

GÉRARD HÉBUTERNE

Systemes à temps discret et réseaux ATM

RAIRO. Recherche opérationnelle, tome 32, n° 3 (1998),
p. 211-225

http://www.numdam.org/item?id=RO_1998__32_3_211_0

© AFCET, 1998, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

SYSTÈMES À TEMPS DISCRET ET RÉSEAUX ATM

par Gérard HÉBUTERNE

Résumé. – *L'analyse des performances des réseaux ATM fait appel aux modèles à temps discret. Nous étudions quelques aspects de la théorie des files d'attente à temps discret, liés aux comportements particuliers de synchronisation inhérents aux systèmes physiques, ou bien induits par la modélisation elle-même. Le but de cet article de synthèse est de mettre l'accent sur la spécificité de cette branche de la théorie des files d'attentes et sur quelques pièges que cachent les études de performance des systèmes à temps discret.* © Elsevier, Paris

Abstract. – *Performance analysis of ATM networks relies on discrete-time queueing models. We discuss various aspects of this theory related to synchronization effects, which may be already present in the system or may be introduced by the modeling approach. This survey paper intends to emphasize the specificity of discrete-time queueing theory and to bring attention to misleading conclusions it could lead to.* © Elsevier, Paris

1. INTRODUCTION

L'analyse des performances des systèmes temps réel réclame un effort de modélisation, qui vise à en isoler les particularités et à définir les critères sur lesquels leur Qualité de Service (QS) sera évaluée. L'outil idéal du modélisateur est évidemment la *théorie des files d'attentes*, née au début du siècle avec les travaux d'A.K. Erlang, suscités par le développement du téléphone. Les applications en gestion des stocks ont suivi, mais cette discipline a connu son épanouissement véritable à partir des années 70, avec l'arrivée de l'informatique: calculateurs fonctionnant en temps partagé – cf le *modèle de Scherr, 1967* (kleinrock 1976), puis réseaux téléinformatiques.

Ce n'est pas ici le lieu d'une histoire de cette théorie. On notera simplement que la majeure partie des efforts s'est portée sur la construction et la résolution de *modèles à temps continu*.

Institut National des Télécommunications, Dept. Réseaux et Services des Télécomms, 9, rue Charles Fourier, 91011 Evry cedex, France, Tel: +33 (0)1 60 76 45 83, Fax: +33 (0)1 60 76 42 91. Email: gerard.hebuterne@int-evry.fr

L'arrivée de la technique ATM va bouleverser le paysage. L'évaluation des performances des réseaux large bande fondés sur l'ATM conduit en effet à adopter la théorie des "*files à temps discret*" comme technique de base de la modélisation. Cette approche rend bien compte, en effet, de l'aspect périodique et synchrone des phénomènes (au niveau de la *cellule ATM*). Ainsi, l'ATM introduit d'emblée le temps discret au cœur de son fonctionnement, obligeant le modélisateur à se placer dans ce formalisme, ou bien à justifier le passage en continu par des argumentations plus ou moins construites ("modèle fluide", par exemple). La théorie des systèmes à temps discret, qui n'a bien sûr pas été ignorée des pionniers (voir par exemple la bibliographie rapportée dans (Hunter 1983)), n'a cependant reçu qu'une attention très mesurée jusqu'à l'arrivée de l'ATM.

Il est possible de construire une théorie des systèmes à temps discret calquée sur la théorie classique à temps continu, le processus de Poisson se transformant à cette occasion en processus de Bernoulli, etc. Cette approche revient à voir le temps discret comme un simple "raffinement" des modèles classiques à temps continu, qui donnerait une précision meilleure aux prévisions des études. Cette vue, sans être strictement erronée, se révèle quelque peu réductrice, comme on va tenter de le montrer.

C'est que le temps discret porte avec lui un certain nombre de caractéristiques nouvelles, que la modélisation se doit d'intégrer: événements simultanés, phénomènes de synchronisation. Cet article veut défendre la thèse de la spécificité des modèles à temps discret. Plus précisément,

- Certains phénomènes liés à la synchronisation ne peuvent être mis en lumière que par des modèles à temps discret.
- Une utilisation inappropriée des modèles discrets pourra introduire des phénomènes parasites.

La théorie des systèmes à temps discret a une déjà longue histoire – même si la bibliographie la concernant est restée modeste. On se reportera par exemple aux références citées dans (Hunter 1983, Bruneel et Kim 1993). La plupart des travaux appartiennent à la tendance évoquée plus haut, et voient le temps discret comme un raffinement du temps continu. La théorie des files à temps discret a trouvé ses applications essentielles dans les domaines des ateliers de production, de la gestion des stocks ou de la génétique. Dans le domaine des télécommunications, elle a eu peu d'usages (Kobayashi et al. 1977, Minoli 1978) jusqu'à l'étude des réseaux ATM (Louvion et al. 1988, Gravey et al. 1990, Cost224 1992) – pour ne citer que quelques références.

Dans cet exposé, nous passons en revue les résultats classiques obtenus dans ce domaine, et nous présentons ces phénomènes typiques du

temps discret et en montrons quelques conséquences. Le problème de la MODELISATION d'un système discret sera abordé au travers de l'exemple classique de l'étage de multiplexage ATM, qui donne lieu à une représentation à temps discret complexe.

2. SYSTÈMES À TEMPS DISCRET

La définition d'un système à temps discret n'a probablement pas de sens. Ou plutôt, elle ressort du philosophe qui s'interrogera sur la nature fondamentale de l'Univers (est-il continu ou discret, à la fois en termes de ses particules ou du temps ?). On peut cependant parler de "modèle" à temps discret. Dans une telle approche, on choisit de définir le modèle par une suite de points $(T_n, n = 1, 2, \dots)$ sur l'axe des temps. Les clients qui arrivent ou qui quittent le système, et plus généralement les *événements* concernant le modèle, ne peuvent intervenir qu'aux instant T_n . Pour faire référence à l'ATM, qui motive cet exposé, les points correspondent aux "tops d'horloge" qui marquent les instants d'émission des cellules. Plus généralement, tous les mécanismes *slottés* (Aloha slotté, TDMA, etc.) conduisent à cette représentation.

2.1. Le processus de Bernoulli

On définira ainsi un "processus des arrivées" sur un axe slotté. L'exemple le plus simple en est fourni par le processus de Bernoulli, noté aussi "processus d'arrivées géométrique". A chaque instant T_n , le processus tire au sort selon une probabilité q la possible arrivée d'un client, indépendamment des arrivées précédentes (processus sans mémoire). On peut donner la loi de probabilité du nombre des arrivées pendant une période de n tops:

$$P\{k \text{ arrivées en } n \text{ slots}\} = \binom{n}{k} (1-q)^{n-k} q^k \quad (1)$$

De même, on donnera la loi de l'intervalle entre 2 arrivées successives (d'où dérive – improprement d'ailleurs – le nom de *processus géométrique*).

$$P(k) = q(1-q)^{k-1}$$

L'intervalle moyen a une durée $1/q$. Si l'on définit Δ comme étant la durée du slot, on peut interpréter la formule comme donnant la probabilité d'une durée $T = k\Delta$ entre arrivées. Faisant alors tendre Δ vers 0, de telle sorte que $q/\Delta = \lambda$ reste constant, on voit assez facilement que la loi précédente admet comme limite un processus de Poisson. Le même résultat serait obtenu en partant de la formule 1.

2.2. Un exemple de file à temps discret: le système Géo/Géo/1

De façon identique, le temps de service sera représenté par un nombre entier de slots, de sorte que les sortie de clients se feront sur les tops. Le service le plus simple sera géométrique: la fin d'un service en cours se produit avec une probabilité indépendante de l'âge du service.

Supposons des arrivées de Bernoulli (paramètre q) et des services de loi géométrique (paramètre r): on a défini le système Géo/Géo/1, dont l'analyse sera obtenue facilement, en se plaçant aux tops, "après" les arrivées et départs (instants T_n^+). L'analyse est élémentaire, et se trouve dans tous les bons ouvrages (Hunter 1983, Bruneel et Kim 1993). On a notamment la probabilité stationnaire:

$$P(n) = \rho(1 - \alpha)\alpha^{n-1}$$

$$P(0) = 1 - \rho$$

avec $\rho = q/r$ et $\alpha = \frac{q(1-r)}{r(1-q)}$.

Encore une fois, faisant $q, r \rightarrow 0$, avec q/r constant, l'expression tend vers la probabilité stationnaire de la file M/M/1.

Cette remarque tendrait à faire apparaître l'analyse en temps discret comme un simple raffinement, tout juste utile à augmenter la précision de l'analyse en temps continu.

3. LA SIMULTANÉITÉ

En réalité, le principe de discrétisation introduit une difficulté, qui se manifesterait en particulier si le système étudié est de capacité finie. La modélisation à temps discret introduit la possibilité d'événements simultanés: une arrivée et une fin de service se produisent au même instant T_n . Le modèle à temps continu interdit cette possibilité (l'événement correspondant y a une probabilité "nulle").

Le problème de la simultanité est rarement abordé. Les développements qui suivent sont tirés de (Gravey et Hébuterne 1992). Pour mettre en équations la file à temps discret, il faut lever l'indétermination résultant de la simultanité: les clients en question séjournent-ils ou non ensemble, pour un temps "nul" dans la file ? On convient de distinguer les configurations suivantes.

- L'arrivée se produit "avant" le départ. C'est à dire que, à l'époque T_n (et dans une durée nulle), le processus gérant la file fait entrer le nouvel arrivant puis procède au départ. On a dénommé cette configuration *Arrival First* (AF).

- L'arrivée se produit "après" le départ. C'est la configuration *Departure First* (DF).

Le schéma de la Figure 1 illustre ces deux configurations. Notons T_n , $n \geq 0$, les instants de tops d'horloge. Supposons qu'à un instant T_n une arrivée et un départ se produisent, de sorte qu'il s'agisse d'événements *simultanés*. Dans le cas AF on dira que l'arrivée a lieu à T_n^- et le départ à T_n . Situation inattendue du cas AF, le client arrivant pourra trouver la file non vide mais avoir une attente nulle (le client occupant le serveur quittant le système simultanément). Cette possibilité caractérise la discipline AF. Il faut comprendre la nécessité du distinguo. Soit une file Géo/G/1. Notons N_n^A (resp. N_n^D) le nombre de clients laissés dans la file par le départ à T_n , dans la configuration AF (resp. DF). Evidemment, le départ pourra ne pas être accompagné d'une arrivée. On voit que:

$$N_n^A = N_n^D + \mathbf{1}_{\text{Une arrivée en } T_n}$$

Les événements sont indépendants, la fonction génératrice des distributions de probabilités vérifiera:

$$N^A(z) = N^D(z) \times (1 - q + qz)$$

En particulier, on peut montrer que :

$$P(N^A = 0) = 1 - \rho \quad P(N^D = 0) = \frac{1 - \rho}{1 - q} \tag{2}$$

Dans cette formule, ρ désigne la charge offerte au serveur, produit du paramètre d'arrivée q par le temps de service exprimé en nombre de slots.

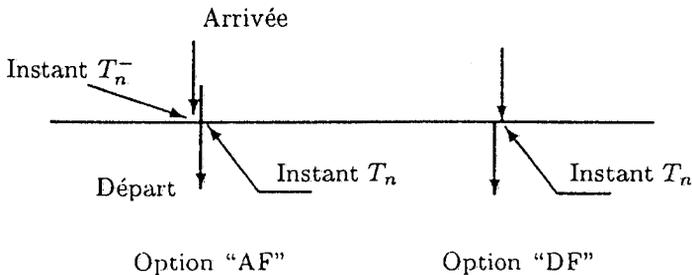


Figure 1. - Événements simultanés: configurations AF et DF.

* Bizarrement, (Hunter 1983), qui note la nuance, appelle *early arrivals* la configuration que nous nommons DF, et *vice versa*.

La forme de N^A est analogue à ce qu'on obtiendrait pour un système $M/G/1$. De façon plus générale, on montre qu'il en est de même pour toute la distribution de probabilité.

Si l'on envisage maintenant le système à capacité finie $Géo/G/1/K$, on se convainc aisément que la configuration AF conduit à une probabilité de rejet plus forte que DF. La mise en équation confirme cette intuition, voir (Gravey et Hébuterne 1992).

Une dernière remarque, quant à la simultanéité. L'analyse des files à arrivées poissonniennes fait un large usage de la propriété PASTA (Poisson Arrivals See Time Averages). Ici, il est possible de démontrer (Gravey et Hébuterne 1992) que la file $Géo/G/1$ avec l'hypothèse AF vérifie la propriété GASTA (*Geometric Arrivals See Time Averages*), mais que l'hypothèse DF viole cette propriété. Cela conduit à voir en la file $Géo/G/1$ sous l'option AF l'analogue à temps discret du système $M/G/1$.

Les effets de la simultanéité auront une grande importance en termes *d'observation* effectuées sur le système: la concordance entre prévisions et observations réclame évidemment un choix cohérent des points de test, dans le modèle et sur le système. Cela nécessite en premier lieu d'avoir mis en lumière ces mécanismes d'événements simultanés.

4. UNE MODÉLISATION DU MULTIPLEXAGE EN ATM

Pour revenir à la technique ATM, le premier usage du modèle sera le dimensionnement d'un multiplexeur. Supposons des arrivées obéissant au processus de Bernoulli. Quel modèle adopter alors, quant au traitement de la simultanéité, $Géo/D/1/K$ AF ou $Géo/D/1/K$ DF ? Pour trancher, il est nécessaire d'analyser en détail le fonctionnement, au niveau matériel, du multiplexeur de base que l'ATM met en œuvre.

4.1. Mécanisme de multiplexage

La discussion qui suit s'appuie sur l'exemple du commutateur du projet Prélude (Devault et al. 1988).

Dans un tel système, M lignes entrantes débitent sur les entrées d'un commutateur, qui aiguille vers la sortie considérée les cellules qui lui sont destinées. La ligne de sortie est équipée d'un tampon (*output queueing*) de taille limitée notée K . Le dimensionnement (quelle valeur pour K ?) constitue la question fondamentale.

Puisque les cellules ont une taille constante, on adopte un service déterministe (Δ désigne le temps d'émission d'une cellule). Les arrivées dans la file résultent de la superposition de M processus discrets, de même

période Δ mais de phases relatives arbitraires et constantes (on néglige pour simplifier les perturbations liées au glissement des différentes bases de temps). En effet, l'arrivée dans la file de sortie d'une cellule en provenance du multiplex entrant m se produit à un instant de la forme $\theta^m + \tau$, où θ^m rend compte du déphasage entre le lien considéré et le multiplex sortant, et où τ est le temps de commutation de la cellule (qu'on peut supposer indépendant du rang m). Une analyse fine des mécanismes en cause a été développée dans (Gounod 1994, Hébuterne et Dupuis 1996).

On représentera le mécanisme des arrivées en divisant le temps-slot du multiplex sortant (temps d'émission de la cellule) en M sous-intervalles, une arrivée pouvant se produire sur chaque sous-intervalle. On a appelé *microslot* chacun de ces points supplémentaires, liés aux références temporelles des liens d'entrée. Supposant pour la simplicité un modèle d'arrivées Bernoulli, le multiplexeur sera représenté par la file Geo/D/1/K.

En réalité, il faut tenir compte du caractère *synchrone*, au niveau cellule, du processus d'émission des cellules. C'est là une propriété très importante de la technique ATM. En bref, si on note t_0 un instant dépendant du multiplexeur, les instants de début d'émission de cellules sont obligatoirement de la forme $t_k = t_0 + k\Delta$. Il en résulte que les cellules entrant à un quelconque des sous-intervalles ne seront pas prises en compte avant le prochain instant t_k de la ligne de sortie. La figure 2 illustre le détail du processus des arrivées sur chaque lien d'entrée et dans la file de sortie (on y a supposé, sans perte de généralité, τ nul – dans le cas général, sa valeur provoque un décalage d'une valeur τ de l'axe du lien sortant). La partie supérieure de la figure représente le processus des arrivées sur 3 multiplex entrants (On suppose $M = 4$, le 4ème multiplex n'étant pas figuré) On remarque la durée constante du slot, pour les 3 multiplex, et également le déphasage, constant, entre chacun d'entre eux.

Le multiplex de sortie, qui fonctionne au même rythme, a cependant son propre synchronisme. De ce fait, les slots du multiplex de sortie (cas a ou b), sont en général sans relation avec les slots des multiplex entrants *même si le calage relatif est fixe*. C'est aux instants de sortie que les cellules peuvent être prises en compte.

Pour décrire le fonctionnement du tampon de sortie, on introduit les *microslots*, instants où les cellules entrent dans le tampon, c'est à dire ici superposition des slots des multiplex entrants. On doit y superposer l'échelle de temps propre au lien de sortie.

Le modèle exact qui résulte de cette remarque est numériquement complexe (chaîne de Markov à $K \times M$ états). Un modèle équivalent peut néanmoins être

construit: les arrivées se produisant sur les sous- intervalles sont regroupées en une arrivée groupée juste avant l'instant de slot, conformément au schéma de la figure2 (b).

Le modèle se notera donc $D^{[x]}/D/1/K$, une arrivée d'un groupe se produisant, de façon déterministe (mais le groupe ayant une taille aléatoire), avant les instants de début de slot, donc selon le mode AF. Avec nos hypothèses, la taille des groupes est distribuée selon une loi binomiale:

$$P(\text{groupe de taille } n) = \binom{M}{n} q^n (1 - q)^{M-n}$$

L'analyse procède par l'écriture de la chaîne de Markov aux instants T_n , qui fournit la probabilité d'occupation stationnaire du tampon. Pour plus de détails sur cette analyse, on consultera (Hébuterne et Dupuis 1996). Le modèle à arrivées groupées est strictement équivalent au modèle complet (du point de vue de la perte de clients), mais d'une complexité beaucoup plus faible (espace des états de dimension K , et non plus $K \times M$).

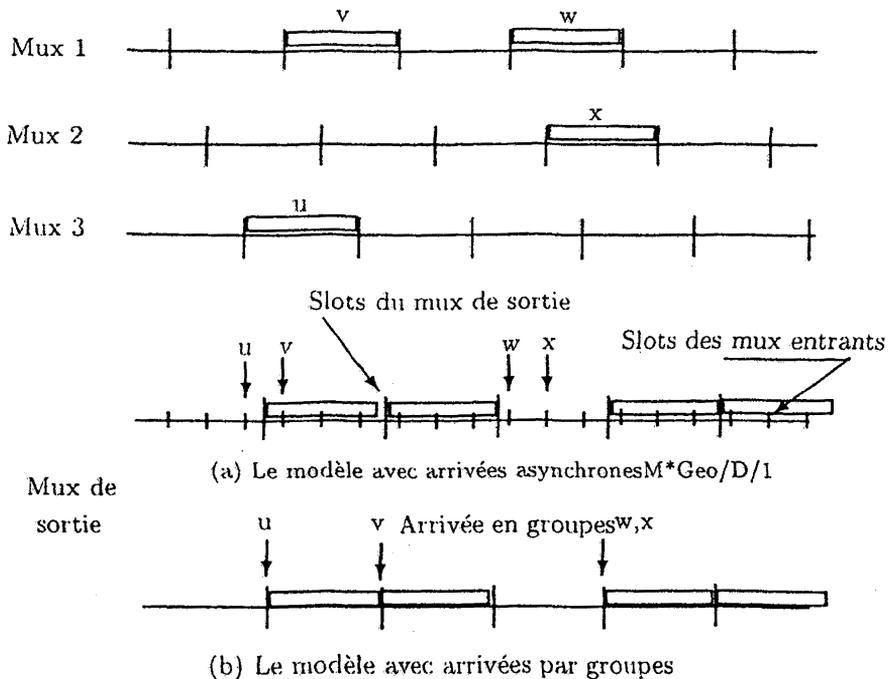


Figure 2. - Mécanisme d'arrivée des cellules au niveau microscopique, et modèle avec arrivées groupées équivalent.

A titre d'illustration des possibilités du modèle, la figure 3 donne les distributions de probabilité des temps d'attente pour les deux modèles. Le modèle avec arrivées groupées ne peut donner que la partie entière de l'attente (il ignore l'attente se produisant entre l'instant d'arrivée et le slot qui suit), ce qui ne constitue pas un réel handicap.

On le voit, le modèle à adopter pour l'étude du dimensionnement d'un multiplexeur ATM est un modèle avec arrivées groupées, sous le mode AF. Cela ne signifie évidemment pas que tous les systèmes seront justiciables du choix AF. Chaque application nécessitera une étude approfondie des mécanismes physiques pour valider l'option AF/DF.

4.2. Des modèles de multiplexage périodiques

Il est possible d'imaginer des modèles dans lesquels l'aspect discret se double d'un comportement périodique particulièrement strict. La technique ATM en a fourni un exemple avec la file $nD/D/1$. Considérons à nouveau notre multiplexeur, en supposant cette fois que les liens entrants portent des sources de trafic de type *Continuous Bit Rate (CBR)*. L'exemple le plus simple provient de la commutation de circuits, où chaque source émet un bloc

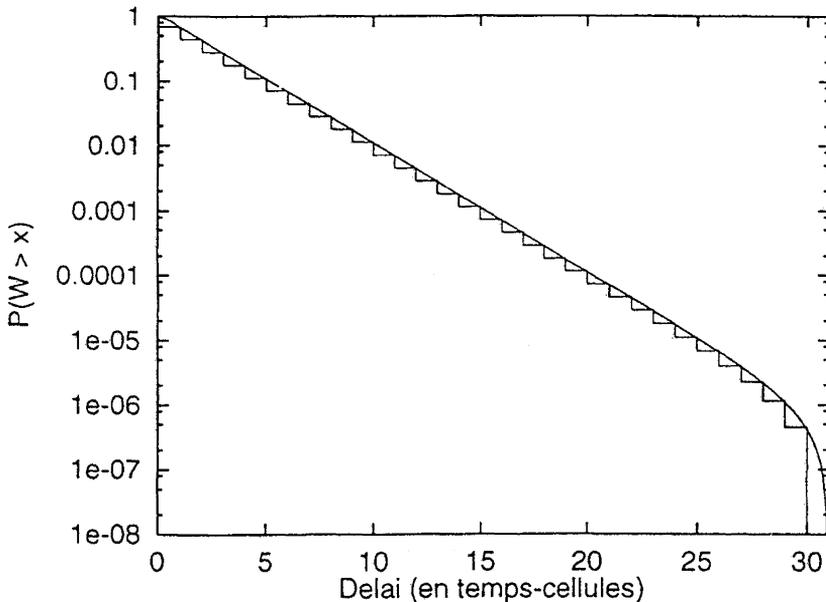


Figure 3. – Distribution comparée des temps d'attente pour les deux modèles. La courbe continue est donné par le modèle microscopique, la courbe en escaliers par le modèle à arrivées groupées.

d'information avec une période T . Les différentes sources n'y ont aucune relation de phase entre elles, et sont multiplexées comme précédemment.

L'analyse statistique de tels systèmes présente des difficultés redoutables. On se trouve en effet en présence de processus qui ont perdu toute propriété d'*ergodicité*: le comportement de l'ensemble reste strictement périodique. Cependant, la valeur des phases de chaque processus est arbitraire, et on peut analyser le phénomène en probabilisant les phases: par exemple, on peut supposer que chaque source démarre à un instant (microslot) uniformément distribué sur l'ensemble des microslots. La référence (Cost224 1992) présente une synthèse des résultats de ce modèle. Il a été généralisé au cas, plus réaliste, où les périodes des sources sont différentes (Virtamo 1990). Là encore, on retrouve la même tendance: si le nombre des sources grandit, le comportement asymptotique de la superposition redevient poissonnien.

5. DES COMPORTEMENTS INÉQUITABLES

5.1. Prise en compte du synchronisme

Revenons au problème de base du dimensionnement d'une file à capacité limitée alimentée par M flux entrants. La modélisation à temps continu, sous les hypothèses élémentaires poissonniennes, conduit à la file $M/G/1/K$, avec pour flux entrant la superposition des flux individuels, qui sont servis de façon identique. C'est à dire que la probabilité de rejet calculée donne la qualité de service éprouvée par chacun des flux entrants.

Considérons à nouveau le système à temps discret, en reprenant pour un moment la représentation microscopique de la figure 2. Notons que *toutes* les cellules provenant d'un lien entrant fixé m arrivent devant la file avec le même décalage. Ainsi, le multiplex de rang 3 offrira toutes ses cellules au 3^e microslot, et donc systématiquement *après* celles des multiplex de rang 1 et 2. Il est facile d'en imaginer les conséquences: supposons que, pendant un intervalle donné la file soit pleine. L'instant T_n suivant voit alors un départ (puisque la durée du slot est le temps de service). Si une cellule arrive alors provenant du multiplex 1, elle pourra toujours pénétrer dans la file. Par contre, notre cellule provenant du multiplex 3 ne sera admise qu'en l'absence d'arrivées sur les liens 1 et 2.

On peut étudier le phénomène à partir du modèle à arrivées groupées, au moyen de l'artifice suivant. Nous cherchons la probabilité de rejet pour les cellules provenant du multiplex m , Π_m . Supposons que le tampon était dans l'état k ($K - k$ places libres) au slot précédent. Une cellule arrivant sera

refusée si $K - k$ cellules ou plus arrivent des $m - 1$ liens qui sont servis avant:

$$\Pi_m = \sum_{k \geq 0} P_k \sum_{j \geq K-k} \binom{m-1}{j} q^j (1-q)^{m-1-j}$$

L'analyse évoquée plus haut a fourni les P_k , probabilités d'occupation stationnaires. La courbe suivante présente un exemple de cet effet: Figure 4. Le système reçoit une charge $\rho = qM = 0.8$, et relie $M = 8$ ou 32 entrées. Les courbes donnent la valeur maximale (Π_M) ainsi que Π_2 (puisque $\Pi_1 = 0$). L'effet est d'autant plus marqué que M augmente, et le rapport Π_M/Π_2 reste à peu près constant. L'analyse fait ressortir l'avantage offert aux multiplex de rang inférieurs. La comparaison des moyennes sur tous les multiplex (non représenté pour clarifier la figure) montre un avantage au cas $M = 8$: de façon générale, un système à temps discret aura un comportement d'autant plus fluide que le pas (la valeur de M) sera faible. Le cas extrême serait celui où $M = 1$, avec évidemment une perte nulle! Lorsque M augmente, le comportement tend vers celui d'une file à temps continu, comme déjà remarqué: la file M/D/1/K constitue une borne pessimiste du multiplexeur de sortie. Cela explique, en particulier, la pente plus forte du

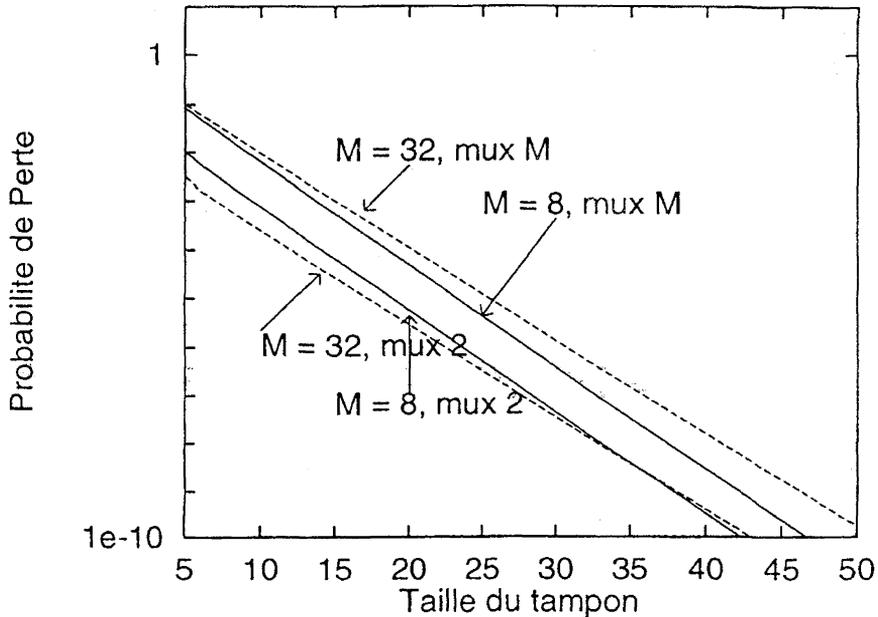


Figure 4. - Comparaison des probabilités de rejet en fonction des rangs des multiplex entrants

cas $M = 8$. La constance du rapport Π_M/Π_2 est, quant à elle, difficile à expliquer.

En réalité, l'analyse d'un système réel doit apporter quelques nuances, par rapport à cet effet brutal qui résulte d'une modélisation à temps discret stricte. On a évoqué plus haut (pour les négliger) les effets de glissement d'horloges. La numérotation des multiplex qu'on a dû introduire n'a rien de canonique, elle résulte en fait du calage des horloges: le multiplex 1 est celui pour lequel l'instant de disponibilité d'une cellule arrivant suit immédiatement le début du slot du multiplex de sortie. A cause du glissement relatif des horloges, la numérotation va évoluer sans cesse, avec un rythme qui dépend de la précision des horloges. La technologie courante offre des précisions dans les gammes $10^{-7} - 10^{-8}$, ce qui correspond à des changements dans l'ordre de numérotation au rythme de la minute.

En conséquence, une connexion établie pour une durée longue pourra mesurer sa QS et lui trouver une valeur nominale, tout en traversant des périodes de rejet significativement plus importants. On le voit, ce phénomène peut amener à reconsidérer la notion même de Qualité de Service.

5.2. L'effet "Conway-Rosenberg"

Un autre comportement inéquitable peut être mis en lumière très simplement. Considérons le modèle du multiplexeur, avec pour simplifier $M = 2$ entrées. Notons q_1 et q_2 les charges offertes par chaque ligne entrante. Observons le système pendant une durée de N slots. On verra arriver Nq_1 (resp. Nq_2) cellules de chaque type, et Nq_1q_2 slots verront l'arrivée simultanée de deux cellules (tous ces comptages en moyenne, évidemment).

Dans le cas où deux cellules arrivent, notons α_i la probabilité que le multiplex i ($i = 1, 2$) soit servi le premier. Si le tampon contient $K - 1$ cellules, c'est alors la probabilité que la cellule de l'autre multiplex soit rejetée. Il y aura ainsi par exemple $Nq_1q_2P_{K-1}\alpha_1$ cellules perdues pour le multiplex 2. En d'autres termes, les probabilités de rejet pour nos deux multiplex s'écriront:

$$\begin{aligned}\Pi_1 &= P_{K-1}q_2\alpha_2 \\ \Pi_2 &= P_{K-1}q_1\alpha_1\end{aligned}\tag{3}$$

(c'est le quotient du nombre moyen de cellules perdues au nombre de cellules offertes). Supposons un traitement du mécanisme d'admission équitable, c'est

à dire $\alpha_1 = \alpha_2 = 1/2$ (les considérations de la Section précédente voudraient que $\alpha = 0, 1$; on fait ici l'hypothèse que l'effet mentionné varie suffisamment vite pour être imperceptible). On a alors la relation suivante:

$$\Pi_1 q_1 = \Pi_2 q_2 \quad (4)$$

C'est un autre comportement inéquitable, avantageant cette fois la source la plus chargée, et qui pourra se conjuguer ou bien combattre l'effet précédent. Là encore, on voit l'impact qu'il aura sur la QS ressentie par les sources.

On peut généraliser à plus de 2 lignes d'entrée. Le phénomène a été étudié dans la référence (Conway et Rosenberg 1996) – d'où le nom qu'on lui a donné. Les auteurs y proposent de neutraliser cet effet par l'introduction d'un mécanisme dit *weighted fair blocking*.

Il faut remarquer que ce phénomène a un aspect positif, à savoir la baisse du taux de rejet global, qui serait donné, ici, par la moyenne pondérée $\Pi = (\Pi_1 q_1 + \Pi_2 q_2)/(q_1 + q_2)$. On se reportera à la référence (Bazanowski et Killat 1995).

6. CONCLUSIONS

Nous avons présenté un panorama de quelques questions soulevées par la mise en œuvre de modèles à temps discret, dans la perspective de l'étude des performances de réseaux ATM. L'étendue du champ d'application de la théorie des files à temps discret dépasse largement le cadre des études ATM. Cependant, l'ATM fournit une illustration privilégiée des conséquences pratiques des phénomènes inhérents au temps discret.

La discussion présentée dans cet exposé s'est focalisée sur les aspects temps discret. On a passé sous silence la possibilité de définir et d'utiliser des modèles (à temps discret) plus sophistiqués. Notamment, le comportement des sources à trafic très variable (sources dites *bursty*) a entraîné l'introduction de modèles d'arrivées de type *Markov modulated Bernoulli Processes (MMBP)* – voir par exemple (Kouvatsos et Fretwell 1995, Cost242 1996). Ces modèles complexifient l'analyse mathématique, sans toutefois remettre en cause l'argumentation qui nous a guidés.

Il faut revenir sur une distinction esquissée en début de l'exposé. Ou bien le modélisateur se trouve en présence d'un système à temps discret, ou bien la modélisation consiste à faire un modèle à temps discret d'un système par essence à temps continu. L'ATM offre un exemple de système intrinsèquement discret. La description du multiplexage en montre les différents aspects : comportement périodique et synchrone des éléments

d'entrée et sortie, accompagné dans le même temps d'une absence de relation de phase entre les entrées et la sortie (le vocabulaire est source de confusion, on parle à la fois de technique asynchrone, et de synchronisme au niveau cellule).

Lorsqu'on se trouve en présence d'un système de ce type, la description même simplifiée doit absolument intégrer les mécanismes d'ordonnement qui accompagnent le synchronisme. Le modèle, quant à lui, doit évidemment se conformer à ces contraintes en les incorporant. Par exemple, on devra spécifier la règle "Arrival First" ou "Departure First".

Ensuite, l'étude des performances du système devra s'attacher à débusquer les phénomènes liés au synchronisme. On en a décrit quelques-uns ici, il s'agit de comportements inévitables engendrés par la périodicité. La discussion de ces effets montre la complexité de leur prise en compte : même si le système appelle un modèle à temps discret, il est probable que les horloges mises en jeu, qui gouvernent la périodicité, amèneront une variation lente de sorte qu'à long terme l'effet de périodicité pourra passer inaperçu. Faut-il alors le prendre en compte, dans la mesure de la QS ? La réponse dépasse le cadre de la théorie; en tout état de cause, elle dépend évidemment du contexte, et des ordres de grandeur.

La discussion de l'analyse du système déterministe $nD/D/1$ conduirait aux mêmes conclusions: les phases relatives des différentes sources, liées aux horloges des équipements, n'ont entre elles aucune relation fixe à long terme, de sorte que l'observation montrera un calage évoluant continûment.

La situation sera quelque peu différente lorsqu'on fera, par commodité, un modèle à temps discret d'un système originellement à temps continu. Cette fois, la modélisation court le risque d'introduire de façon artificielle des phénomènes, du type synchronisation, que l'on a passés en revue ici, mais que le système ignore. Notamment, une discrétisation introduit *de facto* l'occurrence d'événements simultanés parasites, qui peuvent entraîner un biais dans les prévisions qu'on tirera du modèle.

En résumé, les systèmes tout comme les modèles à temps discret présentent, par rapport à leurs homologues à temps continu, des différences qui vont plus loin que les simples raffinements numériques. On doit s'attendre à observer dans de tels systèmes des comportements spécifiques, dont la conséquence pratique pourra être la révision des notions courantes de qualité de service.

Cet exposé vise à éveiller l'intérêt pour la théorie des files à temps discret. Très probablement, ce tour d'horizon n'a pas épuisé l'inventaire des particularités de cette discipline. Une théorie générale des files à temps

discret reste à élaborer, et c'est le souhait de l'auteur que d'avoir suscité des vocations dans ce domaine.

BIBLIOGRAPHIE

- Z. BAZANOWSKI et U. KILLAT, *The superposition of cell streams with geometrically distributed interarrivals in an ATM multiplexer*. IEEE Trans. Com. **42**, février 1995.
- H. BRUNEEL et B. KIM, *Discrete time Models for Communication Systems Including ATM*. Kluwer Academic Publishers, 1993.
- A. E. CONWAY et C. ROSENBERG, *Weighted fair blocking mechanisms for discrete-time multiplexing*. IEEE Infocom'96. Session 8.D. San Francisco, Mars 1996.
- COST 224, *Final Report. Performance evaluation and design of multiservices networks*. J.W. Roberts (editor). Commission of the European Communities, 1992.
- COST 242, *Broadband Network Teletraffic, final Report of Action COST 242* J.W. Roberts, U. Mocci, J. Virtamo (editors). Lecture Notes in Computer Science, 1155. Springer, 1996.
- M. DEVAULT, M. SERVEL et J-Y Cochenneec, *The Prelude experiment: assesments and future prospects*. IEEE JSAC, **6**, décembre 1988.
- P. E. GOUNOD, *Modèles d'évaluation du multiplexage ATM*. Rapport Technique CNET NT/LAA/EIA/15, avril 1994.
- A. GRAVEY, J.R. LOUVION et P. BOYER, *On the Geo/D/1 and Geo/D/1/n queues*. Performance Evaluation, **11**, 1990, p. 117-125.
- A. GRAVEY et G. HÉBUTERNE, *Simultaneity in discrete-time single server queue with Bernoulli inputs*. Performance Evaluation, **14**, p. 123-131, 1992.
- G. HÉBUTERNE et A. DUPUIS, *Microscopic models of ATM multiplexing*. 4th IFIP workshop on performance modelling and evaluation of ATM networks, Illkley, Grande Bretagne, Juillet 1996.
- J. J. HUNTER, *Mathematical Techniques of Applied Probability. Vol 2: discrete time models*. Academic Press, 1983.
- D. KOUVATOS et R. FRETWELL, *Closed form performance distribution of a discrete time GIg/D/1/N queue with correlated traffic*. Sixth IFIP WG6.3 Conference on Computer Networks, Istanbul, Turkey, octobre 1995.
- L. KLEINROCK, *Queueing Systems*. 2 volumes, J. Wiley, 1976.
- H. KOBAYASHI et A. KONHEIM, *Queueing Models for computer communications system analysis*. IEEE Trans. Comm., **25**, pp 2-29, 1977.
- J. R. LOUVION, P. BOYER et A. GRAVEY, *A discrete-time single server queue with Bernoulli arrivals and constant service time*. ITC-12, Turin, avril 1988.
- D. MINOLI, *General geometric arrival-constant server queue with application to packetized voice*. ICC'78 papier 36.6.
- J. T. VIRTAMO, *An exact analysis of the $D_1 + \dots + D_k/D/1$ queue*. 7th ITC Seminar, Morristown, New Jersey, papier 11.1, 1990.