



66

ENGINYERIES
DE LA
TELECOMUNICACIÓ



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

→ **UPCGRAU**

Anàlisi i avaluació de xarxes →

Esteve Pallarès Segarra
Luis J. de la Cruz Llopis



66

ENGINYERIES
DE LA
TELECOMUNICACIÓ



→ UPCGRAU

Anàlisi i evaluació de xarxes →

Esteve Pallarès Segarra
Luis J. de la Cruz Llopis

Primera edició: julio de 2022

© Els autors, 2022

© Iniciativa Digital Politécnica, 2022
Oficina de Publicacions Acadèmiques Digitals de la UPC
Jordi Girona 31,
Edifici Torre Girona, Plant 1, 08034 Barcelona
Tel.: 934 015 885
www.upc.edu/idp
E-mail: info.idp@upc.edu

DL: B 14109-2022
ISBN: 978-84-19184-34-4

Qualsevol forma de reproducció, distribució, comunicació pública o transformació d'aquesta obra només pot realitzar-se amb l'autorització dels seus titulars, excepte excepció prevista en la llei.



Índex

Índex	5
1. Introducció	9
1.1 Xarxes telemàtiques	9
1.2 Model de node de commutació	10
1.3 Anàlisi i avaluació de xarxes	12
2. Modelatge i avaluació de sistemes de transmissió	15
2.1 Introducció	15
2.2 Introducció als processos de Márkov	16
2.3 Processos de naixement i mort	19
2.3.1 Procés de Poisson	25
2.4 Processos de naixement i mort en equilibri	30
2.4.1 Solució general d'equilibri	31
2.5 Sistemes d'espera	33
2.5.1 Sistema M/M/1 El sistema M/M/1 té les característiques següents:	34
2.5.2 Sistema M/M/ ∞	46
2.5.3 Sistema M/M/m	48
2.6 Sistemes amb pèrdues	54
2.6.1 Sistema M/M/1/K	55
2.6.2 Sistema M/M/m/m	59
2.7 Sistemes amb població finita	60
2.7.1 Sistema M/M/1//M	60
2.7.2 Sistema M/M/1/K/M	62
2.8 Sistemes semimarkovians	65
2.8.1 Sistema M/G/1	66
2.8.2 Sistemes amb prioritat	74
2.8.3 Prioritats sense expulsió	76
2.8.4 Prioritats amb expulsió	79



3. Anàlisi de les funcionalitats de la xarxa.....	85
3.1 Teorema de Burke.....	85
3.2 Nombre mitjà de salts que fa un paquet a la xarxa	88
3.3 Temps mitjà de trànsit d'un paquet genèric.....	90
3.4 Repartiment òptim de fluxos.....	91
3.4.1 Repartiment òptim per dos camins.....	92
3.4.2 Repartiment òptim per tres camins o més	97
3.5 Control de congestió	102
3.5.1 Descart de paquets	103
3.5.2 Regulació explícita de la taxa	104
3.5.3 Traffic shaping	104
3.5.4 Mecanismes amb finestra.....	106
4. Tècniques d'accés múltiple	113
4.1 Tècniques aleatòries.....	114
4.1.1 Aloha	114
4.1.2 Carrier Sense Multiple Access (CSMA)	129
4.1.3 Carrier Sense Multiple Access / Collision Detection (CSMA/CD)	141
4.2 Tècniques deterministes	148
4.2.1 Time Division Multiple Access (TDMA)	148
4.2.2 Frequency Division Multiple Access (FDMA).....	150
4.2.3 Sondeig (<i>polling</i>)	151
5. Bibliografia.....	159



→ 1



Introducció

1.1 Xarxes telemàtiques

En una primera aproximació, les xarxes telemàtiques es poden veure com un conjunt de recursos d'interconnexió dissenyats per transportar un trànsit de telecomunicacions originat pels usuaris (persones o dispositius) de la xarxa. Básicamente, estan formades per nodes interconnectats a través d'uns mitjans (amb fil o sense) de transmissió. En els nodes, trobem principalment dos tipus de sistemes: sistemes de transmissió i sistemes de commutació. Els sistemes de transmissió són els responsables de transmetre la informació d'un node a un altre. En canvi, els sistemes de commutació s'encarreguen de realitzar la commutació necessària entre els canals d'entrada i els de sortida del node.

La tècnica de commutació més utilitzada a les xarxes telemàtiques és la commutació de paquets en mode datagrama, que consisteix a fragmentar el missatge que es vol transmetre en fragments més petits i encaminar cada fragment de manera independent dins la xarxa. A cada fragment se li afegeix una capçalera que indica, entre d'altres coses, el destinatari i la posició que ocupa el fragment dins el missatge. El conjunt format per la capçalera i pel fragment s'anomena *paquet*. En el mode datagrama, els paquets són tractats de manera independent dins la xarxa i, per tant, poden seguir rutes diferents i arribar desordenats a la seva destinació, encara que pertanyin al mateix missatge.

Als sistemes de transmissió arriben paquets que provenen de diferents orígens i que pertanyen a diferents missatges. Aquests paquets s'han de repartir el temps d'utilització del mitjà de transmissió, cosa que es coneix com a *multiplexació*. Una tècnica de multiplexació molt utilitzada a les xarxes de commutació de paquets és la multiplexació estadística. Per a aquest tipus de multiplexació, els paquets que s'han de transmetre no tenen prefixat un temps d'utilització del mitjà de transmissió, com passaria en el cas de la multiplexació determinista. En aquest cas, la utilització del mitjà tan sols depèn dels paquets que esperen ser

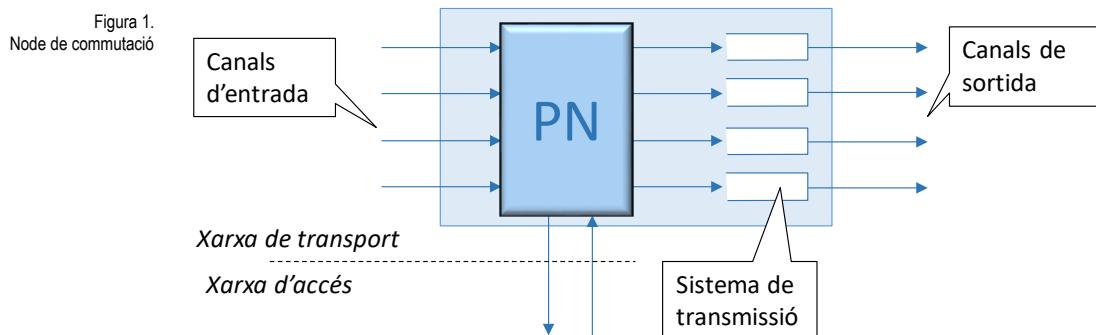


transmesos. Així doncs, quan un paquet arriba al sistema de transmissió, s'emmagatzema en un *buffer* fins que li arriba el torn de ser transmès. Llavors utilitza tot el temps que calgui fins que finalitza la transmissió. Per tant, amb la multiplexació estadística, el mitjà només resta lliure quan ja no queden paquets que esperin ser transmesos. L'ordre de transmissió dels paquets en espera sol ser el mateix que l'ordre d'arribada, però també es poden donar situacions en les quals aquest ordre sigui diferent, com quan es dona prioritat a algun tipus de paquet. En cas que s'assignin prioritats, es pot donar la situació que un paquet prioritari interrompi la transmissió d'un paquet menys prioritari.

Podem fer una classificació genèrica de les xarxes telemàtiques en *xarxes de transport* i *xarxes d'accés*. Les xarxes de transport solen donar cobertura a una àrea extensa i estan formades principalment per nodes de commutació i línies de transmissió. La seva tasca principal és transportar el trànsit per la xarxa. En canvi, les xarxes d'accés són xarxes amb una cobertura geogràfica més reduïda a través de les quals els usuaris poden accedir a les xarxes de transport. Per a aquest tipus de xarxes, hi acostuma a haver un recurs compartit per tots els usuaris de la xarxa (estació base, punt d'accés, satèl·lit, línia de transmissió...). La manera com els usuaris accedeixen a aquest recurs compartit s'anomena **tècnica d'accés múltiple**. Aquestes tècniques es poden classificar en *aleatòries* o *deterministes*. En el cas determinista, un usuari ha de tenir prèviament el mitjà assignat per poder transmetre. En canvi, en el cas aleatori, l'accés es fa quan es necessita, de manera que es poden produir conflictes si dues estacions o més volen fer l'accés simultàniament.

1.2 Model de node de commutació

A la figura 1, es pot veure un esquema simplificat d'un node de commutació. Aquests nodes són dispositius que reben paquets provinents d'altres nodes, n'analitzen la capçalera i decideixen quin és el node següent en la ruta que segueix cada paquet cap a la seva destinació.





Dins el node, tenim el sistema de commutació, format pel processador nodal (PN), el qual decideix per on commutar els paquets que arriben pels ports d'entrada. Un cop presa la decisió, diposita el paquet en el sistema de transmissió corresponent. Els sistemes de transmissió de cada port de sortida estan formats per un *buffer* i un transmissor. Al *buffer* s'emmagatzemen els paquets mentre esperen ser transmesos (multiplexació estadística). El retard que experimenta un paquet mentre es transfereix d'un node a un altre l'anomenem **temps de transferència** i és la suma del temps de commutació en el processador nodal, el temps d'espera en el *buffer*, el temps de transmissió i el temps de propagació del paquet. Usualment, la potència de càlcul del processador nodal està molt per sobre de la velocitat de les línies de transmissió, de manera que el temps de commutació es pot considerar negligible respecte del temps de transmissió. A més, en moltes situacions reals podem considerar que la distància entre nodes també permet negligir el temps de propagació respecte al temps de transmissió d'un paquet. Amb aquestes hipòtesis, el temps de transferència el podem aproximar pel **temps d'espera** més el **temps de transmissió**, i podem modelar el sistema de transmissió tal com es mostra a la figura 2. Així doncs, a l'hora d'avaluar el retard que experimenten els paquets a la xarxa i l'ocupació dels *buffers* dels nodes, caldrà analitzar la multiplexació estadística en els sistemes de transmissió dels nodes.

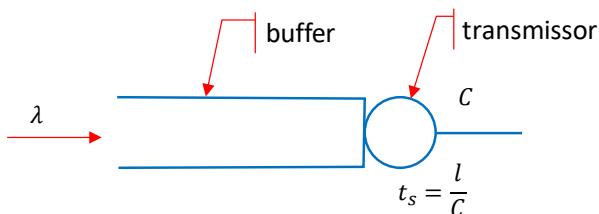


Figura 2.
Model bàsic de
sistema de transmis-
transmissió

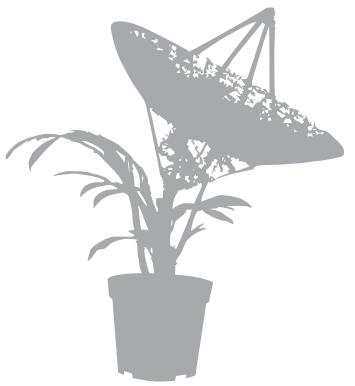
Els paràmetres que determinen aquesta anàlisi són:

- Taxa d'arribades al sistema de transmissió. És el nombre mitjà de paquets que hi arriben cada segon i el denotem amb la lletra λ .
- Model probabilístic del temps entre arribades. El temps entre arribades és una variable aleatòria (t_{ia}) que té una *funció de densitat de probabilitat (PDF)*. La inversa del valor mitjà del temps entre arribades és la taxa d'arribades.
- Capacitat del canal C . És la velocitat del mitjà de transmissió, expressada en bits per segon.
- Longitud dels paquets l (bits). També és una variable aleatòria que determina el temps de transmissió d'un paquet. El temps de transmissió t_s es calcula com la longitud del paquet, dividida per la capacitat del canal.



1.3 Anàlisi i evaluació de xarxes

Als capítols següents, es presenten diferents eines d'anàlisi de les xarxes tele-màtiques. Al capítol 2, es donen eines per avaluar l'ocupació i el retard en els sistemes de transmissió dels nodes. Aquestes eines estan basades en la teoria de cues, la qual modela la multiplexació estadística. Un cop analitzats diferents models de sistemes de transmissió, els resultats s'utilitzaran per fer una anàlisi més global de tota la xarxa. Al capítol 3, es calculen paràmetres com el nombre mitjà de salts que fa un paquet a la xarxa i el temps que triga a anar des del seu node d'origen fins al de destí (**temps de trànsit**). També es defineixen metodologies per optimitzar el repartiment de la càrrega dins la xarxa. Finalment, s'avaluen algunes de les tècniques més comunes per al control de la congestió a les xarxes. Al capítol 4, s'analitzen diferents tècniques d'accés múltiple, tant aleatòries com deterministes. En les tècniques aleatòries, es poden produir col·lisions, és a dir, l'intent d'accés de més d'una estació alhora. En aquest cas, l'anàlisi es basa a trobar la relació entre els intents d'accés de les estacions de la xarxa i el nombre de paquets transmesos amb èxit. A més, també s'avalua el temps des que un paquet vol accedir a la xarxa fins que finalment hi és transmès amb èxit, i en l'anàlisi s'inclou el nombre de retrasmissions del paquet degudes a col·lisions. D'altra banda, si l'accés es fa de forma determinista, no hi ha col·lisions. En aquest cas, l'anàlisi consisteix a calcular el retard que experimenta un paquet des que es genera fins que finalment pot accedir a la xarxa.



→2



Modelatge i avaluació de sistemes de transmissió

2.1 Introducció

Al capítol d'introducció, s'ha vist que un sistema de transmissió, en la seva forma més simplificada, es pot modelar com una cua (buffer de paquets) més un servidor (canal de transmissió). A aquest sistema, hi van arribant paquets que han de ser transmesos pel canal, i el temps entre arribades és aleatori. A més, la grandària (la longitud en bits d'aquests paquets) també és aleatòria, igual que el temps necessari per a la transmissió. Així doncs, com que l'ocupació d'aquest sistema ve determinada per esdeveniments aleatoris, el nombre de paquets dins el sistema també serà aleatori. En altres paraules, el nombre d'elements en el sistema l'hem d'estudiar com una variable aleatòria i la seva evolució temporal dona lloc a un procés estocàstic. A la figura 3, es pot observar una possible realització d'aquest procés, junt amb les variables aleatòries comentades prèviament, per a un sistema de transmissió format per una única cua i un únic canal.

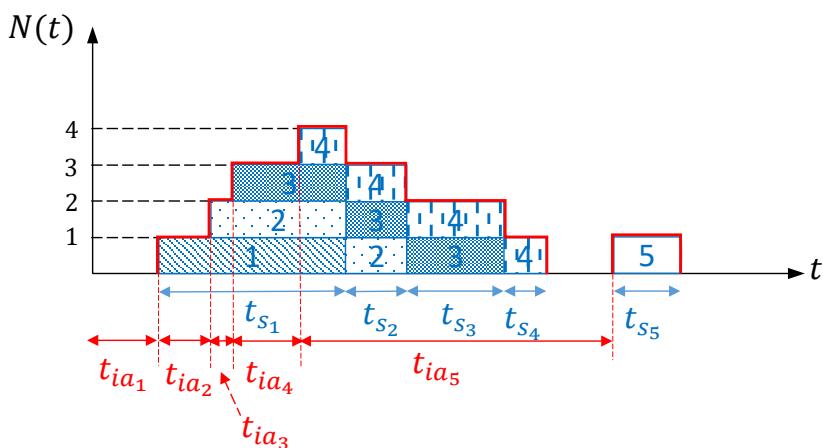


Figura 3.
Realització de l'ocupació d'un sistema de transmissió



Als apartats següents d'aquest capítol, s'estudia un tipus especial de processos estocàstics que s'utilitzen sovint per modelar els sistemes de transmissió. Els processos d'aquest tipus compleixen una sèrie de propietats que els fa especialment interessants en aquest entorn i que permeten reduir la complexitat matemàtica presentada.

2.2 Introducció als processos de Màrkov

Es defineix un procés estocàstic com un conjunt de variables aleatòries X indekses per a un conjunt de valors t anomenats *índexs*. Els possibles valors que pren la variable aleatòria s'anomenen *estats*. Podem entendre un procés estocàstic com una funció $X(t)$ que, per a cada t , pren un valor que no és determinista, és a dir, no es coneix el valor de la funció, però sí la probabilitat de cada possible valor o rang de valors que pot prendre la funció. Podem classificar els processos estocàstics en funció de:

- El valor de l'estat X , el qual pot ser:
 - Continu.
 - Discret; en aquest cas, el procés s'anomena cadena.
- El valor de l'índex t , que pot ser:
 - Continu.
 - Discret; en aquest cas, el procés s'anomena seqüència.

La quantitat de paquets en un sistema de transmissió en funció del temps és un procés discret en estat (la quantitat de paquets), amb índex continu (el temps).

Els processos estocàstics també es poden classificar en funció de les dependències entre les variables aleatòries corresponents a diferents valors de t . Exemples d'aquesta classificació són els processos estacionaris o els processos de Màrkov. Volem centrar-nos en aquest darrer tipus de processos. Un procés amb índex discret és de Màrkov quan es compleix:

$$\begin{aligned} P\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(n-1) = x_{n-1}, \dots, X(1) = x_1\} &= \\ &= P\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \end{aligned}$$

Aquesta definició es pot interpretar de la manera següent. Si n representa l'instant present, aleshores $n+1$ és un instant corresponent al futur i el conjunt $\{n-1, \dots, 1\}$ són els instants del passat. Així doncs, l'equació indica que la probabilitat que l'estat futur sigui x_{n+1} , atès que el procés està a l'estat x_n i ha passat pels estats x_{n-1}, \dots, x_1 , només depèn de l'estat present. És a dir, **el futur del procés només depèn del present, ja que ha oblidat el passat**. Per aquest motiu, els processos de Màrkov s'anomenen processos **sense memòria**.

Com ja s'ha indicat, estem interessats en processos discrets en estat (cadenes) i continus en el temps t . En aquest cas, la definició d'un procés de Márkov és la següent:

$$\begin{aligned} P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1\} &= \\ &= P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\}, \end{aligned}$$

per a qualsevol conjunt $\{t_{n+1}, t_n, \dots, t_1\}$ que compleixi $t_{n+1} < t_n < \dots < t_1$.

La definició indica que, atés que coneixem l'estat del procés per diferents instants de temps (t_n, \dots, t_1) , l'únic instant que aporta informació sobre el futur del procés és l'instant més proper a aquest futur. És a dir, el procés oblide tot el que ha passat en instants anteriors a l'instant més proper al futur.

Imaginem, per exemple, un aventurer que viatja en veler per un arxipèlag. Cada dia, en visita una illa diferent i l'illa escollida tan sols depèn de la direcció del vent. D'aquesta manera, l'illa següent que visitarà l'aventurer només depèn de l'illa on es troba cada matí, independentment de quina hagi estat la ruta per arribar fins aquesta illa. En aquest cas, estem davant d'un procés de Márkov en el qual el futur només depèn del present, ja que ha oblidat el passat.

Perquè un procés compleixi la condició de Márkov, la variable aleatòria que indica el temps de permanència en un estat determinat ha d'estar distribuïda exponencialment. És a dir, si anomenem aquesta variable aleatòria t_e , la seva funció de densitat de probabilitat (*PDF*) és de la forma:

$$f_{t_e}(t) = \lambda e^{-\lambda t}; t \geq 0.$$

Demostrem primer que, si el temps de permanència en un estat és una variable aleatòria exponencial, aleshores el procés és de Márkov (no té memòria). Més endavant demostrarem que les variables aleatòries exponencials són les úniques que compleixen aquesta propietat.

Calculem primer la probabilitat que el temps de permanència en un estat sigui inferior o igual a t_1 :

$$p(t_e \leq t_1) = \int_0^{t_1} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t_1}$$

Aquesta és la probabilitat que el canvi d'estat del procés es produueixi entre els instants 0 i t_1 , és a dir, t_e representa l'instant en què es produeix el canvi d'estat si prenem com a origen del temps l'instant en què el procés entra en l'estat actual. Suposem ara que hem estat observant el procés durant t_0 segons sense que aquest hagi canviat d'estat i volem calcular la probabilitat que el procés canviï d'estat durant els t_1 segons següents. El càlcul serà el següent:¹

¹ Cal recordar que $p(A|B) = p(A \cap B)/p(B)$.

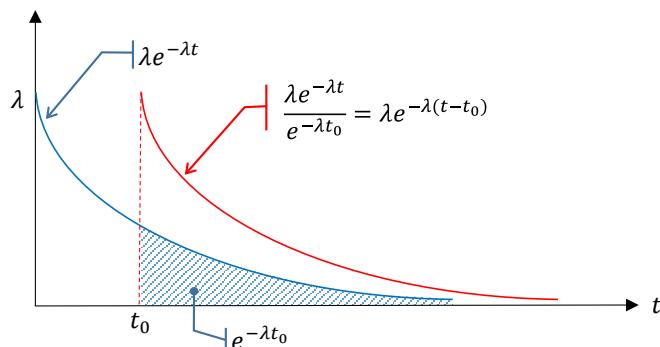


$$\begin{aligned}
 p(t_e \leq t_0 + t_1 | t_e > t_0) &= \frac{p(t_0 < t_e \leq t_0 + t_1)}{p(t_e > t_0)} = \frac{\int_{t_0}^{t_0+t_1} \lambda e^{-\lambda t} dt}{\int_{t_0}^{\infty} \lambda e^{-\lambda t} dt} = \\
 &= \frac{e^{-\lambda t_0} - e^{-\lambda(t_0+t_1)}}{e^{-\lambda t_0}} = 1 - e^{-\lambda t_1}.
 \end{aligned}$$

Ambdós resultats coincideixen. Això vol dir que la probabilitat que el procés canviï d'estat en els t_1 segons següents és la mateixa tant si el procés acaba d'entrar en l'estat actual com si ja porta t_0 segons en aquest estat. El procés no recorda el seu passat, de manera que el seu futur només depèn del seu present, és a dir, no té memòria. Podem afirmar, doncs, que si el temps de permanència en un estat és una variable aleatòria exponencial, el procés és de Márkov.

Podem interpretar geomètricament aquest resultat pensant que, si trunquem els primers t_0 segons de la **PDF** exponencial, n'obtenim una nova exponencial més petita, de manera que la seva àrea és menor que 1, però si l'escalem de manera que l'àrea valgui 1 (dividim per $e^{-\lambda t_0}$), tornem a obtenir la funció original desplaçada a t_0 .

Figura 4.
Propietat *memoryless*
de les variables aleatòries exponencials



Aquesta propietat és molt útil per analitzar els temps residuals. Per exemple, suposem que, transcorregut un temps des de l'inici de la transmissió d'un paquet, volem saber com és la **PDF** del temps que manca per finalitzar la transmissió. Aquest temps l'anomenem *temps residual de transmissió*. Si el temps de transmissió d'un paquet és una variable aleatòria exponencial, aleshores el temps residual també és exponencial, amb la mateixa **PDF** que el temps de transmissió, ja que no té memòria. Això no vol dir que el temps de transmissió i el temps residual tinguin el mateix valor sinó que, si féssim molts experiments i construïssim un histograma amb tots els temps de transmissió i amb tots els temps residuals, n'obtindríem histogrames amb formes semblants. Si normalitzem les àrees d'ambdós histogrames, aleshores seran iguals. Cal tenir en compte que les àrees d'ambdós histogrames abans de normalitzar-los són diferents. El segon histograma conté menys mostres, ja que no inclou les transmissions ja finalitzades en el moment en què mirem el valor del temps residual.

Ara ens cal demostrar que, si un procés és de Márkov, el temps de permanència en un estat només pot ser una variable aleatòria exponencial. Volem demostrar que la funció distribució de probabilitat (**CDF**) del temps de permanència és de la forma:

$$p(t_e \leq t) = F_{t_e}(t) = 1 - e^{-\lambda t}$$

i que, a més, és l'única possible. Si el procés és de Márkov, compleix:

$$p(t_e > t_0 + t_1 | t_e > t_0) = p(t_e > t_1).$$

Manipulant aquesta expressió, s'obté:

$$\frac{p(t_e > t_0 + t_1, t_e > t_0)}{p(t_e > t_0)} = \frac{p(t_e > t_0 + t_1)}{p(t_e > t_0)} = p(t_e > t_1)$$

$$p(t_e > t_0 + t_1) = p(t_e > t_1) \cdot p(t_e > t_0)$$

$$\ln[p(t_e > t_0 + t_1)] = \ln[p(t_e > t_1)] + \ln[p(t_e > t_0)]$$

Es pot demostrar que l'única funció que compleix $g(x + y) = g(x) + g(y)$ és $g(x) = Cx$, essent C una constant. Aleshores, tenim:

$$\ln[p(t_e > t)] = Ct$$

$$p(t_e > t) = e^{Ct}$$

Considerem processos que hauran de canviar d'estat més tard o més d'hora; per tant, $p(t_e > t)$ tendeix a 0 quan t tendeix a ∞ , la qual cosa implica que l'exponent és negatiu, $C = -\lambda$. Finalment, obtenim que, en un procés de Márkov, l'única **CDF** possible pel temps de permanència es correspon amb la **CDF** d'una variable aleatòria exponencial:

$$F_{t_e}(t) = p(t_e \leq t) = 1 - p(t_e > t) = 1 - e^{-\lambda t}.$$

En resum, podem concloure que els tres conceptes següents són equivalents:

Temps en un estat exponencial \Leftrightarrow Procés de Márkov \Leftrightarrow Procés sense memòria

2.3 Processos de naixement i mort

Els processos de naixement i mort s'utilitzen per modelar la grandària d'una població; d'aquí ve el seu nom. Quan diem que el procés $N(t)$ està en l'estat k , és a dir, $N(t) = k$, estem indicant que a l'instant t la població està formada per k membres. Els processos de naixement i mort són un cas particular dels processos de Márkov, els quals es caracteritzen pel fet que els canvis d'estat només es poden produir entre estats veïns, és a dir, en un instant donat la grandària de



la població només pot augmentar o disminuir en una unitat. Per exemple, si estem modelant els estudiants que hi ha en una aula com un procés de naixement i mort, això significa que els estudiants sempre entren o surten de l'aula d'un en un.

Aquests processos també s'utilitzen en la teoria de cues per modelar el nombre de clients que hi ha en una cua. Les cues són uns sistemes que disposen d'un o diversos servidors per donar servei als clients que hi arriben. Quan els servidors estan tots ocupats i arriben nous clients, aquests s'han d'esperar. El lloc on s'esperen els clients que no poden ser atesos pot ser de capacitat finita o infinita. Un cop finalitzat el servei d'un client, aquest abandona el sistema i, si n'hi ha algun altre que espera ser atès, ocupa el servidor que ha quedat lliure. El nombre de clients en el sistema $N(t)$ depèn del temps entre les arribades dels clients i el temps que es triga quan se'ls dona servei. Atès que el temps entre arribades i el temps de servei són variables aleatòries, el nombre de clients dins el sistema és un procés estocàstic. Són exemples d'aquests sistemes les caixes d'un supermercat o el sistema de transmissió d'un node de commutació de paquets. Per al cas d'un sistema de transmissió, els clients que arriben són paquets de dades i el servei que volen obtenir del sistema és la seva transmissió.

Comencem pel sistema més senzill possible, format únicament per una zona d'espera infinita (no hi ha servidors). En aquest cas, els clients arriben al sistema i no en marxen mai, de manera que la població creix indefinidament.

Figura 5.
Sistema de
naixement pur



Això es coneix com un **procés de naixement pur**. En aquest cas, el temps de permanència en un estat coincideix amb el temps entre arribades. Per tal que el procés sigui markovià, cal que el temps entre arribades sigui una variable aleatòria exponencial. Aquesta variable aleatòria la denotem com t_{ia} (*interarrival time*).

$$f_{t_{ia}}(t) = \lambda e^{-\lambda t}; t \geq 0$$

El temps mitjà entre arribades val:

$$T_{ia} = E(t_{ia}) = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

Això vol dir que, de mitjana, el nombre d'arribades cada segon val λ . Aquest valor s'anomena **taxa d'arribada** i les seves unitats són *arribades/segon*, o, simplement, s^{-1} . Per al cas d'un sistema de transmissió, les unitats seran *paquets/segon*. Com es veu a les equacions de dalt, la taxa d'arribada és l'únic paràmetre que cal conèixer per definir el procés d'arribades.

Els processos de naixement i mort es representen mitjançant un diagrama d'estats. Els cercles són els estats i les fletxes indiquen les possibles transicions d'un estat a un altre.

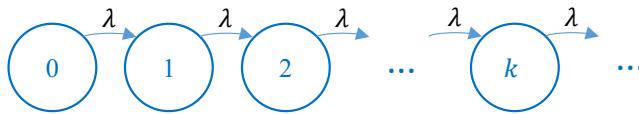


Figura 6.
Diagrama d'estats
d'un procés de naixe-
ment pur amb una
taxa d'arribada
constant

En el diagrama d'estats d'un procés de naixement pur, la fletxa que surt de l'estat k només va a l'estat $k + 1$. El fenomen que provoca que el procés canviï d'estat és l'arribada d'un client al sistema. La probabilitat amb què es produueix una arribada quan el sistema està en un estat depèn de la taxa d'arribades a l'estat. Això s'indica posant la taxa d'arribades damunt la fletxa que surt de l'estat. Es pot donar el cas que les taxes d'arribades siguin diferents en cada estat, tal com es mostra a la figura següent, on λ_k és la taxa d'arribada a l'estat k .

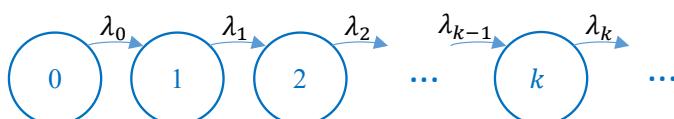


Figura 7.
Diagrama d'estats
d'un procés de
naixement pur

Si afegim un servidor al sistema, la població augmenta quan hi ha una arribada i disminueix quan algun client marxa del sistema, és a dir, un cop ha finalitzat el seu servei. A més, suposem que ambdós esdeveniments són independents, és a dir, que el temps entre arribades de clients és independent del seu temps de servei.



Figura 8.
Sistema de naixe-
ment i mort

Suposem que el temps de servei t_s també és una variable aleatòria exponencial:

$$f_{t_s}(t) = \mu e^{-\mu t}; t \geq 0$$

El temps mitjà de servei val:

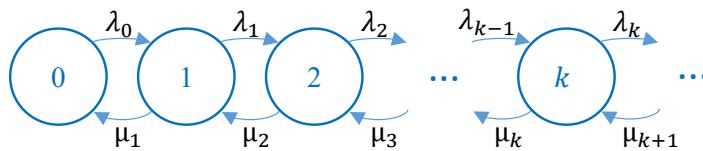
$$T_s = E(t_s) = \int_0^\infty t \mu e^{-\mu t} dt = \frac{1}{\mu}$$



on μ és la **taxa de servei** i les seves unitats són *paquets/segon* o, simplement, s^{-1} . La taxa de servei és el nombre mitjà de clients que pot servir el servidor cada segon. Atès que el servidor no sempre està ocupat, la taxa de servei no es correspon amb el nombre mitjà de clients servits per segon. Si el sistema és estable, el nombre de clients servits per segon ha de ser igual al nombre de clients que hi arriben per segon λ . Per exemple, la caixa d'un supermercat pot servir un client cada minut. La taxa de servei de la caixa és 60 clients/hora. Quan l'observem, veiem que la caixa queda aturada durant 10 minuts cada hora pel fet que no té clients. Això vol dir que el nombre mitjà de clients servits cada hora és 50, que són els que hi arriben. Si hi arribessin més de 60 clients cada hora, la caixa no ho podria suportar. En aquest cas, diríem que el sistema és inestable. Cal dir que, si la cua és finita, és a dir, si no permetem que els clients entrin a la cua quan està plena, aleshores el sistema mai no serà inestable.

De manera genèrica, la taxa de servei pot ser diferent en cada estat; per tant, el diagrama d'estats per a un procés de naixement i mort és el següent:

Figura 9.
Diagrama d'estats
per a un sistema de
naixement i mort



on:

- λ_k és la taxa d'arribades a l'estat k .
- μ_k és la taxa de servei a l'estat k .

Per poder afirmar que en el sistema descrit el nombre d'unitats $N(t)$ és un procés de naixement i mort, ens cal demostrar que el procés és de Márkov. El que sabem és que tant el temps entre arribades com el temps de servei són variables aleatòries exponencials, però encara hem de demostrar que el temps de permanència en un estat també és una variable aleatòria exponencial. Per fer-ho, calculem la **CDF** del temps de permanència a l'estat k .

$$F_{t_{e_k}}(t) = p(t_{e_k} \leq t) = 1 - p(t_{e_k} > t)$$

Quan el sistema està a l'estat k , pot canviar d'estat per dos motius:

- Hi ha una arribada i el sistema passa a l'estat $k + 1$.
- Finalitza un servei i la unitat servida marxa. En aquest cas, el sistema passa a l'estat $k - 1$.

Per tant, el temps de permanència serà més gran que t quan ambdós esdeveniments triguin més de t segons a produir-se:

$$F_{t_{ek}}(t) = 1 - p[(\text{temps fins a l'arribada} > t) \cap (\text{temps fins a la finalització del servei} > t)]$$

Quan el sistema entra a l'estat k , el temps que manca fins a la propera arribada és una variable aleatòria que té la mateixa **PDF** que el temps entre arribades t_{ia_k} ja que, en ser exponencial, el temps entre arribades no té memòria. Podem fer el mateix raonament per al temps que manca fins que finalitzi el servei següent, que tindrà la mateixa **PDF** que el temps de servei t_{sk} . Aleshores:

$$F_{t_{ek}}(t) = 1 - p[(t_{ia_k} > t) \cap (t_{sk} > t)]$$

Com que el temps entre arribades i el de servei són independents, tenim:

$$F_{t_{ek}}(t) = 1 - \int_t^{\infty} \lambda_k e^{-\lambda_k \xi} d\xi \cdot \int_t^{\infty} \mu_k e^{-\mu_k \xi} d\xi = 1 - e^{-(\lambda_k + \mu_k)t}$$

Finalment, calculem la **PDF** del temps de permanència derivant la **CDF** i n'obtem una funció exponencial, tal com volíem demostrar:

$$f_{t_{ek}}(t) = (\lambda_k + \mu_k) e^{-(\lambda_k + \mu_k)t}$$

El temps mitjà de permanència a l'estat k és la inversa de la suma de la taxa d'arribada i la taxa de servei de l'estat k :

$$E(t_{ek}) = \frac{1}{\lambda_k + \mu_k}$$

Càlcul de les probabilitats d'estat

Com ja s'ha dit, per a un procés estocàstic no es pot conèixer el valor del procés en un instant determinat, però sí que podem conèixer la probabilitat que el procés prengui un valor determinat en aquest instant. Definim $P_k(t)$ com la probabilitat que el procés de naixement i mort estigui a l'estat k en un instant t :

$$P_k(t) = p[N(t) = k]$$

Per calcular $P_k(t)$, primer calculem quina és la probabilitat de tenir una arribada en un Δt :

$$p(\text{arribada en } \Delta t) = p(t_{ia} < t + \Delta t \mid t_{ia} > t) = 1 - e^{-\lambda \Delta t}$$

Com que Δt és un valor petit, utilitzem el desenvolupament en sèrie de l'exponencial:²

² L'expressió $o(\Delta t)$ representa una funció que tendeix a 0 més ràpidament que Δt , és a dir, $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$.



$$e^{-\lambda \Delta t} = 1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2} - \frac{(\lambda \Delta t)^3}{3!} + \dots = 1 - \lambda \Delta t + o(\Delta t)$$

Per tant:

$$p(\text{arribada en } \Delta t) = \lambda \Delta t + o(\Delta t)$$

En cas que la taxa d'arribada sigui diferent a cada estat, tindrem una probabilitat diferent per a cadascun:

$$p(\text{arribada en } \Delta t | \text{sistema a l'estat } k) = \lambda_k \Delta t + o(\Delta t)$$

Podem fer el mateix raonament per calcular la probabilitat que un servei finalitzi durant l'interval Δt :

$$p(\text{fi del servei en } \Delta t | \text{sistema a l'estat } k) = \mu_k \Delta t + o(\Delta t).$$

Calculem ara $P_k(t + \Delta t)$. El sistema només podrà estar a l'estat k , a l'instant $t + \Delta t$, si s'ha donat alguna d'aquestes tres situacions:

- A l'instant t , estava a l'estat $k - 1$ i s'ha produït una arribada durant l'interval Δt .
- A l'instant t estava a l'estat $k + 1$ i ha finalitzat un servei durant l'interval Δt .
- A l'instant t estava a l'estat k i no ha arribat cap unitat ni ha finalitzat cap servei durant l'interval Δt .

Per tant, aplicant el teorema de la probabilitat total, tenim:

$$\begin{aligned} P_k(t + \Delta t) &= p(\text{arribada en } \Delta t | \text{sistema a l'estat } k - 1) P_{k-1}(t) + \\ &\quad + p(\text{fi del servei en } \Delta t | \text{sistema a l'estat } k + 1) P_{k+1}(t) + \\ &\quad + p(\text{ni arribada ni servei en } \Delta t | \text{sistema a l'estat } k) P_k(t). \end{aligned}$$

Substituem els valors de les probabilitats i posem junts els termes que van multiplicats per Δt .

$$P_k(t + \Delta t) = \lambda_{k-1} \Delta t P_{k-1}(t) + \mu_{k+1} \Delta t P_{k+1}(t) + [1 - (\lambda_k + \mu_k) \Delta t] P_k(t) + o(\Delta t),$$

$$\begin{aligned} P_k(t + \Delta t) - P_k(t) &= \\ &= \lambda_{k-1} \Delta t P_{k-1}(t) + \mu_{k+1} \Delta t P_{k+1}(t) - (\lambda_k + \mu_k) \Delta t P_k(t) + o(\Delta t) \end{aligned}$$

Dividint ambdós membres per Δt i prenent el límit quan $\Delta t \rightarrow 0$, s'obté el sistema següent d'equacions diferencials:

$$\frac{dP_k(t)}{dt} = \lambda_{k-1} P_{k-1}(t) + \mu_{k+1} P_{k+1}(t) - (\lambda_k + \mu_k) P_k(t); \quad k > 0$$



L'estat $k = 0$ és especial ja que, en aquest estat, no es dona servei (no existeix ni μ_0 ni λ_{-1}). Tenint això en compte, es pot fer el mateix raonament que per a la resta d'estats i s'obté:

$$\frac{dP_0(t)}{dt} = \mu_1 P_1(t) - \lambda_0 P_0(t).$$

Podem interpretar aquest sistema d'equacions diferencials utilitzant les definicions següents:

- Flux entrant a l'estat $k = \lambda_{k-1} P_{k-1}(t) + \mu_{k+1} P_{k+1}(t)$
- Flux sortint de l'estat $k = (\lambda_k + \mu_k) P_k(t)$

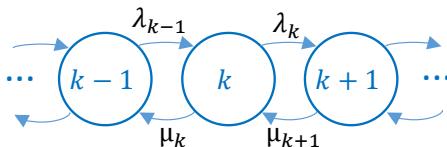


Figura 10.
Flux entrant i sortint
de l'estat k

Aleshores, podem dir que “la variació de la probabilitat de l'estat k respecte del temps és igual al flux entrant menys el flux sortint”.

2.3.1 Procés de Poisson

Resolem els sistema d'equacions diferencials per al cas més senzill possible: un procés de naixement pur amb taxa d'arribada igual per a tots els estats ($\lambda_k = \lambda$; $\mu_k = 0$). En aquest cas, el sistema d'equacions diferencials és:

$$\frac{dP_k(t)}{dt} = \lambda P_{k-1}(t) - \lambda P_k(t); \quad k > 0$$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

L'equació corresponent a l'estat 0 és de variables separables:

$$\frac{dP_0(t)}{P_0(t)} = -\lambda dt$$

$$\ln(P_0(t)) = -\lambda t + C$$

$$P_0(t) = Ae^{-\lambda t}$$

Suposem que inicialment el sistema està buit: $P_0(0) = 1$ i $P_k(0) = 0, \forall k \neq 0$. Amb aquestes condicions iniciales, s'obté:

$$P_0(t) = e^{-\lambda t}$$



Substituïm aquest resultat en el sistema d'equacions per tal de calcular $P_1(t)$:

$$\frac{dP_1(t)}{dt} = \lambda e^{-\lambda t} - \lambda P_1(t)$$

- Solució homogènia: $P_{1H}(t) = Ae^{-\lambda t}$ (es calcula igual que $P_0(t)$)
- Solució particular: $P_{1P}(t) = Bt \cdot e^{-\lambda t}$
- $P_1(t) = P_{1H}(t) + P_{1P}(t) = (A+Bt)e^{-\lambda t}$

Aplicant la condició inicial $P_1(0) = 0$, s'obté $P_1(t) = Bt e^{-\lambda t}$. Finalment, substituint aquesta expressió a l'equació diferencial, tenim:

$$P_1(t) = \lambda t e^{-\lambda t}$$

Podríem seguir així de forma indefinida i trobaríem que la solució general del sistema d'equacions val:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

És fàcil demostrar la validesa d'aquest resultat per inducció, substituint l'expressió en el sistema d'equacions diferencials i comprovant que aquesta també és vàlida per als casos $k = 0, 1$ calculats anteriorment.

El valor de $P_k(t)$ és la probabilitat de tenir k unitats a l'instant t per a un procés de naixement pur amb taxa constant. Com que només hi ha arribades, $P_k(t)$ també indica la probabilitat que s'hagin produït k arribades entre 0 i t , ja que hem suposat que inicialment estava buit. A més, com que es tracta d'un procés sense memòria, aquesta definició es pot fer extensiva a qualsevol interval de durada t ; per tant, $P_k(t)$ és la probabilitat que s'hagin produït k arribades en un interval de durada t .

Demostrem que realment es tracta d'un procés sense memòria comprovant que el temps de permanència en un estat és una variable aleatòria exponencial. En aquest cas, el temps de permanència en un estat coincideix amb el temps entre arribades t_{ia} . En calculem la **CDF**:

$$F_{t_{ia}}(t) = p(t_{ia} \leq t) = 1 - p(t_{ia} > t)$$

Sabem que $p(t_{ia} > t)$ és la probabilitat de tenir 0 arribades en t segons, és a dir, $P_0(t)$; per tant:

$$F_{t_{ia}}(t) = 1 - P_0(t) = 1 - e^{-\lambda t}$$

expressió que es correspon amb la **CDF** d'una variable aleatòria distribuïda exponencialment, tal com es volia demostrar.

La distribució de probabilitats $P_k(t)$ es coneix com la **distribució de Poisson**.

Aquesta distribució és un bon model per a molts fenòmens naturals o socials i és la que utilitzarem en molts casos per modelar les arribades a un sistema de transmissió.

La distribució de Poisson també s'obté a partir d'un **procés de Poisson**, que es defineix de la manera següent:

- Suposem que dividim un interval de temps de durada T en petits subintervals Δt i que, a cadascun d'aquests subintervals, podem tenir una arribada al nostre sistema en les condicions següents:
 1. $p(1\text{ arribada en } \Delta t) = p = \lambda \Delta t + o(\Delta t)$.
 2. $p(0\text{ arribades en } \Delta t) = 1 - p = 1 - \lambda \Delta t + o(\Delta t)$.
 3. El que passi en un Δt és independent de la resta (passats o futurs).

Notar que la condició 1 està relacionada amb el fet que el temps entre arribades és exponencial, la condició 2 implica que és un procés de naixement (ja que en Δt no hi pot haver més d'una arribada) i la condició 3 és imprescindible perquè el procés no tingui memòria.

Amb el procés definit així, en calcular la probabilitat de tenir k arribades en T segons quan Δt tendeix a 0, s'obté la distribució de Poisson, tal com es justifica a continuació. Com que tots els Δt són independents entre ells, podem calcular aquesta probabilitat com a proves repetides de Bernoulli, essent el nombre de proves $n = T / \Delta t$. Per tant, la probabilitat de tenir k arribades en n intents val:

$$p(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Quan prenem el límit per $t \rightarrow 0$, tenim:

$$\begin{aligned} p &\rightarrow 0 \\ n &\rightarrow \infty \\ np &\rightarrow \lambda T \end{aligned}$$

Aleshores, en calcular el límit, tenim:

$$\lim_{\Delta t \rightarrow 0} p(X_n = k) = \lim_{\Delta t \rightarrow 0} \frac{n(n-1)\cdots(n-k+1)}{k!n^k} (np)^k \left[(1-p)^{\frac{1}{p}} \right]^{np} (1-p)^{-k},$$

$$\lim_{t \rightarrow 0} p(X_n = k) = \frac{1}{k!} (\lambda T)^k e^{-\lambda T}$$



Aquesta descripció del procés de Poisson ens permet entendre a partir de quines hipòtesis estem modelant el procés d'arribades a un sistema de transmissió.

Estudiem ara algunes propietats dels processos de Poisson. Comencem calculant el nombre mitjà d'arribades en un interval T :

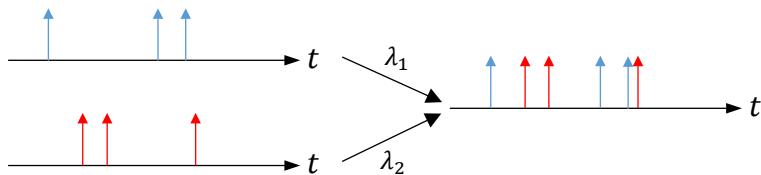
$$E(k) = \sum_{k=0}^{\infty} k P_k(T) = \sum_{k=0}^{\infty} k \frac{(\lambda T)^k}{k!} e^{-\lambda T} = \lambda T e^{-\lambda T} \sum_{k=1}^{\infty} \frac{(\lambda T)^{k-1}}{(k-1)!} = \lambda T e^{-\lambda T} e^{\lambda T}$$

$$E(k) = \lambda T$$

Aquí es torna a veure que λ és el nombre mitjà d'arribades per unitat de temps.

Mirem ara quina és la distribució d'arribades quan combinem dos processos de Poisson independents, amb taxes λ_1 i λ_2 .

Figura 11.
Agregació de fluxos
d'arribades



La probabilitat de tenir k arribades en t segons és la probabilitat de tenir i arribades del procés 1 i $k - i$ arribades del procés 2. Atès que, per a diferents valors de i , els casos són disjunts i que els processos 1 i 2 són independents, tenim:

$$\begin{aligned} P_k(t) &= \sum_{i=0}^k P_{1i}(t) P_{2k-i}(t) = \sum_{i=0}^k \frac{(\lambda_1 t)^i}{i!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{k-i}}{(k-i)!} e^{-\lambda_2 t} = \\ &= \frac{e^{-(\lambda_1 + \lambda_2)t}}{k!} t^k \sum_{i=0}^k \frac{k!}{i!(k-i)!} (\lambda_1)^i (\lambda_2)^{k-i} = \\ &= \frac{e^{-(\lambda_1 + \lambda_2)t}}{k!} t^k \sum_{i=0}^k \binom{k}{i} (\lambda_1)^i (\lambda_2)^{k-i} = \frac{e^{-(\lambda_1 + \lambda_2)t}}{k!} t^k (\lambda_1 + \lambda_2)^k \\ P_k(t) &= \frac{[(\lambda_1 + \lambda_2)t]^k}{k!} e^{-(\lambda_1 + \lambda_2)t} \end{aligned}$$

S'obté una distribució de Poisson de taxa $\lambda = \lambda_1 + \lambda_2$. Aplicant aquest resultat de manera recursiva, tenim que la combinació de n processos de Poisson és un nou procés de Poisson que té com a taxa la suma de les taxes dels n processos. Aquesta propietat serà molt útil a l'hora de multiplexar diferents fluxos d'informació sobre un mateix sistema de transmissió, ja que podrem modelar tots aquests fluxos com un únic flux.



Suposem ara que les arribades d'un procés de Poisson es reparteixen entre dos processos de manera que, quan hi arriba una unitat amb probabilitat α s'envia al procés 1 i amb probabilitat $(1 - \alpha)$, al procés 2.

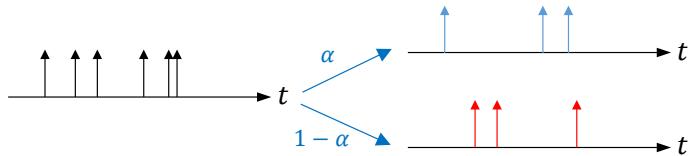


Figura 12.
Separació d'un flux
d'arribades

Volem calcular quina és la probabilitat de tenir k arribades a l'instant t per a cada un d'aquests dos processos. Per al procés 1, tenim que només se li poden haver assignat k arribades si, en total, hi han arribat almenys k unitats. Aplicant el teorema de la probabilitat total, tenim:

$$P_{1k}(t) = p[N_1(t) = k] = \sum_{n=k}^{\infty} p[N_1(t) = k | N(t) = n] p[N(t) = n]$$

La probabilitat d'haver enviat k arribades al procés 1 quan s'han produït en total n arribades la podem calcular a partir de les proves repetides de Bernoulli:

$$p[N_1(t) = k | N(t) = n] = \binom{n}{k} \alpha^k (1 - \alpha)^{n-k}$$

A més, com que el procés original és de Poisson:

$$p[N(t) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Substituint:

$$\begin{aligned} P_{1k}(t) &= \sum_{n=k}^{\infty} \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \\ &= \sum_{n=k}^{\infty} \frac{n!}{k! (n-k)!} \alpha^k (1 - \alpha)^{n-k} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \\ &= \frac{(\alpha \lambda t)^k}{k!} e^{-\lambda t} \sum_{n=k}^{\infty} \frac{[(1 - \alpha) \lambda t]^{n-k}}{(n-k)!} = \frac{(\alpha \lambda t)^k}{k!} e^{-\lambda t} e^{(1-\alpha)\lambda t} \\ P_{1k}(t) &= \frac{(\alpha \lambda t)^k}{k!} e^{-\alpha \lambda t} \end{aligned}$$

i s'obté un nou procés de Poisson de taxa $\lambda_1 = \alpha \lambda$. Fent el mateix raonament, es demostra que el procés 2 és de Poisson amb taxa $\lambda_2 = (1 - \alpha) \lambda$. Cadascun d'aquests dos processos es podria tornar a dividir, de manera que podríem tenir



tantes divisions del procés original com volguéssim. En general, si les arribades d'un procés de Poisson de taxa λ es reparteixen entre n processos, de manera que la probabilitat d'assignació d'una arribada al procés i val α_i , veurem que s'obtenen n processos de Poisson de taxa $\lambda_i = \alpha_i\lambda$. En aquest cas, també es compleix que $\lambda = \sum_{i=1}^n \lambda_i$, ja que estem suposant que $\sum_{i=1}^n \alpha_i = 1$.

A continuació, es presenta un resum de les propietats dels processos de Poisson:

- La probabilitat de tenir k arribades en un interval t val $P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$.
- És un procés sense memòria; per tant, el temps entre arribades t_{ia} és una variable aleatòria exponencial i $E(t_{ia}) = 1/\lambda$.
- Les arribades són igualment probables a tots els instants i la probabilitat de tenir una arribada en Δt val $\lambda \Delta t + o(\Delta t)$.
- El nombre mitjà d'arribades en T segons val λT .
- La combinació de n processos de Poisson amb taxes $\lambda_1, \lambda_2, \dots, \lambda_n$, és un procés de Poisson de taxa $\lambda = \sum_{i=1}^n \lambda_i$.
- Si les arribades d'un procés de Poisson de taxa λ s'assignen a un procés i (on $i \in \{1, \dots, n\}$) amb probabilitat α_i (sent $\sum_{i=1}^n \alpha_i = 1$), es generen n processos que també són de Poisson amb taxa $\lambda_i = \alpha_i\lambda$.

2.4 Processos de naixement i mort en equilibri

Al resoldre el sistema d'equacions diferencials, obtenim les probabilitats d'estat com una funció que varia en el temps. Això no significa necessàriament que el sistema es comporti d'una manera diferent segons l'instant en què sigui observat. En moltes ocasions, els sistemes tenen un comportament estable. El que realment passa és que les $P_k(t)$ parteixen d'una situació inicial (per exemple, al principi el sistema està buit) i, a partir d'aquí, tendeixen a un valor estable, a mesura que el temps va transcorrent. Així doncs, al marge del seu comportament transitori, ens interessa saber el valor de les probabilitats d'estat en règim permanent. Per tant, definim:

$$p_k = \lim_{t \rightarrow \infty} P_k(t)$$

Malgrat que no ho demostrarem formalment, la condició per tal que existeixi aquest límit quan el sistema té una cua infinita és la següent:

$$\forall k > k_0, \frac{\lambda_k}{\mu_k} < 1$$

És a dir, a partir d'un determinat estat k_0 , la taxa d'arribades ha de ser inferior a la taxa a la qual pot servir el sistema. Sembla evident que, si la taxa amb què

arriben clients al sistema és superior a la taxa amb què se'ls pot donar servei, el sistema no podrà assolir una situació estable en règim permanent. Un sistema que compleix la condició d'existència de les probabilitats d'estat en règim permanent s'anomena *ergòdic*. Això significa que aquestes probabilitats es poden mesurar de dues maneres diferents:

- Observem el sistema n vegades de forma aleatòria, de manera que el trobem n_k cops en l'estat k ; aleshores:

$$p_k = \lim_{n \rightarrow \infty} \frac{n_k}{n}$$

- Observem el sistema durant un interval T i mesurem t_k , el temps durant el qual el sistema està en l'estat k ; aleshores:

$$p_k = \lim_{T \rightarrow \infty} \frac{t_k}{T}$$

Tot i que ambdues interpretacions són correctes, utilitzarem principalment la segona. Per tant, direm que p_k indica la fracció de temps en què el sistema es troba a l'estat k .

Els sistemes que estudiem són ergòdics i les probabilitats d'estat tendeixen a un valor fix en règim permanent. En aquest cas, el sistema d'equacions diferencials es converteix en un sistema d'equacions lineals, ja que la derivada s'anul·la, atès que, en règim permanent, la probabilitat no varia amb el temps. Ho podem expressar dient que el flux entrant en un estat és igual al flux sortint.

$$\lambda_{k-1} p_{k-1} + \mu_{k+1} p_{k+1} = (\lambda_k + \mu_k) p_k; \quad k > 0$$

$$\mu_1 p_1 = \lambda_0 p_0$$

N'obtenim un sistema d'equacions lineals més fàcil de resoldre que el sistema d'equacions diferencials. Tenim tantes equacions com estats; per tant, a priori, hauríem de poder calcular totes les p_k . En aquest cas, el sistema d'equacions és linealment dependent, de manera que ens cal una equació addicional per poder resoldre'l. Aquesta equació ens la dona el fet que les probabilitats d'estat han de sumar 1:

$$\sum_{k=0}^{\infty} p_k = 1$$

2.4.1 Solució general d'equilibri

Comencem solucionant el sistema prenent l'equació corresponent a l'estat 0:

$$\mu_1 p_1 = \lambda_0 p_0$$



$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

Com que el sistema d'equacions és linealment dependent, posem les probabilitats d'estat en funció de p_0 . A continuació, substituïm el resultat anterior a l'equació que es correspon amb l'estat 1 i calculem p_2 en funció de p_0 :

$$\lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1) p_1$$

$$\mu_2 p_2 = \lambda_1 p_1$$

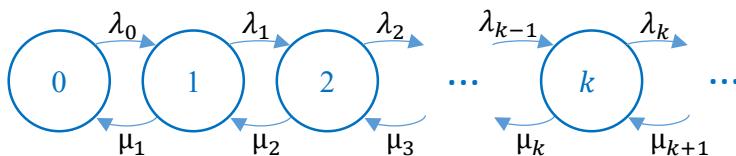
$$p_2 = \frac{\lambda_1}{\mu_2} p_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0$$

Seguint aquest procediment fins a un estat k genèric, tindrem:

$$p_k = \frac{\lambda_{k-1} \cdot \lambda_{k-2} \cdots \lambda_1 \cdot \lambda_0}{\mu_k \cdot \mu_{k-1} \cdots \mu_2 \cdot \mu_1} p_0 = p_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}; \quad k > 0$$

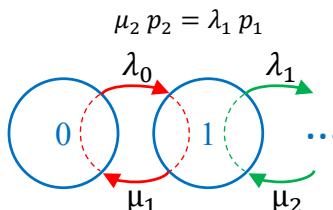
Aquesta expressió es pot demostrar per inducció substituint el terme general de p_{k-1} i p_k a l'equació i comprovant que se n'obté el terme general de p_{k+1} . Una manera fàcil de recordar aquest resultat és observant el diagrama d'estats, ja que la relació entre les probabilitats dels estats 0 i k és el producte de les λ_i , dividit pel producte de les μ_i que hi ha entre aquests dos estats.

Figura 13.
Diagrama d'estats
per a un sistema de
naixement i mort



També podem obtenir la solució general d'equilibri mirant la figura 14 i fent el raonament següent: si el flux que entra a l'estat 0 és igual al flux que en surt, aleshores el flux que surt de l'estat 1 per la seva esquerra ha de ser igual al flux que li entra per l'esquerra ($\mu_1 p_1 = \lambda_0 p_0$). Així doncs, el flux entrant a l'estat 1 per la dreta només pot ser igual al sortint de l'estat 1 per la dreta:

Figura 14.
Flux entrant i sortint
dels estats 0 i 1



Obtenim el mateix resultat si sumem les equacions corresponents als estats 0 i 1, ja que els fluxos entrants són al primer membre de les equacions i els fluxos

sortints, al segon. En sumar-les, els fluxos que surten d'un estat i entren a l'altre apareixen als dos membres de la nova equació i s'anulen.

Continuant amb aquest raonament fins a arribar a l'estat k , tindrem:

$$\mu_k p_k = \lambda_{k-1} p_{k-1}$$

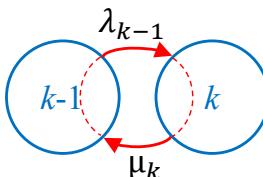


Figura 15.
Relació de fluxos entre els estats $k - 1$ i k

És com si en mig de cada estat hi hagués una frontera, de manera que els fluxos d'una banda no es poden barrejar amb els fluxos de l'altra. Aquest resultat també s'obté si sumem les k primeres equacions del sistema (des de l'estat 0 fins a l'estat $k - 1$). Aïllant, tenim:

$$p_k = \frac{\lambda_{k-1}}{\mu_k} p_{k-1}$$

Si apliquem aquesta equació de manera recursiva fins a p_0 , obtenim la solució general de p_k que hem vist abans. De fet, si no arribem fins a p_0 , tenim que la relació entre les probabilitats de dos estats qualssevol és el producte de les λ_i , dividit pel producte de les μ_i que hi ha entre els dos quan mirem el diagrama d'estats.

Finalment, per completar la solució, ens cal trobar el valor de p_0 . Ho fem utilitzant l'equació $\sum_{k=0}^{\infty} p_k = 1$:

$$p_0 \sum_{k=0}^{\infty} \frac{p_k}{p_0} = 1$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1} \cdot \lambda_{k-2} \cdots \lambda_1 \cdot \lambda_0}{\mu_k \cdot \mu_{k-1} \cdots \mu_2 \cdot \mu_1} \right]^{-1}$$

2.5 Sistemes d'espera

Per definir les característiques d'un sistema de cues, s'utilitza la notació de Kendall, que consisteix en un conjunt ordenat de paràmetres, separats per una barra inclinada "/". Cada paràmetre defineix una característica del sistema. A continuació, s'indica el significat de cada paràmetre:

1. Distribució del temps entre arribades
2. Distribució del temps de servei



3. Quantitat de servidors
4. Capacitat màxima del sistema (nombre màxim de clients que suporta el sistema)
5. Grandària de la població que vol accedir al sistema

Els paràmetres 3, 4 i 5 prenen valors sencers. Concretament, si els paràmetres 4 o 5 no s'indiquen, és que prenen un valor infinit. Els paràmetres 1 i 2 prenen principalment els valors següents:

- M: markovià; indica que la distribució de la variable és exponencial.
- D: determinista; indica que la variable sempre pren un valor fix.
- G: genèric; per a qualsevol distribució de la variable.

Per exemple, el sistema M/M/5 és un sistema pel qual el temps entre arribades i el temps de servei són exponencials, té cinc servidors, una cua infinita i hi accedeix una població també infinita.

Comencem analitzant sistemes d'espera que disposen de prou espai a la cua d'espera per tal que qualsevol client que arribi al sistema hi pugui entrar fins que li arriba el torn d'obtenir-ne un servidor. Per a un sistema de transmissió seria equivalent a pensar que el sistema té un *buffer* infinit, de manera que no hi ha pèrdues de paquets. A més, tant les arribades com el servei són exponencials i, per tant, podem analitzar aquests sistemes utilitzant la solució general d'equilibri dels processos de Mèrkov. Segons la

notació de Kendall, aquests sistemes són del tipus M/M/m.

2.5.1 Sistema M/M/1 El sistema M/M/1 té les característiques següents:

- El temps entre arribades és una variable aleatòria exponencial.
- El temps de servei és una variable aleatòria exponencial.
- Només hi ha un servidor.
- La cua és infinita.

Figura 16.
Sistema M/M/1



Com que les arribades són de Poisson i només hi ha un servidor, la taxa d'arribada i la taxa de servei són iguals a tots els estats. A més, com que la cua és



infinita, el sistema té infinitos estats. Si dibuixem el diagrama d'estats, tenim el següent:

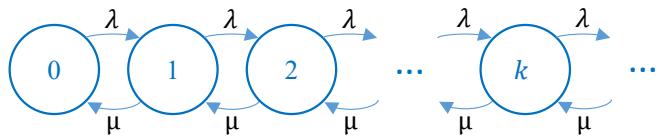


Figura 17.
Diagrama d'estats
del sistema M/M/1

A partir del diagrama, es poden obtenir les probabilitats d'estat utilitzant la solució general en equilibri.

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} = \frac{\lambda^k}{\mu^k} p_0 = \rho^k p_0$$

Hem definit ρ com el quocient entre les taxes d'arribada i de servei. Com es detalla més endavant, aquesta ρ és el factor d'utilització del sistema, que ens indica la relació entre el temps en què el sistema està treballant i el temps disponible del sistema per poder treballar. Calculem p_0 :

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^{\infty} \rho^k \right]^{-1}$$

Per tal que les p_k no siguin nul·les, s'ha de complir $p_0 \neq 0$. És a dir, la suma de la progressió geomètrica que apareix a la fórmula de dalt ha de ser convergent, cosa que es compleix si la seva raó ρ és menor que la unitat:

$$p_k \neq 0 \Rightarrow p_0 \neq 0 \Rightarrow \sum_{k=0}^{\infty} \rho^k < \infty \Rightarrow \rho < 1$$

Suposem que es compleix aquesta condició, la qual cosa vol dir que estem resolent el sistema per al cas que la taxa d'arribada sigui menor a la taxa de servei ($\lambda < \mu$). Cal recordar que aquesta és la condició d'existència de les probabilitats d'estat p_k en règim permanent. En aquest cas, si calculem la suma de la progressió geomètrica, tenim:

$$p_0 = \left[\frac{1}{1 - \rho} \right]^{-1} = 1 - \rho$$

Finalment:

$$p_k = (1 - \rho) \rho^k$$

A partir d'aquest resultat, podem calcular diferents paràmetres del sistema.



Nombre mitjà d'unitats al sistema

$$\begin{aligned}
 N &= \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k (1 - \rho) \rho^k = (1 - \rho) \rho \sum_{k=0}^{\infty} k \rho^{k-1} = (1 - \rho) \rho \sum_{k=0}^{\infty} \frac{d}{dp} \rho^k = \\
 &= (1 - \rho) \rho \frac{d}{dp} \left[\sum_{k=0}^{\infty} \rho^k \right] = (1 - \rho) \rho \frac{d}{dp} \left[\frac{1}{1 - \rho} \right] = \frac{(1 - \rho) \rho}{(1 - \rho)^2} \\
 N &= \frac{\rho}{1 - \rho}
 \end{aligned}$$

Observeu que, si la quantitat d'arribades tendeix a zero ($\rho \rightarrow 0$), el sistema es buida ($N \rightarrow 0$). D'altra banda, si la taxa d'arribades s'aproxima a la taxa de servei ($\rho \rightarrow 1$), el sistema esdevé inestable ($N \rightarrow \infty$).

Nombre mitjà d'unitats al servidor

$$N_s = 0 p_0 + 1(1 - p_0) = 1 - p_0$$

$$N_s = \rho$$

Nombre mitjà d'unitats en espera

$$N_w = 0 p_0 + 0 p_1 + 1 p_2 + 2 p_3 + \dots = \sum_{k=1}^{\infty} (k - 1) p_k = \sum_{k=0}^{\infty} k p_k - \sum_{k=1}^{\infty} p_k = N - N_s$$

Se n'obté un resultat trivial, el nombre mitjà d'unitats a tot el sistema és igual al nombre mitjà d'unitats en cua, més el nombre mitjà d'unitats al servidor:

$$N = N_w + N_s$$

Finalment:

$$N_w = N - N_s = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

Fórmula de Little

La fórmula de Little relaciona el nombre mitjà d'unitats en un sistema amb el temps mitjà d'estada d'una unitat en el sistema. Mirem com calcular aquests dos paràmetres a partir d'una realització del procés $N(t)$.

Al procés de la figura 18, a cada flanc de pujada hi ha una arribada i a cada flanc de baixada, la finalització d'un servei. Els requadres sota de la corba $N(t)$ que tenen el mateix farciment corresponen al temps en què una unitat és al sistema. El número dins el requadre indica l'ordre d'arribada de la unitat. A l'exemple, la unitat 3 hi arriba quan hi ha dues unitats al sistema i marxa deixant-hi una unitat. És per això que els requadres que representen la unitat 3 ocupen l'interval des



que el sistema entra a l'estat 3 fins que retorna a l'estat 1. L'àrea d'aquests requadres és el temps que la unitat 3 és al sistema. Per tant, des d'aquest punt de vista, l'àrea total sota $N(t)$ és la suma dels temps d'estada al sistema de totes les unitats.

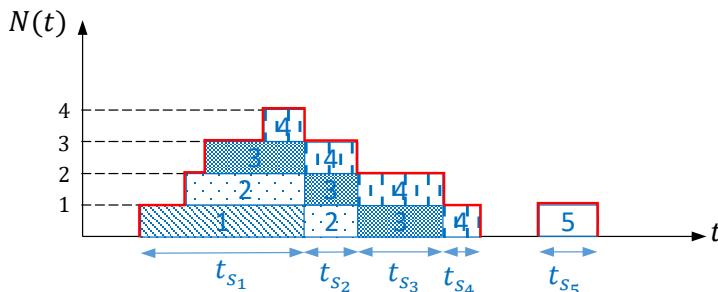


Figura 18.
Relació entre el nombre d'unitats i el temps d'estada al sistema

Si volem calcular el temps mitjà en què una unitat és al sistema, hem de dividir la suma de tots els temps pel nombre d'unitats que han passat pel sistema. Definim $A(t)$ com el procés que indica el nombre d'arribades al sistema en funció del temps. Aleshores, podem expressar el temps mitjà d'estada al sistema com:

$$T = \lim_{T_{obs} \rightarrow \infty} \frac{1}{A(T_{obs})} \int_{T_{obs}} N(t) dt$$

Ara mirem el procés $N(t)$ des d'un altre punt de vista. A la figura 19, l'altura de cada requadre indica el nombre d'unitats al sistema i la seva amplada en el temps durant el qual el sistema té aquest nombre d'unitats. La variable aleatòria t_k indica el temps que el sistema es manté a l'estat k . A l'exemple de la figura, el sistema passa tres vegades per l'estat 1 i, per tant, la variable t_1 pren tres valors diferents. Definim T_k com la suma d'aquests t_k mentre estem observant el sistema, és a dir, T_k és el temps en què el sistema és a l'estat k durant el període d'observació.

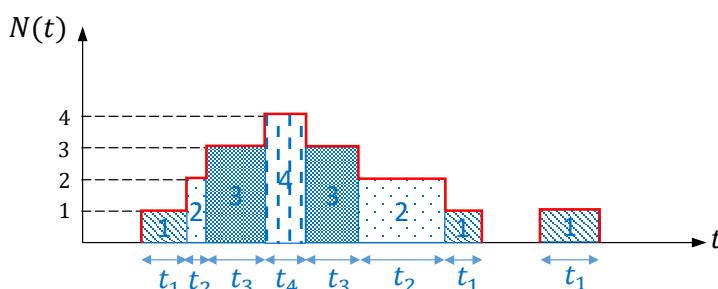


Figura 19.
Relació entre el nombre d'unitats i el temps en què el sistema és en un estat determinat

Si volem calcular el nombre mitjà d'unitats del sistema, hem de ponderar cada estat per la seva probabilitat, és a dir, per la fracció de temps en què el sistema és en aquest estat,

$$N = \lim_{T_{obs} \rightarrow \infty} \sum_{k=0}^{\infty} k \frac{T_k}{T_{obs}}$$



Però $\sum_{k=0}^{\infty} k T_k$ és l'àrea sota $N(t)$; per tant:

$$N = \lim_{T_{obs} \rightarrow \infty} \frac{1}{T_{obs}} \int_{T_{obs}} N(t) dt$$

Per relacionar el nombre mitjà d'unitats i el temps mitjà d'estada al sistema, calculem el quocient entre ambdues magnituds:

$$\frac{N}{T} = \lim_{T_{obs} \rightarrow \infty} \frac{A(T_{obs})}{T_{obs}}$$

Aquest límit és la definició de la taxa d'arribades λ ; així doncs, finalment obtenim la fórmula de Little:

$$N = \lambda T$$

Podem interpretar la fórmula de Little de la manera següent: quan un client arriba al sistema, veu que aquest té un nombre mitjà d'unitats N . Després d'haver estat al sistema, ha de deixar el mateix nombre mitjà d'unitats, atès que el sistema és estable. Aquest nombre mitjà d'unitats que deixa és el nombre mitjà d'arribades mentre el client ha estat al sistema i val λT .

Podem aplicar la fórmula de Little independentment de la distribució del temps entre arribades, la distribució del temps de servei i el nombre de servidors. L'única condició per poder aplicar-la és que el sistema sigui estable, és a dir, que la taxa que entra al sistema sigui igual a la taxa que en surt. Evidentment, no en pot sortir més del que hi entra i, si en surt menys, el sistema esdevé inestable en règim permanent. No s'ha de confondre la taxa que surt del sistema amb la taxa de servei del sistema. El sistema no està servint tota l'estona, ja que pot quedar buit ($p_0 > 0$). La taxa de sortida serà igual a la taxa de servei, multiplicada per la fracció de temps en què el sistema està servint $\lambda = \mu(1 - p_0)$.

Temps mitjà al sistema

Utilitzem la fórmula de Little per calcular el temps mitjà que una unitat es manté al sistema M/M/1:

$$T = \frac{N}{\lambda} = \frac{\rho}{1 - \rho} \frac{1}{\lambda} = \frac{1/\mu}{1 - \rho}$$

Expressem aquest temps en funció del temps mitjà de servei, $T_s = 1/\mu$.

$$T = \frac{T_s}{1 - \rho}$$

Temps mitjà al servidor

Podem aplicar Little al subsistema format únicament pel servidor, el qual té una taxa d'entrada i de sortida igual a λ . En aquest cas, se n'obté un resultat

trivial; el temps mitjà d'estada d'una unitat al servidor és igual al temps mitjà de servei T_s .

Temps mitjà d'espera

El temps mitjà d'espera és el temps mitjà en què una unitat és a la cua. Apliquem la fórmula de Little al subsistema format únicament per la cua i calculem el temps mitjà d'espera.

$$T_w = \frac{N_w}{\lambda} = \frac{\rho^2}{1 - \rho} \frac{1}{\lambda} = \frac{\rho}{1 - \rho} T_s$$

Factor d'utilització

El factor d'utilització del sistema es defineix com la relació entre el temps en què els servidors estan ocupats i el temps màxim en què aquests ho podrien estar. Podem entendre el primer terme com el treball que fa el sistema (mesurant el treball en unitats de temps) i el segon terme com el treball màxim que pot fer. Així doncs, si observem un sistema durant T_{obs} i aquest serveix n_T unitats, el temps durant el qual els servidors han estat ocupats és igual a la suma dels temps de servei de les unitats servides:

$$\sum_{i=1}^{n_T} t_{s_i}$$

Aquest valor és aproximat, ja que hi podria haver unitats parcialment servides que no estaríem comptabilitzant, però serà més exacte com més gran sigui el temps d'observació, de manera que, en el límit, quan el temps d'observació tendeix a infinit, podem considerar-lo exacte. Cal notar que, per a un sistema amb més d'un servidor, aquest temps podria ser més gran que el temps d'observació T_{obs} .

D'altra banda, el temps màxim en què els servidors poden estar ocupats val $m T_{obs}$, essent m el nombre de servidors del sistema.

Per tant, podem calcular el factor d'utilització com:

$$\rho = \lim_{T_{obs} \rightarrow \infty} \frac{\sum_{i=1}^{n_T} t_{s_i}}{m T_{obs}} = \frac{1}{m} \lim_{T_{obs} \rightarrow \infty} \frac{n_T}{T_{obs}} \frac{\sum_{i=1}^{n_T} t_{s_i}}{n_T}$$

Cal notar que, en el límit, el terme n_T/T_{obs} és la taxa cursada pel sistema i l'altre terme és el temps mitjà de servei; per tant, finalment tenim:

$$\rho = \frac{\lambda T_s}{m}$$

Si només tenim un servidor, el factor d'utilització val $\rho = \lambda T_s = \lambda/\mu$. El valor ρ que hem definit en calcular les probabilitats d'estat per a un sistema M/M/1 és, doncs, el factor d'utilització.



Exemple: Un mecanògraf rep $\lambda = 15$ pàgines manuscrites/hora per tal que les mecanografiï en un ordinador. El mecanògraf triga $T_s = 2$ minuts a mecanografiar cada pàgina. En calcular la utilització del mecanògraf, tenim:

$$\rho = 15 \frac{\text{pàgines}}{\text{hora}} 2 \frac{\text{minuts de feina}}{\text{pàgina}} = 30 \frac{\text{minuts de feina}}{\text{hora}} = 0,5 \frac{\text{hores de feina}}{\text{hora}}$$

Una utilització igual a 0,5 significa que només s'utilitza el 50% de la capacitat de treball del sistema. En cas que hi haguessin dos mecanògrafs, el factor d'utilització seria 0,25, cosa que significa que, si repartim la feina d'una manera igual entre tots dos, només estan treballant el 25% del temps o, el que és el mateix, només s'aprofita el 25% de la capacitat de treball del sistema.

Cal notar que un sistema només pot ser estable si el treball que hi entra és inferior a la capacitat del sistema per realitzar-lo, és a dir, si $\rho < 1$.

Suposem un sistema genèric amb un únic servidor (un sistema G/G/1, segons la notació de Kendall). En aquest cas:

$$\rho = \lim_{T_{obs} \rightarrow \infty} \frac{\sum_{i=1}^{n_T} t_{s_i}}{T_{obs}}$$

és la fracció de temps durant la qual el sistema ha estat ocupat, és a dir, la fracció de temps durant la qual el sistema no està buit (almenys, hi ha una unitat). Per tant, tenim que el factor d'utilització també el podem expressar com:

$$\rho = 1 - p_0$$

Així doncs, el resultat que hem trobat per a un sistema M/M/1 es pot generalitzar per a un sistema G/G/1.

A més, per a un sistema G/G/1, el factor d'utilització també és el nombre mitjà d'unitats al servidor:

$$N_s = 0 p_0 + 1(1 - p_0) = 1 - p_0 = \rho$$

Poisson Arrivals See Time Average: PASTA

A l'hora d'avaluar diferents característiques dels sistemes de cues, moltes vegades ens interessarà saber com està el sistema quan hi arriba una unitat. És a dir, volem conèixer quina és la probabilitat que una unitat que arriba a l'instant t trobi el sistema a l'estat k . Aquesta probabilitat la denotarem com $R_k(t)$. Si definim $A(t, t + \Delta t)$ com l'esdeveniment "arribada al sistema entre t i $t + \Delta t$ ", podem expressar aquesta probabilitat de la manera següent:

$$R_k(t) = \lim_{\Delta t \rightarrow 0} p[N(t) = k | A(t, t + \Delta t)]$$



Igual com passa amb les probabilitats d'estat, aquesta probabilitat s'estabilitza en règim permanent, de manera que es defineix:

$$r_k = \lim_{t \rightarrow \infty} R_k(t)$$

A priori, podríem pensar que $P_k(t)$ i $R_k(t)$ coincideixin, però això no és cert, ja que depèn de la manera com arribin les unitats al sistema. Podem comprovar-ho amb un contraexemple, el sistema D/D/1 pel qual el temps entre arribades i el temps de servei són deterministes.

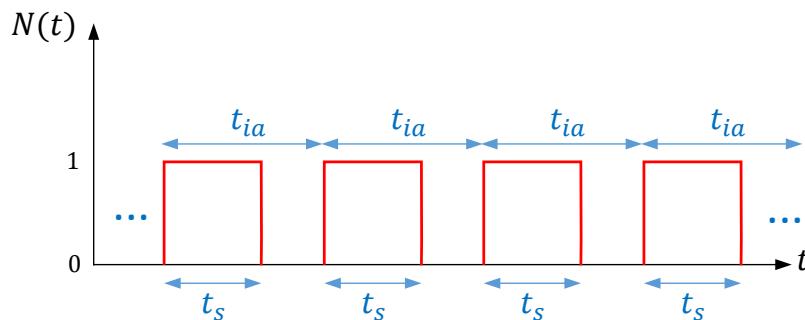


Figura 20. Realització del nombre d'unitats per a un sistema D/D/1

Tal com es veu a la figura 20, la condició perquè aquest sistema sigui estable és $t_{ia} > t_s$. Les inverses d'aquestes variables deterministes són la taxa d'arribada i la taxa de servei, respectivament; per tant, la condició d'estabilitat és $\lambda < \mu$ o, el que és el mateix, $\rho < 1$.³

Si calculem la fracció de temps en què el sistema està ocupat amb una unitat o la fracció de temps en què està buit, tenim:

$$p_1 = \frac{t_s}{t_{ia}} = \frac{\lambda}{\mu} = \rho ; \quad p_0 = 1 - \rho$$

D'altra banda, quan una unitat arriba al sistema, sempre el troba buit:

$$r_1 = 0; \quad r_0 = 1$$

Per tant, en aquest cas, és evident que $r_k \neq p_k$

Ara volem calcular $R_k(t)$ quan les arribades són de Poisson. Per fer-ho, primer apliquem el teorema de Bayes:⁴

$$R_k(t) = \lim_{\Delta t \rightarrow 0} p[N(t) = k | A(t, t + \Delta t)] = \lim_{\Delta t \rightarrow 0} \frac{p[A(t, t + \Delta t) | N(t) = k] p[N(t) = k]}{p[A(t, t + \Delta t)]}$$

³ El sistema D/D/1 és l'únic amb un servidor que continua essent estable quan $\rho = 1$.

⁴ $p(A|B)p(B) = p(B|A)p(A)$.



Quan el temps entre arribades és exponencial, la probabilitat de tenir una arribada quan el sistema és a l'estat k val:

$$p[A(t, t + \Delta t) | N(t) = k] = \lambda_k \Delta t + o(\Delta t)$$

Si, a més, les arribades són de Poisson, la taxa és la mateixa per a tots els estats ($\lambda_k = \lambda; \forall k$) i, per tant:

$$p[A(t, t + \Delta t) | N(t) = k] = p[A(t, t + \Delta t)] = \lambda \Delta t + o(\Delta t)$$

Substituint, s'obté:

$$R_k(t) = p[N(t) = k] = P_k(t)$$

I, en el límit, quan $t \rightarrow \infty$:

$$r_k = p_k$$

Trobem, doncs, que les arribades de Poisson “veuen” que la probabilitat de trobar el sistema a l'estat k coincideix amb la fracció de temps en què el sistema és a l'estat k . D'aquí ve el nom d'aquesta propietat: “*Poisson Arrivals See Time Average*”.

Càlcul de la **PDF** del temps de transferència en un sistema M/M/1

A partir de les probabilitats d'estat, hem pogut calcular el temps mitjà d'estada d'una unitat en un sistema M/M/1. Ara volem calcular quina és la **PDF** d'aquesta variable t_{tr} .

Prèviament ens cal conèixer quina és la **PDF** de la suma de k variables aleatòries exponencials i independents. Per fer aquest càlcul, definim:

$$t_k = \sum_{i=1}^k t_{s_i}; \quad f_{t_{s_i}}(t) = \mu e^{-\mu t}$$

Sabem que la **PDF** de la suma de dues variables aleatòries independents és la convolució de les seves **PDF**; per tant, per al cas $k = 2$, tenim:

$$f_{t_2}(t) = f_{t_{s_1}}(t) * f_{t_{s_2}}(t) = \int_0^t \mu e^{-\mu \xi} \mu e^{-\mu(t-\xi)} d\xi = \mu e^{-\mu t} \mu \int_0^t d\xi = \mu e^{-\mu t} \mu t$$

Aprofitem aquest resultat per calcular la **PDF** de la suma de tres exponencials ($k = 3$):

$$\begin{aligned} f_{t_3}(t) &= f_{t_{s_1}+t_{s_2}}(t) * f_{t_{s_3}}(t) = \int_0^t \mu e^{-\mu \xi} \mu \xi \mu e^{-\mu(t-\xi)} d\xi = \mu e^{-\mu t} \mu^2 \int_0^t \xi d\xi = \\ &= \mu e^{-\mu t} \frac{(\mu t)^2}{2} \end{aligned}$$

En general, per a k exponencials, tindrem:

$$f_{t_k}(t) = \mu e^{-\mu t} \frac{(\mu t)^{k-1}}{(k-1)!}$$

Aquesta funció es coneix com la funció de densitat de probabilitat d'Erlang. Utilitzarem aquest resultat per calcular la **PDF** del temps de transferència i del temps d'espera en un sistema M/M/1. Per fer-ho, mirem quant val el temps de transferència d'una unitat que arriba al sistema en funció del nombre d'unitats que troba n .

- Quan una arribada troba el sistema buit, el seu temps de transferència és el seu temps de servei: $t_{tr|n=0} = t_s$.
- Quan una arribada troba una unitat en el sistema, el seu temps de transferència és el temps residual de servei de la unitat que troba, més el seu propi temps de servei: $t_{tr|n=1} = tr_s + t_s$.
- Quan una arribada troba dues unitats en el sistema, el seu temps de transferència és el temps residual de servei de la unitat que troba en el servidor, més el temps de servei de la unitat que és a la cua, més el seu propi temps de servei: $t_{tr|n=2} = tr_s + t_{s_1} + t_{s_2}$.

En general, quan una arribada troba k unitats, el seu temps de transferència val:

$$t_{tr|n=k} = tr_s + \sum_{i=1}^k t_{s_i}$$

Com que el temps de servei és una variable aleatòria exponencial, no té memòria i, per tant, el temps residual de servei continua essent una variable aleatòria exponencial amb la mateixa **PDF**. Així doncs, podem dir que el temps de transferència és la suma de $k+1$ variables aleatòries exponencials iguals:

$$t_{tr|n=k} = t_{k+1}$$

$$f_{t_{tr|n=k}}(t) = f_{t_{k+1}}(t) = \mu e^{-\mu t} \frac{(\mu t)^k}{k!}$$

Continuant amb el càlcul de la **PDF** del temps de transferència, construïm ara una taula en què indiquem:

- n : nombre d'unitats que troba una arribada.
- $t_{tr|n}$: temps de transferència d'una unitat que, quan arriba, troba n unitats al sistema.
- $f_{t_{tr|n}}(t)$: **PDF** de $t_{tr|n}$.
- r_n : probabilitat d'arribar i trobar n unitats al sistema. Per la propietat PASTA, que hem vist anteriorment, $r_n = p_n$.



n	$t_{tr n}$	$f_{t_{tr n}}(t)$	r_n
0	t_s	$\mu e^{-\mu t}$	$(1 - \rho)$
1	$tr_s + t_s$	$\mu e^{-\mu t} \mu t$	$(1 - \rho)\rho$
2	$tr_s + \sum_{i=1}^2 t_{s_i}$	$\mu e^{-\mu t} \frac{(\mu t)^2}{2}$	$(1 - \rho)\rho^2$
\vdots	\vdots	\vdots	\vdots
k	$tr_s + \sum_{i=1}^k t_{s_i}$	$\mu e^{-\mu t} \frac{(\mu t)^k}{k!}$	$(1 - \rho)\rho^k$

Aplicant el teorema de la probabilitat total, calculem la **PDF** del temps de transferència:

$$f_{t_{tr}}(t) = \sum_{k=0}^{\infty} f_{t_{tr|n=k}}(t) r_k = \sum_{k=0}^{\infty} \mu e^{-\mu t} \frac{(\mu t)^k}{k!} (1 - \rho)\rho^k = (1 - \rho)\mu e^{-\mu t} \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!},$$

$$f_{t_{tr}}(t) = (1 - \rho)\mu e^{-(1-\rho)\mu t}$$

Així doncs, veiem que el temps de transferència també és una variable aleatòria exponencial. Si calculem el temps mitjà de transferència, obtenim el mateix resultat que havíem vist a partir de les probabilitats d'estat:

$$T = T_{tr} = E(t_{tr}) = \frac{1}{(1 - \rho)\mu} = \frac{T_s}{(1 - \rho)}$$

Càlcul de la **PDF** del temps d'espera en un sistema M/M/1

Igual que hem fet per al temps de transferència, ara calculem la **PDF** del temps d'espera t_w . Mirem, en primer lloc, quant val aquest temps en funció del nombre d'unitats que troba una arribada en el sistema.

- Quan una arribada troba el sistema buit, no espera: $t_{w|n=0} = 0$.
- Quan una arribada troba una unitat al sistema, el seu temps d'espera és el temps residual de servei de la unitat que hi troba: $t_{w|n=1} = tr_s$.
- Quan una arribada troba dues unitats al sistema, el seu temps d'espera és el temps residual de servei de la unitat que troba al servidor, més el temps de servei de la unitat que és a la cua: $t_{w|n=2} = tr_s + t_{s_1}$.

En general, quan una arribada troba k unitats ($k > 0$), el seu temps d'espera val:

$$t_{w|n=k} = tr_s + \sum_{i=1}^{k-1} t_{s_i}$$

El temps residual de servei continua essent una variable aleatòria exponencial i, per tant, per a $k > 0$, el temps d'espera és la suma de k variables aleatòries exponencials, $t_{w|n=k} = t_k$. A més, la **PDF** d'una variable determinista és una delta de Dirac centrada en el valor de la variable i, aleshores, $f_{t_w|n=0}(t) = \delta(t)$.

Igual que hem fet per al temps de transferència, construïm la taula següent:

n	$t_{w n}$	$f_{t_w n}(t)$	r_n
0	0	$\delta(t)$	$(1 - \rho)$
1	tr_s	$\mu e^{-\mu t}$	$(1 - \rho)\rho$
2	$tr_s + t_s$	$\mu e^{-\mu t} \mu t$	$(1 - \rho)\rho^2$
\vdots	\vdots	\vdots	\vdots
k	$tr_s + \sum_{i=1}^{k-1} t_{s_i}$	$\mu e^{-\mu t} \frac{(\mu t)^{k-1}}{(k-1)!}$	$(1 - \rho)\rho^k$

Aplicant el teorema de la probabilitat total, calculem la **PDF** del temps d'espera:

$$\begin{aligned}
 f_{t_w}(t) &= \sum_{k=0}^{\infty} f_{t_w|n=k}(t) r_k = (1 - \rho)\delta(t) + \sum_{k=1}^{\infty} \mu e^{-\mu t} \frac{(\mu t)^{k-1}}{(k-1)!} (1 - \rho)\rho^k = \\
 &= (1 - \rho)\delta(t) + (1 - \rho) \rho \mu e^{-\mu t} \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} \\
 f_{t_w}(t) &= (1 - \rho)\delta(t) + \rho(1 - \rho)\mu e^{-(1-\rho)\mu t}
 \end{aligned}$$

Podem interpretar el resultat de la manera següent: per la propietat PASTA, $(1 - \rho)$ és la probabilitat que una arribada trobi el sistema buit i ρ la probabilitat que el trobi ple. Quan una arribada troba el sistema buit, el temps d'espera és nul i, si el troba ocupat, el temps d'espera és una variable aleatòria exponencial:

$$\begin{aligned}
 f_{t_w|n=0}(t) &= \delta(t) \\
 f_{t_w|n \neq 0}(t) &= (1 - \rho)\mu e^{-(1-\rho)\mu t} \\
 f_{t_w}(t) &= f_{t_w|n=0}(t) p(n=0) + f_{t_w|n \neq 0}(t) p(n \neq 0)
 \end{aligned}$$

Si calculem el temps mitjà d'espera, n'obtenim el mateix resultat que amb les probabilitats d'estat:

$$T_w = E(t_w) = (1 - \rho) \int_0^{\infty} t \delta(t) dt + \rho \int_0^{\infty} t (1 - \rho)\mu e^{-(1-\rho)\mu t} dt = \frac{\rho}{(1 - \rho)\mu} =$$



$$= \frac{\rho}{(1 - \rho)} T_s.$$

A partir de la funció de densitat de probabilitat, podem obtenir la funció de probabilitat acumulada:

$$\begin{aligned} F_{t_w}(t) &= \int_0^t (1 - \rho) \delta(\tau) d\tau + \int_0^t \rho(1 - \rho)\mu e^{-(1-\rho)\mu\tau} d\tau = \\ &= (1 - \rho) + \rho(1 - \rho)\mu \int_0^t e^{-(1-\rho)\mu\tau} d\tau = 1 - \rho e^{-(1-\rho)\mu t} \end{aligned}$$

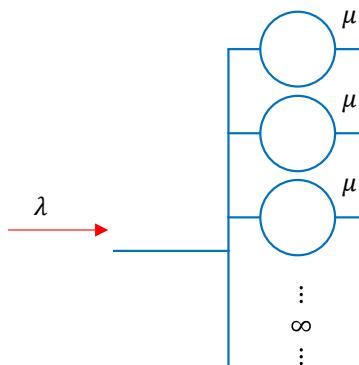
I la funció complementària:

$$p[t_w > t] = 1 - p[t_w \leq t] = 1 - F_{t_w}(t) = \rho e^{-(1-\rho)\mu t}$$

2.5.2 Sistema M/M/ ∞

Modelem ara un sistema en el qual podem considerar que el nombre de serveis és suficientment gran perquè qualsevol nova sol·licitud de transmissió pugui ser atesa directament per un servidor. Per tant, podríem considerar que aquest sistema està format per un nombre infinit de servidors, de manera que, quan una unitat arriba al sistema, sempre hi troba un servidor buit (no cal *buffer*).

Figura 21.
Sistema M/M/ ∞



Quan només tenim un servidor, el temps entre sortides quan el servidor està ocupat és el temps de servei. Ara ens cal conèixer com és el temps entre sortides quan hi ha k servidors ocupats. Definim t_{id_k} com el temps entre sortides quan hi ha k servidors ocupats⁵ i en calculem la *CDF*:

⁵ El subíndex *id* indica *interdeparture*, t_{id} és l'*interdeparture time*.



$$F_{t_{id_k}}(t) = p(t_{id_k} \leq t) = 1 - p(t_{id_k} > t) = \\ = 1 - p[(t_{s_1} > t) \cap (t_{s_2} > t) \cap \dots \cap (t_{s_k} > t)]$$

Per a k servidors independents:

$$F_{t_{id_k}}(t) = 1 - \prod_{i=1}^k p(t_{s_i} > t)$$

Com que els temps de servei són exponencials, $p(t_{s_i} > t) = e^{-\mu_i t}$. Suposem que tots els servidors són iguals $\mu_i = \mu$; aleshores:

$$F_{t_{id_k}}(t) = 1 - \prod_{i=1}^k e^{-\mu t} = 1 - e^{-k\mu t}$$

Per tant, la seva **PDF** val:

$$f_{t_{id_k}}(t) = k\mu e^{-k\mu t}$$

N'obtenim que el temps entre sortides és una variable aleatòria exponencial. Això vol dir que tenir k servidors ocupats és equivalent a tenir un únic servidor que treballa k vegades més ràpid, ja que la seva taxa de servei és, en aquest cas, $k\mu$.

A partir d'aquest resultat, podem dibuixar el diagrama d'estats i calcular-ne les probabilitats d'estat:

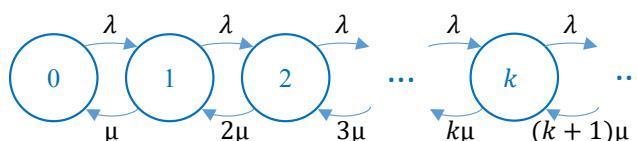


Figura 22.
Diagrama d'estats
del sistema M/M/∞

$$p_k = \frac{(\lambda/\mu)^k}{k!} p_0$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^{\infty} \frac{(\lambda/\mu)^k}{k!} \right]^{-1} = e^{-\lambda/\mu}$$



$$p_k = \frac{(\lambda/\mu)^k}{k!} e^{-\lambda/\mu}$$

El resultat obtingut coincideix amb la probabilitat de tenir k arribades de Poisson en un temps $1/\mu$. Podem aprofitar aquesta coincidència per calcular el nombre mitjà d'unitats en el sistema. Per un procés de Poisson, tenim que el nombre mitjà d'arribades en T segons val λT . En aquest cas, tenim que el nombre mitjà d'unitats en el sistema val:

$$N = \frac{\lambda}{\mu}$$

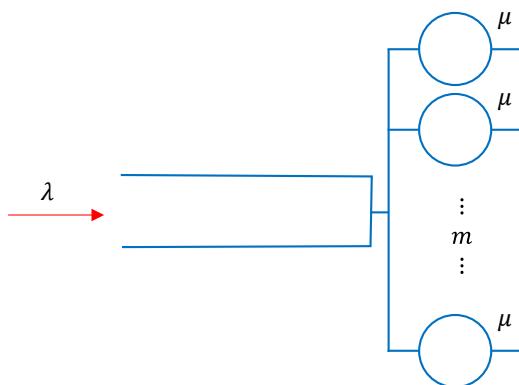
Aplicant Little a aquest resultat, s'obté que el temps mitjà d'estada al sistema val T_s , la qual cosa és evident.

Per aquest sistema, no té gaire sentit parlar de factor d'utilització, ja que la capacitat de treball del sistema és infinita ($m \rightarrow \infty$) i, per tant, la utilització del sistema sempre és 0.

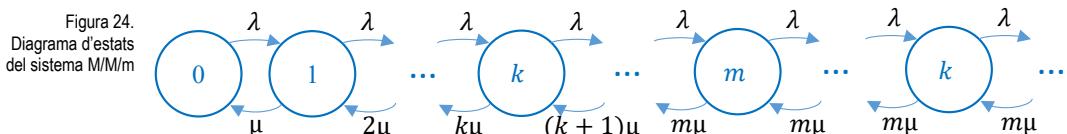
2.5.3 Sistema M/M/m

Aquest és un sistema d'espera amb un nombre finit de servidors.

Figura 23.
Sistema M/M/m



A continuació, modelem un sistema que disposa d'un nombre finit m de servidors. Per construir el diagrama de transició d'estats, en aquest cas cal distingir entre els estats per als quals hi ha servidors lliures ($k < m$) i els estats per als quals tots els servidors estan ocupats ($k \geq m$).





A partir del diagrama d'estats, i utilitzant l'expressió coneguda per als processos de naixement i mort en equilibri, trobem les probabilitats d'estat:

$$p_k = \begin{cases} \frac{(\lambda/\mu)^k}{k!} p_0 = \frac{(m\rho)^k}{k!} p_0; & k < m \\ \frac{(\lambda/\mu)^k}{m! m^{k-m}} p_0 = \frac{m^m}{m!} \rho^k p_0; & k \geq m \end{cases}$$

essent:

$$\rho = \frac{\lambda}{m\mu}$$

Finalment, trobem p_0 :

$$p_0 = \left[\sum_{k=0}^{\infty} p_k \right]^{-1} = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{m^m}{m!} \sum_{k=m}^{\infty} \rho^k \right]^{-1} = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{(1-\rho)m!} \right]^{-1}$$

El valor $A = m\rho = \lambda/\mu$ és la quantitat de treball que entra al sistema per unitat de temps (segons de feina/segon). A l'estudi del teletrànsit de les xarxes de comunicacions, aquest valor s'anomena *trànsit ofert* i es mesura en erlangs.

Probabilitat de demora

Un paràmetre d'interès en aquest sistema és la probabilitat que una arribada s'hagi d'esperar. Aquest valor es coneix com a *probabilitat de demora*. Aquesta situació es produeix sempre que una unitat en arribar troba tots els servidors ocupats. Aplicant PASTA, tenim:

$$P_D = \sum_{i=m}^{\infty} r_i = \sum_{i=m}^{\infty} p_i = \frac{m^m}{m!} p_0 \sum_{i=m}^{\infty} \rho^i = \frac{(m\rho)^m}{(1-\rho)m!} p_0$$

Substituint el valor de p_0 :

$$P_D = \frac{\frac{(m\rho)^m}{(1-\rho)m!}}{\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{(1-\rho)m!}}$$

Aquesta expressió es coneix com la fórmula Erlang C i s'utilitza per dimensionar sistemes amb espera. Per exemple, una centraleta telefònica en què m indica el nombre de línies telefòniques; T_s , la durada mitjana d'una conversa, i λ , les peticions per unitat de temps que fan els usuaris (trucades/hora). Es pot dimensionar el nombre de línies telefòniques per tal que la probabilitat de demora



quedi fitada a un valor determinat. El problema de la fórmula Erlang C és que resulta impossible aïllar m ; per tant, el que es fa és tabular P_D en funció de m i el trànsit ofert. Com ja s'ha comentat, el trànsit ofert és la taxa amb la qual la feina arriba al sistema (segons d'ocupació de les línies telefòniques/segon), es mesura en erlangs i es calcula:

$$A = \frac{\lambda}{\mu} = \lambda T_s = m\rho$$

La condició perquè el sistema sigui estable és que el trànsit ofert sigui inferior al trànsit que pot cursar el sistema:

$$A < m \Leftrightarrow \rho < 1$$

Un sistema M/M/1 és un cas particular d'un sistema M/M/m. Si calculem la probabilitat de demora per al cas $m = 1$, s'obté que aquesta val ρ , com era d'esperar, ja que, tenint en compte la propietat PASTA, ρ és la probabilitat d'arribar i trobar el sistema M/M/1 ocupat.

A continuació, fem diferents càlculs per aquest sistema i n'expresssem el resultat en funció de P_D . Això fa que la interpretació dels resultats sigui molt més intuïtiva i, a la vegada, evita haver de manipular expressions matemàtiques feixugues.

Nombre mitjà d'unitats i temps mitjà d'estada als servidors

Com que cada unitat només és atesa per un servidor, el temps mitjà d'estada d'una unitat en el conjunt format pels m servidors és el seu temps de servei. Aleshores, aplicant Little:

$$N_s = \lambda T_s = m\rho = A$$

També podem calcular el nombre mitjà d'unitats que hi ha al conjunt format pels m servidors de la manera següent:

$$N_s = \sum_{k=0}^{m-1} k p_k + m \sum_{k=m}^{\infty} p_k = \sum_{k=0}^{m-1} k \frac{(m\rho)^k}{k!} p_0 + m \sum_{k=m}^{\infty} \frac{m^m}{m!} \rho^k p_0$$

Arranjant els sumatoris i substituint el valor de p_0 , finalment també se n'obté:

$$N_s = m\rho = A$$

Nombre mitjà d'unitats en espera i temps mitjà d'espera

Quan el sistema té menys de m unitats, ningú no espera i, quan el nombre d'unitats k és mes gran o igual a m , hi ha $k - m$ unitats en espera; per tant:

$$\begin{aligned}
 N_w &= \sum_{k=m}^{\infty} (k-m)p_k = \sum_{k=m}^{\infty} (k-m) \frac{m^m}{m!} \rho^k p_0 = \frac{m^m}{m!} p_0 \sum_{k=0}^{\infty} k \rho^{k+m} = \\
 &= \frac{(m\rho)^m}{m!} p_0 \sum_{k=0}^{\infty} k \rho^k
 \end{aligned}$$

El sumatori que apareix a l'expressió anterior l'hem resolt en el càlcul del nombre mitjà d'unitats en un sistema M/M/1. Aprofitant el resultat, tenim:

$$N_w = \frac{(m\rho)^m}{m!} p_0 \frac{\rho}{(1-\rho)^2}$$

Si ho expressem en funció de la probabilitat de demora, finalment se n'obté:

$$N_w = \frac{\rho}{1-\rho} P_D = \frac{A}{m-A} P_D$$

Observeu que, si $m = 1$, s'obté el mateix resultat que per al sistema M/M/1.

Alicant la fórmula de Little, calculem el temps mitjà d'espera:

$$T_w = \frac{N_w}{\lambda} = \frac{T_s}{m(1-\rho)} P_D = \frac{T_s}{m-A} P_D$$

Nombre mitjà d'unitats al sistema i temps mitjà al sistema

$$N = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{m-1} k p_k + \sum_{k=m}^{\infty} k p_k$$

Sumant i restant el terme $\sum_{k=m}^{\infty} m p_k$, se n'obté un resultat trivial:

$$N = \sum_{k=0}^{m-1} k p_k + \sum_{k=m}^{\infty} m p_k + \sum_{k=m}^{\infty} (k-m) p_k = N_s + N_w$$

Aprofitant els càlculs de N_s i N_w , tenim:

$$N = m\rho + \frac{\rho}{1-\rho} P_D = A + \frac{A}{m-A} P_D$$

Alicant Little per calcular el temps mitjà d'estada al sistema, s'arriba a un resultat també evident:

$$T = T_s + T_w$$

Substituint:



$$T = \left(1 + \frac{P_D}{m(1-\rho)}\right) T_s = \left(1 + \frac{P_D}{m-A}\right) T_s$$

Càlcul de la **PDF** del temps d'espera en un sistema M/M/m

Per calcular la **PDF** del temps d'espera t_w , mirem quant val aquest temps en funció del nombre d'unitats que troba una unitat quan arriba al sistema.

- Quan una arribada troba el sistema amb menys de m unitats, no espera:

$$t_{w|n=0} = \dots = t_{w|n=m-1} = 0$$

- Quan una arribada troba m unitats al sistema, el seu temps d'espera és el temps residual entre sortides: $t_{w|n=m} = tr_{id}$
- Quan una arribada troba $m+1$ unitats al sistema, el seu temps d'espera és el temps residual entre sortides més un nou temps entre sortides: $t_{w|n=m+1} = tr_{id} + t_{id}$

En general, quan una arribada troba $m+k$ unitats, el seu temps d'espera val:

$$t_{w|n=m+k} = tr_{id} + \sum_{j=1}^k t_{id_j}$$

Com ja hem vist, quan hi ha m servidors ocupats, el temps entre sortides és una variable aleatòria exponencial amb:

$$f_{t_{id}}(t) = m\mu e^{-m\mu t}$$

El temps residual entre sortides quan hi ha m servidors ocupats també és una variable aleatòria exponencial amb la mateixa **PDF**. Així doncs, el temps d'espera d'una unitat que troba $m+k$ unitats quan arriba és la suma de $k+1$ variables aleatòries exponencials independents. Recordant la funció de densitat de probabilitat d'Erlang, tenim:

$$f_{t_{w|n=m+k}}(t) = m\mu e^{-m\mu t} \frac{(m\mu t)^k}{k!}; k \geq 0$$

Per al cas en què una arribada trobi menys de m servidors ocupats:

$$f_{t_{w|n=k}}(t) = \delta(t); k = 0, 1, \dots, m-1$$

Ara calculem la probabilitat amb què una arribada troba el sistema en cadascun dels estats descrits. Per PASTA, tenim que $r_k = p_k$. La probabilitat que una arribada trobi menys de m servidors ocupats val:



$$\sum_{k=0}^{m-1} p_k = 1 - P_D$$

La probabilitat que una arribada trobi $m+k$ servidors ocupats val:

$$p_{m+k} = \frac{m^m}{m!} p_0 \rho^{m+k} = p_m \rho^k$$

Relacionem p_m amb la probabilitat de demora:

$$P_D = \sum_{k=0}^{\infty} p_{m+k} = p_m \sum_{k=0}^{\infty} \rho^k = \frac{p_m}{1-\rho} \Rightarrow p_m = (1-\rho)P_D$$

Substituint p_m , finalment se n'obté:

$$p_{m+k} = (1-\rho) \rho^k P_D$$

Per aplicar el teorema de la probabilitat total, construïm una taula amb tot el que ja hem dit, essent:

- n : nombre d'unitats que troba una arribada
- $t_{w|n}$: temps d'espera d'una unitat que, en arribar, troba n unitats al sistema
- $f_{t_w|n}(t)$: PDF de $t_{w|n}$
- r_n : probabilitat d'arribar i trobar n unitats al sistema. Per PASTA, $r_n = p_n$

n	$t_{w n}$	$f_{t_w n}(t)$	r_n
$0, 1, \dots, m-1$	0	$\delta(t)$	$1 - P_D$
m	tr_{id}	$m\mu e^{-m\mu t}$	$(1-\rho)P_D$
$m+1$	$tr_{id} + t_{id}$	$m\mu e^{-m\mu t} m\mu t$	$(1-\rho)\rho P_D$
\vdots	\vdots	\vdots	\vdots
$m+k$	$tr_{id} + \sum_{j=1}^k t_{id_j}$	$m\mu e^{-m\mu t} \frac{(m\mu t)^k}{k!}$	$(1-\rho)\rho^k P_D$

Apliquem el teorema de la probabilitat total per calcular la PDF del temps d'espera:

$$\begin{aligned} f_{t_w}(t) &= \sum_{k=0}^{\infty} f_{t_w|n=k}(t) r_k = (1 - P_D) \delta(t) + \sum_{k=0}^{\infty} m\mu e^{-m\mu t} \frac{(m\mu t)^k}{k!} (1 - \rho) \rho^k P_D = \\ &= (1 - P_D) \delta(t) + P_D (1 - \rho) m\mu e^{-m\mu t} \sum_{k=0}^{\infty} \frac{(m\mu t)^k}{k!} \end{aligned}$$



$$f_{t_w}(t) = (1 - P_D)\delta(t) + P_D(1 - \rho)m\mu e^{-(1-\rho)m\mu t}$$

Podem fer la interpretació següent d'aquest resultat. El valor $1 - P_D$ és la probabilitat que una arribada trobi algun servidor lliure i P_D , la probabilitat que els trobi tots ocupats. Quan una arribada troba algun servidor lliure, el temps d'espera és nul i, si els troba tots ocupats, el temps d'espera és una variable aleatòria exponencial:

$$f_{t_w|n < m}(t) = \delta(t)$$

$$f_{t_w|n \geq m}(t) = (1 - \rho)m\mu e^{-(1-\rho)m\mu t}$$

$$f_{t_w}(t) = f_{t_w|n < m}(t) p(n < m) + f_{t_w|n \geq m}(t) p(n \geq m)$$

Si calculem el temps mitjà d'espera, n'obtenim el resultat que hem vist anteriorment:

$$\begin{aligned} T_w = E(t_w) &= (1 - P_D) \int_0^{\infty} t \delta(t) dt + P_D \int_0^{\infty} t(1 - \rho)m\mu e^{-(1-\rho)m\mu t} dt = \\ &= \frac{P_D}{(1 - \rho)m\mu} \end{aligned}$$

Aprofitant que coneixem la funció de densitat de probabilitat, en podem obtenir la funció de probabilitat acumulada:

$$\begin{aligned} F_{t_w}(t) &= \int_0^t (1 - P_D)\delta(\tau) d\tau + \int_0^t P_D(1 - \rho)m\mu e^{-(1-\rho)m\mu\tau} d\tau = \\ &= (1 - P_D) + P_D(1 - \rho)m\mu \int_0^t e^{-(1-\rho)m\mu\tau} d\tau = 1 - P_D e^{-(1-\rho)m\mu t} \end{aligned}$$

I la funció complementària:

$$p[t_w > t] = 1 - p[t_w \leq t] = 1 - F_{t_w}(t) = P_D e^{-(1-\rho)m\mu t}$$

2.6 Sistemes amb pèrdues

Suposem ara que la zona d'espera és finita, de manera que, si un client hi arriba i la troba totalment plena, no pot entrar al sistema. En aquest cas, parlem de sistemes amb pèrdues, ja que els clients que no poden entrar al sistema "es perden". En la notació de Kendall, indiquem amb un nou paràmetre la quantitat màxima de clients que admet el sistema. Per exemple, un sistema M/M/2/5 és un sistema que admet un màxim de cinc clients i, com que té dos servidors, la



zona d'emmagatzematge és per a tres clients. A continuació, estudiem diferents sistemes d'aquest tipus.

2.6.1 Sistema M/M/1/K

Aquest sistema està format per un servidor i una cua de grandària $Q = K - 1$. En ser un sistema on només caben K unitats, no totes les unitats que arriben al sistema poden entrar-hi i, per tant, tenim una taxa perduda (λ_p) i una taxa cursada (λ_c).

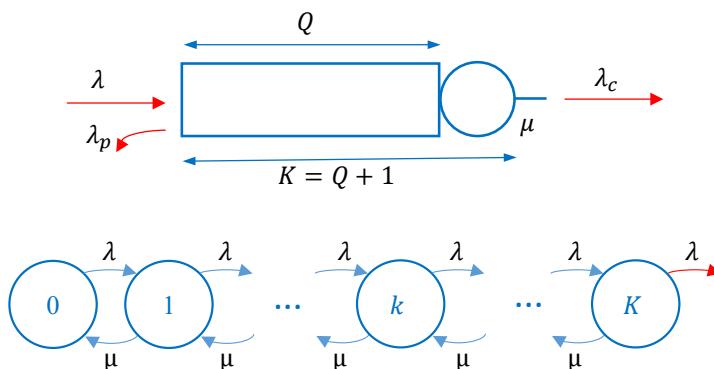


Figura 25.
Sistema M/M/1/K i el seu diagrama d'estats

En aquest cas, el diagrama d'estats està format per una cadena finita. A l'estat K , hi arriba una taxa λ , la qual no porta el sistema a cap nou estat. Tot i que aquest valor no ens és útil per calcular les probabilitats d'estat, ho indiquem al diagrama, ja que, com veurem més endavant, aquesta taxa està relacionada amb les pèrdues.

Calculem les probabilitats d'estat:

$$p_k = \frac{\lambda^k}{\mu^k} p_0 = \rho^k p_0; k \leq K$$

Hem definit ρ com el quocient entre les taxes d'arribada i de servei. Per a valors de $k > K$, la probabilitat p_k val 0.

Calculem p_0 :

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^K \rho^k \right]^{-1} = \frac{1 - \rho}{1 - \rho^{K+1}}$$

Observem que, a diferència dels sistemes estudiats amb cua infinita, en aquest cas no és necessari imposar la condició $\rho < 1$.

Finalment:



$$p_k = \frac{1 - \rho}{1 - \rho^{K+1}} \rho^k$$

Nombre mitjà d'unitats al sistema, a la cua i al servidor

El nombre mitjà d'unitats als diferents subsistemes es troba a partir de les probabilitats d'estat. A continuació, se n'indica el resultat, sense entrar en el detall del càclul. Pel nombre mitjà d'unitats al servidor, tenim:

$$N_s = 1 - p_0 = \frac{\rho - \rho^{K+1}}{1 - \rho^{K+1}}$$

El nombre mitjà d'unitats al sistema val:

$$N = \sum_{k=0}^K k p_k = \frac{\rho}{1 - \rho} - (K + 1) \frac{\rho^{K+1}}{1 - \rho^{K+1}}$$

I el nombre mitjà d'unitats a la cua és:

$$N_w = N - N_s = \frac{\rho}{1 - \rho} - \frac{\rho + K\rho^{K+1}}{1 - \rho^{K+1}}$$

Probabilitat de pèrdua, taxa cursada i taxa perduda

De la taxa que hi arriba, una part és cursada pel sistema, mentre que l'altra es perd. Parlem, doncs, de taxa oferta λ_{of} , taxa cursada λ_c i taxa perduda λ_p de manera que:

$$\lambda_{of} = \lambda = \lambda_c + \lambda_p$$

La probabilitat de pèrdua p_p és la probabilitat que una unitat que arriba al sistema no hi pugui entrar perquè el troba ple. Si les arribades són de Poisson, per la propietat PASTA tenim que la probabilitat que una arribada trobi el sistema ple coincideix amb la probabilitat que el sistema estigui ple:

$$p_p = r_K = p_K$$

D'altra banda, també podem calcular la probabilitat de pèrdua com el nombre d'unitats perdudes, dividit pel nombre d'arribades quan observem el sistema durant un temps molt llarg:

$$p_p = \lim_{T_{obs} \rightarrow \infty} \frac{\text{unitats perdudes}}{\text{unitats que hi arriben}} = \lim_{T_{obs} \rightarrow \infty} \frac{\text{unitats perdudes}/T_{obs}}{\text{unitats que hi arriben}/T_{obs}} = \frac{\lambda_p}{\lambda}$$

Per tant, podem calcular la taxa perduda i la taxa cursada:

$$\lambda_p = \lambda p_K$$

$$\lambda_c = \lambda - \lambda_p = \lambda(1 - p_K)$$

Pel sistema M/M/1/K, tenim:

$$\lambda_p = \lambda \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}$$

$$\lambda_c = \lambda \frac{1 - \rho^K}{1 - \rho^{K+1}}$$

Factor d'utilització per als sistemes amb pèrdues

El factor d'utilització es defineix com el treball que entra al sistema, dividit per la capacitat màxima de treball que pot fer el sistema. En ser un sistema amb pèrdues, no tot el treball que hi arriba hi pot entrar. Aleshores, per calcular el factor d'utilització per a un sistema amb pèrdues, hem de considerar només la taxa que cursa el sistema:

$$\rho_c = \frac{\lambda_c}{m\mu}$$

En enginyeria de teletrànsit, s'anomena *trànsit ofert* el quotient:

$$A = \frac{\lambda}{\mu}$$

Aquest valor indica el ritme amb el qual el treball arriba al sistema. Recordem que mesurem el treball amb unitats de temps; per tant, el trànsit ofert són els segons de feina que arriben a cada segon. El trànsit que aconsegueix entrar al sistema s'anomena *trànsit cursat* i val:

$$A_c = \frac{\lambda_c}{\mu}$$

Finalment, el trànsit que no aconsegueix entrar-hi s'anomena *trànsit perdut*.

La relació entre el factor d'utilització i el trànsit cursat és la següent:

$$\rho_c = \frac{\lambda_c}{m\mu} = \frac{A_c}{m}$$

Aquest factor d'utilització sempre és menor que 1. Els sistemes amb pèrdues són estables per si sols, ja que no accepten la feina que no poden fer. Malgrat que el trànsit ofert sigui més gran que el ritme màxim de treball que pot fer el sistema ($A > m \Leftrightarrow \rho > 1$), el trànsit cursat sempre és més petit que aquest ritme ($A_c < m \Leftrightarrow \rho_c < 1$).

En qualsevol sistema estable, el nombre mitjà d'unitats que hi entren per segon ha de ser igual al nombre mitjà d'unitats que en surten per segon. El nombre



mitjà d'unitats que hi entren per segon és la taxa cursada i ja hem vist que val $\lambda_c = \lambda(1 - p_K)$. Si el sistema només té un servidor, serveix a una taxa μ quan està ocupat i, per tant, el nombre mitjà d'unitats que en surten per segon val $\mu(1 - p_0)$. Noteu que $1 - p_0$ és la fracció de temps durant la qual el sistema està ocupat. Com que ambdues taxes són iguals, tenim:

$$\lambda(1 - p_K) = \mu(1 - p_0)$$

Aquest resultat el podem comprovar substituint els valors de p_0 i p_K obtinguts per un sistema M/M/1/K. Aïllant, s'obté que el factor d'utilització per a un sistema M/M/1/K és, igual que en els sistemes de cua infinita, la probabilitat que el sistema estigui ocupat:

$$\lambda_c = \mu(1 - p_0) \Rightarrow \rho_c = \frac{\lambda_c}{\mu} = (1 - p_0)$$

Temps d'estada al sistema i al servidor, i temps d'espera

La fórmula de Little relaciona el nombre mitjà d'unitats dins el sistema amb el temps d'estada de les unitats que aconsegueixen entrar-hi. La relació entre aquests dos paràmetres és la taxa d'unitats que entren al sistema, és a dir, la *taxa cursada*:

$$N = \lambda_c T$$

El temps mitjà al servidor val T_s com era d'esperar:

$$\frac{N_s}{\lambda_c} = \frac{\frac{\rho - \rho^{K+1}}{1 - \rho^{K+1}}}{\frac{\lambda}{1 - \rho^K} \frac{1 - \rho^K}{1 - \rho^{K+1}}} = \frac{1}{\mu} = T_s$$

Com es pot observar, hauria estat més fàcil partir d'aquest resultat trivial a l'hora de calcular N_s .

Per al càlcul del temps mitjà al sistema i del temps mitjà d'espera, utilitzem la fórmula de Little:

$$T = \frac{N}{\lambda_c}, \quad T_w = \frac{N_w}{\lambda_c}$$

Cas especial $\rho = 1$

Com ja s'ha indicat, el sistema M/M/1/K és estable, independentment del trànsit ofert. Això vol dir que ρ pot ser major o igual a 1. Per al cas $\rho = 1$, la solució de les probabilitats d'estat p_k és indeterminada. Podríem calcular el límit quan $\rho \rightarrow 1$, però és més fàcil tornar a calcular les p_k quan $\rho = 1$:



$$p_k = \rho^k p_0 = p_0 = \frac{1}{K+1}$$

Com es pot veure, tots els estats són equiprobables. En aquest cas:

$$N = \frac{K}{2}, N_s = \frac{K}{K+1}, N_w = \frac{K(K-1)}{2(K+1)} \text{ i } \lambda_c = \lambda \frac{K}{K+1}$$

2.6.2 Sistema M/M/m/m

Aquest és un sistema format per m servidors i sense *buffer*.

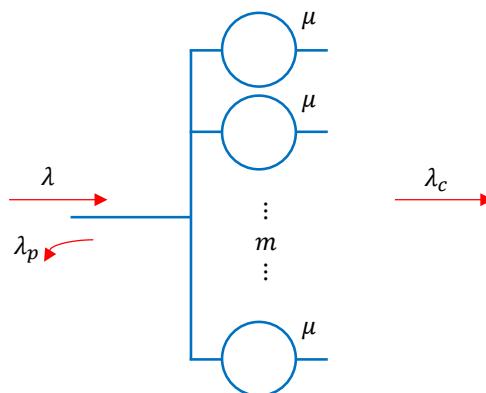


Figura 26.
Sistema M/M/m/m

La taxa de servei és proporcional al nombre de servidors ocupats i, per tant, tenim un diagrama d'estats finit de la forma:

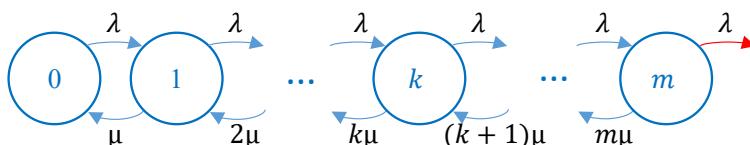


Figura 27.
Diagrama d'estats
del sistema M/M/m/m

Calculem les probabilitats d'estat:

$$p_k = \frac{(\lambda/\mu)^k}{k!} p_0$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^m \frac{(\lambda/\mu)^k}{k!} \right]^{-1}$$

La probabilitat que una arribada trobi tot el sistema ocupat val:



$$p_m = \frac{\frac{(\lambda/\mu)^m}{m!}}{\sum_{k=0}^m \frac{(\lambda/\mu)^k}{k!}}$$

Aquesta expressió es coneix com la fórmula d'Erlang B. Té la mateixa utilitat que la fórmula d'Erlang C, però en aquest cas es calcula la probabilitat de pèrdua, en comptes de la probabilitat d'espera.

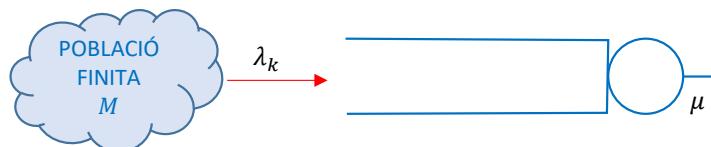
2.7 Sistemes amb població finita

Als sistemes amb població finita, el nombre de clients que hi volen accedir és finit. Això implica que el nombre de peticions sobre el sistema va disminuint a mesura que s'omple, ja que els clients que estan essent servits o en espera no generen noves peticions. En la notació de Kendall, indiquem a la cinquena posició quina és la grandària de la població que vol accedir al sistema. Per exemple, un sistema M/M/4/ ∞ /7 té 4 servidors, una zona d'espera infinita i només 7 clients poden utilitzar el sistema, de manera que quan tots 7 són a dins ja no es reben noves peticions. En aquest cas, la notació de Kendall també admet el format M/M/4//7 ja que, si no s'indica la grandària del sistema, se suposa infinita.

2.7.1 Sistema M/M/1//M

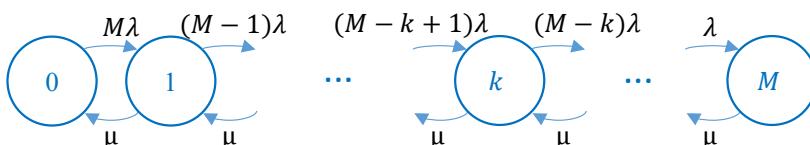
Per a aquest sistema, la grandària de la població val M . A més, suposem que cada membre de la població genera λ peticions/s.

Figura 28.
Sistema M/M/1//M



Dibuixem el diagrama d'estats i en calculem les probabilitats d'estat. El diagrama d'estats és finit, ja que la població també ho és, de manera que l'estat màxim es correspon amb el cas en què tots els membres de la població tenen una petició dins el sistema.

Figura 29.
Diagrama d'estats
del sistema M/M/1//M



En calculem les probabilitats d'estat:

$$p_k = \frac{\lambda^k M(M-1)(M-2)\cdots(M-k+1)}{\mu^k} p_0 = \left(\frac{\lambda}{\mu}\right)^k \frac{M!}{(M-k)!} p_0; k \leq M$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^M \left(\frac{\lambda}{\mu}\right)^k \frac{M!}{(M-k)!} \right]^{-1}$$

Igual que en altres sistemes, podem calcular N , N_s i N_w a partir de les probabilitats d'estat. Hem omès el càlcul de fòrmules genèriques, ja que se n'obtenen expressions feixugues, sense gaire interès i, en molts casos, si el nombre d'estats és petit, pot resultar més fàcil fer els sumatoris directament a partir de les dades del problema que es vol resoldre en lloc d'utilitzar aquestes expressions. També podem calcular T i T_w a partir de la fórmula de Little, però, per fer-ho, hem de calcular prèviament la taxa cursada.

Taxa oferta i taxa cursada

En sistemes amb població finita, la taxa d'arribada és diferent en funció de l'estat en què es troba el sistema. En aquest cas, si volem calcular la taxa total que hi arriba, haurem de ponderar la taxa a cada estat per la fracció de temps en què el sistema es manté en aquest estat. Aquestes fraccions de temps són les probabilitats d'estat i, per tant:

$$\lambda_{of} = \sum_{k=0}^{M-1} \lambda_k p_k$$

Aquesta taxa s'anomena **taxa oferta** al sistema i és el nombre mitjà d'unitats per segon que arriben al sistema. Com que no hi ha pèrdues, tota la taxa que hi arriba es cursa i, per tant, en aquest cas, la taxa oferta és igual a la **taxa cursada** i és la que s'utilitza a la fórmula de Little. Com que tots els paquets que entren al sistema també en surten, podem calcular la taxa cursada a la sortida del sistema, és a dir, ponderant les taxes de servei per la fracció de temps en què el sistema està servint:

$$\lambda_c = \sum_{k=1}^M \mu_k p_k$$

Podem comprovar que, per al cas M/M/1//M, ambdues maneres de calcular la taxa cursada coincideixen:

$$\sum_{k=0}^{M-1} \lambda_k p_k = \sum_{k=0}^{M-1} \lambda(M-k) \left(\frac{\lambda}{\mu}\right)^k \frac{M!}{(M-k)!} p_0 = \sum_{k=0}^{M-1} \mu \left(\frac{\lambda}{\mu}\right)^{k+1} \frac{M!}{(M-k-1)!} p_0 =$$



$$= \sum_{k=1}^M \mu \left(\frac{\lambda}{\mu} \right)^k \frac{M!}{(M-k)!} p_0 = \sum_{k=1}^M \mu_k p_k$$

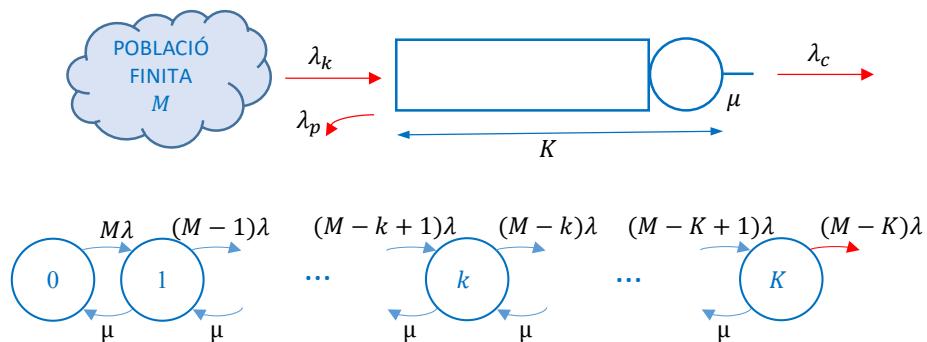
Per aquest sistema, és molt més fàcil fer el càlcul de la taxa cursada a la sortida que a l'entrada, atès que la taxa de servei és la mateixa a tots els estats:

$$\lambda_c = \sum_{k=1}^M \mu_k p_k = \mu(1 - p_0)$$

2.7.2 Sistema M/M/1/K/M

En aquest cas, tenim un sistema en què la població és finita i, a més, hi ha pèrdues. Per tal de veure l'efecte de les pèrdues, suposarem $M > K$, ja que, en cas contrari, mai no hi hauria pèrdues i l'anàlisi seria la mateixa que en el cas M/M/1//M.

Figura 30.
Sistema M/M/1/K/M i
el seu diagrama d'estats



A partir del diagrama d'estats, tenim:

$$p_k = \frac{\lambda^k M(M-1)(M-2) \cdots (M-k+1)}{\mu^k} p_0 = \left(\frac{\lambda}{\mu} \right)^k \frac{M!}{(M-k)!} p_0; k \leq M$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{p_k}{p_0} \right]^{-1} = \left[\sum_{k=0}^K \left(\frac{\lambda}{\mu} \right)^k \frac{M!}{(M-k)!} \right]^{-1}$$

S'obté el mateix resultat que en el cas M/M/1/∞/M, a excepció del càlcul de p_0 .

Taxa oferta, taxa cursada i taxa perduda

La **taxa oferta** al sistema és la ponderació de la taxa d'arribada a cada estat, amb les probabilitats d'estat:



$$\lambda_{of} = \sum_{k=0}^K \lambda_k p_k$$

Com que és un sistema amb pèrdues, no tota la taxa que hi arriba es cursa; només es cursa aquella taxa que hi arriba i que no troba el sistema totalment ple. Durant la fracció de temps en què el sistema està ple, la taxa que hi entra és nul·la. Així doncs, per calcular la **taxa cursada**, hem de ponderar les taxes d'arribada a cada estat amb les probabilitats d'estat, però només en els casos en què el sistema no està totalment ple:

$$\lambda_c = \sum_{k=0}^{K-1} \lambda_k p_k$$

També podem calcular aquesta taxa a la sortida com la taxa de servei a cada estat, ponderada per la fracció de temps en què el sistema està en aquest estat. Tenint en compte que a l'estat 0 la taxa de servei és nul·la:

$$\lambda_c = \sum_{k=1}^K \mu_k p_k$$

És fàcil comprovar que ambdues expressions coincideixen a partir de la propietat que havíem vist quan calculàvem la solució general d'equilibri:

$$\mu_k p_k = \lambda_{k-1} p_{k-1}$$

La taxa oferta que no es cursa es perd i, per tant, definim la **taxa perduda** com la taxa d'arribada quan el sistema és ple, ponderada per la fracció de temps en què el sistema és ple:

$$\lambda_p = \lambda_K p_K$$

Evidentment, tota la taxa que hi arriba o bé es cursa, o bé es perd:

$$\lambda_{of} = \lambda_c + \lambda_p$$

Gràficament, ho podem expressar de la manera següent:

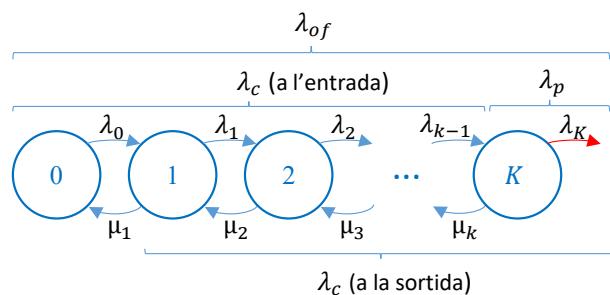


Figura 31.
Taxa oferta, taxa cursada i taxa perduda



Extensió de la propietat PASTA

Quan les arribades són de Poisson, la taxa d'arribades és la mateixa per a tots els estats. Això implica que les arribades visiten igualment tots els estats i, per tant, la probabilitat que una arribada trobi el sistema a l'estat k coincideix amb la fracció de temps en què el sistema és a l'estat k , $r_k = p_k$. Si la taxa d'arribades no és la mateixa a tots els estats, vol dir que hi ha estats que tenen més arribades que d'altres, de manera que la probabilitat que una arribada trobi el sistema en un estat determinat no depèn només de la fracció de temps en què el sistema està en aquell estat si no que, a més, ha de ser més gran com més gran sigui la taxa d'arribades a aquest estat. Sembla lòtic pensar que cal ponderar les probabilitats d'estat per un factor que sigui proporcional a la taxa d'arribades a l'estat.

Recordem la demostració de la propietat PASTA:

$$R_k(t) = \lim_{\Delta t \rightarrow 0} p[N(t) = k | A(t, t + \Delta t)] = \lim_{\Delta t \rightarrow 0} \frac{p[A(t, t + \Delta t) | N(t) = k] p[N(t) = k]}{p[A(t, t + \Delta t)]}$$

La probabilitat de tenir una arribada a l'estat k val:

$$p[A(t, t + \Delta t) | N(t) = k] = \lambda_k \Delta t + o(\Delta t)$$

Per calcular la probabilitat de tenir una arribada (sigui quin sigui l'estat), hem d'utilitzar el teorema de la probabilitat total:

$$p[A(t, t + \Delta t)] = \sum_{i=0}^{\infty} p[A(t, t + \Delta t) | N(t) = i] p[N(t) = i] = \sum_{i=0}^{\infty} \lambda_i \Delta t P_i(t) + o(\Delta t)$$

Substituint:

$$R_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\lambda_k \Delta t P_k(t) + o(\Delta t)}{\sum_{i=0}^{\infty} \lambda_i \Delta t P_i(t) + o(\Delta t)} = \frac{\lambda_k P_k(t)}{\sum_{i=0}^{\infty} \lambda_i P_i(t)}$$

I, en el límit, quan $t \rightarrow \infty$:

$$r_k = \frac{\lambda_k}{\sum_{i=0}^{\infty} \lambda_i} p_i = \frac{\lambda_k}{\lambda_{of}} p_k$$

El quocient λ_k / λ_{of} és igual al quocient entre el nombre d'arribades a l'estat k i el nombre total d'arribades quan observem el sistema durant un temps molt gran. Podem dir que és la probabilitat que una de les arribades al sistema triada a l'atzar s'hagi produït quan aquest estava a l'estat k . Tenim, doncs, que la probabilitat d'arribar i trobar el sistema a l'estat k és proporcional a la fracció de

temps en què el sistema és en aquest estat i proporcional a la probabilitat que una arribada qualsevol s'hagi produït en aquest estat.

Aquesta és una extensió de la propietat PASTA ja que, si les arribades són de Poisson, $\lambda_k = \lambda_{of}$ i, en aquest cas, s'obté que $r_k = p_k$.

2.8 Sistemes semimarkovians

Fins ara, hem estudiat sistemes markovians pels quals el temps entre arribades t_{ia} és una variable aleatòria exponencial i el temps de servei t_s també és una variable aleatòria exponencial. Des del punt de vista analític, els sistemes markovians presenten els avantatges següents:

- Quan el servidor està ocupat i hi ha una arribada, troba que la unitat que hi ha al servidor està parcialment servida, però el temps que manca perquè finalitzi el seu servei tr_s és una variable aleatòria que té la mateixa **PDF** que t_s . Per tant, per la unitat que hi arriba, és equivalent al fet que el servei comencés just al moment en què hi arriba (no té memòria).
- Quan finalitza el servei d'una unitat, el temps que manca fins a la pròpria arribada tr_{ia} és una variable aleatòria que té la mateixa **PDF** que t_{ia} . Al moment en què finalitza el servei, el sistema no recorda que porta algun temps sense tenir cap arribada.

Aquests dos avantatges els podem resumir amb la frase: “És un sistema sense memòria.”

El fet que ambdós temps residuals tinguin la mateixa **PDF** que la variable de la qual són residuals facilita el tractament matemàtic d'aquests sistemes. Si el sistema tingués memòria, no n'hi hauria prou per definir cada estat amb el nombre d'unitats. Per exemple, si es produeix una arribada i el sistema passa de l'estat k a l'estat $k + 1$, per poder caracteritzar l'evolució del sistema caldria indicar també quanta estona porta al servidor la unitat que està essent servida. En canvi, això no passa amb els sistemes sense memòria.

En un sistema de transmissió d'una xarxa de commutació de paquets, suposar que el temps de transmissió (temps de servei) és una variable aleatòria exponencial no sempre és cert. En plantegem un parell d'exemples:

1. Paquets de longitud constant. En aquest cas, tenim un sistema M/D/1 amb una PDF del t_s , que és una delta de Dirac centrada en el valor determinista del temps de transmissió.
2. Agregació de dos tipus de paquets:
 - Tipus 1: amb taxa d'arribada λ_1 i **PDF** del temps de servei $f_{t_{s1}}(t)$.
 - Tipus 2: amb taxa d'arribada λ_2 i **PDF** del temps de servei $f_{t_{s2}}(t)$.



En aquest cas, aplicant el teorema de la probabilitat total, s'obté la **PDF** del temps de servei d'un paquet qualsevol: $f_{t_s}(t) = p_1 f_{t_{s_1}}(t) + p_2 f_{t_{s_2}}(t)$, essent $p_1 = \lambda_1/(\lambda_1 + \lambda_2)$ la fracció de paquets de tipus 1 i $p_2 = \lambda_2/(\lambda_1 + \lambda_2)$ la fracció de paquets de tipus 2.

Volem analitzar sistemes pels quals el temps de servei no sigui una variable aleatòria exponencial. La dificultat principal a l'hora d'analitzar aquests sistemes és el càlcul del temps residual de servei.

2.8.1 Sistema M/G/1

Segons la notació de Kendall, el sistema M/G/1 es caracteritza per:

- Arribades markovianes: Poisson de taxa λ
- Temps de servei genèric: t_s amb $f_{t_s}(t)$ qualsevol
- 1 servidor
- Cua infinita

Per analitzar aquest sistema, calculem el **temps mitjà de transferència** d'una unitat que arriba al sistema. Com sempre, aquest temps serà el temps mitjà de servei, més el temps mitjà d'espera:

$$T = T_s + T_w$$

El càlcul del temps mitjà de servei és trivial si se'n coneix la seva **PDF**, $T_s = E(t_s)$. El temps d'espera d'una unitat i que arriba al sistema es calcula de la manera següent:

$$t_{w_i} = t_{wo_i} + \sum_{j=1}^{n_{w_i}} t_{s_j}$$

essent:

- t_{wo_i} : temps residual de servei que "veu" la unitat i quan arriba al sistema
- n_{w_i} : nombre d'unitats a la cua que troba la unitat i quan arriba al sistema
- t_{s_j} : temps de servei de la unitat j que és a la cua quan arriba la unitat i

Com que n_{w_i} és independent de t_{s_j} , quan es calcula el valor esperat de t_{w_i} se n'obté:

$$T_w = T_{wo} + N_w T_s$$

amb:

- T_w : temps mitjà d'espera
- T_{wo} : temps mitjà residual de servei
- N_w : nombre mitjà d'unitats en espera



A més, la fórmula de Little també ens relaciona N_w i T_w :

$$N_w = \lambda T_w$$

Substituint N_w i aïllant T_w , tenim:

$$T_w = \frac{T_{wo}}{1 - \rho}$$

El valor $\rho = \lambda T_s$ és el factor d'utilització del servidor. Recordem que, per a un sistema G/G/1, aquest valor indica la probabilitat que el servidor estigui ocupat.

Per calcular T_{wo} , hem de tenir en compte que, quan una unitat hi arriba, es poden donar dues situacions:

- L'arribada troba el servidor buit. Aquesta situació s'esdevé amb una probabilitat $1 - \rho$ i, en aquest cas, el temps residual de servei val 0.
- L'arribada troba el servidor ocupat. Aquesta situació s'esdevé amb una probabilitat ρ i, en aquest cas, anomenem z el temps residual de servei.

Per tant, hem definit la variable aleatòria z com el temps residual de servei que "veuen" les unitats que, quan arriben al sistema, el troben ocupat.

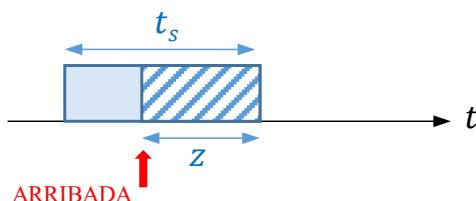


Figura 32.
temps residual de
servei que "veuen"
les unitats que, quan
arriben al sistema, el
troben ocupat

Aleshores, podem calcular el temps residual com:

$$T_{wo} = (1 - \rho) \cdot 0 + \rho \cdot E(z) = \rho \cdot E(z)$$

Dibuixem el procés estocàstic $Z(t)$, que representa el temps que manca perquè finalitzi el servei de la unitat que és al servidor.

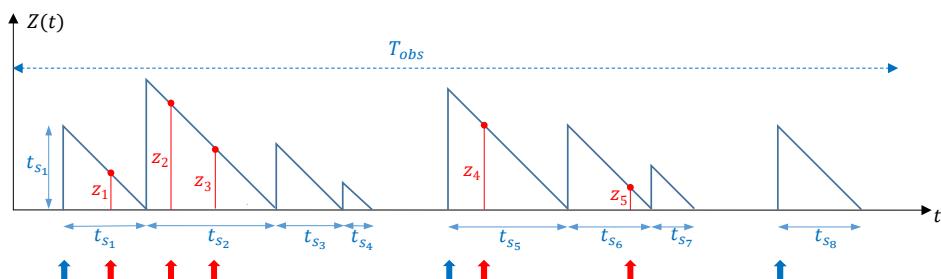
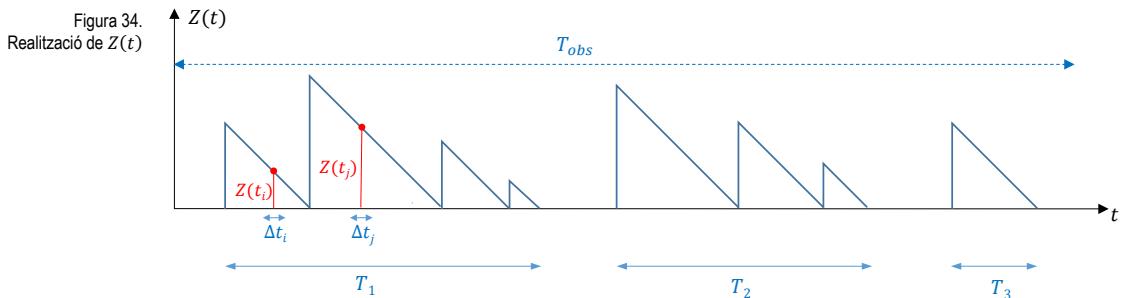


Figura 33.
Realització de $Z(t)$



Quan una unitat entra al servidor, $Z(t)$ és igual al seu temps de servei i va disminuint, de manera que, transcorregut aquest temps de servei, $Z(t)$ val 0. Per això, qualsevol realització del procés té forma de dent de serra. Les fletxes indiquen les arribades: les blaves es corresponen amb aquelles que troben el servidor buit i no esperen; les vermelles troben el servidor ocupat i veuen un temps residual de servei z_i . El valor mitjà de les z_i és el valor $E(z)$ que volem calcular.

Per trobar una expressió analítica de $E(z)$, dividim el temps durant el qual el servidor està ocupat en intervals infinitesimals Δt_i . A la figura 34, el temps durant el qual el servidor està ocupat val $T_1 + T_2 + T_3$; en general, ho indiquem com $\sum T_k$.



Imaginem dos instants de temps, t_i i t_j , que cauen dins els intervals Δt_i i Δt_j , respectivament. Com que les arribades són de Poisson, tots els Δt_i tenen la mateixa probabilitat de tenir una arribada i, per tant, la probabilitat que una arribada vegi un residual $Z(t_i)$ o $Z(t_j)$ és la mateixa. Aquesta probabilitat la calculem com la divisió entre casos favorables i casos possibles:

$$p_i = \frac{\Delta t_i}{\sum T_k}$$

Aleshores, podem calcular $E(z)$ de la manera següent:

$$\begin{aligned} E(z) &= \lim_{T_{obs} \rightarrow \infty} \left[\lim_{\Delta t_i \rightarrow 0} \sum_i Z(t_i) p_i \right] = \lim_{T_{obs} \rightarrow \infty} \frac{1}{\sum T_k} \left[\lim_{\Delta t_i \rightarrow 0} \sum_i Z(t_i) \Delta t_i \right] = \\ &= \lim_{T_{obs} \rightarrow \infty} \left[\frac{1}{\sum T_k} \int_{\sum T_k} Z(t) dt \right] \end{aligned}$$

Cal notar que només integrem el procés $Z(t)$ en els intervals pels quals el servidor està ocupat, per la pròpia definició de z .

Com que $Z(t)$ és un procés estocàstic, resulta impossible calcular la integral i $\sum T_k$ de manera analítica. Per fer-ho, s'aprofita la geometria de $Z(t)$ i es calcula la integral com l'àrea dels triangles del procés. Cada triangle té una àrea igual a $t_{s_i}^2/2$.

Definim $M(T_{obs})$ com el nombre de serveis que han finalitzat en T_{obs} ; aleshores, tenim:

$$E(z) = \lim_{T_{obs} \rightarrow \infty} \left[\frac{1}{\sum_{i=1}^{M(T_{obs})} t_{s_i}} \sum_{i=1}^{M(T_{obs})} \frac{t_{s_i}^2}{2} \right] = \lim_{T_{obs} \rightarrow \infty} \left[\frac{\sum_{i=1}^{M(T_{obs})} t_{s_i}^2}{2 \sum_{i=1}^{M(T_{obs})} t_{s_i}} \right]$$

Dividint el numerador i el denominador per $M(T_{obs})$ i prenent el límit, n'obtenim el primer i el segon moments de t_s :

$$E(z) = \frac{E(t_s^2)}{2E(t_s)} = \frac{E(t_s^2)}{2T_s}$$

Vegem-ne un parell d'exemples:

- Si t_s és exponencial $E(t_s^2) = 2T_s^2$, en aquest cas:

$$E(z) = T_s$$

El valor mitjà del temps residual de servei coincideix amb el temps mitjà de servei, com era d'esperar, ja que t_s no té memòria.

- Si t_s és determinista $E(t_s^2) = T_s^2$, en aquest cas:

$$E(z) = T_s/2$$

El valor mitjà del temps residual de servei coincideix amb la meitat del temps de servei. Com que t_s sempre pren el mateix valor i tots els instants són igualment probables a l'hora de tenir una arribada, tindrem el mateix nombre d'arribades a la primera meitat del servei que a la segona. Si fem la mitjana del residual que veu cada arribada, se n'obté la meitat del temps de servei.

Finalment, podem calcular el temps residual de servei incloent-hi també aquelles arribades que troben el servidor buit:

$$T_{wo} = \rho E(z) = \frac{\lambda E(t_s^2)}{2}$$

Podem obtenir directament aquesta expressió si, en el càlcul de la integral, considerem tot el T_{obs} , en comptes de considerar només els intervals en què el servidor està ocupat:



$$T_{wo} = \lim_{T_{obs} \rightarrow \infty} \left[\frac{1}{T_{obs}} \int_{T_{obs}} Z(t) dt \right] = \lim_{T_{obs} \rightarrow \infty} \left[\frac{M(T_{obs})}{T_{obs}} \frac{1}{M(T_{obs})} \sum_{i=1}^{M(T_{obs})} \frac{t_{si}^2}{2} \right] = \frac{\lambda E(t_s^2)}{2}$$

A l'expressió anterior, el límit de $M(T_{obs})/T_{obs}$ és la taxa d'arribada λ .

Finalment, en calculem el temps d'espera i el temps de transferència substituint el valor de T_{wo} :

$$T_w = \frac{\lambda E(t_s^2)}{2(1 - \rho)}$$

$$T = T_s + \frac{\lambda E(t_s^2)}{2(1 - \rho)}$$

L'expressió del temps de transferència es coneix com la fórmula de Pollaczek-Khinchine però, en la seva formulació original, s'expressa en funció del coeficient de variació del temps de servei. El coeficient de variació d'una variable aleatòria t_s es defineix com:

$$C_{t_s} = \frac{\sigma_{t_s}}{T_s}$$

Recordant que la relació entre la variància i el segon moment d'una variable aleatòria val $\sigma_{t_s}^2 = E(t_s^2) - T_s^2$, podem expressar el segon moment com $E(t_s^2) = T_s^2(1 + C_{t_s}^2)$ i, finalment, el temps de transferència:

$$T = T_s + \rho \frac{T_s(1 + C_{t_s}^2)}{2(1 - \rho)}$$

Amb la fórmula de Pollaczek-Khinchine, es pot calcular el temps d'espera i el temps de transferència per a qualsevol distribució del temps de servei; tan sols cal conèixer el primer i el segon moments de t_s . A continuació, se'n mostren un parell d'exemples:

- Si es calcula el temps d'espera i el temps de transferència per al cas que t_s sigui exponencial, se n'obtenen les mateixes expressions que havíem vist quan analitzàvem el sistema M/M/1. En aquest cas, tenim que $E(t_s^2) = 2T_s^2$ i, per tant:

$$T = T_s + \frac{\rho}{1 - \rho} T_s = \frac{T_s}{1 - \rho}$$

- També podem analitzar un sistema M/D/1. En aquest cas, t_s és determinista i es compleix que $E(t_s^2) = T_s^2$. Aleshores:

$$T = T_s + \frac{\rho}{2(1 - \rho)} T_s = \frac{2 - \rho}{2(1 - \rho)} T_s$$



Per completar l'anàlisi dels sistemes M/G/1, podem calcular N_s , N_w i N . Apliquant la fórmula de Little:

$$N_s = \lambda T_s = \rho; \quad N_w = \lambda T_w = \frac{\lambda^2 E(t_s^2)}{2(1 - \rho)}; \quad N = \lambda T = \rho + \frac{\lambda^2 E(t_s^2)}{2(1 - \rho)}$$

Temps de transferència quan es fa l'agregació de diferents tipus de paquets

Per al càlcul del temps de transferència en un sistema M/G/1, utilitzem la fórmula de Pollaczek-Khinchine, per la qual només ens cal conèixer λ , $E(t_s)$ i $E(t_s^2)$. Suposem el mateix sistema al qual arriben dos tipus de paquets dels quals coneixem:

- Tipus 1: taxa d'arribada λ_1 ; 1r i 2n moments del temps de servei T_{s_1} i $E(t_{s_1}^2)$
- Tipus 2: taxa d'arribada λ_2 ; 1r i 2n moments del temps de servei T_{s_2} i $E(t_{s_2}^2)$

A la figura 35 s'observen arribades de tipus 1 i 2. El temps residual que veu cada arribada no depèn del seu tipus, sinó del tipus d'unitat que troba al servidor quan hi arriba.

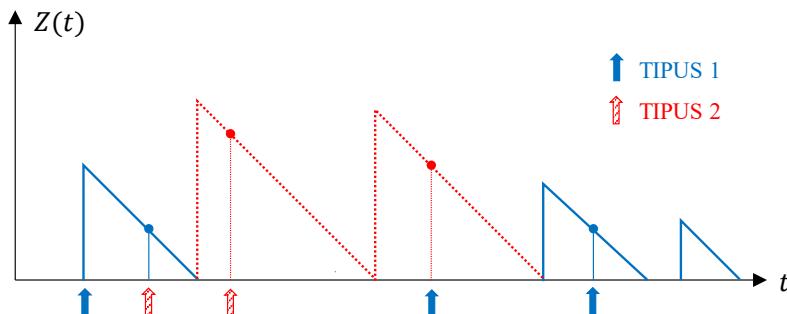


Figura 35.
Realització de $Z(t)$
amb dos tipus de paquets

Construïm una taula amb el residual que veu una unitat quan hi arriba en funció del tipus que troba al servidor:

Tipus que troba quan arriba	Temps residual de servei que observa	Probabilitat d'arribar i trobar aquest tipus (PASTA)
1	$E(z_1) = \frac{E(t_{s_1}^2)}{2T_{s_1}}$	$\rho_1 = \lambda_1 T_{s_1}$
2	$E(z_2) = \frac{E(t_{s_2}^2)}{2T_{s_2}}$	$\rho_2 = \lambda_2 T_{s_2}$
Buit	0	$1 - (\rho_1 + \rho_2)$



Aleshores, el temps residual de servei val:

$$T_{wo} = \rho_1 E(z_1) + \rho_2 E(z_2) + 0 (1 - (\rho_1 + \rho_2))$$

$$T_{wo} = \sum_{i=1}^2 \left[\frac{\lambda_i E(t_{si}^2)}{2} \right]$$

Si multipliquem i dividim per $\lambda = \lambda_1 + \lambda_2$, n'obtenim la mateixa expressió que quan tenim només un tipus de paquet:

$$T_{wo} = \frac{\lambda E(t_s^2)}{2}$$

$$E(t_s^2) = \sum_{i=1}^2 p_i E(t_{si}^2)$$

$$p_i = \frac{\lambda_i}{\lambda}$$

El valor p_i és la probabilitat que una arribada sigui de tipus i ; per tant, en ponderar el segon moment del temps de servei de cada tipus per aquesta probabilitat, se n'obté el segon moment del temps de servei d'un paquet genèric.

Per al càlcul del temps mitjà d'espera, hem de tenir en compte que serà el mateix per a ambdós tipus, ja que només depèn del que troben els paquets quan arriben al sistema, independentment de quin sigui el tipus del paquet que hi arriba. En aquest cas, per fer el càlcul del temps d'espera d'una unitat que arriba, cal tenir en compte el temps residual de servei i la transmissió dels paquets que troba a la cua, tant si són de tipus 1 com de tipus 2:

$$T_w = T_{wo} + N_{w_1} T_{s_1} + N_{w_2} T_{s_2}$$

Utilitzant la fórmula de Little i tenint en compte que el temps mitjà d'espera és el mateix per a ambdós tipus:

$$T_w = T_{wo} + \lambda_1 T_w T_{s_1} + \lambda_2 T_w T_{s_2} = T_{wo} + T_w (\rho_1 + \rho_2)$$

$$T_w = \frac{T_{wo}}{1 - \rho}$$

El valor $\rho_i = \lambda_i T_{s_i}$ és la fracció de temps en què el servidor està ocupat amb una unitat de tipus i . La suma dels ρ_i és la fracció total de temps en que el servidor està ocupat:

$$\rho = \rho_1 + \rho_2 = \lambda_1 T_{s_1} + \lambda_2 T_{s_2}$$

El sistema és estable si $\rho < 1$. Si multipliquem i dividim per $\lambda = \lambda_1 + \lambda_2$, se n'obté:

$$\rho = \lambda T_s$$

$$T_s = \sum_{i=1}^2 p_i T_{s_i}$$

Tornem a calcular el primer moment del temps de servei d'un paquet genèric com la ponderació dels primers moments del temps de servei per a cada tipus.

Podem generalitzar el càlcul per n tipus de paquets. En aquest cas, podem continuar utilitzant la fórmula de Pollaczek-Khinchine:

$$T = T_s + \frac{\lambda E(t_s^2)}{2(1 - \rho)}$$

tenint en compte que:

$$\lambda = \sum_{i=1}^n \lambda_i$$

$$\rho = \sum_{i=1}^n \rho_i = \lambda T_s$$

$$T_s = \sum_{i=1}^n p_i T_{s_i}$$

$$E(t_s^2) = \sum_{i=1}^n p_i E(t_{s_i}^2)$$

$$p_i = \frac{\lambda_i}{\lambda}$$

El temps mitjà de transferència de l'expressió anterior fa referència a un paquet genèric de l'agregació. Per calcular el temps mitjà de transferència d'un tipus concret (tipus i), cal tenir en compte que tots els tipus tenen el mateix temps mitjà d'espera, però un temps de servei diferent:

$$T_i = T_{s_i} + \frac{\lambda \cdot E(t_{s_i}^2)}{2(1 - \rho)}$$

Aplicant Little, podem conèixer el nombre mitjà d'unitats de cada tipus en els diferents elements de la cua:

$$N_{s_i} = \lambda_i T_{s_i} = \rho_i$$



$$N_{w_i} = \lambda_i T_w = \lambda_i \frac{\lambda E(t_s^2)}{2(1 - \rho)}$$

$$N_i = \lambda_i T_i = \rho_i + \lambda_i \frac{\lambda E(t_s^2)}{2(1 - \rho)}$$

El nombre mitjà d'unitats de l'agregació dels n tipus val:

$$N_s = \sum_{i=1}^n N_{s_i} = \lambda T_s = \rho$$

$$N_w = \sum_{i=1}^n N_{w_i} = \lambda T_w = \frac{\lambda^2 E(t_s^2)}{2(1 - \rho)}$$

$$N = \sum_{i=1}^n N_i = \lambda T = \rho + \frac{\lambda^2 E(t_s^2)}{2(1 - \rho)}$$

2.8.2 Sistemes amb prioritat

L'ordre amb què se serveixen les unitats que arriben a una cua es pot basar en diferents estratègies, com ara les següents:

- **First Come First Served (FCFS)**: les unitats se serveixen per ordre d'arribada (equivalent al concepte FIFO).
- **Last Come First Served (LCFS)**: se serveix primer la darrera unitat que hi arriba (equivalent al concepte LIFO).
- **Aleatori**: s'escull la unitat que s'ha de servir de manera aleatòria d'entre les que esperen a la cua.
- **Prioritats**: es classifiquen les unitats en diferents classes i se serveixen primer les de més prioritat.

Tots els sistemes analitzats fins ara utilitzaven la disciplina FCFS. En aquest apartat, s'analitzen sistemes amb prioritat.

En aquest cas, els paquets que arriben a un sistema de transmissió es classifiquem en n classes o categories. Utilitzem la notació següent: el paquets de categoria 1 són els de més prioritat i el de categoria n , els de menys prioritat.

Es tracta d'un sistema semimarkovià ja que, per a cada tipus, les arribades són de Poisson i el temps de servei és una variable aleatòria genèrica de la qual coneixem el primer i el segon moments. Per tant, per a cada categoria, sabem quant valen λ_i , T_{s_i} i $E(t_{s_i}^2)$.

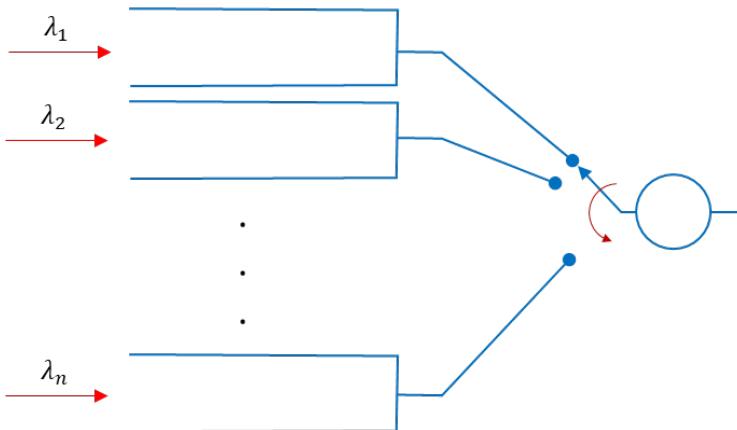


Figura 36.
Sistema amb prioritats

Quan un paquet arriba al sistema, en primer lloc passa per un bloc classificador, que el col·loca al *buffer* corresponent a la seva categoria. El servidor no dona servei als paquets de categoria i fins que els *buffers* amb prioritat superior (de la categoria 1 a la categoria $i - 1$) estan completament buits. Dins de cada categoria, la disciplina de servei és FCFS.

Si considerem les cues corresponents a cada categoria de grandària infinita, la condició perquè el sistema sigui estable és que no s'intenti utilitzar el servidor més del 100% del temps:

$$\rho = \sum_{i=1}^n \lambda_i T_{s_i} < 1$$

Dins de la disciplina de servei amb prioritats, es pot treballar de dues maneres diferents:

- *Prioritats sense expulsió (non-preemptive case)*. Cap paquet no pot interrompre el servei d'un altre, encara que sigui de prioritat superior.
- *Prioritats amb expulsió (preemptive case)*. Els paquets amb més prioritat poden interrompre el servei d'un paquet de menys prioritat. El paquet interromput torna a la primera posició del *buffer* corresponent a la seva classe (serà el primer de la seva classe a obtenir el servidor). En aquest cas, fem una subclassificació en funció de com es torna a donar servei al paquet expulsat:
 - *Preemptive resume*. El servei continua des del punt on es va deixar, és a dir, els bits que ja s'han transmès no es tornen a transmetre.
 - *Preemptive restart*. El servei s'inicia de nou, és a dir, es torna a transmetre tot el paquet de nou.

Només estudiarem els casos **non-preemptive** i **preemptive resume**, que anomenarem **prioritats sense expulsió** i **prioritats amb expulsió**, respectivament.



2.8.3 Prioritats sense expulsió

Per fer l'anàlisi d'aquest sistema, calculem T_k , el temps mitjà de transferència d'un paquet de categoria k . Aquest temps té quatre components:

- T_{w0k} : espera deguda al paquet que s'està servint quan el paquet de categoria k arriba al sistema. És el temps mitjà residual de servei que veu un paquet de categoria k .
- T_{wAk} : espera deguda als paquets que hi ha a la cua quan el paquet de categoria k arriba al sistema i s'han de servir abans que ell. Són paquets de prioritat superior o igual a k (de 1 a k).
- T_{wBk} : espera deguda als paquets que arriben al sistema mentre el paquet de categoria k està en espera i s'han de servir abans que ell. Són paquets de prioritat superior a k (de 1 a $k - 1$).
- T_{sk} : temps mitjà de servei del paquet de categoria k . És $E(t_{sk})$.

Els tres primers corresponen a l'espera i el darrer, al servei:

$$T_k = T_{w_k} + T_{s_k}$$

$$T_{w_k} = T_{w0_k} + T_{wA_k} + T_{wB_k}$$

T_{w0k} depèn del paquet que s'està servint quan n'arriba un de categoria k . Com que no hi ha expulsions, totes les categories veuen el mateix temps mitjà residual de servei (no depèn de k). De fet, és el mateix temps mitjà residual que observaria una agregació de n tipus de paquets en un sistema M/G/1:

$$T_{w0k} = T_{w0} = \sum_{i=1}^n \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$$

Per al càlcul de T_{wAk} , definim la variable aleatòria següent:

- n_{ik} : nombre de paquets de categoria i a la cua quan arriba un paquet de categoria k .

De mitjana, el temps que es triga a servir aquests paquets val $N_{ik} T_{s_i}$, essent $N_{ik} = E(n_{ik})$. Tenint en compte totes les categories que s'han de servir abans que el paquet de categoria k , se n'obté:

$$T_{wAk} = \sum_{i=1}^k N_{ik} T_{s_i}$$

Per la propietat PASTA, el nombre mitjà de paquets a la cua de categoria i amb què es troba un paquet quan hi arriba és igual al nombre mitjà de paquets de categoria i que hi ha a la cua. Per tant, aplicant Little:

$$N_{ik} = N_{wi} = \lambda_i T_{wi}$$

Observeu que N_{ik} no depèn de k . Substituint N_{ik} , tenim:

$$T_{wA_k} = \sum_{i=1}^k \rho_i T_{wi}$$

Per al càlcul de T_{wB_k} , definim la variable aleatòria següent:

- m_{ik} : nombre de paquets de categoria i que hi arriben mentre el paquet de categoria k espera a la cua.

De mitjana, el temps que es triga a servir aquests paquets val $M_{ik} T_{si}$, essent $M_{ik} = E(m_{ik})$. Tenint en compte totes les categories que s'han de servir abans que el paquet de categoria k , se n'obté:

$$T_{wB_k} = \sum_{i=1}^{k-1} M_{ik} T_{si}$$

Com que la taxa d'arribades és λ_i , el nombre mitjà de paquets de **categoria i que arriben** mentre el de **categoria k espera a la cua** val:

$$M_{ik} = \lambda_i T_{wk}$$

En aquest cas, M_{ik} sí que depèn de k . Substituint M_{ik} , tenim:

$$T_{wB_k} = T_{wk} \sum_{i=1}^{k-1} \rho_i$$

Finalment, en podem calcular el temps d'espera:

$$T_{wk} = T_{w0} + \sum_{i=1}^k \rho_i T_{wi} + T_{wk} \sum_{i=1}^{k-1} \rho_i$$

En el segon membre, el primer sumatori té un terme T_{wk} . El traiem i el sumem al segon sumatori. Fent això, podrem aïllar T_{wk} :

$$T_{wk} = T_{w0} + \sum_{i=1}^{k-1} \rho_i T_{wi} + T_{wk} \sum_{i=1}^k \rho_i$$

$$T_{wk} = \frac{T_{w0} + \sum_{i=1}^{k-1} \rho_i T_{wi}}{(1 - \sum_{i=1}^k \rho_i)}$$



N'hem obtingut un sistema d'equacions en què cada T_{w_k} es calcula a partir del temps d'espera de les categories amb més prioritat ($i = 1, 2, \dots, k - 1$).

Resolem aquest sistema per a dues categories de paquets ($n = 2$):

$$\begin{aligned} T_{w_1} &= \frac{T_{w_0}}{1 - \rho_1} \\ T_{w_2} &= \frac{T_{w_0} + \rho_1 T_{w_1}}{1 - \rho_1 - \rho_2} = \frac{T_{w_0} + \rho_1 \frac{T_{w_0}}{1 - \rho_1}}{1 - \rho_1 - \rho_2} = \frac{T_{w_0}}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} \\ T_{w_0} &= \frac{\lambda_1 E(t_1^2)}{2} + \frac{\lambda_2 E(t_2^2)}{2} \end{aligned}$$

A diferència del sistema M/G/1, ara el temps d'espera és diferent per a cada categoria. Evidentment, això és el que es pretén quan s'apliquen diferents prioritats als diferents fluxos de paquets de dades. El temps d'espera corresponent a la categoria més prioritària té una fórmula molt semblant al temps d'espera d'un sistema M/G/1 al qual només arribés aquesta categoria. La diferència és el temps residual de servei, que també inclou els paquets menys prioritaris.

Per aïllar els T_{w_k} i trobar la solució general del sistema d'equacions, fem el següent. Prenem les equacions en la forma:

$$T_{w_k} \cdot \left(1 - \sum_{i=1}^k \rho_i \right) = T_{w_0} + \sum_{i=1}^{k-1} \rho_i T_{w_i}$$

i restem les equacions corresponents a $k - 1$ i k :

$$\begin{aligned} T_{w_{k-1}} \cdot \left(1 - \sum_{i=1}^{k-1} \rho_i \right) - T_{w_k} \cdot \left(1 - \sum_{i=1}^k \rho_i \right) &= -\rho_{k-1} T_{w_{k-1}} \\ T_{w_{k-1}} \cdot \left(1 - \sum_{i=1}^{k-2} \rho_i \right) &= T_{w_k} \cdot \left(1 - \sum_{i=1}^k \rho_i \right) \\ T_{w_k} &= \frac{\left(1 - \sum_{i=1}^{k-2} \rho_i \right)}{\left(1 - \sum_{i=1}^k \rho_i \right)} T_{w_{k-1}} \end{aligned}$$

Utilitzant l'expressió anterior de forma recursiva, s'obté:

$$T_{w_k} = \frac{\left(1 - \sum_{i=1}^j \rho_i \right) \left(1 - \sum_{i=1}^{j-1} \rho_i \right)}{\left(1 - \sum_{i=1}^k \rho_i \right) \left(1 - \sum_{i=1}^{k-1} \rho_i \right)} T_{w_j}; \quad j < k - 1$$

En particular, per a $j = 1$ o $j = 2$, s'obté la solució de T_w :

$$T_{w_k} = \frac{T_{w0}}{(1 - \sum_{i=1}^k \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)}$$

$$T_{w0} = \sum_{i=1}^n \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$$

Finalment, podem calcular el temps de transferència per a un paquet de categoria k :

$$T_k = T_{w_k} + T_{s_k} = \frac{T_{w0}}{(1 - \sum_{i=1}^k \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)} + T_{s_k}$$

2.8.4 Prioritats amb expulsió

Per fer l'anàlisi d'aquest sistema, calculem T_k , el temps mitjà de transferència d'un paquet de categoria k . Aquest temps té, en principi, els mateixos components que per al cas sense expulsió:

- T_{wo_k} : espera deguda al paquet que s'està servint quan el paquet de categoria k arriba al sistema. És el temps residual de servei mitjà que veu un paquet de categoria k .
- T_{wA_k} : espera deguda als paquets que hi ha a la cua quan el paquet de categoria k arriba al sistema i s'han de servir abans que ell. Són paquets de prioritat superior o igual a k (de 1 a k).
- T_{wB_k} : espera deguda als paquets que arriben al sistema mentre el paquet de categoria k és al sistema i s'han de servir abans que ell. Són paquets de prioritat superior a k (de 1 a $k-1$). Cal notar que, a diferència del cas sense expulsió, els paquets que arriben mentre el paquet de categoria k és al servidor l'interrompen i n'augmenten l'espera.
- T_{s_k} : temps mitjà de servei del paquet de categoria k . És $E(t_{s_k})$.

Els tres primers corresponen a l'espera i el darrer, al servei:

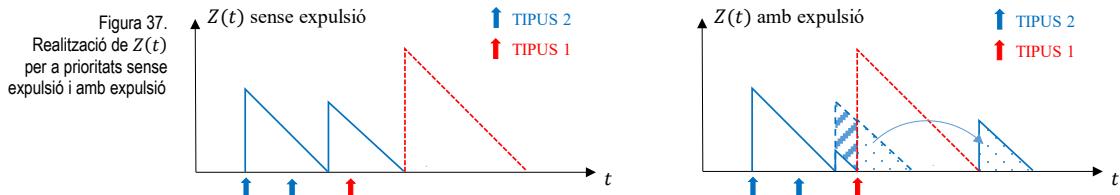
$$T_k = T_{w_k} + T_{s_k}$$

$$T_{w_k} = T_{wo_k} + T_{wA_k} + T_{wB_k}$$

T_{wo_k} depèn del paquet que s'està servint quan n'arriba un de categoria k . En haver-hi expulsions, els paquets més prioritaris “no veuen” els menys prioritaris. Per als paquets amb més prioritat, és com si els menys prioritaris no hi fossin. Per tant, totes les categories no veuen el mateix temps mitjà residual de servei, sinó que aquest depèn de k . A la figura 37, es mostra la funció $Z(t)$ per a dues categories, comparant el cas sense expulsió amb el cas amb expulsió. S'observa que, tot i que el servidor està ocupat durant la mateixa estona, les àrees en ambdós casos no coincideixen. L'àrea per al cas amb expulsió és més



petita (desapareix la zona ratllada). Així doncs, no podem aprofitar els càlculs que havíem fet per al cas sense expulsió.



Quan hi arriba un paquet de categoria k , pot trobar a la cua paquets que han estat interromputs i, per tant, que estan parcialment servits, de manera que el nombre mitjà de paquets de categoria i que troba i el temps de servei d'aquests paquets no es pot avaluar com en un cas sense interrupcions. Això fa que el càlcul de T_{wA_k} tampoc no sigui trivial.

Per tal de calcular el temps mitjà d'espera per al cas d'un sistema amb prioritats amb expulsió, cal calcular $T_{wo_k} + T_{wA_k}$ conjuntament, encara que no sapiguem quant valen aquests dos temps per separat. Per fer-ho, cal tenir presents dues coses:

- Els paquets més prioritaris “no veuen” els menys prioritaris.
- El servidor estarà ocupat durant la mateixa estona, tant si dona prioritats com si no.

Quan arriba un paquet de categoria k a un sistema que utilitza prioritats amb expulsió, aquest “veu” que el treball que ha de fer el servidor abans de donar-li servei és:⁶

$$U_k^{PP} = T_{wo_k}^{PP} + T_{wA_k}^{PP}$$

Aquest treball és el que volem calcular. Com que els paquets de prioritat inferior (de la categoria $k+1$ a n) mai no endarrereixen els de prioritat superior (de la categoria 1 a k), podem afirmar que aquest treball coincideix amb el que trobaria un paquet de tipus k quan arriba a un sistema FCFS on només hi ha paquets de tipus 1 a k . Estem utilitzant el fet que els paquets amb més prioritat “no veuen” els menys prioritaris i que el servidor estarà ocupat durant la mateixa estona, tant si dona prioritats com si no. És equivalent a pensar que els bits no transmesos (treball pendent del servidor) no depenen de l'ordre en què es transmeten (amb prioritat o sense). Des del punt de vista del servidor, a ell li manca la mateixa feina, tant si dona prioritat com si no.

⁶ U indica *unfinished work*. El superíndex PP indica *preemptive priority*.

Per a un sistema FCFS amb k tipus de paquet, quan un paquet hi arriba, troba que el treball que ha de fer el servidor abans de ser servit és el seu temps d'espera:

$$U_k^{FCFS} = T_{w_k}^{FCFS} = \frac{T_{w0_k}^{FCFS}}{1 - \sum_{i=1}^k \rho_i}$$

$$T_{w0_k}^{FCFS} = \sum_{i=1}^k \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$$

Cal notar que $T_{w0_k}^{FCFS}$ és el temps residual de servei quan hi ha k tipus de paquets en un sistema FCFS i no representa el temps residual de servei d'un paquet de categoria k per a un sistema amb prioritats amb expulsió. En un sistema FCFS, tots els paquets "veuen" el mateix residual, però en un sistema amb prioritats amb expulsió, no.

Així doncs, com que $U_k^{PP} = U_k^{FCFS}$, finalment s'obté:

$$T_{w0_k}^{PP} + T_{wA_k}^{PP} = \frac{T_{w0_k}^{FCFS}}{1 - \sum_{i=1}^k \rho_i}$$

Passem ara al càlcul del terme T_{wB_k} . Per fer-ho, definim la variable aleatòria següent.

- m_{ik} : nombre de paquets de categoria i que hi arriben mentre el paquet de categoria k és al sistema.

De mitjana, el temps que es triga a servir aquests paquets val $M_{ik} T_{s_i}$, essent $M_{ik} = E(m_{ik})$. Tenint en compte totes les categories que s'han de servir abans que el paquet de categoria k , s'obté:

$$T_{wB_k} = \sum_{i=1}^{k-1} M_{ik} T_{s_i}$$

Com que la taxa d'arribades és λ_i , el nombre mitjà de paquets de **categoria i que hi arriben** mentre el de **categoria k és al sistema** val:

$$M_{ik} = \lambda_i T_k$$

Substituint M_{ik} , tenim:

$$T_{wB_k} = T_k \sum_{i=1}^{k-1} \rho_i$$

Amb els resultats obtinguts, podem calcular el temps de transferència:



$$T_k = \frac{T_{w0k}^{FCFS}}{1 - \sum_{i=1}^k \rho_i} + T_k \sum_{i=1}^{k-1} \rho_i + T_{s_k}$$

Un cop feta aquesta demostració, prescindim de les paraules *PP* i *FCFS* a la fórmula final del temps de transferència per tal de simplificar-ne la notació. Finalment, aïllant el valor del temps de transferència, tenim:

$$T_k = \frac{T_{w0k} + T_{s_k}(1 - \sum_{i=1}^k \rho_i)}{(1 - \sum_{i=1}^k \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)}$$

$$T_{w0k} = \sum_{i=1}^k \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$$

A l'expressió de T_k no es diferencia entre el temps mitjà d'espera i el temps mitjà de servei; per tant, el temps mitjà d'espera es calcula fent:

$$T_{w_k} = T_k - T_{s_k}$$

Com que la categoria més prioritària no "veu" la resta de tipus, si en calculem el temps de transferència n'obtenim el temps de transferència d'un sistema M/G/1 on només hi ha aquest tipus de paquets:

$$T_1 = \frac{\lambda_1 E(t_{s_1}^2)}{2(1 - \rho_1)} + T_{s_1}$$

A continuació, es mostra una taula comparativa de les anàlisis per les diferents disciplines de servei.

FCFS	$T_k = \frac{T_{w0}}{(1 - \sum_{i=1}^n \rho_i)} + T_{s_k}$	$T_{w0} = \sum_{i=1}^n \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$
Prioritats sense expulsió	$T_k = \frac{T_{w0}}{(1 - \sum_{i=1}^k \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)} + T_{s_k}$	$T_{w0} = \sum_{i=1}^n \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$
Prioritats amb expulsió	$T_k = \frac{T_{w0k} + T_{s_k}(1 - \sum_{i=1}^k \rho_i)}{(1 - \sum_{i=1}^k \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)}$	$T_{w0k} = \sum_{i=1}^k \left[\frac{\lambda_i E(t_{s_i}^2)}{2} \right]$

Observacions:



- FCFS i prioritats sense expulsió tenen el mateix temps residual de servei, el qual no depèn del tipus o de la categoria.
- El temps residual de servei per prioritats amb expulsió depèn de la categoria.
- Per FCFS, el temps d'espera no depèn del tipus; pels sistemes amb prioritat, sí que depèn de la categoria.

→3



Anàlisi de les funcionalitats de la xarxa

Al capítol anterior, s'han modelat els sistemes de transmissió utilitzant la teoria de cues. En aquest capítol, s'analitzen diferents aspectes de les xarxes telemàtiques en la seva globalitat. Per fer-ho, s'utilitzen els models que hem vist al capítol anterior per avaluar els sistemes de transmissió de la xarxa, connectats entre si, en el que es coneix com a *xarxes de cues*. Principalment, si tenim un sistema de transmissió amb un únic canal, el modelem, en una primera aproximació, com una cua M/M/1, ja que és un model força estudiat, té un tractament matemàticament senzill i s'ajusta amb precisió a diferents situacions. Per descomptat, dependent de la precisió que necessitem en els nostres resultats, en molts casos es necessitaran altres models, amb un tractament analític més complex, que no són objecte d'aquest curs de grau.

En primer lloc, estudiarem el comportament de la xarxa més senzilla possible, formada únicament per la connexió en cascada de dos sistemes de transmissió.

3.1 Teorema de Burke

Suposem dos sistemes de transmissió de dos nodes adjacents, de manera que els paquets que surten del sistema de transmissió del primer node entren en el sistema de transmissió del segon node:

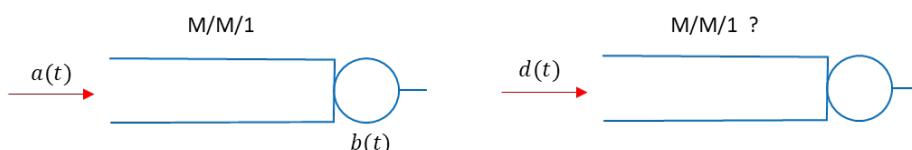


Figura 38.
Teorema de Burke

El primer node el modelem com un sistema M/M/1, és a dir:



$$a(t) = f_{t_{ia}}(t) = \lambda e^{-\lambda t}; t \geq 0 \text{ i } b(t) = f_{t_s}(t) = \mu e^{-\mu t}; t \geq 0$$

Volem saber si és possible modelar el segon node com un sistema M/M/1. És evident que la longitud dels paquets no canvia; per tant, el temps de transmissió en el segon sistema serà exponencial (potser amb una μ diferent, si les capacitats dels canals són diferents). També està clar que la taxa d'arribades al segon sistema serà λ , ja que el primer sistema ni crea ni destrueix paquets. Malgrat això, ens manca saber si el temps entre arribades al segon sistema és exponencial.

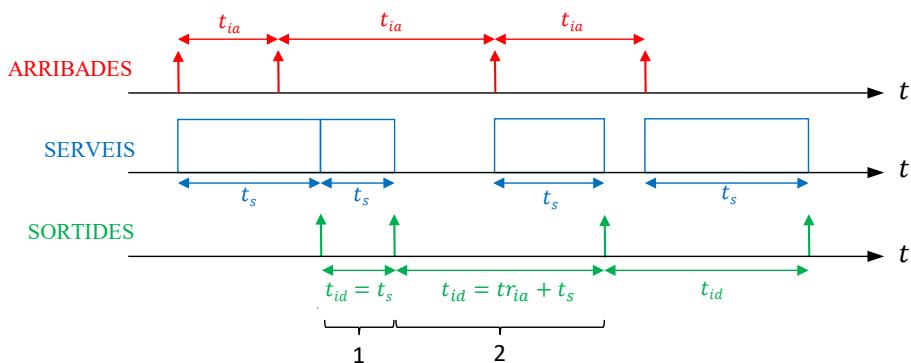
Definim la variable aleatòria següent:

- t_{id} : temps entre sortides (*interdeparture time*), amb $f_{t_{id}}(t) = d(t)$.

El temps entre sortides del primer sistema es correspon amb el temps entre arribades del segon. Per tant, si $d(t)$ és exponencial; en principi, podríem modelar el segon sistema com una cua M/M/1.

A la figura 39, es mostren les arribades, els serveis i les sortides del primer sistema.

Figura 39.
Temps entre arribades, temps de servei i temps entre sortides



Se'n distingeixen dos casos:

- Cas 1: entre dues arribades, la segona es produeix abans que finalitzi el servei de la primera, és a dir, la segona arribada troba el sistema ocupat. Aquest cas es produeix amb una probabilitat ρ i el temps entre sortides és igual al temps de servei.
- Cas 2: entre dues arribades, la segona es produeix després d'haver finalitzat el servei de la primera, és a dir, la segona arribada troba el sistema buit. Aquest cas es produeix amb una probabilitat $1 - \rho$ i el temps entre sortides és un temps residual de t_{ia} més el temps de servei.

Aplicant-hi el teorema de la probabilitat total:



$$f_{tid}(t) = \rho f_{ts}(t) + (1 - \rho) f_{tr_{ia}+t_s}(t)$$

Tenint en compte que el temps residual del temps entre arribades té la mateixa **PDF** que el temps entre arribades (en ser exponencial) i que la **PDF** de la suma de dues variables aleatòries és la convolució de les seves **PDF**, tenim:

$$d(t) = \rho b(t) + (1 - \rho) (a(t) * b(t))$$

Convertim la convolució en un producte passant l'equació al domini transformat de Laplace. Primer transformem $a(t)$ i $b(t)$:

$$a(t) = \lambda e^{-\lambda t}; t \geq 0 \xrightarrow{\mathcal{L}} A(s) = \frac{\lambda}{s + \lambda}$$

$$b(t) = \mu e^{-\mu t}; t \geq 0 \xrightarrow{\mathcal{L}} B(s) = \frac{\mu}{s + \mu}$$

Substituint i simplificant l'expressió:

$$\begin{aligned} D(s) &= \rho B(s) + (1 - \rho) (A(s) B(s)) = \rho \frac{\mu}{s + \mu} + (1 - \rho) \left(\frac{\lambda}{s + \lambda} \frac{\mu}{s + \mu} \right) = \\ &= \frac{\mu}{s + \mu} \left(\frac{\rho(s + \lambda) + (1 - \rho)\lambda}{s + \lambda} \right) = \frac{\mu}{s + \mu} \left(\frac{\rho s + \lambda}{s + \lambda} \right) = \\ &= \frac{1}{s + \mu} \left(\frac{\lambda s + \lambda \mu}{s + \lambda} \right) = \frac{\lambda}{s + \lambda} \end{aligned}$$

Tornant al domini temporal, n'obtenim que el temps entre sortides és també una variable aleatòria exponencial:

$$D(s) = \frac{\lambda}{s + \lambda} \xrightarrow{\mathcal{L}^{-1}} d(t) = \lambda e^{-\lambda t}; t \geq 0$$

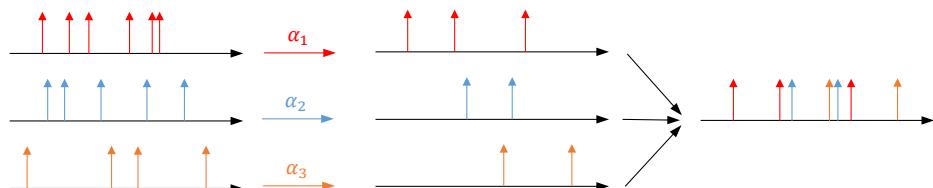
Per tant, el temps entre arribades al segon sistema és una variable aleatòria exponencial. Per tal de poder modelar el segon sistema com una M/M/1, tan sols manca que el temps entre arribades i el temps de servei siguin variables aleatòries independents. Cal recordar que aquesta propietat s'utilitza per demostrar que el temps de permanència en un estat és exponencial i, en conseqüència, que el procés és markovià. De l'equació de $d(t)$, s'observa que el temps entre sortides no és independent del de servei i, per tant, el segon sistema no es pot modelar com una cua M/M/1.

No obstant això, en xarxes més complexes, no tots els paquets provinents d'un node antecessor s'encaminen vers el mateix destí i no tots els paquets que s'en-caminen vers el mateix destí provenen del mateix node antecessor. Tenint en compte això, podem dir que les arribades al sistema de transmissió d'un node estan formades per l'agregació d'una fracció de les sortides dels nodes anteces-sors. El fet de prendre només una fracció dels paquets (dividir) i ajuntar-la amb



altres fraccions (superposar) fa que el temps entre arribades al sistema de transmissió i el temps de servei dels paquets siguin cada cop més independents. Com més gran sigui el nombre de divisions i superposicions, major serà la seva independència. A més, cal recordar que la divisió i la superposició de processos de Poisson continua essent un procés de Poisson i, per tant, el temps entre arribades té una distribució exponencial de taxa $\sum_{i=1}^n \alpha_i \lambda_i$, on n representa el nombre de nodes antecessors; λ_i , la taxa d'arribada al node provinent de cada un d'ells, i α_i , la fracció de paquets del node antecessor i que s'envia al sistema de transmissió.

Figura 40.
Aleatorització de les
arribades a un sistema de transmissió

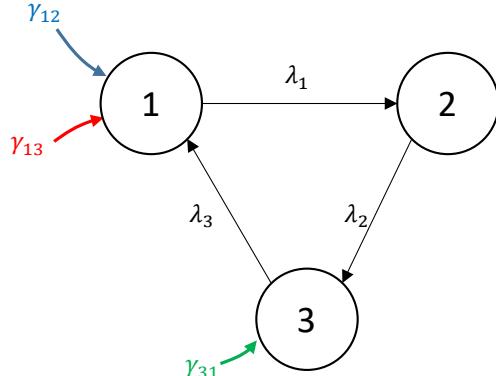


Aleshores, si el temps entre arribades continua essent exponencial i considerem que és independent del temps de servei, podem fer l'aproximació de modelar els sistemes de transmissió de la xarxa com una cua M/M/1. A la resta del capítol, utilitzarem el model M/M/1 per avaluar els sistemes de transmissió d'una xarxa.

3.2 Nombre mitjà de salts que fa un paquet a la xarxa

A l'exemple de la figura 41, es mostra una xarxa de transport amb $N = 3$ nodes i $M = 3$ canals. En aquesta xarxa, l'encaminament es fa de manera circular en sentit horari. El paràmetre γ_{ij} indica el trànsit (en paquets per segon) que entra al node i , destinat al node j . Aquest trànsit no es genera a la xarxa de transport, sinó que prové de xarxes d'accés i està destinat a altres xarxes d'accés.

Figura 41.
Exemple de topologia



Definim el trànsit total que entra a la xarxa com:



$$\gamma = \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \gamma_{ij}$$

A l'exemple, $\gamma = \gamma_{12} + \gamma_{13} + \gamma_{31}$.

Numerem els canals i calculem els paquets per segon que es transmeten per cada canal, de manera que λ_i fa referència al canal i .

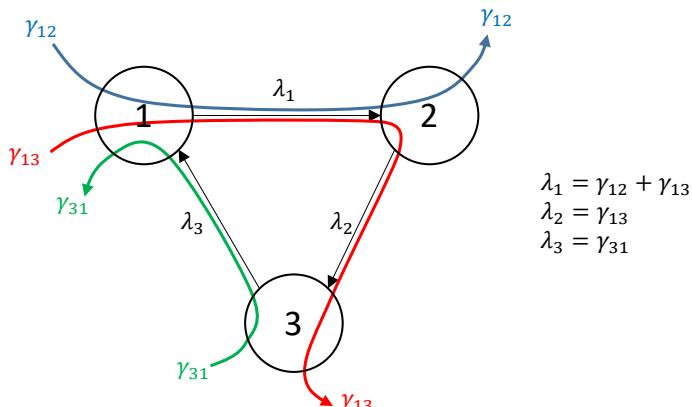


Figura 42.
Encaminament dels paquets

En aquest exemple, hi ha paquets que només fan un salt i hi ha paquets que en fan dos. Per calcular el nombre mitjà de salts que fa un paquet genèric, cal tenir en compte la fracció de paquets que fa un salt i la fracció de paquets que en fa dos:

$$H = \frac{\gamma_{12}}{\gamma} 1 + \frac{\gamma_{13}}{\gamma} 2 + \frac{\gamma_{31}}{\gamma} 1$$

El fet que el trànsit γ_{13} faci dos salts significa que contribueix al trànsit de dos canals (λ_1 i λ_2). Per tant, descomponent l'expressió anterior i reagrupant, se n'obté:

$$H = \frac{1}{\gamma} [(\gamma_{12} + \gamma_{13}) + \gamma_{13} + \gamma_{31}] = \frac{1}{\gamma} [\lambda_1 + \lambda_2 + \lambda_3]$$

Aquest raonament el podem utilitzar en qualsevol altra topologia i conjunt de trànsits d'entrada a la xarxa. Així doncs, el nombre mitjà de salts que fa un paquet genèric en una xarxa val:

$$H = \frac{1}{\gamma} \sum_{i=1}^M \lambda_i$$

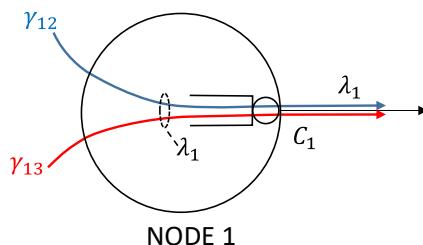
essent M el nombre de canals de la topologia.



3.3 Temps mitjà de trànsit d'un paquet genèric

El temps de trànsit d'un paquet és la suma dels seus temps de transferència. Si neglimitem els temps de commutació i els temps de propagació, el temps de transferència és el temps que el paquet és al sistema de transmissió (espera + transmissió).

Figura 43.
Paràmetres del sistema de transmissió



Dels sistemes de transmissió, en coneixem:

- λ_i : taxa d'arribada al canal i (paquets/s)
- C_i : capacitat del canal i (bps)
- L : longitud mitjana dels paquets (bits)

A partir d'aquests paràmetres, tenim que el temps mitjà de transmissió, que es correspon amb el temps mitjà de servei, es pot calcular com:

$$T_{s_i} = \frac{L}{C_i}$$

Si modelem els sistemes de transmissió com una cua M/M/1, el temps de transferència pel canal i val:

$$T_i = \frac{T_{s_i}}{1 - \rho_i} = \frac{L}{C_i - \lambda_i L}$$

Seguint amb l'exemple anterior, tenim que el temps mitjà de trànsit es calcula ponderant cada temps de trànsit per la fracció de paquets corresponent:

$$\begin{aligned} T_N &= \frac{\gamma_{12}}{\gamma} T_1 + \frac{\gamma_{13}}{\gamma} (T_1 + T_2) + \frac{\gamma_{31}}{\gamma} T_3 = \frac{1}{\gamma} [(\gamma_{12} + \gamma_{13}) T_1 + \gamma_{13} T_2 + \gamma_{31} T_3] \\ T_N &= \frac{1}{\gamma} [\lambda_1 T_1 + \lambda_2 T_2 + \lambda_3 T_3] \end{aligned}$$

El temps de trànsit dels paquets que fan més d'un salt està format per la suma dels temps de transferència de cada salt. En agrupar tots els trànsits γ_{ij} corresponents a un mateix temps de transferència T_k , estem sumant tots els trànsits



a través del canal k i n'obtenim el valor de λ_k . Atès que això passa per a qualsevol topologia o conjunt de trànsits entrants, podem generalitzar l'expressió anterior:

$$T_N = \frac{1}{\gamma} \sum_{i=1}^M \lambda_i T_i$$

Aquesta expressió es pot obtenir també si interpretarem la xarxa de transport com tot un sistema al qual entren γ paquets/s (provinents de les xarxes d'accés) i en surten γ paquets/s (destinats a les xarxes d'accés).

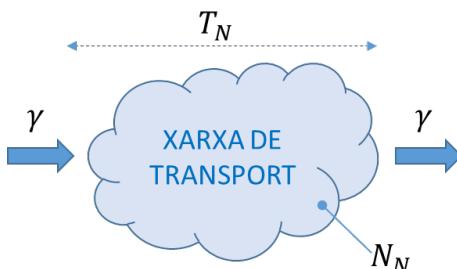


Figura 44.
Temps de trànsit

El nombre mitjà de paquets a la xarxa és la suma del nombre mitjà de paquets que hi ha a cada sistema de transmissió, que podem calcular aplicant la llei de Little.

$$N_N = \sum_{i=1}^M N_i = \sum_{i=1}^M \lambda_i T_i$$

També podem calcular el nombre mitjà de paquets a la xarxa aplicant la mateixa llei al sistema format per tota la xarxa, és a dir, $N_N = \gamma T_N$. Igualant ambdues expressions i aïllant el temps mitjà de trànsit, n'obtenim el mateix resultat.

$$T_N = \frac{1}{\gamma} \sum_{i=1}^M \lambda_i T_i$$

3.4 Repartiment òptim de fluxos

A les xarxes de transport, és molt possible que hi hagi més d'una ruta per encaminar els paquets cap al seu destí. Repartir el trànsit que es vol enviar entre les possibles rutes té alguns avantatges:

- Millora la fiabilitat de la xarxa. Si hi ha fallida en alguna ruta, no es perd tota la informació.
- Reparteix la càrrega evitant que alguns enllaços es puguin sobrecarregar.

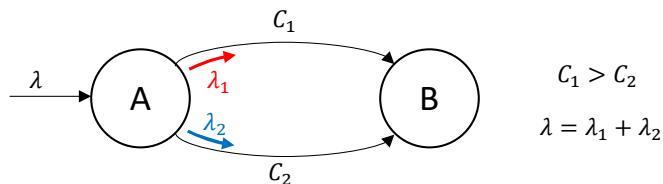


En aquest apartat, estudiem com repartir el trànsit entre les diferents rutes per tal d'optimitzar algun paràmetre de la xarxa. Nosaltres concretament optimitzem el temps mitjà de trànsit.

3.4.1 Repartiment òptim per dos camins

Comencem fent l'estudi per a una topologia ben senzilla i posteriorment l'anirem complicant. Aquest cas senzill està format per dos nodes A (origen) i B (destí), units per dos canals de capacitat C_1 i C_2 . Sense pèrdua de generalitat, suposem que $C_1 > C_2$. El node A ha d'enviar λ paquets/s al node B. Aquest trànsit s'ha de repartir entre el canal 1 i el canal 2, de manera que per a cada canal s'envien λ_1 i λ_2 paquets/s, respectivament. L'objectiu es trobar els valors de λ_1 i λ_2 que optimitzen el temps mitjà de trànsit.

Figura 45.
Repartiment de fluxos per dos camins



Modelem cada sistema de transmissió com una cua M/M/1. El temps mitjà de trànsit és igual als temps mitjans de transferència dels canals, ponderats per la fracció de paquets que s'envia per cada camí.

$$T_{AB} = \frac{\lambda_1}{\lambda} \frac{L}{C_1 - \lambda_1 L} + \frac{\lambda_2}{\lambda} \frac{L}{C_2 - \lambda_2 L}$$

Observeu que aquesta expressió es correspon amb l'obtinguda a l'apartat anterior. Per comoditat a l'hora de treballar amb aquesta equació, definim $f_i = \lambda_i L$. Aquest paràmetre representa el flux de bits enviat per cada camí, expressat en bps.

$$T_{AB} = \frac{1}{\lambda} \left(\frac{f_1}{C_1 - f_1} + \frac{f_2}{C_2 - f_2} \right) = \frac{1}{\lambda} F(f_1, f_2)$$

Atès que λ és un valor fix, optimitzar T_{AB} equival a optimitzar $F(f_1, f_2)$. Aquesta funció és un camp escalar i és diferent per a cada topologia. Per a aquesta topologia senzilla val:

$$F(f_1, f_2) = \frac{f_1}{C_1 - f_1} + \frac{f_2}{C_2 - f_2}$$

A més, s'ha de complir la condició de lligadura $f = \lambda L = f_1 + f_2$. Recordeu que el mètode utilitzat per optimitzar un camp escalar sotmès a una condició de lligadura són els multiplicadors de Lagrange.



La condició de lligadura és l'equació d'una recta i ens limita els valors f_1 i f_2 on hem de buscar l'òptim. En aquest cas, l'òptim ha de ser un punt que caigui sobre el segment de la figura 46. Estem suposant que no hi ha fluxos negatius ($f_i > 0$).

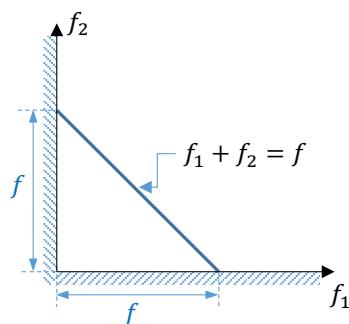


Figura 46.
Condició de lligadura

Suposem que el flux f que arriba al node A és pràcticament nul i que va augmentant progressivament. Vegem, en aquest cas, com seria el repartiment òptim del flux. Inicialment, tot el flux s'envia pel canal 1 (el de més capacitat). A mesura que el flux entrant augmenta, s'aproxima a un valor llindar f_t a partir del qual es comença a utilitzar el canal 2. Aquest comportament es mostra a la figura següent:

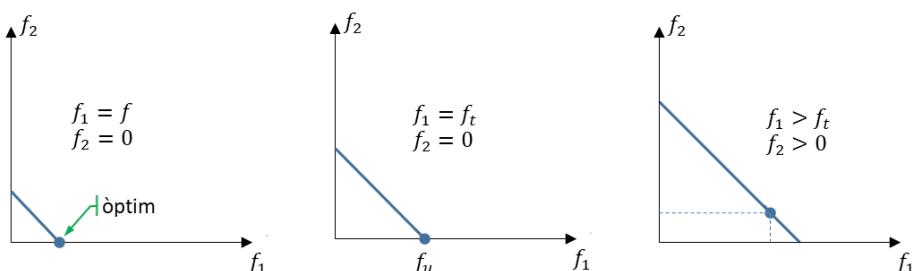


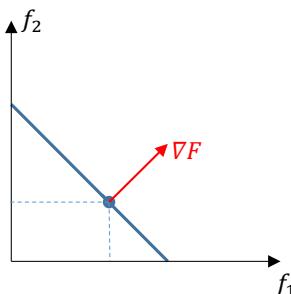
Figura 47.
Variació de l'òptim a
mesura que augmenta el flux que hi
arriba

Per calcular el repartiment òptim de f_1 i f_2 quan el flux que hi arriba és superior al flux llindar ($f > f_u$), hem d'utilitzar el mètode dels multiplicadors de Lagrange. Aquest mètode indica que, en l'òptim, el gradient del camp $F(f_1, f_2)$ ha de ser paral·lel al gradient de la condició de lligadura. Cal recordar que el gradient d'una funció és un vector que indica la direcció de variació màxima de la funció. A més, la projecció del gradient en qualsevol direcció indica quant val la derivada de la funció en aquesta direcció. Atès que el segment on hem de buscar la solució es correspon amb una col·lecció de punts pels quals $f_1 + f_2$ es manté constant, la direcció de màxima variació de la condició de lligadura ha de ser un vector que s'allunyi al més ràpidament possible d'aquest segment, és a dir, el gradient de la lligadura és perpendicular al segment.



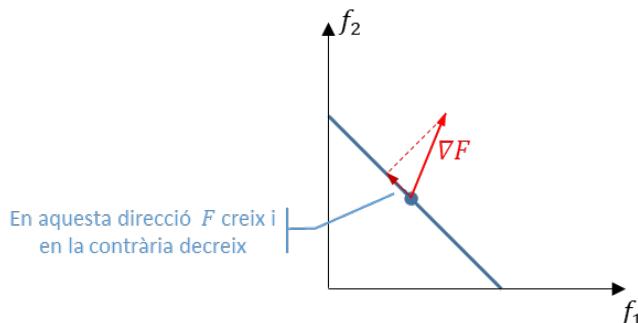
Per tant, en el punt òptim, s'ha de complir que el gradient de $F(f_1, f_2)$ ha de ser perpendicular al segment.

Figura 48.
Condició del punt òptim



Vegem ara una manera més intuïtiva d'entendre aquest concepte. Sabem que el punt òptim que estem buscant ha d'estar sobre el segment. Imaginem que escollim un punt sobre aquest segment pel qual $\nabla F(f_1, f_2)$ no és perpendicular al segment.

Figura 49.
Punt no òptim



En aquest cas, podem projectar el gradient sobre el segment. El mòdul del vector projectat ens indica el valor de la derivada en la direcció del segment. Com que aquesta derivada no és nul·la, si ens movem sobre el segment en aquesta direcció, la funció $F(f_1, f_2)$ creix i, si ens movem en sentit contrari, decreix; per tant, és evident que no estem en el punt òptim. Ens hem de moure sobre el segment en la direcció contrària a la projecció del gradient (ja que volem trobar el mínim de $F(f_1, f_2)$) fins que arribem a un punt pel qual el gradient sigui perpendicular al segment, de manera que no es pugui projectar ja que, mentre es pugui projectar, no haurem trobat l'òptim.

El vector $(1,1)$ és un dels possibles vectors perpendiculars al segment; per tant, ∇F ha de ser paral·lel a aquest vector:

$$\left(\frac{\partial F}{\partial f_1}, \frac{\partial F}{\partial f_2} \right) = \alpha(1,1)$$

Finalment, trobem la condició que ens permet calcular els valors de f_1 i f_2 òptims:



$$\frac{\partial F}{\partial f_1} = \frac{\partial F}{\partial f_2}$$

Aquesta condició és independent de la topologia de xarxa que tinguem. L'única condició que hem utilitzat per deduir-la és $f = f_1 + f_2$; per tant, es pot utilitzar per a qualsevol topologia, com la que es mostra a la figura.

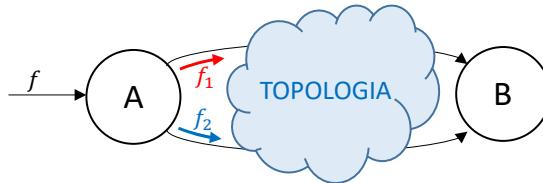


Figura 50.
Topologia genèrica
amb dos camins

Diferents topologies determinen diferents funcions $F(f_1, f_2)$. Si apliquem la condició per trobar l'òptim a la topologia senzilla que estem analitzant, tenim:

$$\frac{C_1}{(C_1 - f_1)^2} = \frac{C_2}{(C_2 - f_2)^2}$$

Utilitzem aquest resultat per calcular el flux llindar en aquest cas. Per fer-ho, utilitzem el fet que al llindar es compleix $f_1 = f_t$ i $f_2 = 0$. Substituint i aïllant f_t , se n'obté:

$$f_t = C_1 - \sqrt{C_1 C_2}$$

Per calcular com hem de repartir el flux f per optimitzar el temps de trànsit, hem de resoldre el sistema d'equacions següent:

$$\begin{cases} \frac{C_1}{(C_1 - f_1)^2} = \frac{C_2}{(C_2 - f_2)^2} \\ f = f_1 + f_2 \end{cases}$$

De la primera equació, traient l'arrel quadrada i invertint ambdós membres, n'obtem:

$$\frac{C_i - f_i}{\sqrt{C_i}} = K \Rightarrow f_i = C_i - K\sqrt{C_i}, \quad i = 1, 2$$

on K és una constant que calculem substituint f_i a la segona equació:

$$f = C_1 + C_2 - K(\sqrt{C_1} + \sqrt{C_2}) \Rightarrow K = \frac{C_1 + C_2 - f}{\sqrt{C_1} + \sqrt{C_2}}$$

Substituint la K :

$$f_i = C_i + \frac{\sqrt{C_i}}{\sqrt{C_1} + \sqrt{C_2}} [f - (C_1 + C_2)], \quad i = 1, 2$$



Aquest és el flux que s'envia per cada canal quan el flux que hi arriba és superior al flux llindar ($f > f_t$). Com que el flux que s'envia pel canal 1 és superior al flux llindar, volem expressar f_1 com la suma del flux llindar més un altre flux que hem de calcular. Per fer-ho, sumem i restem el terme $\sqrt{C_1} + \sqrt{C_2}$ a l'expressió de f_1 :

$$\begin{aligned}f_1 &= C_1 - \sqrt{C_1 C_2} + \frac{\sqrt{C_i}}{\sqrt{C_1} + \sqrt{C_2}} [f - (C_1 + C_2)] + \sqrt{C_1 C_2} = \\&= f_u + \frac{\sqrt{C_1}}{\sqrt{C_1} + \sqrt{C_2}} [f - (C_1 + C_2) + \sqrt{C_2}(\sqrt{C_1} + \sqrt{C_2})] = \\&= f_u + \frac{\sqrt{C_1}}{\sqrt{C_1} + \sqrt{C_2}} [f - C_1 + \sqrt{C_1 C_2}]\end{aligned}$$

Finalment:

$$f_1 = f_t + \frac{\sqrt{C_1}}{\sqrt{C_1} + \sqrt{C_2}} (f - f_t)$$

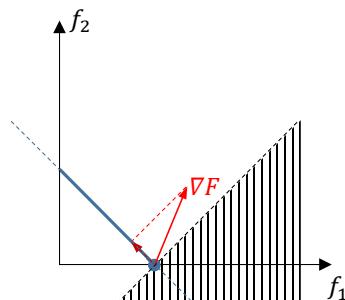
Atès que $f_1 + f_2 = f$, se n'obté:

$$f_2 = 0 + \frac{\sqrt{C_2}}{\sqrt{C_1} + \sqrt{C_2}} (f - f_t).$$

Les expressions anteriors es poden interpretar de la manera següent: "Per sota del flux llindar, tot el flux s'envia pel canal ràpid i, un cop superat el llindar, l'excés respecte al llindar es reparteix entre ambdós canals proporcionalment a l'arrel quadrada de les capitats".

Per acabar d'entendre què succeeix amb ∇F quan el flux que hi arriba està per sota del llindar, ens fixem en la figura següent:

Figura 51.
Gradient de
 $F(f_1, f_2)$ per sota
del flux llindar



El ∇F no pot estar a la zona ratllada ja que, si hi estigués, es podria projectar sobre la recta de manera que, recta avall, la funció F creix i, recta amunt, decreix. En aquest cas, seria possible pujar per la recta ($f_2 > 0$) i trobar un valor de F més



petit sobre el segment. Si ∇F és a la zona no ratllada i el projectem sobre la recta, tindrem que la funció F creix recta amunt ($f_2 > 0$) i decreix recta avall ($f_2 < 0$), però no podem anar recta avall, ja que el flux f_2 no pot ser negatiu i, per tant, l'òptim s'obté per a $f_2 = 0$. Matemàticament, el fet que el gradient no estigui a la zona ratllada es pot expressar amb la inequació següent:

$$\frac{\partial F}{\partial f_1} < \frac{\partial F}{\partial f_2}$$

3.4.2 Repartiment òptim per tres camins o més

Volem optimitzar el temps de trànsit de la xarxa de la figura 52:

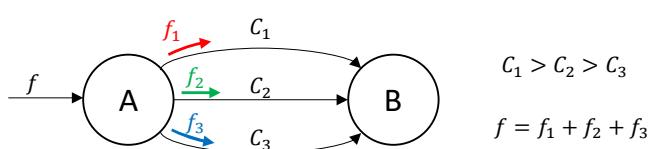


Figura 52.
Repartiment de fluxos per tres camins

$$T_{AB} = \frac{f_1}{f} \frac{L}{C_1 - f_1} + \frac{f_2}{f} \frac{L}{C_2 - f_2} + \frac{f_3}{f} \frac{L}{C_3 - f_3} = \frac{L}{f} \sum_{i=1}^3 \frac{f_i}{C_i - f_i}$$

$$F(f_1, f_2, f_3) = \sum_{i=1}^3 \frac{f_i}{C_i - f_i}$$

Hem de trobar el mínim del camp escalar F amb les equacions següents de lligadura:

$$f = \sum_{i=1}^3 f_i ; \quad f_i \geq 0; \quad i = 1,2,3$$

En aquest cas, tenim dos llindars, f_{t1} i f_{t2} . Per sota del primer llindar, només s'utilitza el canal 1 i, per sota del segon, només s'utilitzen els canals 1 i 2. A continuació, es descriuen les tres possibles situacions en funció del valor de f .

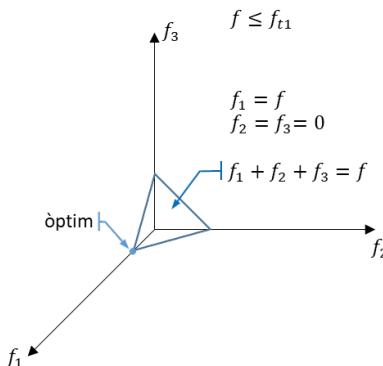
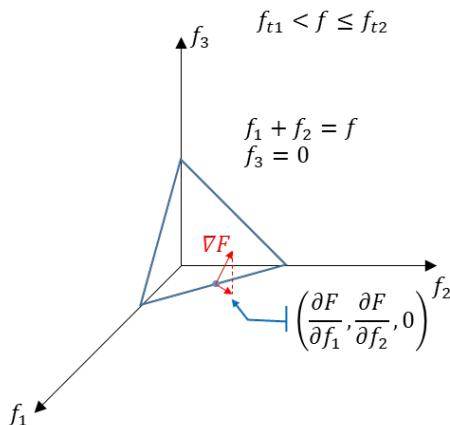


Figura 53.
Òptim per sota del primer llindar



L'equació $f = \sum_{i=1}^3 f_i$ defineix un pla i la intersecció d'aquest pla amb els plans cartesians ($f_i = 0; i = 1,2,3$) determina un triangle sobre el qual hem de trobar l'òptim. Els punts de tall del triangle amb els eixos (els vèrtexs del triangle) estan a una distància f de l'origen de coordenades. Quan el flux no supera el primer llindar ($f \leq f_{t1}$), aquest es transmet tot pel canal 1.

Figura 54.
Òptim entre el primer
llindar i el segon

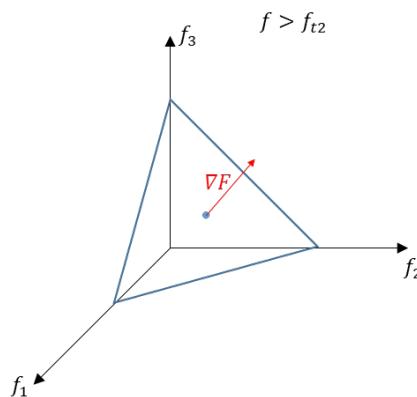


Quan el flux supera aquest llindar, es comença a utilitzar el canal 2 i, mentre no supera el segon llindar ($f_{t1} < f \leq f_{t2}$), el canal 3 no s'utilitza ($f_3 = 0$). En aquest cas, ∇F ha de ser perpendicular al costat inferior del triangle, de manera que no es pugui projectar sobre aquest costat (si es pogués projectar, voldria dir que ens podríem moure sobre aquest segment i trobar un valor més petit de F). Si projectem aquest gradient sobre el pla format pels eixos f_1 i f_2 , n'obtenim un vector format per les dues primeres components del gradient i un component nul. Aquest vector també és perpendicular al segment inferior. Per tant, en aquest cas, la condició que s'ha de complir és la mateixa que quan teníem només dos canals.

$$\frac{\partial F}{\partial f_1} = \frac{\partial F}{\partial f_2}$$

Aquesta condició també es compleix just al primer llindar, és a dir, quan $f = f_{t1}$.

Figura 55.
Òptim per sobre del
segon llindar





Quan el flux supera el segon llindar, f_3 deixa de valdre 0 i la solució òptima abandona el segment inferior del triangle. En aquest cas, el gradient al punt òptim ha de ser perpendicular al pla ja que, si no fos així, el podríem projectar sobre el pla i, en aquest cas, ens podríem moure sobre el pla fins a un valor més petit de F . El vector $(1,1,1)$ és perpendicular al pla; per tant, la condició que ha de complir el gradient per ser paral·lel a aquest vector és que tots els seus components siguin iguals.

$$\frac{\partial F}{\partial f_1} = \frac{\partial F}{\partial f_2} = \frac{\partial F}{\partial f_3}$$

Aquesta condició també es compleix just al segon llindar, és a dir, quan $f = f_{t2}$.

A la taula següent, es mostren les equacions que ens permeten calcular els llindars i els repartiments de flux per a diferents valors de f .

$f < f_{t1}$	$f_{t1} \leq f < f_{t2}$	$f \geq f_{t2}$
$\frac{\partial F}{\partial f_1} < \frac{\partial F}{\partial f_2}; \frac{\partial F}{\partial f_1} < \frac{\partial F}{\partial f_3}$	$\frac{\partial F}{\partial f_1} = \frac{\partial F}{\partial f_2} < \frac{\partial F}{\partial f_3}$	$\frac{\partial F}{\partial f_1} = \frac{\partial F}{\partial f_2} = \frac{\partial F}{\partial f_3}$
$f_1 = f$	$f_1 > 0$	$f_1 > 0$
$f_2 = 0$	$f_2 \geq 0$; igual si $f = f_{t1}$	$f_2 > 0$
$f_3 = 0$	$f_3 = 0$	$f_3 \geq 0$; igual si $f = f_{t2}$

Observeu que l'única condició que hem utilitzat per deduir la taula anterior és $f = \sum_{i=1}^3 f_i$, independentment de la funció $F(f_1, f_2, f_3)$. Per tant, la podem aplicar per a qualsevol topologia amb tres camins.

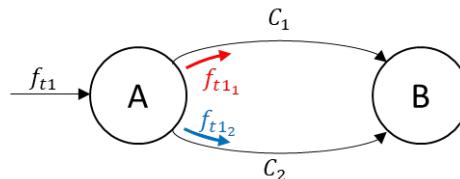
Calculem els llindars i el repartiment del flux per a l'exemple del principi del capítol, on hi ha tres canals directes entre els nodes d'origen i de destí.

Per a $f < f_{t1}$, la solució és trivial; tot el flux es transmet pel canal 1.

Quan $f_{t1} \leq f < f_{t2}$, tenim la mateixa situació que quan teníem dos canals. Escrivim el resultat obtingut amb una nomenclatura més compacta per tal de generalitzar-ne el resultat. En aquest cas, el flux que hi arriba s'ha de dividir en dos fluxos. Podem interpretar que, quan el flux que hi arriba és igual al primer llindar ($f = f_{t1}$), també es divideix, però, en aquest cas, $f_{t1_1} = f_{t1}$ i $f_{t1_2} = 0$.



Figura 56.
Repartiment quan el flux està entre el primer llindar i el segon



Amb aquesta interpretació, podem reescriure l'equació del llindar de la manera següent:

$$f_{t1} = \sum_{i=1}^2 f_{t1_i}$$

$$f_{t1_i} = C_i - \sqrt{C_i C_2}; i = 1,2$$

Observeu que, amb aquesta formulació: $f_{t1_2} = 0$ i $f_{t1} = f_{t1_1} = C_1 - \sqrt{C_1 C_2}$

El repartiment de flux també el podem escriure com:

$$f_i = f_{t1_i} + \frac{\sqrt{C_i}}{\sum_{j=1}^2 \sqrt{C_j}} (f - f_{t1}); i = 1,2$$

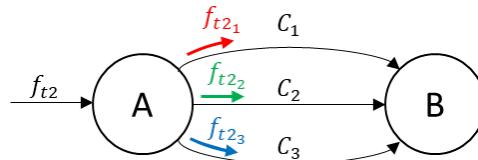
Aquesta nomenclatura s'adapta molt bé al concepte: "Un cop superat el llindar, l'excés respecte al llindar es reparteix entre ambdós canals proporcionalment a l'arrel quadrada de les capacitats."

Per al cas $f \geq f_{t2}$, el flux que hi arriba es reparteix entre els tres canals i ha de complir l'equació següent:

$$\left. \begin{aligned} \frac{\partial F}{\partial f_1} &= \frac{\partial F}{\partial f_2} = \frac{\partial F}{\partial f_3} \\ F(f_1, f_2, f_3) &= \sum_{i=1}^3 \frac{f_i}{C_i - f_i} \end{aligned} \right\} \Rightarrow \frac{C_1}{(C_1 - f_1)^2} = \frac{C_2}{(C_2 - f_2)^2} = \frac{C_3}{(C_3 - f_3)^2}$$

Al llindar, el flux també es reparteix pels tres canals: $f_{t2} = \sum_{i=1}^3 f_{t2_i}$, essent $f_{t2_3} = 0$.

Figura 57.
Repartiment quan el flux està per sobre del segon llindar



Per tant, substituint a l'equació anterior:



$$\frac{C_i}{(C_i - f_{t2i})^2} = \frac{1}{C_3}; \quad i = 1, 2$$

i, aïllant:

$$f_{t2i} = C_i - \sqrt{C_i C_3}; \quad i = 1, 2, 3$$

$$f_{t2} = \sum_{i=1}^3 f_{t2i}$$

Per calcular el repartiment del flux, resolem el sistema d'equacions següent:

$$\begin{cases} \frac{C_i}{(C_i - f_i)^2} = \text{constant;} \\ f = \sum_{i=1}^3 f_i; \quad f_i \geq 0; \quad i = 1, 2, 3. \end{cases}$$

De la primera equació, traient l'arrel quadrada i invertint ambdós membres, se n'obté:

$$\frac{C_i - f_i}{\sqrt{C_i}} = K \Rightarrow f_i = C_i - K\sqrt{C_i}, \quad i = 1, 2, 3$$

on K és una constant que calculem substituint f_i a la segona equació:

$$f = \sum_{j=1}^3 C_j - K \sum_{j=1}^3 \sqrt{C_j} \Rightarrow K = \frac{(\sum_{j=1}^3 C_j) - f}{\sum_{j=1}^3 \sqrt{C_j}}$$

Substituint el valor de K :

$$f_i = C_i + \frac{\sqrt{C_i}}{\sum_{j=1}^3 \sqrt{C_j}} \left(f - \sum_{j=1}^3 C_j \right), \quad i = 1, 2, 3$$

Ho relacionem amb el flux llindar sumant i restant el terme $\sqrt{C_i C_3}$:

$$f_i = C_i - \sqrt{C_i C_3} + \frac{\sqrt{C_i}}{\sum_{j=1}^3 \sqrt{C_j}} \left(f - \sum_{j=1}^3 C_j + \sqrt{C_i C_3} \right)$$

$$f_i = f_{t2i} + \frac{\sqrt{C_i}}{\sum_{j=1}^3 \sqrt{C_j}} \left(f - \left(\sum_{j=1}^3 C_j + \sqrt{C_3} \sum_{j=1}^3 \sqrt{C_j} \right) \right)$$



Finalment:

$$f_i = f_{t2i} + \frac{\sqrt{C_i}}{\sum_{j=1}^3 \sqrt{C_j}} (f - f_{t2}); \quad i = 1, 2, 3$$

Aquesta expressió també es pot interpretar com: "Un cop superat el segon lliniar, l'excés respecte al lliniar es reparteix entre els tres canals proporcionalment a l'arrel quadrada de les seves capitats."

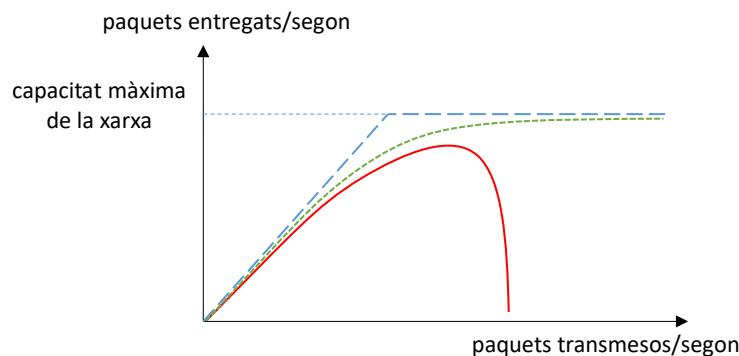
El procediment que es presenta per a tres canals es pot generalitzar per a quatre canals o més. Per a una topologia amb n canals directes entre l'origen i el destí, la solució és immediata i es poden generalitzar els resultats obtinguts per a tres canals. Per a d'altres topologies, cal resoldre les equacions:

$$\begin{cases} \frac{\partial F}{\partial f_i} = \text{constant}; \quad i = 1, 2, \dots, n \\ f = \sum_{i=1}^n f_i; \quad f_i \geq 0; \quad i = 1, 2, \dots, n \end{cases}$$

3.5 Control de congestió

S'entén per *congestió* en una xarxa de comunicacions la degradació de les prestacions ofertes deguda a la presència d'un volum excessiu de trànsit. A la figura següent, es mostra com és la taxa de paquets entregats al destí respecte de la taxa de paquets enviats en una xarxa.

Figura 58.
Relació entre els paquets transmesos i els paquets entregats



Atès que la xarxa té una capacitat màxima, no podem superar la línia de ratlles blava. La línia de punts verda representa el comportament que voldríem, just per sota de les limitacions de la xarxa. La línia sòlida en vermell dibuixa el comportament real d'una xarxa en què, a partir d'un volum de trànsit, el nombre de paquets entregats disminueix. De fet, superat aquest volum màxim, tota la xarxa es congestionada ja que, en disminuir el nombre de paquets entregats, el nombre



de retransmissions augmenta, és a dir, augmenta el volum de paquets transmèsos dins la xarxa, i això fa que el nombre de paquets entregats acabi tendint a zero.

La congestió és un problema global de tota la xarxa, ja que, si un node es congestionava (té els seus *buffers* plens), comença a perdre paquets del seu node antecessor. Si, per exemple, el node antecessor necessita rebre ACK que li confirmen la recepció correcta dels paquets, no pot alliberar els seus *buffers* i també es congestionava. Aquest fenomen es propaga i finalment es pot congestionar tota la xarxa.

Els mecanismes de control de congestió intenten evitar que la xarxa arribi a una situació de congestió. Aquests mecanismes els podem classificar en:

- Preventius (o de llaç obert): Actuen tota l'estona sobre el trànsit de la xarxa en prevenció d'una situació de congestió.
- Reactius (o de llaç tancat): Actuen quan en algun punt de la xarxa es detecta una situació d'alerta de congestió.

A continuació, es mostren els principals mecanismes de control de la congestió i s'utilitzen les tècniques de la teoria de cues per avaluar-los.

3.5.1 Descart de paquets

Aquesta és una tècnica reactiva en la qual els nodes fan una assignació dinàmica de la grandària dels *buffers* a cada sistema de transmissió en funció de l'estat de la xarxa. Quan un paquet hi arriba i ha de ser transmès per un sistema de transmissió que té el *buffer* ple, es descarta. L'assignació de la grandària dels *buffers* es pot fer de diverses maneres:

- Assignació dinàmica: s'assigna un *buffer* més gran a les línies de sortida que cursen més trànsit (les que en necessiten més). Pot ser que una línia acapari els recursos, la qual cosa pot ser injusta per a la resta.
- Assignació fixa: tots els sistemes de transmissió tenen la mateixa grandària de *buffer*. Es pot donar el cas de fer un ús ineficient dels recursos assignant més *buffer* del necessari a uns sistemes de transmissió que no el necessiten.
- Assignació subòptima: barreja les dues tècniques anterior. Tots els sistemes de transmissió tenen assignat un *buffer* amb una grandària fixa i la resta de la memòria del node s'assigna dinàmicament a cada sistema de transmissió en funció de les seves necessitats.

Per tal d'avaluar aquesta tècnica, es modela el sistema de transmissió com un sistema markovià amb un *buffer* finit i s'estudia l'ocupació del canal o la taxa cursada de paquets en funció de la grandària del *buffer*.



3.5.2 Regulació explícita de la taxa

Aquesta també és una tècnica reactiva en la qual els modes monitoritzen els seus sistemes de transmissió. Quan detecten una situació d'alerta de congestió (alta ocupació dels *buffers*, utilització del canal per sobre d'un límit...), envien un paquet regulador a l'estació origen del trànsit que ha causat l'alerta. Quan l'estació origen rep el paquet regulador, disminueix el trànsit que introduceix a la xarxa en un percentatge determinat. Si, transcorregut un temps (*timeout*), l'estació origen no rep cap paquet regulador torna a augmentar el trànsit que envia a la xarxa.

Per tal d'avaluar aquesta tècnica, una possibilitat és modelar el sistema de transmissió com un sistema markovià per què la taxa d'arribada disminueix a mesura que l'ocupació del sistema augmenta.

3.5.3 Traffic shaping

Aquesta és una tècnica preventiva que consisteix a controlar el volum de trànsit a l'entrada de la xarxa, de manera que aquest entri a una taxa determinada, i es descarta el trànsit quan se supera aquesta taxa. Es pretén "donar forma" al trànsit a l'entrada de la xarxa. Principalment, hi ha dos algorismes que conformen el trànsit: *leaky bucket* i *token bucket*.

Leaky bucket

Aquest algorisme consisteix en un *buffer* on s'emmagatzemen els paquets que han d'entrar a la xarxa. Cada determinat interval temps es deixa que un nombre determinat de paquets entri a la xarxa; d'aquesta manera, els paquets hi entren a una taxa constant. Si els paquets superen la capacitat del *buffer* es descarten. Aquest mecanisme és similar a una gallada amb un forat. A la gallada, hi cau el líquid i aquest en surt a una taxa constant pel degoteig. Quan el líquid supera la capacitat de la gallada, es perd. D'aquí ve el nom de *leaky bucket* ("gallada amb degoteig").

Si, durant molt de temps, no s'introduceix trànsit a la xarxa, potser seria convenient recordar-ho i permetre l'entrada d'una ràfega quan es vulgui accedir a la xarxa. Per fer-ho, s'utilitza un mecanisme basat en crèdits (o permisos), anomenat *token bucket*.

Token bucket ("per permisos")

Aquest mecanisme també consisteix en un *buffer* on s'emmagatzemen els paquets que han d'entrar a la xarxa. Cada determinat temps, es genera un permís o crèdit anomenat *token*, que permet que un nombre determinat de paquets entri a la xarxa. D'aquesta manera, es regula l'entrada de paquets a la xarxa. Quan no hi ha trànsit, els permisos s'emmagatzemen en un *buffer*. Així, quan hi hagi trànsit, cada permís permetrà que un nombre determinat de paquets entri a la xarxa i es genererà una ràfega. El *buffer* de permisos és finit i delimita la grandària màxima de la ràfega de paquets que pot entrar a la xarxa. La taxa a la



qual es generen els permisos determina la taxa d'entrada de paquets a la xarxa. Podem interpretar el *leaky bucket* com un *token bucket* en el qual el *buffer* de permisos és nul.

Podem avaluar aquestes tècniques si considerem que tant la generació de paquets com la de permisos responen a un procés de Poisson i que cada permís habilita l'entrada d'un paquet a la xarxa. En aquest cas, podem construir una cadena de Márkov. Definim cada estat per una parella ordenada de valors que indica el nombre de paquets i el nombre de permisos. Cal tenir en compte que sempre algun dels dos valors ha de valdre zero, ja que cada permís permet l'entrada d'un paquet a la xarxa i, per tant, no hi pot haver al mateix temps permisos i paquets en els *buffers*.

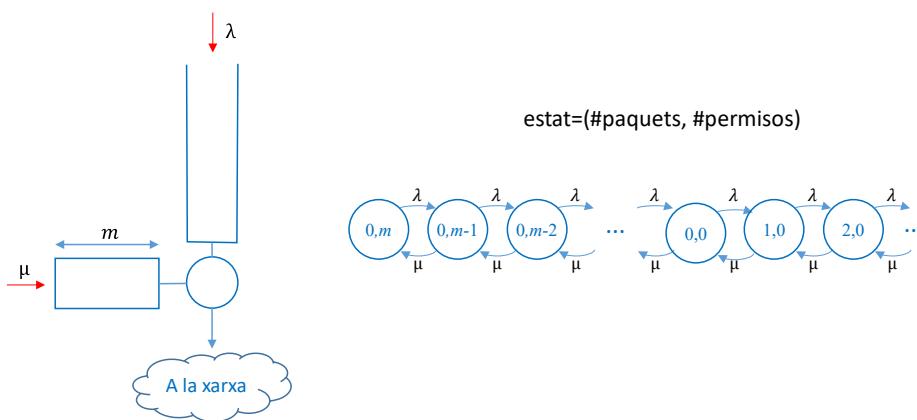


Figura 59.
Anàlisi del control de congestió amb permisos

A la figura 59, el temps entre arribades de paquets i el temps entre arribades de permisos són variables aleatòries exponencials de taxa λ paquets/s i μ permisos/s, respectivament. Quan hi arriba un paquet, o bé desapareix un permís o bé augmenta el nombre de paquets en el *buffer* de paquets. En ambdós casos, el sistema evoluciona cap a la dreta i ho fa amb una taxa λ . Anàlogament, quan hi arriba un permís, el sistema evoluciona cap a l'esquerra, amb una taxa μ . Com que el *buffer* de permisos és finit i de grandària m , l'estat de l'extrem esquerre de la cadena és l'estat $(0, m)$. Si el *buffer* de paquets és infinit, la cadena és infinita cap a la dreta; si no, la cadena serà finita. Un cop plantejat el problema, el resolem aplicant els coneixements que tenim de sistemes markovians.

Per exemple, per al cas d'un *buffer* de paquets infinit, tindríem un sistema modelable com M/M/1, amb les probabilitats d'estat següents:

$$p_{0,m} = 1 - \frac{\lambda}{\mu}$$

$$p_{0,m-k} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^k, k = 1..m$$



$$p_{k,0} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{m+k}, k > 0$$

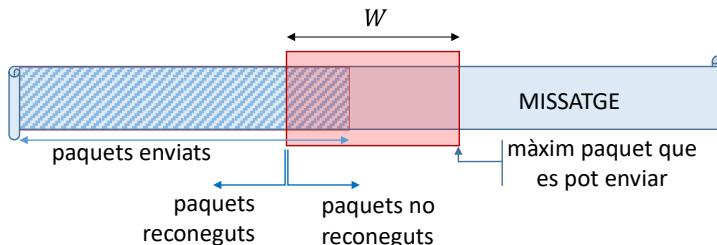
Un valor interessant és la probabilitat que un paquet hagi d'esperar a accedir a la xarxa, cosa que, tenint en compte la propietat PASTA, en aquest cas es calcularia com:

$$P_{d,access} = \sum_{k=0}^{\infty} p_{k,0} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^m \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k = \left(\frac{\lambda}{\mu}\right)^m$$

3.5.4 Mecanismes amb finestra

Aquest és un mecanisme que s'aplica en moltes ocasions extrem a extrem entre el node d'origen i el node de destí. És a dir, la responsabilitat del control de la congestió a la xarxa es deixa als seus usuaris. Una possibilitat és utilitzar un protocol de transport amb una finestra lliscant per controlar el volum de trànsit a la xarxa. La finestra de l'emissor és un *buffer* de grandària W on s'emmagatzem els paquets transmesos. Aquest *buffer* no s'allibera fins que es rep el reconeixement (ACK) del node de destí.

Figura 60.
Finestra lliscant



Es defineix el RTT (*round-trip time*) com el temps d'anada del paquet fins al node de destí més el temps de tornada del seu ACK. Com més gran sigui el RTT, més congestionada estarà la xarxa. Suposem que el temps de transmissió dels paquets és fix i val T_s . Per avaluar aquest mecanisme, calclem quina és la taxa de transmissió de paquets del node d'origen al node de destí.

Finestra fixa

Si la finestra és fixa i té una grandària W (mesurada en paquets), se'n distingeixen dues situacions:

- El temps de transmissió d'una finestra és inferior a RTT, és a dir, $WT_s < RTT$.
- El temps de transmissió d'una finestra és superior a RTT, és a dir, $WT_s > RTT$.

Ambdues situacions es mostren a la figura següent. Per a cadascuna de les situacions, es dibuixa un eix de temps per a l'estació d'origen i l'estació de destí. A la part superior de l'eix, es mostren els paquets que transmet l'estació (TX) i, a la part inferior, els paquets que rep (RX). En aquest cas, la finestra és per a cinc paquets.

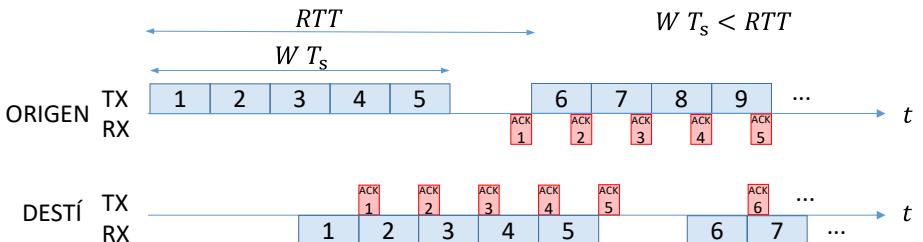
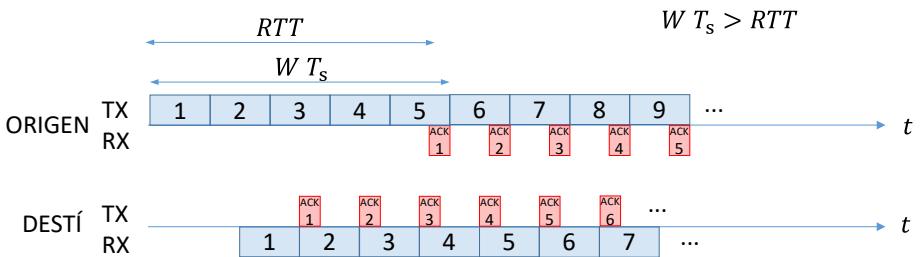


Figura 61.
Taxa de transmissió per mecanisme amb finestra



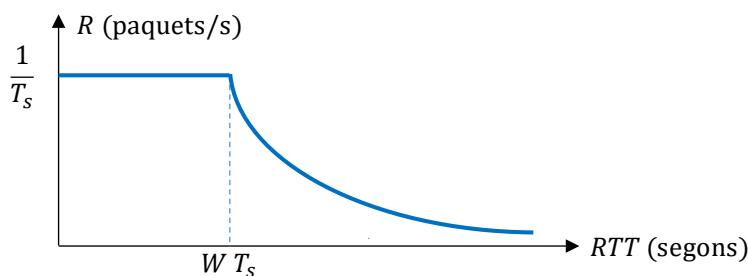
Quan el temps de transmissió d'una finestra és inferior al RTT, només es poden transmetre W paquets cada RTT segons. D'altra banda, si el RTT es inferior al temps de transmissió d'una finestra, rebem l'ACK abans d'haver transmès tota la finestra i podem anar-la lliscant de manera que els paquets es poden transmetre de manera continuada. En aquest cas, la taxa de transmissió és un paquet cada T_s segons.

$$\begin{cases} R = \frac{W}{RTT} & \text{si } W T_s < RTT \\ R = \frac{1}{T_s} & \text{si } W T_s > RTT \end{cases}$$

Ambdues expressions coincideixen per $W T_s = RTT$. Si ho representem gràficament així:



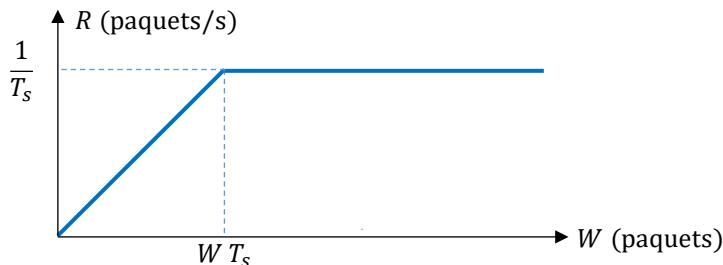
Figura 62.
Relació entre la taxa de transmissió i RTT



Com s'observa a la figura 62, quan la xarxa comença a congestionar-se i el seu RTT augmenta, el node d'origen disminueix la taxa de transmissió.

També és possible fer una interpretació mantenint constant el valor de RTT i fent variar la grandària de la finestra. Com es pot observar a la figura següent, grandàries de finestra per sota del líindar $W = RTT/T_s$ redueixen la taxa de transmissió, mentre que amb grandàries superiors al líindar la taxa se satura en el seu valor màxim.

Figura 63.
Relació entre la taxa de transmissió i el RTT



Finestra variable

De les equacions anteriors, s'observa que, donat un RTT, podem controlar la taxa de transmissió modificant la grandària de la finestra. D'aquesta manera, l'estació d'origen pot anar variant la grandària de la seva finestra, en prevenció d'una situació de congestió. L'objectiu és que la taxa de transmissió sigui la més gran possible, sense arribar mai a una situació de congestió. Un algorisme bastant utilitzat a la capa de transport és l'anomenat *additive increase/multiplicative decrease*. Aquest algorisme va augmentant progressivament la taxa de transmissió mentre l'estació d'origen no té cap alerta de congestió i, en cas d'alerta, disminueix la taxa de transmissió dràsticament. El fenomen que causa l'alerta de congestió a l'estació d'origen és la no-recepció de l'ACK corresponent a un paquet. L'estació va fent una estimació del RTT i, un cop transcorregut aquest temps després de l'emissió d'un paquet, si no ha rebut l'ACK corresponent, considera que el paquet s'ha perdut. Així doncs, quan tots els paquets corresponents a una finestra han estat reconeguts, la grandària de la finestra augmenta



en una unitat (*additive increase*) i, si no rep l'ACK d'algún paquet (paquet perdut), la grandària disminueix a la meitat (*multiplicative decrease*). Quan l'algorisme es posa en funcionament, per evitar que la taxa de transmissió augmenti molt lentament, la grandària de la finestra augmenta exponencialment fins que es produeix la pèrdua d'un paquet. A partir d'aquest punt, la grandària de la finestra es controla amb l'algorisme *additive increase/multiplicative decrease*. Per tant, inicialment tenim un petit transitori que ens fixa el punt de partida de l'algorisme.

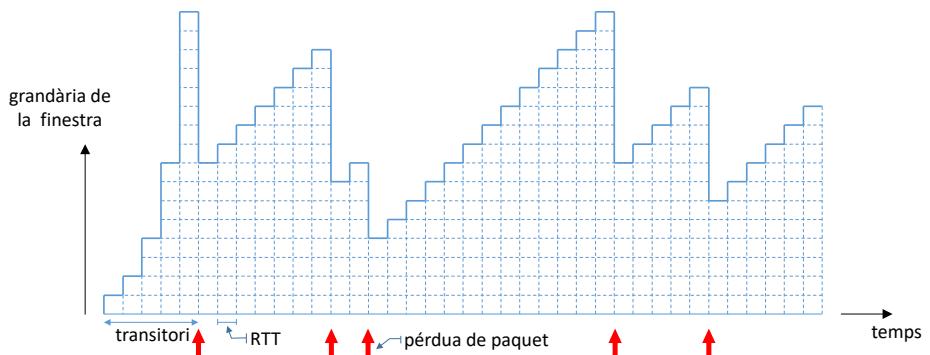


Figura 64.
Additive in-
crease/multiplicative
decrease

Dins de cada cicle, podem calcular la taxa de transmissió com el quocient entre el nombre de paquets transmesos en el cicle i la seva durada. Però, com que cada cicle és diferent, aquesta taxa de transmissió també serà diferent. Per tal de fer un càlcul del valor mitjà de la taxa de transmissió, es fa l'aproximació següent: es considera que la taxa mitjana de transmissió serà molt semblant a la que obtindríem si tots els cicles fossin iguals.

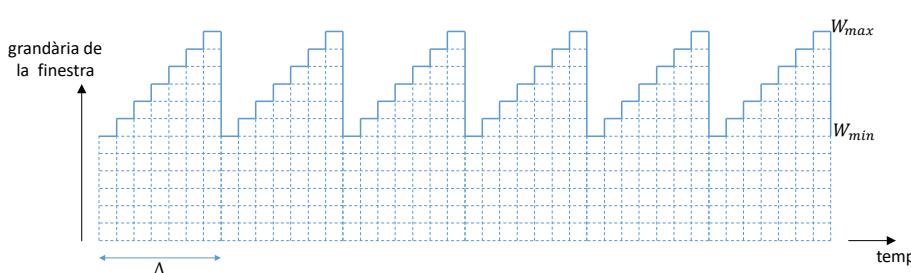
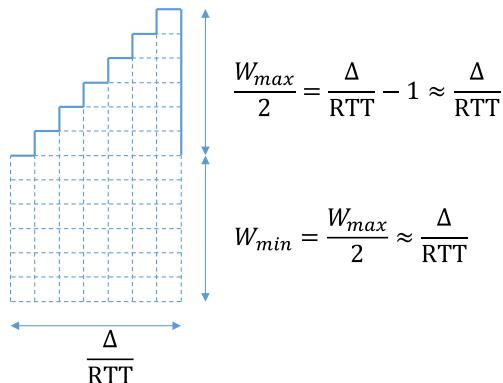


Figura 65.
Model aproximat per
analitzar additive in-
crease/multiplicative
decrease

En aquest cas, la grandària de la finestra màxima (W_{max}), la grandària de la finestra mínima (W_{min}) i la durada d'un cicle (Δ) són les mateixes a tots els cicles. Amb aquesta aproximació, com que tots els cicles són iguals, podem calcular la taxa de transmissió considerant només un cicle.



Figura 66.
Relació entre
 W_{max} , W_{min} , Δ i
RTT



A la figura 66, l'amplada de la base es correspon amb el nombre de RTT que hi ha en un cicle i val Δ/RTT . Podem relacionar la grandària màxima de la finestra amb la durada d'un cicle:

$$\frac{W_{max}}{2} = \frac{\Delta}{RTT} - 1 \approx \frac{\Delta}{RTT}$$

El nombre de quadrats que hi ha a la figura d'un cicle representa el nombre de paquets transmesos en el cicle. Per comptar-los, tenim en compte que cada cicle està format per una base rectangular i a sobre té un triangle. L'altura del rectangle i del triangle és la mateixa i val $W_{max}/2$, i la base val Δ/RTT . Per tant, utilitzant l'aproximació anterior, tenim que el nombre de paquets transmesos és:

$$\frac{W_{max}}{2} \frac{\Delta}{RTT} + \frac{1}{2} \frac{W_{max}}{2} \frac{\Delta}{RTT} = \frac{3W_{max}^2}{8}$$

Dividint el nombre de paquets perduts en un cicle pel nombre total de paquets en un cicle, s'obté la probabilitat de pèrdua. Fent la suposició que a cada cicle s'ha percut un únic paquet:

$$pp = \frac{8}{3W_{max}^2}$$

Aïllant la grandària màxima de la finestra, se n'obté:

$$W_{max} = \sqrt{\frac{8}{3pp}}$$

La durada d'un cicle la calculem a partir de l'aproximació que n'hem fet:



$$\Delta \approx \frac{W_{max}}{2} RTT = \sqrt{\frac{2}{3 pp}} RTT$$

Finalment, la taxa de transmissió dels paquets la calculem com el quocient entre el nombre de paquets dins un cicle, dividit per la seva durada:

$$R = \frac{1/pp}{\Delta} = \frac{1}{RTT} \sqrt{\frac{3}{2 pp}}$$

→ 4



Tècniques d'accés múltiple

Als capítols anteriors, ens hem centrat en les xarxes de transport, les quals normalment són xarxes commutades que utilitzen la tècnica de commutació de paquets. A les xarxes d'accés, l'escenari és diferent; normalment, hi ha un recurs que comparteixen diferents estacions, cosa que permet l'accés a la xarxa. Pot ser tant una estació base com un punt d'accés, un satèl·lit o un medi físic compartit com un cable. Atès que el recurs és compartit, cal establir la manera com les diferents estacions accedeixen al recurs quan el necessiten, és a dir, la *tècnica d'accés* al recurs.

Aquestes tècniques es poden classificar en:

- Tècniques aleatòries. Les estacions hi accedeixen quan ho necessiten. En aquest cas, es poden produir conflictes quan més d'una estació vol accedir al recurs simultàniament. Aquests conflictes els anomenem *col·lisions*.
- Tècniques deterministes. Hi ha un ordre preestablert a l'hora d'accendir al recurs, de manera que s'elimina la possibilitat que es produixin col·lisions.

L'avantatge de les tècniques deterministes és que no s'hi produeixen col·lisions. L'inconvenient és que poden estar assignant el recurs compartit a una estació que no el necessita, amb la qual cosa no s'aprofita bé la utilització del recurs. En general, es podria dir que, en situacions amb un alt volum de trànsit, són millors les tècniques deterministes i, quan el volum de trànsit és baix, són millors les aleatòries. Per fer un símil, podem pensar en l'accés a una cruïlla: quan hi ha molt de trànsit de vehicles, el millor és posar-hi un semàfor (determinista) i, quan n'hi ha poc, un "cedir el pas" (una mica més aleatori).



4.1 Tècniques aleatòries

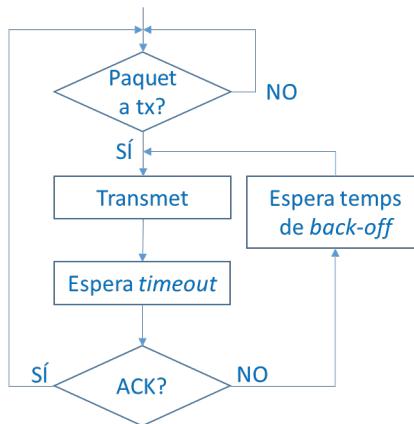
D'entre tècniques aleatòries s'analitza la tècnica Aloha, tant en la seva versió pura com en la versió amb ranures temporals. També s'analitza la tècnica d'accés amb escolta de portadora CSMA, però només el cas no persistent i sense ranures. Finalment, s'analitza la tècnica d'accés amb escolta de portadora amb detecció de col·lisió CSMA/CD, també en la modalitat no persistent i sense ranures.

4.1.1 Aloha

El funcionament de la tècnica Aloha és el següent:

- Quan una estació té la necessitat d'accendir-hi, ho fa.
- L'estació transmissora ha de tenir coneixement de si la transmissió s'ha realitzat amb èxit o s'ha produït una col·lisió. Una possibilitat és esperar la recepció d'un reconeixement (*acknowledge*, ACK) transmès per l'estació receptora.
- Si, transcorregut un *timeout*, l'estació transmissora no rep l'ACK, ha de procedir a retransmetre el paquet. Per tal d'evitar una nova col·lisió, abans de fer la retransmissió s'espera durant un temps aleatori anomenat *temps de back-off*.
- Després d'aquest temps, torna a fer un nou intent d'accés.
- L'algorisme es repeteix fins que l'estació transmissora rep un ACK.

Figura 67.
Funcionament de la
tècnica Aloha



Anàlisi de la tècnica Aloha pur amb un nombre infinit d'estacions

Per fer-ne l'anàlisi, utilitzem les hipòtesis següents:

- El nombre d'estacions és infinit.
- Tots els paquets tenen la mateixa longitud, és a dir, el seu temps de transmissió és constant i val T_t .

- Els ACK es transmeten per un canal a part de les dades, de manera que no es produeixen col·lisions entre ells i els paquets.
- En canal està lliure de soroll; si no es rep l'ACK és a causa d'una col·lisió.
- Els intents d'accés al canal són un procés de Poisson de taxa Λ . La probabilitat de tenir k intents d'accés en t segons val:

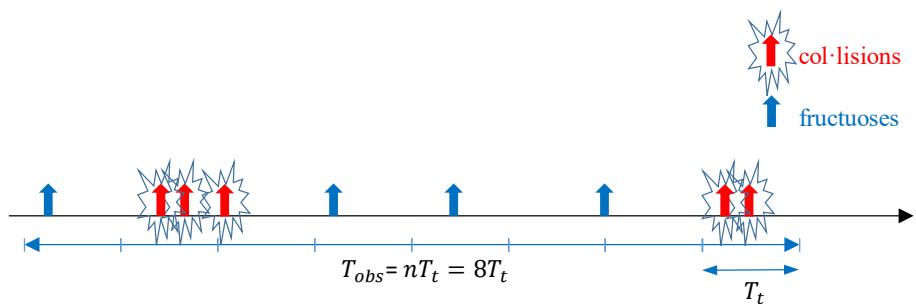
$$P_k(t) = \frac{(\Lambda t)^k}{k!} e^{-\Lambda t}$$

Λ inclou:

- paquets nous generats per les estacions
- retransmissions de paquets que havien col·lidit

Com que el nombre d'estacions és infinit, la taxa d'intents d'accés Λ és constant. És a dir, malgrat que hi ha estacions que no hi intenten accedir perquè estan a l'espera d'un ACK, la taxa no disminueix. No és un cas de població finita.

A la figura 68, es mostren els intents d'accés de les estacions en un temps d'observació T_{obs} . Quan la separació temporal entre dos intents és inferior a T_t , s'ha produït una col·lisió entre ells.



$$\begin{aligned} l &= 4 \text{ transmissions fructuosas en } T_{obs} \\ m &= 9 \text{ intents d'accés en } T_{obs} \\ n &= 8 \text{ } T_t \text{ en } T_{obs} \end{aligned}$$

Figura 68.
Exemple de funcionament de la tècnica Aloha

Utilitzem l'exemple de la figura per explicar un parell de conceptes.

Es defineix el **cabal S** com el **nombre mitjà de transmissions fructuoses (sense col·lisió) cada T_t** . Mirant l'exemple de la figura, en una primera aproximació el podríem calcular com:

$$S \approx \frac{l}{n} = \frac{4}{8} = 0,5$$



El valor obtingut no seria una aproximació si el temps d'observació fos infinit. El cabal sempre serà menor que 1 ja que no és possible tenir més d'una transmissió fructuosa en el temps de transmissió d'un paquet. Si a l'expressió anterior multipliquem el numerador i el denominador per T_t , obtenim el quotient entre el temps durant el qual s'està transmeten fructuosament en el T_{obs} , dividit pel T_{obs} . Per tant, el cabal també ens indica la **fracció de temps durant la qual s'està transmetent fructuosament**.

Si dividim el cabal per T_t , obtenim el nombre de transmissions fructuosas per segon. Si la tècnica Aloha és estable, és a dir, no es congestionada, el nombre de transmissions fructuosas per segon ha de coincidir amb el nombre mitjà de paquets nous per segon que generen les estacions. La taxa de paquets nous de totes les estacions l'anomenem λ . Per tant:

$$S = \lambda T_t$$

La càrrega G es defineix com el **nombre mitjà d'intents d'accés cada T_t** . Mirant l'exemple de la figura 68, en una primera aproximació el podríem calcular com:

$$G \approx \frac{m}{n} = \frac{9}{8} = 1,125$$

El valor de la càrrega pot ser més gran que 1, la qual cosa significa que es fa més d'un intent d'accés en el temps de transmissió d'un paquet. Si a l'expressió anterior multipliquem el numerador i el denominador per T_t , obtenim el quotient entre el temps durant el qual es vol ocupar el recurs compartit en el T_{obs} , dividit pel T_{obs} . Per tant, la càrrega també ens indica la **fracció de temps durant la qual es vol ocupar el recurs compartit**. Tenint en compte aquesta darrera interpretació de G , la càrrega també s'anomena *trànsit ofert al canal* (els segons de feina oferts al canal cada segon).

Si dividim la càrrega per T_t , obtenim el nombre mitjà d'intents d'accés per segon, és a dir, Λ i, per tant:

$$G = \Lambda T_t$$

També ho podem interpretar com el nombre mitjà d'intents en T_t per a un procés de Poisson de taxa Λ .

Per tal de fer l'anàlisi de la tècnica Aloha, volem relacionar els valors de S i G , la qual cosa és equivalent a relacionar λ i Λ .

Suposem que una de les infinites estacions fa un intent d'accés i volem conèixer la probabilitat que el seu intent sigui fructuós. Ho podem calcular dividint els casos favorables pels casos possibles:

$$p(\text{intent fructuós}) = \lim_{T_{obs} \rightarrow \infty} \frac{\lambda T_{obs}}{\Lambda T_{obs}} = \frac{\lambda}{\Lambda} = \frac{S}{G}$$

$$S = G p(\text{intent fructuós})$$

Vegem ara quines condicions s'han de donar perquè aquest intent sigui fructuós.

A la figura 69, es mostra la transmissió del paquet objecte d'estudi i els límits de les transmissions del paquet anterior i posterior per tal que no col·lideixin amb ell.



Figura 69.
Interval de vulnerabilitat per la tècnica Aloha

Si el paquet anterior es transmet més tard o el paquet posterior es transmet abans, es produeix una col·lisió amb el paquet objecte d'estudi. Per tant, aquests límits defineixen l'interval de vulnerabilitat, de manera que qualsevol intent d'accés dins aquest interval provoca una col·lisió amb el paquet objecte d'estudi. Aquest interval té una durada de $2T_t$. Així doncs, la probabilitat que l'intent de transmissió sigui fructuós coincideix amb la probabilitat que la resta d'estacions no hi intentin accedir en l'interval de vulnerabilitat.

$$p(\text{intent fructuós}) = p(0 \text{ intents d'accés en } 2T_t) = \frac{(\Lambda 2T_t)^0}{0!} e^{-\Lambda 2T_t} = e^{-2G}$$

Finalment, ajuntant ambdues equacions, se n'obté:

$$S = G e^{-2G}$$

A continuació, es mostra la representació gràfica del cabal en funció de la càrrega. El cabal màxim que podem tenir val $1/2e = 0,1839$, quan la $G = 0,5$.

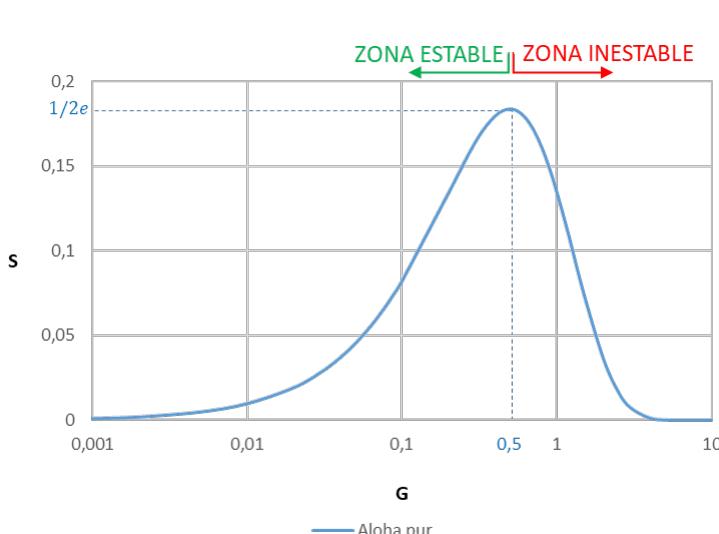


Figura 70.
Relació entre la càrrega i el cabal per la tècnica Aloha pur



A la gràfica, es defineixen dues zones, la zona estable (quan la càrrega és inferior a 0,5) i la zona inestable (quan és superior). Suposem per un moment que el nombre d'intents d'accés augmenta, de manera que s'entra a la zona inestable. En aquest cas, el cabal disminueix, és a dir, el nombre de paquets transmesos fructuosament disminueix. Això implica que les retrasmissons augmenten i, per tant, el nombre d'intents d'accés també. Com es pot observar, aquest fenomen es realimenta fins al punt que el cabal tendeix a 0 (s'ha produït una situació de congestió). És per això que la zona de la dreta s'anomena *zona inestable*. Per tant, tenim:

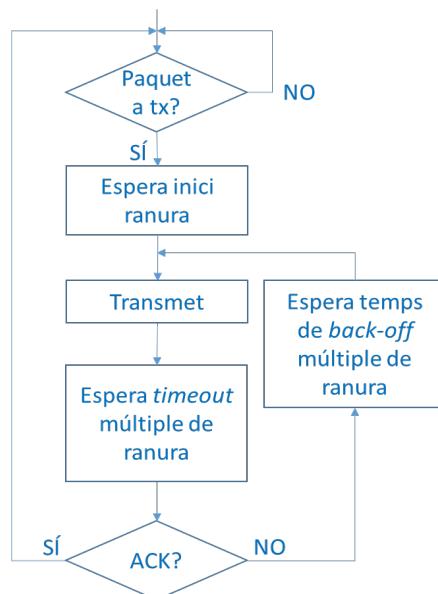
- A la zona estable, el nombre de paquets transmesos fructuosament cada segon coincideix amb el nombre de paquets nous generats per les estacions cada segon: $S = \lambda T_t$
- A la zona inestable, el nombre de paquets transmesos fructuosament és nul ($S = 0$) i el nombre de paquets per segon generats per les estacions supera el valor màxim que admet la xarxa (λ_{\max}).
- S'entra a la zona inestable quan les estacions generen més paquets nous per segon que el que admet la tècnica Aloha pur

$$\lambda_{\max} = \frac{S_{\max}}{T_t} = \frac{1}{2e T_t}$$

Aloha amb ranures (S-Aloha)

L'Aloha amb ranures és una variació respecte l'Aloha pur. S'expressa com *S-Aloha* (*slotted Aloha*). El seu funcionament és igual al de l'Aloha pur però, a més:

Figura 71.
Funcionament de la
tècnica Aloha amb
ranures



- El temps es divideix en ranures temporals (*slots*), de manera que, quan les estacions volen accedir al recurs compartit, han d'esperar a l'inici de la ranura abans de poder fer la transmissió. La durada d'una ranura és igual al temps de transmissió d'un paquet.

Anàlisi Aloha amb ranures per a un nombre infinit d'estacions

Per fer l'anàlisi, s'utilitzen les mateixes hipòtesis que per a l'Aloha pur.

Suposem que una de les infinites estacions fa un intent d'accés i volem conèixer la probabilitat que el seu intent sigui fructuós. Sabem que aquesta probabilitat relaciona S i G :

$$S = G p(\text{intent fructuós})$$

Vegem ara quines condicions s'han de donar perquè aquest intent sigui fructuós. Qualsevol intent que es produueixi dins de l'interval de vulnerabilitat que es mostra a la figura 72 s'esperarà a l'inici de la ranura següent i es produirà una col·lisió amb el paquet objecte d'estudi. Qualsevol intent fora d'aquest interval no col·lidirà amb el paquet esmentat. L'avantatge de l'Aloha amb ranures respecte l'Aloha pur és que disminueix l'interval de vulnerabilitat a la meitat. El desavantatge és que les estacions han d'estar sincronitzades i conèixer els instants d'inici de les ranures, cosa que incrementa la complexitat del sistema.

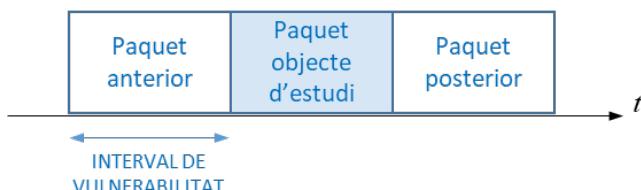


Figura 72.
Interval de vulnerabilitat per la tècnica Aloha amb ranures

Així doncs, la probabilitat que l'intent de transmissió sigui fructuós coincideix amb la probabilitat que la resta d'estacions no hi intentin accedir en l'interval de vulnerabilitat de durada T_t .

$$p(\text{intent fructuós}) = p(0 \text{ intents d'accés en } T_t) = \frac{(\Lambda T_t)^0}{0!} e^{-\Lambda T_t} = e^{-G}$$

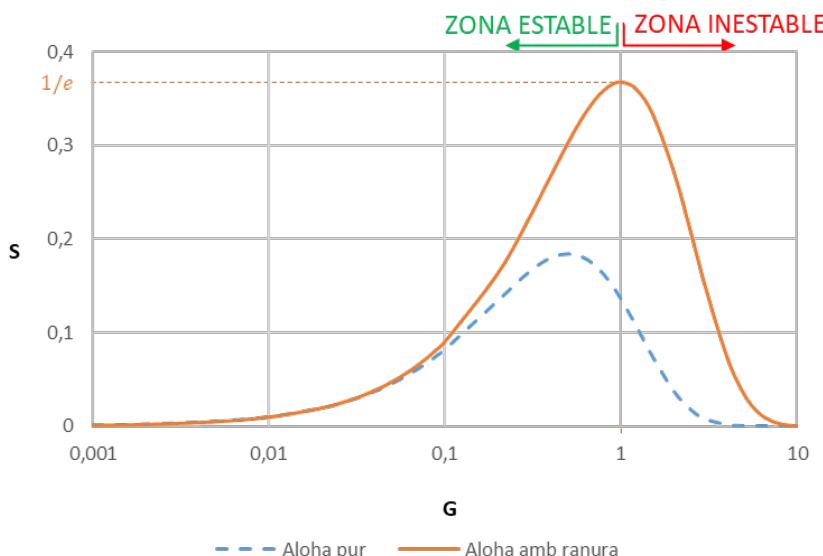
Finalment, ajuntant ambdues equacions, se n'obté:

$$S = G e^{-G}$$

La representació gràfica del cabal en funció de la càrrega ens dona una corba semblant al cas Aloha pur, però en aquest cas el cabal màxim val $1/e = 0,3678$, quan la $G = 1$.



Figura 73.
Relació entre la càrrega i el cabal per les tècniques Aloha pur i Aloha amb ranures



També hi ha una zona estable a l'esquerra del màxim i una zona inestable a la dreta del màxim. La taxa màxima que poden generar les estacions per no entrar a la zona inestable val:

$$\lambda_{\max} = \frac{S_{\max}}{T_t} = \frac{1}{e T_t}$$

Anàlisi Aloha amb ranures per a un nombre finit d'estacions

Fem ara l'anàlisi per a un nombre finit d'estacions. Al contrari que abans, comencem pel cas Aloha amb ranures.

Per fer-ne l'anàlisi, utilitzem les hipòtesis següents:

- Hi ha M estacions, totes iguals i independents.
- Tots els paquets tenen la mateixa longitud, és a dir, el seu temps de transmissió és constant i val T_t .
- Els ACK es transmeten per un canal a part de les dades, de manera que no es produeixen col·lisions entre ells i els paquets.
- El canal està lliure de soroll; si no es rep l'ACK és perquè hi ha una col·lisió.

A la figura 74, es mostren els intents d'accés per part de l'estació i en un temps d'observació T_{obs} .

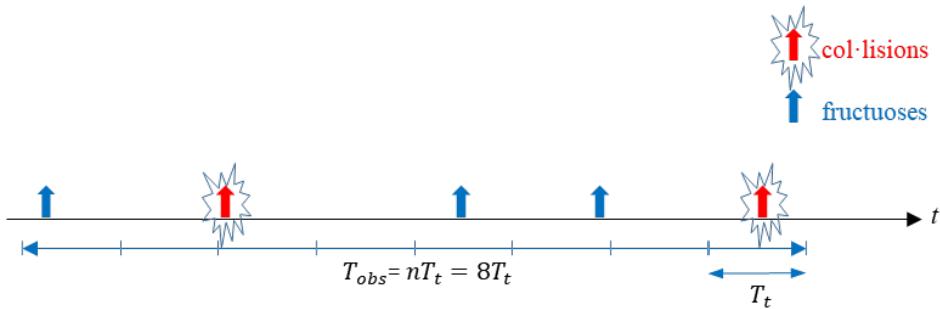


Figura 74.
Exemple de funcionament de la tècnica Aloha amb ranures, vista per a una estació

$$\begin{aligned} l_i &= 3 \text{ transmissions fructuoses per part de l'estació } i \text{ en } T_{obs} \\ m_i &= 5 \text{ intents d'accés per part de l'estació } i \text{ en } T_{obs} \\ n &= 8 \quad T_t \text{ en } T_{obs} \end{aligned}$$

Fem les definicions següents:

- S_i : nombre mitjà de transmissions fructuoses cada T_t per part de l'estació i . Mirant l'exemple de la figura 74, en una primera aproximació el podríem calcular com:

$$S_i \approx \frac{l_i}{n}$$

En el límit, quan el temps d'observació tendeix a infinit, aquest valor es pot interpretar com la **probabilitat que l'estació i transmeti fructuosament en T_t** , ja que són els casos favorables l_i , dividits pels casos possibles n . A més, si a l'expressió anterior multipliquem el numerador i el denominador per T_t , n'obtenim el quotient entre el temps durant el qual l'estació i està transmetent fructuosament en el T_{obs} i el T_{obs} . Per tant, S_i també indica la **fracció de temps durant la qual l'estació i està transmetent fructuosament**.

Si sumem el nombre mitjà de transmissions fructuoses cada T_t de totes les estacions, n'obtenim el cabal $S = \sum_{i=1}^M S_i$.

- G_i : nombre mitjà d'intents d'accés cada T_t per part de l'estació i . Observant l'exemple de la figura, en una primera aproximació el podríem calcular com:

$$G_i \approx \frac{m_i}{n}$$

En el límit, quan el temps d'observació tendeix a infinit, aquest valor es pot interpretar com la **probabilitat que l'estació i intenti accedir en T_t** , ja que són els casos favorables m_i , dividits pels casos possibles n . A més, si a l'expressió anterior multipliquem el numerador i el denominador per T_t , n'obtenim el quotient entre el temps durant el qual l'estació i vol ocupar el recurs compartit en el T_{obs} i el T_{obs} . Per tant, G_i també ens



indica la fracció de temps durant la qual l'estació i vol ocupar el recurs compartit.

Si sumem el nombre mitjà d'intents d'accés cada T_t de totes les estacions, n'obtenim la càrrega $G = \sum_{i=1}^M G_i$.

La probabilitat que l'estació i transmeti fructuosament en T_t és igual a la probabilitat que l'estació i ho intenti en T_t i la resta d'estacions no ho intentin en l'interval de vulnerabilitat. Ja hem vist que, per al cas Aloha amb ranura, l'interval de vulnerabilitat val T_t . Per tant, tenim:

- S_i : probabilitat que l'estació i transmeti fructuosament en T_t .
- G_i : probabilitat que l'estació i ho intenti en T_t .
- $1 - G_j$: probabilitat que l'estació j no ho intenti en T_t .

Com que les estacions són independents:

$$S_i = G_i \prod_{\substack{j=1 \\ j \neq i}}^M (1 - G_j)$$

Per al cas en què totes les estacions són iguals:

$$S = M S_i$$

$$G = M G_i$$

$$S_i = G_i (1 - G_i)^{M-1}$$

Tenint en compte aquestes tres equacions, finalment podem relacionar S i G :

$$S = G \left(1 - \frac{G}{M}\right)^{M-1}$$

Si calculem el límit quan M tendeix a infinit, n'obtenim:

$$S = G e^{-G}$$

Anàlisi Aloha pur amb un nombre finit d'estacions

Per fer aquesta anàlisi, utilitzem les hipòtesis següents:

- Hi ha M estacions, totes iguals i independents.
- Tots els paquets tenen la mateixa longitud, és a dir, el seu temps de transmissió és constant i val T_t .
- Si una estació transmet durant T_t segons, no ho fa en els T_t segons següents, ja que està esperant l'ACK.

- Els ACK es transmeten per un canal a part de les dades, de manera que no es produeixen col·lisions entre ells i els paquets.
- El canal està lliure de soroll; si no es rep l'ACK és perquè hi ha una col·lisió.

Com ja hem vist, per al cas Aloha pur l'interval de vulnerabilitat val $2T_t$. Dividim aquest interval en dos subintervals de durada T_t , que anomenem T_{t1} i T_{t2} . La probabilitat que l'estació j hi intenti accedir en l'interval de vulnerabilitat és la probabilitat que ho intenti en el subinterval T_{t1} o en el subinterval T_{t2} .

$$\begin{aligned} p(\text{intent en } T_{t1} \cup \text{intent en } T_{t2}) &= \\ &= p(\text{intent en } T_{t1}) + p(\text{intent en } T_{t2}) - p(\text{intent en } T_{t1} \cap \text{intent en } T_{t2}) \end{aligned}$$

Per hipòtesi, una estació no intenta accedir en dos T_t consecutius; per tant, la probabilitat que ho intenti en T_{t1} i també en T_{t2} val zero ($p(\text{intent en } T_{t1} \cap \text{intent en } T_{t2}) = 0$). Aleshores:

$$p(\text{intent en } T_{t1} \cup \text{intent en } T_{t2}) = p(\text{intent en } T_{t1}) + p(\text{intent en } T_{t2}) = 2G_j$$

Ara ja podem calcular la probabilitat que l'estació j **no** intenti accedir en l'interval de vulnerabilitat i valgui $1 - 2G_j$.

Per trobar la relació entre S i G , utilitzem l'equació següent, que s'obté del fet que la probabilitat que l'estació i transmeti fructuosament en T_t és igual a la probabilitat que l'estació i ho intenti en T_t i la resta d'estacions no ho intentin en l'interval de vulnerabilitat.

$$S_i = G_i \prod_{\substack{j=1 \\ j \neq i}}^M (1 - 2G_j)$$

A l'equació anterior, hem considerat que les estacions són independents. Per al cas en que totes siguin iguals, tenim:

$$S_i = G_i (1 - 2G_i)^{M-1}$$

$$S = G \left(1 - \frac{2G}{M}\right)^{M-1}$$

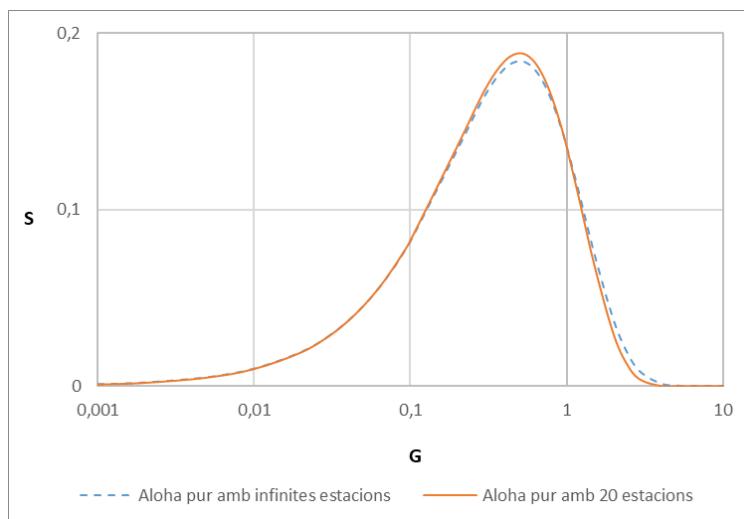
Si calculem el límit quan M tendeix a infinit, n'obtenim:

$$S = G e^{-2G}$$



Figura 75.
Comparació de l'anàlisi de la tècnica Aloha pur per a 20 estacions i per a un nombre infinit d'estacions

Les equacions corresponents a un nombre infinit d'estacions són més fàcils de manipular matemàticament; és per això que en molts problemes es treballa amb aquestes equacions, malgrat que el nombre d'estacions sempre és finit. A la figura 75, es comparen gràficament les equacions de la tècnica Aloha pur per a un nombre finit de 20 estacions i un nombre infinit d'estacions. Com es pot observar, ambdues gràfiques són molt semblants a la zona estable.



Temps de servei per la tècnica Aloha

En la teoria de cues, el temps de servei és el temps des que una unitat entra al servidor fins que hi pot entrar la següent. A les xarxes de transport, el servei consisteix només en la transmissió del paquet; per tant, el temps de servei i el temps de transmissió són el mateix en aquest cas. A les xarxes d'accés, el servei ha de considerar més aspectes, ja que no es pot iniciar el servei del paquet següent immediatament després de l'intent de transmissió del paquet anterior. Així doncs, en el càlcul del temps de servei, s'ha de prendre en consideració tot això:

- Temps de transmissió
- Temps d'espera de l'ACK
- *Timeouts*
- Temps de *back-off*
- Temps de les retransmissions
- Temps de propagació

En general, per la tècnica Aloha, el temps mitjà de servei es pot expressar de la manera següent:

$$T_s = T_{s,sc} + N_{r/p} T_{RT}$$

essent:

- $T_{s,sc}$: temps mitjà de servei sense col·lisió
- $N_{r/p}$: nombre mitjà de retransmissions per paquet
- T_{RT} : temps mitjà entre retransmissions

Aquesta expressió s'ha de particularitzar pel tipus de tècnica (Aloha pur o amb ranura) i per a diferents escenaris.

Independentment de l'escenari, el nombre mitjà de transmissions que es fa per a cada paquet es pot calcular com:

$$\lim_{T_{OBS} \rightarrow \infty} \frac{\text{nombre total d'intents de transmissió}}{\text{nombre de transmissions fructuoses}} = \lim_{T_{OBS} \rightarrow \infty} \frac{\Lambda T_{OBS}}{\lambda T_{OBS}} = \frac{\Lambda}{\lambda} = \frac{G}{S}$$

D'aquestes, una és fructuosa i la resta són retransmissions; per tant:

$$N_{r/p} = \frac{G}{S} - 1$$

Els valors de $T_{s,sc}$ i T_{RT} depenen de l'escenari on s'estigui utilitzant la tècnica Aloha. Analitzem el temps de servei a dos escenaris diferents: l'escenari satèl·lit i l'escenari terrestre.

Temps de servei per a un escenari satèl·lit

Les característiques de l'escenari satèl·lit són les següents:

- Les estacions terrenes transmeten paquets cap al satèl·lit pel canal de pujada i el satèl·lit els retransmet pel canal de baixada (diferent del de pujada) cap al *hub*, i totes les estacions terrenes.
- El *hub* permet l'accés cap a la xarxa de transport.
- Les distàncies són grans i el retard de propagació és significatiu.
- La retransmissió del paquet de baixada el podria fer servir d'ACK l'estació que ha originat el paquet (i aquest és el cas que considerarem).

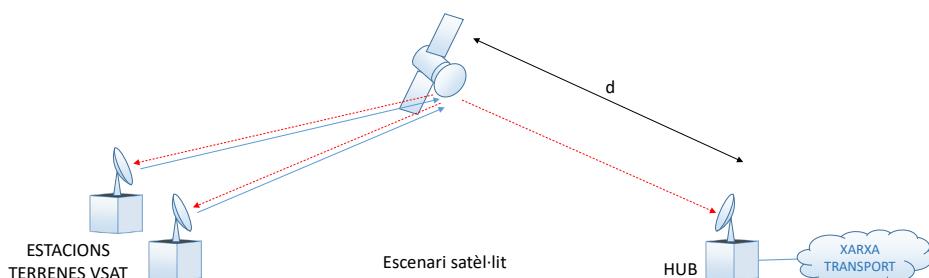
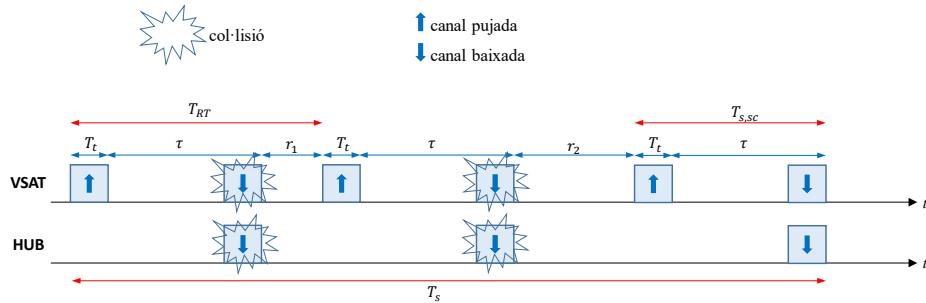


Figura 76.
Escenari satèl·lit



La figura següent mostra el temps de servei d'un paquet per al cas Aloha pur en l'escenari satèl·lit.

Figura 77.
Exemple de funcionament de la tècnica Aloha pur per a l'escenari satèl·lit



L'estació VSAT transmet el paquet i, transcorregut el temps de propagació τ , torna a rebre el paquet. La propagació inclou l'anada al satèl·lit més la tornada del satèl·lit. Si s'ha produït una col·lisió al satèl·lit, ho detecta i espera un temps aleatori de *back-off* r_i abans de tornar-ho a intentar.

En aquest cas, el temps mitjà entre retrasmissions val $T_{RT} = T_t + \tau + R$, essent $R = E(r_i)$. Quan no hi ha col·lisió, el temps mitjà de servei val $T_{s,sc} = T_t + \tau$. Així doncs, el temps de servei per al cas Aloha pur val:

$$T_s = T_t + \tau + (e^{2G} - 1)(T_t + \tau + R)$$

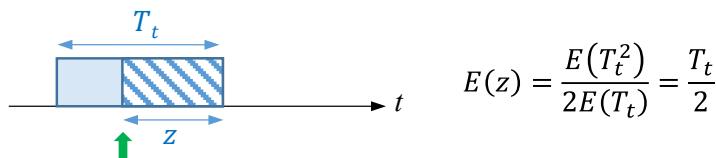
Podem expressar el temps de servei com el temps de transmissió, multiplicat per un factor més gran que 1.

$$T_s = T_t[1 + a + (e^{2G} - 1)(1 + a + r)]$$

essent $a = \tau/T_t$ el temps de propagació normalitzat i $r = R/T_t$ el temps mitjà de *back-off* normalitzat.

Per al cas de l'Aloha amb ranures, quan una estació vol transmetre un paquet, primer ha d'esperar l'inici d'una ranura. Aquest temps és el *temps de sincronisme de ranura* i val $T_t/2$, atès que T_t és determinista.

Figura 78.
Temps de sincronisme de ranura



A la figura, es mostra un temps de servei per al cas Aloha amb ranura.

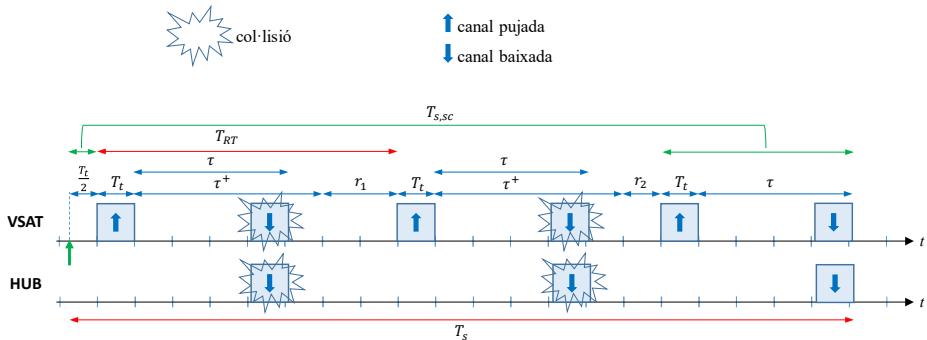


Figura 79.
Exemple de funcionament de la tècnica Aloha amb ranura per a l'escenari satèl·lit

Com que el retard de propagació no és un múltiple del temps de ranura, hem de calcular el valor τ^+ . El valor normalitzat de τ^+ ha de ser un nombre sencer. De fet, és el sencer immediatament superior al valor normalitzat de τ .

$$a^+ = \frac{\tau^+}{T_t} = [a] = \left\lceil \frac{\tau}{T_t} \right\rceil \Rightarrow \tau^+ = \left\lceil \frac{\tau}{T_t} \right\rceil T_t$$

Finalment, trobem el temps mitjà de servei:

$$T_s = T_{s,sc} + N_{r/p} T_{RT}$$

$$T_s = \frac{T_t}{2} + T_t + \tau + (e^G - 1)(T_t + \tau^+ + R)$$

La seva expressió normalitzada val:

$$T_s = T_t \left[\frac{1}{2} + 1 + a + (e^G - 1)(1 + a^+ + r) \right]$$

Temps de servei per escenari terrestre

Les característiques de l'escenari terrestre són:

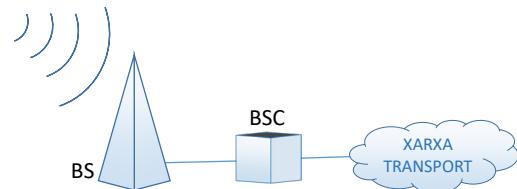
- Les estacions transmeten paquets a l'estació base (BS) pel canal d'anada i l'estació base transmet els ACK a les estacions per un canal de tornada diferent.
- L'estació base fa l'accés a la xarxa de transport a través del *base station controller* (BSC).
- Les distàncies són petites i el retard de propagació és negligible.



Figura 80.
Escenari terrestre

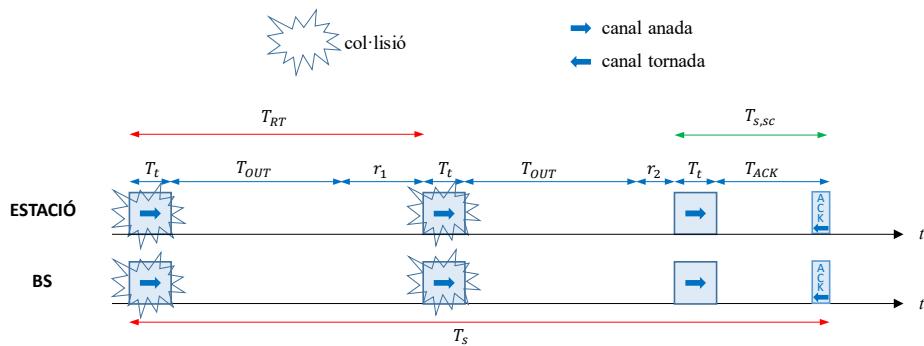


Escenari terrestre



A la figura següent, es mostra el temps de servei d'un paquet pel cas Aloha pur i l'escenari terrestre.

Figura 81.
Exemple de funcio-
nement de la tècnica
Aloha per a l'escenari
terrestre



L'estació transmet un paquet i, si després d'un temps de *timeout* (T_{OUT}) no rep l'ACK, es retira durant el temps de *back-off* (r_i) abans de tornar-ho a intentar. Tan bon punt rep l'ACK, pot atendre el paquet següent i, per tant, finalitza el servei. El temps que triga a rebre l'ACK quan no hi ha col·lisió és T_{ACK} . En aquest cas, el temps de servei val:

$$T_s = T_t + T_{ACK} + (e^{2G} - 1)(T_t + T_{OUT} + R)$$

Si normalitzem T_{ACK} i T_{OUT} , i els anomenem t_{ack} i t_{out} , respectivament:

$$T_s = T_t[1 + t_{ack} + (e^{2G} - 1)(1 + t_{out} + r)]$$

Per al cas d'Aloha amb ranures i escenari terrestre, tenim la mateixa situació però, a més, hem de tenir en compte el temps de sincronisme de ranura. Els temps de *back-off* i el T_{OUT} són múltiples del temps de ranura T_t .

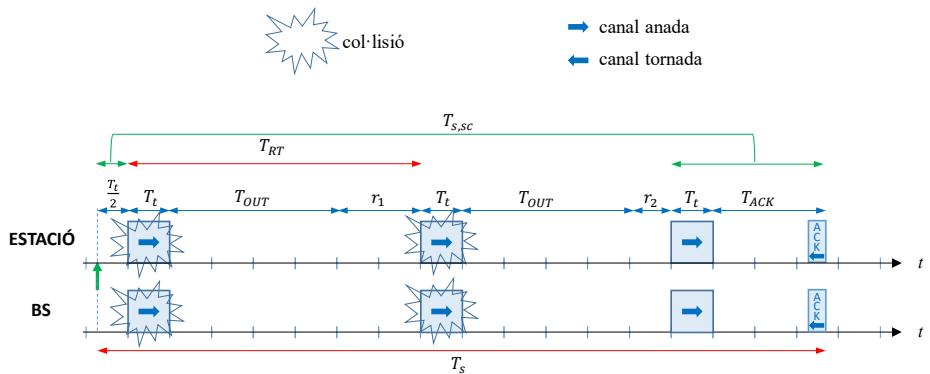


Figura 82.
Exemple de funcionament de la tècnica Aloha amb ranures per a l'escenari terrestre

En aquest cas:

$$T_s = \frac{T_t}{2} + T_t + T_{ACK} + (e^G - 1)(T_t + T_{OUT} + R)$$

$$T_s = T_t \left[\frac{1}{2} + 1 + t_{ack} + (e^G - 1)(1 + t_{out} + r) \right]$$

4.1.2 Carrier Sense Multiple Access (CSMA)

En aquesta tècnica, les estacions disposen del hardware necessari per tal d'escoltar el medi i detectar si està ocupat o no abans d'iniciar la transmissió d'un paquet. El fet d'escoltar les transmissions d'altres estacions redueix l'interval de vulnerabilitat. Per tal que la tècnica sigui efectiva, cal que el temps de propagació sigui el més petit possible ja que, quan una estació comença a transmetre, les altres no se n'adonen fins que el senyal s'ha propagat fins a la seva posició, i podrien començar a transmetre els seus propis paquets, amb l'aparició conseqüent de col·lisions. És especialment rellevant el fet que el temps de propagació sigui inferior al temps de transmissió d'un paquet. En cas contrari, les estacions finalitzen la transmissió d'un paquet abans que la resta pugui escoltar-lo i el fet de poder escoltar el medi no suposa cap millora.

Igual que en la tècnica Aloha, es pot treballar:

- amb ranures
- sense ranures

A diferència de l'Aloha, per al cas amb ranures la durada d'aquests és de l'ordre del temps de propagació τ .

En funció de l'estrategia d'escolta de la portadora, també podem classificar els algorismes CSMA de la manera següent:

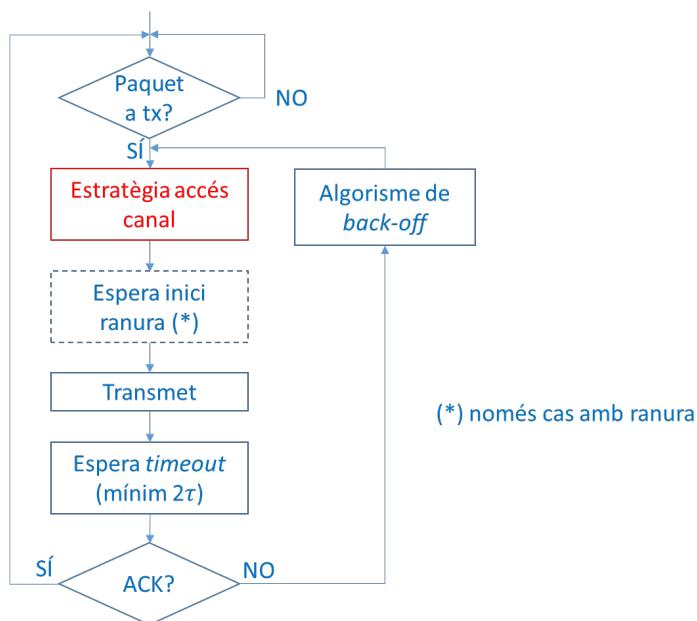


- No persistent: si una estació troba ocupat el canal no persisteix, sinó que es retira un temps de *back-off* abans de tornar a escoltar el canal.
- 1-persistent: persisteix amb probabilitat 1. Quan el canal queda lliure, hi accedeix.
- p-persistent: quan el canal està lliure, hi accedeix amb una probabilitat p .

L'estratègia no persistent minimitza les possibles col·lisions entre paquets, però pot causar un ús inefficient del recurs compartit. En canvi, l'estratègia 1-persistent pot augmentar el nombre de col·lisions ja que, si hi ha molt trànsit, és molt probable que més d'una estació hi vulgui accedir quan el canal quedí lliure. L'estratègia p-persistent és una solució de compromís entre les dues anterior.

A continuació, es mostra el diagrama de flux del funcionament de la tècnica CSMA.

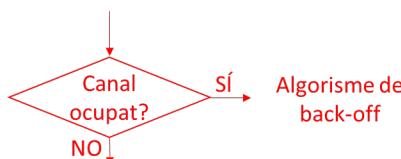
Figura 83.
Funcionament de la
tècnica CSMA



El bloc corresponent a l'estratègia d'accés al canal és diferent en cada cas:

- No persistent

Figura 84.
Estratègia d'accés
per al cas no persis-
tent



- 1-persistent

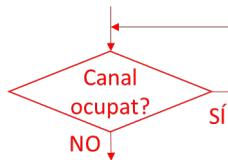


Figura 85.
Estratègia d'accés per al cas 1-persistent

- p-persistent

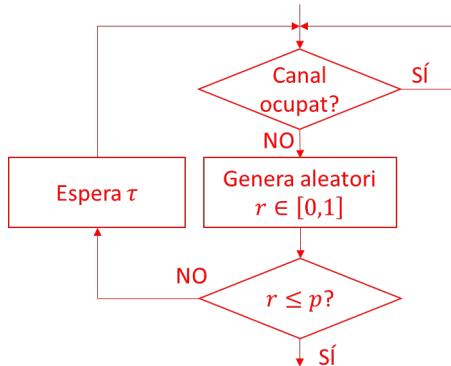


Figura 86.
Estratègia d'accés per al cas p-persistent

Anàlisi CSMA (no persistent, sense ranures)

Per fer l'anàlisi, utilitzem les hipòtesis següents:

- El nombre d'estacions és infinit.
- Els intents d'accés al canal són un procés de Poisson de taxa Λ . Per CSMA, cada escolta del canal és un intent d'accés. La probabilitat de tenir k intents d'accés en t segons val:

$$P_k(t) = \frac{(\Lambda t)^k}{k!} e^{-\Lambda t}$$

Λ inclou:

- **paquets nous**: escolta per transmetre un paquet generat per una estació.
- **col·lisions**: escolta per retransmetre un paquet que havia col·lidit.
- **bloquejos**: escolta per retransmetre un paquet que havia trobat el canal ocupat.

Com que el nombre d'estacions és infinit, la taxa d'intents d'accés Λ és constant.

- Tots els paquets tenen la mateixa longitud, és a dir, el seu temps de transmissió és constant i val T_t .



- El retard de propagació entre dues estacions qualssevol és constant i val τ . Aquesta hipòtesi simplifica l'anàlisi, però suposa que totes les estacions són equidistants i això normalment no és cert. En un cas real, podríem utilitzar el màxim retard de propagació, de manera que la xarxa sempre tindria un comportament millor que el resultat de l'anàlisi.
- Els ACK es transmeten per un canal a part de les dades, de manera que no es produeixen col·lisions entre ells i els paquets.
- L'escolta del canal és instantània.
- El canal està lliure de soroll; si no es rep l'ACK és perquè hi ha una col·lisió.

Igual que per a la tècnica Aloha, volem relacionar el cabal S i la càrrega G , la qual cosa equival a relacionar la taxa d'intents d'accés Λ i la taxa de paquets nous generats per les estacions λ , ja que:

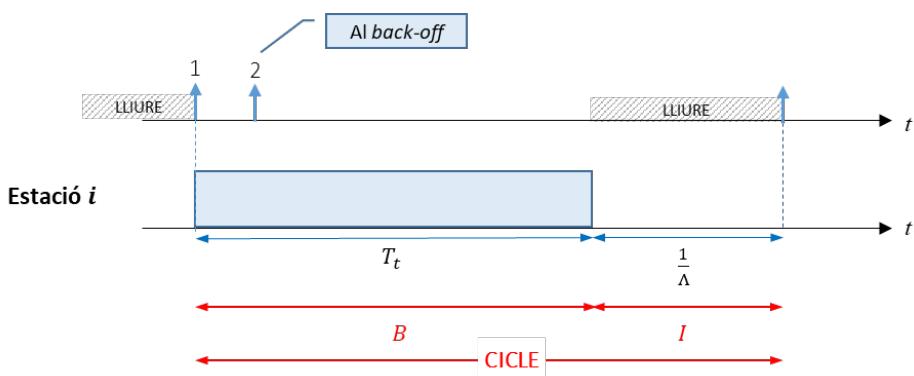
$$G = \Lambda T_t$$

$$S = \lambda T_t$$

Cal recordar que la darrera equació només és vàlida quan la tècnica es comporta d'una manera estable.

Inicialment, simplificarem l'anàlisi suposant que el retard de propagació és nul. En aquest cas, mai no hi haurà col·lisions ja que, quan una estació iniciï una transmissió, la resta ho escoltarà immediatament i no hi accedirà. El que sí que hi haurà són bloquejos (l'estació que intenta accedir al mitjà per transmetre-hi se'l troba ocupat i, per tant, se li bloqueja l'accés). A la figura següent, es mostra la transmissió d'un paquet feta per l'estació i . A l'eix temporal superior, es mostren els intents d'accés (escoltes) i els intervals en què les estacions "veuen" el canal lliure. L'estació que fa l'intent d'accés 2, en escoltar que el canal està ocupat, se'n retira (*back-off*), ja que estem analitzant el cas no persistent.

Figura 87.
Cicle CSMA quan el
retard de propagació
és nul



Quan *totes* les estacions escolten que no hi ha transmissions, diem que el canal està lliure. I quan *alguna* escolta que hi ha alguna transmissió, diem que està ocupat. Cada cop que el canal passa d'estar lliure a estar ocupat, finalitza un cicle i en comença un de nou. Dins de cada cicle, definim les variables aleatòries següents:

- **Període ocupat:** temps dins un cicle en què *alguna* estació escolta que el canal és ocupat. És una variable aleatòria que denotem amb la lletra b i el seu valor esperat val $B = E(b)$.
- **Període lliure:** temps dins un cicle en què *totes* les estacions escolten que el canal és lliure. És una variable aleatòria que denotem amb la lletra i i el seu valor esperat val $I = E(i)$.
- Temps dins un cicle en què es transmet un paquet fructuosament. És una variable aleatòria que denotem amb la lletra u i el seu valor esperat val $U = E(u)$.

Observem que, tal com els hem definit, el període lliure i el període ocupat són intervals de temps disjunts i la seva suma és la durada d'un cicle.

El valor mitjà del període lliure és el temps residual entre intents. Com que els intents d'accés són un procés de Poisson, el valor mitjà del període lliure val:

$$I = \frac{1}{\Lambda}$$

Recordem que el temps mitjà entre intents és igual a la inversa de la taxa Λ però, a més, com que és un procés sense memòria, el temps residual entre intents té el mateix valor, de mitjana.

Si el temps de propagació és nul, les variables B i U són deterministes i valen:

$$B = U = T_t$$

El cabal és la fracció de temps durant la qual s'està transmetent fructuosament; per tant:

$$S = \frac{U}{B + I} = \frac{T_t}{T_t + \frac{1}{\Lambda}} = \frac{\Lambda T_t}{\Lambda T_t + 1}$$

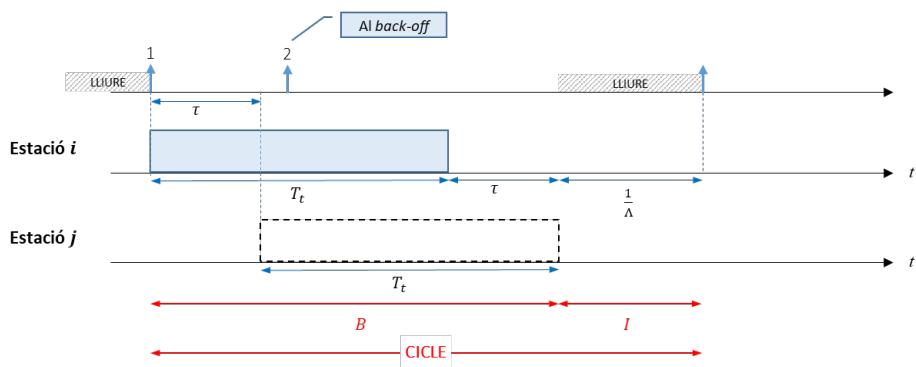
$$S = \frac{G}{G + 1}$$

En cas que el retard de propagació no sigui nul, es pot donar la situació que una estació transmeti quan el canal està ocupat ja que, a causa d'aquest retard, encara no pot escoltar el paquet que és a la xarxa. Així doncs, si hi ha un segon intent d'accés durant els τ segons següents al primer intent d'accés dins el cicle, es produeix una col·lisió. L'interval de vulnerabilitat per la tècnica CSMA val τ .



Estudiem ambdues situacions: quan hi ha col·lisió i quan no hi ha col·lisió. Comencem amb el cas en què no hi ha col·lisió:

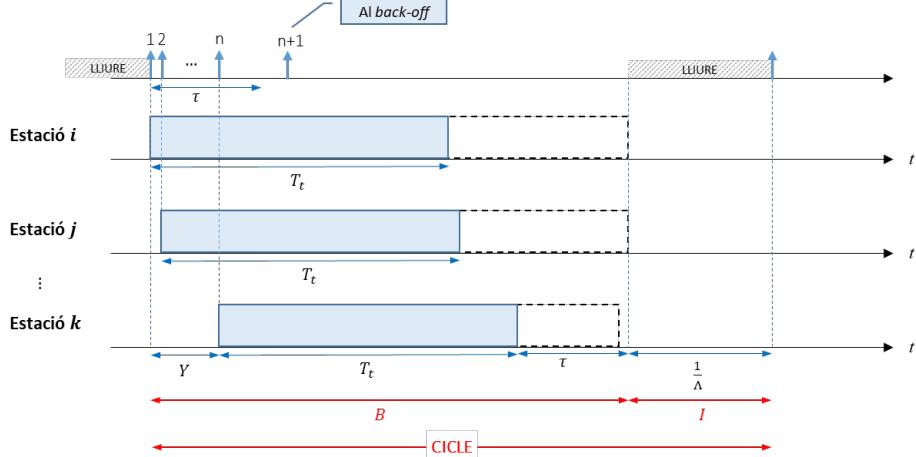
Figura 88.
Cicle CSMA quan no
hi ha col·lisió



A la figura 89, les transmissions pròpies de cada estació s'il·lustren de color sòlid i les transmissions que escolten altres estacions, amb una línia puntejada. L'estació i inicia la transmissió. El segon intent d'accés el fa l'estació j després que hagin passat més de τ segons des de l'inici de la transmissió de l'estació i (fora de l'interval de vulnerabilitat). Aleshores, l'estació j escolta que el canal està ocupat i es retira (*back-off*). D'aquesta manera, no hi ha col·lisió. Tot i que l'estació i finalitza la seva transmissió als primers T_t segons del cicle, la resta d'estacions escolten que el canal està ocupat durant els τ segons següents. Per tant, el període ocupat en aquest cas val $B = T_t + \tau$. Com ja havíem vist, el valor mitjà del període lliure val $I = 1/\Lambda$.

Vegem ara quant val el període ocupat quan es produeix una col·lisió.

Figura 89.
Cicle CSMA quan hi
ha col·lisió



Es produeix una col·lisió quan alguna estació intenta accedir dins l'interval de vulnerabilitat τ . La darrera que ho fa és la que determina la durada del període

ocupat. Definim la variable aleatòria y com l'instant en què es produeix el darrer intent d'accés dins l'interval de vulnerabilitat (prenem l'origen del temps a l'inici del cicle). Aleshores, el valor mitjà del període ocupat per al cas amb col·lisió val:

$$B = Y + T_t + \tau$$

Podem interpretar que el valor de B per al cas sense col·lisió és un cas particular d'aquesta expressió en la qual $y = 0$. És equivalent a pensar que, per al cas sense col·lisió, la darrera estació que fa un intent dins l'interval de vulnerabilitat sempre és la primera que ho intenta.

Com ja sabem, el valor mitjà del període lliure val $I = 1/\Lambda$. Per acabar de calcular quant val el cabal, ens manca calcular U i B i, per tant, Y .

La variable aleatòria u és el temps durant el qual es transmet fructuosament dins d'un cicle i pren els valors següents:

- $u = 0$, en cas de col·lisió
- $u = T_t$, en cas contrari

Per tant:

$$U = 0 p(\text{col·lisió}) + T_t p(\text{no col·lisió}) = T_t p(\text{no col·lisió})$$

Quan una estació inicia un cicle, la probabilitat que no hi hagi col·lisió és la probabilitat que la resta d'estacions no intentin accedir al canal durant l'interval de vulnerabilitat τ .

$$p(\text{no col·lisió}) = \frac{(\Lambda\tau)^0}{0!} e^{-\Lambda\tau} = e^{-\Lambda\tau}$$

Per tant:

$$U = T_t e^{-\Lambda\tau}$$

Per calcular el valor mitjà del període ocupat, cal calcular Y . Per fer-ho, primer calculem la funció de distribució de probabilitat (**CDF**) d'aquesta variable.

La probabilitat que el darrer intent que es fa dins l'interval de vulnerabilitat caigui dins aquest subinterval és la probabilitat de no tenir nous intents d'accés en l'interval $[x, \tau]$

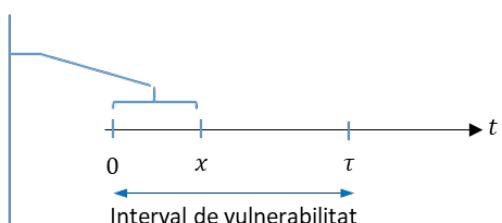


Figura 90.
Càlcul de la CDF de la variable aleatòria y

$$F_y(x) = p(y \leq x) = p(0 \text{ intents en } [x, \tau])$$



$$F_y(x) = p(0 \text{ intents en } (\tau - x) \text{ segons}) = \frac{(\Lambda(\tau - x))^0}{0!} e^{-\Lambda(\tau - x)} = e^{-\Lambda(\tau - x)}$$

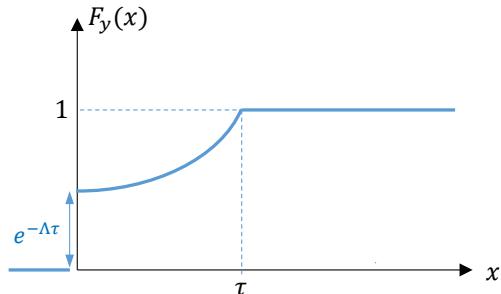
Aquest resultat és vàlid per a valors de x en l'interval $(0, \tau)$. Podem trobar els valors de $F_y(x)$ fora d'aquest interval, tenint en compte la pròpia definició de y i el fet que, com a mínim, sempre hi ha un intent d'accés (el que es produeix a l'instant 0). Aleshores:

- $p(y < 0) = 0$, el darrer intent dins $(0, \tau)$ no es pot produir en instants negatius;
- $p(y \leq \tau) = 1$, el darrer intent dins $(0, \tau)$ segur que ja s'ha produït a l'instant τ .

Si ho posem tot junt:

Figura 91.
CDF de la variable aleatòria y

$$F_y(x) = \begin{cases} 0; & \text{si } x < 0 \\ e^{-\Lambda(\tau-x)}; & \text{si } 0 < x \leq \tau \\ 1; & \text{si } x > \tau \end{cases}$$



S'observa que a l'origen hi ha una discontinuïtat; això significa:

$$p(y = 0) = F_y(0^+) - F_y(0^-) = e^{-\Lambda\tau}$$

Altrament dit, la probabilitat que $y = 0$ és la probabilitat de tenir 0 arribades en l'interval de vulnerabilitat o, el que és el mateix, és la probabilitat de no tenir col·lisió. Per tant, com ja hem dit, l'expressió del període ocupat per al cas amb col·lisió també la podem utilitzar per al cas sense col·lisió, ja que en aquest cas $y = 0$.

La funció densitat de probabilitat val:

$$f_y(x) = \frac{dF_y(x)}{dx} = e^{-\Lambda\tau}\delta(x) + \Lambda e^{-\Lambda(\tau-x)}; \quad 0 \leq x \leq \tau$$

Observeu que la discontinuïtat a l'origen fa que aparegui una delta de Dirac a la derivada. Fora de l'interval $0 \leq x \leq \tau$, la derivada és nul·la.

Per trobar el valor esperat, fem:

$$Y = E(y) = \int_0^\tau x f_y(x) dx = \int_0^\tau x [e^{-\Lambda\tau} \delta(x) + \Lambda e^{-\Lambda(\tau-x)}] dx$$

Resolent la integral, se n'obté:

$$Y = \tau - \frac{1}{\Lambda}(1 - e^{-\Lambda\tau})$$

Amb el resultat obtingut, trobem el període ocupat:

$$B = Y + T_t + \tau = T_t + 2\tau - \frac{1}{\Lambda}(1 - e^{-\Lambda\tau})$$

I ara ja se'n pot calcular el cabal:

$$S = \frac{U}{B + I} = \frac{T_t \cdot e^{-\Lambda\tau}}{T_t + 2\tau - \frac{1}{\Lambda}(1 - e^{-\Lambda\tau}) + \frac{1}{\Lambda}} = \frac{\Lambda T_t \cdot e^{-\Lambda\tau}}{\Lambda T_t + 2\Lambda\tau + e^{-\Lambda\tau}}$$

Com que $\Lambda T_t = G$ i $\Lambda\tau = aG$, finalment se n'obté:

$$S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}}$$

Quan el retard de propagació tendeix a 0, aquesta expressió tendeix a l'obtinguda pel cabal sense retard de propagació.

Temps de servei per CSMA (no persistent, sense ranures)

Igual com hem fet pel mètode Aloha, calculem el temps de servei per la tècnica d'accés CSMA no persistent sense ranures.

Primer fem el càlcul per a un retard de propagació nul. En aquest cas, no hi ha col·lisions, però sí bloquejos. Es pot calcular el temps de servei com:

$$T_s = T_{s,sb} + N_{b/p} T_{RT}$$

essent:

- $T_{s,sb}$: temps mitjà de servei sense bloqueig
- $N_{b/p}$: nombre mitjà de bloquejos per paquet
- T_{RT} : temps mitjà entre bloquejos

El nombre mitjà d'intents d'accés per paquet val $N_{ac/p} = G/S$, dels quals tots menys un (la transmissió finalment correcta) corresponen a bloquejos; per tant:

$$N_{b/p} = N_{ac/p} - 1 = \frac{G}{S} - 1$$



A més, com que:

$$S = \frac{G}{G+1} \Rightarrow N_{b/p} = G$$

En ser CSMA no persistent, quan es produeix un bloqueig l'estació es retira durant un temps de *back-off*, així doncs, el temps entre bloquejos val $T_{RT} = R$.

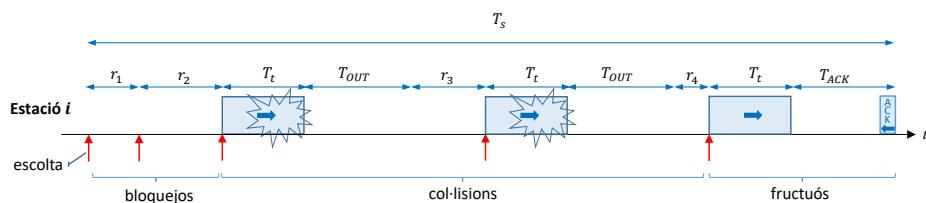
Si no hi ha bloqueig, el temps de servei és igual al temps de transmissió més el temps que triguem a rebre l'ACK: $T_{s,sb} = T_t + T_{ACK}$. Tot i que no hi ha col·lisions, per ser coherent amb el funcionament de la tècnica CSMA cal esperar l'ACK.

Finalment:

$$T_s = T_t + T_{ACK} + G R$$

En cas que el temps de propagació no sigui nul, hem de distingir entre bloquejos i col·lisions. A la figura 92, es mostra el temps de servei d'un paquet.

Figura 92.
Exemple d'accés
amb CSMA i temps
de servei



Podem expressar el temps de servei com:

$$T_s = N_{b/p} R + N_{c/p} (T_t + T_{OUT} + R) + T_t + T_{ACK}$$

essent:

- $N_{b/p}$: nombre mitjà de bloquejos per paquet
- $N_{c/p}$: nombre mitjà de col·lisions per paquet

Evidentment, el nombre mitjà d'intents d'accés per paquet val:

$$N_{ac/p} = N_{b/p} + N_{c/p} + 1$$

Per calcular $N_{b/p}$ i $N_{c/p}$ observem a la figura 93 següent les taxes de bloquejos i les taxes de col·lisions en el CSMA:

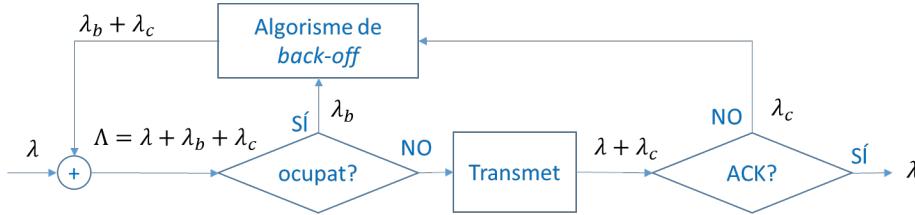


Figura 93.
Relació entre
 λ , λ_b , λ_c i Λ

essent:

- λ : paquets nous/s. A la zona estable, coincideix amb les transmissions fructuoses/s.
- λ_b : bloquejos/s
- λ_c : col·lisions/s
- Λ : intents/s

Definim la probabilitat de bloqueig:

$$P_b = \frac{\lambda_b}{\Lambda}$$

Calculem $N_{b/p}$ en funció de P_b :

$$N_{b/p} = \frac{\lambda_b}{\lambda} = \frac{\Lambda}{\lambda} P_b$$

$$N_{b/p} = \frac{G}{S} P_b$$

L'expressió anterior indica que el nombre mitjà de bloquejos per paquet és igual al nombre mitjà d'intents d'accés per paquet, multiplicat per la probabilitat de bloqueig.

Ara calculem $N_{c/p}$:

$$\lambda_c = \Lambda - \lambda_b - \lambda = (1 - P_b)\Lambda - \lambda$$

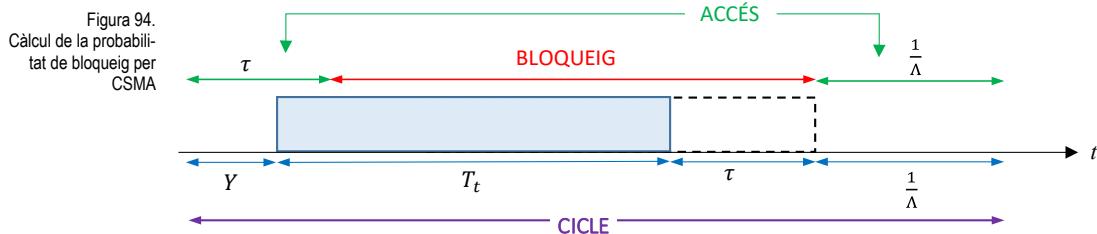
$$N_{c/p} = \frac{\lambda_c}{\lambda} = \frac{(1 - P_b)\Lambda}{\lambda} - 1$$

$$N_{c/p} = \frac{G}{S} (1 - P_b) - 1$$

L'expressió anterior indica que el nombre mitjà de col·lisions per paquet és igual al nombre mitjà d'intents d'accés per paquet que no es bloquegen, menys un. Aquest un es correspon amb el paquet que es transmet fructuosament.



Ara hem de trobar una altra manera de calcular la probabilitat de bloqueig. Per fer-ho, ens fixem en el cicle de la figura següent. En color sòlid, s'il·lustra la transmissió de la darrera estació que hi accedeix dins l'interval de vulnerabilitat τ i, amb una línia puntejada, el temps durant el qual la resta d'estacions encara escolten que el canal està ocupat.



Tan bon punt una estació ha iniciat el cicle, la resta d'estacions no es bloquejaran si fan un intent en els primers τ segons (interval de vulnerabilitat) o en els darrers $1/\Lambda$ segons (periode lliure). A la figura 94, aquests dos intervals es representen en color verd. En canvi, experimentaran un bloqueig si intenten accedir a la resta del cicle (a la figura, en vermell). És a dir, en els intervals verds, escoltaran que el canal està lliure i, en els vermells, que està ocupat. Aleshores, tenim:

$$(1 - P_b) = \frac{\tau + \frac{1}{\Lambda}}{CIRCLE}, \quad P_b = \frac{Y + T_t}{CIRCLE}$$

Recordeu que:

$$S = \frac{U}{CIRCLE} = \frac{T_t e^{-aG}}{CIRCLE} \Rightarrow \frac{1}{CIRCLE} = \frac{S}{T_t} e^{aG}$$

Per tant:

$$(1 - P_b) = \left(a + \frac{1}{G} \right) S e^{aG}$$

$$P_b = \left(\frac{Y}{T_t} + 1 \right) S e^{aG} = \left(a - \frac{1}{G} (1 - e^{-aG}) + 1 \right) S e^{aG}$$

Substituint en $N_{b/p}$ i $N_{c/p}$, se n'obté:

$$N_{b/p} = \frac{G}{S} P_b = \left(a - \frac{1}{G} (1 - e^{-aG}) + 1 \right) G e^{aG} = (a + 1) G e^{aG} - e^{aG} + 1$$

$$N_{c/p} = \frac{G}{S} (1 - P_b) - 1 = \left(a + \frac{1}{G} \right) G e^{aG} - 1 = (aG + 1) e^{aG} - 1$$

Utilitzem aquestes expressions per calcular T_s :

$$T_s = [(a + 1) G e^{aG} - e^{aG} + 1] R + [(aG + 1) e^{aG} - 1] (T_t + T_{OUT} + R) + T_t + T_{ACK}$$

Simplificant, n'obtenim finalment:

$$T_s = [(2a + 1)Ge^{aG}]R + [(aG + 1)e^{aG} - 1](T_t + T_{OUT}) + T_t + T_{ACK}$$

Si considerem que el retard de propagació és nul ($a = 0$), se n'obté l'expressió que havíem calculat anteriorment:

$$T_s = T_t + t_{ACK} + G R$$

És interessant observar que, quan un intent d'accés no és bloquejat, és a dir, accedeix al canal i transmet, pot passar que es produueixi una col·lisió o que la transmissió sigui correcta, amb les probabilitats següents:

$$P_{acc} = (1 - P_b) = \frac{\tau + \frac{1}{\Lambda}}{CIRCLE} = P_{col} + P_S$$

La transmissió provocarà una col·lisió si es realitza dins els primers τ segons del cicle, o bé en l'interval lliure i, a més, hi ha un nou intent d'accés per part d'una altra estació en els següents τ segons:

$$P_{col} = P(\text{intent en els } \tau \text{ primers segons}) + P(\text{intent en període lliure})(1 - P(0 \text{ intents en } \tau))$$

$$P_{col} = \frac{\tau + \frac{1}{\Lambda}(1 - e^{-aG})}{CIRCLE} = \frac{\tau + \frac{T_t}{G}(1 - e^{-aG})}{2\tau + T_t + \frac{T_t}{G}e^{-aG}} = \frac{1 + aG - e^{-aG}}{G(1 + 2a) + e^{-aG}}$$

D'altra banda, la transmissió tindrà èxit si es realitza dins l'interval lliure i no hi ha cap altre intent d'accés en els següents τ segons:

$$P_S = \frac{\frac{1}{\Lambda}e^{-aG}}{CIRCLE} = \frac{e^{-aG}}{G(1 + 2a) + e^{-aG}}$$

Evidentment, s'ha de complir:

$$P_b + P_{acc} = P_b + P_{col} + P_S = 1$$

Observeu que l'expressió obtinguda per a la probabilitat de transmissió amb èxit és coherent amb l'obtinguda a l'apartat anterior per al cabal:

$$S = G P_S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}}$$

4.1.3 Carrier Sense Multiple Access / Collision Detection (CSMA/CD)

Aquesta tècnica funciona igual que la tècnica CSMA però, a més, les estacions continuen escoltant el canal després d'haver iniciat la transmissió. D'aquesta



manera, si el nivell de potència del senyal al canal és molt alt, detecten que s'ha produït una col·lisió. En aquest cas, avorten la transmissió del paquet per minimitzar el temps d'ocupació del canal amb transmissions infructuoses. Aquest darrer fet comportarà una millora en termes de cabal de CSMA/CD respecte a CSMA.

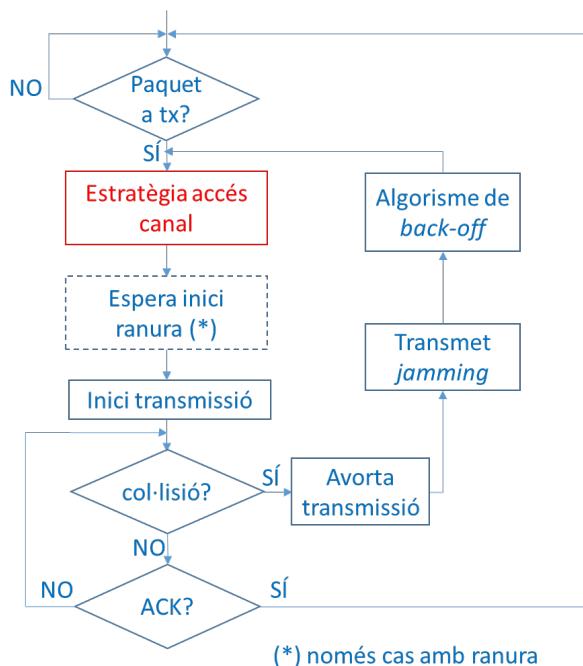
Igual que amb el mecanisme CSMA, l'estrategia d'accés al canal pot ser:

- no persistent
- 1-persistent
- p-persistent

A més, es pot treballar amb ranures o sense, la qual cosa dona lloc a sis variacions de CSMA/CD.

A continuació, es mostra el diagrama de flux del funcionament de la tècnica CSMA.

Figura 95.
Funcionament de la
tècnica CSMA/CD



En aquest cas, no cal esperar l'ACK de l'estació receptora ja que, si no es detecta col·lisió, se suposa que el paquet s'ha transmès correctament. En algunes implementacions de la tècnica CSMA/CD, quan una estació detecta una col·lisió, injecta un senyal sorollós al canal per garantir que la resta d'estacions també detecten la col·lisió. Aquest senyal s'anomena *jamming*.

Anàlisi CSMA/CD (no persistent, sense ranures)

Analitzem aquesta tècnica només per al cas no persistent i sense ranures temporals. Per fer-ho, utilitzem les hipòtesis següents:

- El nombre d'estacions és infinit.
- Els intents d'accés al canal són un procés de Poisson de taxa Λ . *Per CSMA/CD, cada escolta del canal és un intent d'accés.* La probabilitat de tenir k intents d'accés en t segons val:

$$P_k(t) = \frac{(\Lambda t)^k}{k!} e^{-\Lambda t}$$

Λ inclou:

- **paquets nous:** escolta per transmetre un paquet generat per una estació.
- **col·lisions:** escolta per retransmetre un paquet que havia col·lidit.
- **bloquejos:** escolta per retransmetre un paquet que havia trobat el canal ocupat.

Com que el nombre d'estacions és infinit, la taxa d'intents d'accés Λ és constant.

- Tots els paquets tenen la mateixa longitud, és a dir, el seu temps de transmissió és constant i val T_t .
- El retard de propagació entre dues estacions qualssevol és constant i val τ .
- L'escola del canal és instantània.
- La detecció d'una col·lisió és instantània.
- El temps de transmissió de *jamming* val J i és el mateix a totes les estacions. El cas $J = 0$ es correspon amb el fet que no es transmet *jamming*.

Igual que en els casos anteriors, per fer l'anàlisi relacionem el cabal S i la càrrega G :

$$G = \Lambda T_t$$

$$S = \lambda T_t$$

Recordem que la darrera equació només és vàlida quan la tècnica es comporta de manera estable.

Si el temps de propagació fos nul, mai no es produirien col·lisions, de manera que el fet de detectar la col·lisió no serviria de res. Per tant, en aquest cas, les



tècniques CSMA i CSMA/CD coincideixen. Així doncs, l'anàlisi de la tècnica CSMA/CD per al cas en què considerem que el temps de propagació és nul és la mateixa que la que ja s'ha fet per CSMA.

L'anàlisi que fem a continuació és per a un temps de propagació no nul. Igual que en CSMA, calculem el valor mitjà de les variables aleatòries següents, definides dins un cicle:

- B , valor mitjà del període ocupat dins un cicle.
- I , valor mitjà del període lliure dins un cicle.
- U , valor mitjà del temps de transmissió fructuosa dins un cicle.

Aleshores, podem calcular el cabal com la fracció de temps durant la qual es transmet fructuosament dins un cicle:

$$S = \frac{U}{CICLE} = \frac{U}{B + I}$$

Els valors de I i de U es calculen igual que en el cas CSMA. El valor mitjà del període lliure és el temps mitjà residual entre intents d'accés:

$$I = \frac{1}{\Lambda}$$

El temps mitjà de transmissió fructuosa és:

$$U = T_t p(\text{cicle amb èxit}) = T_t e^{-\Lambda \tau}$$

Recordem que la probabilitat de no col·lisió en un cicle és la probabilitat de no tenir cap intent d'accés en l'interval de vulnerabilitat τ .

Per calcular el valor mitjà del període ocupat, cal distingir entre el cas amb col·lisió i el cas sense col·lisió. Si en un cicle no hi ha col·lisió, alguna de les estacions veurà que el canal està ocupat durant un període igual a $T_t + \tau$, com passava en el cas CSMA. D'altra banda, definim la variable aleatòria b_c com la durada del període ocupat en cas de col·lisió i n'expresssem el valor esperat amb majúscules $B_c = E(b_c)$. Així doncs, podem expressar el valor mitjà del període ocupat com:

$$B = (T_t + \tau) p(\text{cicle amb èxit}) + B_c p(\text{cicle sense èxit})$$

$$B = (T_t + \tau)e^{-\Lambda \tau} + B_c (1 - e^{-\Lambda \tau})$$

Per calcular el valor de B_c ens fixem en la figura següent. A, B i C representen tres estacions de la xarxa que volen transmetre. L'eix horitzontal indica la separació entre estacions i l'eix vertical, el temps. Inicialment el canal està lliure i l'estació A és la primera que hi accedeix. A causa de la propagació, inicialment les estacions B i C no "l'escolten". A l'inici de la transmissió de l'estació A, comença l'interval de vulnerabilitat de durada τ , de manera que qualsevol intent

d'accés dins aquest interval provocarà una col·lisió. L'estació B és la primera que hi accedeix dins l'interval de vulnerabilitat i l'estació C, la darrera, de manera que la resta d'intents d'accés estarien inclosos entre aquests dos. L'instant d'accés de l'estació B és una variable aleatòria que anomenem z i l'instant d'accés de l'estació C l'anomenem y . Recordem que en CSMA anomenàvem y l'instant en què es produïa el darrer intent d'accés dins l'interval de vulnerabilitat.

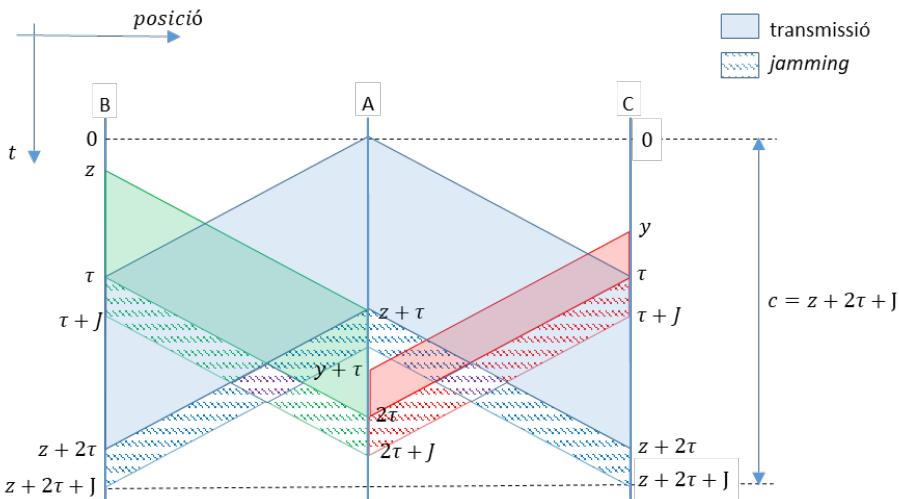


Figura 96.
Càlcul del temps de cicle per la tècnica CSMA/CD

Com es veu a la figura 96, la darrera estació que detecta la col·lisió és l'estació A. Per tant, en ser ella la darrera a avortar la transmissió, el període ocupat en cas de col·lisió depèn de l'instant en què l'estació A detecta la col·lisió. Aquesta detecció de la col·lisió és deguda al primer accés dins l'interval de vulnerabilitat, és a dir, al fet que l'estació A escolta l'estació B. Així doncs, com mostra la figura, la durada del període ocupat en cas de col·lisió depèn de l'instant en què l'estació B hi accedeix, és a dir, depèn de z :

$$b_c = 2\tau + J + z$$

Observeu la diferència amb el cas CSMA, en el qual el període ocupat depenia de l'instant en què es produïa el darrer accés dins l'interval de vulnerabilitat y .

Així doncs, per calcular C ens cal calcular $Z = E(z)$. Per fer-ho, comencem calculant la **CDF** de la variable z :

$$F_z(x) = p(z \leq x) = 1 - p(z > x), x \in [0, \tau]$$

La variable z representa l'instant de temps en el qual es produeix el primer intent d'accés en l'interval de vulnerabilitat, atès que hi ha una col·lisió. Que es produeixi una col·lisió significa que almenys hi ha un intent d'accés en $[0, \tau]$; per tant:

$$p(z > x) = p(0 \text{ intents en } [0, x] \mid \text{almenys un intent en } [0, \tau]) =$$



$$\begin{aligned}
 &= \frac{p(0 \text{ intents en } [0, x] \cap \text{almenys un intent en } [0, \tau])}{p(\text{almenys un intent en } [0, \tau])} = \\
 &= \frac{p(0 \text{ intents en } [0, x] \cap \text{almenys un intent en } (x, \tau])}{p(\text{almenys un intent en } [0, \tau])}.
 \end{aligned}$$

Els intents d'accés són un procés de Poisson pel qual el que passi en un interval és independent del que passi en un altre si ambdós intervals són disjunts. Així doncs:

$$\begin{aligned}
 p(z > x) &= \frac{p(0 \text{ intents en } [0, x]) p(\text{almenys un intent en } (x, \tau])}{p(\text{almenys un intent en } [0, \tau])} = \\
 &= \frac{p(0 \text{ intents en } [0, x]) (1 - p(0 \text{ intents en } (x, \tau]))}{1 - p(0 \text{ intents en } [0, \tau])} = \\
 &= \frac{e^{-\Lambda x} (1 - e^{-\Lambda(\tau-x)})}{1 - e^{-\Lambda\tau}} = \frac{e^{-\Lambda x} - e^{-\Lambda\tau}}{1 - e^{-\Lambda\tau}}
 \end{aligned}$$

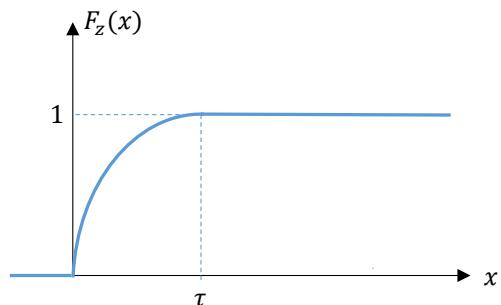
Ara calculem la **CDF** $F_z(x) = 1 - p(z > x)$:

$$F_z(x) = \frac{1 - e^{-\Lambda x}}{1 - e^{-\Lambda\tau}}, x \in [0, \tau]$$

Per la pròpia definició de z , $F_z(x)$ val 0 si $x < 0$ i val 1 si $x > \tau$. A més, es pot comprovar que a l'expressió anterior tenim que $F_z(0) = 0$ i $F_z(\tau) = 1$. Per tant, $F_z(x)$ és contínua i la seva derivada és nula fora de l'intervall $[0, \tau]$.

Figura 97.
CDF de la variable z

$$F_z(x) = \begin{cases} 0; & \text{si } x < 0 \\ \frac{1 - e^{-\Lambda x}}{1 - e^{-\Lambda\tau}}; & \text{si } 0 < x \leq \tau \\ 1; & \text{si } x > \tau \end{cases}$$



Calculem la **PDF** de z derivant-ne la **CDF**:

$$f_z(x) = \frac{\Lambda e^{-\Lambda x}}{1 - e^{-\Lambda\tau}}, x \in [0, \tau]$$

Ara podem calcular el valor esperat de z :

$$E(z) = \frac{\Lambda}{1 - e^{-\Lambda\tau}} \int_0^\tau x e^{-\Lambda x} dx = \frac{1}{\Lambda} - \tau \frac{e^{-\Lambda\tau}}{1 - e^{-\Lambda\tau}}$$

A continuació, substituïm aquest valor per trobar quant val B_c :

$$B_c = 2\tau + J + \frac{1}{\Lambda} - \tau \frac{e^{-\Lambda\tau}}{1 - e^{-\Lambda\tau}}$$

Amb el valor de B_c calculem el valor mitjà del període ocupat:

$$\begin{aligned} B &= (T_t + \tau)e^{-\Lambda\tau} + \left(2\tau + J + \frac{1}{\Lambda} - \tau \frac{e^{-\Lambda\tau}}{1 - e^{-\Lambda\tau}}\right)(1 - e^{-\Lambda\tau}) = \\ &= T_t e^{-\Lambda\tau} + \left(2\tau + J + \frac{1}{\Lambda}\right)(1 - e^{-\Lambda\tau}) \end{aligned}$$

Definim $\gamma = J/\tau$:

$$B = T_t e^{-\Lambda\tau} + \left(\tau(2 + \gamma) + \frac{1}{\Lambda}\right)(1 - e^{-\Lambda\tau})$$

Ho posem en funció de $G = \Lambda T_t$ i el retard de propagació normalitzat, $a = \tau/T_t$. Aleshores:

$$B = T_t [e^{-aG} + \left(a(2 + \gamma) + \frac{1}{G}\right)(1 - e^{-aG})]$$

Fem el mateix per a I i U :

$$I = \frac{1}{\Lambda} = \frac{T_t}{G}, \quad U = T_t e^{-\Lambda\tau} = T_t e^{-aG}$$

Ara ja podem expressar el cabal en funció de G :

$$S = \frac{U}{B + I} = \frac{e^{-aG}}{e^{-aG} + \left(a(2 + \gamma) + \frac{1}{G}\right)(1 - e^{-aG}) + \frac{1}{G}}$$

Arrangem l'expressió multiplicant el numerador i el denominador per G i, finalment, n'obtenim:

$$S = \frac{Ge^{-aG}}{(G - 1)e^{-aG} + aG(2 + \gamma)(1 - e^{-aG}) + 2}$$

Per al cas que negligim el retard de propagació ($a = 0$), n'obtenim el mateix resultat que per la tècnica CSMA, amb $a = 0$:

$$S = \frac{G}{G + 1}.$$



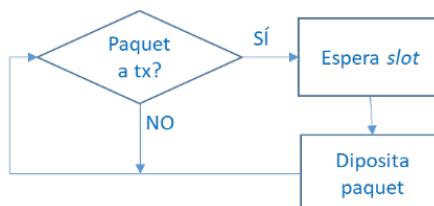
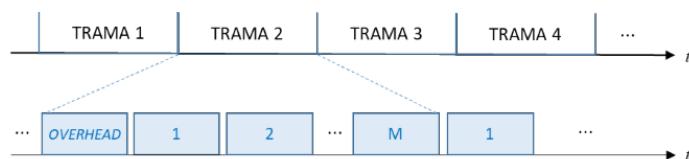
4.2 Tècniques deterministes

D'entre les tècniques deterministes, s'analitzarà la tècnica d'accés per multiplexació en el temps (TDMA) i per multiplexació en freqüència (FDMA). També s'analitzarà la tècnica d'accés per sondeig o *polling*.

4.2.1 Time Division Multiple Access (TDMA)

Aquesta tècnica consisteix a dividir el temps en trames de durada fixa. Les trames se subdividen en ranures temporals (*time slots*) i s'assigna una ranura a cada estació. Dins la trama, també hi poden haver *slots* per senyalització o sincronisme, que suposen un d'*overhead*. Les estacions hi accedeixen per ordre, transmetent en el seu *time slot*. Quan hi han accedit totes, es torna a començar per la primera. Entre els *time slots*, hi pot haver un petit temps de guarda per evitar solapaments entre les transmissions de paquets de les estacions.

Figura 98.
Funcionament de la
tècnica TDMA



Anàlisi de la tècnica TDMA

En aquesta anàlisi, calclem el temps mitjà de transferència d'un paquet. Per fer-ho, utilitzem les hipòtesis següents:

- La xarxa està formada per M estacions iguals.
- El canal té una capacitat de C bps.
- Els paquets són de longitud fixa igual a L bits.
- Cada estació i genera λ_i paquets/s, amb un procés de Poisson (totes iguals).
- No hi ha l'*overhead* a les trames.
- El temps de guarda és nul.
- Cada paquet espera a la cua de la seva estació fins que pot ser transmès.

A la figura següent, es mostra com és la transmissió d'una trama amb aquestes hipòtesis. La variable T_t indica el temps de transmissió d'un paquet i la variable T_{TR} , el temps de transmissió d'una trama.

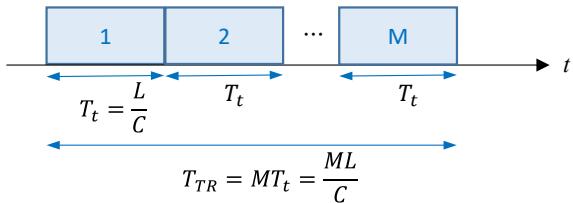


Figura 99.
Transmissió d'una
trama TDMA

El temps de transferència té tres components:

- T_{w0} : temps mitjà de sincronització de ranura. És el temps que transcorre des que es genera un paquet a l'estació fins que a aquesta li arriba el torn de transmetre.
- $N_w T_{TR}$: temps mitjà de transmissió dels paquets que hi ha a la cua per davant del paquet que es vol transmetre. N_w és el nombre mitjà de paquets que hi ha a la cua (hauríem d'utilitzar el nombre mitjà de paquets que es troba a la cua quan es genera però, per la propietat PASTA, coincideix amb N_w). Observeu que només es transmet un paquet cada temps de trama.
- T_t : temps de transmissió del paquet.

Les dues primeres corresponen al temps d'espera i la darrera, al temps de servei:

$$T = T_w + T_t = T_{w0} + N_w \cdot T_{TR} + T_t$$

Comencem calculant el temps mitjà d'espera T_w . Aplicant la llei de Little, tenim que:

$$N_w = \lambda_i T_w$$

A més:

$$T_w = T_{w0} + \lambda_i T_w T_{TR}$$

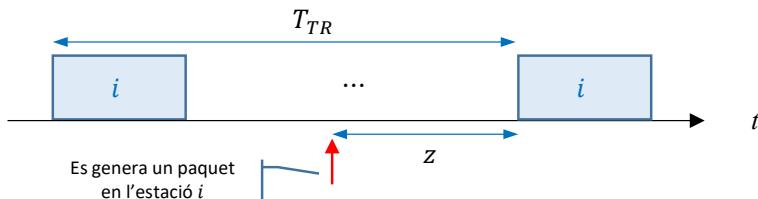
$$T_w = \frac{T_{w0}}{1 - \lambda_i T_{TR}} = \frac{T_{w0}}{1 - \lambda_i M T_t} = \frac{T_{w0}}{1 - \rho}$$

En aquesta expressió, ρ és el trànsit generat per totes les estacions (hores de transmissió cada hora). Observeu que, quan ρ tendeix a 1, l'espera és infinita. El fet que ρ valgui 1 significa que s'aprofiten tots els *slots* i, per tant, sempre hi ha d'haver un paquet disponible per omplir un *slot*. L'única manera que això passi és que totes les estacions tinguin la cua plena fins a l'infinít.



Calculem ara el temps mitjà de sincronisme de ranura T_{w0} . A la figura, z representa el temps des que es genera un paquet a l'estació i fins que aquesta diposita un paquet al seu slot.

Figura 100.
Temps de sincronisme
de ranura



Aleshores:

$$T_{w0} = E(z) = \frac{T_{TR}^2}{2T_{TR}} = \frac{T_{TR}}{2} = \frac{ML}{2C}$$

Finalment, calculem el temps d'espera i el de transferència:

$$T_w = \frac{M}{2(1-\rho)} \frac{L}{C}$$

$$T = \frac{M}{2(1-\rho)} \frac{L}{C} + \frac{L}{C} \quad \text{amb } \rho = \frac{M\lambda_i L}{C}$$

4.2.2 Frequency Division Multiple Access (FDMA)

En aquesta tècnica, l'espectre del canal es divideix en subcanals i s'assiguals un subcanal a cada estació. Cada estació pot transmetre quan vulgui, ja que disposa d'un canal per a ella sola. A diferència del TDMA, ha de transmetre els paquets a una velocitat M vegades inferior, en disposar només d'una M -èsima part de l'ample de banda de canal (essent M el nombre d'estacions).

Anàlisi del FDMA

Per analitzar el FDMA, calculem el temps mitjà de transferència d'un paquet. Per fer-ho, utilitzem les hipòtesis següents:

- La xarxa està formada per M estacions iguals.
- El canal té un ample de banda de BW Hz i s'utilitza una modulació que proporciona una capacitat de C bps.
- La longitud mitjana dels paquets és igual a L bits.
- Cada estació i genera λ_i paquets/s, amb un procés de Poisson (totes iguals).
- Cada paquet espera a la cua de la seva estació fins que pot ser transmès.

Cada estació disposa d'una amplada de banda de BW/M Hz; aleshores, la velocitat de transmissió de cada estació és C/M bps. A més, la longitud dels paquets és constant; per tant, cada estació es pot modelar com un sistema M/D/1.

La utilització d'un canal val:

$$\rho = \lambda_i T_t = \lambda_i \frac{L}{C/M} = \frac{M\lambda_i L}{C}$$

Observeu que la ρ es pot interpretar per a totes les estacions juntes o per a una única estació. Com que cada estació té un canal diferenciat, totes poden transmetre simultàniament; per tant, la fracció de temps en què està transmetent l'estació i és igual a la fracció de temps en què estan ocupant el canal totes les estacions (recordem que totes les estacions són iguals).

Finalment, aplicant la fórmula de Pollaczek-Khinchine, tenim:

$$T_w = \frac{\lambda_i T_t^2}{2(1 - \rho)} = \frac{\rho}{2(1 - \rho)} \frac{ML}{C}$$

$$T = \frac{\rho}{2(1 - \rho)} \frac{ML}{C} + \frac{ML}{C}$$

Si resten els temps de transferència del FDMA i el TDMA, obtenim que la relació entre els seus temps de transferència és:

$$T_{FDMA} = T_{TDMA} + \frac{L}{C} \left[\frac{M}{2} - 1 \right]$$

De l'expressió es dedueix que, per a més de dues estacions, el temps de transferència del FDMA és més gran que el del TDMA.

4.2.3 Sondeig (*polling*)

A les xarxes d'accés per sondeig, el control d'accés es fa de manera centralitzada. Una estació central va preguntant a la resta d'estacions si tenen informació per transmetre. Aquest sondeig es fa seguint un ordre predeterminat. Les estacions emmagatzemem la informació pendent de transmetre a la seva memòria fins que són sondejades. Quan són sondejades, disposen de tota la capacitat de canal per transmetre la seva informació. A causa del sondeig, hi ha un trànsit d'*overhead*.



Figura 101.
Escenari d'un meca-nisme per sondeig



Anàlisi de la tècnica per sondeig

Per analitzar la tècnica per sondeig, calculem el temps mitjà de transferència d'un paquet. Per fer-ho, utilitzem les hipòtesis següents:

- Hi ha M estacions iguals i una estació central.
- Tot el trànsit d'informació va des de les estacions cap a l'estació central. El paquet de sondeig va de l'estació central a les estacions.
- Cada estació genera λ_i paquets/s, amb un procés de Poisson (totes iguals).
- La longitud dels paquets és una variable aleatòria de la qual es coneix la funció densitat de probabilitat.
- El paquet de sondeig té longitud fixa.
- Els paquets s'emmagatzemen a cada estació fins que són sondejades i, quan ho són, buiden completament el seu *buffer* (això es coneix com a *sistema exhaustiu*).
- Al final de la transmissió, les estacions afegeixen un sufix per indicar que no tenen més informació per transmetre. Aquest sufix té una longitud fixa.

Es defineixen els conceptes següents:

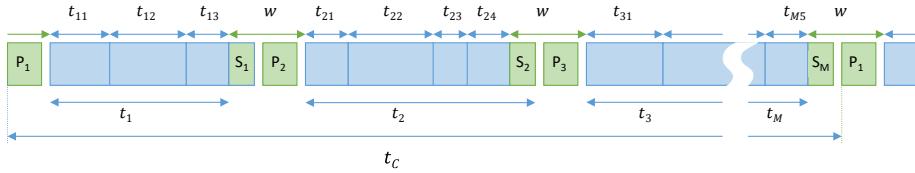
- **Walk time** (o temps de passeig): temps necessari per transferir el sondeig d'una estació a una altra, és a dir, és el temps des que una estació deixa de transmetre paquets (inici del sufix) fins que la següent en comença a transmetre. Aquest temps inclou el temps que requereixen totes les funcions necessàries per transferir el sondeig: retard de propagació, temps de transmissió del paquet de sondeig i del sufix.

El denotem amb la lletra w i, a partir de les hipòtesis, tenim que és constant per a tots els sondejos.

- **Temps de cicle:** temps total necessari per sondejar cada estació i tornar a l'estació inicial en la seqüència de sondeig, és a dir, és el temps des que se sondea una estació fins que aquesta torna a ser sondejada.

És una variable aleatòria que denotem t_c . El seu valor esperat el denotem amb majúscules, $T_c = E(t_c)$.

Abans de començar l'anàlisi del temps de transferència, calculem quant val el temps mitjà de cicle.


 Figura 102.
Temps de cicle

A la figura 102, es mostra un cicle on P_i representa el paquet de sondeig cap a l'estació i , i S_i és el sufíx que afegeix l'estació i per indicar que no té res més per transmetre. Cada parella de sufíx i sondeig consecutius formen un *walk time* (aquí també s'hauria d'incloure qualsevol retard de propagació entre l'estació central i una estació qualsevol). L'estació i transmet durant un temps t_i , que és la suma dels temps de transmissió dels paquets que hi ha al *buffer* ($t_{i1} + t_{i2} + \dots$).

Aleshores:

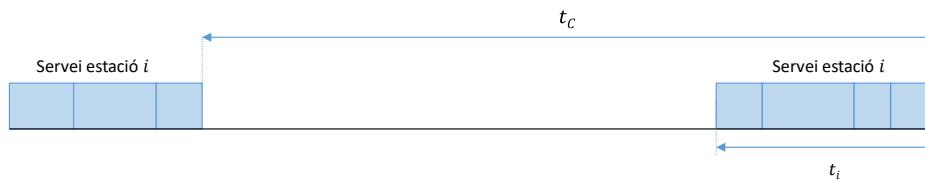
$$t_c = \sum_{i=1}^M w + t_i \Rightarrow T_c = E(t_c) = Mw + \sum_{i=1}^M E(t_i)$$

Per calcular el valor mitjà de t_i , cal saber el nombre mitjà de paquets que transmet l'estació i quan buida el seu *buffer* N_i , i multiplicar-lo pel temps mitjà de transmissió d'un paquet T_t .

$$E(t_i) = N_i T_t$$

Tal com es veu a la figura següent, N_i és igual al nombre mitjà de paquets que genera l'estació i en un cicle, ja que l'estació buida completament el seu *buffer* a cada cicle.

$$N_i = \lambda_i \cdot T_c$$


 Figura 103.
Nombre mitjà de paquets transmesos en un cicle

Per tant:

$$E(t_i) = \lambda_i \cdot T_c \cdot T_t = \rho_i \cdot T_c$$

$$T_c = Mw + T_c \sum_{i=1}^M \rho_i$$

Finalment:

$$T_c = \frac{Mw}{1 - \rho}$$



Observeu que ρ_i és la fracció de temps durant la qual l'estació i està ocupant el canal i ρ és la fracció de temps durant la qual totes les estacions ocupen el canal (és la suma de totes les ρ_i). Fixeu-vos que el temps de cicle és finit mentre $\rho < 1$, és a dir, mentre l'ocupació del canal per part de les estacions és inferior al 100% del temps. Si manipulem l'expressió anterior, tenim:

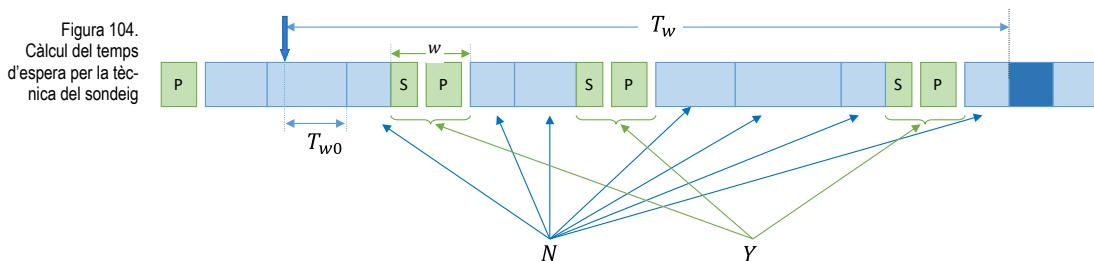
$$1 = \frac{Mw}{T_c} + \rho$$

El primer sumand és la fracció de temps durant la qual s'està sondejant i el segon sumand és la fracció de temps durant la qual s'està transmetent informació. Com que sempre s'està ocupant el canal, ja sigui amb sondejos o amb transmissions, la seva suma val 1. Així doncs, mentre ρ és inferior a 1, el temps de cicle s'adapta. Si ρ augmenta, la fracció de temps durant la qual se sondea ha de disminuir, cosa que s'aconsegueix augmentant el temps de cicle.

En cas que no hi hagi trànsit per transmetre ($\rho = 0$), el temps de cicle val $T^{(0)} = Mw$. En aquest cas, el cicle està format únicament per M sondejos. Alguns autors utilitzen aquest valor a l'hora d'expressar el temps de cicle:

$$T_c = \frac{T^{(0)}}{1 - \rho}$$

Per fer el càlcul del temps mitjà de transferència d'un paquet, primer se'n calcula el **temps mitjà d'espera al buffer**, és a dir, el temps des que es genera el paquet fins que s'inicia la seva transmissió. Aquest temps d'espera es mostra a la figura següent:



essent:

- T_w : temps mitjà d'espera al buffer.
- T_{w0} : temps mitjà residual de la unitat que s'estava transmetent en el moment en què es genera el paquet.
- N : nombre mitjà de paquets que seran transmesos abans que el paquet objecte d'estudi.

- Y : nombre mitjà de sondejos que es realitzaran abans que es transmeti el paquet objecte d'estudi.

A partir de la figura, tenim que el temps mitjà d'espera val:

$$T_w = T_{w0} + Yw + NT_t$$

Ens cal trobar T_{w0} , Y i N . Comencem per N .

En condicions d'estabilitat, el nombre mitjà de paquets en espera a tota la xarxa ha de ser un valor constant. No té sentit pensar que el nombre mitjà de paquets en espera a tota la xarxa creix o decreix. Mentre una estació buida el *buffer* perquè és el seu torn de transmissió, la resta els van omplint amb nous paquets fins que els hi arribi el torn per poder-los transmetre. Per tant, el nombre mitjà de paquets transmesos durant un interval de temps serà igual al nombre mitjà de paquets generats per totes les estacions durant el mateix interval de temps. Així doncs, el nombre mitjà de paquets que seran transmesos abans que el paquet objecte d'estudi ha de coincidir amb el nombre de paquets que es generen a la xarxa mentre el paquet objecte d'estudi s'espera:

$$N = M\lambda_i T_w$$

Substituint a l'equació del temps d'espera:

$$T_w = T_{w0} + Yw + M\lambda_i T_w T_t$$

$$T_w = \frac{T_{w0} + Yw}{1 - \rho}$$

Per al càlcul de T_{w0} , cal tenir en compte que, quan es genera un paquet, a la xarxa poden estar succeint dues coses:

- El canal està ocupat amb la transmissió d'un paquet. Això passa amb una probabilitat ρ .
- El canal està ocupat amb el sondeig en alguna estació. Això passa amb una probabilitat $1 - \rho$.

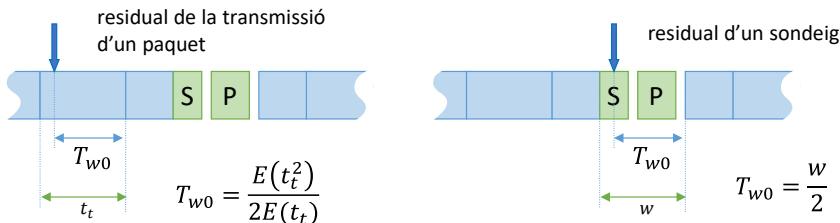


Figura 105.
Temps residuals

Com que el *walk time* és constant, el residual durant un sondeig és $w/2$. Per tant, el valor mitjà del temps residual val:



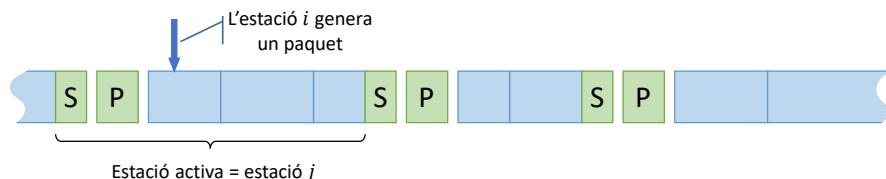
$$T_{w0} = \rho \frac{E(t_t^2)}{2E(t_t)} + (1 - \rho) \frac{w}{2} = \frac{\lambda E(t_t^2)}{2} + (1 - \rho) \frac{w}{2}$$

Observeu que el valor de λ que apareix a l'expressió es refereix a la taxa d'arribada de totes les estacions de la xarxa, és a dir, $\lambda = \sum_{i=1}^M \lambda_i$.

Finalment, calclem el valor de Y . Quan es genera un paquet a l'estació i (paquet objecte d'estudi), qualsevol estació, inclosa la i , podria estar transmetent o bé essent sondejada. Aquesta estació l'anomenem *estació activa*. Com que les arribades són de Poisson, tots els instants són igualment probables; a més, com que totes les estacions són iguals, quan es hi arriba el paquet objecte d'estudi totes les estacions tenen la mateixa probabilitat de ser l'estació activa.

$$p(\text{estació activa} = \text{estació } j) = \frac{1}{M}$$

Figura 106.
Càlcul del valor Y



Suposem que l'estació j és l'estació activa mentre es produeix una arribada a l'estació i ; aleshores:

- Si $i = j$, hi haurà 0 sondejos abans de la transmissió del paquet objecte d'estudi; per tant, en aquest cas, $y = 0$.
- Si $i = j + 1$, hi haurà 1 sondeig abans de la transmissió del paquet objecte d'estudi; per tant, en aquest cas, $y = 1$.
- Si $i = j + 2$, hi haurà 2 sondejos abans de la transmissió del paquet objecte d'estudi; per tant, en aquest cas, $y = 2$.
- ...
- Si $i = j + M - 1$, hi haurà $M - 1$ sondejos abans de la transmissió del paquet objecte d'estudi; per tant, en aquest cas, $y = M - 1$.

Tots aquests casos es produeixen amb la mateixa probabilitat i, en conseqüència, el seu valor mitjà val:

$$Y = \frac{1}{M} \sum_{i=0}^{M-1} i = \frac{1}{M} \frac{(M-1)M}{2} = \frac{M-1}{2}$$

Substituint els valors de T_{w0} i Y , se n'obté:

$$T_w = \frac{1}{1-\rho} \left[\frac{\lambda E(t_t^2)}{2} + (1+\rho) \frac{w}{2} + \frac{M-1}{2} w \right]$$

Finalment, el temps d'espera val:

$$T_w = \frac{\lambda E(t_t^2)}{2(1-\rho)} + \frac{(M-\rho)w}{2(1-\rho)}$$

Tan bon punt tenim el temps mitjà d'espera, és immediat trobar el temps de transferència,

$$T = T_w + T_t$$

Finalment, aplicant la fórmula de Little, podem calcular el nombre mitjà de paquets a la cua en cada estació:

$$N_{w_i} = \lambda_i T_w$$





Bibliografia

BERTSEKAS, D. P.; GALLAGER, R. G. (1992): *Data Networks*. 2a ed. Englewood Cliffs, NJ: Prentice-Hall.

GROSS, D. et al. (2008): *Fundamentals of Queueing Theory*. 4a ed. John Wiley & Sons.

HAMMOND J. L.; O'REILLY, P.J.P. (1986): *Performance Analysis of Local Computer Networks*. Addison-Wesley.

KLEINROCK, L. (1975): *Queueing Systems. Theory*. Vol. 1. Nova York: John Wiley & Sons.