

Solución al problema de encaminamiento dinámico en redes ATM para tráfico constante CBR mediante un algoritmo de búsqueda tabú

Joaquín R. Fernández V.¹, Pablo Cortés¹, José Guadix¹

¹ Dpto. de Organización de Empresas. Escuela Técnica Superior de Ingeniería Industrial de Sevilla

Keywords: búsqueda tabú, encaminamiento dinámico, red ATM.

1. Introducción

La tecnología de banda ancha ATM (modo de transferencia asíncrono) basada en conmutación de paquetes de longitud fija (células ATM) agrupa las ventajas de la conmutación de circuitos y paquetes en una sola, proporcionando una velocidad de transmisión de información muy alta y versatilidad a la hora de implementar nuevas aplicaciones. Ofrece cuatro tipos de servicio diferentes:

- Servicio de tasa de bit constante (CBR).
- Servicio de tasa de bit variable (VBR).
- Servicio de tasa de bit disponible (ABR).
- Servicio de tasa de bit no especificada (UBR).

De todos ellos CBR es el más restrictivo de todos los tipos de tráfico porque está diseñado para aplicaciones en tiempo real (audio o video).

Mediante las pérdidas de células a través de la red y el retraso sufrido por las que consiguen llegar al destino se comprueba si la información recibida cumple con el grado de servicio (GoS) contratado. La mayoría de estas pérdidas y los retrasos se producen en las colas de los conmutadores ATM porque el medio de transmisión compuesto por haces de fibra óptica proporciona unos retrasos y unas pérdidas prácticamente nulos.

Al ser las pérdidas en el medio casi inexistentes, los mecanismos de corrección frente a pérdidas como el control de flujo por ventana entre nodos adyacentes y la retransmisión de células que dan lugar a un tráfico de red demasiado intenso se han podido eliminar en las redes ATM para aumentar la velocidad de transmisión.

2. Objetivo

Se pretende optimizar el encaminamiento dinámico para una red ATM con una topología dada y tráfico constante CBR mediante un algoritmo de búsqueda tabú, de forma que la aplicación del algoritmo permita elegir en las tablas de encaminamiento el camino por el que se deben establecer las conexiones atendiendo a las pérdidas de células producidas en la red pero satisfaciendo siempre una condición mínima sobre el retraso que asegure el grado de servicio (GoS). La importancia de un encaminamiento eficiente es doble:

- Si la distribución del flujo no se realiza correctamente algunos enlaces pueden sobrecargarse y otros infrautilizarse dando lugar a transmisiones lentas y defectuosas.

- Si se minimizan las pérdidas la comunicación ocupa los recursos de la red durante menos tiempo y su rendimiento aumenta.

3. Justificación de la optimización de las pérdidas y la acotación del retraso

Dada la importancia del retraso en ATM se podría optar por una función objetivo que lo minimizase como se propone en Amiri et al (1998). Sin embargo las redes ATM garantizan los servicios para todas las conexiones durante todo el horizonte temporal, negociando con el usuario para cada intento de conexión los parámetros del servicio de manera que cuando no se pueden asegurar la conexión no se lleva a cabo. Así, se deduce que mientras las pérdidas sean pequeñas el retraso se encontrará dentro de los límites establecidos pues la red garantizó al inicio de la conexión la calidad en el servicio y de no ser posible cumplirlo la hubiese rechazado. Pero si por el contrario el número de células perdidas es significativo, las células que llegan lo hacen con retardos pequeños pero no toda la información que requiere el servicio alcanza el destino (aún cuando la red sea sumamente rápida).

Por todo ello el retraso sufrido por una célula al atravesar la red se impone como restricción de diseño que supuestamente se negoció al inicio de la conexión y no como parte de la función objetivo del modelo. Además, debido a que la naturaleza del retraso es la misma que la de las pérdidas y depende en exclusiva del número de conmutadores que atraviesan los datos que a su vez depende del encaminamiento empleado, el modelo planteado para minimizar las pérdidas conlleva de forma implícita e indirecta la optimización de los retardos sin la necesidad de incluir una restricción específica al respecto.

4. Metodología

Una red ATM se puede representar mediante un grafo $G = (N, E)$ donde E es el conjunto de enlaces comunicación y N el conjunto de nodos (conmutadores) cada uno de ellos compuesto por un conjunto de búferes con capacidad finita para almacenar células y un mecanismo de interconexión formado por todas las conexiones internas que existen en el conmutador para que cualquier célula que entre en el sistema pueda ser conmutada a su salida correspondiente.

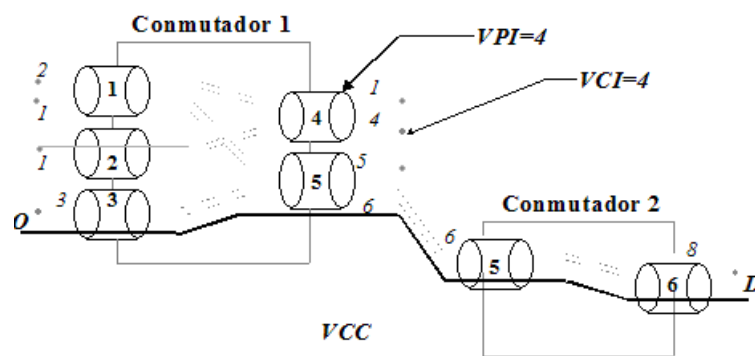


Figura 1. Concatenación de caminos virtuales

Para caracterizar las pérdidas en forma de modelo matemático, cada buffer se representa como una cadena de Markov, tomando como base la probabilidad de pérdida de célula que se producen en el interior de los conmutadores.

4.1. Escalas de tiempo

Medova (1998) define tres niveles temporales distintos para redes de banda ancha: nivel de conexión, nivel de ráfaga y nivel de célula. En el modelo propuesto se trabaja sólo con las escalas de tiempo en las que sea posible identificar la ocurrencia de algún proceso estocástico

relevante. Éstas son el **nivel de conexión** al cuál van referidas las llamadas (periodos de tiempo en los que permanece invariable el camino) y el **nivel de célula** en el que se mide el flujo y las pérdidas de los conmutadores.

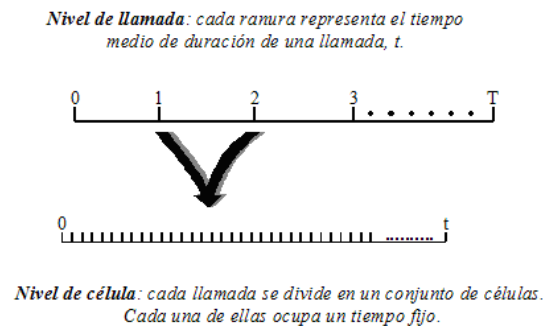


Figura 2. Diferentes escalas de tiempo.

5. Componentes de la red

El modelo propuesto considera tres elementos necesarios y suficientes para caracterizar toda red ATM:

5.1. Enlaces

Sirven de nexo entre nodos adyacentes. La capacidad de transmisión típica de un enlace STM-1 es de 155,52 Mb/s.

5.2. Conmutadores

Los conmutadores en ATM son sistemas que funcionan mediante divisiones de la transmisión en intervalos temporales denominados “ranuras” (slots) de duración el tiempo de transmisión de una célula (aprox. $2,8\mu\text{seg.}$ a la velocidad de transmisión 155,52 Mb/s). De la amplia gama de conmutadores existentes se toman conmutadores con colas a la salida por ser los más utilizados.

En una red ATM existe sincronización, de forma que las células se sirven al inicio de un tiempo de ranura y abandonan el buffer siempre al final de otra ranura. Por ello, una célula que llegue durante la ranura n tendrá que esperar como mínimo a la $n+1$ para ser servida.

Así, al abandonarse el buffer siempre al final del tiempo de célula la forma de gestión del buffer es de “llegadas primero”, siendo el instante de llegada de una célula a una ranura irrelevante pero considerando que ésta se produce siempre antes de la partida de la célula que este en la cabecera del buffer.

5.3. Buffer

Un buffer se puede caracterizar como un sistema de colas descrito por un patrón de servicio, el número de servidores, la capacidad del sistema, la disciplina de cola y el patrón de llegadas:

- **El patrón de servicio:** se describe mediante el tiempo requerido para servir a un cliente s (tiempo de célula) o por su inversa la tasa de servicio μ . En ATM siempre se transmite aunque el servidor se encuentre desocupado porque el flujo debe ser siempre continuo y los servidores sincronizados y deterministas.

- **El número de servidores:** determina cuantos canales hay en servicio o cuantos clientes pueden ser servidos simultáneamente. En la redes ATM las colas son exclusivas de modo que cada servidor se encarga de ofrecer servicio sólo a los clientes que tiene en su cola particular.

- **La capacidad del sistema:** es la capacidad de las colas y de los canales de servicio.
- **El patrón de llegadas de clientes al sistema:** se caracteriza por el número medio de llegadas por unidad de tiempo ($\bar{\epsilon}$) o por el tiempo medio entre llegadas de los clientes. En ATM las células llegan en lotes por lo que el número de células que llega en cada ranura de tiempo puede variar.

Atendiendo a la notación Kendall el sistema se describe como M/D/1/K donde las colas son FIFO y el tiempo entre llegadas viene dado por una distribución exponencial negativa (Pitts et al, 1996).

5.3.1. Llegadas por lotes

Considerando el tiempo discreto el patrón de llegadas forma un proceso de Bernoulli en el que la probabilidad de que haya una llegada en una ranura es p y la probabilidad de que la ranura esté vacía $1-p$. Para un sistema de N ranuras la distribución de las llegadas se define por una Binomial donde la probabilidad de que se den k llegadas en N ranuras es:

$$P(k) = \frac{N!}{(N-k)!k!} (1-p)^{N-k} p^k \quad (1)$$

Pero si se tiene en cuenta que las llegadas son en lotes y se define por p no la llegada de una célula sino la de un grupo de varias (a priori se desconoce cuantas) y se define por M el número de puertos a la entrada del conmutador, el número de células que pueden recibirse en un lote variará entre 0 y M y el número de células por ranura $a(k)$ se puede describir como:

$$a(k) = \frac{M!}{(M-k)!k!} (1-p)^{M-k} p^k \quad (2)$$

La tasa total de llegadas es de Mp células por ranura de tiempo y si M aumenta y p disminuye la distribución Binomial tiende a la de Poisson.

6. Ecuaciones de balance

Cada puerto de salida de los conmutadores se comporta como un servidor con una cola de tamaño b que se puede considerar como una cadena de Markov de b estados descritos de la forma $s(j)$ donde j es el número de células en cola. Las transiciones posibles hacia un estado i cualquiera se muestran en la figura 3:

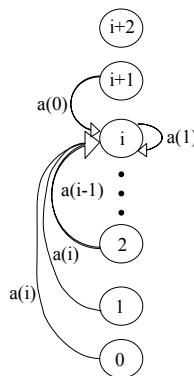


Figura 3. Diagrama de estado del sistema

Como las llegadas de las células son por lotes dado un estado j todos los estados que le suceden son accesibles, sin embargo como el sistema sólo es capaz de servir una célula por cada ranura de tiempo, dado un estado j la única transición posible hacia un estado precedente sería hacia el inmediatamente anterior y sólo en caso de que no llegue ninguna célula nueva.

7. Modelo del problema

En las subsecciones siguientes se describen los parámetros, variables, datos de partida, restricciones y función objetivo del modelo:

7.6.1. Parámetros del modelo

- **N**: Conjunto de nodos de la red.
- **E**: Conjunto de enlaces de la red.
- **M**: Conjunto de pares origen-destino que pueden establecer la conexión.
- **H(m)**: Conjunto de todos los caminos posibles que pueden establecer una conexión entre un par de nodos origen-destino.
- **T**: Horizonte global temporal a nivel de conexión.
- $t \in T$: Duración de una llamada. Periodos de tiempo en los que se divide una conexión y en los que permanece invariable el camino de transmisión.
- $\tau \in t$: Tiempo de célula. Periodo de tiempo en el que se subdivide cada llamada. Las pérdidas son referidas a este nivel.
- **N(h)**: Conjunto de nodos que forman parte del camino h
- **E(h)**: Conjunto de enlaces del camino h
- **A(i)**: Conjunto de nodos antecesores al nodo i .
- **D(i)**: Conjunto de nodos sucesores al nodo i .

7.6.2. Variables del modelo

- $P_h^{m,t}$: Variable binaria que indica si la conexión entre el par m (O-D) se realiza a través del camino h durante el intervalo de conexión $t \in T$.
- $X_{h,ij}^{m,\tau}$: Flujo asociado al par m que discurre a través el enlace (i,j) del camino h en dirección de i a j durante el intervalo $\tau \in t$ de célula. Se expresa en células por ranura y representa el flujo real de células.
- $I_{h,ij}^{m,\tau}$: Pérdida de células que se produce en el enlace (i,j) del camino h debido al nodo i durante el intervalo $\tau \in t$ de célula. Se expresa en células por ranura.
- F_{ij}^τ : Flujo total de células que en teoría debería atravesar el enlace i - j desde el nodo i hacia el nodo j durante el intervalo $\tau \in t$ de célula. Se expresa en células por ranura. Y representa el flujo ideal de células.

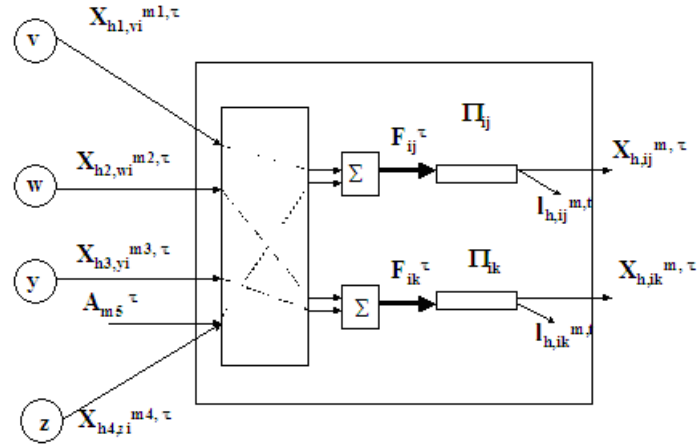


Figura 4. Relación entre las variables del modelo

7.6.3. Datos de partida:

- Q_{ij} : Capacidad del enlace (i,j) expresada en células por ranura de tiempo.
- A_m^τ : Demanda del par m en el intervalo $\tau \in t$ expresada en células/ranura.
- $GOS|_{cel}^m$: Grado de servicio, máximo número permitido de células perdidas para la conexión m medido en células/ranura.

7.6.4. Restricciones:

$$1) \quad \sum_{h \in H(m)} P_h^{m,t} = 1 \quad \forall m \in M, \forall t \in T$$

Si cada conexión se subdivide en varias llamadas, cada llamada entre el par de nodos $O(m)$ - $D(m)$ se debe realizar a través de sólo un camino de entre todos los disponibles.

$$2) \quad X_{h,ij}^{m,\tau} + l_{h,ij}^{m,\tau} = \begin{cases} A_m^\tau P_h^{m,t} & \text{si } i = O(m) \\ X_{h,ki}^{m,\tau-1} & \text{si } i \neq O(m) \end{cases}$$

$$\forall \tau \subset t, \forall t \in T, \forall h \in H(m), \forall m \in M, \forall i \in N: \forall (i,j) \in E(h), \forall (k,i) \in E(h), /i,j,k \in N(h)$$

Balace de flujo para todo nodo de la red: el flujo a través del enlace (i,j) del camino h que une el par m más las pérdidas de células que se producen en el buffer de salida del nodo i debe ser igual a todo el tráfico creado en el origen que se ha enviado por el camino h caso de ser el nodo i el origen de la comunicación, o igual a todo el flujo que llega al nodo i procedente del nodo k caso de no ser éste origen de la comunicación (el trafico proveniente de los nodos anteriores llega retrasado debido a los conmutadores).

$$3) \quad F_{ij}^\tau = \sum_{m \in M} \sum_{\substack{h \in H(m), \\ (q,i) \in E(h): j \in N(h)}} \sum_{q \in A(i)} X_{h,qi}^{m,\tau-1} + \sum_{m \in M / i=O(m)} A_m^\tau \quad \forall (i,j) \in E, \forall \tau \subset t, \forall t \in T$$

El flujo total que atraviesa el enlace i-j comprende los flujos procedentes de aquellos nodos predecesores al i que deban ser encaminados hacia el j para poder alcanzar sus destinos así como los generados en ese mismo nodo caso de ser origen de una conexión. Se considera de nuevo el retraso producido por el conmutador anterior en el flujo entrante de la red.

$$4) \sum_{m \in M} \sum_{h \in H(m)} l_{h,ij}^{m,\tau} \geq L_{ij}^{\tau} (F_{ij}^{\tau}) \quad \forall i \in N, \forall (i, j) \in E, \forall \tau \subset t, \forall t \in T$$

Las variables $l_{h,ij}^{m,\tau}$ reflejan al menos las pérdidas debidas al desbordamiento de las colas de los puertos de salida pues las pérdidas reales contabilizadas en cada puerto de salida deben ser mayor o igual que las pérdidas estimadas para ese puerto por la función de pérdidas de los conmutadores. Aunque se consideren otras posibles causas de pérdidas de células L_{ij}^{τ} en las redes ATM, éstas no son ni con mucho tan importantes como las pérdidas debidas al desbordamiento de los buffers.

La función L está relacionada con la estructura interna de los conmutadores. Depende del tamaño del buffer y de las colas.

$$5) \sum_{m \in M} \sum_{h \in H(m): (i,j) \in E(h)} X_{h,ij}^{m,\tau} \leq Q_{ij} \quad \forall (i, j) \in E, \forall \tau \subset t, \forall t \in T$$

El flujo total que atraviesa un enlace (i,j) no puede superar su capacidad.

$$6) \sum_{\tau \subset t} \sum_{h \in H(m)} \sum_{(i,j) \in E(h): j \in D(i)} l_{h,ij}^{m,\tau} \leq GOS_{cel}^m \quad \forall m \in M, \forall t \in T$$

Condición restrictiva sobre las pérdidas: el grado de servicio contratado por una conexión es el límite máximo de células que se pueden perder en todo el camino que recorre el flujo de una conexión, de forma que para cada par m se debe cumplir que el conjunto de pérdidas que se producen a lo largo de todo el trayecto h no supera la cota impuesta.

$$7) l_{h,ij}^{m,\tau} \geq 0 \quad \forall (i, j) \in E, \forall m \in M, \forall \tau \subset t, \forall t \in T$$

$$8) X_{h,ij}^{m,\tau} \geq 0 \quad \forall (i, j) \in E, \forall m \in M, \forall \tau \subset t, \forall t \in T$$

$$9) P_h^{m,t} \in (0,1) \quad \forall m \in M, \forall h \in H(m), \forall t \in T$$

7.6.5. Función Objetivo:

El objetivo es lograr un encaminamiento óptimo minimizando las pérdidas totales en todos los enlaces y para todo el horizonte. Luego atendiendo a todas las conexiones que soporta la red:

$$FO: \quad MIN \sum_{t \in T} \sum_{\tau \subset t} \sum_{(i,j) \in E} \sum_{m \in M} \sum_{h \in H(m)} l_{h,ij}^{m,t}$$

8. Algoritmo de búsqueda tabú

Dada la imposibilidad de optimizar conjuntamente pérdidas y retraso, se diseña un algoritmo para minimizar las pérdidas en la red asegurando un grado de servicio mínimo. Así, la aplicación del algoritmo decide como se establecen las conexiones a lo largo de la red. Del tamaño de las redes reales y la consecuente dimensión del espacio de soluciones posibles surge la necesidad de emplear un procedimiento metaheurístico que obtenga una solución de calidad en un tiempo razonable. El algoritmo de búsqueda tabú explora el espacio de soluciones tomando como base la suposición de que una mala elección estratégica puede producir más información que una buena elección al azar. Así, las soluciones poseen distintas características (atributos) que bajo ciertas condiciones pueden declararse "tabú", siendo

obviadas entonces por el proceso de búsqueda que de esta manera posee la facultad de dirigir la exploración.

Inicialmente se calcula la solución de rutas mínimas para comenzar el proceso de búsqueda tabú desde una “región” del espacio que se presupone repleta de soluciones de calidad. Dicha solución encamina todas las conexiones de manera que las células atraviesan el número menor posible de nodos durante su recorrido desde el nodo origen hasta el nodo destino. Después se ejecuta de forma iterativa una exploración del espacio basada en una memoria a corto plazo y otra a largo plazo, en la que se lleva a cabo una búsqueda mediante “movimientos” de una solución a otra y la indagación en sus entornos respectivos. Según esto, se constituye una “cadena de soluciones” formada por soluciones encadenadas adyacentes. La memoria a corto plazo se encarga de encadenar estas soluciones de manera óptima en el espacio, conformando así una trayectoria consecuencia de la calidad de las soluciones y del filtro tabú que conlleva una exploración inteligente. Su funcionamiento se basa en la creación y examen de calidad de una región de vecindad (soluciones similares) constituida entorno a una solución generadora que forma parte de la “cadena” donde sólo las soluciones que no posean ninguna ruta tabú pueden participar en el proceso.

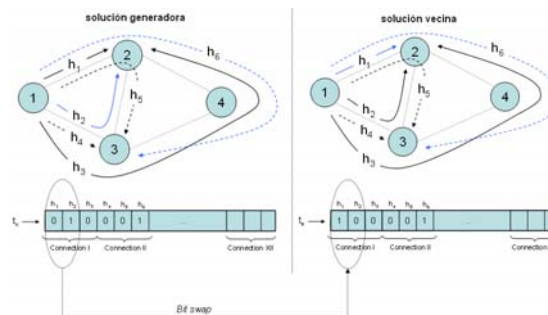


Figura 5. Solución generadora (Izq.) y solución vecina (Der.)

En cambio la memoria a largo plazo cuestiona periódicamente si la trayectoria seguida por el “camino” es aceptable o si por el contrario sería más conveniente retroceder a lo largo de esta “cadena” para redirigir la exploración en otro sentido hacia otras regiones del espacio.



Figura 6. Exploración del espacio de soluciones

8.1. Mecanismo tabú

Las rutas que forman parte de la resolución del problema, cuando se encuentran presentes de forma consecutiva y repetitiva en las soluciones, son marcadas como tabú o prohibidas, de forma que las siguientes soluciones visitadas no puedan contenerlas y la exploración no se vicie. El conjunto de rutas que en un determinado momento son designadas como tabú conforma un grupo dinámico a lo largo del proceso. La frecuencia tabú fija el número de

veces que se puede repetir en una solución un atributo determinado (ruta) antes de que le sea adjudicado el estatus de tabú, mientras que la tenencia tabú de un atributo indica el tiempo (medido en número de iteraciones) que permanecerá activo como tabú.

9. Análisis y comparación con un algoritmo genético

Para evaluar la calidad del método propuesto basado en búsqueda tabú, se simularon distintas condiciones de tráfico CBR en un conjunto de redes cuyo tamaño va desde 10 a 35 nodos y se compararon los resultados obtenidos con los de un algoritmo genético que también resolvía el mismo problema. En las figuras 7 y 8 se muestran los resultados obtenidos en función de la calidad de la solución obtenida y el tiempo de cómputo necesitado para alcanzarla:

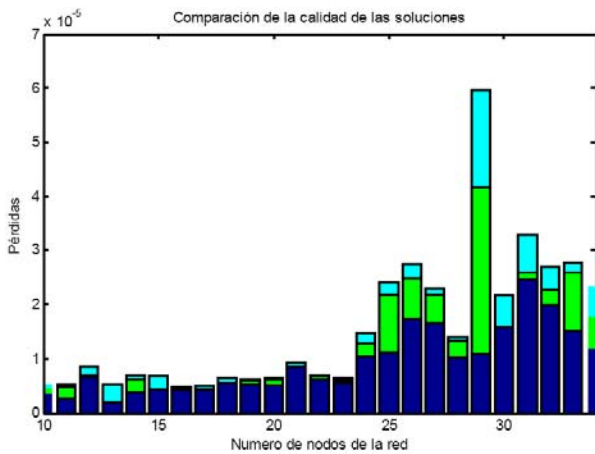


Figura 7. Comparación de la calidad de las soluciones Alg. Ruta Mín. (Celeste), Alg. G (Verde) y B. Tabú (azul)

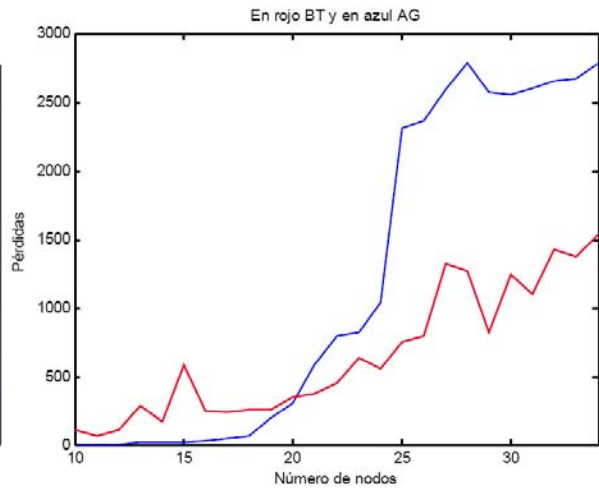


Figura 8. Comparación del tiempo de cómputo

10. Conclusiones

Sin cambiar de tecnología o de topología es posible reducir las pérdidas de la red mediante un correcto encaminamiento dinámico de las comunicaciones que optimice al máximo la distribución de los recursos. Lógicamente a medida que el número de nodos de una red (conmutadores) es mayor también lo es el tiempo necesario para computar todos los cálculos. Para problemas con cargas de tráfico similares también aumentan las pérdidas conforme las redes tienen más nodos debido a que el número de procesos de conmutación aumenta y aunque también aumenten los enlaces disponibles entre nodos que alivian la carga mediante el reparto del flujo, estos no compensan el aumento del número de nodos a la hora de minimizar las pérdidas.

Para redes de un tamaño superior a los veinte nodos (aprox.) la aplicación de la búsqueda tabú a la resolución del problema ofrece una mejora muy significativa respecto del algoritmo genético tanto en cuestión de calidad de la solución final como en cuestión de tiempo. Para redes pequeñas de diez o quince nodos existen diferencias aunque no tan significativas como para redes mayores. La única ventaja del algoritmo genético es que suele ofrecer mejor tiempo de ejecución del programa si la red es pequeña aunque ofreciendo una solución de calidad menor.

Referencias

- J. Xu; S.Y. Chiu; F. Glover (1997). Tabu search for dynamic routing communications network design, *Telecommunication Systems* 8 (1) 55-77.
- P. Cortés; J. Muñuzuri; J. Larrañeta; L. Onieva (1990). A genetic algorithm based on cell loss for dynamic routing in ATM networks, in: *Soft Computing in Industry*, Berlin, pp. 627-640.
- Pitts, J.M., Schormans, J.A. (1996): "Introduction to ATM design and performance", Wiley.
- Medova (1998): "Chance-constrained stochastic programming".The Institute of Management.