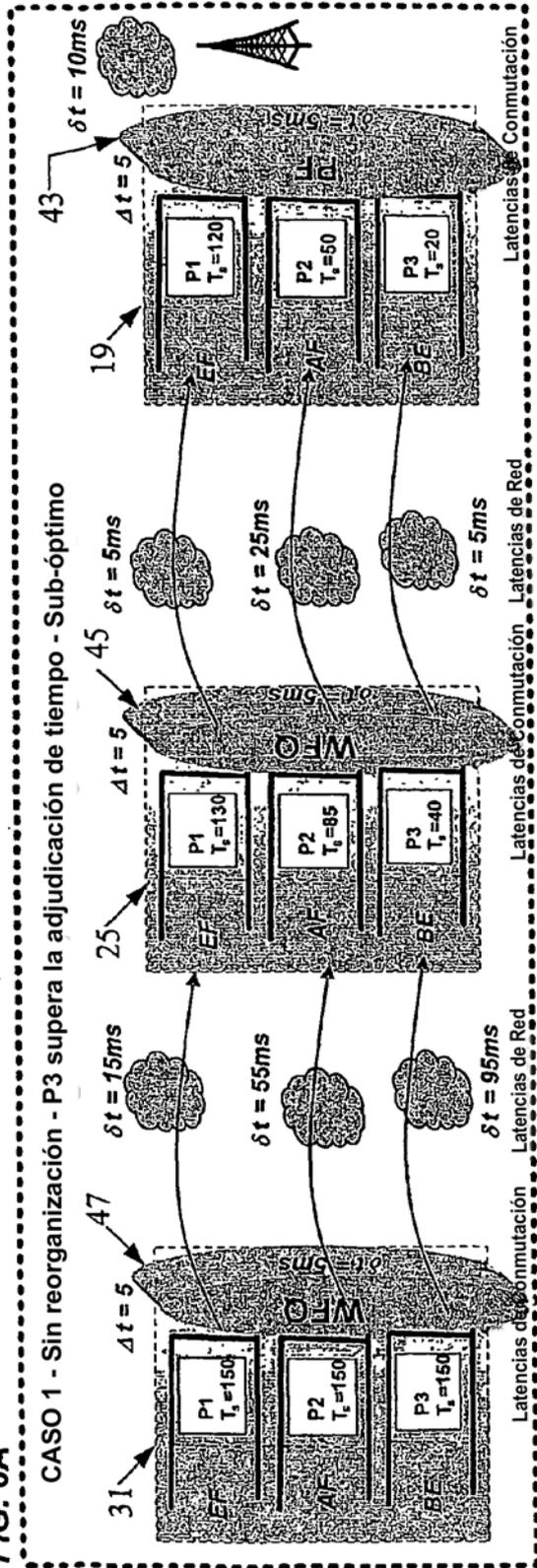


FIG. 5

FIG. 6A



CASO 2 - Se reorganiza P3 → P3 cumple la adjudicación de tiempo - Óptimo

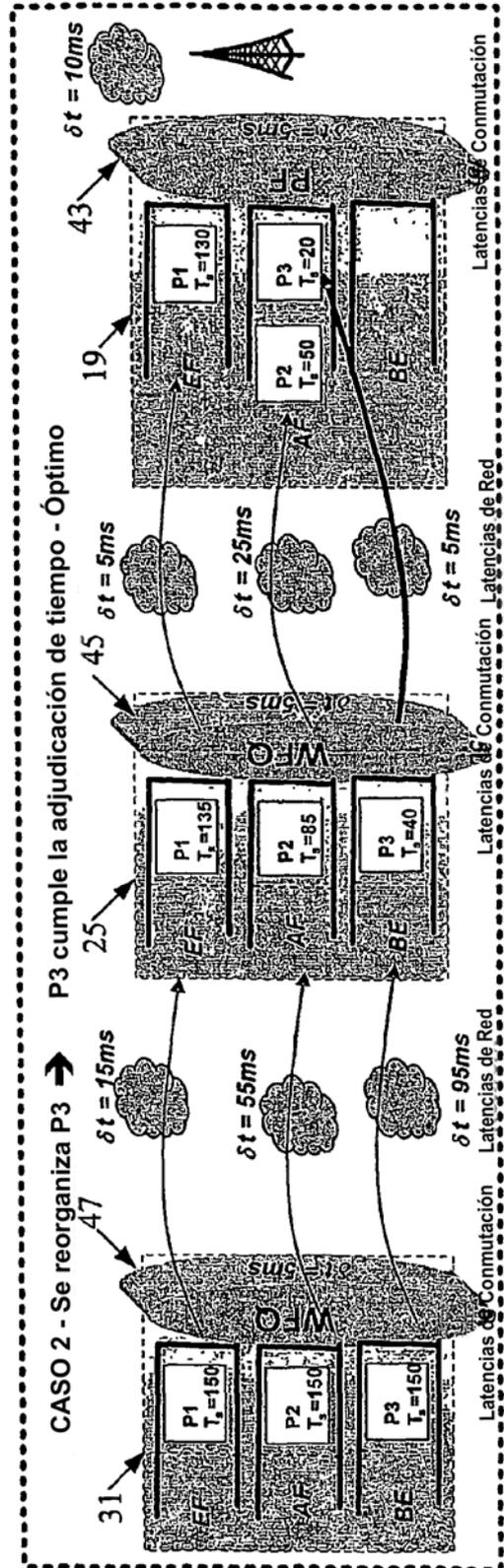


FIG. 6B

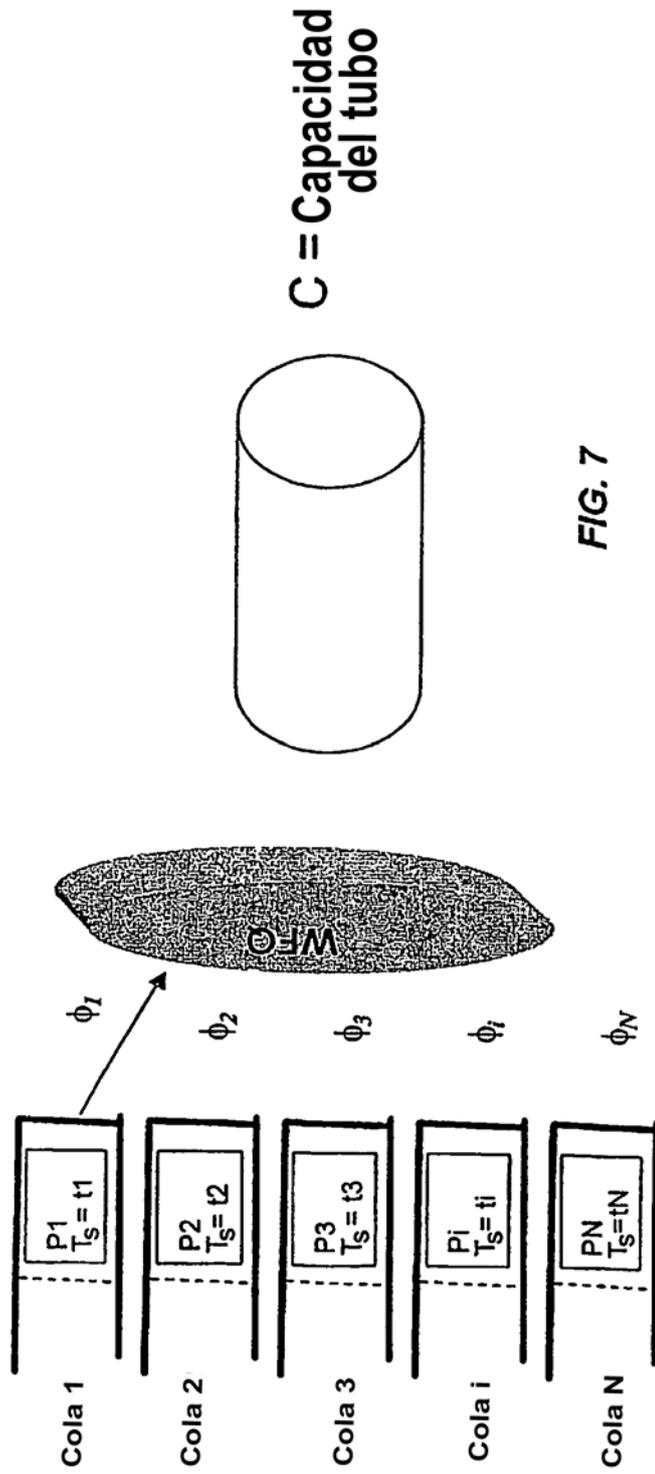


FIG. 7

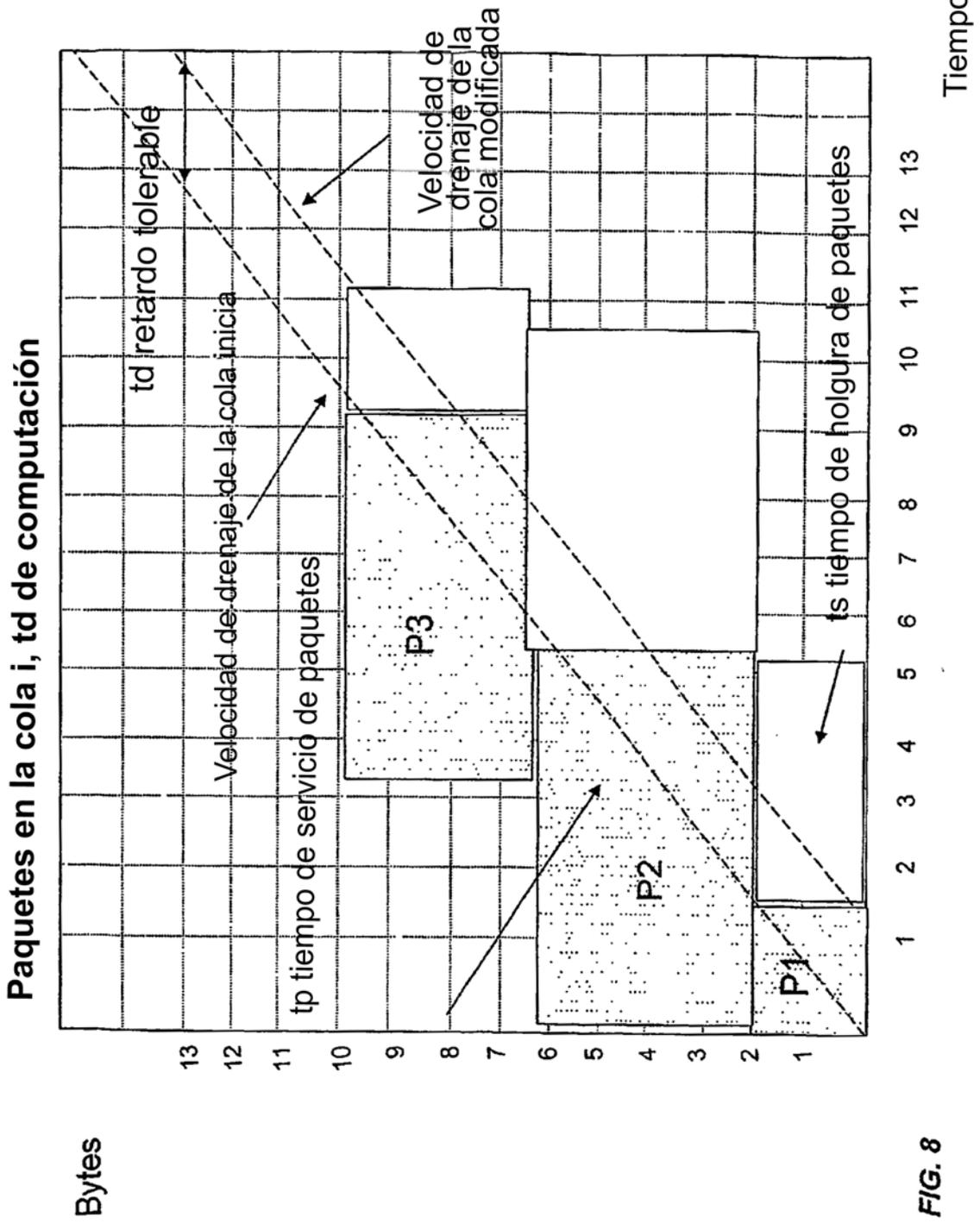


FIG. 8

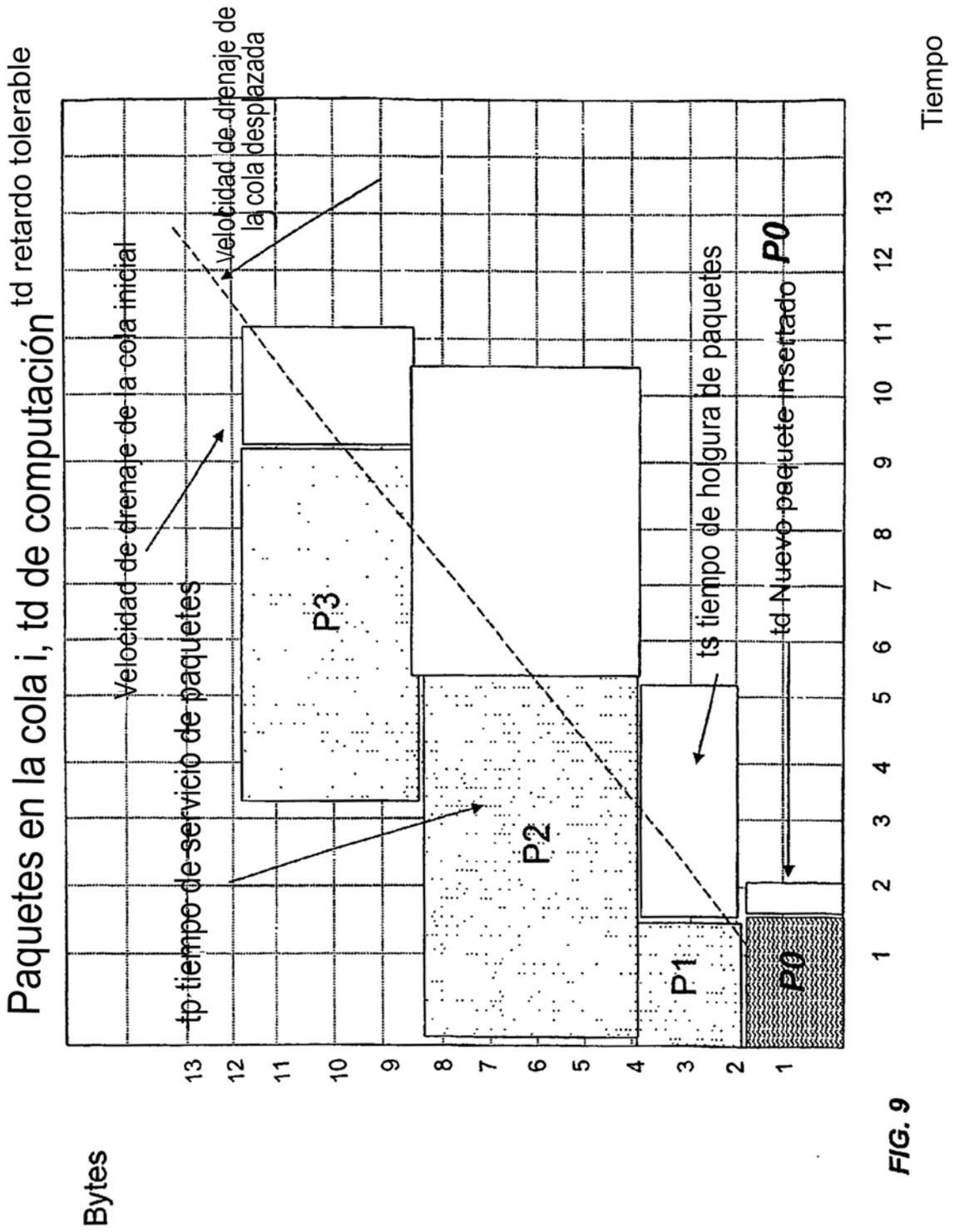


FIG. 9