

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 607 465**

51 Int. Cl.:

H04L 12/803 (2013.01)

H04L 12/801 (2013.01)

H04L 12/841 (2013.01)

H04L 29/08 (2006.01)

H04L 12/835 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.02.2008 PCT/IB2008/000351**

87 Fecha y número de publicación internacional: **28.08.2008 WO08102230**

96 Fecha de presentación y número de la solicitud europea: **15.02.2008 E 08719141 (7)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 2116014**

54 Título: **Sistema y método para el equilibrado de cargas de información**

30 Prioridad:

20.02.2007 US 707870

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

31.03.2017

73 Titular/es:

**AMADEUS S.A.S. (100.0%)
485 Route du Pin Montard, Les Bouillides, BP 69
06902 Sophia Antipolis Cede, FR**

72 Inventor/es:

**DANIEL, JEROME y
DEGUET, CYRIL**

74 Agente/Representante:

SUGRAÑES MOLINÉ, Pedro

ES 2 607 465 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

Sistema y método para el equilibrado de cargas de información

5 Antecedentes de la invención

1. Campo de la invención

10 La presente invención se refiere a mecanismos dinámicos de equilibrado de carga y estrangulamiento. Más específicamente, la presente invención se refiere a un factor de carga notificado por la aplicación a la que se dirige que influye en un algoritmo de equilibrado de cargas en tiempo real.

2. Explicación de la información antecedente

15 El equilibrado de cargas en un Bus de Servicio Empresarial (ESB del inglés "Enterprise Services Bus") implica dos características: un destino y una ruta. Una ruta puede distribuir el tráfico a varios destinos. Los ajustes del ESB permiten la asignación de ponderaciones de equilibrado de carga a los destinos a los que se dirige por una ruta. Por ejemplo, para una ruta dada, dos destinos pueden ponderarse con 6 y 4, respectivamente. Basándose en esta ponderación, el 60 % del tráfico será enrutado al primer destino, y el 40 % al otro. Este tipo de equilibrado de carga se denomina equilibrado de carga estático. Dicho equilibrado de carga estático no es sensible a las necesidades del sistema subyacente.

25 El documento US-A- 2005/120095 describe un aparato y método para la distribución del tráfico a través de un grupo de máquinas usando estadísticas de la instancia de aplicación. Para realizar el equilibrado de carga de acuerdo con la presente invención, se proporciona un método de generación de ponderaciones para desplazar las distribuciones de equilibrado de carga. Las instancias de aplicación para las que se distribuye el tráfico, o el medio de la aplicación, se dotan para establecer ciertas métricas acerca de la aplicación mientras está en ejecución. La dotación de la instancia de aplicación proporcionará estadísticas de aplicación tal como el número de transacciones con éxito, tiempos de respuesta de la aplicación, topología de la aplicación, importancia de las transacciones que se están procesando, tiempo durante el que la aplicación está bloqueada esperando recursos, datos de consumo de recursos, y similares. Estas métricas se recogen, procesan, y se presentan a continuación como conjunto de ponderaciones al aparato de equilibrado de carga para gobernar su distribución de tráfico. Con dichas métricas de aplicación disponibles, el tráfico puede repartirse basándose en el estado actual de las instancias de aplicación y otros casos de aplicación en la trayectoria de las transacciones, la probabilidad de que la instancia de aplicación complete la solicitud, o incluso objetivos orientados a actividades comerciales de nivel más alto.

De acuerdo con la invención, un método para el enrutado de mensajes en un sistema, que comprende:

- 40 determinar un patrón de distribución inicial para el enrutado de mensajes a diferentes destinos;
- determinar un parámetro de cola basándose en al menos una cantidad actual de uso del sistema y un uso potencial máximo del sistema de un sitio objetivo en un destino;
- determinar un parámetro de tiempo basándose en una cantidad de tiempo que ha estado esperando para su procesamiento un mensaje en una aplicación en el destino;
- 45 determinar un parámetro de carga basándose en la combinación del parámetro de tiempo y del parámetro de cola; y
- modificar el patrón de distribución inicial basándose en la siguiente fórmula:

$$Wd(i) = Ws(i) (1 - LF(i))$$

50 en donde:

- Ws(i) es una ponderación tal como se establece por el patrón de distribución inicial para un destino P(i) particular en una ruta;
- LF(i) es el parámetro de carga para el destino particular; y
- 55 Wd(i) es una ponderación actualizada, para el destino P(i) particular en la ruta, para la revisión del patrón de distribución;

60 en el que la ruta consiste en un conjunto de destinos, y la ponderación afecta al flujo de datos a cada destino de la ruta; y
en el que:

el parámetro de carga para un destino particular, LF(i), se calcula mediante la aplicación de la fórmula:

$$LF(i) = Wq LFq + Wet LFet$$

65

en donde:

Wq y Wet son ponderaciones personalizables que permiten la personalización del sistema;
 LFq es un factor inicial de carga de cola; y
 LFet es un factor inicial de existencia en el tiempo;

dicha determinación del parámetro de cola, que se basa en el factor inicial de carga de cola, comprende la aplicación de la fórmula:

$$LFq = QUP / HLFM$$

en donde:

LFq es el factor inicial de carga de cola;
 HLFM es un porcentaje preestablecido del número máximo de mensajes que una cola puede procesar; y
 QUP representa el número de mensajes en la cola en comparación con su capacidad; y

dicha determinación del parámetro de tiempo, que se basa en el factor inicial de existencia en el tiempo, comprende la aplicación de la fórmula:

$$LFet = (ATT - TTT) / TTS$$

en donde:

LFet es el factor inicial de existencia en el tiempo;
 ATT se calcula basándose en el tiempo de transacción medio calculado por la aplicación;
 TTT es un tiempo mínimo predeterminado; y
 TTS es un parámetro configurable que define cómo de rápido reacciona el parámetro del factor de carga a un incremento en el tiempo de transacción.

El parámetro de carga puede basarse en al menos una combinación del parámetro de tiempo y del parámetro de cola, o al menos una combinación ponderada del parámetro de tiempo y el parámetro de cola. El método puede incluir adicionalmente la determinación de un período de sobrecarga del sistema y descarte, durante el período, de mensajes que coincidan con criterios predeterminados. El método puede basarse en un algoritmo de Bresenham de trazado de líneas para calcular sustancialmente en tiempo real una distribución de los mensajes. La modificación puede comprender adicionalmente la aplicación de la fórmula:

$$N_i \leq k \frac{w_i}{\sum w_i}$$

en donde k es el número de mensajes ya recibidos por el ESB, Ni es el número de mensajes enviados al destino i, y Wi es la ponderación del destino i.

Pueden determinarse otras realizaciones de ejemplo y ventajas de la presente invención mediante la revisión de la presente divulgación y los dibujos adjuntos.

Breve descripción de los dibujos

La presente invención se describe adicionalmente en la descripción detallada que sigue, en referencia a la pluralidad de dibujos que se han hecho notar por medio de ejemplos no limitativos de ciertas realizaciones de la presente invención, en los que los números de referencia iguales representan elementos iguales a todo lo largo de las diversas vistas de los dibujos, y en los que:

La Fig. 1 es una visión general de una realización de la invención con factores de carga iguales;
 la Fig. 2 es una visión general de una realización de la invención con factores de carga desiguales; y
 la Fig. 3 es un gráfico que muestra la distribución de mensajes para la realización de la Fig. 2.

Descripción detallada de la realización de ejemplo

Las particularidades mostradas en el presente documento lo son a modo de ejemplo y solamente para finalidades de explicación ilustrativa de las realizaciones de la presente invención y se presentan para proporcionar lo que se cree es la descripción más útil y fácilmente entendida de los principios y aspectos conceptuales de la presente invención. En este sentido, no se realiza ningún intento para mostrar detalles estructurales de la presente invención con más detalle del que es necesario para el entendimiento fundamental de la presente invención, la descripción tomada con los dibujos hace evidente para los expertos en la materia cómo pueden realizarse en la práctica las diversas formas

de la presente invención.

El equilibrado dinámico de cargas implica cambios de la distribución del equilibrado estático de cargas basándose en el comportamiento del sitio objetivo en el destino. El sitio objetivo y la red intercambian en tiempo real un factor de carga que representa la evolución de las cargas en el sitio objetivo. Este factor de carga se usa para actualizar la distribución de ponderaciones. La metodología de equilibrado de cargas de la realización preferida se implementa preferentemente como parte de un Bus de Servicio Empresarial (ESB), aunque es igualmente aplicable a otros entornos.

Preferentemente hay alguna tasa mínima a la que se actualiza el factor de carga. A modo de ejemplo no limitativo, el tiempo mínimo entre dos intercambios de factor de carga podría establecerse en 1 segundo. Cualquier tiempo mínimo apropiado podría seleccionarse con una actualización correspondiente de los factores de carga sustancialmente en tiempo real. Siempre que el ESB no recibe un nuevo factor de carga, supone que el factor de carga permanece sin cambios.

El factor de carga de una aplicación dada afecta a la ponderación estática $W_s(i)$ del destino (P_i) correspondiente en la ruta. Si el factor de carga se considera como un valor entre 0 y 1 (o un porcentaje correspondiente), entonces la ponderación dinámica $W_d(i)$ para el destino será:

$$W_d(i) = W_s(i) (1 - LF(i))$$

Ocasionalmente el sitio objetivo puede estar tan sobrecargado ($LF(i) = 1$) que no envíe actualizaciones del factor de carga. El ESB no recibirá por lo tanto mensajes hasta que disminuya el factor de carga, cualquiera que sea su ponderación estática. Si no se envían los mensajes, el algoritmo de enrutado global permanece así el mismo, tanto si se usa como si no el equilibrado dinámico de cargas. Puede usarse un algoritmo basado en Bresenham más eficiente para evitar los más pesados cálculos en coma flotante cada vez que se actualizan los factores de carga.

El factor de carga se calcula preferentemente en el lado de la aplicación y basándose en al menos dos criterios: (1) el número de mensajes en cola esperando a ser procesados, y (2) el tiempo de existencia medio. La ponderación de estos dos sub-factores puede configurarse para el cálculo del factor de carga del nodo global.

El primer factor se calcula basándose en el número de mensajes en espera a ser procesados. Para este cálculo, una "marca de agua de alto factor de carga" es un porcentaje preestablecido definido por el usuario u operador del número máximo de mensajes que una cola puede procesar, y un "porcentaje de uso de cola" representa el número de mensajes en la cola en comparación con su capacidad.

El factor preliminar de carga de cola se calcula en consecuencia tal como sigue:

$$LF_q = QUP / HLFM$$

en donde:

- LF_q es el factor preliminar de carga de cola;
- QUP es el porcentaje de uso de cola; y
- HLFM es la marca de agua de alto factor de carga.

En la mayor parte de los casos, el factor preliminar de carga de cola puede ser el factor final de carga de cola. Hay preferentemente excepciones, sin embargo, particularmente cerca de los límites superior e inferior de los valores potenciales para LF_q. Dos de dichas excepciones son las siguientes:

- Cuando el LF_q preliminar está por debajo de un porcentaje mínimo, el LF_q final se fija a ese valor mínimo. Ejemplos no limitativos del porcentaje mínimo es el 10 %, y el valor mínimo es 0.
- Cuando QUP = HLF, entonces LF_q alcanza su valor máximo. Si el LF_q preliminar es aún mayor que un porcentaje máximo durante el procesamiento posterior, entonces el LF_q final se establece al valor máximo (como si QUP = HLFM). Ejemplos no limitativos del porcentaje máximo es el 95 % y el valor máximo es 1.

Pueden usarse cualesquiera valores y/o porcentajes apropiados. El ajuste puede ser para un único evento, iterativamente hasta que las condiciones cesen (por ejemplo, LF_q permanece en el valor máximo hasta que cae por debajo del porcentaje máximo), o para un número fijo de iteraciones. Los porcentajes y valores pueden ser fijos o variables.

El segundo factor, al que se hace referencia en el presente documento como el factor de carga de tiempo de existencia, representa el tiempo empleado por un mensaje en la aplicación para su procesamiento. Se actualiza preferentemente cada vez que se procesa un mensaje de acuerdo con la siguiente fórmula

$$LF_{et} = (ATT - TTT) / TTS$$

en donde:

LFet es el factor de tiempo de existencia;

5 el Tiempo de Transacción Medio (ATT) es el tiempo de transacción medio (preferentemente en un período de integración de 2 segundos) (tiempo entre la recepción de la consulta del mensaje y la respuesta al mensaje) calculado por la aplicación;

el Umbral del Tiempo de Transacción (TTT) es un tiempo mínimo preestablecido definido por el usuario o el operador; y

10 la Escala del Tiempo de Transacción (TTS) es un parámetro configurable que define cómo de rápido reacciona el factor de carga a un incremento en el tiempo de transacción.

Si el ATT es menor que el TTT, entonces el factor de carga se establece a un valor mínimo. Este valor es preferentemente 0.

15 Basándose en los cálculos anteriores, el factor de carga total viene dado por la siguiente fórmula:

$$LF = Wq LFq + Wet LFet$$

20 Wq y Wet son ponderaciones personalizables (factores de multiplicación) que permiten la personalización del sistema. Puede aplicarse cualquier ponderación, incluyendo una ponderación de uno.

El factor de carga se actualiza preferentemente para cada mensaje procesado, y notificado al ESB cuando se devuelve el mensaje de respuesta.

25 En referencia a la Fig. 1, el ESB 110 tiene uno o más enrutadores 120 (denominados en el presente documento colectivamente como "enrutador") ocupados con el equilibrado de cargas. El enrutador gestiona una o más rutas, en donde cada ruta maneja uno o más destinos. El enrutador selecciona una ruta basándose en el mensaje a distribuir, y a continuación aplica una metodología de algoritmo de equilibrado de cargas.

30 Cada vez que debe enviarse un mensaje en una ruta, el enrutador 110 envía el mensaje sobre el primer destino elegible basándose en la variable de decisión. Si no hay conexión de red establecida para el destino elegido, el algoritmo continúa el bucle hasta que encuentra un destino conectado; si aún no encuentra un destino, el ESB 110 replica con un mensaje de error indicando que no puede alcanzarse ningún destino.

35 Cada destino asociado a una ruta tiene una ponderación estática, que se fija estáticamente por configuración. Por omisión, se asocia un factor de carga igual a 0 para cada ponderación estática. La ponderación dinámica es la ponderación estática multiplicada por (1 – factor de carga). El factor de carga se calcula usando la metodología indicada en el presente documento.

40 Cada destino se asocia con una variable de decisión para determinar si el destino es elegible para enviar un mensaje. Esta variable de decisión es preferentemente el número de mensajes enviados a este destino menos el número teórico de mensajes que deberían haberse enviado, basándose en el reparto de ponderaciones; un mensaje puede enviarse a un destino solo si esta diferencia es negativa.

45 Un ejemplo de la operación de una realización de la presente invención se muestra en las Figs. 1-3. En cada caso, una ruta consiste en un conjunto de destinos, y una ponderación afecta al flujo de tráfico de cada destino de la ruta. En la realización, la aplicación alcanzada por la ruta R1 se distribuye en tres nodos A, B y C. De ese modo en el lado del ESB, la ruta R1 contiene 3 destinos a alcanzar por los 3 nodos de aplicación. El factor de carga calculado en cada nodo ($LF = Wq LFq + Wet LFet$ explicado anteriormente) se notifica al ESB, que afecta a las ponderaciones asignadas a cada destino. Dado que el factor de carga en este ejemplo es cero, se asignan ponderaciones W iguales a los nodos de destino individuales como sigue:

55 Dest. A: $W_a = 10$
 Dest. B: $W_b = 10$
 Dest. C: $W_c = 10$

Los mensajes se distribuyen a los nodos basándose en las ponderaciones anteriores y la fórmula siguiente:

60
$$N_i \leq k \frac{w_i}{\sum w_i}$$

en donde:

k es el número de mensajes ya recibidos por el ESB (correspondiente aproximadamente al tiempo);

N_i es el número de mensajes enviados al destino i; y

W_i es la ponderación del destino i .

5 La aplicación de la ecuación asegura que el número de mensajes enviados a un destino I no excede un valor teórico, proporcional a la ponderación de este destino. En este caso, las ponderaciones iguales de los destinos crean una distribución igual de mensajes (por ejemplo, 33 %) a cada nodo de destino.

10 En referencia ahora a la Fig. 2, los mensajes que se acumulan en varias colas asociadas con los nodos inducen un cambio correspondiente en los diversos factores de carga. La Fig. 2 muestra un ejemplo en el que el factor de carga para los nodos A, B y C se incrementa desde 0 a 0,5, 0,7 y 0,9, respectivamente. Esto cambia la ponderación de los diferentes destinos, tal como sigue:

15 Dest. A: $W_a = 10 \times (1-0,5) = 5$

Dest. B: $W_b = 10 \times (1-0,7) = 3$

Dest. C: $W_c = 10 \times (1-0,9) = 1$

Basándose en estas ponderaciones, el 56 % (5/9) de los mensajes deben enviarse al nodo A, el 33 % (3/9) de los mensajes deben enviarse al nodo B, y el 11 % (1/9) de los mensajes deben enviarse al nodo C. La distribución resultante se muestra en la Fig. 3.

20 Las tres líneas diagonales en la Fig. 3 representan el flujo teórico de mensajes a cada destino. Las tres curvas escalonadas representan el flujo real (cada escalón corresponde a un mensaje que se envía a un destino). La metodología de equilibrado de cargas resultante se aproxima por lo tanto a una línea con una curva escalonada. Esto permite que se use el algoritmo de Bresenham para realizar un cálculo rápido con números enteros en lugar de números en coma flotante.

25 Una característica opcional de las metodologías anteriores es el estrangulamiento, en el que se descartan los mensajes de baja prioridad cuando la aplicación global está sobrecargada. Por ejemplo, si el factor de carga de todos los destinos por detrás de una ruta supera un umbral de factor de carga dado, todos los mensajes que coinciden con algún criterio (tipo de servicio u originador, por ejemplo) se descartan.

30 Se ha de observar que los ejemplos precedentes se han proporcionado meramente con la finalidad de explicación y de ninguna manera deben interpretarse como limitativos de la presente invención. Aunque la presente invención se ha descrito con referencia a ciertas realizaciones, se entiende que las palabras que se han usado en el presente documento son palabras de descripción e ilustración, en lugar de palabras de limitación. Pueden realizarse cambios, dentro del ámbito de las reivindicaciones adjuntas, tal como se establecen actualmente y tal como se modifiquen, sin apartarse del alcance de la presente invención en sus aspectos. Aunque la presente invención se ha descrito en el presente documento con referencia a medios, materiales y realizaciones particulares, la presente invención no está dirigida a limitarse a las particularidades divulgadas en el presente documento; por el contrario, la presente invención se extiende a todas las estructuras, métodos y usos funcionalmente equivalentes, tal como están dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un método para el enrutado de mensajes en un sistema, que comprende:

- 5 determinar un patrón de distribución inicial para el enrutado de mensajes a diferentes destinos;
- determinar un parámetro de cola basándose en al menos una cantidad actual de uso del sistema y un uso potencial máximo del sistema de un sitio objetivo en un destino;
- determinar un parámetro de tiempo basándose en una cantidad de tiempo que ha estado esperando para su procesamiento un mensaje en una aplicación en el destino;
- 10 determinar un parámetro de carga basándose en la combinación del parámetro de tiempo y del parámetro de cola; y
- modificar el patrón de distribución inicial basándose en la siguiente fórmula:

$$Wd(i) = Ws(i) (1 - LF(i))$$

15 en donde:

- Ws(i) es una ponderación tal como se establece por el patrón de distribución inicial para un destino P(i) particular en una ruta;
- 20 LF(i) es el parámetro de carga para el destino particular; y
- Wd(i) es una ponderación actualizada, para el destino P(i) particular en la ruta, para la revisión del patrón de distribución;
- en el que la ruta consiste en un conjunto de destinos, y la ponderación afecta al flujo de datos a cada destino de la ruta; y
- 25 en el que:

el parámetro de carga para un destino particular, LF(i), se calcula mediante la aplicación de la fórmula:

$$LF(i) = Wq LFq + Wet LFet$$

30 en donde:

- Wq y Wet son ponderaciones personalizables que permiten la personalización del sistema;
- 35 LFq es un factor inicial de carga de cola; y
- LFet es un factor inicial de existencia en el tiempo;
- dicha determinación del parámetro de cola, que se basa en el factor inicial de carga de cola, comprende la aplicación de la fórmula:

$$LFq = QUP / HLFM$$

40 en donde:

- LFq es un factor inicial de carga de cola;
- 45 HLFM es un porcentaje preestablecido del número máximo de mensajes que una cola puede procesar; y
- QUP representa el número de mensajes en la cola en comparación con su capacidad;
- caracterizado por que**
- dicha determinación del parámetro de tiempo, que se basa en el factor inicial de existencia en el tiempo, comprende la aplicación de la fórmula:

$$50 LFet = (ATT - TTT) / TTS$$

en donde:

- 55 LFet es un factor inicial de existencia en el tiempo;
- ATT se calcula basándose en el tiempo de transacción medio calculado por la aplicación;
- TTT es un tiempo mínimo predeterminado; y
- TTS es un parámetro configurable que define cómo de rápido reacciona el parámetro del factor de carga a un incremento en el tiempo de transacción;
- 60 en el que dicha modificación comprende adicionalmente la aplicación de la fórmula:

$$N_i \leq k \frac{w_i}{\sum w_i}$$

en donde k es el número de mensajes ya recibidos por un ESB, N_i es el número de mensajes enviados al destino i , y W_i es la ponderación del destino i ; y los mensajes se distribuyen a los destinos basándose en sus ponderaciones.

- 5 2. El método de la reivindicación 1, en el que el parámetro de carga se basa en al menos una combinación del parámetro de tiempo y del parámetro de cola, o en al menos una combinación ponderada del parámetro de tiempo y el parámetro de cola.
- 10 3. El método de la reivindicación 1, que comprende adicionalmente:
 determinar un período de sobrecarga del sistema y
 descartar, durante el período, mensajes que coincidan con criterios predeterminados.
- 15 4. El método de la reivindicación 3, que comprende adicionalmente:
 descartar, durante el período, mensajes que coincidan con criterios predeterminados.
- 20 5. El método de la reivindicación 1 en el que el método se basa en un algoritmo de Bresenham de trazado de líneas para calcular sustancialmente en tiempo real una distribución de los mensajes.

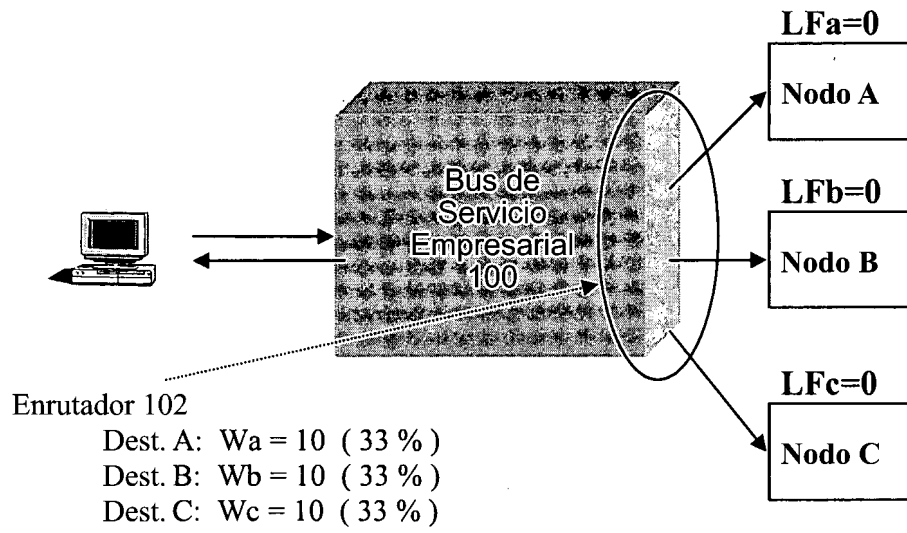
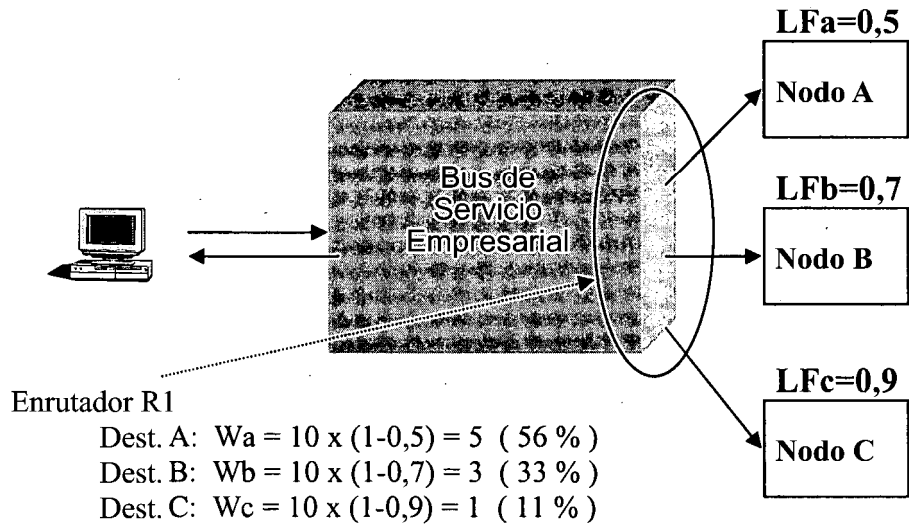


Fig.1

Fig. 2



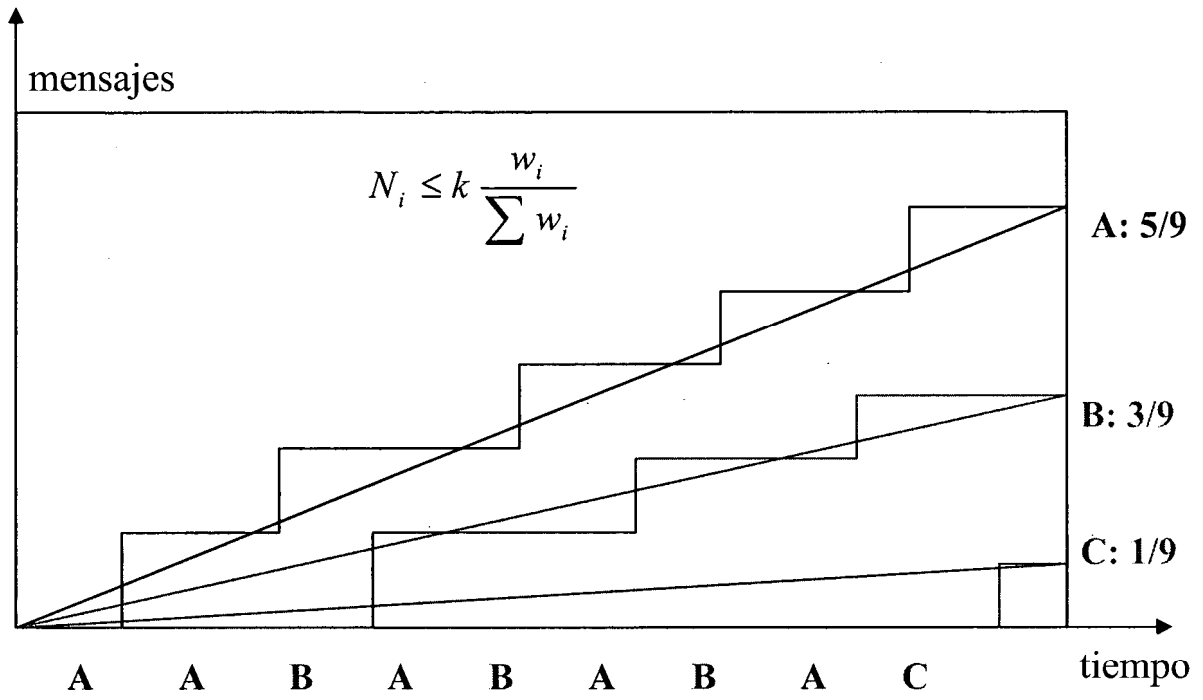


Fig. 3