

# Bornes stochastiques pour le temps de réponse des services Web composites

Samir Youcef

LAMSADE, Université de Paris Dauphine  
Place du Maréchal de Lattre de Tassigny  
75775 Paris Cedex 16, France

samir.youcef@lamsade.dauphine.fr

## Keywords

Services Web, Qualité de Service, Borne stochastique

## 1. INTRODUCTION

Un service Web désigne essentiellement une application (ou un programme) mise à disposition sur Internet par un fournisseur de service, et accessible par les clients à travers des protocoles Internet standards. Cependant, Pour certains types d'application, il est nécessaire de combiner un ensemble de services Web en services plus complexes (services Web agrégés ou composites) afin de répondre à des exigences plus complexes [4].

La qualité de service (**QoS**) est une combinaison de plusieurs mesures telles que le débit, la disponibilité et le temps de réponse [3]. La dégradation de ces mesures dans des applications critiques peut engendrer de sérieuses conséquences [1].

L'évaluation de performance des services Web est un problème difficile à cause du fait que les méthodes d'analyse ne se montrent que de peu d'intérêt lorsque l'on prend en compte leur fonctionnement. Notre objectif est justement de proposer une nouvelle approche de mesurer les contraintes de QoS des services Web, en l'occurrence le temps de réponse. Cette approche repose sur la comparaison stochastique. Nous proposons des modèles bornants pour encadrer le temps de réponse d'un service Web composite.

## 2. DESCRIPTION DU MODÈLE

Nous considérons, dans cette section, un service Web composé de  $n$  services élémentaires (notés de 1 à  $n$ ). Un service élémentaire  $i$  est modélisé par une file d'attente  $M/M/1/B_i$  où  $B_i - 1$  est la capacité de la file  $i$  (voir figure 1).

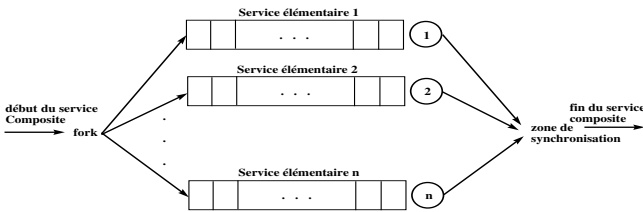


Figure 1: Services Web composites

Nous pouvons alors associer une chaîne de Markov à temps continu,  $\{N(t), t \geq 0\}$ , à ce système, dont l'espace d'état est décrit par le vecteur  $N = (N_1, N_2, \dots, N_n)$ , avec  $N_i$  le nombre de clients dans la file  $i$ . Ainsi, le nombre de clients

en attente dans la zone de synchronisation est donné par  $\sum_{i=1}^n (M - N_i)$ , avec  $M = \max\{N_1, \dots, N_n\}$ . Le temps moyen de réponse d'un service composite est alors donné par :

$$T = \frac{\bar{N}}{\beta} \quad (1)$$

avec  $\bar{N}$  et  $\beta = n\lambda$  sont respectivement le nombre moyen de services élémentaires dans le système et le taux des arrivées. Le nombre moyen de services élémentaires dans le système est donné par  $\bar{N} = \sum_{e \in E} \pi(e)f(e)$ , où  $E$  est l'ensemble des états de la chaîne de Markov  $\{N(t), t \geq 0\}$ . La fonction  $f(e)$  peut être exprimée comme suit :

$$f(e) = n \times M$$

## 3. ORDRE STOCHASTIQUE

La méthode des ordres stochastiques pour l'étude des performances consiste à encadrer l'indice de performance considéré. Dans ce travail, nous avons utilisé l'ordre stochastique fort noté  $\preceq_{st}$ . Il existe plusieurs ordres stochastiques tel que  $\preceq_{st}$ ,  $\preceq_{icx}$ ,  $\preceq_{wk}$ , ... [5].

DEFINITION 1. Soit  $X, Y$  deux variables aléatoires positives à valeurs dans  $E$ .

$$X \preceq_{st} Y \iff \forall a \in E, P(X > a) \leq P(Y > a).$$

PROPOSITION 1. Soit  $X, Y$  deux variables aléatoires positives à valeur dans  $E$ , alors :

$$X \preceq_{st} Y \iff E(f(X)) \leq E(f(Y)) \forall f \text{ fonction croissante.}$$

Dans le cas de l'utilisation de l'ordre  $\preceq_{st}$ , on a les propriétés suivantes connues sous le nom de couplage des trajectoires :

THEOREM 1.  $P \preceq_{st} Q$  si et seulement s'il existe des variables aléatoires  $X$  et  $Y$  qui ont respectivement les distributions  $P$  et  $Q$ , tel que  $X \preceq Y$  presque sûrement.

Les chaînes de Markov associées aux différents systèmes étant différentes, nous devons alors définir un espace d'états commun noté  $\epsilon$ . Cet espace est défini comme suit :

DEFINITION 2. Soit  $X = (X_1, \dots, X_n) \in \epsilon$  tels que :

$$X_i = \begin{cases} 1, & \text{s'i y a des clients dans la file } i, \\ 0, & \text{sinon.} \end{cases}$$

Définissons, maintenant, le préordre  $\preceq$  sur  $\epsilon$  que nous avons utilisé pour la comparaison.

DEFINITION 3. Soient  $X = (X_1, \dots, X_n)$ ,  $Y = (Y_1, \dots, Y_n) \in \epsilon$  tel que :

$$\begin{cases} X \preceq Y \text{ si } X_1 \leq Y_1, \dots, X_n \leq Y_n \\ X = Y \text{ si } X_1 = Y_1, \dots, X_n = Y_n, \end{cases}$$

Cet ordre a été choisi pour la comparaison de l'indice de performance qui nous intéresse, c'est à dire le temps de réponse. Intuitivement, si  $X$  et  $Y$  sont deux états tel que  $X \preceq Y$ , nous pouvons dire que le temps de réponse à l'état  $Y$  est supérieur au temps de réponse à l'état  $X$ .

L'espace de comparaison ainsi que le préordre définis sur cet espace permet de faire une comparaison des images des systèmes considérés, c'est à dire le système original et les deux systèmes bornants, au sens de l'ordre  $\preceq_{st}$ . Pour cela, nous utilisons le couplage des trajectoires. Définissons, alors, les applications surjectives qui projettent les espaces d'états des différents systèmes sur  $\epsilon$ . Notons par  $N^{inf}$ ,  $N$  et  $N^{sup}$  les espaces d'états respectivement du système fournissant la borne inférieure, du système original et du système fournissant la borne supérieure. Ces applications sont définies comme suit  $\phi : N^{inf} \rightarrow \epsilon$ ,  $\alpha : N \rightarrow \epsilon$  et  $\beta : N^{sup} \rightarrow \epsilon$ . Il reste donc à montrer les relations d'ordre stochastique entre les images des chaînes définies, c'est à dire :

$$\phi(N^{inf}(t)) \preceq_{st} \alpha(N(t)) \preceq_{st} \beta(N^{sup}(t)).$$

#### 4. BORNE SUPÉRIEURE

Que ce soit pour la borne supérieure ou inférieure notre point de départ était les travaux de Fourneau [2]. Nous simplifions le système en supprimant certains tampons du système original (ces files sont réduite à des services) et en supposant qu'un client servi est immédiatement remplacé par un autre client. Une autre façon de voir est de supposer que ces files ne sont jamais vides. Ainsi, la résolution de la chaîne de Markov associée à ce système est plus simple, que celle associée au système original, puisqu'on ne s'intéresse pas à l'évolution de ces files. Les capacités des autres files restent inchangées (i.e  $B_i^{sup} = B_i$ ). Notons  $N^{sup} = (N_1^{sup}, \dots, N_n^{sup})$  l'état du système où  $N_i^{sup}$  est le nombre de requêtes à traiter par le service élémentaire  $i$ . Soit  $S$  l'ensemble des tampons supprimés, alors :

$$\forall t \geq 0, N_j^{sup}(t) = \begin{cases} 1, & \text{si } j \in S, \\ N_j(t), & \text{sinon.} \end{cases}$$

Soit  $N_i^{sup} = (N_{i,1}^{sup}, \dots, N_{i,n}^{sup})$  l'état du système immédiatement après l'événement  $i$ . Si le  $(i+1)^{ème}$  événement est l'arrivée d'une requête, alors :

$$N_{i+1,j} = \begin{cases} 1, & \text{si } j \in S, \\ N_{i,j}^{sup} + \mathbf{1}_{\{N_{i,j}^{sup} < B_j\}}, & \text{sinon.} \end{cases}$$

Si le  $(i+1)^{ème}$  événement est un service du serveur  $h$  :

$$N_{i+1,j} = \begin{cases} 1, & \text{si } j \in S, \\ N_{i,j}, & \text{si } j \neq h \text{ et } j \notin S, \\ (N_{i,h} - 1)^+, & \text{si } j=h \text{ et } j \notin S. \end{cases}$$

où  $(x)^+ = \max(0, x)$ .

THEOREM 2. Si  $\alpha(N(0)) \preceq_{st} \beta(N^{sup}(0))$ , alors  $\alpha(N(t)) \preceq_{st} \beta(N^{sup}(t))$ .

La comparaison stochastique des ces chaînes par les fonctions d'états  $\alpha$  et  $\beta$  permet de déduire une comparaison stochastique des distributions stationnaires :

PROPOSITION 2.  $\alpha(\Pi) \preceq_{st} \beta(\Pi^{sup})$  où  $\Pi$  désigne la distribution stationnaire.

Ainsi, on en déduit l'inégalité  $T \leq T^{sup}$  sur le temps de réponse, qui est une conséquence directe de la proposition 2 et de la fonction 1 (fonction croissante par rapport à l'espace de comparaison  $\epsilon$ ).

#### 5. BORNE INFÉRIEURE

Pour la borne inférieure, nous simplifions le système original en considérant le même réseau mais avec des capacités des files inférieures. Nous considérons donc un système où :

$$B_i^{inf} \leq B_i, 1 \leq i \leq n$$

Soit  $N^{inf} = (N_1^{inf}, \dots, N_n^{inf})$  l'état du système où  $N_i^{inf}$  le nombre de clients dans la  $i^{ème}$  file.  $N_i^{inf} = (N_{i,1}^{inf}, \dots, N_{i,n}^{inf})$  l'état du système immédiatement après l'événement  $i$ . Si le  $(i+1)^{ème}$  événement est l'arrivée d'une requête, alors :

$$N_{i+1,j}^{inf} = N_{i,j}^{inf} + \mathbf{1}_{\{N_{i,j}^{inf} < B_j^{inf}\}}$$

Si le  $(i+1)^{ème}$  événement est un service du serveur  $j$  :

$$N_{i+1,j}^{inf} = \begin{cases} N_{i,j}^{inf}, & \text{si } j \neq h, \\ (N_{i,j}^{inf} - 1)^+, & \text{sinon.} \end{cases}$$

THEOREM 3. Si  $\phi(N^{inf}(0)) \preceq_{st} \alpha(N(0))$ , alors  $\phi(N^{inf}(t)) \preceq_{st} \alpha(N(t))$ .

La comparaison stochastique des ces chaînes par les fonctions d'états  $\phi$  et  $\alpha$  permet de déduire une comparaison stochastique des distributions stationnaires :

PROPOSITION 3.  $\phi(\Pi^{inf}) \preceq_{st} \alpha(\Pi)$  où  $\Pi$  désigne la distribution stationnaire.

Ainsi, on en déduit l'inégalité  $T^{inf} \leq T$  sur le temps de réponse, qui est un conséquence directe de la proposition 3 et de la fonction 1 (fonction croissante par rapport à l'espace de comparaison  $\epsilon$ ).

#### 6. CONCLUSION

Nous avons présenté, dans ce résumé, des modèles bornants pour le temps de réponse des services Web composés. L'intérêt de notre approche provient dans le fait que l'on peut trouver un compromis entre la qualité de la borne et la complexité de résolution.

#### 7. REFERENCES

- [1] I. Silva-Lepe F. Curbera and S. Weerawarana. On the integration of heterogeneous web service partners. *In International Semantic Web Conference*, 2001.
- [2] L. Mokdad et N. Pekergin J.M. Fourneau. Bounding the loss rates in a multistage atm switch. *9th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 193–205, (TOOLS'97), Saint-Malo, France, LNCS 1245.
- [3] D. Menasce. Response time analysis of composite web services. *IEEE Internte Computing*, (1), January/February 2004.
- [4] S. Narayanan. Simulation, verification and automated composition of web services. in proceedings of the eleventh international world wide web. *conference, Honolulu, HI*, 2002.
- [5] D. Stoyan. Comparison methods for queue and other stochastic models. *J-Wiley and Son*, 1976.