CHAPITRE 3

Performance et Dimensionnement des Réseaux

III.1. Introduction

L'évolution permanente des systèmes et réseaux informatiques et des télécommunications souligne le besoin croissant d'outils facilitant l'étude de leur comportement. Il est nécessaire d'avoir une assistance dans les phases de conception (comparaison des solutions), de développement (dimensionnement) et d'utilisation (gestion des flux et de la QoS : Quality of Service) des systèmes et réseaux. En effet, le développement d'un système complexe demande non seulement une modélisation qualitative pour vérifier sa correction logique mais aussi une validation a priori des performances du système lors de la phase de conception. En plus, lorsque ces systèmes possèdent des contraintes temporelles (applications temps réel, par exemple), nous devons inclure parmi les éléments à retenir des considérations de performances qui ne peuvent être abordées avec rigueur que grâce à l'utilisation de techniques quantitatives pour l'évaluation de performances.

Les réseaux informatiques et de télécommunications numériques sont conçus comme des ressources à partager par plusieurs utilisateurs. Afin de gérer le problème d'accès multiples à une ressource, on utilise une mémoire tampon pour stocker momentanément des demandes que l'on n'arrive pas à traiter instantanément. Dans un équipement réseau, chaque fois il y a une mémoire tampon, il y a des attentes de traitement. Il est donc naturel de modéliser un réseau sous le formalisme de files d'attente. Ces délais d'attente dépendront de l'ensemble des trafics entrants. Selon la connaissance complète ou partielle dont on dispose sur l'application, leur évaluation peut donner des résultats du type exact, probabiliste ou de bornes. La figure suivante (Figure 3.1) met en évidence les attentes dans un réseau.

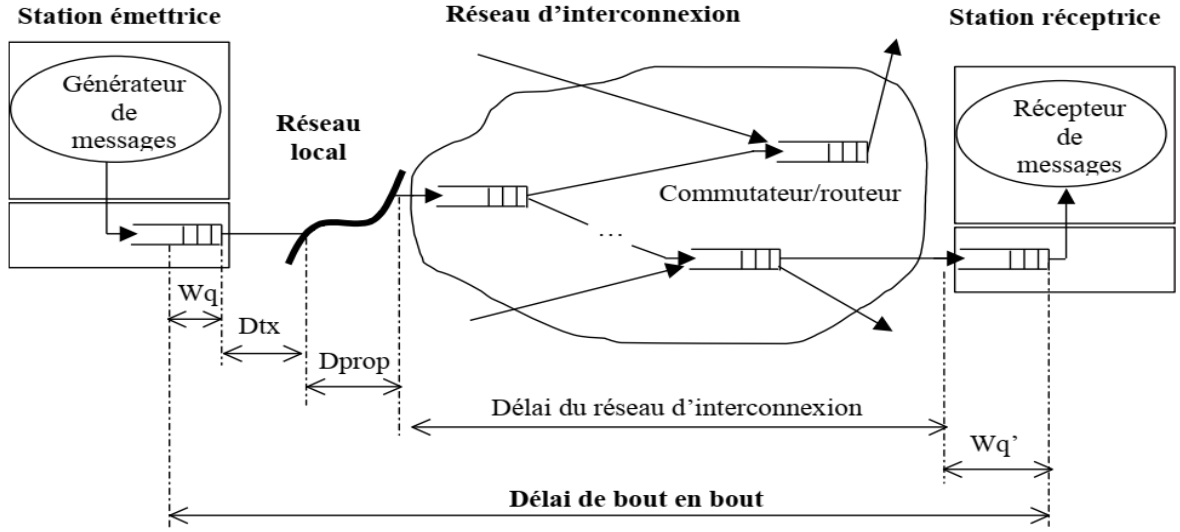


Figure 3.1 : Les attentes dans un réseau de transmission de données

III.2. Généralités

Pour satisfaire les besoins de télécommunication des entreprises, les opérateurs (prestataires de service de télécommunication) offrent deux types de service, les services supports et les téléservices.

Les **services supports** consistent à fournir à un abonné (l'entreprise) un lien de communication (support physique ou virtuel) entre deux points (ligne point à point) ou entre un et plusieurs points (ligne multipoint). Les services offerts peuvent être simples comme la location d'un lien (liaison louée). L'opérateur met alors à disposition de l'entreprise un lien et les organes d'extrémités d'accès (modems). Lorsque ce support est numérique l'opérateur ne s'engage que sur le débit nominal du lien et le taux d'erreur. Quand il s'agit d'un lien analogique seule la bande passante et le rapport signal à bruit sont garantis. Les liaisons louées sont utilisées pour réaliser des liaisons point à point et les réseaux privés d'entreprise.

Une alternative à la réalisation d'un réseau privé consiste à relier les divers établissements d'une entreprise via un réseau de transport public. L'opérateur assure alors le raccordement des sites de l'abonné à son réseau. Les caractéristiques du raccordement dépendent alors du réseau de l'opérateur (débit, protocole...).

Les **téléservices** constituent une offre plus élaborée. L'opérateur fournit alors un moyen complet de communication. Dans cette catégorie, on trouve la téléphonie analogique (RTC) et numérique (RNIS), les services de messagerie, les services télématiques (vidéotex)...

Le type de service requis est déterminé en fonction des services attendus (téléservices) ou de critères techniques (services supports). La politique tarifaire de l'opérateur, l'évolution prévisible des besoins et la reprise de l'existant sont les éléments déterminants devant guider le choix.

L'ingénierie des réseaux recouvre l'ensemble des moyens mis en œuvre pour le choix d'une solution de télécommunication et la réalisation d'un réseau de télécommunication (topologie, dimensionnement, évaluation des performances, optimisation...).

III.3. Eléments d'architecture des réseaux

III.3.1. Structure de base des réseaux

En principe, un réseau comporte deux sous-ensembles : le **réseau dorsal** ou réseau de transit (backbone) et le **réseau de desserte** ou réseau capillaire ou encore réseau de bas niveau (figure 3.2). Le premier est un réseau de concentration qui assure la mise en relation (connectivité) des différentes composantes du réseau de desserte. Le second assure la distribution du service aux abonnés, il est composé d'un ensemble de liens et de concentrateurs. Un niveau de concentration élevé optimise le réseau mais le fragilise.

Le choix de la localisation et du nombre de points d'accès détermine le coût de raccordement des abonnés (sites terminaux de l'entreprise ou abonnés d'un opérateur). De ce fait, il existe une relation étroite entre le graphe du réseau dorsal et le graphe du réseau de desserte.

La réalisation d'un réseau privé dans le contexte d'une entreprise donnée est relativement simple, l'emplacement des points d'accès au réseau est parfaitement déterminé. Dans ces réseaux, ce sont

les points de concentration et le mode de liaison entre les différents sites qu'il convient d'optimiser. La topologie est généralement arborescente, le maillage n'intervient qu'en second lieu et essentiellement pour des critères de sécurité (redondance de liens).

Celle des réseaux publics est plus complexe, non seulement les points d'accès sont nombreux, mais les points de concentration ne sont pas prédéterminés et les contraintes de qualité de service sont plus strictes. La solution optimale est généralement le résultat d'un compromis obtenu par itérations successives en formulant différentes hypothèses de positionnement des nœuds d'accès et de concentration.

Quelles que soient les hypothèses formulées, un réseau résultera de compromis entre performance, fiabilité et coût. En effet, un niveau de performance élevé nécessite un maillage élevé du réseau dorsal et un faible niveau de concentration dans le réseau de desserte, ce qui en accroît la fiabilité mais en augmente aussi le coût.

Il existe de nombreuses méthodes pour optimiser la topologie des réseaux selon un critère prédéterminé (coût, débit, performance...). Certaines essaient de trouver la solution optimale, elles sont difficiles à mettre en œuvre et résistent mal à l'introduction de nouvelles contraintes ou à une évolution des besoins. D'autres méthodes, dites heuristiques, ne cherchent pas à déterminer la solution optimale mais à s'en rapprocher le plus possible. Elles prennent facilement en compte de nombreuses contraintes. Leur principe est simple, à partir d'hypothèses sur le positionnement des composants, elles les réunissent deux à deux par le lien de moindre coût.

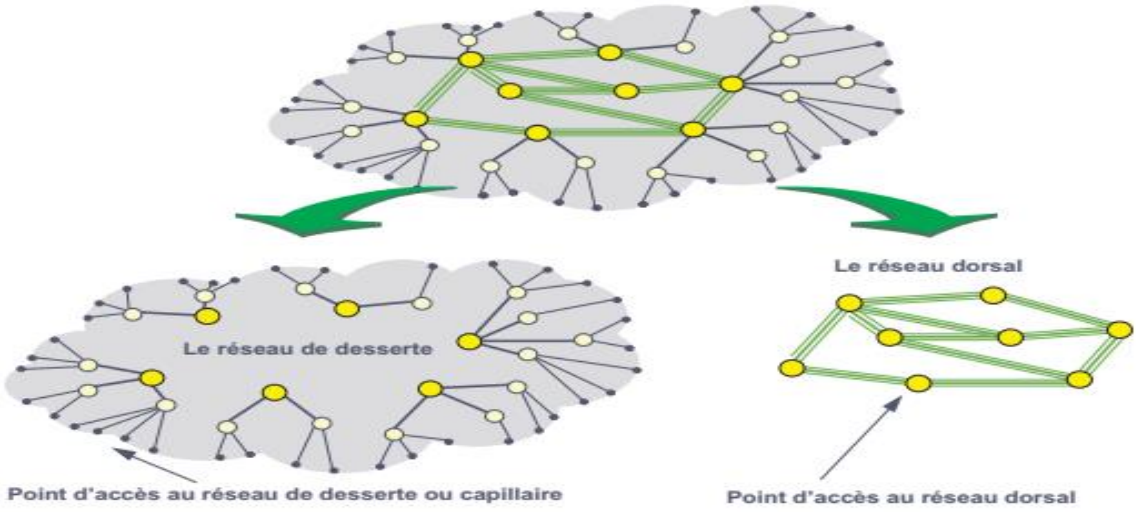


Figure 3.2 : Architecture générale d'un réseau.

III.3.2. Conception du réseau de desserte

Du fait de la densité des points de concentration et des liaisons, le réseau de desserte est généralement le plus coûteux à réaliser. Aussi, c'est à partir de lui que la localisation des points d'accès au réseau dorsal est déterminée. Ainsi, le réseau de desserte est étudié nœud par nœud en formulant diverses hypothèses sur la localisation des nœuds du niveau supérieur. Après avoir défini les niveaux de concentration, on affecte les concentrateurs de niveau N aux concentrateurs de niveaux N – 1, en respectant les contraintes de nombre de branches (fiabilité), de débit, de coût...

Les deux méthodes exposées ci-dessous dérivent de l'algorithme de Prim et de Krustal. L'algorithme de Prim sous contrainte consiste à :

1) formuler une hypothèse sur le positionnement du nœud de concentration principal (accès au réseau dorsal...),

2) définir les points à raccorder,

3) déterminer le coût ($\mathbf{C_{xy}}$) de toutes les liaisons une à une (matrice des coûts) entre les différents nœuds du réseau,

4) trier les liens par ordre croissant des coûts,

5) considérer le nœud de concentration comme nœud central et l'inclure en premier dans l'arbre,

6) tant que les contraintes fixées pour les raccordements sont respectées (débit maximal à concentrer, nombre de liens regroupés...) relier le point dont le coût de connexion est le plus faible à l'un des points déjà contenus dans l'arbre,

7) réitérer l'action précédente jusqu'à ce que tous les nœuds soient reliés, en ne retenant que les liaisons ne faisant pas de boucle et qui respectent les contraintes imposées.

La mise en œuvre de l'algorithme de Kruskal sous contrainte est aussi simple, elle comprend les étapes suivantes, les 4 premières étapes étant identiques à celles de l'algorithme de Prim :

1) construire la première branche en réunissant les 2 nœuds dont le coût de la liaison est le plus faible ;

2) réitérer l'action précédente jusqu'à ce que tous les nœuds soient reliés, en ne retenant que les liaisons ne faisant pas de boucle et qui respectent les contraintes imposées (débit global, nombre de liens sur un même point de concentration...) ;

3) quand pour une branche, un regroupement ne peut plus être opéré (contrainte atteinte), le regroupement est considéré comme un sous-arbre qui sera relié au nœud de concentration supérieur élu, par son nœud le plus proche.

	a	b	c	d	e	f	g	h	i	j	k
a		35	40	74	69	42	40	64	60	25	43
b			31	61	80	75	55	95	91	60	76
c				35	50	84	32	88	95	84	82
d					49	89	51	111	124	96	116
e						58	29	74	96	77	97
f							38	23	37	29	41
g								60	74	50	70
h									25	41	44
i										35	27
j											20

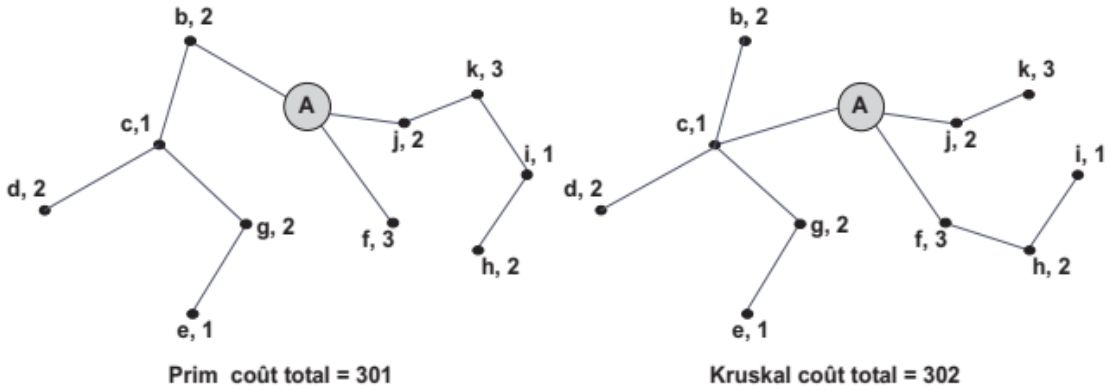


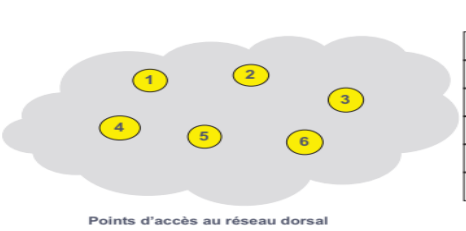
Figure 3.3 : Matrice des coûts et arbres à coût minimal de Prim et Kruskal.

La figure 3.3 illustre ces deux méthodes. Le coût retenu est la distance séparant les différents nœuds. Le nœud « A » pouvant représenter le nœud de concentration de niveau supérieur ou le site informatique principal d'une entreprise vers lequel tous les sites doivent converger. La contrainte retenue ici a été le débit, celui-ci ne devant pas excéder pour cet exemple. Le débit requis par chaque nœud figure à droite de son identification. Bien que les graphes résultants diffèrent, les coûts sont ici très proches.

III.3.3. Conception du réseau dorsal

La recherche de performance et de fiabilité conduit à multiplier les liens inter-nœuds (maillage) de façon à assurer une redondance de route (fiabilité) et à minimiser le nombre de sauts (performance). Plusieurs méthodes existent, l'une des plus simples, la méthode de Steiglitz, donne des résultats satisfaisants pour la plupart des réseaux d'entreprise. Cette méthode, illustrée figure 3.4, consiste à

1) définir le positionnement de tous les nœuds et leur connectivité (nombre de liens aboutissant à un nœud),
2) attribuer, arbitrairement, un poids à chaque nœud,
3) relier le nœud de plus faible connectivité (en cas d'égalité prendre le nœud de plus faible poids) à son voisin de moindre coût,
4) répéter l'opération tant que la connectivité de chaque nœud n'est pas atteinte,
5) modifier les poids attribués, refaire le graphe jusqu'à trouver celui de coût minimal.



	1	2	3	4	5	6
1		52	101	40	52	95
2	52		50	85	57	55
3	101	50		123	80	32
4	40	85	123		48	107
5	52	57	80	48		60
6	95	55	32	107	60	

Matrice des coûts

Figure 3.4 : Conception du réseau dorsal.

Au début, tous les nœuds ayant la même connectivité (0), on prend le nœud de poids le plus faible soit 1, le nœud 4 est le voisin à coût minimal, on relie 1 à 4. Ensuite, on prend le suivant 2 (poids le plus faible) on le relie à 3 (coût le plus faible)... En final on obtient le réseau de la figure 3.5 (gauche).

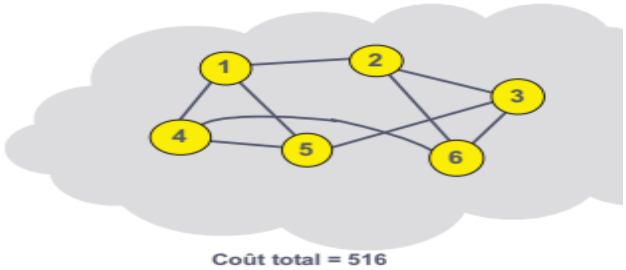


Figure 3.5 : Graphe du réseau dorsal.

En principe, la réalisation d'un réseau d'entreprise est plus simple : les points de concentration coïncident avec les points de consommation et sont déterminés (établissements de l'entreprise), beaucoup d'entreprises mettent en œuvre une topologie étoile autour de leur centre informatique principal. Lorsque les contraintes sont faibles en termes de débit, mais fortes en termes de coût, l'algorithme de Kruskal vu précédemment donne d'excellents résultats.

III.4. Dimensionnement et évaluation des performance

III.4.1. Généralités

Evaluer les performances d'un réseau est une tâche essentielle, tant lors de la conception que durant tout le cycle de vie du système. Indépendamment de mesures concrètes (débit, temps de transit...), on utilise généralement un modèle mathématique du réseau (modélisation) permettant de déterminer le comportement du réseau en fonction de certains paramètres et de leur variation (trafic, état d'un lien...). Les éléments pris en compte diffèrent selon que le réseau est en mode circuits ou en mode paquets (figure 3.6) :

- Dans **les réseaux en mode circuits** (circuits physiques ou IT, Intervalle de Temps), la bande passante est affectée en permanence à un utilisateur, le mode de transfert est dit synchrone (bande passante fixée par le réseau). La problématique consiste à définir le nombre de liens (IT) en fonction du nombre de sessions utilisateurs par unité de temps (taux d'arrivée) et de la durée moyenne d'une session.
- Dans **les réseaux en mode paquets**, il n'y a pas de ressource affectée, la bande passante est partagée. Il n'y a pas de relation directe entre le débit de la source et celui du réseau, le mode de transfert est dit asynchrone. Soumis à des trafics sporadiques, et souvent sous dimensionnés par rapport au trafic total à écouler, ils sont sensibles à la congestion. Les réseaux en mode connecté avec réservation de ressource relèvent à la fois du modèle des réseaux en mode circuit pour le nombre de connexions admises et du modèle des réseaux en mode paquet en ce qui concerne les performances.

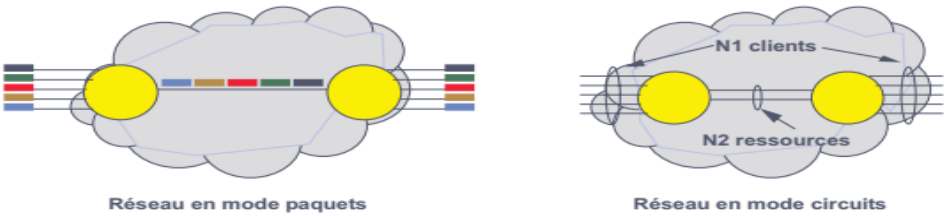


Figure 3.6 : Les deux types de réseaux.

III.4.2. Les réseaux en mode circuit

A. Notion d'intensité de trafic

Lorsqu'un système établit une connexion permanente, le nombre de liens à établir est simple à définir. Si N sites doivent être reliés simultanément à un autre site, généralement désigné par site central, il est nécessaire de disposer de N lignes. Mais lorsque les systèmes d'extrémité n'établissent qu'une connexion temporaire, ce serait un gâchis de ressource que d'équiper le site central d'autant de lignes qu'il y a de possibilité de mise en relation. C'est notamment le cas pour un serveur de messagerie auquel se connectent une fois par jour les itinérants des sociétés pour y relever leur courrier, transférer leurs commandes et éventuellement télécharger les nouveaux tarifs.

Partant du principe que les sources sont indépendantes les unes des autres (les arrivées sont supposées suivre une loi de Poisson), il n'est pas improbable que, dans certaines circonstances, les arrivées dépassent les capacités de traitement du système. Dans ces conditions, les requêtes en surnombre sont soit éliminées soit mises en attente de libération d'un organe de traitement

(support, ou autre). Sur ce principe, Erlang (mathématicien danois) a développé deux modèles l'un dit à refus ou blocage **(modèle B)**, l'autre dit à attente **(modèle C)**. Les modèles ont été élaborés à partir de l'expression de la quantité de trafic à satisfaire dénommée **intensité de trafic** et exprimée en erlang par la relation :

$$E = \frac{1}{T}\int_0^T n(t)\,dt \quad \textbf{ou plus simplement} \quad E = \frac{NT}{3600}$$

T : est exprimé en secondes, **E** : charge de trafic en erlang, **N** : nombre de sessions par heure

Ainsi, 1 erlang représente l'occupation permanente d'un organe pendant 1 heure, ou l'occupation effective de 2 organes à 50% (1/2 heure chacun) pendant la même période. L'intensité de trafic sur une ligne représente le taux de connexion de celle-ci, c'est-à-dire le temps d'appropriation de l'organe par unité de temps.

B. Modèle d'Erlang à refus (modèle B)

➢ Formule d'Erlang

Dans un système dit à pénurie de ressource, c'est-à-dire que le nombre de clients **(n)** est supérieur au nombre d'organes de traitement **(m)**, Erlang a établi que la probabilité **p** de refus d'appel (facteur de blocage) par suite d'encombrement ou autre pour un trafic à écouler de **E** erlang est donnée par la relation (figure 3.7) :

$$p = \frac{E^m/m!}{\sum_{k=0}^{k=m} E^k/k!}$$

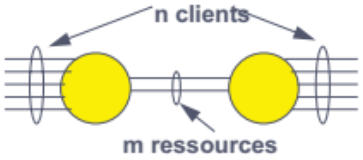


Figure 3.7 : Formule d'Erlang à refus.

Cette formule a permis d'établir un abaque de dimensionnement appelé abaque d'Erlang représenté figure 3.8.

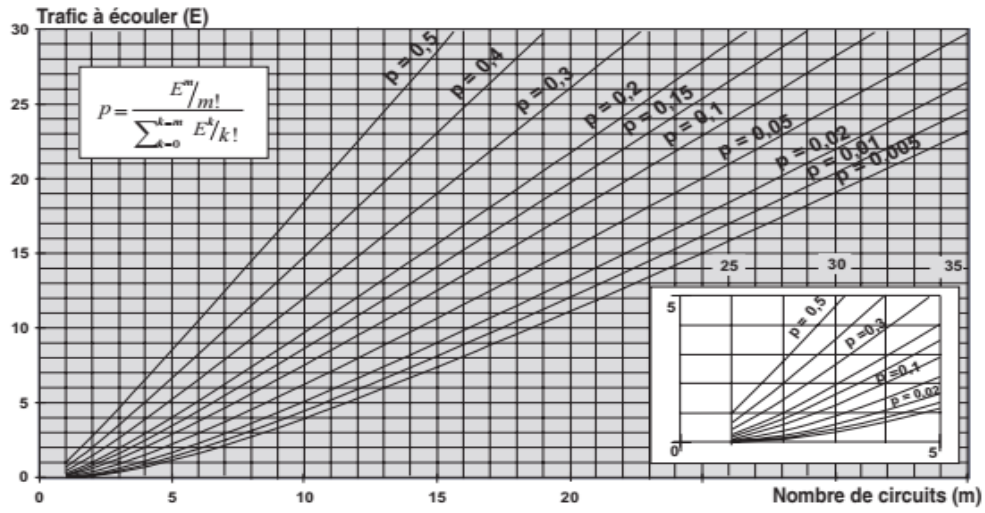


Figure 3.8 : Abaque d'Erlang (refus)

➢ Trafic demandé, trafic écoulé, trafic perdu

Par suite des collisions d'appel (organes occupés), une partie du trafic à écouler ou trafic demandé est perdue. Le trafic perdu et celui demandé sont liés par la relation :

Trafic perdu = Trafic demandé × p

où **p** est la probabilité de refus d'appel.

Le trafic réellement écoulé est alors :

Trafic écoulé = Trafic demandé - Trafic perdu = Trafic demandé(1 - p)

Le tableau 3.1 fournit quelques exemples pour un trafic de 0,9 erlang lorsque la ressource passe de 1 à 5 organes.

Trafic demandé	Ressources (m)	Probabilité de perte	Trafic perdu	Trafic écoulé
0,9 E	1	0,47	0,43	0,47
0,9 E	2	0,17	0,16	0,74
0,9 E	3	0,05	0,05	0,85
0,9 E	4	0,01	0,01	0,89
0,9 E	5	0,002	0,00	0,90

Tableau 3.1 : Trafic demandé, perdu et écoulé.

C. Modèle d'Erlang à attente (Modèle C)

Le modèle à attente suppose que toute demande ne pouvant être satisfaite est mise dans une file d'attente de capacité infinie et sera traitée dès qu'un organe se libère. C'est le cas notamment des processus dans les systèmes relais (routeur...). Cependant, on peut admettre que ce modèle s'applique aussi aux appels entrants des centraux téléphoniques d'entreprise. En effet, dans ces centraux, en cas d'occupation de l'appelé, l'appel entrant n'est généralement pas refusé mais mis en attente. Ce procédé monopolise une ressource en arrivée et n'écoule aucun trafic. Dans le modèle d'Erlang **C**, la probabilité $\mathbf{p_a}$ de mise en attente est donnée par la relation

$$p_a = \frac{(E^m/m!)\cdot\left(\frac{m}{m-E}\right)}{\left[\sum_{k=0}^{k=m-1} E^k/k!\right] + \left[(E^m/m!)\cdot\left(\frac{m}{m-E}\right)\right]}$$

III.4.3. Les réseaux en mode paquets

A. Principe de la modélisation des réseaux

L'échange d'information entre deux applications ou entre une application et un terminal est généralement variable en termes de volume à échanger et de fréquence des échanges. Si l'on prend en compte l'écart de débit, entre les systèmes locaux et les liens d'interconnexion, l'organe d'interconnexion du système de transmission (routeur, FRAD, modem...) constitue un goulet d'étranglement. Il en résulte que les messages ne seront pas transmis instantanément, ils seront placés dans une file d'attente (queue) avant d'être traités par le système de transmission. En première approximation, on peut modéliser un système de transmission comme étant constitué d'une simple ligne reliant les deux systèmes d'extrémité (figure 3.9).

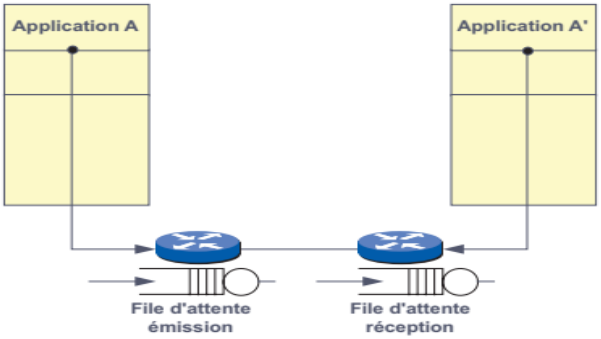


Figure 3.9 : Modélisation simplifiée d'un système de transmission.

Le temps de réponse du système pour un message peut s'exprimer par la relation :

$$\mathbf{T_r = T_p + T_t + T_q}$$

où $\mathbf{T_r}$ est le temps de réponse du système, $\mathbf{T_p}$ le temps de propagation sur le support, $\mathbf{T_t}$ le temps de traitement par les systèmes d'interconnexion, $\mathbf{T_q}$ le temps de séjour dans les systèmes ou temps de queue.

Lorsque le nombre de messages à traiter augmente, $\mathbf{T_q}$ devient prépondérant par rapport aux autres éléments, ces derniers pourront alors être négligés. Le problème consiste donc à calculer le temps de queue en fonction des caractéristiques du système (nombre d'items traités par unité de temps ou **m**), de la longueur du message (**L**) et du débit de la ligne du système (**D**). Il est possible aussi, à partir de ces éléments, de déterminer les caractéristiques du système pour répondre aux contraintes temporelles des applications.

B. Notions de files d'attente

➢ Relation de base, formule de Little

Dans un système à file d'attente en équilibre (figure 3.10), c'est-à-dire que le nombre moyen d'entrées (l) est égal au nombre moyen de sorties par unité de temps, le nombre moyen d'items (clients) dans le système N est donné par la relation :

Figure 3.10 : Formule de Little.

Si la condition d'équilibre est respectée, la relation de Little s'applique à tous les types de files d'attente.

➢ Les files d'attente M/M/1/∞

Le modèle **M/M/1/∞** est le modèle plus simple retenu pour la modélisation des réseaux. Selon la notation de Kendal cette file est caractérisée par :

- M : processus Markovien en entrée (distribution exponentielle des arrivées),
- M : processus Markovien en sortie,
- 1 : il n'y a qu'un seul processeur (système mono-serveur),
- ∞ : la file d'attente a une capacité infinie, aucun item entrant n'est perdu.

Dans ce type de processus, on considère les arrivées indépendantes les unes des autres, ce qui est généralement le cas des processus informatiques où les systèmes d'extrémité s'échangent des messages indépendamment les uns des autres.

Dans le système **M/M/1/∞** (figure 3.11), on désigne par **λ** le taux d'arrivée des items dans la file, par **ta** le temps de séjour dans la file d'attente, par **ts** le temps de traitement par le système (son inverse **1/ts** représente le nombre d'items traités par unité de temps ou taux de service et est désigné par **μ**) et **tq** le temps qui s'écoule entre l'entrée d'un item dans le système et sa sortie (temps de séjour). Le nombre d'items en attente de traitement **Na** et le nombre d'items en cours de traitement **Ns** sont une fonction directe du temps d'attente ou de traitement et du taux d'arrivée :

$$\mathbf{Na = \lambda ta \text{ et } Ns = \lambda ts}$$

Le nombre d'items N dans le système peut s'écrire (décomposition de la formule de Little) :

$$N = Na + Ns = \lambda\,(ta + ts) \qquad \text{(relation 1)}$$

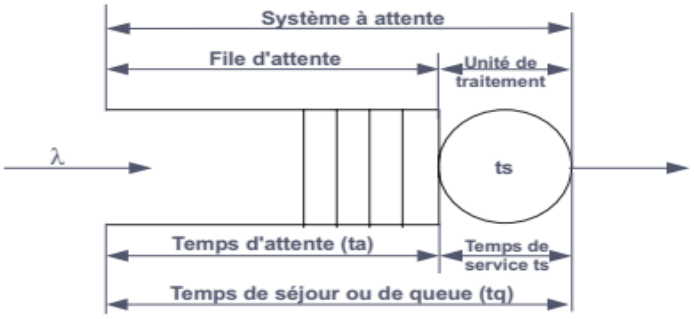


Figure 3.11 : Modélisation d'une file M/M/1/∞.

Pour tout nouvel entrant dans le système, le temps d'attente correspond au temps nécessaire pour traiter tous les items déjà présents dans la file soit :

$$ta = N \cdot ts \qquad \text{(relation 2)}$$

Dans la relation 1, en remplaçant la valeur de **ta** par celle obtenue dans la relation 2 :

$$N = \lambda\,(ta + ts) = \lambda\,(Isys \cdot ts + ts) = N \cdot \lambda\, ts + \lambda\, ts$$
$$N - N \cdot \lambda\, ts = \lambda ts$$
$$N(1 - \lambda\, ts) = \lambda ts$$
$$N = \lambda\, ts/(1 - \lambda ts)$$

Si on pose $\mu = 1/ts$ (taux de service du système), la charge du système r est représentée par la relation :

$$\rho = \lambda/\mu = \lambda ts$$

on obtient :

$$N = \frac{\rho}{1-\rho}$$

En reportant cette valeur dans la relation 2, on obtient :

$$ta = \frac{\rho}{1-\rho} ts$$

Le temps de séjour dans le système ou temps de queue est alors

$$tq = ta + ts = ts(1 + \frac{\rho}{1-\rho}) = ts(\frac{1}{1-\rho})$$

Étudions la variation du temps de séjour dans la file d'attente quand la charge du système varie de 10 à 100% par pas de 10%.

Le temps de séjour dans la file est relativement faible jusqu'à une charge de 50%. Au-dessus de cette charge, le temps de séjour croît très vite (tableau 3.2). Compte tenu qu'un système est généralement étudié pour un trafic moyen, pour garantir, lors de trafic de crête, des performances acceptables, il convient de le dimensionner pour que la charge moyenne de celui-ci n'excède pas 50%.

Charge	10%	20%	30%	40 %	50 %	60 %	70 %	80 %	90 %	100 %
r/(1-r)	0,1/0,9	0,2/0,8	0,3/0,7	0,4/0,6	0,5/0,5	0,6/0,4	0,7/0,3	0,8/0,2	0,9/0,1	1/0
ta=ts . r/(1-r)	0,11	0,25	0,42	0,66	1	1,5	2,3	4	9	∞
	Temps d'attente acceptable					**Temps d'attente prohibitif**				

Tableau 3.2 Évolution du temps de séjour en fonction de la charge.

➢ **Application au calcul du débit d'une ligne**

En exprimant la charge d'une ligne en fonction des caractéristiques du message (L : longueur en bits) et du système (D : débit de la ligne en bit/s), on obtient :

$$\rho = \frac{\lambda}{\mu} = \frac{\lambda}{\frac{1}{ts}} = \lambda ts = \frac{\lambda L}{D}$$

Le temps séjour dans le système ou temps de queue est alors :

$$tq = ts\left(\frac{1}{1-\rho}\right) = \frac{L}{D}\left(\frac{1}{1-\frac{\lambda L}{D}}\right) = \frac{L}{D-\lambda L}$$

avec **D > λL**

Mais, compte tenu que nous venons d'établir qu'un système ne doit pas être chargé à plus de 50%, le débit minimal requis (D_{min}) pour la ligne est de :

$$D_{min} \geq 2\lambda L$$

➢ **File d'attente en série**

Lorsque deux files d'attente sont en série, si la file d'attente d'entrée est du type **M/M/1/∞**, les arrivées suivent une loi de Poisson ainsi que les sorties. Dans ces conditions, les entrées de la suivante respectent une loi de Poisson ; cette file est donc aussi poissonnière et de type **M/M/1/∞.** Le temps de traversée global est la somme des temps de traversée de chacune des files (figure 3.12).

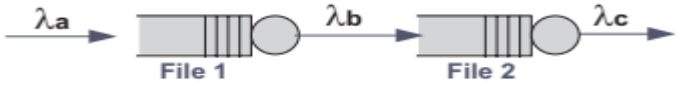


Figure 3.12 : Réseau ouvert de files d'attente en série.

Un réseau peut être modélisé comme une succession de files d'attente, et chaque élément actif du réseau (routeur, commutateur, FRAD...) peut lui-même être considéré comme deux files d'attente en série (figure 3.13).

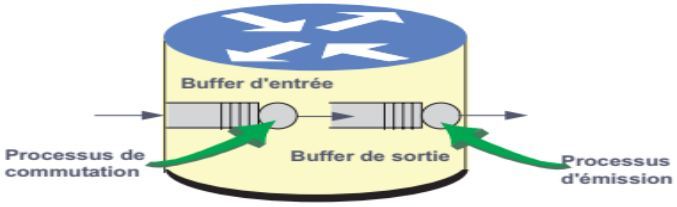


Figure 3.13 Modélisation simplifiée d'un élément actif.

➢ **File d'attente à entrées multiples**

On montre, que si les messages ont la même longueur moyenne sur toutes les voies incidentes, la file d'attente est équivalente à une file d'attente à une seule entrée où le taux d'arrivée est la somme des taux d'arrivée : $\lambda = \sum \lambda_i$ (principe de la superposition, figure 3.14).

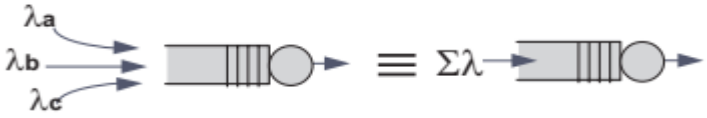


Figure 3.14 : File d'attente avec entrées multiples.

Par exemple, une file d'attente à entrées multiples peut correspondre à un nœud intermédiaire sur lequel convergent plusieurs liens.

➢ **Les files d'attente M/M/1/K**

Le modèle **M/M/1/∞** considère la file d'attente comme infinie, c'est-à-dire que tout item entrant sera traité, cette hypothèse simplificatrice est satisfaisante dans la majorité des cas. Mais lorsque le réseau est chargé, les files d'attente physiques (tampon ou buffer) n'étant pas de taille illimitée, la probabilité pour qu'un item entrant ne puisse être accepté n'est pas nulle (figure 3.15).

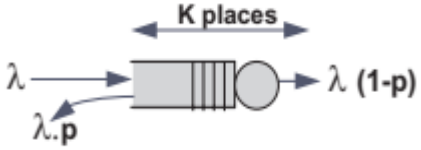


Figure 3.15 File d'attente à capacité limitée.

Si **K** est la taille de la file d'attente et que celle-ci contient déjà **n** items, la probabilité pour qu'un nouvel arrivant soit perdu ($\mathbf{p_n}$) est :

$$si \quad \rho \neq 1 \quad p_n = \frac{\rho^n (1-\rho)}{1-\rho^{K+1}} \quad avec \quad 0 \leq n \leq K$$

$$si \quad \rho = 1 \quad p_n = \frac{1}{K+1} \quad avec \quad 0 \leq n \leq K$$

Application à la modélisation d'un réseau

Soit le réseau représenté par la figure 3.16, si on admet que tout item entrant est sortant (système en équilibre), il peut être modélisé simplement en remplaçant chaque élément actif par une file d'attente. Il est alors possible à partir d'un trafic d'entrée (E) de déterminer le comportement du réseau pour les données en transit de E vers S.

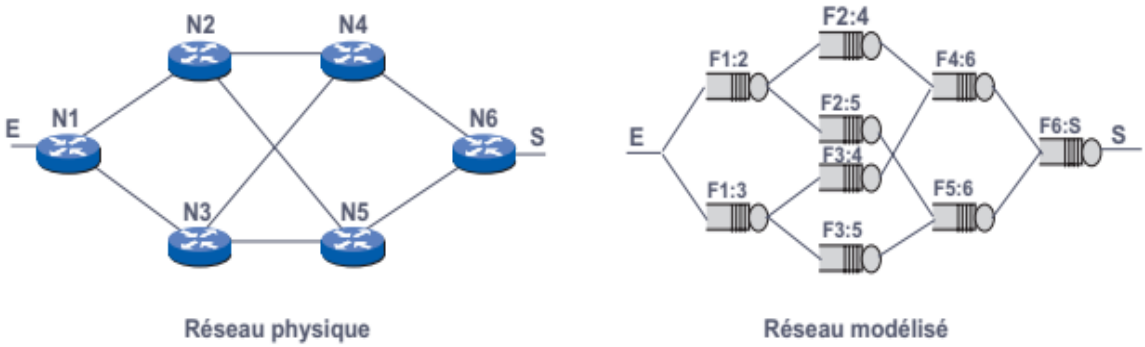


Figure 3.16 : Modélisation d'un réseau.

Pour définir le comportement global du réseau, il faut quantifier le trafic entre tous ses points d'entrée et de sortie (matrice de trafic). En superposant chacun des trafics élémentaires dans chaque file, on en détermine le comportement point à point puis on en déduit le comportement global du système.

III.5. Conclusion

Ce chapitre a pour objectif d'initier aux méthodes utilisées pour la réalisation, le dimensionnement et les mesures de performances des réseaux. Ces techniques conduisent à construire des modèles mathématiques de plus en plus élaborés. Devant la complexité des réseaux et les exigences des utilisateurs, les modèles deviennent de plus en plus complexes et la recherche de nouvelles techniques de modélisation permanente.