(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 547 338**

(51) Int. CI.:

H04L 29/08 (2006.01)
H04L 29/06 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **30.09.2011 E 11183520 (3)**

(97) Fecha y número de publicación de la concesión europea: **26.08.2015 EP 2575323**

(54) Título: **Suministro de contenido de un servidor a un cliente**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
05.10.2015

(73) Titular/es:

DEUTSCHE TELEKOM AG (100.0%)
Friedrich-Ebert-Allee 140
53113 Bonn, DE

(72) Inventor/es:

HUI, PAN y
PETZ, AGOSTON

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

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

Suministro de contenido de un servidor a un cliente

Sector técnico

La presente invención se refiere a un método, preferiblemente implementado en un ordenador, de suministro de contenido de un servidor a un cliente y, más particularmente, a un servidor dispuesto para ello, un cliente dispuesto para ello, así como programas informáticos para implementar tal servidor o tal cliente.

Antecedentes

Debido en gran medida a la creciente popularidad de los dispositivos electrónicos móviles de alto rendimiento tales como particularmente los teléfonos inteligentes y otros, así como los contratos de acceso a datos móviles de alta capacidad tales como las suscripciones a datos de tarifa plana y otros, el tráfico de datos móviles se ha incrementado mucho recientemente. En consecuencia, las demandas de redes inalámbricas tales como las reces celulares o las redes WiFi están creciendo también de forma continua, de manera que se están produciendo frecuentemente problemas de capacidad.

Un ejemplo de una situación en la cual están implicadas demandas particularmente elevadas para tráfico de datos celulares se refiere al video bajo demanda en móviles. Servicios tales como la plataforma de difusión de películas en la web, YouTube o los ofrecidos por los canales de TV local son cada vez más populares. Con la llegada de los dispositivos móviles de alto rendimiento tales como los teléfonos inteligentes, los usuarios también desean utilizar estos servicios en sus dispositivos móviles, sobrecargando las redes celulares. Potenciales similitudes locales en las solicitudes para este contenido, tal como ordenadores en el mismo tren, todos deseando captar el episodio de la última noche de una popular serie de TV o ver el último video viral, pueden provocar problemas de funcionamiento o incluso el colapso de la red móvil correspondiente, particularmente la red celular.

Otro ejemplo de tal situación se refiere a eventos con grandes multitudes. Algunos eventos conllevan multitudes en áreas en las que las redes celulares u otras redes inalámbricas están dimensionadas para menos gente. Tales eventos incluyen grandes eventos deportivos en exteriores, tales como maratones, que se despliegan en una extensa área, haciendo difícil el despliegue de capacidad extra en una ubicación particular, bodas reales o un evento popular, donde se reúnen grandes multitudes que desean igualmente poder ver el evento en sus dispositivos móviles. Muchas personas de la multitud tendrán intereses similares y solicitarán los mismos datos.

En este contexto, los trabajos recientes han explorado el potencial de la descarga de datos desde redes inalámbricas de un comparativamente bajo rendimiento, tal como redes celulares antiguas, a modernas redes inalámbricas de un comparativamente alto rendimiento, tales como las redes WiFi. Si los datos no son críticos con el tiempo, puede conseguirse un ahorro mediante transmisiones diferidas y el uso de técnicas de transmisión en red que toleran el retardo [5]. Recientes avances en las redes centradas en la información [1] convierten al contenido en un actor importante en una red, y la captación de contenido popular en las estaciones de base y en los dispositivos móviles o en los nodos móviles reduce la carga en las redes de acceso compartido, que a menudo pueden ser un cuello de botella [3].

Las llamadas técnicas de descarga pueden solucionar los problemas de capacidad que plagan las redes celulares u otras redes móviles inalámbricas [6, 2]. Aunque hoy en día la descarga de datos se considera desde las perspectivas teórica y conceptual, se ha hecho poco en soportar tales técnicas en redes reales. Las arquitecturas de red celulares y/u otras arquitecturas de red inalámbricas a menudo no detectan suficientemente la red y el contexto de datos para tomar decisiones de descarga inteligentes.

Por lo tanto, existe la necesidad de un sistema eficientemente implementable para el suministro de contenido en una red al menos parcialmente inalámbrica capaz de proporcionar contenido de manera eficiente en la infraestructura de red existente, satisfaciendo las expectativas del usuario para un oportuno suministro.

El documento US 2009/0088188 A1 se refiere a un método, dispositivo móvil de comunicación y sistema para una descarga selectiva a un dispositivo de comunicación móvil.

Descripción de la invención

De acuerdo con la invención, esta necesidad es resuelta mediante un método tal como se define mediante las características de la reivindicación 1 independiente, por un servidor tal como se define mediante las características de la reivindicación 6 independiente, y por un programa informático tal como se define mediante las características de la reivindicación 9 independiente. Realizaciones preferidas están sujetas a las reivindicaciones dependientes.

En particular, lo esencial de la invención es lo siguiente: Un método de suministro de contenido a través de una red de un servidor a un cliente que tiene una primera interfaz de red comprende: proporcionar el cliente una solicitud de contenido al servidor a través de un primer vector en el que el primer vector está asociado a la primera interfaz de red; recibir el servidor información de contexto acerca del contexto del cliente; separar el servidor el contenido para

ser suministrado al cliente de acuerdo con la solicitud en contenido pequeño y contenido grande; definir el servidor un procedimiento de transferencia para el contenido grande basándose en la información de contexto; proporcionar el servidor una primera respuesta al cliente por medio del primer vector, donde la primera respuesta comprende el contenido pequeño e información acerca del procedimiento de transferencia del contenido grande; y recibir el cliente el contenido grande de acuerdo con la información acerca del procedimiento de transferencia del contenido grande. El método puede ser particularmente un método implementado por ordenador.

En el contexto de la invención, el término "contenido" se refiere a información de cualquier formato que pueda proporcionar valor para un usuario final o audiencia y que pueda ser solicitado a través de un cliente. Generalmente, tal contenido puede ser proporcionado a través de un medio adecuado tal como televisión, discos compactos de audio o de datos, así como eventos en directo tales como conferencias y montajes escénicos. No obstante, en el contexto de la invención, el contenido puede ser proporcionado particularmente a través de la red, que puede ser o comprender Internet y particularmente la red extensa mundial (WWW). Por ejemplo, el contenido puede ser suministrado utilizando un lenguaje de marcado tal como el lenguaje de marcado de hipertexto (HTML – HyperText Markup Language, en inglés) que posiblemente incluye objetos embebidos u otros.

Así, el contenido puede particularmente ser uno o una pluralidad de archivos de datos. Tales archivos de datos de contenido pueden ser archivos de datos de HTML o cuadros de HTML, archivos de datos de imagen, por ejemplo, en el formato de la junta del grupo de expertos en fotografía (JPG o JPEG – Joint Photographic Experts Group, en inglés) o en el formato de intercambio de gráficos (GIF – Graphics Interchange Format, en inglés), archivos de datos de música, por ejemplo, en los grupos de expertos en imágenes en movimiento (MPEG – Moving Pictures Experts Groups, en inglés)-1 o el formato de capa III de audio MPEG-2 denominado comúnmente MP3, archivos de datos de películas, por ejemplo en el formato QuickTime (MOV), en el formato MPEG-4 (MP4) o en el formato de formato de onda de choque (SWF – ShockWave Flash, en inglés), o archivos de programa embebidos tales como applets de Java u otros.

En conexión con el contenido para ser suministrado al cliente, el término "pequeño" se utiliza en esta memoria para referirse particularmente a contenido que es lo suficientemente pequeño en tamaño para posibilitar un suministro eficiente a través del primer vector. Contenido pequeño puede, por ejemplo, comprender cuadros de HTML, teletipos de noticias, avances, imágenes de baja resolución u otros. Por ejemplo, en redes en las que al menos segmentos y en particular el primer vector están implementados como redes celulares o redes de 3G, el contenido puede estar definido como pequeño si su tamaño es menor de aproximadamente 512 Kilobytes, o menor de aproximadamente 1 Megabyte (MB), o menor de aproximadamente 2 MB, o menor de aproximadamente 5 MB, o menor de aproximadamente 10 MB u otros. Como contrapartida, el término "grande" tal como se utiliza en esta memoria en conexión con el contenido para ser suministrado, puede referirse particularmente al contenido que posiblemente no puede ser suministrado a través del primer vector con un rendimiento satisfactorio. En particular, el contenido grande puede ser cualquier contenido que no sea pequeño según se ha especificado anteriormente y puede, por ejemplo, comprender música, videos de transmisión en directo, imágenes de alta resolución y otros.

Servidor en el contexto de la invención puede referirse a un ordenador que está conectado a una red a través de una interfaz de red y que proporciona servicios a otros nodos o dispositivos dentro de la red o que está interconectado a la red. El servidor puede ser un servidor web que puede particularmente referirse a ordenadores que proporcionan servicios web a través de internet y particularmente a través de WWW. Tales servicios web pueden comprender suministro de contenido a través de la WWW o proporcionar aplicaciones web a través de la WWW. Cliente en el contexto de la invención puede referirse particularmente a un dispositivo electrónico preferiblemente móvil que tiene la intención de recibir servicios sobre una red preferiblemente al menos parcialmente inalámbrica. La primera interfaz de red puede estar adaptada para permitir la conexión del cliente a una red inalámbrica y particularmente a una red celular tal como 3G. Por ejemplo, clientes pueden ser teléfonos inteligentes, asistentes digitales personales (PDA – Personal Digital Assistant, en inglés), dispositivos de transmisión en tiempo real de multimedios, teléfonos móviles, ordenadores portátiles, ordenadores de tableta, dispositivos de sistema de localización global (GPS – Global Positioning System, en inglés), combinaciones de los mismos u otros. El término "solicitud" tal como se utiliza en esta memoria se refiere a cualquier solicitud adecuada para el suministro de contenido a través de la red. Por ejemplo, tal solicitud puede ser una solicitud de HTML.

En el contexto de la invención, el término "vector" puede referirse particularmente a una tecnología de red o con más detalle a una ruta dentro de una tecnología de red específica. Por ejemplo, un vector puede estar basado en una red celular tal como 3G u otra o una red de área local inalámbrica (WLAN – Wireless Local Area Network, en inglés) tal como WiFi u otra. El término "red celular" tal como se utiliza en esta memoria puede referirse particularmente a redes de radio distribuidas en áreas denominadas células. Por ello, cada célula puede estar servida por al menos un transmisor receptor de ubicación fija conocido como un sitio de célula o estación de base. Esto permite que un gran número de dispositivos móviles o transmisores receptores portátiles tales como, por ejemplo, teléfonos móviles, localizadores, ordenadores portátiles, ordenadores de tableta u otros, se comuniquen en cualquier parte de la red a través de las estaciones de base incluso si algunos de los dispositivos móviles se están desplazando en más de una célula durante la transmisión. El término "3G" tal como se utiliza en esta memoria puede referirse particularmente a las telecomunicaciones de móviles de tercera generación que son una generación de estándares para teléfonos móviles y servicios de telecomunicación móviles que cumplen las especificaciones Telecomunicaciones Móviles Internacionales-2000 (IMT-2000 – International Mobile Telecommunications-2000, en inglés) de la Unión de

Telecomunicación Internacional (ITU – International Telecommunication Union, en inglés). En 3G, los servicios de aplicación pueden incluir teléfono de voz inalámbrico de área extensa, acceso a internet para móviles, videollamadas y TV para móviles en entornos móviles. Recientes versiones del 3G, por ejemplo denotadas como 3,5G o 3,75G, pueden proporcionar también acceso de banda ancha de varios Megabits por segundo para dispositivos móviles. El término "WiFi" tal como se utiliza en esta memoria puede referirse particularmente a un mecanismo para conectar de manera inalámbrica dispositivos móviles de tales como ordenadores personales, consolas de videojuegos, teléfonos inteligentes o reproductores de audio digitales, a la internet, a través de un punto de acceso a red inalámbrica o punto de conexión. El término "teléfono inteligente" tal como se utiliza en esta memoria puede referirse particularmente a un teléfono móvil de gama alta que combina varias tecnologías y un teléfono móvil. Tales tecnologías de dispositivos electrónicos pueden ser, por ejemplo, reproductores de medios portátiles, cámaras, aplicaciones de acceso a internet tales como navegadores web, navegación por GPS, acceso por WiFi y por banda ancha para móviles u otra y combinaciones de los mismos. El término "contexto" tal como se utiliza en esta memoria puede referirse particularmente a contexto de datos y contexto de red del cliente. Puede comprender información acerca del dispositivo del cliente, de su entorno local, que puede cambiar de manera continua mientras el cliente se está desplazando, y de sus enlaces de red o interfaces de red y su disponibilidad.

El método de acuerdo con la invención permite transferencias de contenido de descarga eficientes que afectarían negativamente al rendimiento de una red comparativamente débil o a un segmento de la misma. En particular, puede permitir la descarga de contenido grande o pesado desde una red celular. El término "descarga" tal como se utiliza en esta memoria se refiere a evitar la transferencia de contenido a través de un vector, pero permitiendo la recepción del citado contenido de otro modo. Así, el cliente puede ser suministrado con un rendimiento comparativamente alto, de manera que su usuario pueda ser suministrado a tiempo.

Proporcionando información acerca del contexto del cliente al servidor, el método puede permitir hacer uso de patrones de movilidad de cliente y predicciones para encontrar recursos de red adecuados para ser dirigidos en el procedimiento de transferencia del contenido grande y permitir con ello un suministro avanzado de contenido. Puede particularmente permitir el suministro de una arquitectura eficiente para utilizar contexto de usuario y propiedades de datos en la toma de decisiones de descarga. Aún más, el método puede permitir el suministro de un modo comparativamente simple para que los desarrolladores prioricen contenido para mostrar que puede ser fácilmente descargado y que es crítico con el tiempo. Puede asimismo permitir una implementación comparativamente fácil.

En el método de acuerdo con la invención, puede proporcionarse una arquitectura de servidor web que detecta la movilidad de los consumidores de contenido y el contexto de la red y el contenido de datos. Más detalladamente, esta arquitectura puede hacer posible el envío de diferentes partes de contenido mediante diferentes tecnologías de red, permitiendo la descarga de contenido grande o contenido de alto consumo de recursos desde las redes celulares, en particular si tal contenido no es crítico con el tiempo. Los clientes o dispositivos móviles pueden recombinar las partes antes de proporcionarlas al usuario. Este desacoplamiento de contenido con respecto a las solicitudes del usuario puede permitir también un suministro avanzado de contenido a sitios en los que estará el usuario, en lugar de donde el usuario se encuentra actualmente. En el entorno altamente móvil de hoy en día, este guardado previo en una memoria oculta puede aprovechar al máximo las oportunidades de descarga.

El método de acuerdo con la invención puede habilitar la descarga de datos basada en contexto sin requerir nuevo software en las torres celulares y con cambios comparativamente pequeños en los clientes y servidores. Esto puede permitir una comparativamente rápida adopción de descarga de datos y puede facilitar la carga de trabajo de los programadores de la red. En el método de acuerdo con la invención el contenido (web) puede dividirse en dos partes conceptuales: contenido grande, tal como imágenes, videos de transmisión en tiempo real, música, etc., y el resto, por ejemplo de los cuadros de HTML que en sí mismos pueden contener elementos de menor contenido, tal como teletipos de noticias, avances, etc. Asimismo, pueden distinguirse dos vectores de suministro (ver a continuación), tales como 3G o cualquier comunicación de red celular y el vector de descarga de contenido que puede tomar muchas formas, aunque puede requerir alguna memoria oculta de contenido local y puede basarse en WiFi para su suministro de último salto. En particular, el contenido pequeño puede ser transferido en 3G y el contenido grande puede ser descargado cuando resulte beneficioso y eventualmente con restricciones de retardo. Por ello, no es necesario terminar y reiniciar las sesiones activas cuando se cambien las tecnologías, lo que puede perturbar la experiencia del usuario y puede conducir a una nueva transferencia de páginas parcialmente recibidas. Por el contrario, cuando un vector de suministro alternativo está disponible, el servidor puede descargar la mayor parte de la comunicación pero proporcionar aún una sesión continua con mínimo uso de 3G.

La información de contexto puede ser recibida por el servidor de varias maneras tal como proporcionada por nodos que tienen información relevante acerca del cliente y otros. Preferiblemente, la solicitud proporcionada por el cliente al servidor comprende la información de contexto acerca del contexto del cliente. Por ello, la información de contexto puede particularmente ser parte de una solicitud de HTML en la que pueden utilizarse etiquetas de contexto específicas. Tal inclusión de la información de contexto en la solicitud permite una eficiente provisión de toda la información necesaria del cliente al servidor.

Preferiblemente, el cliente comprende una segunda interfaz de red, la información acerca del procedimiento de transferencia del contenido grande comprende información acerca de un segundo vector y el cliente recibe el contenido grande a través del segundo vector, donde el segundo vector está asociado a la segunda interfaz de red.

En particular, la segunda interfaz de red puede ser una interfaz de WLAN o una interfaz de WiFi. En tal situación, como está comúnmente implementado en dispositivos móviles tales como los teléfonos inteligentes hoy en día, el suministro de datos o de contenido puede ser eficientemente descargado del primer vector comparativamente de baja gama, por ejemplo, asociado a una interfaz de red celular, al segundo vector comparativamente de alta gama, por ejemplo, asociado a una interfaz de red de WiFi. Para el usuario del cliente, el contenido puede parecer de rendimiento comparativamente alto o en un momento adecuado en el cliente incluso aunque la solicitud se proporcione a través del primer vector.

Por ello, el servidor preferiblemente proporciona el contenido grande a un nodo de red que es alcanzable por el cliente a través del segundo vector. Tal nodo de red puede, por ejemplo, ser un punto de acceso, por ejemplo conectado al servidor mediante una conexión por cable y alcanzable de manera inalámbrica por el cliente. Puede ser también un punto de acceso que no es alcanzable por el cliente cuando se ha proporcionado la solicitud, sino cuando el contenido va a ser suministrado. El servidor puede en este caso incluir información del punto de acceso en la información acerca del procedimiento de transferencia del contenido grande, en el que el movimiento del cliente es estimado y se elige un punto de acceso adecuado.

Preferiblemente, el servidor proporciona una segunda respuesta al cliente a través del segundo vector, que comprende el contenido grande. Así, el cliente puede recibir la primera respuesta que comprende el contenido pequeño a través del primer vector, por ejemplo, una red celular, y la segunda respuesta que comprende el contenido grande a través del segundo vector, por ejemplo, una red WiFi.

Otro aspecto de la invención se refiere a un servidor que comprende una unidad de procesamiento de solicitud, una unidad de gestión de contexto y una unidad de procesamiento de respuesta, en el que: la unidad de procesamiento de solicitud está dispuesta para recibir una solicitud de contenido desde un cliente a través de un primer vector, para recibir información de contexto acerca del contexto del cliente y para proporcionar la información de contexto a la unidad de gestión de contexto; la unidad de gestión de contexto está dispuesta para definir un procedimiento de transferencia para el contenido grande basándose en la información de contexto y para proporcionar información acerca del procedimiento de transferencia del contenido grande a la unidad de procesamiento de respuesta; y una unidad de procesamiento de respuesta comprende reglas de separación predefinidas y está dispuesta para separar el contenido que va a ser entregado al cliente en contenido pequeño y contenido grande aplicando las reglas de separación predefinidas, y para proporcionar una primera respuesta al cliente por medio del primer vector que comprende el contenido pequeño, e información acerca del procedimiento de transferencia del contenido grande. Como se ha descrito anteriormente, el servidor puede particularmente ser un ordenador que ejecuta un servicio web y que tiene una interfaz de red conectada a una red. Tal servidor permite implementar de manera eficiente un sistema para aplicar el método descrito anteriormente y conseguir los correspondientes efectos ventajosos descritos anteriormente.

Por ello, la unidad de procesamiento de solicitud preferiblemente está dispuesta para separar la información de contexto de la solicitud del cliente. Así, el servidor puede separar de manera eficiente la información de contexto de la solicitud y procesarla. Preferiblemente, la unidad de gestión de contexto está dispuesta para hacer un seguimiento del cliente por medio de un identificador de cliente. El identificador de cliente puede, por ejemplo, ser una dirección de protocolo de internet (IP – Internet Protocol, en inglés) y en particular una dirección de IP de red celular. Esto permite al servidor coordinar de manera eficiente el procesamiento separado de contexto y contenido en particular también en situaciones en las que deben procesarse varias solicitudes al mismo tiempo.

Otro aspecto más de la invención se refiere a un programa informático que comprende medios de código de programa que están dispuestos para implementar las características del servidor descritas anteriormente cuando es ejecutado. Tal programa informático permite una eficiente y conveniente distribución e implementación del servidor descrito.

Otro aspecto más de la invención se refiere a un cliente que comprende una unidad de generación de solicitud, una unidad de procesamiento de respuesta y una unidad de obtención de contexto, en el que: la unidad de obtención de contexto está dispuesta para obtener información de contexto acerca del contexto del cliente y para proporcionar la información acerca del contexto del cliente; la unidad de generación de solicitud está dispuesta para proporcionar una solicitud de contenido a un servidor a través del primer vector, comprendiendo la solicitud la información acerca del contexto del cliente; y la unidad de procesamiento de respuesta está dispuesta para recibir una primera respuesta desde el servidor a través del primer vector que comprende contenido pequeño e información acerca de un procedimiento de transferencia de contenido grande y para recibir el contenido grande de acuerdo con la información acerca del procedimiento de transferencia del contenido grande. Como se ha descrito anteriormente, el cliente puede particularmente ser un dispositivo móvil tal como un teléfono inteligente u otro. Tal cliente permite una eficiente implementación del sistema para aplicar el método descrito anteriormente y alcanzar los correspondientes efectos ventajosos descritos anteriormente.

Preferiblemente, la unidad de generación de solicitud es simplemente un componente de complemento de un programa navegador. Con tal componente de complemento el cliente puede ser eficientemente implementado, pudiendo utilizarse en él tecnologías web estándar ya existentes en el programa navegador.

Otro aspecto más de la invención se refiere a un programa informático que comprende medios de código de programa que están dispuestos para implementar las características del cliente descritas anteriormente cuando es ejecutado. Tal programa informático permite una eficiente y conveniente distribución e implementación del cliente descrito.

Otro aspecto de la presente descripción se refiere a un sistema que comprende un cliente tal como el descrito anteriormente y a un cliente tal como el descrito anteriormente. Independientemente de este sistema, el cliente y el servidor pueden estar correspondientemente dispuestos de manera que juntos aplican e implementan al máximo el método de acuerdo con la invención.

Estos y otros aspectos de la invención resultarán evidentes y claros con referencia a la realización o realizaciones descrita o descritas en lo que sigue en esta memoria.

Breve descripción de los dibujos

El método, el servidor, el cliente y los programas informáticos de acuerdo con la invención se describen con más detalle en lo que sigue a modo de realizaciones de ejemplo y con referencia a los dibujos adjuntos, en los cuales:

la Fig. 1 muestra una arquitectura general para implementar una realización del método de acuerdo con la invención con una realización de un servidor de acuerdo con la invención y una realización de un cliente de acuerdo con la invención;

la Fig. 2 muestra una arquitectura de servidor de una realización del servidor de acuerdo con la otra invención;

la Fig. 3 muestra los resultados de latencia con su desviación estándar para diferentes puntos de acceso mediante WiFi a partir de un primer experimento de una realización de la invención;

la Fig. 4 muestra el uso del ancho de banda de 3G durante una operación regular en comparación con la operación que incluye descarga de datos a partir de un segundo experimento de una realización de la invención; y

la Fig. 5 muestra el uso de 3G a lo largo del tiempo a partir de un tercer experimento de una realización de la invención.

Modo o modos para llevar a cabo la invención

Las arquitecturas de red modernas a menudo consisten en un servidor web de alto rendimiento, por ejemplo, escrito en C/C++, con aplicaciones web, por ejemplo, a menudo escritas en lenguajes orientados al objeto tales como Python u otros. La arquitectura descrita en esta memoria como una realización de la invención dispone una capa media ligera entre estos componentes en el servidor para permitir una descarga de contenido expresivo basada en el contexto del usuario y de los datos. No es necesario ningún software adicional en las torres celulares y los desarrolladores web no necesitan enfrentarse a un nuevo sistema no familiar y complejo.

La Fig. 1 muestra la arquitectura de alto nivel. Las solicitudes del cliente son enviadas sobre 3G, aunque solicitudes grandes como descargas de archivos podrían ser también descargadas. La respuesta del servidor es dividida en Respuesta' y Respuesta'' para ser enviada sobre 3G como primer vector, y el vector de descarga como segundo vector, respectivamente. Respuesta' contiene cuadros de HTML particularmente ser servidos de la manera común con las etiquetas de contenido reescritas para señalar a futuras ubicaciones de descarga. Respuesta'' contiene el contenido grande con sus meta-datos.

Con respecto al contexto de obtención y utilización para la descarga de datos, existen dos problemas claves: i) qué contexto obtener e ii) cómo distribuir y responder al contexto. Existen muchos mecanismos para detectar, almacenar y agregar contexto conocido en el sector, pero la cuestión clave pregunta qué tipos de información de contexto son particularmente útiles en este escenario. Por ejemplo, cuando una memoria oculta de red se basa en nodos de contacto oportunistas, el contexto puede incluir conocimiento acerca de los patrones de contacto tales como duración de contactos o número de contactos únicos. Como alternativa, cuando las aplicaciones intentan utilizar la descarga para aprovechar el que múltiples usuarios están solicitando el mismo contenido, el contexto puede incluir actividades agregadas de grupos de nodos locales. El permitir que un nodo o cliente obtenga este contexto, se basa en servicios y técnicas existentes [6, 7] que pueden obtener información acerca de un dispositivo, de su entorno local y de sus enlaces de red. El permitir que los clientes obtengan información de contexto agregada, se basa en un reciente trabajo de evaluación de contexto de en grupo dinámico [4] que puede razonar acerca de similitudes entre usuarios, vecindad de la red y gráficos sociales.

Básicamente, es asunto del cliente el determinar qué contexto obtener y compartir con el servidor. Esto depende de las capacidades de obtención de contexto del cliente y de los requisitos de una solicitud. Como en todos los usos de contexto similares, el cliente y el servidor deben acordar a priori el lenguaje utilizado para describir el contexto. El contexto que un usuario obtiene acerca de su situación debe ser enviado al servidor web para permitir una descarga

de datos inteligente. Un contexto de cliente puede ser puesto a continuación de la solicitud del cliente. La arquitectura de acuerdo con la invención aprovecha los planteamientos existentes para resumir sucintamente el contexto [4] para evitar que la transmisión de contexto sobrecargue las conexiones 3G. Información de contexto agregada acerca de un grupo de nodos o clientes puede ser compartida de manera similar. De manera alternativa, el contexto puede ser compartido a través de la memoria oculta de la red y de nuevo hacia el servidor a través del proceso inverso de suministrar la Respuesta'.

En la arquitectura de la realización de la invención, se suministra contenido crítico con el tiempo de la manera tradicional a través de la red celular como primer vector, pero puede enviarse otro contenido selectivamente mediante un mecanismo alternativo, es decir, la descarga de contenido o segundo vector. Esta alternativa puede basarse en principios de red tolerante con el retardo que permiten una comunicación asíncrona, oportunista.

La Fig. 2 muestra una arquitectura de lado del servidor que tiene tres componentes principales: un procesador de solicitud, un gestor de contexto y un procesador de respuesta. Estos tres componentes están dispuestos en una capa media directamente por debajo de la aplicación web. El cliente inserta su contexto en sus solicitudes de HTML. El procesador de solicitud busca etiquetas de contexto específicas, las separa de la solicitud y envía la información de contexto al gestor de contexto, el cual hace un seguimiento de cada usuario mediante su dirección de IP de 3G. En caso contrario, solicita continuar como se conoce comúnmente, sin requerir cambios a la propia aplicación web. Una vez que se ha generado la respuesta de HTML, el procesador de respuesta reescribe la respuesta de acuerdo con reglas predefinidas y el contexto de usuario proporcionado por el gestor de contenido, posiblemente eliminando algo del contenido y enlaces del contenido y enviando esos datos sobre un vector de suministro de descarga como segundo vector.

Como ejemplo de operación, pueden considerarse los siguientes dos elementos de contenido:

1. <a href=\http://server-location.url/media/largeVideoFile.flv" id=embedStreamPlayer>

2. <a href \http://server-location.url/images/largeImageFile.jpg">

Normalmente, ambos elementos de contenido serían servidos por el servidor web sobre la red de 3G. El procesador de respuesta puede reescribir lo anterior en una respuesta para:

1. <a href=\http://localhost/tmp-cache/largeVideoFile.flv" id= embedLocalStreamPlayer>
2. <a href=\http://localhost/tmp-cache/largeImageFile.jpg''>
3. < LOOKUP CONTENT \dtn://*/largeImageFile.jpg", \dtn://*/largeVideoFile.flv" USING DTN BPQ>

Así, las url de archivo de video grande y archivo de imagen grande apuntan ahora a una memoria oculta local, y el reproductor de video embebido ha cambiado a un servicio de transmisión en tiempo real local. El cliente busca los dos elementos de contenido en la memoria oculta de la red DTN local utilizando la extensión de BPQ y, una vez que llega el contenido, el cliente puede transmitir el video en tiempo real desde su propia memoria oculta local.

Con respecto a la arquitectura del lado del cliente, los clientes deben proporcionar contexto al servidor. Para ello, los clientes comprenden dos elementos: un agregador de contexto y un complemento de navegador. El agregador de contexto enumera los recursos y los servicios disponibles, hace un seguimiento de la disponibilidad del vector de suministro de descarga y calcula las posibilidades de descarga futuras. El contexto resultante es enviado entonces a los servidores a través del complemento de navegador que automáticamente detecta los servidores que tienen capacidades de descarga de datos e inserta la información de contexto del agregador de contexto en solicitudes de HTTP. Una agregación de contexto y un procesamiento más centralizados pueden resultar deseables, por ejemplo proporcionados por las portadoras celulares. La arquitectura de la realización de la invención puede ser aún válida, puesto que tal contexto puede ser enviado fácilmente a los clientes, que pueden a continuación enviarlo a los servidores. La privacidad de los datos es una ventaja natural del contexto de cliente controlado, es decir, el usuario mantiene un completo control sobre qué contexto comparte y con quién.

En una posible implementación que subyace también en los ejemplos que se describen a continuación, para el servidor web y la interfaz Apache se utiliza puesto que es un estándar de fuente abierta para el despliegue de servidores web e interconectados con la aplicación web utilizando Interfaz de puerta de enlace de servidor web (WSGI – Web Server Gateway Interface, en inglés) Python. La capa media y el gestor de contexto están implementados en la Arquitectura web Django. En la implementación del gestor de contexto, el cliente codifica tres partes de contexto en cada solicitud de HTTP: ¿Descarga?, DestIPAddr, DRNEndpointID.

Cuando el gestor de contexto en el cliente determina que el suministro avanzado móvil de contenido resulta beneficioso, solicita al servidor que inicie la descarga del contenido y proporciona una dirección de IP de destino a la que debe ser enviado el contenido. Esto solo tiene que ser un destino que ejecute un encaminador de protocolo de paquetes que pueda aceptar y enviar paquetes, y pueda así actuar como una memoria oculta de contenido en red. El cliente proporciona asimismo su identificador de destino de protocolo de paquetes único globalmente. El servidor encapsula los elementos de contenido solicitados en su propio paquete, que está dirigido al DTNEndpointID del cliente y es enviado sobre la Internet a la IPAddr proporcionada. Una vez que el paquete se ha transferido al destino, un encaminamiento oportunista de salto a salto suministrará el contenido.

Para la aplicación web, se utiliza un sitio web de red social completamente funcional construido en la arquitectura de desarrollo rápido de aplicación web Pinax, con una base de datos del lado del servidor (back-end, en inglés) y un sistema de archivos estático para el contenido (imágenes, video, etc.). Para el servicio de descarga de contenido, se utiliza la implementación de referencia DTN2 del Protocolo de paquetes (Bundle protocol, en inglés) para implementar el vector de suministro de descarga de datos. Esto permite dirigir el contenido independiente de la dirección de IP actual de un usuario (en lugar de utilizar su ID de destino globalmente único) y poner el contenido previamente en memoria oculta en lugares que un nodo podría visitar con tal de que allí exista al menos un anfitrión que implemente el protocolo de paquetes y pueda aceptar y enviar paquetes. Si existe una red de tales nodos, entonces el contenido es diseminado a todos los nodos de acuerdo con las reglas de envío de DTN2. Cuando el usuario eventualmente entra en contacto con un nodo que está poniendo en memoria oculta su contenido, el contenido es suministrado.

En lo que sigue, se describe un primer experimento relativo a la descarga solo de 3G frente a datos, por medio de la Fig. 3. Por este medio, la descarga de contenido no solo libera un costoso ancho de banda celular, sino que puede suministrar contenido más rápidamente incluso que sin colocación previa en memoria oculta. En este primer experimento, se emiten 50 solicitudes para tres páginas web diferentes con cuadros de HTML mínimos que contienen imágenes de tamaños 512 kilobytes, 1 MB y 5 MB. El cliente es un ordenador portátil de Linux con un modem de 3G de USB y una tarjeta de WiFi, situado en Europa, y el servidor está situado en el centro de Estados Unidos. La latencia de suministro media se mide utilizando conectividad de 3G y utilizando la descarga de contenido protocolo de paquetes basado en el vector de descarga. El último enlace de salto está por encima de 802,1 1b a través de un punto de acceso por WiFi. La Fig. 3 muestra los resultados de latencia con su desviación estándar para dos puntos de acceso por WiFi diferentes, uno con una potencia de señal relativamente baja y muchos usuarios, y uno con buena potencia de señal y comparativamente menos usuarios. La WiFi no es necesariamente más rápida que 3G [2], aunque esto refuerza la discusión de la descarga de datos basada en contexto del usuario: si la conexión celular es actualmente costosa (por ejemplo, el contexto está en itinerancia), la latencia extra puede merecer la pena frente al dinero ahorrado. A la inversa, si los datos son de alta prioridad, utilizar 3G (dependiendo del ancho de banda disponible de la WiFi) puede ser mejor.

El segundo experimento se refiere al ahorro de ancho de banda de 3G. La Fig. 4 muestra el uso del ancho de banda de 3G durante una operación regular y con descarga de datos. El cliente realiza tres solicitudes para una página que contiene un archivo de video de 16 MB. En el caso solo de 3G, el video es transmitido en tiempo real. En el caso de descarga de datos, el video es agrupado en paquetes y enviado al cliente utilizando nuestro vector de suministro de contenido DTN2. El salto final está por encima de 802.11 g desde un punto de acceso de WiFi. El cliente recibe el paquete, lo inserta en su memoria oculta local y el video es transmitido localmente en tiempo real. Como muestran los resultados, el ahorro del 3G es significativo, es decir, tres órdenes de magnitud. Además, la latencia del establecimiento de la conexión WiFi y el acuse de recibo del registro del cliente de protocolo de paquetes es de solo unos cuantos segundos, y no resulta perjudicial para la experiencia del usuario. El flujo en tiempo real del archivo de video local puede ser iniciado tan pronto como el archivo de video empieza a llegar al cliente. No es necesario esperar a que llegue todo el archivo para iniciar la transmisión en tiempo real del descriptor del archivo. La latencia combinada de la solicitud de inicio de video es así solo de unos cuantos segundos más que cuando se utiliza solo 3G.

El tercer experimento 3 se refiere a los beneficios de un suministro avanzado. La implementación también permite un suministro avanzado de contenido para móviles. Si el gestor del contexto de usuario determina que un nodo está a punto de conectarse a una memoria oculta de contenido, puede solicitar que el contenido actual y futuro solicite ser servido por la memoria oculta en lugar de descargado sobre 3G. La Fig. 5 muestra los resultados de un experimento en el cual un usuario realiza 20 solicitudes independientes de páginas web, incluyendo cada una un elemento de contenido de 5 MB. El diagrama del lado izquierdo representa el uso del 3G a lo largo del tiempo si el contenido es solicitado y suministrado solo a través de 3G. El diagrama del lado derecho muestra un escenario en el que tras las cinco primeras solicitudes el usuario determina que se conectará pronto a una memoria oculta de suministro avanzado para móviles implementada mediante DTN2, momento en el cual todos los contenidos deben ser enviados a la memoria oculta. El servidor responde a dos de las solicitudes sobre 3G, independientemente de porque los datos son de alta prioridad de acuerdo con sus etiquetas de meta-datos. El resto es enviado a la memoria oculta de contenido y servido al cliente cuando se conecta por encima de 802.1 1b. Cuando el cliente se desconecta de la memoria oculta de contenido, envía una actualización de contexto, y las restantes cinco solicitudes son servidas sobre 3G. En este experimento, el contexto del usuario fue predeterminado y proporcionado antes del experimento. La arquitectura descrita no solo ahorra ancho de banda de 3G, reduciendo en este caso la carga de 3G

de 108 MB a 60,7 MB para solo una conexión dada a una memoria oculta de contenido de 51 segundos, sino que es capaz de suministrar los veinte elementos de contenido en casi la mitad de tiempo.

Aunque la invención ha sido ilustrada y descrita con detalle en los dibujos y en la descripción anterior, tal ilustración y descripción deben ser consideradas ilustrativas o de ejemplo y no restrictivas. Resultará evidente que personas no expertas en la materia pueden realizar cambios y modificaciones dentro del alcance de las siguientes reivindicaciones. En particular, la presente invención cubre otras realizaciones con cualquier combinación de características de diferentes realizaciones descritas anteriormente y en lo que sigue.

La invención cubre también todas las demás características mostradas en las figuras individualmente, aunque puedan no haberse descrito en la descripción anterior o siguiente. Asimismo, alternativas únicas de las realizaciones descritas en las figuras y en la descripción y alternativas únicas de características de la misma pueden ser excluidas del objeto de la invención.

Además, en las reivindicaciones, las expresiones “que comprende” o “que comprenden” no excluyen otros elementos o etapas, y los artículos indefinidos “un” o “una” no excluyen una pluralidad. Una sola unidad o etapa puede cumplir las funciones de varias características citadas en las reivindicaciones. Los términos “esencialmente”, “aproximadamente” y otros en conexión con un atributo o un valor particularmente definen también exactamente el atributo o exactamente el valor, respectivamente. El término “aproximadamente” en el contexto de un valor o intervalo numérico dado se refiere a un valor o intervalo que se encuentra, por ejemplo, dentro del 20%, dentro del 10%, dentro del 5% o dentro del 2% de un valor o intervalo dado. Un programa informático puede ser almacenado / distribuido en un medio adecuado, tal como un medio de almacenamiento óptico o un medio de estado sólido suministrado junto con o como parte de otro hardware, pero puede ser también distribuido en otras formas, tales como a través de la Internet o de otros sistemas de telecomunicación por cable o inalámbrica. En particular, por ejemplo, un programa informático puede ser un producto de programa informático almacenado en un medio legible por ordenador, cuyo producto de programa informático puede tener código de programa ejecutable por ordenador adaptado para ser ejecutado para implementar un método específico tal como el método de acuerdo con la invención. Cualquier signo de referencia en las reivindicaciones no debe ser considerado como limitativo del alcance.

Publicaciones de referencia

[1] B.Ahlgren y V. Vercellone. Networking of information an information-centric approach to the network of the future. En el Taller de tecnologías de red futuras de ETSI, Marzo de 2010.

[2] A. Balasubramanian, R. Mahajan y A. Venkataramani. Augmenting mobile 3G using WiFi. En Procedimiento de la 8ª Conferencia internacional en sistemas de telefonía móvil, aplicaciones y servicios, 2010.

[3] D. Giustiniano, E. Goma, A. López Toledo, I. Dangerfield, J. Morillo y P. Rodríguez. Fair WLAN backhaul aggregation. En Proc. de MobiCom, 2010.

[4] C. Julien. The context of coordinating groups in dynamic mobile networks. En Proc. de COORDINACIÓN, 2011.

[5] A. Nicholson y B. Noble. BreadCrumbs: Forecasting mobile connectivity. En Proc. de MobiCom, 2008.

[6] A. Petz, T. Jun, N. Roy, C.-L. Fok y C. Julien. Passive network-awareness for dynamic resource-constrained networks. In Proc. de DAIS, 2011.

[7] A. Rahmati y L. Zhong. Context-for-wireless: context-sensitive energy-efficient wireless data transfer. En Procedimientos de la 5ª conferencia internacional de sistemas de telefonía móvil, aplicaciones y servicios, 2007.

REIVINDICACIONES

1. Método de suministrar contenido a través de una red de un servidor a un cliente que tiene una primera interfaz de red, que comprende

proporcionar el cliente una solicitud de contenido al servidor a través de un primer vector, en el que el primer vector se refiere a una ruta dentro de una tecnología de red específica y está asociado con la primera interfaz de red;

recibir el servidor información de contexto acerca del contexto del cliente, en el que la información de contexto comprende información acerca del dispositivo del cliente, de su entorno local, que puede cambiar de manera continua mientras el cliente se está desplazando y, de sus enlaces de red o interfaces de red y su disponibilidad;

separar el servidor el contenido para ser suministrado al cliente de acuerdo con la solicitud en contenido pequeño y contenido grande;

definir el servidor un procedimiento de transferencia para el contenido grande basándose en la información de contexto;

proporcionar el servidor una primera respuesta al cliente a través del primer vector, en el que la primera respuesta comprende el contenido pequeño e información acerca del procedimiento de transferencia del contenido grande; y

recibir el cliente el contenido grande de acuerdo con la información acerca del procedimiento de transferencia del contenido grande.

2. El método de la reivindicación 1, en el que la solicitud proporcionada por el cliente al servidor comprende la información de contexto acerca del contexto del cliente.

3. El método de la reivindicación 1 o 2, en el que el cliente comprende una segunda interfaz de red, la información acerca del procedimiento de transferencia del contenido grande comprende información acerca de un segundo vector y el cliente recibe el contenido grande a través del segundo vector, en el que el segundo vector está asociado con la segunda interfaz de red.

4. El método de la reivindicación 3, en el que el servidor proporciona el contenido grande a un nodo de red que es alcanzable por el cliente a través del segundo vector.

5. El método de la reivindicación 3 o 4, en el que el servidor proporciona una segunda respuesta al cliente a través del segundo vector que comprende el contenido grande.

6. Servidor que comprende una unidad de procesamiento de solicitud, una unidad de gestión de contexto y una unidad de procesamiento de respuesta, en el que

la unidad de procesamiento de solicitud está dispuesta para recibir una solicitud de contenido desde un cliente a través de un primer vector, en el que el primer vector se refiere a una ruta en una tecnología específica, para recibir información de contexto acerca del contexto del cliente y para proporcionar la información de contexto a la unidad de gestión de contexto, en el que la información de contexto comprende información acerca del dispositivo del cliente, de su entorno local que puede cambiar de manera continua mientras el cliente se está desplazando y, de sus enlaces de red o interfaces de red y su disponibilidad;

la unidad de gestión de contexto está dispuesta para definir un procedimiento de transferencia para el contenido grande basándose en la información de contexto y proporcionar información acerca del procedimiento de transferencia del contenido grande a la unidad de procesamiento de respuesta; y

la unidad de procesamiento de respuesta comprende reglas de separación predefinidas y está dispuesta para separar el contenido para ser suministrado al cliente en contenido pequeño y contenido grande aplicando las reglas de separación predefinidas y para proporcionar una primera respuesta al cliente a través del primer vector que comprende el contenido pequeño e información acerca del procedimiento de transferencia del contenido grande.

7. El servidor de la reivindicación 6, en el que la unidad de procesamiento de solicitud está dispuesta para separar la información de contexto de la solicitud del cliente.

8. El servidor de la reivindicación 6 o 7, en el que la unidad de gestión de contexto está dispuesta para hacer un seguimiento del cliente por medio de un identificador de cliente.

9. El programa informático que comprende medios de código de programa que están dispuestos para implementar las características del servidor de acuerdo con una cualquiera de las reivindicaciones 6 a 8 cuando es ejecutado.

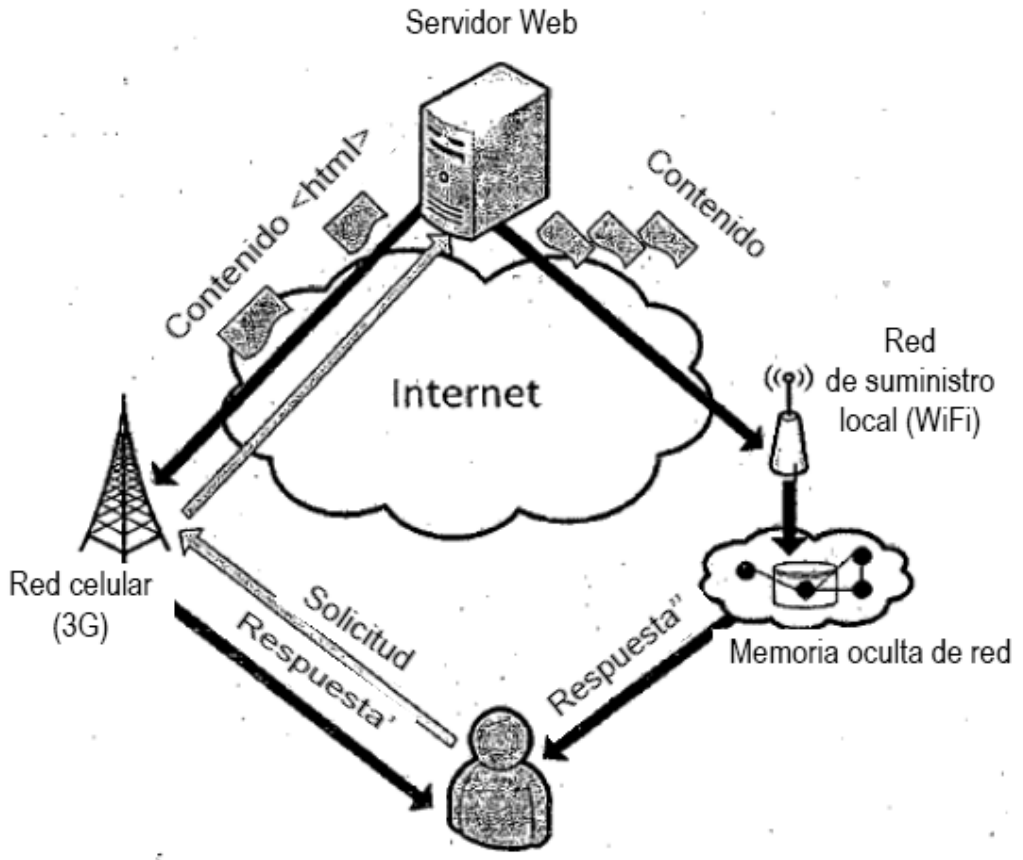
Servidor Web
Contenido <html>
Contenido
Internet
Red
de suministro
local (WiFi)
Red celular
(3G)
Solicitud
Respuesta'
Respuesta"
Memoria oculta de red

Fig. 1

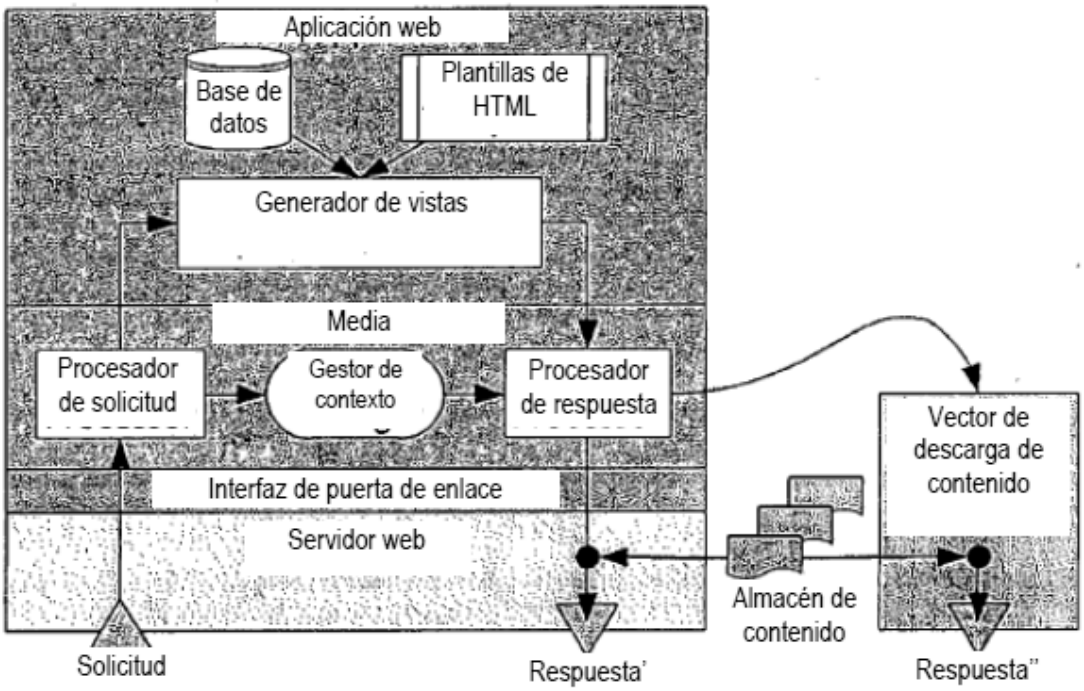
Aplicación web
Base de datos
Plantillas de HTML
Generador de vistas
Media
Procesador de solicitud
Gestor de contexto
Procesador de respuesta
Interfaz de puerta de enlace
Servidor web
Vector de descarga de contenido
Almacén de contenido
Solicitud
Respuesta'
Respuesta"

Fig. 2

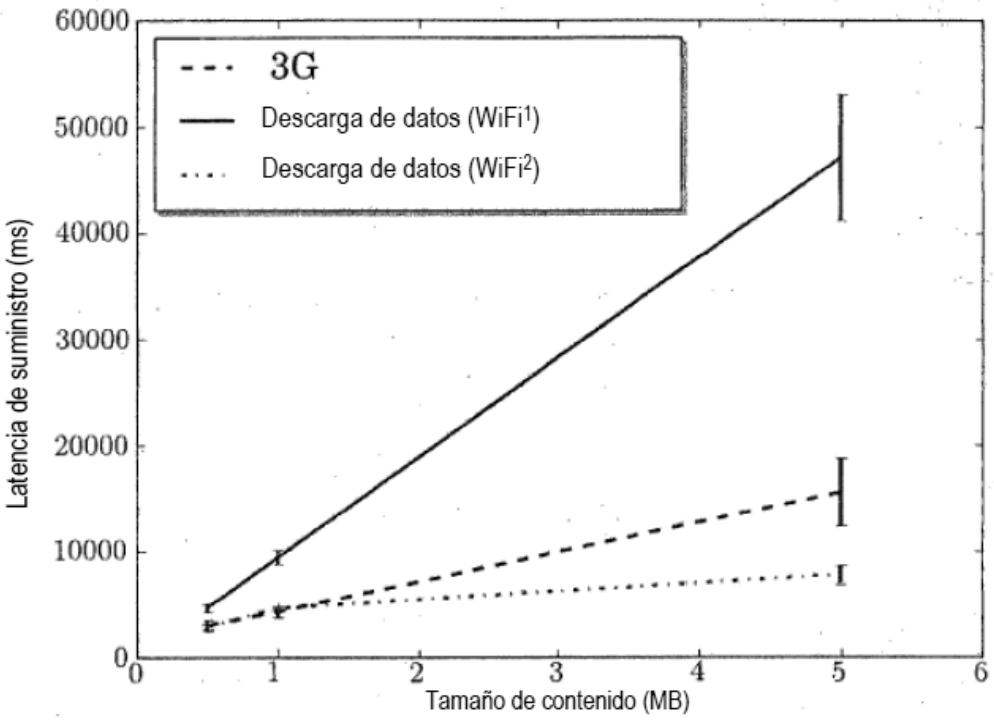
3G
Descarga de datos (WiFi1)
Descarga de datos (WiFi2)
Latencia de suministro (ms)
Tamaño de contenido (MB)
0
10000
20000
30000
40000
50000
60000
0
1
2
3
4
5
6

Fig. 3

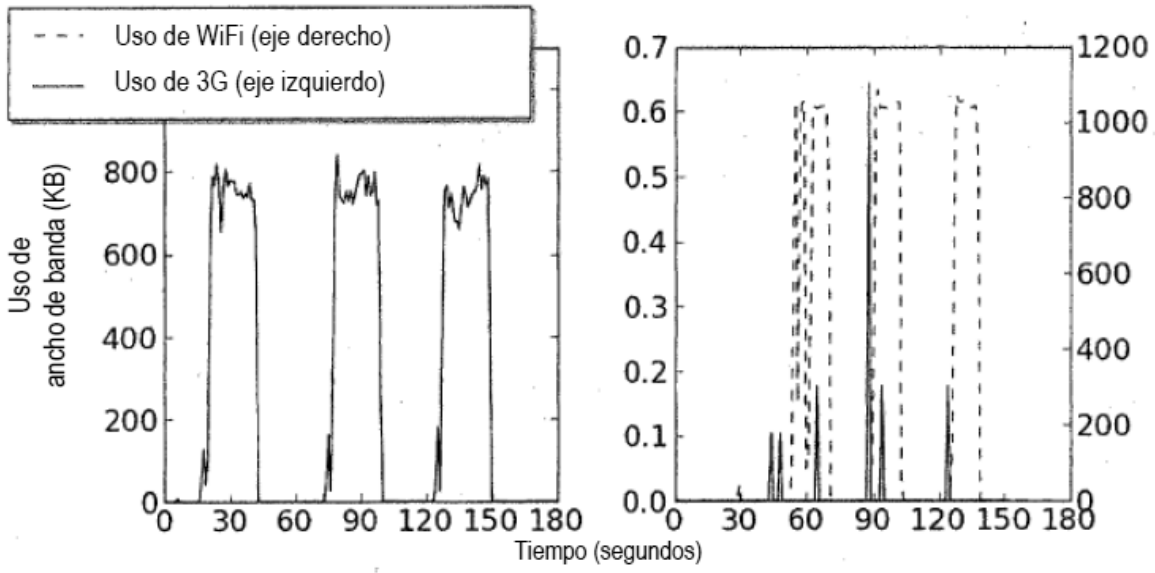
Uso de WiFi (eje derecho)
Uso de 3G (eje izquierdo)
Uso de ancho de banda (KB)
Tiempo (segundos)
0
200
400
600
800
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
200
400
600
800
1000
1200
0
30
60
90
120
150
180

Fig. 4

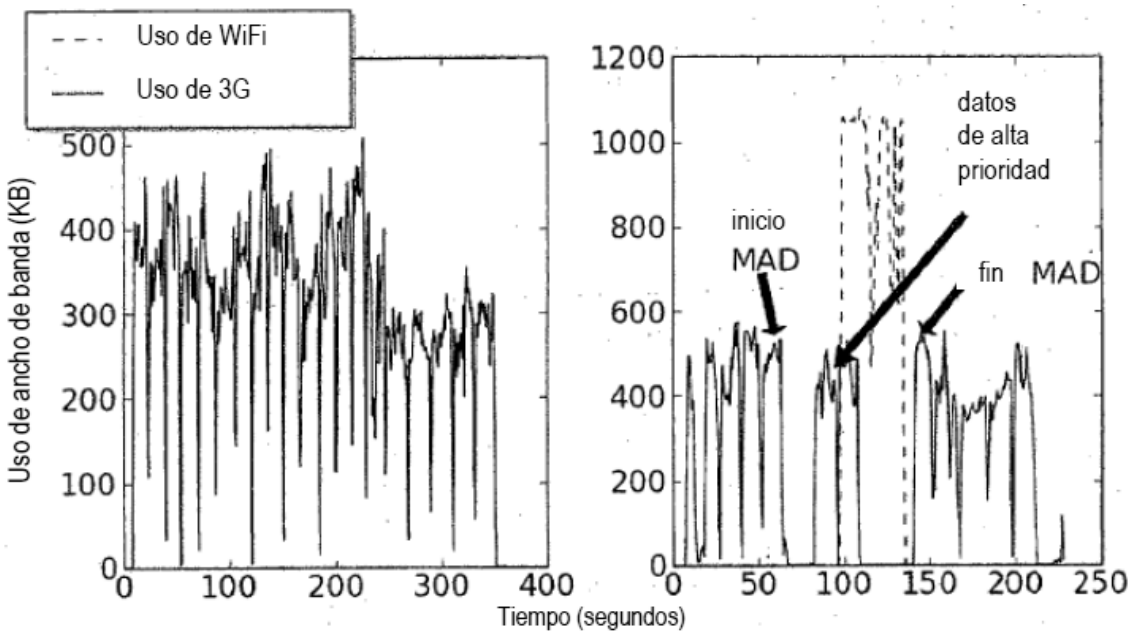
Uso de WiFi
Uso de 3G
Uso de ancho de banda (KB)
Tiempo (segundos)
inicio
MAD
fin
MAD
datos
de alta
prioridad

Fig. 5