

ASIGNACION DE TRANSPORTE PUBLICO CON RESTRICCION DE CAPACIDAD: IMPLEMENTACION Y RESULTADOS

G. Crovetto R., J. de Cea Ch. y J.E. Fernández L.

Pontificia Universidad Católica de Chile

Casilla 6177, Cod. 105, Santiago 22, Chile

RESUMEN

Una de las limitaciones más importantes de los modelos operacionales de asignación a redes de transporte público es la no consideración de las restricciones de capacidad de los vehículos.

En este trabajo, se presenta brevemente un modelo propuesto por De Cea y Fernández, que incorpora tales restricciones en los costos de espera de los usuarios, y un algoritmo de diagonalización para resolver el problema. Asimismo, se describe más extensamente los resultados obtenidos por el modelo en redes de la ciudad de Santiago. Se expone la convergencia del algoritmo, el tiempo de ejecución y memoria requeridos, y las características de los flujos obtenidos en las líneas de buses, comparándolos con una asignación a rutas mínimas.

1 Introducción

El proceso que siguen los modelos de asignación a redes de Transporte Público es conceptualmente muy simple. Primero se definen las alternativas que tiene un usuario para realizar su viaje, de acuerdo a los conceptos de itinerario, ruta o estrategia. Luego, el modelo determina para cada persona o grupo de personas cual es la que minimiza la función de costo generalizado, y asigna estos viajes a la línea o líneas de transporte público involucradas. Una vez que se ha realizado este proceso para todos los individuos, se obtienen los viajes sobre la red.

Sin embargo, todos estos métodos hacen uso de un supuesto en forma implícita, el cual establece que una persona sólo debe esperar al primer vehículo que pase y que pertenezca a una línea dentro de un conjunto de posibilidades. Este supuesto implica ignorar que las líneas tienen una capacidad limitada para movilizar personas en un período. Esto es razonable en países desarrollados, en que la oferta de servicios de transporte es muy buena, pero en países en vías de desarrollo, en que es fácil ver a los vehículos con pasajeros incluso en sus pisaderas, no puede asumirse que en todos los casos bastará esperar al primer bus.

A las redes de transporte público que presentan estos casos, se les denomina "redes congestionadas", puesto que existen muchas personas que desean el mismo servicio. Asimismo, a los modelos recién comentados se les llama "modelos sin congestión" o "modelos sin restricción de capacidad".

Si bien en la literatura han aparecido algunos modelos que consideran los efectos de las restricciones de capacidad (ver por ejemplo Last y Leak 1976, Spiess 1983, Nguyen y Pallotino 1988, etc), ninguno de ellos parece ser adecuado, por cuanto o no tienen una fácil implementación o no consideran el efecto más importante, como lo es el aumento del tiempo de espera.

En este trabajo se presenta la implementación y los resultados del modelo de De Cea y Fernández (1988), el cual considera las demoras sufridas por los usuarios debido a la congestión de pasajeros en los vehículos de locomoción colectiva.

El trabajo se divide en 3 capítulos, siendo el primero de ellos esta breve introducción. En el segundo se presenta brevemente la formulación del modelo de De Cea y Fernández, y su algoritmo de solución. Finalmente, en el capítulo tres se exponen los resultados obtenidos y las principales conclusiones.

2 Formulación Matemática y Algoritmo de Solución

2.1 Definiciones Básicas y Notación

Con el objeto de permitir una mejor comprensión del problema a tratar, se definen a continuación los conceptos de línea, sección de línea, ruta y sección de ruta.

Se denomina línea de transporte público, o simplemente línea, a un servicio de vehículos de una determinada capacidad y frecuencia con un itinerario fijo definido por una secuencia de nodos de la red. Una sección de línea es un tramo de ella comprendido entre dos nodos no necesariamente consecutivos.

Una ruta de transporte público define la forma de realizar un viaje entre un nodo de origen y un nodo de destino. A ella se asocia una secuencia de nodos, en la que se incluyen los nodos de origen y destino del viaje y los nodos de trasbordo (si los hay). Así, esta secuencia define el nodo inicial y final de cada uno de los tramos que conforman la ruta. Cada uno de estos tramos corresponde a una sección de ruta y tiene asociado el conjunto de líneas percibidas como equivalentes por los viajeros para desplazarse en él.

En la figura 2.1 se presenta una red codificada en términos de líneas de transporte público, mientras que la figura 2.2 representa la misma red codificada en términos de secciones de ruta. Ambas codificaciones representan la misma situación. A la primera de ellas se le denomina "Red Real", y a la segunda se le llama "Red Auxiliar".

A continuación se presenta la notación a utilizar.

W	: Conjunto de pares Origen-Destino de la Red
w	: Un elemento de W
P	: Conjunto de rutas para usuarios de Transporte Público
P_w	: Conjunto de rutas para el par Origen-Destino w
p	: Un elemento de P
T	: Conjunto de demandas de viaje
T_w	: Demanda entre el par w
V_s	: Flujo de pasajeros en la sección de ruta s
V	: Vector de flujos de pasajeros en las secciones de ruta
v_s^i	: Flujo de pasajeros en la sección de ruta s , línea i
v	: Vector de flujos de pasajeros en las secciones de línea
h_p	: Flujo de pasajeros en la ruta p
c_s	: Costo de viaje para la sección de ruta s
\bar{c}_s	: Tiempo de viaje en vehículo para la sección de ruta s
δ_{sp}	: Elemento de la matriz de incidencia sección de ruta—ruta: 1 si p pasa por s , 0 si no
f_l	: Frecuencia de la línea l
k_l	: Capacidad de los vehículos de la línea l
f_s	: Frecuencia total en la sección de ruta s
K_s	: Capacidad práctica en la sección de ruta s

2.2 Funciones de Costo

El elemento fundamental de esta formulación lo constituye la naturaleza de las funciones de costo asociadas a cada sección de ruta. En redes de transporte público "congestionadas", como ocurre normalmente en sistemas urbanos de países en desarrollo, la sobrecarga de los servicios se traduce en aumentos importantes de los tiempos de espera de los pasajeros. Por este motivo, en tal situación parece

recomendable suponer que, en general, el tiempo de espera de un viaje sea una función creciente de los flujos en la red.

En general, se tendrá que el tiempo de espera experimentado por un pasajero que usa la sección de ruta s (en su nodo de origen i) dependerá de los siguientes términos:

- V_s : el número de pasajeros que abordan la misma sección de ruta s
- $V_{i,s}^+$: el número total de pasajeros que abordan en i otras secciones de ruta, las cuales tengan líneas que estén en s
- $\bar{V}_{i,s}$: el número de pasajeros que abordaron a las líneas de s antes del nodo i y que se bajarán después de i
- K_s : la capacidad de la sección de ruta s , calculada como la suma de las capacidades de todas sus líneas

El flujo $V_{i,s}^+$ compite con V_s por la capacidad residual de las líneas de la sección s en el nodo i , y el flujo $\bar{V}_{i,s}$ reduce tal capacidad. Así, si se llama $S_{i,s}^+$ al conjunto de secciones que salen del nodo i con excepción de s , y $\bar{S}_{i,s}$ al conjunto de secciones que comienzan antes de i y terminan después de i , puede definirse:

$$\tilde{V}_s = V_{i,s}^+ + \bar{V}_{i,s} = \sum_{l \in s} \sum_{r \in S_{i,s}^+} v_l^r + \sum_{r \in \bar{S}_{i,s}} v_l^r \quad (2.1)$$

donde \tilde{V}_s representa el flujo que compite con V_s por la misma capacidad común. Entonces, las funciones de costo de las secciones de rutas tendrán la siguiente forma general:

$$c_s = \bar{c}_s + \frac{\alpha}{f_s} + PT + \varphi \left(\frac{V_s + \tilde{V}_s}{K_s} \right) \quad (2.2)$$

Una posibilidad para la función φ es la forma de potencia usada en las funciones del BPR:

$$\varphi = \beta \left(\frac{V_s + \tilde{V}_s}{K_s} \right)^n \quad (2.3)$$

2.3 Restricciones Generales

Cualquier conjunto de flujos factible V en $G(N, S)$ debe satisfacer las siguientes restricciones de conservación, continuidad en la red y no negatividad de los flujos:

$$\sum_{p \in P_w} h_p = T_w \quad \forall w \in W \quad (2.4)$$

$$\sum_{p \in P} h_p \delta_{sp} = V_s \quad \forall s \in S \quad (2.5)$$

$$V_s \geq 0 \quad \forall s \in S \quad (2.6)$$

Asimismo, se asume que el flujo de una sección de ruta se reparte entre las líneas que la componen en forma proporcional a sus frecuencias, esto es:

$$v_s^i = \frac{f_i}{\sum_{k \in s} f_k} V_s \quad (2.7)$$

Se denominará Ω al conjunto de flujos que cumplen con las restricciones anteriores. De acuerdo al primer principio de Wardrop (1977), un flujo factible se llamará de equilibrio si satisface la siguiente restricción sobre la red $G(N, S)$:

$$C(V^*)(V^* - V) \leq 0, \forall V \in \Omega \quad (2.8)$$

2.4 Método de Solución

El enfoque de solución implementado en este caso es el propuesto por De Cea y Fernández y se basa en la utilización del conocido método de "Diagonalización" (ver Florian y Spikss 1982, Abdulaal et Le Blanc 1979, etc), que es un procedimiento

iterativo del tipo Jacobi, usado para resolver ecuaciones no lineales, ver Pang and Chang (1982). Dentro de cada iteración de la diagonalización debe resolverse un problema de programación no lineal con funciones de costo separables, para lo cual se usa el método de Frank-Wolfe.

En breve, el método consiste en producir, en cada iteración, un vector de funciones de costo "diagonalizado", (la función de costo asociada a una sección de ruta dada depende sólo de su flujo) con lo que el problema se transforma en un problema simétrico de equilibrio de tráfico. Si se denota por $\hat{C}(V)$ el vector diagonalizado de funciones de costo, en cada iteración se resuelve un problema del tipo:

$$\hat{C}(V^*)(V^* - V) \leq 0, \forall V \in \Omega \quad (2.9)$$

Dado que en $\hat{C}(V)$ cada función de costo $\hat{c}(V)$ depende sólo de su flujo propio V_s , el jacobiano obtenido es diagonal, y por tanto simétrico. Bajo estas condiciones existe un problema de optimización equivalente, el cual se presenta en la formulación (2.10).

$$\begin{aligned} \text{Min} \sum_{s \in S} \int_0^{V_s} \hat{c}(x) dx \\ \text{s.a. 2.4, 2.5, 2.6, 2.7} \end{aligned} \quad (2.10)$$

El problema (2.10) es resuelto mediante el conocido método de Frank-Wolfe.

El algoritmo completo de solución del problema puede resumirse como sigue:

Etapa 0 : (Inicialización) Encontrar una solución factible (\bar{V}, \bar{v})

Etapa 1 : Diagonalizar $C(V)$ en la solución (\bar{V}, \bar{v})

Etapa 2 : Resolver (2.10) para obtener (\hat{V}, \hat{v})

Etapa 3 : Test de Parada. Si (\bar{V}, \bar{v}) y (\hat{V}, \hat{v}) son suficientemente cercanos, parar; si no

Etapa 4 : Hacer $(\bar{V}, \bar{v}) = (\hat{V}, \hat{v})$ y volver a la etapa 1

3 Resultados

El algoritmo descrito en el capítulo anterior fue implementado en el lenguaje FORTRAN-77 en un computador MICROVAX-II, bajo sistema operativo VMS. Fue probado con dos redes de transporte público correspondientes a la ciudad de Santiago, para los años 1977 y 1986, con las cuales se ha trabajado en investigaciones anteriores.

La red de 1977 consta de 245 nodos y 444 líneas unidireccionales que generan 75103 secciones de línea, agrupadas en 19.078 secciones de ruta. Por su parte, la red de 1986 tiene 785 nodos y 606 líneas unidireccionales, lo que produce 521.104 secciones de línea y 167.399 secciones de ruta.

Las funciones de costo utilizadas para estas pruebas fueron del tipo siguiente:

$$c_s = \bar{c}_s + \frac{\alpha}{f_s} + PT + \beta \left(\frac{V_s + \tilde{V}_s}{K_s} \right)^n \quad (3.1)$$

En donde \tilde{V}_s es el flujo que compite con V_s por la capacidad K_s , PT es la penalidad de trasbordo (se asumió un valor de 45 minutos), y α , β , n son parámetros a calibrar. Se consideró un valor de 0.5 para α ; para β se tomaron dos valores: 10 y 60; y n se hizo variar entre 2 y 5.

3.1 Convergencia del Algoritmo

Si se denomina V^k al vector de flujos en secciones de ruta obtenido en la diagonalización k , resulta natural definir el siguiente error relativo:

$$\epsilon_1 = \frac{\|V^{k+1} - V^k\|}{\|V^k\|} \quad (3.2)$$

El algoritmo presentado, con algunas ligeras variaciones, resultó ser convergente en gran parte de los casos. Sin embargo, debe destacarse que cuando

se hizo pruebas con redes muy sobrecargadas, por ejemplo al doble del flujo normal, el algoritmo no convergió.

En los casos en que el algoritmo determinó la solución de equilibrio, se requirieron entre 6 y 12 diagonalizaciones para llevar este criterio a un valor menor que el 5 por ciento, dependiendo de las iteraciones que se efectuaban en el método de Frank-Wolfe y de los valores de β y n . El criterio tuvo en general un comportamiento monótono decreciente, excepto en algunos casos en que se exigió sólo una iteración de Frank-Wolfe.

Pudo comprobarse asimismo que basta realizar dos de estas iteraciones, por cuanto la evolución del criterio de parada no se ve afectada si se realizan más iteraciones. Fue descartado realizar sólo una tanto por el comportamiento del criterio de error, como porque los resultados a los que se llega eran ligeramente diferentes a los que se llega con dos o más iteraciones de Frank-Wolfe.

3.2 Memoria Requerida y Tiempos de Ejecución

La versión del algoritmo con que se trabajó hace un uso intensivo de memoria, ya que mantiene toda la información relevante en las estructuras de datos para representar la red real y la auxiliar. La sola representación de la información, principalmente secciones de línea y de ruta, ocupa 2.5 megabytes para la red de Santiago de 1977, y 17.8 megabytes para la de 1986. Estos niveles de uso de memoria eran imposibles de manejar hace algunos años; sin embargo, hoy en día pueden utilizarse gracias a los avances tecnológicos en los dispositivos de memoria, y a las técnicas de memoria virtual difundidas actualmente.

El tiempo de ejecución del algoritmo depende de los parámetros n y β y del número de iteraciones de Frank-Wolfe exigidas en cada diagonalización. Para la red de 1977, este dato fluctuó entre 1190 y 1377 segundos para el computador MICROVAX-II.

Cabe destacar al respecto que EMME/2, un paquete comercial de asignación a redes de transporte público que utiliza el concepto de estrategia mínima, demora para la misma red y computador 263 segundos (ver De Cea et al 1987), solucionando el problema sin restricción de capacidad.

3.3 Calidad de los Resultados

Para poder analizar el conjunto de flujos en líneas y segmentos de línea, debe establecerse una medida común de carga. Por esto, se definirá la "carga porcentual en un segmento de línea" como la tasa de ocupación, según la siguiente expresión:

$$\frac{v_a}{k_l} \quad \forall a \in l, \forall l \in L \quad (5.15)$$

Se considerarán los siguientes parámetros:

- El porcentaje de segmentos sobrecargados, esto es, que poseen un flujo mayor que la capacidad.
- El porcentaje de líneas que tienen algún segmento sobrecargado. Para estos efectos se considerarán independientes los recorridos de ida y de vuelta.
- La carga media porcentual de los segmentos de línea, definida según la expresión:

$$\frac{1}{\text{total de segmentos}} \sum_{a \in l \in L} \frac{v_a}{k_l} \quad (3.3)$$

- La desviación standard de esta carga porcentual
- El máximo flujo porcentual en un segmento, definido como:

$$\text{MAX}_{a \in l \in L} \frac{v_a}{k_l} \quad (3.4)$$

- El promedio de carga en los segmentos sobrecargados
- La desviación estandard de la carga en los segmentos sobrecargados

Las tablas 3.1 y 3.2 muestran estos valores para los casos con y sin congestión. Puede observarse que al considerar la congestión, la carga media resulta mayor, pero su desviación standard es mucho menor y el valor máximo de la carga disminuye notablemente. Por otra parte, el promedio de carga en los segmentos sobrecargados disminuye y que su desviación standard también baja. Sin embargo, el número de segmentos de línea sobrecargados y el número de líneas cargadas aumentan.

n	β	% líneas con sobrecarga	% segmentos con sobrecarga	% carga segmento más sobrecargado
sin cong.		30.2	13.1	510.7
2	10	30.0	13.5	374.5
3	10	32.4	14.0	343.0
4	10	33.1	14.3	294.6
5	10	34.9	14.9	282.9
2	60	32.2	14.1	254.2
3	60	34.2	14.0	227.7
4	60	36.0	14.3	218.4
5	60	34.7	13.8	229.7

Tabla 3.1: Medidas de Flujo en Líneas y Segmentos de Líneas

n	β	% medio carga segmentos	Desviación carga segmentos	% medio carga segmentos sobrecargados	Desviación carga segmentos sob.
sin cong.		45.7	74.1	224.5	117.2
2	10	41.3	55.7	159.4	49.5
3	10	41.3	51.6	146.4	39.3
4	10	41.6	50.7	142.6	37.1
5	10	41.8	47.9	134.1	28.3
2	60	47.5	46.1	135.3	29.8
3	60	47.4	44.7	130.8	26.5
4	60	47.5	45.4	131.6	26.8
5	60	47.3	46.4	134.7	28.1

Tabla 3.2 : Medidas de Flujo en Líneas y Segmentos de Líneas

Todo lo anterior indica que el modelo está efectivamente redistribuyendo los flujos desde las rutas mínimas hacia otras alternativas, las cuales pueden quedar saturadas. De esta forma, si bien se eliminan cargas muy altas, también aparecen sobrecargas en otras líneas, dependiendo de las rutas alternativas que existan en la red para los viajes que provocan las sobrecargas.

Esto último puede verse claramente en la figura 3.1, donde se muestra la carga normalizada de los segmentos para una asignación sin congestión, frente a otra con congestión. La nube de puntos indica que los segmentos con muy alta sobrecarga en la situación sin congestión, disminuyen notoriamente en el caso con congestión. Asimismo, puede verse, en cambio, que un alto número de segmentos de línea aumenta su carga desde flujo casi nulo a altos valores.

En la figura 3.2 se muestra los histogramas de la carga normalizada bajo las situaciones con y sin congestión. Se presentan dos claros efectos : una disminución del número de segmentos con baja carga, y una reducción de la "cola" de segmentos con flujos muy altos. También puede apreciarse que a pesar de la penalización, siguen habiendo segmentos con alto valor de flujo.

Por último, se hará un breve análisis de los flujos en las líneas de transporte público. Cuando no se consideran los efectos de la congestión, suelen existir líneas de transporte a las que se les asigna muy poco flujo. Para observar este efecto, se presentan las figuras 3.3 y 3.4. En ellas se compara el modelo sin congestión, y un modelo con congestión considerando $\beta = 60$ y $n = 4$. Cada línea se caracteriza por su flujo máximo (primer gráfico), y por el promedio de carga en todos sus segmentos (segundo gráfico), representándose en forma de histogramas. Puede observarse que el número de líneas con poco flujo disminuye notablemente, al igual que aquellas que sobrepasan su capacidad. De esta forma, al considerar el efecto de la congestión se obtiene una mejor repartición de los viajes.

3.4 Conclusiones

En esta sección se expondrán las principales conclusiones que se pueden obtener de la implementación del modelo, y se harán algunas recomendaciones sobre investigaciones futuras.

En cuanto a la implementación del modelo de asignación que considera los efectos de congestión, pueden plantearse algunas conclusiones importantes. Los tiempos de respuesta son muy buenos. De hecho, son comparables a tiempos obtenidos por modelos que no consideran las restricciones de capacidad. Centrándose en el uso de memoria, redes detalladas como la de Santiago 1986 requieren en la actualidad un uso bastante grande de este recurso. Sin embargo, se han identificado algunas posibilidades de fácil implementación (ver Crovetto 1989), que podrían reducir sustancialmente estos requerimientos (de 17 a 5 Megabytes). Evidentemente, existe una situación de compromiso entre memoria y tiempo de ejecución que deberá considerarse en cada aplicación específica.

Como ya se mencionó, hubo ocasiones en que el algoritmo no convergió. De esta forma, la convergencia no está asegurada teóricamente, y lo que es peor, hay casos prácticos en que no se da. Esto motiva a investigar correcciones del algoritmo de diagonalización que pudiesen servir bajo sobrecargas excesivas, y a implementar otros algoritmos que resuelvan el modelo simplificado de De Cea y Fernández.

Los resultados obtenidos han sido mejores que los producidos por algoritmos de rutas mínimas, aunque en algunos casos presenta todavía sobrecargas excesivas. Frente a este hecho se debe estudiar muchos aspectos. Primero, debe tratarse con mucho más cuidado el problema de los nodos de acceso a las zonas. Tal como están definidos en la red de 1977, es imposible que cualquier modelo que exija la conservación de flujos no entregue resultados con sobrecarga. En efecto, hay zonas con gran volumen de viajes que tienen acceso a pocos nodos, por lo cual las líneas que pasan por ellos quedan inevitablemente sobrecargadas. Debe estudiarse, entonces, cual es la situación real de los usuarios de esas zonas. Posibles explicaciones son

que existan más nodos de acceso, o haya más líneas que pasen por ellos, o los viajes simplemente no se estén realizando. Por otro lado, el concepto de "capacidad de un vehículo", que determina finalmente las capacidades de las líneas y de las secciones de ruta, debe ser precisado. En efecto, estos datos deben tener valores muy confiables para poder así analizar los valores obtenidos.

También, se debe incluir algunas conclusiones sobre la función de penalización que se escogió para representar el aumento de los tiempos de espera. Esta contiene dos parámetros que deben ser calibrados: n y β .

En los resultados obtenidos y para el rango de valores considerados, parece tener mucha más influencia el valor de β que el de n . En efecto, cuando se varió n dejando β constante, se obtenía una distribución muy parecida, produciéndose sólo un efecto de dispersión. En cambio, cuando se alteró el valor de β , se obtuvo una asignación distinta. Este efecto será muy importante cuando se desee calibrar este modelo para aproximarse a una situación real.

REFERENCIAS

ABDULAAL,M. y LEBLANC,L.J. (1979) Methods for Combining Modal Split and Equilibrium Assignment Models. Trans. Sci., Vol. 13, 192-314.

Crovetto, G. (1989) Implementación Computacional y prueba de un modelo de Asignación a Redes Congestionadas de Transporte Público. Departamento de Ingeniería de Transporte, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago.

DE CEA,J. Y FERNANDEZ,J.E. (1988) Equilibrium Assignment on Congested Transit Networks. Centre de Recherche sur les Transports, Publication No. 584, Université de Montréal.

DE CEA,J., ZUBIETA,L., FLORIAN,M. y BUNSTER,P. (1987) Optimal Strategies and Optimal Routes in Public Transit Assignment Models. Centre de Recherche sur les Transports, Publication No. 555, Université de Montréal.

FLORIAN,M. y SPIESS,H. (1982) The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems. Trans. Res., Vol. 16B, 447-483.

LAST,A. y LEAK,S.E. (1976) TRANSEPT: A Bus Model. Traffic Engineering and Control, Vol. 5, 14-20.

NGUYEN,S. y PALLOTINO,S. (1988) Equilibrium Traffic Assignment for Large Scale Transit Networks. European Journal of Operational Research por aparecer.

PANG,J.M. y CHAN,D. (1982) Iterative Methods for Variational and Complementary Problems. Math. Prog., Vol. 24, 284-313.

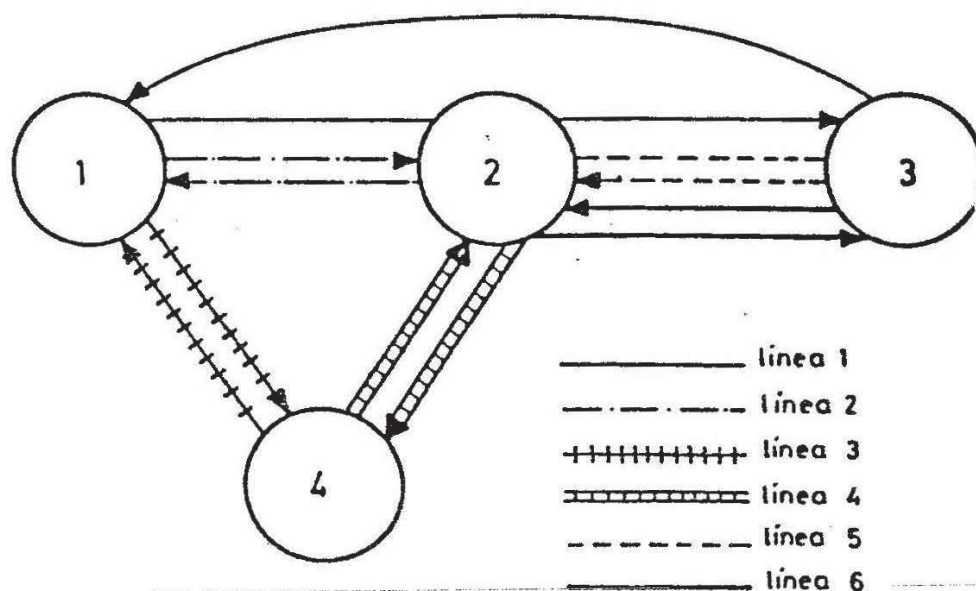


FIG.2.1 RED CODIFICADA EN TERMINOS DE LINEAS DE TRANSPORTE PUBLICO.

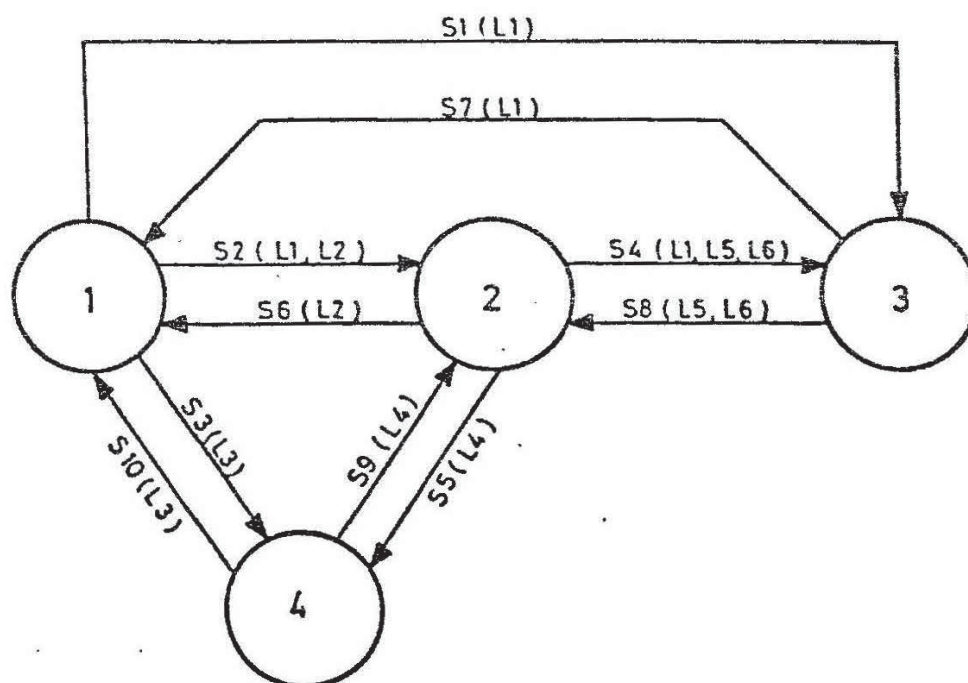


FIG.2.2 RED CODIFICADA EN TERMINOS DE SECCIONES DE RUTA

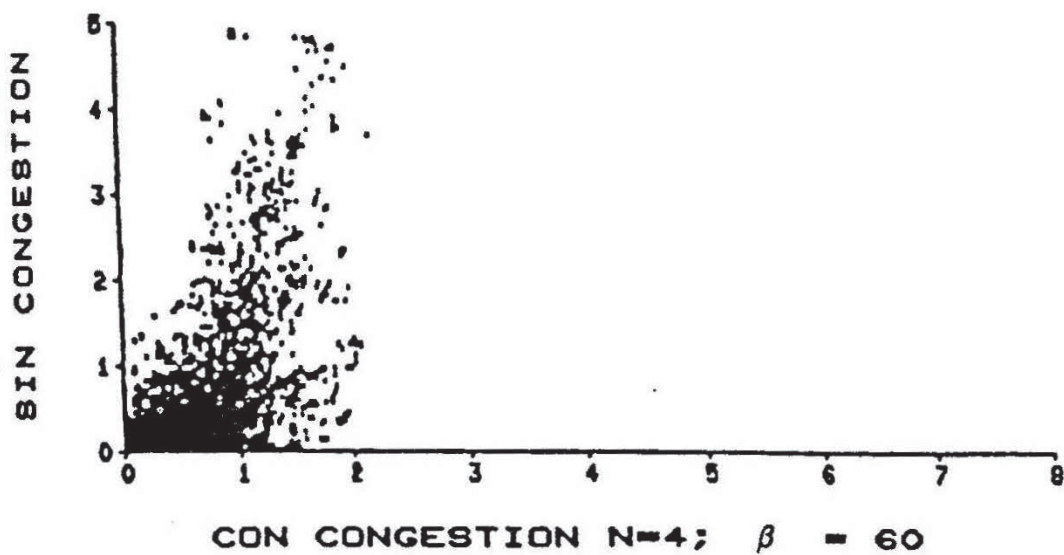


Figura 3.1 Carga en segmenos de Línea

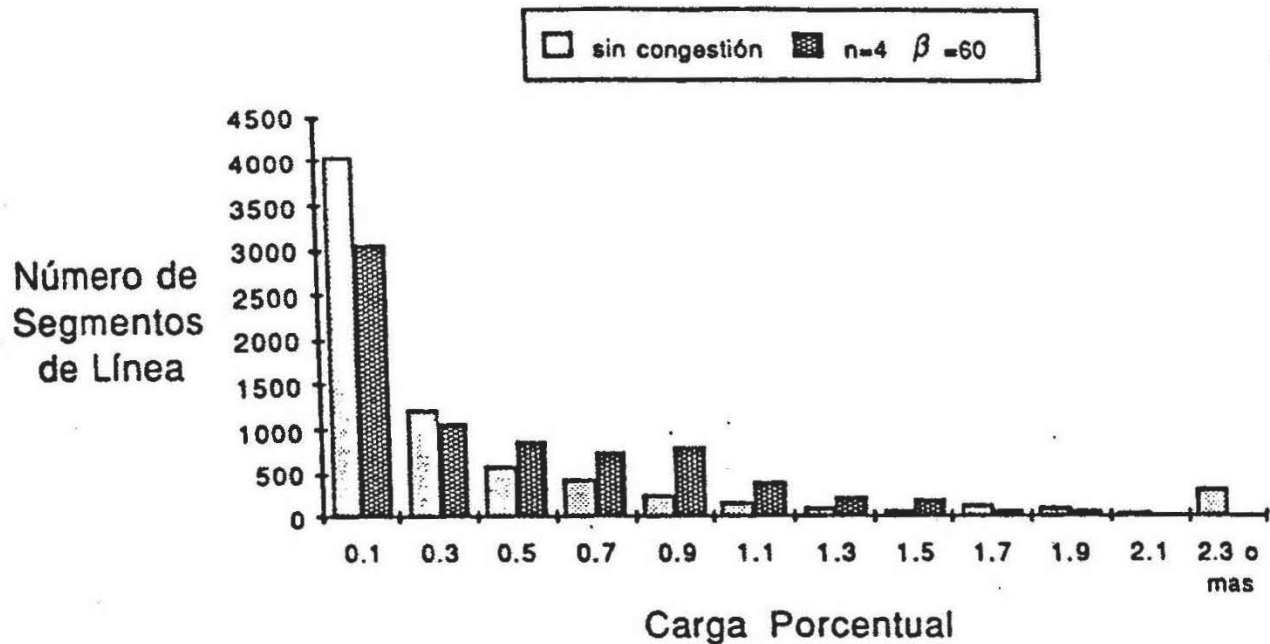


Figura 3.2 Histograma de carga en Segmentos de Línea

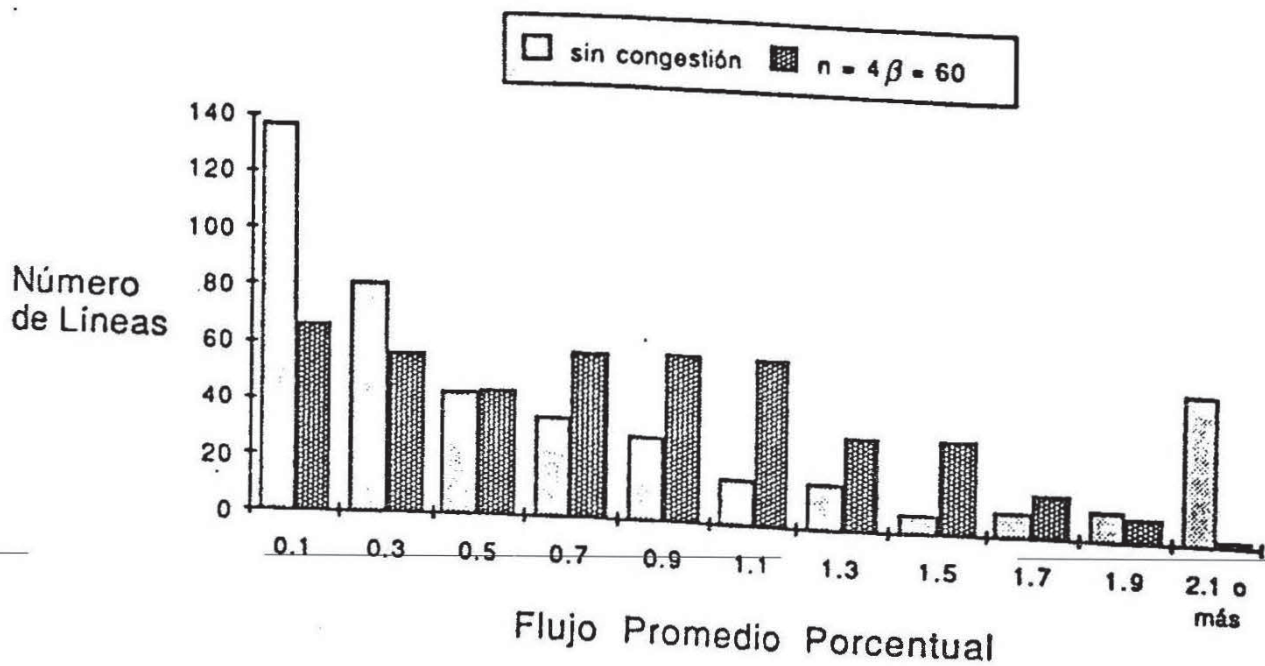


Figura 3.3 Flujo promedio en Líneas

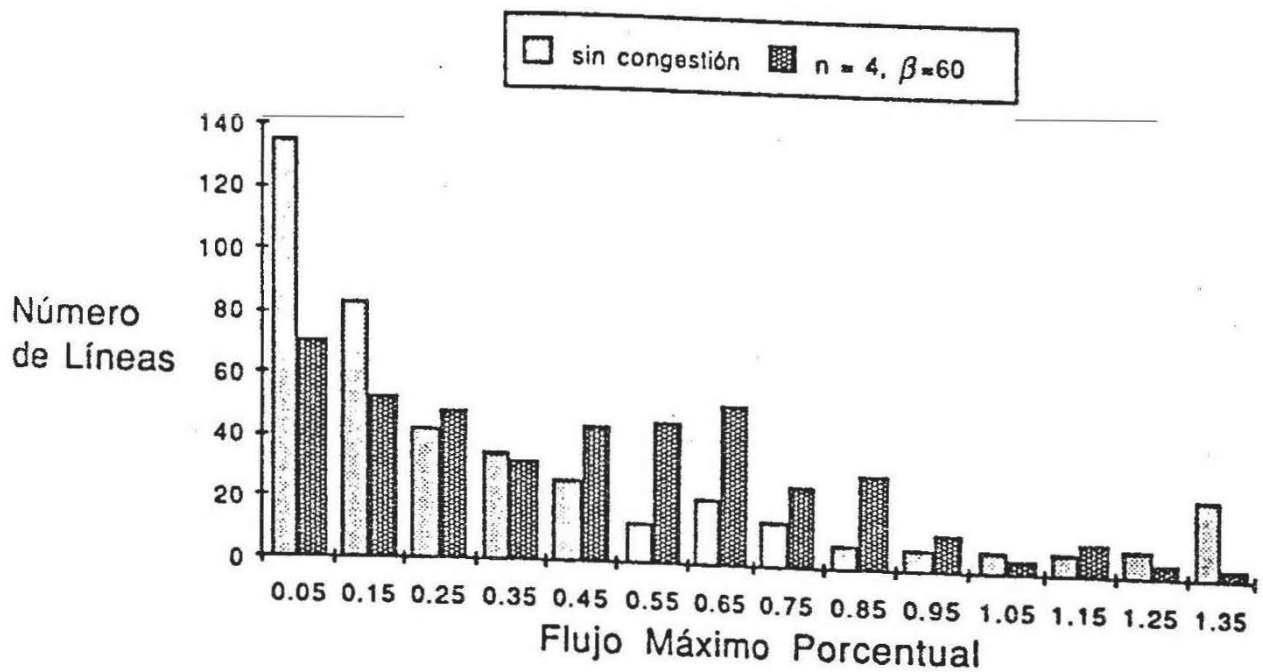


Figura 3.4 Flujo máximo en Líneas