

ASIGNACIÓN DE CAPACIDADES DE CANALES

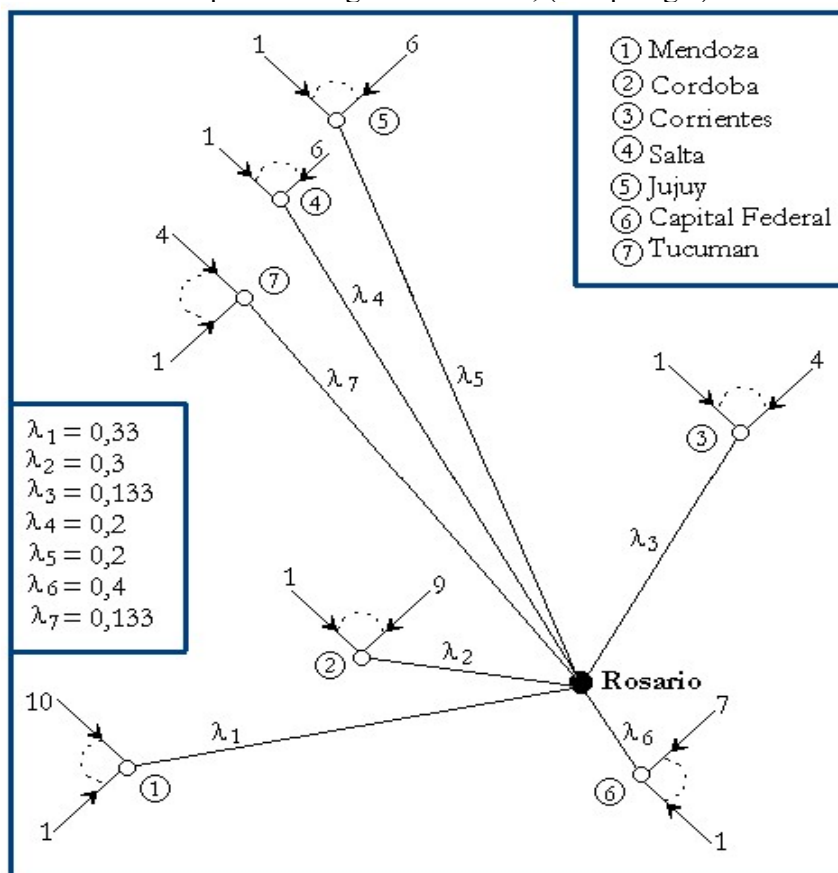
Enfocaremos el problema de determinar: ¿qué capacidad debe tener cada enlace (bps) para suministrar una determinada performance?

Se supone una cierta distribución estadística de tráfico (promedio de longitud de mensajes, frecuencia de ingreso, promedio del número de mensajes entre dos puntos de la red).

Se comienza con la asignación de capacidades para luego abordar temas tales como: elección de topología, conexión entre concentradores, ruteo de mensajes, etc.

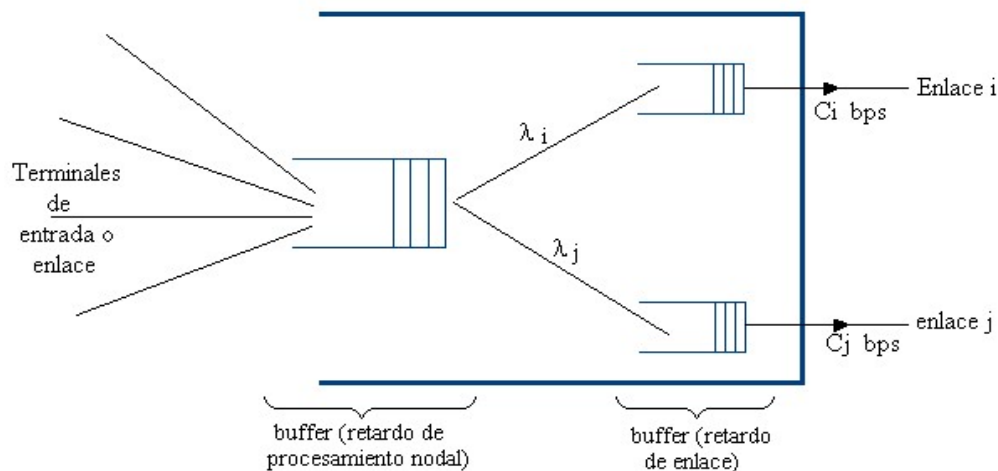
Existen distintos criterios para la asignación de capacidades. Se adopta el criterio debido a L. KLEINROCK, que supone el costo linealmente proporcional a la capacidad. Esto es, el costo equivale a una determinada capacidad total de la red. Luego, se asignarán las capacidades enlace por enlace, con el criterio de tiempo promedio de retardo mínimo. Esto da lugar a la llamada “estrategia de asignación de capacidades de la raíz cuadrada”. En la cual la capacidad de cada enlace es proporcional a la raíz cuadrada del flujo de tráfico a través del enlace.

Consideramos el caso más simple, se tienen 7 ciudades, cada una de ellas con un determinado número de terminales de datos (TTY) a conectarse a una computadora central en Rosario. Se adopta la configuración de red, (la topología) estrella.



Cada terminal (TTY) se supone que producen un mensaje, en promedio, 30 seg. Cada mensaje se supone con una longitud promedio de 120 bits. El objetivo es determinar las capacidades troncales (C_i), en bps, para cada uno de los siete enlaces.

Antes de comenzar analizaremos que sucede en el concentrador en un nodo típico, que se utiliza para combinar los mensajes entrantes y rutearlos, luego, hacia el enlace apropiado (previamente hubo efectuado los procesamientos y almacenamientos necesarios). El modelo más simple supone todos los puestos de entrada explorados instantáneamente, con mensajes en un buffer FIFO (se considera multiplexado estadístico o asincrónico). Luego de un cierto procesamiento los mensajes destinados al enlace i se rutean uno a la vez sobre la línea de salida del concentrador (en el ejemplo resulta un único buffer debido a que en configuración estrella cada concentrador tiene un solo enlace de salida).



Para la asignación de capacidades ignoremos detalles de las operaciones de concentración y almacenamiento. Una simplificación importante surge de suponer que el tiempo de procesamiento nodal es despreciable frente al retardo de espera de enlace (línea disponible).

El modelo se reducirá a:



La frecuencia promedio de mensajes λ_i del enlace i es la suma de las frecuencias de los mensajes de todas las líneas entrantes al nodo. Por ejemplo: en la red estrella de la figura cada terminal tiene la misma frecuencia promedio de mensajes ($1/30$ mensajes/seg.), entonces, λ_i para el i -ésimo enlace es el número de terminales conectados a este por el número de mensajes por segundo. Para el enlace 1 se tiene, en promedio:

$$\lambda_1 = \frac{10}{30} \text{ mes/seg} = \underline{\underline{0,33 \text{ mes/seg}}}$$

en correspondencia con los 10 terminales conectados al concentrador de Mendoza.

La teoría de la cola puede aplicarse a este modelo. Se supone para ello, frecuencia de arribo de mensajes Poisson con mensajes de longitud exponencialmente distribuidos con una longitud promedio de mensajes de $1/\eta_i$ bits/mensajes y un buffer de capacidad infinita, resultando un tiempo de retardo promedio del mensaje de:

$$T_i = \frac{1}{\mu_i - \lambda_i}$$

donde: μ_i es la frecuencia de mensajes por segundo a la salida, no siendo más que la capacidad del canal (C_i) –en bps- al multiplicarla por la longitud promedio en bits en promedio de los mensajes –longitud promedio $1/\eta_i$ -, luego:

$$\mu_i = \eta_i C_i$$

luego:

$$T_i = \frac{1}{\eta_i C_i - \lambda_i}$$

Este retardo incluye el tiempo necesario para transmitir un mensaje más el retardo del mensaje en el buffer. Los buffer reales presentan una probabilidad de bloque inferior a $10e^{-3}$, por lo tanto pueden suponerse como buffer infinito.

Cuando de la frecuencia de flujo de tráfico es cercana al valor $\eta_i C_i$ el tiempo de retardo crece sin límites. Es común definir un parámetro de intensidad de tráfico que por tener un único servidor coincide con el factor de utilización $\rightarrow (u_i = \rho_i)$, luego:

$$\rho_i = \frac{\lambda_i}{\eta_i C_i} < 1$$

En términos de este parámetro:

$$T_i = \frac{1}{\eta_i C_i (1 - \rho_i)}$$

Para $\rho_i < 1$ el retardo será $T_i = 1/\eta_i C_i$ exactamente el tiempo necesario para transmitir el mensaje completo. Para $\rho_i \rightarrow 1$, T_i resulta muy grande.

Si $\lambda_i = 0,33$ mens./seg. y $1/\eta_i = 120$ bits por mensaje, y si $C_i = 100$ bps se produce un retardo por mensaje transmitido sobre el canal de salida de:

$$T_i = \frac{1}{100 / 120 - 0,33} \text{ seg./mens.} \cong \underline{\underline{2 \text{ seg./mens.}}}$$

De este tiempo el necesario para transmitir el mensaje será:

$$\frac{1}{\eta_i C_i} = \underline{\underline{1,2 \text{ seg.}}}$$

Y el tiempo de almacenamiento es:

$$(2 - 1,2) \text{ seg.} = \underline{\underline{0,8 \text{ seg.}}}$$

Resulta:

$$\rho_1 = \frac{\lambda_1}{\eta_1 C_1} = \frac{0,33}{0,83} = 0,4$$

El tiempo de retardo sobre el canal Rosario-Buenos Aires con la misma capacidad (100 bps) se tiene:

$$T_6 = \frac{1}{0,833 (1 - 0,48)} = \underline{\underline{2,3 \text{ seg.}}}$$

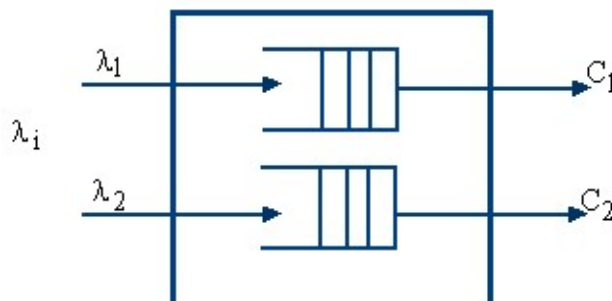
Mientras que entre Corrientes y Rosario será, para la misma capacidad:

$$T_3 = \frac{1}{0,833 (1 - 0,16)} = \underline{\underline{1,43 \text{ seg.}}}$$

Incrementando la capacidad del canal a 900 bps, disminuye el tiempo de retardo, resultando:

$$\left. \begin{aligned} T_1 &= \frac{1}{7,5 (1 - 0,044)} = \underline{\underline{0,14 \text{ seg.}}} \\ T_2 &= \underline{\underline{0,141 \text{ seg.}}} \\ T_3 &= \underline{\underline{0,136 \text{ seg.}}} \end{aligned} \right\} \begin{array}{l} \text{tiempo de transmision} \\ \text{del mensaje } 0,133 \text{ seg.} \end{array}$$

Considerando la red como un conjunto de enlaces (uno por nodo), el buffer asociado con el enlace i-esimo tiene una frecuencia de arribo de λ_i y la longitud promedio de mensaje emitido será $1/\eta_i$, se tiene:



Se considera, que esta forma, la red como una “caja negra”, observando los arribos a la caja, se tienen δ mensajes ingresando a la caja por unidad de tiempo, donde:

$$\delta = \sum_{i=0}^n \lambda_i$$

La caja negra se comporta como una cola para la cual, por el teorema de Little, tenemos:

$$N = \delta E[T] = E[n]$$

Para el promedio estadístico se tiene:

$$E[n] = \sum_{i=1}^n E[n_i] = \sum_{i=1}^n \lambda_i E[T_i]$$

Sumado sobre todos los buffer de la red y aplicando el teorema de Little a cada uno, nos da:

$$E[T] = \frac{1}{\delta} \sum_{i=1}^x \lambda_i E[T_i] \quad (*)$$

Con la ecuación para tiempo de retardo, podemos iniciar el manejo del problema de asignación de capacidades de canales en una red (figura anterior).

Suponemos el objetivo es minimizar el tiempo de retardo promedio sobre la red completa con una capacidad total supuesta fija. La frecuencia total de la red será:

$$\begin{aligned} \delta &= \sum_{i=1}^x \lambda_i = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 \\ \delta &= \underline{\underline{1,7 \text{ mens./seg.}}} \end{aligned}$$

El, retardo promedio de mensajes para la red se puede calcular por medio de la ecuación(*).

$$E[T] = \bar{T} = \frac{1}{\delta} \sum_{i=1}^x \lambda_i E[T_i]$$

Donde $E[T]$ presenta el tiempo de retardo “pesado” con los T_i en el rango de 1 a 7. Suponiendo la capacidad total como la suma de las capacidades C_i , resulta:

$$T_i = \frac{1}{\eta_i C_i - \lambda_i}$$

Este valor se toma constante, debiendo elegirse los C_i tal que minimicen el valor $E[T]$. Estos valores se pueden hallar utilizando los multiplicadores de Lagrange: se puede demostrar que el mínimo se puede hallar minimizando $E[T] + \alpha C$. El multiplicador de Lagrange α se elige para asegurar la constancia de C .

Minimizando por diferenciación de $E[T] + \alpha C$ con respecto a C_i , se encuentra la solución óptima dada por:

$$C_i^{\text{opt}} = \frac{\lambda_i}{\eta_i} + \frac{C(1-\rho)\sqrt{\lambda_i/\eta_i}}{\sum_{j=1}^n \sqrt{\lambda_j/\eta_j}} \quad (**)$$

ρ cumple la función de parámetro intensidad de tráfico para toda la red.

$$\delta = \frac{1}{C} \sum_{i=1}^x \lambda_i / \eta_i$$

Con esta relación de capacidades, canal por canal es posible sustituir en la ecuación (**) y luego hallar T_i para cada canal. Luego se puede calcular la expresión T_{\min} , siendo:

$$\overline{T}_{\min} = \frac{\left(\sum_{i=1}^x \sqrt{\lambda_i / \eta_i} \right)^2}{\delta C (1 - \rho)}$$

Esta asignación de capacidad se denomina la “regla de asignación de la raíz cuadrada del flujo de información. El T_{\min} depende inversamente de C (o sea del costo). Siempre es posible disminuir el tiempo de retardo, aumentando el valor de C y con ello el costo. La ecuación de T_{\min} da el mínimo tiempo de retardo posible para un determinado costo.

La ecuación de C_i opt. Tiene dos partes:

✓ La primera parte λ_i / η_i es el valor mínimo necesario que puede ser dado a C_i para habilitar el tráfico sobre el canal donde λ_i es el promedio de mensajes ha ser manejados y $1/\eta_i$ su longitud promedio, luego el cociente representa el promedio de flujo ofrecido sobre el canal i en bps. Debe ser por supuesto menor que C_i .

✓ La segunda parte indica la capacidad remanente según la estrategia de la raíz cuadrada.

¿Como se aplica, ahora, la estrategia de la raíz cuadrada al circuito de la figura 1?

La inversa de la longitud $\eta_i = \eta = 1/120$ para todos los enlaces. Entonces, y en base a cada λ_i se halla:

$$\rho C = \sum_{i=1}^7 \lambda_i / \eta_i = \frac{1}{\delta} \sum_{i=1}^x \lambda_i = 1,7 \text{ mens./seg.} \cdot 120 \text{ bits/mens}$$

$$\rho C = \underline{\underline{204 \text{ bps.}}}$$

Luego, para $C = 4500$ bps. Como la capacidad total de la red, resulta un $\rho = 0,0453$. –En la práctica se realizan varias elecciones de C , y cada una corresponde a diferentes costos, se encuentra el valor de tiempo de retardo mínimo para cada una y se grafica el tiempo de retardo vs. el costo. De ello se obtiene la C apropiada para un cierto tiempo mínimo.- Así se obtiene:

	Enlace	$C_i^{opt.}$ (bps)	T_i (seg)
<i>Caso Optimo</i>	1	774	0,163
	2	733	0,172
	3	480	0,26
	4	593	0,21
	5	593	0,21
	6	852	0,15
	7	480	0,26

$$\sum_{i=1}^7 \sqrt{\lambda_i / \eta_i} \cong 37$$

$$\overline{T}_{min} = 0,19 \text{ seg.}$$

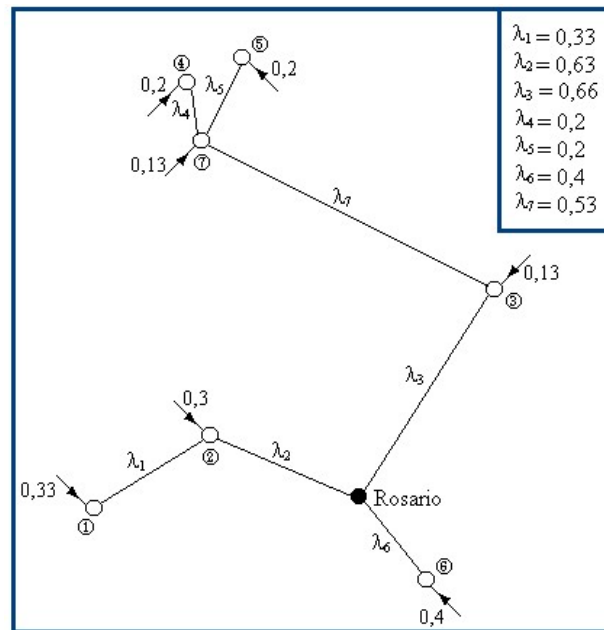
Luego, se debe seleccionar valores de capacidad reales. En este caso se seleccionan canales de 450 o 900 bps.

Enlace	C_i (bps)	T_i (seg)	$\rho_{opt.}$	Costo (u\$s)
1	900	0,140	900	1.080
2	900	0,139	400	480
3	450	0,276	700	700
4	450	0,282	1100	1.100
5	450	0,282	1200	1.200
6	900	0,141	350	420
7	450	0,276	900	900
Costo Total				<u>\$ 5.880</u>

$$\overline{T}_{discreta} = 0,195 \text{ seg.}$$

El costo de la línea de 450 bps es de aproximadamente 1 U\$S por Km, y el de 900 bps es de 1,2 U\$S por Km. Es claro que el costo sería menor si adoptamos una topología árbol.

Consideremos ahora el mismo problema, pero con una estructura árbol para observar como se modifican los parámetros de la red y el costo. Inmediatamente se percibe que la disminución de los Kilómetros de canales es notable.



Para analizar esta red y determinar la asignación de capacidad apropiada, debemos tener en cuenta en el cálculo del retardo de la cola no solamente en el de los mensajes de los terminales que entran en cada nodo sino también los mensajes de concentradores previos. Suponiendo multiplex estadístico, la frecuencia de arribos, las frecuencias de arribos se pueden sumar para determinar la frecuencia de arribo total al nodo, pero las estadísticas de mensajes sobre los sucesivos enlaces están ahora relacionadas a otros. Esto es verdad si vale la suposición de independencia, esto es, se ignoran las dependencias entre las distintas colas.

Bajo esta suposición los mensajes se suponen generados independientemente por cada nodo con longitudes independientes seleccionadas de una distribución exponencial. Luego se suma simplemente las frecuencias de arribos. De los enlaces previos a aquellos que entran al concentrador por el puerto de entrada, tratándolos como si fuesen generados por terminales equivalentes.

Resulta así:

Enlace	Caso Optimo	Caso Discreto			
	Ci (bps)	Ti (seg)	P _{opt.}	Costo (u\$s)	
1	577	450	0,29	550	550
2	834	900	0,146	400	480
3	834	900	0,146	700	840
4	440	450	0,28	250	250
5	440	450	0,28	350	350
6	636	450	0,288	350	350
7	740	900	0,144	700	840
		Costo Total			<u>3.660</u>
$\bar{T}_{min} = 0,3385 \text{ seg.}$		$\bar{T}_{discreta} = 0,35 \text{ seg}$			

Notar que esta red resulta con un tiempo promedio mayor que en la topología estrella. El costo es menor por la eliminación de cuatro largos tramos. Podemos reducir el retardo incrementando la capacidad. Puede suceder de que el costo no crezca tanto como en la topología estrella y obtener tiempos de retardos similares en ambas estructuras.