

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 524 881**

51 Int. Cl.:

H04L 29/08 (2006.01)

H04L 29/06 (2006.01)

H04L 29/14 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.10.2009 E 09174140 (5)**

97 Fecha y número de publicación de la concesión europea: **01.10.2014 EP 2180659**

54 Título: **Detección de una sobrecarga de un servidor SIP y su control**

30 Prioridad:

27.10.2008 US 259011

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

15.12.2014

73 Titular/es:

**BROADSOFT, INC. (100.0%)
9737 Washingtonian Boulevard, Suite 350
GAITHERSBURG, MD 20878, US**

72 Inventor/es:

**GRIFFITH, LAURA y
KUSHNIR, MARK**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 524 881 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

Detección de una sobrecarga de un servidor SIP y su control

5 Campo de la invención

La presente invención se refiere al campo de redes de comunicaciones.

Antecedentes de la invención

10 Los servicios de comunicaciones actuales suelen basarse en la red telefónica conmutada pública (PSTN), que proporciona servidores vocales de circuitos conmutados, transmisión por fax en tiempo real y correo de voz basado en la oficina central (CO) o sistemas de Internet, intranet o extranet basados en el protocolo Internet (IP), que prestan servicios tales como correo electrónico.

15 Los sistemas de red PSTN y basados en el protocolo IP se basan en diferentes arquitecturas y protocolos. La red PSTN suele referirse a un conjunto de sistemas interconectados utilizados por las diferentes compañías telefónicas y administraciones de correos, teléfonos y telégrafos (PTTs) a escala mundial. Los sistemas de red PSTN utilizan circuitos especiales. Es decir, cuando se marca un número de teléfono, se activa la búsqueda en la red de un circuito dedicado que solamente pueda utilizarse por esa llamada telefónica y existen hasta que se termine dicha llamada telefónica. De este modo, aún cuando no exista ninguna información a transmitir como parte de la transmisión (p.e., los intervinientes en las llamadas están en silencio), el circuito está todavía en uso por la llamada.

25 En lugar de enviar transmisiones a través de un circuito dedicado, los sistemas basados en IP empaquetan las transmisiones con direcciones autónomas que luego se cargan a través de una pasarela en la red en donde se retransmiten entre enrutadores hasta que alcanzan su destino final. Esta formación de paquetes de transmisiones permite un transporte mucho más eficiente de las transmisiones puesto que el espacio vacío en una transmisión (en donde ninguna información está presente) puede rellenarse por transmisiones adicionales.

30 Las empresas de telecomunicaciones tradicionales y los proveedores de servicios se desplazan desde una red PSTN a una infraestructura de IP basada en paquetes. Los contenidos ofrecidos a través de IP incluyen, a modo de ejemplo, transmisión de voz, fax, mensajería instantánea, televisión digital, etc. En el caso de transmisión vocal, la tecnología es referida como voz sobre IP (VOIP).

35 Un aspecto para una red de telefónica puede ser el control de llamadas, incluyendo el establecimiento de llamadas y desmontaje operativo. Un sistema de VOIP opera en la ausencia de circuitos dedicados, pudiendo el control de llamadas utilizarse para evitar paradas técnicas del servicio, degradación de la calidad del servicio (QoS) y pérdida de ingresos. El control de llamadas puede conseguirse mediante un protocolo de iniciación de sesión (SIP).

40 SIP es un protocolo de señalización que utiliza mensajes de texto. Estas son dos entidades SIP básicas: agentes de usuarios SIP (UAs) y servidores SIP. Los servidores SIP pueden agruparse, además, en servidores proxy para encaminamiento de sesiones y servidores de registro para registro de UA.

45 SIP es un protocolo de nivel de aplicación que puede ejecutarse en la parte superior de una capa de transporte, tal como el protocolo de control de transmisión (TCP) o el protocolo de diagrama de usuarios (UDP).

50 SIP se utiliza ampliamente para establecer y mantener sesiones de comunicaciones multimedia tales como llamadas de voz y de vídeo a través de Internet. Otros ejemplos de aplicación de SIP incluyen videoconferencias, flujos de distribución multimedia, mensajería instantánea, información de presencia y juegos en línea. A modo de ejemplo, la utilización de SIP para mensajería instantánea y presencia implica una carga adicional sobre la infraestructura de control de llamadas basadas en SIP.

55 En noviembre del año 2000, SIP fue aceptado como un protocolo de señalización del denominado Proyecto de Asociación de la 3ª Generación (3GPP) y elemento permanente de la arquitectura del subsistema multimedia IP (IMS) para servicios multimedia de flujos basados en IP en sistemas celulares. Este desarrollo significa que la infraestructura de control de la infraestructura de control de llamadas basadas en SIP necesita gestionar la movilidad de terminal entre diferentes límites celulares, además del establecimiento y mantenimiento de llamadas. Tanto el soporte para usuarios móviles como la transferencia representan cargas adicionales sobre la infraestructura de control de llamadas.

60 La solicitud de patente internacional con número de publicación WO 98/57504 da a conocer un método de prevención de sobrecarga en un punto de control de servicio (SCP). La solicitud de patente de Estados Unidos, con número de publicación 2007/0233896 da a conocer un método para impedir la sobrecarga en una red basada en SIP que comprende una pluralidad de servidores. La solicitud de patente de Estados Unidos, con número de publicación 2005/0105464 A1 da a conocer un método para la gestión diferenciada de mensajes de control de llamadas de VoIP según su importancia funcionalidad.

65

Sumario de la invención

Según la presente invención, se da a conocer un aparato y método según se establece en las reivindicaciones adjuntas. Otras características de la invención serán evidentes a partir de las reivindicaciones subordinadas y la descripción que sigue.

La invención puede incluir un método para detectar y mitigar la sobrecarga del servidor SIP. El método puede proseguir recibiendo mensajes SIP en el servidor SIP procedentes de una red de comunicación, procesando los mensajes SIP recibidos, detectando una sobrecarga basada en una condición de iniciación operativa como resultado de la recepción y procesamiento de mensajes SIP y la transición del servidor SIP a un estado de sobrecarga y la aplicación de acciones de control adecuadas para mitigar la sobrecarga detectada.

La invención puede incluir un soporte legible por ordenador o legible por máquina para la puesta en práctica del método para la detección y mitigación de la sobrecarga del servidor SIP. Un "soporte legible por ordenador" o "soporte legible por máquina" puede referirse a cualquier dispositivo de memorización utilizado para memorizar datos accesibles por un ordenador. Un soporte legible por ordenador, a modo de ejemplo, puede incluir: una unidad de disco duro; un disco flexible; un disco magneto-óptico; un disco óptico; una cinta magnética; una memoria extraíble instantánea y un circuito integrado de memoria.

La invención puede dar a conocer un sistema para la comunicación multimedia a través de una red. El sistema puede incluir una conexión de red para recibir y transmitir mensajes SIP entre un servidor SIP la red, un módulo de procesamiento en el servidor SIP para procesar y servir mensajes SIP recibidos y un módulo de detección de sobrecarga para realizar el control de sobrecarga para el servidor SIP.

Breve descripción de los dibujos

Las formas de realización de la invención se describirán a continuación en relación con los dibujos asociados.

La Figura 1 ilustra las colas de espera internas y las iniciativas de procesos asociadas con un servidor SIP en una forma de realización, a modo de ejemplo, de la invención.

La Figura 2 ilustra los componentes de un módulo de control de sobrecarga en una forma de realización, a modo de ejemplo, de la invención.

La Figura 3 ilustra, además, la composición del servidor SIP y el flujo de trabajo en una forma de realización, a modo de ejemplo, de la invención.

La Figura 4 ilustra los estados de sobrecarga y diagrama de transición de estados para un servidor SIP en una forma de realización, a modo de ejemplo, de la invención.

La Figura 5A ilustra un ordenador de servidor, a modo de ejemplo, en relación con una red de comunicación en una forma de realización, a modo de ejemplo, de la invención.

La Figura 5B ilustra un sistema informático del servidor, a modo de ejemplo, que comprende ordenadores de servidores múltiples en una forma de realización a modo de ejemplo, de la invención.

La Figura 5C ilustra un sistema informático de servidor, a modo de ejemplo, en una forma de realización, a modo de ejemplo, de la invención.

Descripción detallada de las formas de realización

En una forma de realización, a modo de ejemplo, de la invención, un módulo de control de sobrecarga se utiliza para detectar la sobrecarga del servidor SIP y para aplicar acciones de control al respecto. Para la detección de sobrecarga, el módulo compara al menos una métrica de recursos con respecto a un umbral. Para el control de la sobrecarga, el módulo efectúa la transición del servidor SIP a un estado diferente después de que se detecte una sobrecarga y luego, aplica un conjunto de acciones de control asociadas con ese estado.

La Figura 1 ilustra las colas de espera internas y las iniciativas de procesos asociadas con el servidor SIP 150 en una forma de realización, a modo de ejemplo, de la invención. El servidor SIP 150 puede incluir, a modo de ejemplo, una cola de espera de decodificador SIP rápido 101, un decodificador SIP rápido 102, una cola de espera de decodificador SIP CallP 103, una cola de espera de decodificador SIP no CallP 104, una pluralidad de decodificadores CallP 105, una pluralidad de decodificadores SIP no CallP 106, una pluralidad de colas de espera de adaptador CallP 107, una pluralidad de colas de espera de adaptador no CallP 108, una pluralidad de iniciativas operativas de adaptador CallP 109, una pluralidad de iniciativas operativas de adaptador no-CallP 110, una cola de espera de codificador SIP 111, un codificador SIP 112 y un módulo de control de sobrecarga 114.

El mensaje SIP entrante 100 puede tener una marca temporal con el tiempo más actual y colocarse en la cola de espera de decodificador SIP rápido 101. El decodificador SIP rápido 102 puede realizar un análisis sintáctico rápido del mensaje SIP entrante 100 para determinar si el mensaje es una demanda o respuesta de procesamiento de llamada (CallP) (p.e., INVITE) o una demanda relacionada con el procesamiento de llamadas (no CallP) o su respuesta (p.e., REGISTER); los mensajes relacionados con CallP pueden enrutarse a la cola de espera del decodificador SIP CallP 103 mientras que los mensajes relacionados con no CallP pueden enrutarse a la cola de espera del decodificador SIP no CallP 104. Los mensajes pueden ser objeto de marca temporal con el tiempo más actual de nuevo cuando se colocan en las colas de espera respectivas del decodificador 103 y 104 para indicar el tiempo de entrada preciso.

El decodificador SIP CallP 105 puede efectuar un análisis sintáctico de un mensaje en la cola de espera del decodificador SIP CallP 103 para determinar si el mensaje es una demanda o respuesta asociada con una sesión existente o si el mensaje es una nueva demanda. Dependiendo del resultado, el contenido del mensaje objeto de análisis sintáctico puede colocarse en la cola de espera de adaptador CallP 107 en donde los mensajes pueden ser de nuevo objeto de marca temporal con tiempo más actual para indicar el tiempo de entrada preciso en la cola de espera de adaptador CallP 107. Durante este proceso, los mensajes pertenecientes a sesiones existentes pueden asignarse a la cola de espera de adaptador CallP 107 asociada con la iniciativa operativa de adaptador CallP 109 que gestiona esa sesión ya existente.

La iniciativa operativa de adaptador tipo CallP 109 puede extraer mensajes de la cola de espera de adaptador CallP 107, procesar los mensajes y aplicar servicios cuando sea adecuado. Cuando un mensaje SIP saliente se requiere en referencia a una sesión, la iniciativa operativa de adaptador 109 puede transmitir la información requerida en un mensaje a la cola de espera del codificador SIP 111 en donde el mensaje puede ser nuevo objeto de marca temporal con el tiempo más actual para indicar el tiempo de entrada preciso en la cola de espera del codificador SIP 111.

Asimismo, el decodificador SIP no CallP 106 puede realizar el análisis sintáctico de un mensaje en la cola de espera del decodificador SIP no CallP 104 para determinar si el mensaje es una demanda o respuesta asociada con una sesión existente o si el mensaje es una nueva demanda. Dependiendo del resultado, el contenido del mensaje objeto de análisis sintáctico puede colocarse en una cola de espera de adaptador no CallP 108 en donde los mensajes pueden de nuevo ser objeto de marca temporal con el tiempo más actual para indicar el tiempo de entrada preciso en la cola de espera de adaptador no CallP 108. Durante este proceso, los mensajes pertenecientes a sesiones existentes pueden asignarse a la cola de espera de adaptador no CallP 108 asociada a la iniciativa operativa de adaptador no CallP 110 que gestiona esa sesión ya existente.

La iniciativa operativa de adaptador no CallP 110 puede extraer mensajes desde la cola de espera de adaptador no CallP 108, procesarlos y aplicar servicios cuando sea adecuado. Cuando se requiere un mensaje SIP saliente en referencia a una sesión, la iniciativa operativa de adaptador no CallP 110 puede transmitir la información requerida en un mensaje a la cola de espera de codificador SIP 111, en donde el mensaje puede de nuevo ser objeto de marca temporal con el tiempo más actual para indicar la hora de entrada precisa en la cola de espera del codificador SIP 111.

El codificador SIP 112 puede extraer mensajes desde la cola de espera del codificador SIP 111, construir un mensaje SIP saliente y luego, enviar el mensaje SIP saliente 113.

El módulo de control de sobrecarga 114 puede utilizarse para detectar la sobrecarga del servidor SIP 150 y para aplicar acciones de control para el servidor SIP 150 cuando se detecta una sobrecarga.

La Figura 2 ilustra los componentes del módulo de control de sobrecarga 114 en una forma de realización, a modo de ejemplo, de la invención. El servidor SIP puede considerarse como que incluye, a modo de ejemplo, un módulo de procesamiento 200 y un módulo de control de sobrecarga 114. El módulo de control de sobrecarga 114 puede incluir, a modo de ejemplo, un monitor 205, un comparador 206 y una interfaz de software 203. El módulo de procesamiento 200 puede incluir, a modo de ejemplo, la cola de espera del decodificador SIP rápido 101, un decodificador SIP rápido, una cola de espera de decodificador SIP CallP 103, una cola de espera de decodificador SIP no CallP 104, una pluralidad de decodificadores CallP 105, una pluralidad de decodificadores SIP no CallP 106, una pluralidad de colas de espera de adaptador CallP 107, una pluralidad de colas de espera de adaptador no CallP 108, una pluralidad de iniciativa operativa de adaptador CallP 109, una pluralidad de iniciativas operativas de adaptador no CallP 110, una cola de espera de codificador SIP 111 y un codificador SIP 112. Para fines de ilustración solamente, el módulo de procesamiento 200 se ilustra en la Figura 2 con solamente el decodificador CallP 105 y el decodificador SIP no CallP 106.

Mientras el mensaje SIP entrante recibido 100 se está procesando por el módulo de procesamiento 200 del servidor SIP 150, el módulo de control de sobrecarga 114 puede efectuar un seguimiento continuo de la métrica de recursos como resultado de la ejecución del módulo de procesamiento 200. Este seguimiento puede realizarse por el monitor 205 que puede leer la métrica de recursos desde el módulo de procesamiento 200 a través de la ruta 208. El comparador 206 puede comparar una métrica de recursos objeto de seguimiento con una instancia operativa de

umbrales 209. La interfaz de software 207 puede recibir el resultado de la comparación y enviar señales al módulo de procesamiento 200 por intermedio de rutas 115 y 116. La métrica de recursos, objeto de seguimiento, puede incluir, a modo de ejemplo, un retardo de cola de espera de CallP 201, un retardo de cola de espera no CallP 202, la utilización de CPU 203, la utilización de la memoria de procesos 204 y otras informaciones.

5 A modo de ejemplo, el retardo de la cola de espera CallP 201 puede referirse al tiempo en que un mensaje permanece en la cola de espera 107. El retardo de cola de espera CallP 201 puede supervisarse comparando el tiempo actual cuando el mensaje se extrae de la cola de espera con el tiempo de entrada en la cola de espera del mensaje. A modo de ejemplo, si el retardo de la cola de espera CallP actual 201 sobrepasa un umbral de retardo de
10 cola de espera, el módulo de control de sobrecarga 114 puede efectuar la transición del servidor SIP 150 a un estado de sobrecarga CallP adecuado en donde puedan aplicarse acciones de control para mitigar la sobrecarga.

15 A modo de ejemplo, el retardo de cola de espera no CallP 202 puede referirse al tiempo en que un mensaje permanece en la cola de espera 108. El retardo de cola de espera no CallP 202 puede supervisarse comparando el tiempo actual en que el mensaje se extrae de la cola de espera con el tiempo de entrada de la cola de espera del mensaje. A modo de ejemplo, si el retardo de cola de espera no CallP actual 202 sobrepasa un umbral de retardo de
20 cola de espera, el módulo de control de sobrecarga 114 puede efectuar la transición del servidor SIP 150 a un estado de sobrecarga no CallP adecuado en donde se pueden aplicar acciones de control para mitigar la sobrecarga.

25 Un retardo de cola de espera medio renovable puede calcularse para el retardo de cola de espera CallP 201 y para el retardo de cola de espera no CallP 202. El retardo de cola de espera medio renovable puede calcularse a partir de una ventana móvil que cubre varios retardos de cola de espera. El retardo de cola de espera medio renovable puede utilizarse para suavizar un retardo de cola de espera inusual. El módulo de control de sobrecarga 114 puede
30 comparar un retardo de cola de espera medio renovable con respecto a un umbral de retardo de cola de espera.

35 En la realización, a modo de ejemplo, del retardo de cola de espera medio renovable CallP, si el umbral para entrar en un estado de sobrecarga es 500 mseg, y si el retardo de cola de espera medio renovable de CallP actual para la cola de espera de adaptador CallP 107 es de 550 milisegundos, el servidor SIP 150 puede efectuar la transición a un estado de sobrecarga CallP. El módulo de control de sobrecarga 114, por intermedio de la interfaz de software 207 y la ruta de transmisión 116, pueden formar al decodificador SIP CallP 105 para, a modo de ejemplo, bloquear el 50 % de las llamadas nuevas. Cuando el decodificador SIP CallP 105 recibe la señal, el decodificador SIP CallP 105 podrá, a modo de ejemplo, rechazar el 50 % de las demandas de llamadas nuevas entrantes generando una
40 respuesta de "503 servicio no disponible" y enviando directamente la respuesta, a través de la ruta de transmisión 118, a la cola de espera del codificador SIP 111, con lo que se deriva la iniciativa operativa del adaptador de CallP 109.

45 En la realización, a modo de ejemplo, del retardo de cola de espera medio renovable no CallP, si el umbral para entrar en un estado de sobrecarga es 500 milisegundos y el retardo de cola de espera medio renovable no CallP para la cola de espera de adaptador no CallP 108 es 550 milisegundos, el servidor SIP 150 puede efectuar la transición a un estado de sobrecarga no CallP. El módulo de control de sobrecarga 114 puede informar al decodificador SIP no CallP 106, mediante una interfaz de software 207 a través de la ruta de transmisión 115 para, a modo de ejemplo, bloquear el 50 % de las nuevas demandas SIP no CallP. Cuando el decodificador SIP no CallP 106 recibe la señal, el decodificador SIP no CallP 106 podrá, a modo de ejemplo, rechazar el 50 % de las nuevas
50 demandas entrantes generando una respuesta de "503 servicio no disponible" y enviando directamente la respuesta, a través de la ruta de transmisión 117 a la cola de espera del codificador SIP 111, con lo que se deriva la iniciativa operativa de adaptador no CallP 110.

55 A modo de ejemplo, la utilización de la unidad CPU 203 puede referirse al porcentaje del tiempo de la CPU ocupado por la ejecución de los componentes del programa de aplicación del servidor 150 en un ordenador o sistema informático. A modo de ejemplo, si la utilización de la CPU actual 203 sobrepasa un umbral de utilización de la CPU, el módulo de control de sobrecarga 114 puede efectuar la transición del servidor SIP 150 al estado de sobrecarga CallP o no CallP adecuado, en donde pueden aplicarse acciones de control para mitigar la sobrecarga. En el caso de sobrecarga de utilización de la unidad CPU en CallP, el módulo de control de la sobrecarga 114 puede informar al decodificador SIP CallP 105 y al decodificador SIP no CallP 106, mediante la interfaz de software 207 por intermedio de la ruta de transmisión 115 y 116 respectivamente para, a modo de ejemplo, bloquear el 100 % de las nuevas demandas. Cuando el decodificador SIP CallP 105 y el decodificador SIP no CallP 106 reciben la señal, ambos decodificadores pueden, a modo de ejemplo, rechazar el 100 % de las nuevas demandas entrantes generando una
60 respuesta de "503 servicio no disponible" y enviando directamente la respuesta por intermedio de la ruta de transmisión 118, a la cola de espera del codificador SIP 111, con lo que se elude operativamente la iniciativa del adaptador CallP 109 y la iniciativa operativa del adaptador no CallP 110.

65 A modo de ejemplo, la utilización de memoria de procesos 204 puede referirse a la cantidad de memoria de hardware de ordenador consumida ejecutando los componentes del programa de aplicación del servidor 150 en un ordenador o sistema informático. A modo de ejemplo, si la utilización de memoria de procesos actual 204 sobrepasa un umbral de utilización de memoria, el módulo de control de sobrecarga 114 puede efectuar la transición del

servidor SIP 150 a un estado de sobrecarga adecuado en donde pueden aplicarse acciones de control para mitigar la sobrecarga. En el caso de sobrecarga de utilización de memoria de procesos, el módulo de control de sobrecarga 114 puede informar al decodificador SIP CallP 105 y al decodificador SIP no CallP 106, por intermedio de la interfaz de software 207 a través de las rutas de transmisión 115 y 116 respectivamente para, a modo de ejemplo, bloquear el 100 % de las nuevas demandas. Cuando el decodificador SIP CallP 105 y el decodificador SIP no CallP 106 reciben la señal, ambos decodificadores podrán, a modo de ejemplo, rechazar el 100 % de las nuevas demandas entrantes generando una respuesta "503 servicio no disponible" y enviando directamente la respuesta, a través de las rutas de transmisión 118, a la cola de espera del codificador SIP 111, con lo que se elude la iniciativa operativa del adaptador CallP 109 y la iniciativa operativa del adaptador no CallP 110.

Según se ilustra en las Figuras 1 y 2, el módulo de control de sobrecarga 114 puede ser parte del servidor SIP 150. En otra forma de realización, a modo de ejemplo, el módulo de control de sobrecarga 114 puede ser un módulo autónomo fuera del servidor SIP 150 y puede realizar un control de sobrecarga para múltiples servidores SIP.

El umbral 209 se ilustra en la Figura 2 como siendo exterior al servidor SIP 150. A modo de ejemplo, el umbral puede residir en un fichero de configuración que puede leerse por el servidor SIP 150 durante la iniciación. Sin embargo, el umbral 209 puede residir también en el interior del servidor SIP 150. A modo de ejemplo, el umbral 209 puede definirse por el código de software que pone en práctica el servidor SIP 150. Además, el umbral 209 puede proporcionar más umbrales utilizados por el comparador 206 para las diversas métricas de recursos del monitor 205.

La Figura 3 ilustra, además, la composición del servidor SIP 150 y el flujo de trabajo en una forma de realización, a modo de ejemplo, de la invención en el bloque 301, el módulo de procesamiento 200 puede recibir mensajes SIP entrantes 100. En el bloque 302, el módulo de procesamiento 200 puede procesar los mensajes recibidos para generar mensajes salientes 113. En el bloque 303, el módulo de control de sobrecarga 114 puede supervisar las métricas de recursos resultantes de la ejecución del programa del módulo de procesamiento 200 en un ordenador o sistema informático. En el bloque 304, el módulo de control de sobrecarga 114 puede comparar una métrica de recursos supervisados con respecto a un umbral entre los umbrales 209 para determinar si existe una sobrecarga. En el bloque 305, el módulo de control de sobrecarga 114 puede efectuar la transición del servidor SIP 150 a un estado de sobrecarga si se detecta una sobrecarga.

La Figura 4 ilustra los estados de sobrecarga y los diagramas de transición de estados para un servidor SIP en una forma de realización, a modo de ejemplo, de la invención. Los estados, a modo de ejemplo, del servidor SIP 150 pueden incluir un estado verde 400, un estado amarillo no CallP 401, un estado rojo no CallP 402, un estado amarillo CallP 403 y un estado rojo CallP 404. Más o menos estados pueden incluirse para el servidor SIP 150 según pueda desearse. A modo de ejemplo, menos estados pueden utilizarse para concentrar acciones de control y más estados pueden utilizarse para efectuar la difusión de las acciones de control.

El servidor 150 puede iniciar su actividad en el estado verde 400, que puede ser un estado no de sobrecarga (o normal). Pueden existir cuatro estados de sobrecarga disponibles, esto es, estado amarillo no CallP 401, estado rojo no CallP 402, estado amarillo CallP 403 y estado rojo CallP 404. Cada estado 401-404 puede tener un conjunto de acciones de control asociadas que pueden aplicarse para mitigar la sobrecarga una vez que el servidor SIP 150 entre en estado de sobrecarga específico.

Las acciones de control en cada uno de los estados 401-404 pueden incluir acciones de control internas y acciones de control externas. Las acciones de control internas pueden representar acciones realizadas en el servidor SIP 150 de forma interna, tales como, a modo de ejemplo, aumentar automáticamente los tiempos de nueva tentativa de SIP para derivar el tráfico correspondiente y evitar la congestión de mensajes, desechando los mensajes extraídos de la cola de espera que son más antiguos que una edad de mensaje máxima predefinida, doblando todos los temporizadores de avance de rutas, desactivando automáticamente las tareas de mantenimiento programas, desactivando todas las funciones de registro, etc. Las acciones de control externas pueden representar acciones de control retransmitidas a dispositivos próximos del servidor SIP 150 tales como, a modo de ejemplo, bloquear un 50 % de las demandas nuevas reenviando un mensaje "503 servicio no disponible" con o sin una cabecera de 'vuelva a intentarlo después', un mensaje "302 temporalmente desplazado" redireccionando la llamada a otro servidor SIP 150, reenviando ninguna respuesta a la demanda de SIP próximo en el tráfico para estimular un avance de ruta, etc.

A modo de ejemplo, la acción de control asociada con el estado amarillo no CallP 401 puede incluir: rechazar un 50 % de mensajes SIP no CallP y/o descartar mensajes más antiguos que una edad de mensaje máxima desde las colas de espera 101 y/o 111.

A modo de ejemplo, la acción de control asociada con un estado rojo no CallP 402 puede incluir: rechazar un 100 % de los mensajes SIP no CallP y/o descartar los mensajes más antiguos que una edad de mensaje máxima desde las colas de espera 101 y/o 111.

A modo de ejemplo, la acción de control asociada con el estado amarillo CallP 403 puede incluir: rechazar un 50 % de llamadas nuevas; doblar los temporizadores de reintentos SIP y encaminar temporizadores de avance; desactivar las tareas de mantenimiento automáticamente programadas; desactivar todas las funciones de registro y/o aplicar

todas las acciones de control de estado rojo no CallP.

A modo de ejemplo, la acción de control asociada con el estado rojo CallP 404 puede incluir: rechazar un 100 % de las llamadas nuevas; cuadruplicar los temporizadores de reintentos de SIP y encaminar los temporizadores de avance y aplicar todas las acciones de control de estado amarillo no CallP.

Las acciones de control asociadas con cualquiera de los estados de sobrecarga de CallP 403-404 pueden incluir todas las acciones de control para el estado rojo no CallP 402.

A partir del estado verde 400, un servidor SIP 150 puede realizar la escalada a cualquiera de los estados de sobrecarga existentes 401-404 sobre la base de las condiciones de iniciación operativa asociadas con métricas de recursos del sistema (p.e., retardo de cola de espera de CallP 201, retardo de cola de espera no CallP 202, utilización de memoria de procesos 204, utilización de unidad CPU 203, etc.). Las condiciones de iniciación operativa no CallP pueden impulsar los estados de sobrecarga de no CallP 401 y 402 y las condiciones de iniciación operativa de CallP pueden impulsar los estados de sobrecarga de CallP 403 y 404. Diferenciando los mensajes de CallP y no CallP, el servidor SIP 150 puede ser capaz de tratar de ajustar una sobrecarga impulsada no CallP (p.e., de tipo REGISTER) sin impactar sobre el procesamiento de llamada normal.

A modo de ejemplo, las condiciones de iniciativa operativa para la escalada del servidor SIP 150 desde el verde 400 al estado amarillo no CallP 401, por intermedio de la ruta 405 puede incluir el retardo de la cola de espera no CallP 202 que está por encima de un umbral de retardo de cola de espera para introducir el estado amarillo no CallP 401. A modo de ejemplo, las condiciones de iniciativa operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado rojo no CallP 402 a través de la ruta 406 puede incluir el retardo de cola de espera no CallP 202 que está por encima de un umbral de retardo para introducir el estado rojo no CallP 402, siendo el umbral para introducir el estado rojo no CallP 402 más alto que el umbral para introducir el estado amarillo no CallP 401.

A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado amarillo CallP 403 a través de la ruta 407 puede incluir el retardo de cola de espera CallP 201 que está por encima de un umbral de retardo de cola de espera para introducir el estado amarillo CallP 403. A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado rojo CallP 404 a través de la ruta 408 puede incluir el retardo de cola de espera CallP 201 que está por encima de un umbral de retardo de cola de espera para introducir el estado rojo CallP 404, el umbral del retardo de cola de espera CallP 201 para introducir el estado rojo CallP 404 que es más alto que el umbral de retardo de cola de espera CallP 201 para introducir el estado amarillo CallP 403.

A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado amarillo CallP capa superior 403 a través de la ruta 407 puede incluir la utilización de memoria de procesos 204 que está por encima de un umbral de utilización de memoria para introducir el estado amarillo CallP 403. A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado rojo CallP 404 a través de la ruta 408 puede incluir la utilización de memoria de procesos 204 que está por encima de un umbral de utilización de memoria para la introducción del estado rojo CallP 404, siendo el umbral de utilización de memoria de procesos 204 para introducir el estado rojo CallP 404 más alto que el umbral de utilización de memoria de procesos 204 para introducir el estado amarillo CallP 403.

A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado amarillo CallP 403 a través de la ruta 407 puede incluir la utilización de la CPU 203 que está por encima de un umbral de utilización de la CPU para introducir el estado amarillo CallP 403. A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado verde 400 al estado rojo CallP 404 a través de la ruta 408 puede incluir la utilización de la CPU 203 que está por encima de un umbral de utilización de la CPU para introducir el estado rojo CallP 404, siendo el umbral de utilización de la CPU 203 para introducir el estado rojo CallP 404 más alto que el umbral de utilización de la CPU 203 para introducir el estado amarillo CallP 403.

Una vez efectuada la transición a uno de los estados de sobrecarga 401-404, el servidor SIP 150 puede permanecer en ese estado durante un periodo de retención mínimo. Durante el periodo de retención, si empeoran las condiciones de iniciación operativa para la escalada del estado (p.e., retardo de la cola de espera o utilización de memoria de procesos), entonces, el servidor SIP 150 puede efectuar una escalada inmediata a un nuevo estado de sobrecarga con un umbral más alto correspondiente a una carga más pesada. Una vez que termine el periodo de retención, si permanece el conjunto de condiciones de iniciación operativa para la escala de estado (p.e., retardo de cola de espera o utilización de memoria de procesos), el servidor SIP 150 puede efectuar la reposición operativa del temporizador del periodo de retención y permanecer en ese estado de sobrecarga. Por el contrario, si mejora el conjunto de condiciones de iniciación operativa para la escalada de estados (p.e., retardo de cola de espera o utilización de memoria de procesos) (p.e., cae por debajo del nivel de umbral que causó la escalada de estados), el servidor SIP 150 puede efectuar la transición a un estado con un nivel más bajo de umbral o el estado verde 400.

A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado

amarillo no CallP 401 al estado rojo no CallP 402 a través de la ruta 413 puede incluir: retardo de cola de espera no CallP 202 que está por encima de un umbral de retardo de cola de espera para entrar en el estado rojo no CallP 402.

5 A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado amarillo CallP 403 al estado rojo CallP 404 a través de la ruta 414 puede incluir: retardo de cola de espera CallP 201 que está por encima de un umbral de retardo de cola de espera para introducir el estado rojo CallP 404.

10 A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado amarillo CallP 403 al estado rojo CallP 404 a través de la ruta 414 puede incluir: utilización de memoria de procesos 204 que está por encima de un umbral de utilización de memoria para introducir el estado rojo CallP 404.

15 A modo de ejemplo, las condiciones de iniciación operativa para la escalada del servidor SIP 150 desde el estado amarillo CallP 403 al estado rojo CallP 404 a través de la ruta 414 puede incluir: utilización de la unidad CPU 203 que está por encima de un umbral de utilización de la CPU para introducir el estado rojo CallP 404.

20 Una vez efectuada la transición a un estado de sobrecarga, el servidor SIP 150 puede permanecer en ese estado de sobrecarga durante un periodo de retención mínimo. Una vez que termine el periodo de retención, si ya no existe el conjunto de condiciones de iniciación operativa para la escalada de estados (esto es, retardo de cola de espera, utilización de memoria de procesos y/o utilización de la CPU), entonces, las acciones de control asociadas con ese estado pueden eliminarse y el servidor SIP 150 puede volver al estado verde 400 o a un estado cuyas condiciones de iniciación operativa asociadas comprendan umbrales más bajos.

25 A modo de ejemplo, las condiciones de iniciación operativa para hacer retornar el servidor SIP 150 desde el estado amarillo no CallP 401 al estado verde 400 a través de la ruta 412 puede incluir: retardo de cola de espera no CallP 202 que sea menor que un umbral de retardo de cola de espera para introducir el estado amarillo no CallP 401 y el servidor SIP 150 que ha ocupado el estado amarillo no CallP 401 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado rojo no CallP 402 al estado verde 400 a través de la ruta 411 puede incluir: un retardo de cola de espera no CallP 202 que sea menor que el umbral del retardo de cola de espera para la introducción del estado amarillo no CallP 401 y el servidor SIP 150 que ha ocupado el estado rojo no CallP 402 durante un periodo de retención mínimo, a modo de ejemplo, las condiciones de iniciación para efectuar la transición del servidor SIP 150 desde el estado rojo no CallP 402 al estado amarillo no CallP 401 a través de la ruta 415 puede incluir: un retardo de cola de espera no CallP 202 que es menor que un umbral de retardo de cola de espera para introducir el estado rojo no CallP 402 y no menor que el umbral de retardo de cola de espera para introducir el estado amarillo no CallP 401 y el servidor SIP 150 que ha ocupado el estado rojo no CallP 402 durante un periodo de retención mínimo, siendo el umbral del retardo de cola de espera no CallP 202 para entrar en el estado rojo no CallP 402 más alto que el umbral del retardo de cola de espera no CallP 202 para introducir el estado amarillo no CallP 401.

40 A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado amarillo CallP 403 al estado verde 400 a través de la ruta 410 puede incluir: retardo de cola de espera CallP 201 que es menor que un umbral del retardo de cola de espera para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado amarillo CallP 403 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado rojo CallP 404 al estado verde 400 a través de la ruta 409 puede incluir: un retardo de cola de espera CallP 201 que es menor que el umbral del retardo de cola de espera para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado rojo CallP 404 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para efectuar la transición del servidor SIP 150 desde el estado rojo CallP 404 al estado amarillo CallP 403 a través de la ruta 416 puede incluir: retardo de la cola de espera CallP 201 que es menor que un umbral de retardo de cola de espera para introducir el estado rojo CallP 404 y no menor que el umbral de retardo de cola de espera CallP 201 para introducir el estado amarillo CallP 403.

55 A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado amarillo CallP 403 al estado verde 400 a través de la ruta 410 puede incluir: utilización de la memoria de procesos 204 que sea menor que un umbral de utilización de memoria para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado amarillo CallP 403 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado rojo CallP 404 al estado verde 400 a través de la ruta 409 puede incluir: utilización de memoria de procesos 204 que es menor que el umbral de utilización de memoria para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado rojo CallP 404 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para efectuar la transición del servidor SIP 150 desde el estado rojo CallP 404 al estado amarillo CallP 403 a través de la ruta 416 puede incluir: utilización de memoria de procesos 204 que es menor que un umbral de utilización de memoria para introducir el estado rojo CallP 404 y no menor que el umbral de utilización de

memoria para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado rojo CallP 404 durante un periodo de retención mínimo, siendo el umbral de utilización de memoria de procesos 204 para introducir el estado rojo CallP 404 más alto que el umbral de utilización de memoria de procesos 204 para introducir el estado amarillo CallP 403.

5 A modo de ejemplo, las condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado amarillo CallP 403 al estado verde 400 a través de la ruta 410 puede incluir: utilización de la CPU 203 que sea menor que un umbral de utilización de la CPU para introducir el estado amarillo CallP 404 y el servidor SIP 150 que ha ocupado el estado amarillo CallP 403 durante un periodo de retención mínimo. A modo de ejemplo, las
10 condiciones de iniciación operativa para el retorno del servidor SIP 150 desde el estado rojo CallP 404 al estado verde 400 a través de la ruta 409 puede incluir: utilización de la CPU 203 que sea menor que el umbral de utilización de la CPU para introducir el estado amarillo no CallP 403 y el servidor SIP 150 que ha ocupado el estado rojo CallP 404 durante un periodo de retención mínimo. A modo de ejemplo, las condiciones de iniciación operativa para
15 efectuar la transición del servidor SIP 150 desde el estado rojo CallP 404 al estado amarillo CallP 403 a través de la ruta 416 puede incluir: una utilización de la CPU 203 que sea menor que un umbral de utilización de la CPU para introducir el estado rojo CallP 404 y no menor que el umbral de utilización de la CPU para introducir el estado amarillo CallP 403 y el servidor SIP 150 que ha ocupado el estado rojo CallP 404 durante un periodo de retención mínimo, siendo el umbral de utilización de la CPU 203, para introducir el estado rojo CallP 404, más alto que el umbral de utilización de la CPU 203 para introducir el estado amarillo CallP 403.

20 A modo de ejemplo de la transición de estados de sobrecarga, un servidor SIP 150 puede comenzar en el estado verde 400 durante un tráfico normal en horas de ocupación. El retardo de la cola de espera CallP medio renovable 201 para la cola de espera del adaptador CallP, puede alcanzar 1000 milisegundos que es mayor que un umbral de 980 milisegundos para introducir el estado amarillo CallP 403. El servidor SIP 150 puede efectuar la transición, a
25 través de la ruta 407, al estado amarillo CallP 403 y puede aplicar acciones de control asociadas con el estado amarillo CallP 403. El retardo de cola de espera CallP medio renovable 201 para la cola de espera del adaptador CallP 107 puede aumentarse ahora a 2000 milisegundos que es mayor que un umbral de 1580 milisegundos para introducir el estado rojo CallP 404. El servidor SIP 150 puede efectuar la transición, a través de la ruta 414, al estado rojo CallP 404 y puede aplicar acciones de control asociadas con el estado rojo CallP 404. El retardo de cola de
30 espera CallP medio renovable 201 para la cola de espera del adaptador CallP puede reducirse ahora a 1000 milisegundos. El valor del retardo puede ser suficiente para hacer retornar el servidor SIP 150 al estado amarillo CallP 403. Si el servidor SIP 150 no ha concluido el periodo de retención mínimo en el estado rojo CallP 404, el servidor SIP 150 puede permanecer en el estado rojo CallP 404. Después de que termine el periodo de retención mínimo para el estado rojo CallP 404, el retardo de la cola de espera CallP medio renovable 201 para la cola de
35 espera del adaptador CallP 107 puede reducirse a 200 milisegundos, que es inferior al umbral de retardo de 980 milisegundos para introducir el estado amarillo CallP 403 y el servidor SIP 150 puede retornar al estado verde 400. El servidor SIP 150 puede introducir el estado verde 400 a través de la ruta 209 y eliminar todas las acciones de control anteriormente aplicadas.

40 El servidor SIP 150 solamente puede estar en un estado de sobrecarga en un momento determinado. Aunque los estados disponibles pueden ser mutuamente exclusivos en su existencia, puesto que las acciones de control asociadas con los estados de sobrecarga CallP 403 y 404 pueden incluir todas las acciones de control asociadas con el estado rojo no CallP 402, los estados de sobrecarga CallP 403 y 404 pueden ser superestados que abarcan los estados de sobrecarga no CallP 401 y 402. De este modo, tan pronto como el servidor SIP 150 entra en un
45 estado de sobrecarga CallP 403 o 404, pueden cesar de aplicarse los estados de sobrecarga CallP 401 y 402. Si, cuando el servidor SIP 150 sale de un estado de sobrecarga CallP 403 o 404 y continúan las condiciones garantizando un estado de sobrecarga no CallP 401 o 402, el servidor puede introducir el estado de sobrecarga no CallP retornando al estado verde 400 y efectuando una transición inmediata al estado de sobrecarga no CallP.

50 Un servidor SIP proxy se refiere a un servidor que enruta las sesiones SIP. Un servidor SIP de registro se refiere a un servidor SIP que gestiona el registro de agentes de usuarios. El servidor SIP 150 puede ser, a la vez, un servidor proxy y un servidor de registro diferenciando dos tipos de mensajes y estados. A modo de ejemplo, el tipo de mensajes CallP es para enrutamiento de sesiones. A modo de ejemplo, el tipo de estado de sobrecarga no CallP es para gestionar una sobrecarga de registro de agentes de usuarios. De este modo, el servidor SIP 150, dado a
55 conocer por la invención, puede ser, a la vez, un servidor proxy un servidor de registro.

El servidor SIP 150 puede ponerse en práctica como un ordenador o un sistema informático. Un servidor SIP puede ser también una aplicación de servidor SIP en un ordenador o sistema informático. Una aplicación de servidor SIP puede ser un programa informático que acepte mensajes SIP desde una red, procese los mensajes y aplique el servicio cuando sea adecuado.
60

Un ordenador o sistema informático puede tener al menos el hardware siguiente: un procesador (p.e., CPU); una pluralidad de memorias; una conexión de red y un soporte legible por ordenador que contiene software para ejecutarse por el ordenador o sistema informático. A modo de ejemplo, la conexión de red de hardware puede incluir una tarjeta de interfaz de red (NIC). Asimismo, a modo de ejemplo, el soporte legible por ordenador puede incluir: una unidad de disco duro, un disco flexible; un disco magneto-óptico; un disco óptico; tal como un CD-ROM o un
65

DVD o un disco Blue-ray; una cinta magnética; una memoria extraíble instantánea; un circuito integrado de memoria y/o otros tipos de medios de soporte que pueden memorizar instrucciones legible por ordenador.

5 La Figura 5A ilustra un ordenador de servidor, a modo de ejemplo, en relación con una red de comunicaciones en una forma de realización a modo de ejemplo, de la invención. La red de comunicación 500 puede incluir, a la vez, sistemas de red PSTN y basados en IP. El ordenador de servidor 501 puede recibir mensaje SIP entrantes 100 desde la red de comunicación 500 a través de la ruta 502 y entregar mensajes SIP salientes 113 a través de la ruta 503 a la red de comunicación 500.

10 La Figura 5B ilustra un sistema informático de servidor, a modo de ejemplo, que comprende múltiples ordenadores de servidores en una forma de realización, a modo de ejemplo, de la invención. Una parte llamante 504 puede efectuar una llamada en una red de comunicación 500 con la parte llamante 534 como receptora. El enlace 506 entre la parte llamante 504 y la red de comunicación 500 puede ser bidireccional. Un ordenador de servidor SIP 508 puede recibir mensajes SIP entrantes 100 desde la red de comunicación 500 a través del enlace 507 y puede
15 procesar los mensajes SIP entrantes 100. El enlace 507 entre el ordenador de servidor 508 y la red de comunicación 500 puede ser bidireccional. El ordenador de servidor puede enviar algunos de los mensajes SIP salientes 113 a al menos uno de los ordenadores de servidor SIP próximos en el sistema informático del servidor, esto es, 510, 512 y 514 para un procesamiento adicional antes de que los mensajes SIP salientes 113 puedan enrutarse a la red de comunicación 500. Un mensaje SIP puede desplazarse, de este modo, a través de múltiples ordenadores de servidor SIP en ruta hacia la parte llamante 534. Los enlaces 509, 511, 513, 516, 517 y 518 entre ordenadores de servidores SIP 508, 510, 512 y 514 pueden ser bidireccionales y pueden transmitir paquetes IP. Un ordenador de servidor SIP próximo 510 o 512 o 514 puede reenviar también algunos de los mensajes SIP salientes 113 a otro ordenador de servidor SIP próximo. De este modo, los mensajes SIP asociados con la misma sesión pueden desplazarse a través de diferentes ordenadores de servidor SIP en ruta hacia la parte llamante 534. Pueden existir al
20 menos dos partes llamantes 534 y 504 en una sesión de llamada de Internet. El enlace 505 entre la parte llamante 534 y la red de comunicación 500 puede ser bidireccional.

La Figura 5C ilustra una configuración, a modo de ejemplo, de partes llamantes procedentes de diferentes redes en una forma de realización, a modo de ejemplo, de la invención. Una parte llamante 521 puede efectuar en una
30 llamada en una red de comunicación 519 con la parte llamante 522 como receptora. El enlace 532 entre la parte llamante 521 y la red de comunicación 519 puede ser bidireccional. Puesto que la parte llamante 522 puede residir en una red de área local (LAN) 520 detrás de un dispositivo cortafuegos, firewall, los mensajes SIP entrantes 100 pueden filtrarse en un ordenador de periferia 523. El enlace 527 entre el ordenador de periferia 523 y la red de comunicación 519 puede ser bidireccional. Posteriormente, los mensajes SIP pueden servirse por un servidor SIP en el ordenador de periferia 523 o un ordenador de servidor SIP 524, 525 o 526 en la red LAN 520. En adelante, los mensajes SIP salientes 113 pueden generarse en ruta a la parte llamante 522. Los enlaces 528-513 y 533 entre la red LAN 520 y un ordenador de servidor SIP en la red LAN 520 pueden ser bidireccionales.

40 Cuando se pone en práctica como un ordenador o sistema informático, el servidor SIP 150 puede ser generalmente aplicable hacia las comunicaciones multimedia a través de Internet que pueden utilizar una combinación de diferentes formas de contenidos incluyendo, a modo de ejemplo, texto, audio/voz, imágenes fijas, animación, vídeo y/o forma de contenidos de interactividad.

45 Cuando se pone en práctica como un programa informático de aplicación del servidor SIP, la invención puede materializarse como un soporte legible por ordenador que contiene códigos de software para su ejecución en un ordenador. El software puede codificar el módulo de procesamiento 200 y/o módulo de control de sobrecarga 114.

A modo de ejemplo, el módulo de procesamiento 200 puede ponerse en práctica como software, ejecutable en un ordenador, para recibir mensajes SIP y procesar mensajes SIP. Las instrucciones pueden codificar los componentes
50 ilustrados en el servidor SIP 150 de la Figura 1. A modo de ejemplo, las colas de espera 101, 103, 104, 107, 108 y 111 pueden ponerse en práctica como estructuras de datos de tipo "primero en entrar, primero en salir". A modo de ejemplo, un decodificador SIP rápido 102 puede ponerse en práctica como instrucciones para mensajes SIP recibidos del tipo de difusión en CallP relacionado o no CallP relacionado. A modo de ejemplo, los decodificadores 105 y 106 pueden ponerse en práctica como instrucciones para colocar contenidos de mensajes, objeto de análisis sintáctico, en la cola de espera asociada con una sesión existente a la que pertenece el mensaje o en una nueva cola de espera correspondiente a una nueva sesión. A modo de ejemplo, las iniciativas operativas de adaptador 109 y 110 pueden ponerse en práctica como instrucciones para procesar mensajes en cola de espera y aplicar servicios cuando sea adecuado. A modo de ejemplo, el codificador 112 puede ponerse en práctica como instrucciones para
55 construir mensajes SIP salientes y enviar los mensajes SIP salientes.

60 La forma de realización del soporte de la aplicación del servidor SIP puede codificar también el módulo de control de sobrecarga 114 según se ilustra en las Figuras 2 y 3. A modo de ejemplo, el módulo de control de sobrecarga 114 puede ponerse en práctica como instrucciones para supervisar las métricas de recursos (p.e., retardo de cola de espera CallP 201, retardo de cola de espera no CallP 202, utilización de la CPU 203, utilización de memoria de procesos 204), para comparar una métrica de recursos supervisada con respecto a un umbral y para efectuar la transición del estado del servidor SIP cuando sea adecuado. Las instrucciones pueden incluir también una interfaz
65

de software 207 para las comunicaciones con el módulo de procesamiento 200, de modo que el módulo de control de sobrecarga 114 pueda enviar señales a decodificadores 115 y 116 en el módulo de procesamiento 200. La interfaz de software puede ponerse en práctica como conectores, interfaz de paso de mensajes (MPI) y/o memoria compartida.

5 Cuando se pone en práctica como un programa informático de aplicación de servidor SIP, la información del estado de servidor puede codificarse en la pila o el segmento de datos asociado con el programa de aplicación de servidor SIP que se ejecuta como un proceso en un ordenador o sistema informático. Además, el programa informático puede incluir instrucciones que definen acciones de control asociadas con cada estado que pueden activarse a través de una interrupción de software basada en reglas o un gestor de excepciones. Varios lenguajes de programación, a modo de ejemplo, JAVA, C y C++, tienen soporte para la interrupción del software y el gestor de excepciones.

15 El tiempo de entrada de cola de espera puede registrarse como una etiqueta en el mensaje SIP a insertarse en una cola de espera. Cuando se pone en práctica como software, la etiqueta puede ser un campo en la estructura de datos del mensaje SIP. El tiempo de entrada de cola de espera puede registrarse también en cualquier lugar, a modo de ejemplo, una estructura de datos de tabla de consulta en la pila o segmento de datos asociado con el programa de aplicación del servidor al que puede acceder el módulo de control de sobrecarga 114. El tiempo de entrada de cola de espera puede ser objeto de lectura desde el hardware del ordenador que ejecuta el programa de aplicación del servidor SIP, a modo de ejemplo, el reloj del sistema en la placa principal del ordenador. El tiempo de entrada de cola de espera puede ser también objeto de lectura desde un dispositivo distante, a modo de ejemplo, un servidor de tiempo de red que ejecuta el protocolo de tiempo de red (NTP) o el protocolo de tiempo de red simple (SNTP). Además de un formato de coma flotante, el tiempo de entrada de cola de espera codificado puede redondearse también a un formato de punto fijo o un formato ASCII para ahorrar tiempo en el cálculo informático del tiempo del retardo de cola de espera.

20 La utilización de la CPU 203 y la utilización de la memoria de procesos 204 pueden obtenerse consultando el sistema operativo en el ordenador que ejecuta la aplicación del servidor SIP mediante, a modo de ejemplo, interfaz de programas de aplicación (APIs). Varios lenguajes y herramientas de programación pueden poner en práctica interfaces APIs para acceder a la información de utilización de la CPU 203 y la utilización de memoria de procesos 204 asociada con un programa informático que se ejecuta como un proceso en un ordenador. A modo de ejemplo, Microsoft Visual Studio tiene una API de GetProcessMemoryInfo para un programa de usuario para obtener información de la cantidad de memoria utilizada por un proceso. El módulo de control de sobrecarga 114 puede obtener, incluso directamente, la utilización de la CPU 203 y/o la utilización de la memoria de procesos 204 de un proceso mediante el estado del árbol denominado /proc bajo el número de identificación de proceso de dicho proceso en un ordenador que ejecuta el sistema operativo LINUX.

30 Los umbrales 209 utilizados para comparación con respecto a las métricas de recursos supervisados 201-204 pueden ser configurables. Los umbrales pueden codificarse en un fichero de configuración o memorizarse en una base de datos para ser objeto de lectura por una aplicación de servidor SIP que materializa la invención. Los umbrales pueden introducirse también o modificarse por un usuario humano u otro programa informático cuando la aplicación del servidor, que materializa la invención, se está ejecutando como un proceso en un ordenador. A modo de ejemplo, una interfaz de usuario gráfica (GUI) puede utilizarse para un ser humano para introducir los umbrales de métricas de recursos supervisadas. A modo de ejemplo, una interfaz de software, en la forma de conectores, interfaz de paso de mensajes y/o memoria compartida pueden utilizarse por otro programa informático para enviar los umbrales para métricas de recursos supervisadas a la aplicación del servidor que materializa la invención que se ejecuta como un procesos.

45 Los periodos de retención mínimos que estipulan el tiempo más corto que un servidor SIP 150 pueda permanecer en un estado de sobrecarga puede ser configurable de la misma manera.

50 La edad máxima de mensajes SIP, más allá de la que pueden descartarse los mensajes SIP en un servidor SIP de sobrecarga, puede ser configurable de la misma manera.

55 La aplicación de servidor puede ponerse en práctica como otro servidor SIP monolítico 150 según se ilustra en la Figura 1 que comprende un módulo de procesamiento 200 y un módulo de control de sobrecarga 114. La utilización compartida de datos y comunicaciones puede ser simple lo que puede permitir una integración estricta.

60 La aplicación del servidor puede ponerse en práctica también como módulo de procesamiento separado 200 y módulo de control de sobrecarga 114. El enfoque modular puede permitir a un módulo de control de sobrecarga 114 realizar un control de sobrecarga para varios servidores SIP 150, pudiendo cada uno de ellos diseñarse para diferentes fines que pueden dar lugar a diferentes niveles de umbrales para métricas de recursos. A modo de ejemplo, los usuarios preferentes que pagan más cuotas pueden designarse para un servidor SIP 150 con umbrales de niveles más bajos. A modo de ejemplo, un servidor SIP 150 puede dedicarse a 911 llamadas con niveles más bajos de umbrales. Además, el módulo de control de sobrecarga 114 puede realizar el control de sobrecarga para servidores SIP 150 residentes en diferentes ordenadores. De este modo, un enfoque modular puede mejorar la

flexibilidad y granularidad en la calidad del servicio.

Las utilidades, a modo de ejemplo, y formas de realización aquí descritas son a modo de ejemplo no limitadores.

- 5 La invención se describe en detalle con respecto a formas de realización, a modo de ejemplo, y será ahora evidente a partir de lo que antecede, para los expertos en esta técnica, que se pueden realizar cambios y modificaciones sin desviarse por ello del alcance de protección de la invención en sus aspectos más amplios.
- 10 La totalidad de las técnicas dadas a conocer en esta especificación (incluyendo cualesquiera reivindicaciones adjuntas, resumen y dibujos) y/o la totalidad de las etapas de cualquier método o proceso así dado a conocer, pueden combinarse en cualquier combinación, excepto las combinaciones en donde al menos de algunas de dichas características y/o etapas son mutuamente exclusivas.
- 15 Cada característica dada a conocer en esta especificación (incluyendo cualesquiera reivindicaciones, resumen y dibujos adjuntos) pueden sustituirse por características alternativas que presten el mismo servicio, equivalentes o con un fin similar, a no ser que se establezca expresamente de otro modo. En consecuencia, a menos que se establezca expresamente en sentido contrario, cada característica dada a conocer es, a modo de ejemplo, solamente una serie genérica de características equivalentes o similares.
- 20 La invención no está restringida a los detalles de las formas de realización anteriores. La invención se extiende a cualquier novedad inventiva o cualquier combinación de novedades inventivas, de las características dadas a conocer en esta especificación (incluyendo cualesquiera reivindicaciones adjuntas, resumen y dibujos) o a cualquier novedad operativa, o cualquier combinación de novedades operativas, de las etapas de cualquier método o proceso así dado a conocer.
- 25

REIVINDICACIONES

1. Un método para gestionar la sobrecarga para un servidor bajo protocolo de iniciación de sesión, SIP, que comprende:

la recepción, al nivel del servidor SIP, de mensajes SIP procedentes de una red de comunicación;

el procesamiento, en el servidor SIP, de los mensajes SIP recibidos con el fin de obtener mensajes SIP procesados, en donde la etapa de procesamiento comprende, además:

(a) el análisis sintáctico de mensajes SIP entrantes recibidos de modo que se clasifiquen los mensajes como relativo a un procesamiento no vinculado a una llamada, denominado no CallP, o como relativos a un procesamiento vinculado a una llamada, denominado CallP, y la colocación de los mensajes SIP clasificados en una de entre una cola de espera de decodificador no CallP o en una cola de espera de decodificador CallP sobre la base de la clasificación;

(b) el análisis sintáctico de un mensaje SIP puesto en cola de espera procedente de la cola de espera del decodificador no CallP y/o de la cola de espera de decodificador CallP de modo que se determine si el mensaje SIP, puesto en cola de espera, es una nueva demanda que no está asociada con sesiones existentes, una demanda asociada a una sesión existente o una respuesta asociada a la sesión existente y la puesta en cola de espera, en una nueva cola de espera de adaptador asociada a una nueva sesión o en una cola de espera de adaptador, existente asociada a una sesión también existente, del mensaje SIP puesto en cola de espera, sobre la base de la determinación;

(c) la extracción de un mensaje desde una cola de espera de adaptador, el procesamiento del mensaje y la aplicación del servicio que corresponde al mensaje;

(d) la transmisión, a una cola de espera de codificador, de información requerida en relación con una sesión existente; y

(e) la extracción de información de la cola de espera de codificador, el ensamblado de un mensaje SIP saliente y el envío del mensaje SIP saliente;

la detección de una sobrecarga sobre la base de una condición de iniciativa operativa asociada a por lo menos un recurso a la escala del sistema del servidor SIP, en donde la detección se realiza como resultado de la recepción de mensajes SIP procedentes de la red de comunicación en el servidor SIP y el procesamiento de los mensajes SIP recibidos en el servidor SIP;

la transición del servidor SIP a un estado de sobrecarga y la aplicación de acciones de control asociadas cuando se detecta la sobrecarga para mitigar la magnitud de la sobrecarga.

2. El método según se establece en la reivindicación 1, en donde el recurso a nivel del sistema del servidor SIP comprende al menos una de: utilización de memoria de procesos o utilización de la unidad central de procesamiento, CPU.

3. El método según se establece en la reivindicación 1 o la reivindicación 2, en donde la condición de iniciación operativa comprende al menos una de:

comparación de la utilización de memoria de procesos con un umbral de utilización de memoria o

la comparación de la utilización de la CPU con un umbral de utilización de la CPU.

4. El método según la reivindicación 1 que comprende, además:

descartar uno o más mensajes extraídos desde la cola de espera de adaptador existente, en donde los uno o más mensajes descartados son más antiguos que una edad de mensaje máxima predefinida;

doblaje de uno o más tiempos de avance de rutas;

desactivar una o más tareas de mantenimiento automáticamente programadas o

desactivar una o más funciones de registro.

5. El método según se establece en cualquier reivindicación precedente, en donde el estado de sobrecarga es uno de un primer estado de sobrecarga y un segundo estado de sobrecarga en donde el primer estado de sobrecarga tiene un primer conjunto de condiciones de iniciación operativa asociadas y el segundo estado de

sobrecarga tiene un segundo conjunto de condiciones de iniciación operativa asociadas, siendo los umbrales en el primer conjunto de condiciones de iniciación operativa no más altos que los umbrales en el segundo conjunto de condiciones de iniciación operativa.

5 **6.** El método según se establece en la reivindicación 5 que comprende, además, al menos una de las etapas siguientes:

10 efectuar la transición del servidor SIP en un estado normal al primer estado de sobrecarga cuando el primer conjunto de condiciones de iniciación operativa asociadas con el primer estado de sobrecarga se cumplen pero el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga no se cumplen;

efectuar la transición del servidor SIP en el estado normal al segundo estado de sobrecarga cuando el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se cumplen;

15 efectuar la transición del servidor SIP en el primer estado de sobrecarga al segundo estado de sobrecarga cuando el conjunto de las condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se cumplen;

20 efectuar la transición del servidor SIP en el primero o segundo estado de sobrecarga de nuevo al estado normal cuando los conjuntos de condiciones de iniciación operativa de sobrecarga se hacen no válidos y el servidor SIP ha ocupado el primero o segundo estado de sobrecarga durante un periodo de retención mínimo; o

25 efectuar la transición del servidor SIP en el segundo estado de sobrecarga al primer estado de sobrecarga cuando el servidor SIP ha ocupado el segundo estado de sobrecarga durante un periodo de retención mínimo y el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se hizo no válido, pero el primer conjunto de condiciones de iniciación operativa asociadas con el primer estado de sobrecarga son todavía válidas.

30 **7.** El método según se establece en cualquier reivindicación precedente, en donde un estado de sobrecarga tiene un conjunto de acciones de control asociadas para mitigar la magnitud de la sobrecarga, estando dichas acciones de control dirigidas al servidor SIP de forma interna y/o externa.

8. El método según se establece en la reivindicación 1, en donde se registra una marca temporal para seguimiento cuando un mensaje SIP se introduce en una cola de espera.

35 **9.** Un soporte legible por ordenador que comprende programas informáticos, que cuando se ejecutan por un ordenador, hacen que el ordenador realice operaciones para gestionar la sobrecarga para un servidor de protocolo de iniciación de sesión, SIP, en una red, comprendiendo dicho soporte legible por ordenador:

40 instrucciones para la recepción, en el servidor SIP, de mensajes SIP procedentes de una red de comunicación;

instrucciones para servir los mensajes SIP recibidos en el servidor SIP, que comprenden:

45 (a) instrucciones para el análisis sintáctico de los mensajes SIP entrantes para clasificar los mensajes como uno de procesamiento no de llamada, vinculado con el procesamiento no CallP o el procesamiento de llamada, CallP y colocando los mensajes en una de entre una cola de espera de decodificador no CallP o una cola de espera de decodificador CallP sobre la base de la clasificación;

50 (b) instrucciones para el análisis sintáctico de un mensaje puesto en cola de espera desde al menos una de entre la cola de espera de decodificador no CallP o la cola de espera de decodificador CallP para determinar si el mensaje es una nueva demanda no asociada con la sesión existente, una demanda asociada con una sesión existente o una respuesta asociada con la sesión existente y la colocación del mensaje en una de entre una nueva cola de espera de adaptador asociada con una nueva sesión o una cola de espera de adaptador existente asociada con una sesión existente sobre la base de la determinación;

55 (c) instrucciones para extraer un mensaje desde la cola de espera de adaptador, el procesamiento del mensaje y la aplicación de servicios correspondientes al mensaje;

60 (d) instrucciones para transmitir información requerida en referencia a una sesión existente hacia la cola de espera de codificador; y

(e) instrucciones para extraer información desde la cola de espera del codificador construyendo un mensaje SIP saliente y enviando el mensaje SIP saliente; y

65 instrucciones para efectuar la transición del servidor SIP a un estado de sobrecarga, aplicando acciones de control asociadas y definiendo una condición de iniciación operativa a cumplirse, estando la condición de iniciación operativa asociada con al menos un recurso al nivel del sistema del servidor SIP.

10. El soporte según se establece en la reivindicación 9 que comprende, además:

instrucciones para definir el recurso al nivel del sistema del servidor SIP comprendiendo al menos una de: utilización de memoria de procesos o utilización de la unidad CPU; y

las instrucciones para definir la condición de iniciación operativa que comprende al menos una de:

comparación de la utilización de memoria de procesos con un umbral de utilización de memoria; y

comparación de la utilización de la CPU con un umbral de utilización de la CPU.

11. El soporte según se establece en la reivindicación 9 o 10 que incluye, además:

Instrucciones para definir un primer conjunto de condiciones de iniciación operativa asociadas con un primer estado de sobrecarga y un segundo conjunto de condiciones de iniciación operativa asociadas con un segundo estado de sobrecarga; siendo los umbrales en el primer conjunto de condiciones de iniciación operativa no más altos que los umbrales en el segundo conjunto de condiciones de iniciación operativa.

12. El soporte según se establece en la reivindicación 11 que comprende, además, al menos una de:

instrucciones para efectuar la transición del servidor SIP en un estado normal al primer estado de sobrecarga después de haber determinado que el primer conjunto de condiciones de iniciación operativa en transición asociada con el primer conjunto se cumplen mientras que el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga no se cumplen;

instrucciones para efectuar la transición del servidor SIP en el estado normal al segundo estado de sobrecarga después de determinarse que el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se cumplen;

instrucciones para efectuar la transición del servidor SIP en el primer estado de sobrecarga al segundo estado de sobrecarga después de determinarse que el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se cumplen,

instrucciones para efectuar la transición del servidor SIP en el primero o segundo estado de sobrecarga de nuevo a un estado normal después de determinarse que el servidor SIP ha ocupado el primero o segundo estado de sobrecarga durante un periodo de retención mínimo y los conjuntos de condiciones de iniciación operativa se hicieron no válidas o

instrucciones para efectuar la transición del servidor SIP en el segundo estado de sobrecarga al primer estado de sobrecarga después de determinarse que el servidor SIP ha ocupado el segundo estado de sobrecarga durante un periodo de retención mínimo y el segundo conjunto de condiciones de iniciación operativa asociadas con el segundo estado de sobrecarga se hicieron no válidas pero el primer conjunto de condiciones de iniciación operativa asociadas con el primer estado de sobrecarga siguen siendo válidas.

13. El soporte según se establece en cualquiera de las reivindicaciones 9 a 12, que comprende, además:

instrucciones para definir un conjunto de acciones de control asociadas con un estado de sobrecarga para mitigar la magnitud de la sobrecarga, comprendiendo dicho conjunto de acciones de control unas acciones de control dirigidas hacia el servidor SIP, de forma interna y/o externa.

14. El soporte según se establece en la reivindicación 9 que comprende, además:

instrucciones para obtener una marca temporal desde al menos uno de entre el ordenador o la red; e

instrucciones para registrar la marca temporal para registro del momento en que un mensaje SIP se coloca en una cola de espera.

15. Un sistema para comunicación multimedia a través de una red que comprende:

una conexión de red para recibir y transmitir una pluralidad de mensajes SIP entre un servidor SIP y la red;

un módulo de procesamiento en el servidor SIP para procesar y servir los mensajes SIP recibidos que comprende:

una primera cola de espera para retener los mensajes SIP recibidos;

un primer decodificador para clasificar los mensajes SIP recibidos en al menos dos tipos de mensajes;

- una cola de espera de decodificador para cada tipo de mensaje para retener los mensajes SIP clasificados;
- una pluralidad de decodificadores SIP para agrupar mensajes SIP desde la cola de espera del decodificador en sesiones existentes o sesiones nuevas;
- 5 una pluralidad de colas de espera de adaptador asociadas con sesiones existentes y sesiones nuevas para retener mensajes SIP;
- 10 una pluralidad de adaptadores, cada uno de ellos asociado con una cola de espera de adaptador, para procesar y servir sesiones;
- una cola de espera de codificador para mantener la codificación de información de mensajes SIP salientes; y
- 15 un codificador para construir y enviar mensajes SIP salientes; y
- un módulo de control de sobrecarga para realizar el control de sobrecarga para el servidor SIP que comprende:
- 20 un monitor para un seguimiento de registro de al menos un recurso al nivel del sistema del servidor SIP como un resultado de la ejecución del módulo de procesamiento para procesar y servir mensajes SIP recibidos; y
- un comparador para determinar si existe sobrecarga si al menos una métrica de recursos en condiciones de registro, como resultado de la ejecución del módulo de procesamiento, es superior a un umbral,
- 25 en donde el módulo de control de sobrecarga efectúa la transición del servidor SIP a un nuevo estado si existe sobrecarga.
- 16.** El sistema según se establece en la reivindicación 15, en donde el módulo de control de sobrecarga comprende, además, una interfaz de software para enviar señales al módulo de procesamiento en donde las señales son para efectuar la transición del estado del servidor SIP.
- 30 **17.** El sistema según se establece en la reivindicación 15 o reivindicación 16, en donde el recurso al nivel del sistema del servidor SIP comprende al menos una de: utilización de memoria de procesos o utilización de la unidad central de procesamiento, CPU.
- 35 **18.** El sistema según se establece en cualquiera de las reivindicaciones 15 a 17, en donde el servidor SIP comprende el módulo de control de sobrecarga y el módulo de procesamiento.
- 19.** El sistema según se establece en cualquiera de las reivindicaciones 15 a 18, en donde el servidor SIP comprende el módulo de procesamiento pero no el módulo de control de sobrecarga.
- 40 **20.** El sistema según se establece en la reivindicación 15, en donde los decodificadores SIP están adaptados, además, para recibir señales procedentes del módulo de control de sobrecarga y después de que se reciba una señal procedente del módulo de control de sobrecarga, para realizar al menos una de las funciones de:
- 45 reenviar al menos un mensaje SIP en las colas de espera de adaptador a la cola de espera de codificador;
- rechazar al menos un mensaje SIP en las colas de espera del decodificador;
- 50 incremento de al menos un temporizador de servidor SIP;
- descartar al menos un mensaje SIP en la primera cola de espera y/o la cola de espera del codificador o
- la generación de al menos un mensaje de error por intermedio del codificador.
- 55 **21.** El sistema según se establece en la reivindicación 20, en donde el temporizador del servidor SIP es uno de entre un temporizador de reintentos o un temporizador de avance de ruta y el mensaje de error está relacionado con uno de entre un servicio no disponible o un servicio temporalmente desplazado.

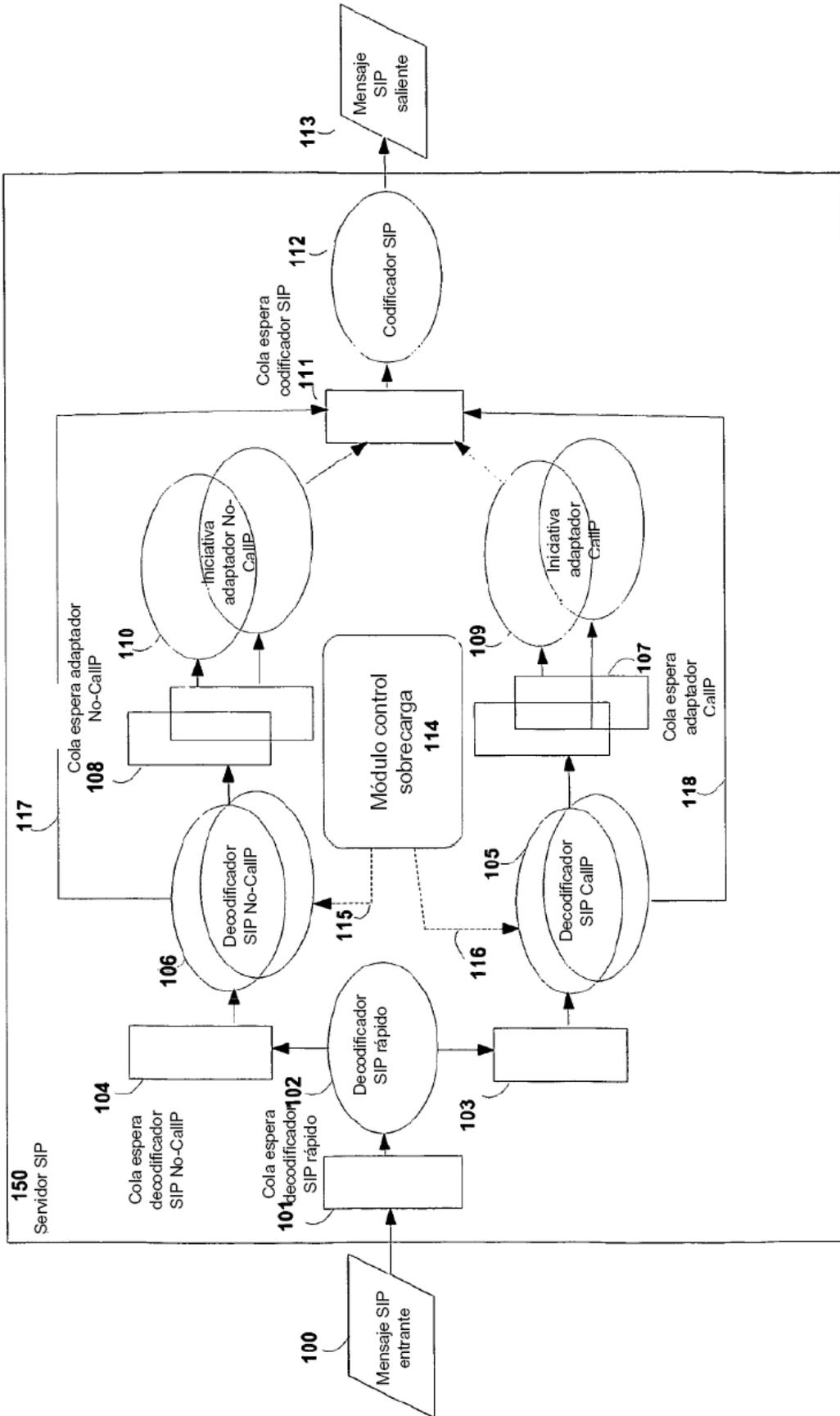


FIG. 1

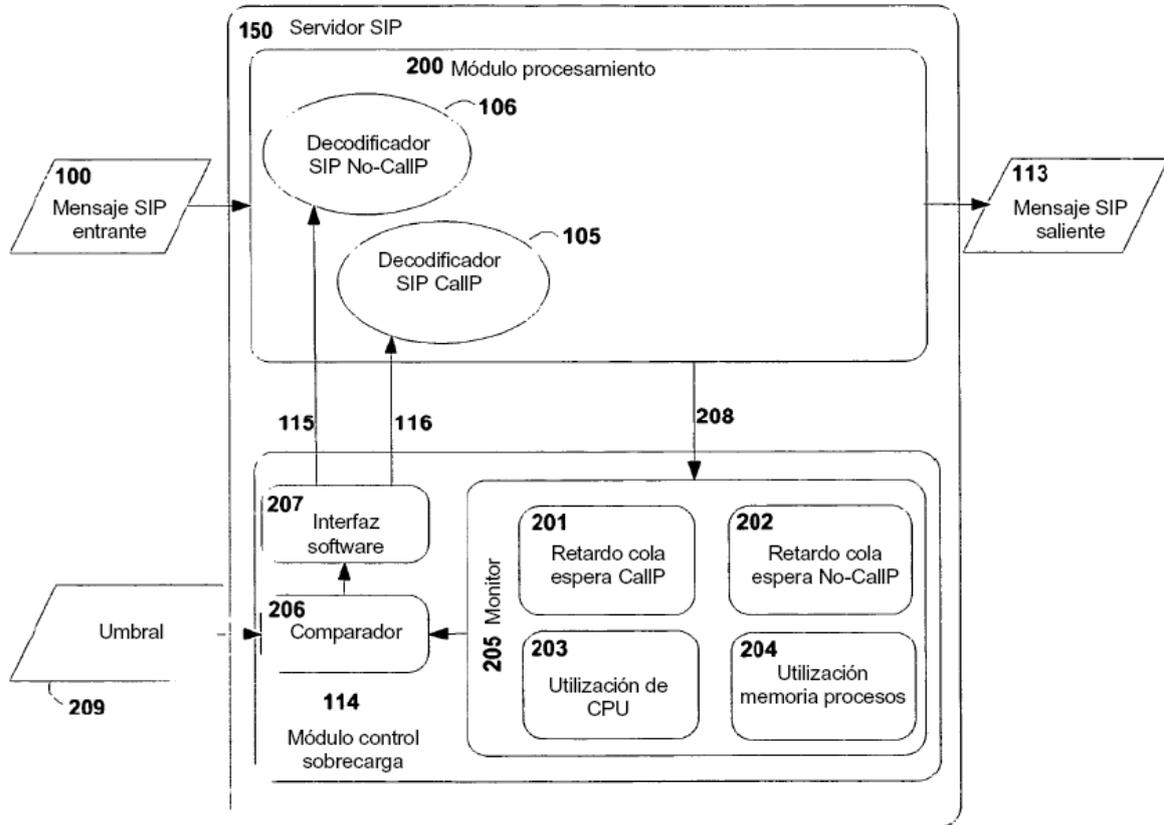


FIG. 2

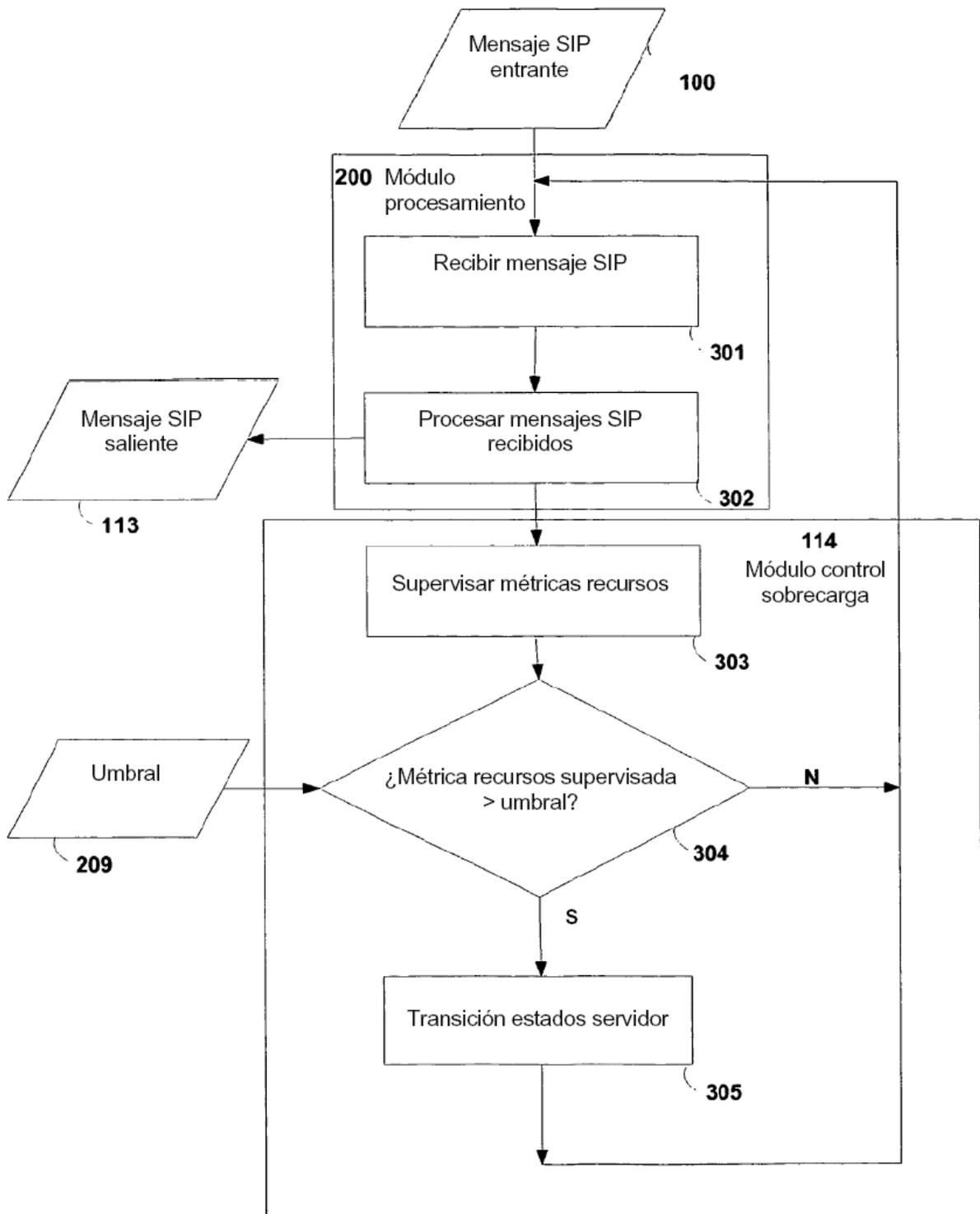


FIG. 3

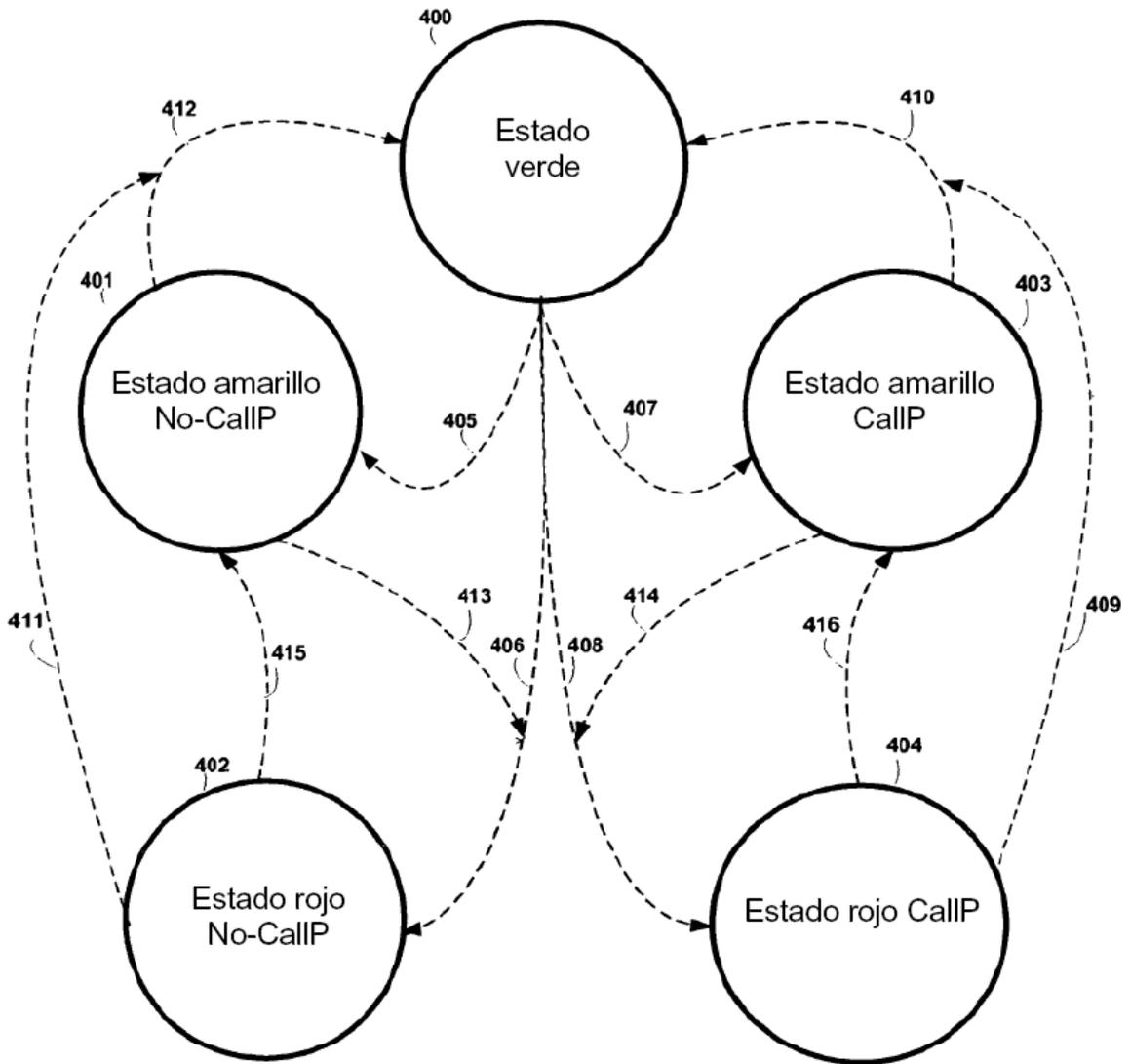


FIG. 4

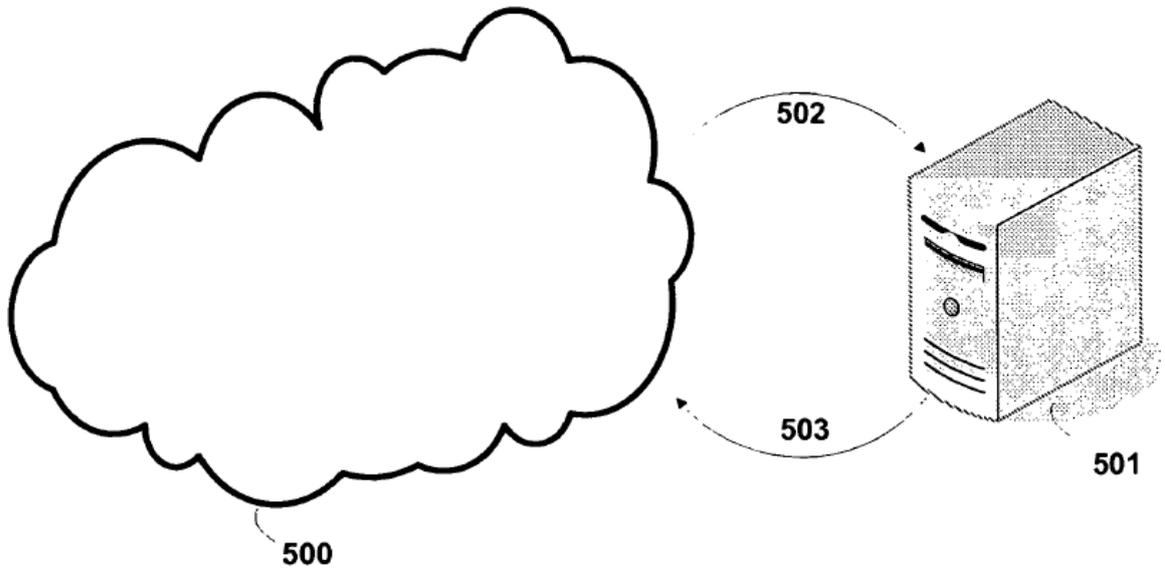


FIG. 5A

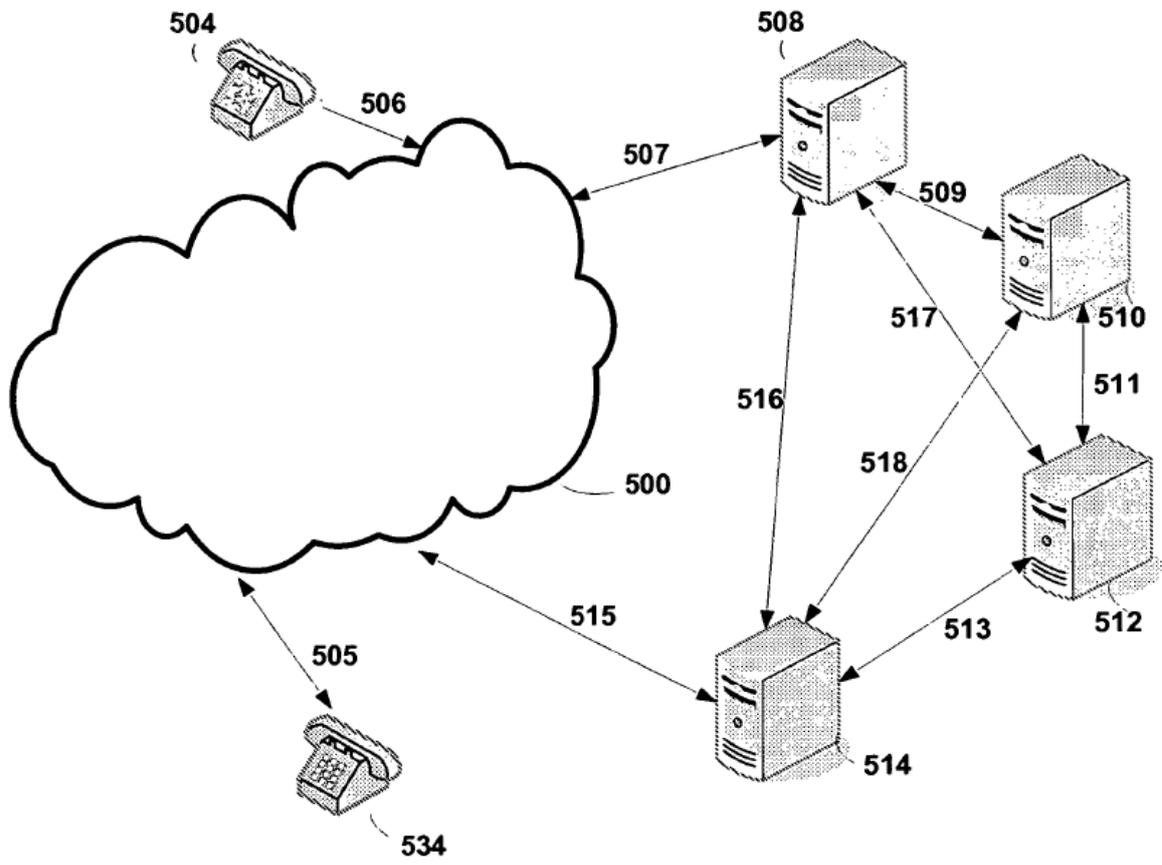


FIG. 5B

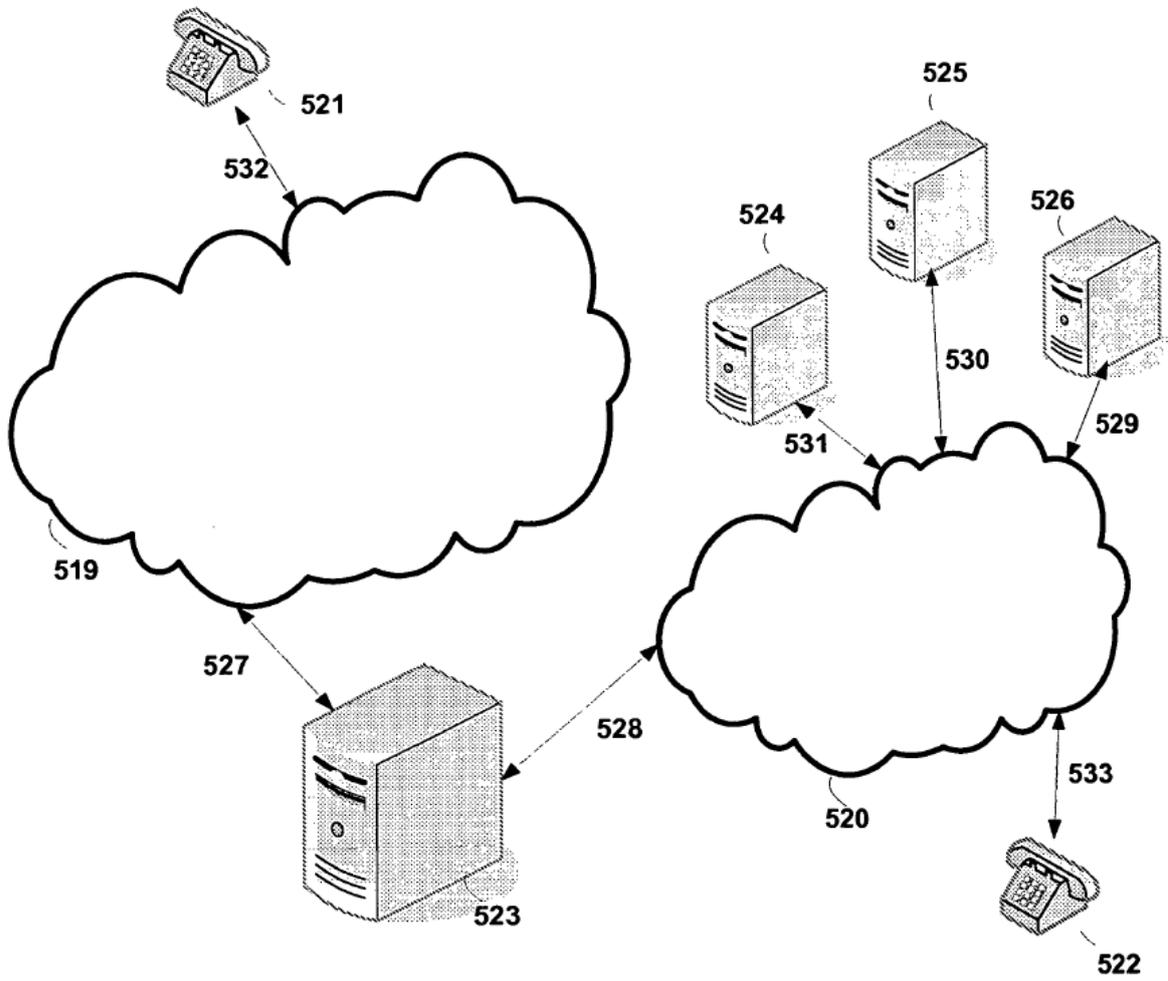


FIG. 5C