

Occupation des ressources partagées dans un système multi-service

T.B. Ravaliminoarimalalason¹, M. Rakotomalala^{1,2}, F. Randimbindrainibe¹

¹EDT « Sciences et Techniques de l'Ingénierie et de l'Innovation » de l'Université d'Antananarivo

²Institut Supérieur de Technologie d'Antananarivo

Correspondant : tokybaz@gmail.com

Résumé

Cet article présente des expressions analytiques des quantités de ressources occupées dans un système fournissant plusieurs services à ces utilisateurs. Ce système est assimilable à une file d'attente multi-service. Les quantités de ressources nécessaires à chaque service sont déterministes. Des analyses effectuées sur ces occupations ont permis de dimensionner la quantité totale de ressources nécessaire qu'on doit déployer sur le serveur de la file d'attente pour servir les utilisateurs à une certaine qualité de service près en terme de probabilité d'encombrement. Rapporté au cas mono-service, on a retrouvé les résultats sur la probabilité d'encombrement dans le modèle de trafic de Fry-Molina à partir des expressions analytiques établis dans ce travail. Les résultats obtenus sont comparés à des simulations d'étude de cas similaires à un système réel.

Mots-clés : File d'attente, Dimensionnement, Multi-service, Occupation, Partage de ressources

I. Introduction

Le dimensionnement des nombres de ressources nécessaires au fonctionnement d'un système garantissant une qualité de service impose a commencé à l'ère de Erlang. Il a établi une fameuse formule d'Erlang B pour déterminer le nombre de canaux nécessaires à un réseau téléphonique [1]. Cette formule ne considère qu'un seul service (le service voix) fourni par un système (réseau téléphonique) aux utilisateurs qui en demandent. Sa formule est déduite de l'étude d'une file d'attente de capacité finie M/M/n/n. Les résultats de ses études ont été développés et étendus par beaucoup de chercheurs comme Palm [2], Vaultot et Chaveau [3], Joys [4], Iversen [5] et beaucoup d'autres.

Actuellement, on rencontre des multiples facettes de services que peuvent fournir les systèmes complexes, si on ne parle que des réseaux de communication par exemple. Iversen [6] a étudié le cas d'un système multi-service en évaluant le taux de congestion d'un tel système doté d'un contrôle d'accès. Des cas de trafic élastique ont été développés sur des systèmes assimilables en file d'attente multi-service par Hanczewski et al. [7]. Dans [7], il a proposé un modèle de file d'attente multi-service à discipline SD-FIFO (State Dependent FIFO). Ce système alloue ses ressources à chaque classe de service d'une façon équité équilibrée. Ils ont encore avancé une généralisation des files d'attente multi-service utilisés en trafic élastique [8].

D'autres études sur le partage équité de ressources ont été avancées par Bonald et al. [9] où toutes les ressources sont disposées à toutes les classes de service et à tous les utilisateurs qui se présentent devant le serveur. Une extension de Haddad et Mazumdar [10] évalue une certaine congestion dans tel système.

Hanczewski et al. [11] ont proposé un autre type de discipline appelé cFIFO (continuous FIFO) pour la file d'attente multi-service. Ils ont donné une approximation du modèle de convolution d'un système de file d'attente multi-service cFIFO tout en mettant possible les utilisations à débit variable en fonction des quantités de ressources initialement inutilisées dans le serveur.

Dans toutes ces littératures, que ce soit dans le partage des ressources disponibles à tous les utilisateurs présents dans le système, ou soit avec de débit variable, de trafic élastique, le problème se pose sur la quantité de ressources que le serveur de la file d'attente multi-service dispose à ces clients. Cet article propose une généralisation de file d'attente multi-service, avec des quantités de ressources fixes pour chaque classe de service, dans le but de déterminer la quantité totale de ressources nécessaires au serveur afin d'assurer une certaine qualité de service.

II. Méthodes

Notre étude est basée sur les déterminations analytiques des occupations de ressources dans une file d'attente multi-service partageant ses ressources à travers des clients y demandant des services. Chaque client peut demander un service i nécessitant une ressource R_i du système pendant une durée aléatoire de moyenne μ_i . Le taux d'arrivée des clients demandant le service i au système est décrit par un processus de Poisson d'intensité λ_i .

En notant par N le nombre total de services que ce système peut fournir à ces clients, supposant que les arrivées des clients de chaque service sont indépendantes, on peut dire que l'arrivée des clients à ce système forme un processus de Poisson dont l'intensité λ est exprimé par $\lambda = \lambda_1 + \dots + \lambda_N = \sum_{i=1}^N \lambda_i$. Le taux p_i de clients demandant le service i est égal à $p_i = \lambda_i / \lambda$ pour chaque i . Ces deux expressions découlent de la propriété de superposition de processus de Poisson indépendants [12].

Soit S_i la distribution de la durée du service i demandé par un client. C'est une variable aléatoire réelle strictement positive de moyenne μ_i . On va noter par S la durée de service demandé par un client quelconque. A partir de la formule des probabilités totales, on peut exprimer que S est égale en loi à $\sum_{i=1}^N p_i S_i$, c'est-à-dire que la fonction de répartition $F_S(s)$ de S

est égale à $F_S(s) = \sum_{i=1}^N p_i F_{S_i}(s)$, où $F_{S_i}(s)$ est la fonction de répartition de S_i . Suite à l'indépendance des variables aléatoires S_i , son espérance est exprimé par $\mu = \sum_{i=1}^N p_i \mu_i$.

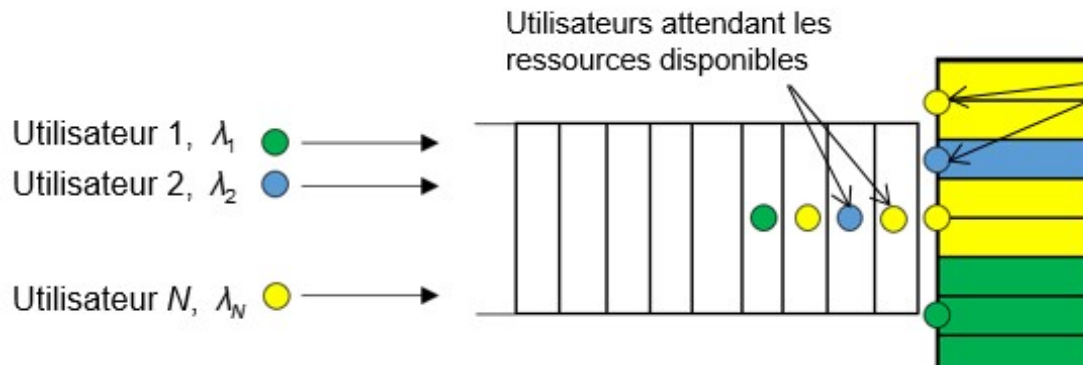


Figure 1. Description du modèle

Le système est alors équivalent à une file d'attente, à serveur unique doté de ressources de quantité R . Les clients arrivent à cette file d'attente selon un processus de Poisson d'intensité λ . La durée de service d'un client quelconque est une variable aléatoire S de moyenne μ . Ces clients vont utiliser une quantité de ressources fixe R_i du serveur de la file selon le type de service i qu'il a demandé. Si des clients sont en train d'être servis par le système, utilisant une ressource r , la somme de toutes les quantités de ressources utilisées par ces clients, un autre client arrivant voulant utiliser une quantité de ressources inférieure à $R - r$ sera tout de suite servi. Si la quantité de ressources demandée par le client arrivant est supérieure à $R - r$, alors il sera mis dans la file d'attente attendant qu'un client servi termine et libère des ressources.

Dans le cas où la quantité de ressources disponible R du serveur est infinie, il s'agit d'une file d'attente $M/GI/\infty$ (en terme des clients). Les clients qui arrivent seront immédiatement servis.

Puisque l'arrivée de ces clients de chaque type de service est aléatoire, la quantité totale de ressources occupées dans le serveur est donc une variable aléatoire. On va déterminer analytiquement sa fonction de répartition pour un cas de ressources infinies dans le serveur. On verra ensuite le cas d'une capacité finie du serveur et à partir de laquelle on va déduire la quantité totale de ressources nécessaires. Un cas particulier de modèle mono-service à partir de l'expression analytique obtenu sera considéré.

Une comparaison des résultats analytiques avec une simulation effectuée sur MATLAB-Simulink validera le modèle ainsi posé.

III. Résultats

III.1. Résultats analytiques

III.1.1. Partage de ressources infinies

On va exprimer la quantité de ressources utilisée par les clients décrits dans le paragraphe II. Notons par π_n la probabilité que le système serve n clients. π_n suit la loi de Poisson de paramètre $\lambda \int_0^{+\infty} (1 - F_S(s)) ds = \lambda \int_0^{+\infty} F_S^c(s) ds$, c'est-à-dire :

$$\pi_n = \frac{\lambda^n}{n!} \cdot \left(\int_0^{+\infty} F_S^c(s) ds \right)^n \cdot \exp\left(-\lambda \int_0^{+\infty} F_S^c(s) ds\right) \quad (1)$$

En remplaçant par les expressions de λ et de F_S dans le cadre de notre système fournissant N services à ses clients, on a donc :

$$\pi_n = \frac{1}{n!} \cdot \left(\int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds \right)^n \cdot \exp\left(-\int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds\right) \quad (2)$$

Conditionnellement au nombre k clients que le serveur est en train servir, on va noter r_k la quantité de ressources utilisée par ces clients. Soit $P(R \leq r)$ la fonction de répartition de la loi de probabilité qu'une quantité r de ressources de ce système soit utilisée par les clients qu'il sert.

On a alors :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \pi_k \cdot P(r_k \leq r) \quad (3)$$

Où $P(r_k \leq r)$ désigne la fonction de répartition de la loi de probabilité que la quantité de ressources r occupée par les k clients soit r_k .

Maintenant on va poser que le système sert k clients. La fonction de répartition de la loi de probabilité que la quantité de ressources utilisée soit r a pour expression $P(r_k \leq r)$.

Un client demandant un service i ($i = 1, \dots, N$) utilise une ressource R_i . Il existe des N -uplets de coefficients (k_1, \dots, k_N) vérifiant $\sum_{i=1}^N k_i \cdot R_i \leq r$ tel que $k_1 + \dots + k_N = k$. k_i indique le nombre de clients utilisant le service i parmi les k clients servis par le système.

$$P(r_k \leq r) = \sum_{\substack{k_1, \dots, k_N \\ k_1 R_1 + \dots + k_N R_N \leq r}} \binom{k}{k_1, \dots, k_N} \cdot p_1^{k_1} \dots p_N^{k_N} \quad (4)$$

Où p_i désigne la probabilité qu'un client choisi au hasard parmi les k clients servis utilise le service i , c'est-à-dire $p_i = \lambda_i / \lambda$.

$$\text{Et } \binom{k}{k_1, \dots, k_N} = \frac{k!}{k_1! k_2! \dots k_N!}.$$

L'expression de la fonction de répartition de la loi de probabilité de l'utilisation de ressources du système est donc donnée par :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \pi_k \cdot \left(\sum_{\substack{k_1, \dots, k_N \\ k_1 R_1 + \dots + k_N R_N \leq r}} \binom{k}{k_1, \dots, k_N} \cdot p_1^{k_1} \dots p_N^{k_N} \right) \quad (5)$$

Où :

$$\pi_k = \frac{1}{k!} \cdot \left(\int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds \right)^k \cdot \exp \left(- \int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds \right)$$

Dans toute la suite, on va simplifier l'écriture de l'équation (5) par :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \sum_{\substack{k_i \\ \sum_{i=1}^N k_i R_i \leq r}} \left(\pi_k \cdot \binom{k}{k_i} \cdot \prod_i \rho_i^{k_i} \right)$$

En adoptant les notations $\vec{k}_i = k_1, \dots, k_N$, $\vec{R}_i = R_1, \dots, R_N$ et $\vec{k}_i \vec{R}_i = k_1 R_1 + \dots + k_N R_N$.

III.1.2. Partage de ressources finies

Pour le cas des ressources finies, il est évident de penser à une file d'attente M/GI/C où C désigne une certaine capacité finie du système. Mais la résolution analytique de cette file d'attente reste encore un problème ouvert, et les solutions proposées jusqu'à maintenant ne sont que des approximations [13]. Et en plus dans notre cas, on a une capacité variable en fonction des utilisateurs servis par le système et de leurs besoins en terme de nombre de ressources, donc ceci va encore empirer la complexité de notre problème.

Dans notre recherche, nous nous focalisons sur le dimensionnement du nombre de ressources nécessaires pour satisfaire les clients dans leurs utilisations des services fournis par le système.

Etant donné que le système a une capacité C en terme de quantités de ressources disponibles qu'il peut attribuer aux utilisateurs. Si toutes les ressources occupées avec le besoin du client suivant est supérieur à la quantité de ressources disponibles dans le système, on peut dire qu'il y a un phénomène de blocage. Deux cas de figure peuvent se présenter : soit le client en question est rejeté, soit il est mis en attente. Dans tous ces deux cas, la capacité en terme de quantité de ressources est insuffisante.

D'un côté, si on va déployer beaucoup de ressources. Quelques une d'entre elles peuvent ne pas être utilisées. Il s'agit donc de gaspillage de ressources. Dans certains cas, les mises à disponibilités de ressources sont payantes et ce sera donc une perte en terme d'investissement. De l'autre côté, le déploiement de peu ressources peuvent faire perdre certains utilisateurs comme nous le décrivons dans le phénomène de blocage précédent. Il est donc nécessaire de bien déterminer la bonne quantité de ressources à mettre en place. Pour cela, on doit adopter un certain seuil de tolérance ϵ tel que la probabilité de rejet ou de mise en attente des utilisateurs soit inférieure à cette tolérance.

Assimilons qu'on a déployé une quantité infinie de ressources. Pour cela, la probabilité qu'on utilise une quantité d'au plus r ressources est donnée par l'équation (5) rappelée ci-dessous :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \sum_{\substack{k_j \\ \sum_{j=1}^N k_j R_j \leq r}} \left(\pi_k \cdot \binom{k}{k_1, \dots, k_N} \cdot \prod_i \rho_i^{k_i} \right)$$

En d'autres termes, la probabilité qu'on utilise une quantité de ressources supérieure à C est donnée par :

$$B = P(R > C) = \sum_{k=0}^{+\infty} \sum_{\substack{k_j \\ \sum_{j=1}^N k_j R_j > C}} \left(\pi_k \cdot \binom{k}{k_1, \dots, k_N} \cdot \prod_i \rho_i^{k_i} \right)$$

On peut dire que c'est la probabilité de blocage du système de capacité C puisqu'à partir de cette quantité de ressources C, il ne peut plus servir d'utilisateur.

Pour un dimensionnement avec une tolérance ε , l'objectif est donc de déterminer une capacité C telle que :

$$B = P(R > C) = \sum_{k=0}^{+\infty} \sum_{\substack{k_j \\ \sum_{j=1}^N k_j R_j > C}} \left(\pi_k \cdot \binom{k}{k_1, \dots, k_N} \cdot \prod_i \rho_i^{k_i} \right) \leq \varepsilon \quad (6)$$

III.1.3. Cas mono-service et ressource unitaire pour chaque utilisateur

Dans cette partie, on va voir le cas d'un système offrant un service unique dont chaque utilisateur servi n'utilise qu'une seule quantité de ressource. Pour le cas de durée de service exponentielle de moyenne $1/\mu$, on va voir qu'à partir des équations (5) et (6), on aboutit au résultat du modèle de trafic BCH (Busy-Calls Held) de Fry-Molina [14][15].

Le nombre de services fourni par le système est alors $N = 1$. Le taux d'arrivée des utilisateurs suit un processus de Poisson d'intensité $\lambda_1 = \lambda$. Puisque la durée de service exponentielle, la fonction de répartition $F_S(s) = F_{S_1}(s)$ de la variable aléatoire S de durée de service est donnée par $F_S(s) = 1 - \exp(-\mu s)$. Le taux des utilisateurs demandant le service 1 est donc $\rho_1 = 1$.

La fonction de répartition de la loi de probabilité de l'utilisation de ressources du système, à partir de l'équation (5), est donné par :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \pi_k \cdot \left(\sum_{\substack{k_1, \dots, k_N \\ k_1 R_1 + \dots + k_N R_N \leq r}} \binom{k}{k_1, \dots, k_N} \cdot \rho_1^{k_1} \dots \rho_N^{k_N} \right) = \sum_{k=0}^{+\infty} \pi_k \cdot \left(\sum_{k_1 \leq r} \binom{k}{k_1} \cdot \rho_1^{k_1} \right) = \sum_{k=0}^{+\infty} \pi_k \cdot 1_{k \leq r}(k) = \sum_{k=0}^r \pi_k$$

Où :

$$\begin{aligned} \pi_k &= \frac{1}{k!} \cdot \left(\int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds \right)^k \cdot \exp\left(-\int_0^{+\infty} \sum_{i=1}^N \lambda_i F_{S_i}^c(s) ds\right) \\ &= \frac{1}{k!} \cdot \left(\int_0^{+\infty} \lambda \exp(-\mu s) ds \right)^k \cdot \exp\left(-\int_0^{+\infty} \lambda \exp(-\mu s) ds\right) \\ &= \frac{1}{k!} \cdot \left(\frac{\lambda}{\mu} \right)^k \cdot \exp\left(-\frac{\lambda}{\mu}\right) = \frac{\rho^k}{k!} \exp(-\rho) \end{aligned}$$

en notant $\lambda / \mu = \rho$.

Ainsi,

$$P(R \leq r) = \sum_{k=0}^r \pi_k = \sum_{k=0}^r \frac{\rho^k}{k!} \exp(-\rho) = \exp(-\rho) \sum_{k=0}^r \frac{\rho^k}{k!}.$$

L'équation (6) devient aussi :

$$\begin{aligned} P(R > C) &= \sum_{k=0}^{+\infty} \sum_{\substack{k_j \\ k_j R_j > C}} \left(\pi_k \cdot \binom{k}{k_j} \cdot \prod_i \rho_i^{k_j} \right) \leq \varepsilon \\ &= \exp(-\rho) \sum_{k=C}^{+\infty} \frac{\rho^k}{k!} \leq \varepsilon \end{aligned}$$

En remplaçant $\exp(-\rho)$ par $1 / \sum_{k=0}^{+\infty} \frac{\rho^k}{k!}$, on retrouve le taux de congestion $\frac{\sum_{k=C}^{+\infty} \frac{\rho^k}{k!}}{\sum_{k=0}^{+\infty} \frac{\rho^k}{k!}}$ posé par Fry-Molina dans leur modèle de trafic.

On note que ce modèle, qui est une alternative de la loi d'Erlang B [1], est utilisé dans certains dimensionnements de téléphonie en Amérique du Nord. Il est constaté que ce modèle reflète beaucoup plus la réalité des trafics observés par rapport au modèle d'Erlang [16].

III.2. Comparaison des résultats analytiques à des simulations

On va voir un exemple de partage de ressources entre les utilisateurs arrivant chez un système de capacité quelconque C dont on déterminera. Le système peut fournir un des $N = 4$ services à chaque utilisateur. Les arrivées de ces utilisateurs à ce système forment des processus de Poisson d'intensité respective $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 2,5$, et $\lambda_4 = 5$ arrivées par unité de temps pour le service $i = 1, 2, 3$, et 4 . Par le principe de superposition des processus de Poisson, ces arrivées forment un seul processus d'intensité $\lambda = \sum_{i=1}^4 \lambda_i = 12,5$.

Les quantités de ressources nécessaires pour ces services sont respectivement $R_1 = 4$, $R_2 = 3$, $R_3 = 2$, et $R_4 = 1$ ressource.

Les durées de services sont supposées aléatoires suivant des lois exponentielles de moyennes respectives $1/\mu_1 = 1/3$, $1/\mu_2 = 1/4$, $1/\mu_3 = 1/5$, et $1/\mu_4 = 1/6$ unités de temps pour les services $i = 1$ à 4 .

D'après l'équation (5), la fonction de répartition de la loi de probabilité de l'utilisation de ressources du système à capacité infinie est égale à :

$$P(R \leq r) = \sum_{k=0}^{+\infty} \pi_k \cdot \left(\sum_{\substack{k_1, \dots, k_4 \\ k_1 R_1 + \dots + k_4 R_4 \leq r}} \binom{k}{k_1, \dots, k_4} \cdot \rho_1^{k_1} \dots \rho_4^{k_4} \right)$$

Tel que $\rho_i = \lambda_i / \lambda$ désigne la probabilité que l'utilisateur arrivant au système va utiliser le service i .

Et :

$$\pi_k = \frac{1}{k!} \cdot \left(\int_0^{+\infty} \sum_{i=1}^4 \lambda_i e^{-\mu_i s} ds \right)^k \cdot \exp \left(- \int_0^{+\infty} \sum_{i=1}^4 \lambda_i e^{-\mu_i s} ds \right) = \frac{1}{k!} \cdot \left(\sum_{i=1}^4 \frac{\lambda_i}{\mu_i} \right)^k \cdot \exp \left(- \sum_{i=1}^4 \frac{\lambda_i}{\mu_i} \right)$$

Les courbes de la figure 2 représentent cette fonction de répartition :

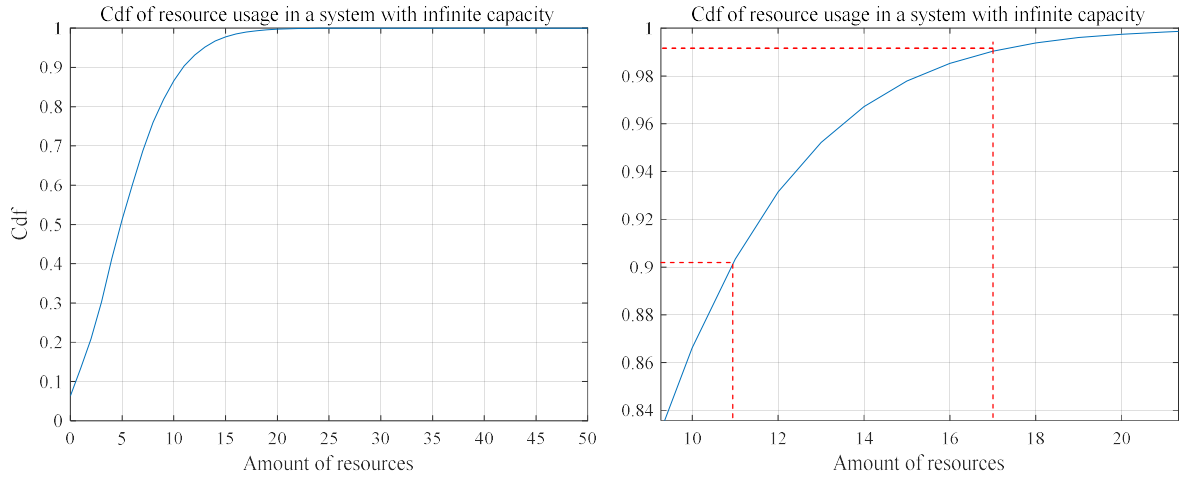


Figure 2. Fonction de répartition de la quantité de ressources occupées pour une capacité infinie

Il est bien évident à partir de la figure 2 que le déploiement de ressources à partir d'une quantité initiale égale à 20 n'influence pas au système. On déduit que la probabilité que la quantité de ressources occupées soit supérieure ou égale à 20 est très infime (avoisine le zéro). Il n'est donc pas nécessaire de doter plus de 20 ressources le système puisqu'elles ne vont pas être toutes utilisées.

A partir de ces courbes et l'équation (6), on peut conclure une probabilité d'encombrement de 10% lors que le système est doté de 11 ressources, et cette probabilité devient 1% avec 17 ressources.

Nous avons utilisé MATLAB – Simulink pour simuler ces deux valeurs de quantités de ressources pour vérifier ces probabilités d'encombrement.

On a utilisé le modèle de la figure 3. Les utilisateurs sont générés à partir des blocs User of service i selon un taux respectif λ_i décrits précédemment.

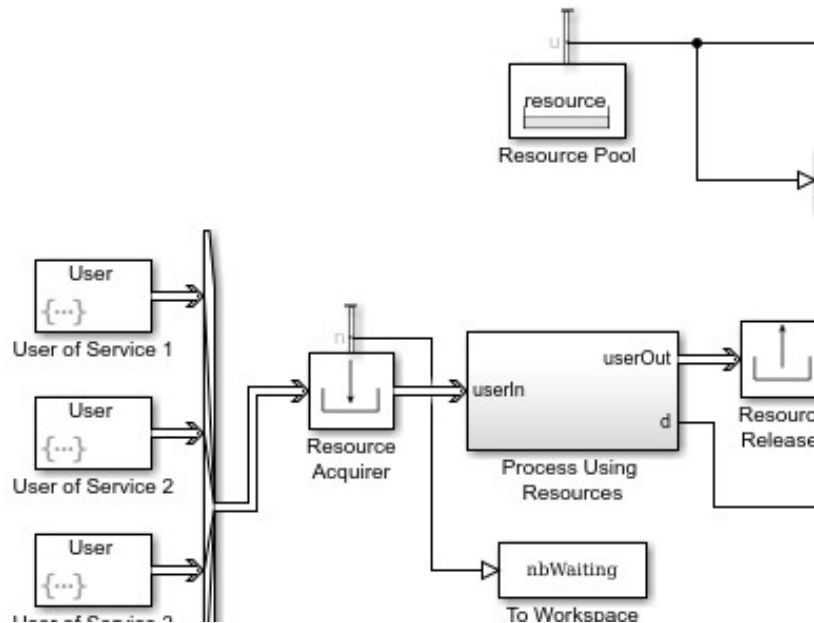


Figure 3. Modèle de partage de ressources d'un système fournissant quatre services à ses utilisateurs

Le calcul de la probabilité d'encombrement se fait à partir du pourcentage de temps durant lequel la quantité de ressources utilisées est égale à la capacité du système. On va donc utiliser le paramètre `NbUsedResources` dans l'algorithme suivant pour déterminer ce pourcentage.

```

count = 0;
capacity = 11;
t = 0:0.01:2000;
nbUsedResourcesResampled = resample(NbUsedResources,t,'zoh');
usedResRsmpl = nbUsedResourcesResampled.Data;
[timeResRsmpl,y]=size(usedResRsmpl);
for i = 1:timeResRsmpl
    if usedResRsmpl(i,1)>= capacity
        count = count + 1;
    end
end
pCongestion = count/timeResRsmpl;
disp('Capacity = 11');
fprintf('The percentage of time the capacity is reached : %0.3g \n',
pCongestion);

```

Nous avons trouvé les pourcentages suivants pour une capacité de 11 ressources, et pour une capacité de 17 ressources à 5000 unités de temps de la simulation.

```

>>
Capacity = 11
The percentage of time the capacity is reached : 0.1073

```

>>

Capacity = 17

The percentage of time the capacity is reached : 0.0166

Ce sont les probabilités qu'un utilisateur quelconque, arrivant à un instant hasard, trouve toutes les ressources occupées.

On a alors trouvé les résultats théoriques de notre modèle explicité par les courbes des figures 2a et 2b qui sont respectivement 10% et 1% pour les capacités 11 et 17. Ces résultats valident le modèle ainsi établis.

IV. Discussions

Les résultats ainsi obtenus peuvent être utilisés dans beaucoup de domaines ayant des services à offrir à des utilisateurs à partir des ressources du fournisseur de service. A titre d'exemple, il s'applique pour le dimensionnement des codes (ressources) d'un réseau de communication mobile de la troisième génération (réseau 3G) qui partagent ces codes à travers ses utilisateurs [17]. Ce type de réseau fournit de multiple de services, comme par exemple :

- Service voix AMR : nécessitant 2 codes
- Service VP : nécessitant 8 codes
- Service PS R99 DL : nécessitant 8 codes
- Service HSDPA : nécessitant 16 codes

Ces résultats restent valables pour le cas des ressources non discrètes. Dans ce cas, la quantité de ressources initialement disponible dans le système est un nombre réel strictement positif. Les quantités des ressources nécessaires à chaque service i peuvent être aussi des nombres réels strictement positifs (non nécessairement un entier naturel non nul). La variable aléatoire désignant la quantité de ressources occupées par les utilisateurs est alors une variable aléatoire réelle positive. Et les expressions des équations (5) et (6) se portent déjà avec les fonctions de répartitions qui caractérisent les variables aléatoires réelles. Et dans ce cas, le domaine d'application est plus vaste si on ne cite que les systèmes de communications modernes partageant des bandes passantes, des puissances, des spectres, ...

Les résultats fournis sont limités dans le cadre des quantités de ressources déterministes par service. On peut les étendre aux cas des quantités aléatoires qu'on rencontre souvent dans les cas réels. L'expression analytique pour le dimensionnement devient plus complexe du fait qu'elle doit être fonction des caractéristiques (fonction densité ou fonction de répartition) des demandes de ressources pour chacun des services que le système peut offrir.

V. Conclusion

Un système multi-service partage ses ressources à travers ses utilisateurs selon les besoins de chacun pour chaque service demandé. Les utilisateurs arrivent sur ce système selon des processus de Poisson caractéristiques de chaque classe de service. Avec des quantités déterministes de ressources nécessaires pour chaque service, le système est assimilable à une file d'attente multi-service M/GI/C en terme de nombre de clients. Deux cas de capacités en terme de quantités de ressources ont été considérés : le cas infini qui nous a permis d'évaluer la distribution d'occupation des ressources de la file d'attente, et le cas fini dont nous avons tiré une probabilité d'encombrement à partir du cas infini. Les expressions analytiques établies nous ont rendus possible le dimensionnement du système en déterminant la quantité totale de ressources nécessaires au serveur étant une probabilité d'encombrement admissible. A partir du résultat théorique et de l'étude de cas présenté, on a trouvé qu'il est bien possible de déterminer cette quantité totale nécessaire puisqu'à partir d'une certaine valeur de la capacité, la caractéristique du système en terme de nombre de ressources occupées ne présente pas beaucoup de variations. Les résultats obtenus se limitent au cas déterministe des ressources par service, et les cas aléatoires peuvent être considérés comme future extension des travaux que nous avons effectués.

Références

- [1] Erlang, A.K. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler (Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges). *Elektroteknikerens*. **1917**, 13, 5-13.
- [2] Palm, C. Table of the Erlang Loss Formula. *Telefonaktiebolaget LM Ericsson*, Stockholm. **1947**.
- [3] Joys, L.A. *Comments on the Engset and Erlang Formulae for Telephone Traffic Losses*. Thesis. Report TF No. 25,/71, Research Establishment, The Norwegian Telecommunications Administration. **1971**.
- [4] Iversen, V.B. The A-formula. *Teleteknik, English ed.* **1980**, 23 (2), 64-79.
- [5] Vaultot, E.; Chaveau, J. Extension de la formule d'Erlang au Cas où le Trafic est Fonction du Nombre d'Abonnés Occupés (Extension of the Erlang formula to the case where the traffic depends on the number of busy subscribers). *Annales de Télécommunications*. **1949**, 4 (8), 319-324.
- [6] Iversen, V.B. The exact evaluation of multi-service loss system with access control. *Teleteknik, English ed.* **1987**, 31 (2), 56-61.
- [7] Hanczewski, S.; Stasiak, M.; Weissenberg, J. A queueing model of a multi-service system with state-dependent distribution of resources for each class of calls. *IEICE Trans. Commun.* **2004**, E97-B (8), 1592-1605.

- [8] Hanczewski, S.; Kmiecik, D.; Stasiak, M.; Weissenberg, J. Multi-service queuing system with elastic traffic. *Proc. IEICE Gen. Conf.* **2016**, BS-3-18
- [9] Bonald, T.; Proutière, A.; Virtamo, J. A queuing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems.* **2006**, 53 (1), 65-84
- [10] Haddad, J.-P.; Mazumdar, R.R. Congestion in large balanced multirate networks. *Queueing Systems.* **2013**, 74 (2), 333-368
- [11] Hanczewski, S.; Kaliszan, A.; Stasiak, M. Convolution Model of a Queueing System with the cFIFO Service Discipline. *Hindawi Publishing Corporation, Mobile Information Systems.* **2016**, 2016, Article ID 2185714, 1-15
- [12] Crane, H.; McCullagh, P. Poisson superposition processes. *Journal of Applied Probability.* **2015**, 52 (4), 1013-1027
- [13] Tijms, H. C.; Van Hoorn, M. H.; Federgruen, A. Approximations for the Steady-State Probabilities in the M/G/c Queue. *Advances in Applied Probability.* **1981**, 13 (1), 186–206
- [14] Fry, T.C. *Probability and its Engineering Uses.* New York. **1928**
- [15] Molina, E.C. Application of the Theory of Probability to Telephone Trunking Problems. *The Bell System Technical Journal.* **1927**, 6 (3), 461–494
- [16] Iversen, V.B. *Teletraffic engineering handbook*, ITU–D SG 2/16 & ITC. **2001**
- [17] Bernhard, H.W.; Seidenberg, P.; Althoff, M.P. *UMTS : The Fundamentals.* John Wiley & Sons. **2003**