DE L'OPTIMUM ECONOMIQUE DANS LES PROBLEMES DE FILES D'ATTENTE (*)

par J. TEGHEM
Université Libre de Bruxelles

1. On sait la grande variété de problèmes pratiques qui ressortissent à la théorie des files d'attente : problèmes de téléphonie, de trafic, de décodage de messages, d'exécution de commandes, de réparation de machines et autres problèmes d'organisation de centres de service divers, salle d'attente, centre de production ou de vente, port, aérodrome, etc. « Waiting-line theory is a management tool » [1].

Les modèles correspondants, quoique fort divers en nature ou en complexité, ont ceci de commun que, dans chacun d'eux, des unités (on dit souvent « clients ») — appels, véhicules, messages, commandes, machines en panne, personnes, navires, avions, etc. — se présentent à un centre de service constitué d'un certain nombre de « stations » (ou « guichets ») et y forment éventuellement une ou des files d'attente. Nombreuses et diverses sont les hypothèses devant être prises en considération pour répondre aux caractéristiques variées des modèles, quant au nombre de catégories de clients, à la composition du centre de service, au nombre et à la disposition des stations, aux disciplines de file, aux modes de gestion du centre et aux lois de probabilité des arrivées de clients et des durées de service. Aussi le nombre de publications consacrées aux problèmes de files d'attente est-il considérable et continue-t-il à s'accroître à une cadence fort rapide.

Les questions posées sont d'ailleurs diverses, elles aussi ; elles concernent ou le nombre de clients présents dans le système, ou le temps d'attente de ces clients, ou le temps d'inactivité des stations, ou le nombre de stations, etc., ou plusieurs de ces grandeurs à la fois, de par leur mutuelle dépendance ou de par la mesure dans laquelle il est possible d'agir sur elles par le choix d'un paramètre ou d'un mode d'organisation.

On peut distinguer, dans l'esprit du chercheur qui se penche sur l'étude de ces problèmes, deux points de vue : celui du mathématicien, plus particulièrement celui du probabiliste, attiré par la grande variété et la difficulté

(*) Conférence faite à la tribune de la SOGESCI le 7 décembre 1964.

analytique des questions, et le point de vue du chercheur « opérationnel », nécessairement préoccupé des aspects économiques de l'étude et de la recherche d'une organisation optimale du système considéré, que celui-ci soit industriel, commercial ou administratif.

Sans doute ces deux points de vue occupent-ils souvent, ou même le plus souvent, simultanément, comme il se doit, l'esprit du chercheur, mais on peut dire que c'est généralement alors, d'après sa formation et le caractère de ses préoccupations dominantes, dans des proportions bien différentes, l'un des aspects prenant habituellement nettement le pas sur l'autre.

Les problèmes de files d'attente, avec les problèmes de gestion de stocks, se prêtent d'ailleurs, mieux sans doute que tous autres problèmes de recherche opérationnelle, à cette distinction

Alors, en effet, que dans les problèmes de recherche opérationnelle qui ressortissent à la programmation proprement dite — qu'il s'agisse de programmation linéaire ou non linéaire ou de programmation dynamique — la fonction économique est intégrée dans les questions analytiques dès le départ des recherches et s'y impose de manière permanente en tant que centre même des préoccupations, dans les problèmes de files d'attente, la recherche analytique se concentre généralement sur la découverte d'une distribution de probabilités, transitoire ou stationnaire, soit du nombre d'unités présentes dans le système, soit du temps d'inactivité des stations, soit de toute autre grandeur susceptible de mériter l'attention, ou, plus modestement, sur la découverte de certains paramètres de ces distributions, la moyenne le plus souvent. Et ce n'est qu'ensuite, se superposant à cette recherche analytique de nature descriptive, qu'est traité, quand il n'est pas, comme le plus souvent, omis, le problème d'optimisation, qui porte sur une fonction du temps ou sur une fonction de coûts, fonction dans laquelle interviennent notamment des coûts d'attente, d'inactivité, de transition d'un type d'activité à un autre, etc.

Cette dichotomie n'est pas seulement commode pour le chercheur « analytique », dont les recherches, débarrassées du souci de considérations économiques permanentes, peuvent alors, avec beaucoup d'audace, être dirigées vers des modèles de plus en plus complexes — embrassant de mieux en mieux les réalités multiples des problèmes — et des questions analytiques de plus en plus difficiles, elle est aussi la seule manière vraiment pratique d'envisager les choses dès le moment où la complexité des hypothèses mathématiques de base ou des composantes économiques des problèmes dépasse certaines limites.

On peut d'ailleurs, à bien des égards, juger de l'efficience d'un système ou des modifications qu'il serait utile d'apporter à son organisation par l'examen des prédictions que fournit l'appareil analytique quant à l'évolution de l'état du système. Les résultats analytiques permettront par exemple, avant toute considération d'ordre économique, de voir qu'une organisation donne lieu à trop d'attente ou à de trop longues périodes d'inactivité des stations, et de déterminer dans quelle mesure il peut être remédié à cet état de fait par des modifications de la durée des services ou du nombre des stations.

Il convient de tenir compte, en outre, de ce que les paramètres de nature économique, tels que des coûts, sont quelquefois difficilement chiffrables. Et enfin, et ceci n'est pas sans rapport avec cela, qu'à un même modèle mathématique peuvent s'adapter des hypothèses économiques diverses. L'étude descriptive du modèle fournit alors un support commun à l'examen des conséquences de ces diverses hypothèses. Ce dernier examen peut se faire lui-même soit mathématiquement, soit numériquement. De tout quoi résultera *in fine* le meilleur choix, ou, tout au moins, un choix intéressant des paramètres d'organisation du système.

Souvent aussi, devant la complexité d'un problème, a-t-on recours, dès le départ de l'étude, à une technique de recherche par simulation. Je ne traiterai pas ici de cette méthode, malgré son utilité et le très grand intérêt qu'elle peut présenter.

Je me résume par l'affirmation de la nécessité, ou de l'utilité même, dans la plupart des cas, d'une dichotomie dans le traitement des problèmes de files d'attente : 1. *étude mathématique* ayant pour objet la description des conséquences de l'adoption d'un modèle d'organisation, cette description constituant ensuite, grâce à des possibilités de prédiction, la base de comparaisons utiles entre des modèles divers ou affectés de paramètres différents; 2. *étude économique,* qui sera, selon les cas, soit analytique, soit numérique, soit simplement constituée des comparaisons auxquelles il vient d'être fait allusion.

La nécessité, voire l'utilité, inscrite dans la nature même de la plupart des problèmes de files d'attente, de séparer l'économique du mathématique proprement dit, explique que les très nombreux ouvrages, mémoires et articles relatifs à ces problèmes sont consacrés uniquement, dans leur grande majorité, à l'aspect mathématique des études.

Et les exceptions, qui deviennent à vrai dire plus nombreuses, ne concernent, à peu de chose près, que des cas élémentaires, l'étude économique se superposant alors simplement à l'étude mathématique, sans s'y intégrer.

Le livre sur les phénomènes d'attente, de Kaufmann et Cruon [2], ne consacre que peu de pages aux problèmes analytiques d'optimisation. Les auteurs soumettent, par contre, plusieurs problèmes concrets aux techniques de simulation. Saaty, dans son important livre sur les « Queues » [3], fait allusion aux problèmes économiques dans l'introduction, pour ne plus du tout en parler ensuite. Morse, dans son ouvrage « Queues, Inventories and Maintenance » [4], ne traite de l'aspect économique qu'à l'occasion de quelques problèmes particuliers relatifs aux modèles les plus simples. Cox et Smith, dans leur livre « Queues » [5], mettent aussi l'accent, dès l'introduction, sur l'aspect mathématique descriptif des études de files d'attente, tout en insistant sur le fait qu'il doit servir de base à la recherche des modifications les plus dignes d'intérêt de l'organisation du système. On peut citer enfin quelques articles récents dans le « Operations Research » et le « Management Science ». Krishnamoorti [6] cite deux exemples intéressants de problèmes économiques simples — ayant trait au choix d'une discipline de file optimale en face de deux stations opérant en parallèle, à des vitesses différentes — mais après avoir fait l'étude mathématique descriptive, il laisse au lecteur le soin de poursuivre. Les hypothèses qu'il adopte rendent, il est vrai, cette tâche aisée. Hillier [7] traite de problèmes économiques industriels où l'on détermine, sur la base de coûts, soit un nombre optimal de stations, soit un couple optimal nombre de stations - taux d'arrivées des unités, soit un couple optimal nombre de stations - vitesse de service. Je dirai encore, par parenthèse, qu'on trouve, dans l'article de Hillier, un bel exemple du danger qu'il peut y avoir à vouloir traiter de problèmes réels à partir de modèles trop simples.

Je conclurai cette introduction en disant que la théorie probabiliste des files d'attente se situe en fait, dans la plupart des modèles dont elle traite, en marge de la recherche opérationnelle proprement dite, si on considère celle-ci comme inséparable d'un objectif d'optimisation. Elle appartient à la recherche opérationnelle, sans doute, par les outils efficaces qu'elle lui apporte, mais elle ne s'y intègre pas de plain-pied.

C'est la raison pour laquelle je citerai, dans les paragraphes suivants, presque à titre d'exceptions, quelques résultats analytiques de recherche d'un optimum, dus à Naor et à certains de ses collaborateurs de l'Institut de Technologie d'Israël, à Haïfa. Ces résultats s'inscrivent dans tout un ensemble de recherches de l'école de Naor. Ils concernent des cas où les coûts sont supportés par la même firme ou administration ou organisation, laquelle a donc intérêt à en minimiser la somme.

Les résultats que je citerai dans le paragraphe 2 ont trait à des questions de priorité, les modèles étudiés ayant ceci de particulier que l'organisation du système, d'une manière plus précise l'attribution des priorités, n'y est pas fixée à priori, mais est basée partiellement sur l'effet, étudié analytiquement, de certains paramètres économiques, de telle sorte qu'on peut parler, à propos de ces problèmes — et c'est ce qui est en contraste avec la manière habituelle — d'un *traitement analytique unifié,* où l'aspect économique s'intègre véritablement dans l'étude mathématique, et ne se superpose pas seulement à elle.

Dans le paragraphe 3, il sera traité d'un modèle mathématico-économique pour lequel la question se pose de rechercher s'il ne peut être remédié à la perte économique résultant du temps d'inactivité d'une station de service, par la fermeture momentanée de celle-ci, l'opération de fermeture, ainsi que la réouverture de la station, impliquant des coûts et exigeant qu'un certain temps y soit consacré. Ce problème peut être illustré par des exemples de collecte et transmission d'informations, de surveillance de machines ou d'exécution de commandes.

Il est à noter aussi qu'il est assimilable à un problème de priorités, une période de fermeture ou d'opération de réouverture pouvant, en effet, être considérée, du point de vue des clients, comme une période au cours de laquelle la station se consacre à d'autres tâches que celle de les servir ou de se tenir prête à les servir.

La dichotomie habituelle, description mathématique-recherche économique de l'optimum, celle-ci faisant suite à la première et s'y superposant, réapparaît dans le traitement de ce modèle, dont l'étude peut donc être considérée plutôt, à cet égard, comme une illustration de la manière générale et être opposée à celle des problèmes du paragraphe 2.

On constatera enfin que certains aspects des modèles traités mériteraient d'être généralisés, et que, plus particulièrement dans le paragraphe 2, il serait utile de prendre en considération des coûts et même des temps de passage de l'occupation de la station par un client d'une catégorie à l'occupation par un client d'une autre catégorie. Ces coûts et temps de « switch-over » seraient les analogues des coûts et temps de fermeture et de réouverture qui ont été intégrés dans le modèle du paragraphe 3.

2.1. Le premier modèle dont il va être question a été considéré par Brosh et Naor dans une communication qu'ils ont faite au mois d'août 1963, à la 34^{me} Session de l'Institut international de Statistique, à Ottawa. L'étude a paru dans les Actes de ce Congrès [8].

2.1.1. *Description du modèle et énoncé du problème.*

Soit un centre de service à une seule station, et soient P catégories de clients, la catégorie 1, ..., la catégorie P, aucune priorité n'étant encore fixée, et les indices 1, ..., P étant donc simplement destinés à distinguer les catégories. Les flux d'arrivées de ces clients sont poissonniens homogènes, de paramètres respectifs $\lambda_1, \ldots, \lambda_P$, c'est-à-dire que la probabilité de l'arrivée en un temps t de j clients de la catégorie i est

$$e^{-\lambda_i t} \frac{(\lambda_i t)^j}{j\,!} .$$

Les durées de service t_i ($i = 1, \ldots, P$) ont, elles, des distributions quelconques, pouvant être différentes d'une catégorie à l'autre; on supposera seulement l'existence des moments des deux premiers ordres $E(t_i)$ et $E(t_i^2)$. On fait encore l'habituelle hypothèse d'indépendance stochastique entre les durées de service, ainsi qu'entre ces durées et les intervalles de temps séparant les arrivées.

Soit c_i ce que coûte l'attente pendant une unité de temps d'un client de la catégorie i. Aucun autre coût n'est pris en considération et le coût total est supposé linéaire en les nombres moyens de clients des diverses catégories se trouvant dans le système.

Le problème est d'établir, parmi les P catégories de clients, un ordre de priorité *non absolue,* qui minimise le coût total en régime stationnaire, c'est-à-dire en régime d'équilibre statistique.

On suppose que la condition de l'existence de ce régime est satisfaite; elle sera explicitée plus bas [en 2.1.2, formule (2)].

Rappelons enfin que l'on entend par priorité non absolue ou non préemptive ou relative une priorité qui ne s'exerce que par rapport aux clients en attente d'être servis. Elle ne s'exerce pas par rapport aux clients dont le service est en cours, de sorte qu'aucun service en cours n'est jamais interrompu. Un client prioritaire bénéficie seulement de l'avantage de prendre place dans la file d'attente devant tous les clients non prioritaires ou de priorité moindre. D'où aussi le nom de « head-of-the-line-discipline ».

2.1.2. *Résolution.*

Soit b_i la fraction du temps pendant laquelle la station sert des clients de la catégorie i ($i = 1, \ldots, P$). Il est bien connu que

$$b_i = \lambda_i E(t_i) \tag{1}$$

(la fraction de temps b_i est égale au produit du nombre moyen d'arrivées de clients de la catégorie i pendant une unité de temps et de la durée moyenne des services de ces mêmes clients).

Soit $b = \sum_i b_i$ la fraction du temps pendant laquelle la station est occupée. On sait que la condition de l'existence d'un régime stationnaire est la non-saturation de la station, c'est-à-dire

$$b < 1 ,$$

ou encore, grâce à (1),

$$\sum_i \lambda_i \, E(t_i) < 1 . \tag{2}$$

Soient encore, mais en supposant cette fois un ordre de priorité fixé, à un indice plus petit correspondant une priorité plus forte,

$$b_{(p)} = \sum_{i=1}^{p} b_i \qquad (p = 1, \ldots, P) ,$$

q_p le nombre moyen de clients de la catégorie p se trouvant dans le système, et w_p le nombre moyen de clients de la catégorie p se trouvant dans la file d'attente proprement dite (station exclue).

Le problème est de fixer l'ordre des priorités de manière à minimiser $\sum_i c_i \, q_i$ ou, ce qui revient au même, puisque $w_i = q_i - b_i$ et que $\sum_i c_i \, b_i$ est une constante, de manière à minimiser

$$\sum_i c_i \, w_i \quad . \tag{3}$$

Or, il découle des résultats que Cobham a établis en 1954, dans son étude des priorités non absolues [9], que

$$w_p = \lambda_p \frac{\sum_i \lambda_i \, E(t_i^2)}{2\,(1 - b_{(p-1)})\,(1 - b_{(p)})} \tag{4}$$

(cette formule a été démontrée aussi, de manières différentes, par Miller [10] et Cox et Smith [5]).

Il résulte de (4) que la permutation de deux catégories p_1 et p_2 ($p_1 < p_2$) dans l'ordre des priorités n'affecte en rien les nombres moyens de clients des catégories d'indices inférieurs à p_1 ou supérieurs à p_2; sont seuls modifiés les w_i des classes intermédiaires ($p_1 \leqslant i \leqslant p_2$). En particulier, si $p_2 = p_1 + 1$, la permutation n'affecte que les nombres moyens des seules deux catégories permutées.

Cela étant, comparons les coûts correspondant aux deux classements

$$(p)rs = \underbrace{\;.\;\;.\;\;.\;\;.\;\;.\;}_{p}\; rs \;\underbrace{\;.\;\;.\;\;.\;\;.\;\;.\;}_{P - p - 2}\,.$$

$$(p)sr = \underbrace{\;.\;\;.\;\;.\;\;.\;\;.\;}_{p}\; sr \;\underbrace{\;.\;\;.\;\;.\;\;.\;\;.\;}_{P - p - 2}\,,$$

dont le second dérive du premier par la permutation des deux catégories r et s.

Les fonctions de coût (3) correspondantes sont

$$L\{(p)rs\} = \sum_{i \neq r,s} c_i\, w_i + c_r\, w_r\{(p)rs\} + c_s\, w_s\{(p)rs\}$$

et

$$L\{(p)sr\} = \sum_{i \neq r,s} c_i\, w_i + c_r\, w_r\{(p)sr\} + c_s\, w_s\{(p)sr\}\,,$$

où les w_i ($i \neq r, s$), et donc aussi la somme $\sum\limits_{i \neq r,s} c_i\, w_i$, ne dépendent pas de l'ordre choisi.

La comparaison des deux fonctions de coût revient ainsi, grâce à la soustraction de termes communs, à celle des deux expressions

$$c_r\,[w_r\{(p)rs\} - w_r\{(p)sr\}]$$

et

$$c_s\,[w_s\{(p)sr\} - w_s\{(p)rs\}]$$

ou encore, si l'on tient compte de (4) et divise par la constante positive $\frac{1}{2} \sum\limits_i \lambda_i\, E(t_i^2)$, à celle de

$$c_r\, \lambda_r \left[\frac{1}{(1 - b_{(p)})\,(1 - b_{(p)} - b_r)} - \frac{1}{(1 - b_{(p)} - b_s)\,(1 - b_{(p)} - b_r - b_s)}\right]$$

et

$$c_s\, \lambda_s \left[\frac{1}{(1 - b_{(p)})\,(1 - b_{(p)} - b_s)} - \frac{1}{(1 - b_{(p)} - b_r)\,(1 - b_{(p)} - b_r - b_s)}\right].$$

Une multiplication par le produit positif

$$(1 - b_{(p)})\,(1 - b_{(p)} - b_r)\,(1 - b_{(p)} - b_s)\,(1 - b_{(p)} - b_r - b_s)$$

et des modifications simples des expressions résultantes conduisent enfin à la comparaison équivalente de

$$- c_r\, \lambda_r\, b_s\,(2 - 2\, b_{(p)} - b_r - b_s)$$

et

$$- c_s\, \lambda_s\, b_r\,(2 - 2\, b_{(p)} - b_r - b_s)\,,$$

c'est-à-dire à celle des quantités

$$-\ c_r\ \lambda_r/b_r \qquad \text{et} \qquad -\ c_s\ \lambda_s/b_s\ ,$$

égales respectivement, grâce à (1), à

$$-\ c_r/\mathrm{E}(t_r) \qquad \text{et} \quad -\ c_s/\mathrm{E}(t_s)\ .$$

Il résulte de ces considérations simples que le premier classement est préférable au second lorsque

$$c_r/\mathrm{E}(t_r)\ >\ c_s/\mathrm{E}(t_s)\ ,$$

et, partant, que l'ordre optimal des priorités correspond à un classement par ordre décroissant des $c_i/\mathrm{E}(t_i)$, la première priorité étant attribuée à la catégorie pour laquelle le quotient $c_i/\mathrm{E}(t_i)$ est le plus grand; ce qui précède montre, en effet, que tout autre classement peut être amélioré par la permutation de deux classes voisines, pour lesquelles cet ordre n'est pas respecté.

2.1.3. *Commentaires et remarques.*

On constate donc que seul avait été décidé à priori le principe de la priorité non absolue, mais que c'est sur la base d'une étude analytique unifiée, dans laquelle les coûts ont joué un rôle primordial, qu'a été arrêtée l'attribution des priorités. La partie économique de la résolution s'est fondue analytiquement dans la partie mathématique, qui est sortie ainsi de son rôle purement descriptif pour contribuer à fixer l'organisation même du système.

Le fait qu'un accroissement de c_i et une diminution de $\mathrm{E}(t_i)$ tendent à faire élever la priorité ne peut évidemment surprendre, mais on ne pouvait prévoir une expression aussi simple de la règle, ni sans doute une aussi grande simplicité dans la résolution, qui est constamment élémentaire, si l'on part de (4). Enfin, ni les $\mathrm{E}(t_i^2)$, dont il a fallu supposer l'existence, ni les λ_i n'interviennent dans la règle. Ceci est sans doute ce qu'il y a de plus surprenant dans le résultat. Il convient cependant de ne pas perdre de vue que les λ_i sont soumis à la condition (2).

Cette règle d'attribution des priorités d'après les valeurs des $c_i/\mathrm{E}(t_i)$ avait été trouvée déjà par Cox et Smith [5], mais Naor et Brosh ont dégagé parfaitement la raison analytique de ses possibilité et facilité d'obtention, et, partant, une condition suffisante pour que d'autres modèles de priorités puissent être traités d'une manière analogue. Cette raison réside simplement dans le fait que la comparaison des fonctions de coût $\mathrm{L}\ \{(p)rs\}$ et $\mathrm{L}\ \{(p)sr\}$ a pu, par des opérations élémentaires, être rendue équivalente à la comparaison de deux expressions $c_r \psi_r$ et $c_s \psi_s$ ne dépendant respectivement que de r et de s. L'ordre optimal des priorités correspond à un classement par ordre de grandeur des nombres $c_i \psi_i$.

2.2. Arrêtons-nous maintenant, avec Brosh et Naor, au régime de la priorité *absolue* ou préemptive, c'est-à-dire s'exerçant même par rapport à un client dont le service est en cours. Ce client doit céder sa place à un client qui lui est prioritaire, dès l'arrivée de celui-ci. Il retourne à la file d'attente, en y prenant cependant la première place, quitte à devoir céder éventuellement celle-ci au moment de l'arrivée d'un autre client, soit au client arrivant, soit au client dont l'arrivée a interrompu son service, selon les degrés de priorité de ces deux derniers. Nous supposerons en outre, en ce qui concerne la reprise des services ayant été interrompus, qu'ils recommencent au point d'interruption : la partie du service déjà accomplie reste valable (« resume situation »).

On peut tenter de traiter ce cas de la manière dont on a traité le précédent, en se servant cette fois, au lieu de (4), d'une formule pour w_p, qui concerne le cas de ce régime de priorité absolue. — Une telle formule a été établie de diverses manières, par Miller [10], Avi-Itzhak et Naor [11], et Naor et Brosh [8]. — Le résultat est que la règle de priorité basée sur le classement des $c_i/\mathrm{E}(t_i)$ ne reste valable que dans le cas particulier où les durées de service ont des distributions exponentielles négatives, c'est-à-dire telles que

$$\text{Prob. } (t_i \leqslant t) = 1 - e^{-\mu_i t}, \quad \mu_i > 0,$$

d'où

$$\mathrm{E}(t_i) = 1/\mu_i \quad \text{et} \quad \mathrm{E}(t_i^2) = 2/\mu_i^2 .$$

Il est aisé de comprendre pourquoi, et même de prévoir, pour le cas général de distributions de durées de service quelconques, un régime de priorité intermédiaire entre le régime de la priorité non absolue et celui de la priorité absolue, qui, tout en étant simple, règle l'attribution des priorités de manière optimale.

La raison du résultat réside dans ce qui différencie les régimes de priorité relative et de priorité absolue en ce qui concerne les moments d'exécution des décisions d'attribution de la plus haute priorité, c'est-à-dire en réalité les moments où il suffirait que soient prises — instantanément — ces décisions. Dans le premier cas, ces moments ne se situent qu'aux fins de service, tandis que dans le second les décisions sont susceptibles de devoir être prises à tout instant des services, de par l'arrivée d'un nouveau client. Or, une décision prise en un pareil instant quelconque et qui ne tiendrait compte que des seuls c_i et $E(t_i)$ ignorerait une part importante de l'information normalement disponible à cet instant quant à l'état du système, à savoir la durée déjà atteinte par le service en cours. On peut donc imaginer et même prévoir que la règle la meilleure sera celle qui, à l'arrivée d'un client, n'attribue la priorité à celui-ci que si le rapport $c_j/E(t_j)$ qui lui correspond dépasse le rapport $c_i/E(\tau_i | T_i)$ qui correspond au client dont le service est en cours, $E(\tau_i | T_i)$ désignant l'espérance mathématique conditionnelle de la durée *restante* τ_i du service en cours, étant donné que la durée déjà atteinte est T_i. On a $\tau_i + T_i = t_i$ et, en général, $E(\tau_i | T_i) < E(t_i)$. Cependant, et ceci achève l'explication des résultats enregistrés, dans le cas particulier des distributions exponentielles négatives, on sait que la distribution de τ_i est identique à celle de t_i — propriété d'oubli — et donc que $E(\tau_i | T_i) = E(t_i)$. La règle optimale revient ainsi, dans ce cas, à celle obtenue pour le cas des priorités relatives, la connaissance de T_i n'apportant aucune information supplémentaire utile.

2.3. Avi-Itzhak, Brosh et Naor [12] ont donné à un tel régime intermédiaire, où, en général, les risques d'interruption d'un service sont d'autant moindres que le service est en cours depuis plus longtemps, le nom de régime de priorité *discrétionnaire*. Discrétionnaire, car on peut considérer que les décisions d'attribution des priorités, plutôt que de relever de l'un des régimes extrêmes et catégoriques de priorité absolue ou non absolue, sont laissées à la discrétion du « gestionnaire » du système, afin qu'il puisse, sur la base des informations qu'il possède, choisir la solution économiquement la meilleure.

Ces auteurs ont fait l'étude du cas particulier de deux catégories de clients, à durées de service constantes (mais pouvant être différentes d'une catégorie à l'autre), et ce dans le cas déjà signalé en 2.2, où un service interrompu est repris au point d'interruption (« resume situation »), comme aussi dans le cas où il doit être repris entièrement (« repeat situation »). Dans la « resume situation », le résultat obtenu confirme les prévisions émises en 2.2.

2.3.1. *Premier cas* (« resume situation »).

Désignons par λ_1 et λ_2 les taux d'arrivées et par S_1 et S_2 les durées de service, en supposant $\lambda_1 S_1 + \lambda_2 S_2 < 1$ (condition d'existence d'une distribution stationnaire) et $c_1/S_1 > c_2/S_2$. Le problème est de rechercher la valeur optimale du paramètre d'organisation Φ tel que si un client de la catégorie 1 se présente au centre pendant le service d'un client de la catégorie 2, ce service est interrompu ou poursuivi selon qu'il est en cours depuis un temps $< \Phi S_2$ ou depuis un temps $\geqslant \Phi S_2$, $\Phi = 0$ et $\Phi = 1$ correspondant respectivement à la priorité « head-of-the-line » et à la priorité absolue. La valeur optimale est $1 - (c_2/c_1) \cdot (S_1/S_2)$ (qui, grâce à l'hypothèse $c_1/S_1 > c_2/S_2$, est bien, comme il se doit, comprise entre 0 et 1). Le service d'un client de la catégorie 2 est donc interrompu s'il est en cours depuis un temps $T_2 < S_2 - (c_2/c_1)\, S_2$, c'est-à-dire tel que $c_1/S_1 > c_2/(S_2 - T_2)$. C'est bien le résultat prévu en 2.2.

Voici maintenant les éléments de la résolution. Il convient de minimiser $c_1 w_1 + c_2 w_2$. Nous allons déterminer les nombres moyens de clients en attente, w_1 et w_2, en fonction de Φ.

La détermination de w_1 se réalise par l'élimination du temps d'attente moyen θ_1 (durée du service exclue) entre deux relations liant ce temps d'attente à w_1. La première de ces relations est une formule classique, établie par Little [13],

$$w_1 = \lambda_1 \theta_1 .$$

La seconde est

$$\theta_1 = w_1 S_1 + b_1 (S_1/2) + b_2 (1 - \Phi)^2 (S_2/2) ,$$

où $b_1 = \lambda_1 S_1$ et $b_2 = \lambda_2 S_2$ sont les fractions du temps pendant lesquelles la station est respectivement occupée par des clients de la catégorie 1 et de la catégorie 2. Le terme $w_1 S_1$ correspond à la durée du service du nombre moyen w_1 de clients de la catégorie 1. Il convient d'y ajouter le temps moyen d'attente dû à la poursuite éventuelle du service éventuellement en cours; il y a une probabilité b_1 que la station soit occupée par un client de la catégorie 1, et ceci donne lieu à une attente moyenne $S_1/2$ — d'où le terme $b_1 (S_1/2)$ — et une probabilité $(1 - \Phi)\, b_2$ que la station soit occupée par un client de la catégorie 2 dont le service est suffisamment avancé pour qu'il puisse être poursuivi, et ceci donne lieu à une attente moyenne $\frac{1}{2} (1 - \Phi)\, S_2$ — d'où le terme $b_2 (1 - \Phi)^2 (S_2/2)$.

L'élimination de θ_1 est immédiate et conduit à une équation linéaire, dont on déduit

$$w_1 = \frac{b_1^2 + \lambda_1 S_2 b_2 (1 - \Phi)^2}{2 (1 - b_1)} = K_1 + \frac{\lambda_1 S_2 b_2 (1 - \Phi)^2}{2 (1 - b_1)} , \qquad (5)$$

où K_1 est un terme indépendant de Φ.

Pour w_2, on a

$$w_2 = \lambda_2 \theta_2 ,$$

où θ_2 se compose du temps d'attente moyen avant le début du service, θ_2', et du temps d'attente moyen dû aux interruptions du service, θ_2''. La première de ces composantes, θ_2', est indépendante de Φ $(0 \leqslant \Phi \leqslant 1)$ et n'intervient donc pas dans la recherche de la valeur optimale de ce paramètre. La seconde, θ_2'', est égale au produit du nombre moyen de clients de la catégorie 1 qui se présentent pendant la partie $\Phi\, S_2$ d'un service S_2, c'est-à-dire $\lambda_1 \Phi\, S_2$, par la moyenne de la perte de temps globale due au service d'un client prioritaire et aux services des clients prioritaires qui arrivent pendant le service de celui-ci, et qui de ce fait précéderont au guichet le client de la catégorie 2. Cette perte de temps globale est égale, en moyenne, à $S_1/(1 - b_1)$, comme on peut le voir assez aisément, en la considérant comme une période d'occupation continue T_b de la station par des clients de la catégorie 1. Si on désigne par T_c une période d'inoccupation continue de la station par des clients de la catégorie 1, on a

$$\frac{E(T_b)}{E\,(T_b + T_c)} = b_1$$

ou

$$E\,(T_b) = \frac{b_1\, E(T_c)}{1 - b_1}$$

Or, T_c est la durée de « survie » d'un intervalle séparant deux arrivées consécutives d'un client de la catégorie 1, et l'on sait qu'un tel intervalle satisfait à une distribution exponentielle négative de paramètre λ_1 ; on a donc

$$E(T_c) = 1/\lambda_1 \,.$$

Il en résulte, par substitution et en tenant compte de ce que $b_1 = \lambda_1 S_1$, que

$$E(T_b) = S_1/(1 - b_1)\,,$$

ce qu'on voulait démontrer [1].

Le bilan de ce qui précède donne

$$\theta_2'' = \lambda_1 \Phi\, S_2 \frac{S_1}{1 - b_1}\,,$$

d'où, puisque $\lambda_2 S_2 = b_2$,

$$w_2 = K_2 + \frac{\Phi\, \lambda_1 S_1 b_2}{1 - b_1}\,, \qquad (6)$$

où K_2 est un terme indépendant de Φ.

[1] La même démonstration donne, dans le cas de durées de service non constantes,

$$E\,(T_b) = \frac{E\,(h_1)}{1 - b_1}\,.$$

La propriété sera exploitée sous cette forme plus générale en 3.3.

L'expression à minimiser, $c_1 w_1 + c_2 w_2$, est donc, grâce à (5) et (6), l'expression du second degré en Φ,

$$K_3 + c_1 \frac{\lambda_1 S_2 b_2 (1 - \Phi)^2}{2(1 - b_1)} + c_2 \frac{\lambda_1 S_1 b_2 \Phi}{1 - b_1}$$

(K_3 = terme indépendant de Φ). Une propriété classique des trinômes du second degré montre dès lors immédiatement que ce minimum est réalisé, comme il a été annoncé plus haut, par

$$\Phi = 1 - (c_2/c_1)(S_1/S_2) .$$

2.3.2. *Deuxième cas* (« repeat situation »).

L'étude nécessite des calculs supplémentaires, dans le détail desquels nous n'entrerons pas. Il faut notamment tenir compte de ce qu'un client de la catégorie 2 n'occupe pas la station pendant un temps S_2, mais bien, en moyenne, pendant le temps

$$S_2 + E(k)\, E(T) ,$$

où k est le nombre d'interruptions d'un service de la catégorie 2 et T ($0 \leqslant T \leqslant \Phi S_2$), pour chacune des interruptions, le temps passé au guichet en pure perte, du fait du régime « repeat ».

Le résultat, aussi, se présente moins simplement. Les auteurs donnent cependant un ensemble de formules qui, dans tout cas numérique, permet le calcul de la valeur optimale de Φ à partir de λ_1, S_1, c_1, λ_2, S_2 et c_2, à l'aide d'une machine à calculer ordinaire.

2.4. Le régime de la priorité absolue ou de la priorité « head-of-the-line » peut, bien entendu, être préféré à tout autre, et notamment à celui d'une priorité discrétionnaire, pour de simples raisons administratives, comme il se pourrait aussi qu'on ne soit pas en mesure de contrôler le temps qu'une unité non prioritaire a déjà passé à la station.

L'étude des régimes de priorité discrétionnaire aura montré que les raisons ainsi invoquées et la non-adoption, qui en résulte, d'un régime de priorité discrétionnaire optimal se paient.

3. On sait qu'une file d'attente ne peut atteindre un état d'équilibre statistique que si pendant une fraction du temps la ou les stations sont inactives. Et dans bien des cas, où l'on désire que la file ne soit pas trop longue, cette fraction peut même être importante. Ainsi, par exemple, pour un centre à une seule station, à arrivées poissonniennes et à durées de service

de même distribution quelconque, le nombre moyen d'unités dans la file d'attente (l'unité éventuellement au service exclue) est, en régime stationnaire, $C\,[(1 - L)^2/L]$, où C est une constante dépendant du coefficient de variation du temps de service et habituellement comprise entre $\frac{1}{2}$ et 1, et où L est la fraction du temps pendant laquelle la station est inoccupée. Si l'on désire n'avoir en moyenne qu'une unité en attente, $(1 - L)^2/L$ doit donc être de l'ordre de grandeur de 0,30 à 0,40.

On conçoit dès lors que dans le cas d'une station qui donne lieu à dépense même lorsqu'elle n'est pas en fonctionnement, la question se pose d'examiner si un système de fermeture et de réouverture de la station n'est pas économiquement préférable à un système d'ouverture permanente.

3.1. *Description du modèle et énoncé du problème.*

Le modèle qui va être décrit dans ce qui suit a été étudié par Naor et Yadin dans une communication qu'ils ont faite au mois d'août 1963, au 10[e] Congrès international du T.I.M.S. (The Institute of Management Sciences), à Tokyo [14].

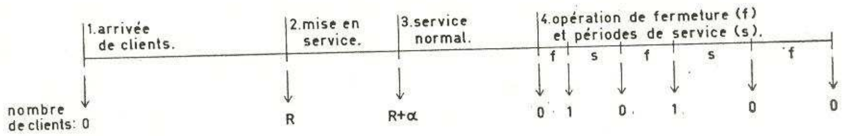

Fig. 1.

Il s'agit d'une seule station. On commence l'opération de fermeture dès le moment où plus aucun client n'est présent, mais on l'interrompt immédiatement si un client se présente avant qu'elle soit achevée, la station reprenant alors normalement son service pour le client dont l'arrivée en a interrompu l'opération de fermeture, ainsi que pour les clients successifs, et ce jusqu'à ce que le centre soit de nouveau vide d'unités à servir. L'opération de fermeture reprend alors depuis son début, c'est-à-dire que le temps qui y a déjà été consacré est perdu. Elle est éventuellement de nouveau interrompue, et ainsi de suite. Lorsque la station est fermée, c'est-à-dire lorsque l'opération de fermeture a pu être achevée, on attend, pour en commencer l'opération de réouverture, que R clients soient présents dans le système. A l'issue de l'opération de réouverture, $R + \alpha$ ($\alpha \geqslant 0$) clients sont présents dans le système. Le service reprend alors normalement, et ainsi de suite. Un cycle complet — que nous supposerons commencer à un instant où l'opération de fermeture vient de s'achever — comprend donc quatre phases : (voir fig. 1) : 1. arrivée de clients, jusqu'à ce que R clients soient présents; 2. opération de réouverture ou remise en service; 3. service normal, jusqu'à ce que plus aucun client ne soit présent; 4. opération de fermeture, interrompue par des périodes de service, de sorte qu'il conviendra de distinguer, dans cette quatrième phase, le temps d'occupation de la station et le temps consacré à sa fermeture. Les périodes d'occupation incluses dans cette quatrième phase commencent avec un client présent, tandis que la phase 3 commence avec $R + \alpha$ clients.

Les arrivées obéissent à une distribution de Poisson de paramètre λ. Les durées de service t_s ont la même fonction de répartition

$$F_s(t) = \text{Prob.}\ (t_s \leqslant t);$$

on suppose l'existence des deux premiers moments $E(t_s)$ et $E(t_s^2)$. Les durées de réouverture ont la même fonction de répartition $F_\alpha(t)$; on suppose l'existence des deux premiers moments $E(t_\alpha)$ et $E(t_\alpha^2)$. Les durées de l'opération de fermeture ont la même fonction de répartition $F_\beta(t)$; on suppose l'existence d'une fonction de fréquence $f_\beta(t)$ et des moments de tous les ordres. On fait l'hypothèse d'indépendance de toutes les variables aléatoires - durées ou intervalles de temps ainsi introduites.

On suppose encore la non-saturation, c'est-à-dire $b = \lambda E(t_s) < 1$, et l'existence des coûts suivants : C_c, coût de séjour dans le centre, par unité de temps et par client; C_T, coût de fermeture et de réouverture, par cycle; C_s, coût par unité de temps, lorsque la station est en service, et aussi pendant des fractions données ψ_α et ψ_β du temps consacré respectivement à la réouverture et à la fermeture.

Le problème est de chercher, en régime stationnaire, la valeur optimale de R ($\geqslant 1$), et de voir si l'adoption de cette valeur donne lieu à une règle de gestion qui soit économiquement préférable à la règle d'ouverture permanente de la station.

3.2. *Résolution* (les détails des calculs les plus longs sont donnés en annexe, en 3.3).

Le coût total se différencie du coût total relatif au régime d'ouverture permanente par deux coûts positifs, dus respectivement à l'allongement de la file et aux opérations de fermeture et réouverture, et un coût négatif, dû à l'élimination des dépenses de fonctionnement pendant la première phase et les fractions $(1 - \psi_\alpha)$ et $(1 - \psi_\beta)$ du temps consacré respectivement aux opérations de réouverture et de fermeture (deuxième phase et partie de la quatrième phase).

Le premier de ces coûts est égal, en moyenne, par unité de temps, à $C_c \cdot \Delta q$, où Δq est la différence entre le nombre moyen q_R de clients présents dans le centre sous le régime de fermeture et de réouverture à paramètre R et le nombre moyen q de clients présents dans le centre sous le régime d'ouverture permanente.

La valeur de q est donnée, en fonction de $b = \lambda\, E(t_s)$ et du coefficient de variation γ_s de t_s, par la formule bien connue de Pollaczek-Khintchine :

$$q = b + \frac{b^2}{1 - b} \, \frac{1 + \gamma_s^2}{2}.$$

Nous démontrerons en 3.3 que

$$q_R = q + \frac{1}{R + \alpha + \beta} \left[\frac{R\,(R - 1)}{2} + \alpha\,(R + \alpha \frac{1 + \gamma_a^2}{2})\right],$$

où γ_a^2 est le coefficient de variation de t_a, α le nombre moyen d'arrivées de clients pendant l'opération de réouverture (deuxième phase), β le nombre moyen de tentatives infructueuses de fermeture pendant la quatrième phase, c'est-à-dire le nombre moyen d'arrivées de clients pendant la partie de la quatrième phase qui est consacrée à l'opération de fermture. α et β se déterminent aisément à partir des données du problème, respectivement par

$$\alpha = \int_0^\infty \lambda t \, d\, F_a(t) = \lambda\, E_a(t)$$

et, avec

$$\pi = \int_0^\infty e^{-\lambda t} f_\beta(t)\; dt = \mathfrak{L}\,(f_\beta, \lambda),$$

qui est la probabilité pour qu'une opération de fermeture ne soit pas interrompue,

$$\beta = \sum_{n=0}^{\infty} n\, \pi\, (1 - \pi)^n = \frac{1 - \pi}{\pi}$$

Le deuxième coût est en moyenne, par unité de temps, C_T/T, où T est la durée moyenne du cycle. On montrera en 3.3, au cours des calculs relatifs à la détermination de q_R, que

$$T = \frac{R + \alpha + \beta}{\lambda\,(1 - b)}. \tag{14}$$

La détermination du troisième coût nécessite la connaissance des durées moyennes des première et deuxième phases et de la partie de la quatrième qui est consacrée à l'opération de fermeture. Les deux premières sont égales respectivement à R/λ (R fois la durée moyenne $1/\lambda$ d'un intervalle entre deux arrivées consécutives) et à $E(t_a) = \alpha/\lambda$. On montrera en 3.3, au cours des calculs relatifs à la détermination de q_R, que l'autre est égale à β/λ.

Il en résulte, grâce à (14), que le troisième coût est en moyenne, par unité de temps,

$$- \frac{C_s (1 - b)}{R + \alpha + \beta} [R + (1 - \psi_\alpha) \cdot \alpha + (1 - \psi_\beta) \cdot \beta].$$

Le total des trois coûts différenciels est ainsi en moyenne, par unité de temps,

$$\Delta C_R = \frac{1}{R + \alpha + \beta} \{C_c [\frac{R (R - 1)}{2} + \alpha (R + \alpha + \frac{1 + \gamma_\alpha^2}{2})$$

$$+ C_T \lambda (1 - b) - C_s (1 - b) [R + (1 - \psi_\alpha) \alpha + (1 - \psi_\beta) \beta] \} .$$

La valeur optimale de R peut être cherchée par annulation de la dérivée, d'où l'équation du second degré

$$R^2 + 2 (\alpha + \beta) R - \{ (\gamma_\alpha^2 - 1) \alpha^2 + 2 (1 - 2 \beta) \alpha$$

$$+ \beta + 2 (1 - b) [(\psi_\alpha \alpha + \psi_\beta \beta) \frac{C_s}{C_c} + \lambda \frac{C_T}{C_c}] \} = 0 ,$$

et comparaison de $\Delta C_{[R^*]}$ et $\Delta C_{[R^*+1]}$, où R^* est une solution de cette équation et $[R^*]$ le plus grand entier qui ne lui soit pas supérieur. Le régime de fermeture et réouverture le meilleur ne sera cependant économiquement préférable au régime d'ouverture permanente que si l'accroissement ΔC correspondant est négatif :

$$\Delta C_{R'} < 0 , \qquad \text{où } R' = [R^*] \text{ ou } [R^* + 1] .$$

3.3. *Annexe :* Calculs relatifs à la résolution 3.2.

On a, par la formule de Little [13],

$$q_R = \lambda \, \theta_R , \tag{15}$$

où θ_R est le temps moyen de séjour d'un client dans le système (durée du service comprise); le calcul de q_R revient donc à celui de θ_R.

Or, si l'on désigne par E_{ki} ($k = 0, 1; i = 0, 1, 2, \ldots$) l'état du système pour k stations disponibles et i clients présents, et par p_{ki} et θ_{ki} respectivement la probabilité pour que le système soit dans l'état E_{ki} et le temps moyen de séjour d'un client qui arrive dans le système lorsque celui-ci est dans l'état E_{ki}, on a

$$\theta_R = \sum_k \sum_i p_{ki} \, \theta_{ki} . \tag{16}$$

Déterminons d'abord les θ_{ki}. Pour les θ_{0i} $(0 \leqslant i \leqslant R-1)$, c'est-à-dire les temps moyens de séjour des clients qui arrivent pendant la première phase, i clients déjà présents, on a

$$\theta_{0i} = \frac{R-1-i}{\lambda} + E(t_a) + (i+1)\, E(t_s), \tag{17}$$

les trois termes du second membre étant les espérances mathématiques respectives du temps nécessaire à l'arrivée de $(R-i-1)$ autres clients — l'opération de réouverture ne commence qu'à l'arrivée du $(R-i-1)^e$ — de la durée de la période de réouverture et de la durée de $(i+1)$ services.

Pour les θ_{0i} $(i \geqslant R)$, c'est-à-dire les temps moyens de séjour des clients qui arrivent pendant la deuxième phase, i clients déjà présents, on a

$$\theta_{0i} = E(\tau_a \mid i) + (i+1)\, E(t_s), \tag{18}$$

où $E(\tau_a \mid i)$ est l'espérance mathématique du temps nécessaire à l'achèvement de l'opération de réouverture, étant donné que i clients se trouvent dans le centre.

Pour les θ_{1i} $(1 \leqslant i < \infty)$, c'est-à-dire les temps moyens de séjour des clients qui arrivent soit pendant la troisième phase, soit pendant la partie d'occupation de la quatrième phase, on a

$$\theta_{1i} = E(\tau_s) + i\, E(t_s), \tag{19}$$

où τ_s est le temps de survie du service en cours.

Pour θ_{10} enfin, c'est-à-dire le temps moyen de séjour d'un client qui arrive pendant la quatrième phase, alors qu'est en cours l'opération de fermeture, on a simplement

$$\theta_{10} = E(t_s). \tag{20}$$

Dans la substitution de (17), ..., (20) dans (16), apparaît le terme $\sum_{i=R}^{\infty} p_{0i}\, E(\tau_a \mid i)$. Or, l'espérance mathématique inconditionnelle $E(\tau_a)$ est égale à

$$\frac{\sum_{i=R}^{\infty} p_{0i}\, E(\tau_a \mid i)}{\sum_{i=R}^{\infty} p_{0i}} \; ;$$

le terme considéré peut donc s'écrire $E(\tau_a) . \sum_{i=R}^{\infty} p_{0i}$.

On a en outre, grâce à une propriété établie par Cox et Smith [15],

$$E(\tau_a) = E(t_a) \cdot \frac{1+\gamma_a^2}{2} \qquad \text{et} \qquad E(\tau_s) = E(t_s) \cdot \frac{1+\gamma_s^2}{2},$$

et enfin — rappel —

$$E(t_a) = \alpha/\lambda \qquad \text{et} \qquad E(t_s) = b/\lambda\,.$$

Il vient ainsi

$$\left\{\begin{aligned} q_R = \{ & \sum_{i=0}^{R-1} p_{0i}\,[(R-1-i) + \alpha + (i+1)\,b] \\ & + \sum_{i=R}^{\infty} p_{1i}\,[b\,\frac{1+\gamma_a^2}{2} + ib] + p_{10}\,b\}\,. \\ & + \sum_{i=1}^{\infty} p_{0i}\,[\alpha\,\frac{1+\gamma_s^2}{2} + (i+1)\,b\,] \end{aligned}\right.$$

Mais

$$\sum_{i=0}^{R-1} p_{0i}\,(i+1)\,b + \sum_{i=R}^{\infty} p_{0i}\,(i+1)\,b + \sum_{i=1}^{\infty} p_{1i}\,ib + p_{10}\,b$$

$$= b\,[\sum_{i=0}^{\infty} ip_{0i} + \sum_{i=1}^{\infty} ip_{1i} + \sum_{i=0}^{\infty} p_{0i} + \sum_{i=0}^{\infty} p_{1i} - \sum_{i=1}^{\infty} p_{1i}] = b\,[q+1-b]\,.$$

On a donc

$$\left\{\begin{aligned} q_R = & \sum_{i=0}^{R-1} p_{0i}\,[\,(R-1-i) + \alpha] \\ & + \alpha\,\frac{1+\gamma_a^2}{2}\,p_{(2)} + b\,\frac{1+\gamma_s^2}{2}\,p_{(3)}\,, \end{aligned}\right. \qquad (21)$$

où $p_{(2)}$ et $p_{(3)}$ sont les probabilités pour que le système se trouve respectivement en deuxième ou troisième phase.

Les probabilités qui figurent dans (21) vont être calculées à l'aide du rapport à T, la durée moyenne du cycle, des durées moyennes de E_{0i} ($i = 0, \dots, R-1$) et des deuxième et troisième phases. La durée moyenne des E_{0i} est $1/\lambda$, l'intervalle de temps moyen entre deux arrivées consécutives; les p_{0i} ($i = 0, \dots, R-1$) valent donc $1/\lambda T$. La durée moyenne de la première phase est R/λ. La durée moyenne de la deuxième phase est $E(t_a) = \alpha/\lambda$; on a donc $p_{(2)} = \alpha/\lambda T$. La durée moyenne de la troisième phase serait $E(t_s)/(1-b)$ — voir la note (1) de 2.3.1 — si un seul client était présent au début de la troisième phase; elle est donc

$$(R + \alpha) \frac{E(t_s)}{1-b} = (R + \alpha)(b/\lambda)$$

et on a

$$p_{(3)} = \frac{(R+\alpha)\, b}{(1-b)\,\lambda T}.$$

La durée moyenne de la partie de la quatrième phase pendant laquelle la station est en fonctionnement est — voir de nouveau la note (1) de 2.3.1—

$$\beta \frac{E(t_s)}{1-b} = \frac{\beta}{1-b}(b/\lambda).$$

Le calcul de la durée moyenne de la partie de la quatrième phase pendant laquelle la station ne sert aucun client est plus délicat. Cette durée moyenne est égale à $\beta\, E(t') + E(t'')$, où t' est la variable aléatoire - durée d'une tentative infructueuse de fermeture et t'' la variable aléatoire - durée d'une tentative fructueuse. La densité de probabilité de t' est

$$\frac{1}{1-\pi}\, \lambda\, e^{-\lambda t}\, [1 - F_\beta(t)],$$

où $(1-\pi)$ représente la probabilité pour que l'opération de fermeture soit interrompue, $\lambda\, e^{-\lambda t}$ la dérivée de la probabilité $(1 - e^{-\lambda t})$ pour qu'il arrive au moins un client au cours de la période t, et $1 - F_\beta(t)$ la probabilité pour que la durée de l'opération de fermeture complète soit supérieure à t. La densité de probabilité de t'' est

$$(1/\pi)\, f_\beta(t)\, e^{-\lambda t},$$

où π représente la probabilité pour que l'opération de fermeture ne soit pas interrompue et $e^{-\lambda t}$ la probabilité pour qu'il n'arrive aucun client au cours de la période t. La durée moyenne recherchée est donc, puisqu'on sait que $\beta = (1-\pi)/\pi$ (voir 3.2),

$$\frac{1}{\pi}\left[\int_0^\infty \lambda t\, e^{-\lambda t} \int_t^\infty f_\beta(u)\, du\, dt + \int_0^\infty t e^{-\lambda t} f_\beta(t)\, dt\right],$$

qui peut se transformer en

$$\frac{1}{\lambda\pi}\int_0^\infty (1 - e^{-\lambda t})\, f_\beta(t)\, dt = \frac{1}{\lambda\pi}[1 - \mathfrak{L}(f_\beta, \lambda)]$$

$$= \frac{1}{\lambda}\;\frac{1-\pi}{\pi} = \frac{\beta}{\lambda} \qquad \text{(voir 3.2).}$$

Par addition des durées moyennes des première, deuxième, troisième et quatrième phases, on a

$$T = \frac{R}{\lambda} + \frac{\alpha}{\lambda} + \frac{R+\alpha}{1-b} \cdot \frac{b}{\lambda} + \frac{\beta}{1-b} \cdot \frac{b}{\lambda} + \frac{\beta}{\lambda}$$

$$= \frac{1}{\lambda}(R+\alpha+\beta)\left[1+\frac{b}{1-b}\right] = \frac{R+\alpha+\beta}{\lambda(1-b)},$$

formule annoncée en (14), et d'où l'on déduit, en retournant à ce qui a déjà été obtenu pour p_{0i} $(0 \leqslant i \leqslant R-1)$, $p_{(2)}$ et $p_{(3)}$,

$$p_{0i} = \frac{1-b}{R+\alpha+\beta},$$

$$p_{(2)} = \frac{\alpha(1-b)}{R+\alpha+\beta},$$

$$p_{(3)} = \frac{(R+\alpha)\,b}{R+\alpha+\beta},$$

expressions qu'il suffit de substituer dans (21) pour obtenir, après une transformation aisée,

$$\left\{\begin{aligned} q_R &= b + \frac{b^2}{1-b}\,\frac{1+\gamma_s^2}{2} + \frac{R}{R+\alpha+\beta} \cdot \frac{R-1}{2} \\ &\quad + \frac{\alpha}{R+\alpha+\beta}\left[R+\alpha\,\frac{1+\gamma_a^2}{2}\right] \\ &= q + \frac{1}{R+\alpha+\beta}\left[\frac{R(R-1)}{2} + \alpha\left(R+\alpha\,\frac{1+\gamma_a^2}{2}\right)\right]. \end{aligned}\right.$$

REFERENCES

[1] W.R. VAN VOORHIS. Waiting-Line Theory as a Management Tool. Operations Research, v. 4, 1956, pp. 221-231.

[2] A. KAUFMANN et R. CRUON. Les phénomènes d'attente. Dunod, Paris, 1961.

[3] Th.L. SAATY. Elements of Queueing Theory, with Applications. McGraw-Hill, New York, 1961.

[4] Ph.M. MORSE. Queues, Inventories and Maintenance. Wiley, New York, 1958.

[5] D.R. COX and W.L. SMITH. Queues. Methuen and Co, London, 1961.

[6] B. KRISHNAMOORTI. On Poisson Queues with heterogeneous Servers. Operations Research, v. 11, 1963, pp. 322-329.

[7] F.S. HILLIER. Economic Models for Industrial Waiting Line Problems. Management Science, v. 10, 1963, pp. 119-130.

[8] I. BROSH and P. NAOR. On Optimal Disciplines in Priority Queueing. Bulletin Inst. Int. de Stat., t. XL, Actes de la trente-quatrième Session, Toronto, 1964, pp. 593-609.

[9] A. COBHAM. Priority Assignment in Waiting Line Problems. Operations Research, v 2, 1954, pp. 91-99.

[10] R.G. MILLER. Priority Queues. Annals of Math. Stat., v. 31, 1960, pp. 86-103.

[11] B. AVI-ITZHAK and P. NAOR. Some Queueing Problems with the Service Station Subject to Breakdown. Operations Research, v. 11, 1963, pp. 303-320.

[12] B. AVI-ITZHAK, I. BROSH and P. NAOR. On Discretionary Priority Queueing. Zeitschrift für Angewandte Mathematik und Mechanik, v. 44, 1964, pp. 235-242.

[13] J.D.C. LITTLE. A Proof for the Queueing Formula : $L = \lambda W$. Operations Research, v. 9, 1961, pp. 383-387.

[14] M. YADIN and P. NAOR. Queueing Systems with a removable Service Station. Opérational Research Quarterly, v. 14, 1963, pp. 393-406.

[15] D.R. COX and W.L. SMITH. On the Superposition of Renewal Processes. Biometrika, v. 41, 1954, pp. 91-99.