Université Abderrahmane Mira de Bejaia

Unité de Recherche LaMOS

Modélisation des Systèmes Réactifs (MSR'19)

Un Nouveau Modèle Analytique pour le Calcul de l'Élasticité dans le Cloud Computing

Assia OUTAMAZIRT[1] Kamel BARKAOUI[2] et Djamil AISSANI[1]

[1] Unité de Recherche LaMOS de l'Université de Bejaia
[2] CEDRIC, CNAM, Paris, France

Mercredi 13 Novembre 2019

Plan de travail

Introduction

Introduction

L'émergence du Cloud Computing a donné naissance à un nouveau type de systèmes autonomiques dits systèmes Cloud élastiques, dans lesquels l'élasticité est un principe de conception clé.

Avantage: Capacité de gérer la quantité de ressources informatiques utilisées selon la charge de travail actuelle tout en maintenant la performance et la qualité de service.

Introduction

Introduction

L'émergence du Cloud Computing a donné naissance à un nouveau type de systèmes autonomiques dits systèmes Cloud élastiques, dans lesquels l'élasticité est un principe de conception clé.

Avantage: Capacité de gérer la quantité de ressources informatiques utilisées selon la charge de travail actuelle tout en maintenant la performance et la qualité de service.

Problématique et motivation

Dans les environnements Cloud (Amazon EC2, Microsoft Azure,. . .), de nombreuses solutions académiques et industrielles ont été proposées pour gérer la quantité de ressources informatiques utilisées de manière automatique, de telle façon à ce que les ressources fournies soient conformes à la demande du système.

Limites

- Des limites en termes de contrôle et évaluation de l'élasticité,
- Ce qui influe sur la disponibilité, la performance des systèmes et la qualité de service.

Problématique et motivation

Comment peut-on assurer une bonne gestion de la consommation de ressources informatiques tout en maintenant une bonne qualité de service dans un système Cloud élastique?

Peut-on adopter des approches analytiques qui peuvent analyser et calculer l'élasticité dans le Cloud Computing d'une manière précise?

Outils

Modélisation stochastique $\rightarrow$ Files dattente.

Outils

Modélisation stochastique $\rightarrow$ Files dattente.

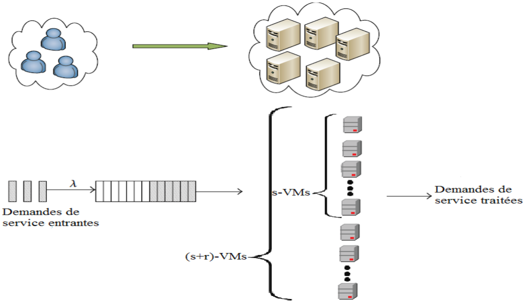
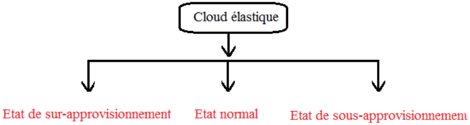
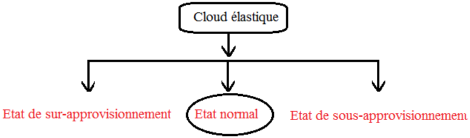
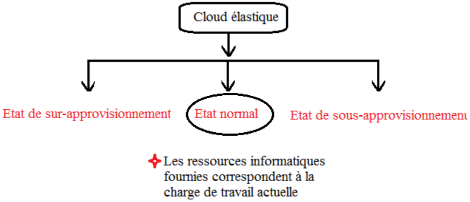
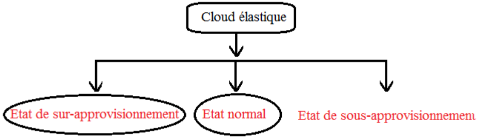


Figure: Modélisation d'un sysème Cloud élastique sous la forme d'un modèle de files d'attente $M/M/s + r/k$, où le nombre de serveurs s'adapte à la charge de travail actuelle.

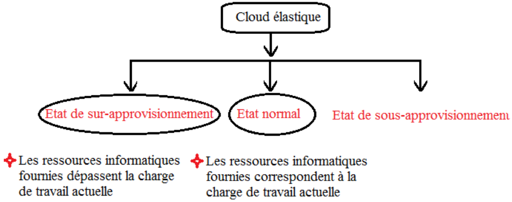

Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnemen

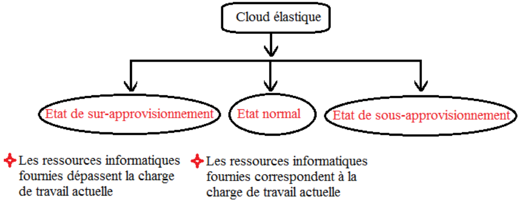

Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnement

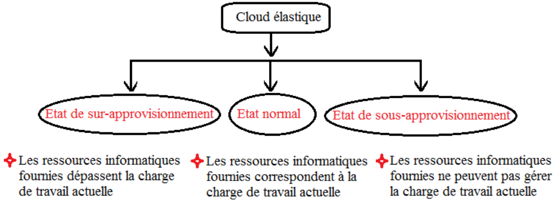

Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnemen
Les ressources informatiques fournies correspondent à la charge de travail actuelle



- Les ressources informatiques fournies correspondent à la charge de travail actuelle


Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnement
Les ressources informatiques fournies dépassent la charge de travail actuelle
Les ressources informatiques fournies correspondent à la charge de travail actuelle


Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnement
Les ressources informatiques fournies dépassent la charge de travail actuelle
Les ressources informatiques fournies correspondent à la charge de travail actuelle


Cloud élastique
Etat de sur-approvisionnement
Etat normal
Etat de sous-approvisionnement
Les ressources informatiques fournies dépassent la charge de travail actuelle
Les ressources informatiques fournies correspondent à la charge de travail actuelle
Les ressources informatiques fournies ne peuvent pas gérer la charge de travail actuelle

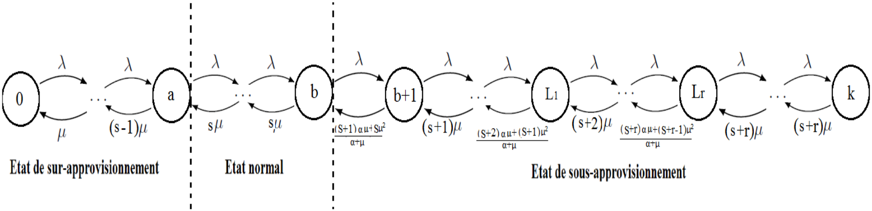


Figure: Diagramme de transition.

$$\lambda\pi_0 = \mu\pi_1 \tag{1}$$

$$(\lambda + min(s-1,i)\mu)\,\pi_i = \lambda\pi_{i-1} + min(s-1,i+1)\mu\pi_{i+1} \text{ pour } 0 \leq i \leq a \tag{2}$$

$$(\lambda + s\mu)\pi_i = \lambda\pi_{i-1} + s\mu\pi_{i+1} \text{ pour } a < i \leq b \tag{3}$$

$$(\lambda + (s+j-1)\mu)\pi_{L_j-1} = \lambda\pi_{L_j-2} + \frac{(s+1)\alpha\mu + c(s+j-1)\mu^2}{\alpha+\mu}\pi_{L_j} \text{ pour } j = \overline{1, r-1} \tag{4}$$

$$\left(\lambda + \frac{(s+j)\alpha\mu + (s+j-1)\mu^2}{\alpha+\mu}\right)\pi_{L_j} = \lambda\pi_{L_j-1} + (s+j)\mu\pi_{L_j+1} \text{ pour } j = \overline{1, r} \tag{5}$$

$$(\lambda + (s+j-1)\mu)\pi_i = \lambda\pi_{i-1} + (s+j-1)\mu\pi_{i+1} \text{ pour } j = \overline{2, r}, \; L_{j-1}+1 \leq i \leq L_j - 2 \tag{6}$$

$$(\lambda + (s+r)\mu)\pi_i = \lambda\pi_{i-1} + (s+r)\mu\pi_{i+1} \text{ pour } L_r + 1 \leq i < k \tag{7}$$

$$(s+r)\mu\pi_k = \lambda\pi_{k-1} \tag{8}$$

$$\pi_i = \begin{cases} \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i \pi_0 \quad \text{pour } 0 \le i \le a+1 \le s; \\ \frac{1}{s^{i-s}s!}\left(\frac{\lambda}{\mu}\right)^i \pi_0 \quad \text{pour } s+1 \le i \le b; \\ \left(\frac{1}{s^{b-s}s!}\left(\frac{\lambda}{\mu}\right)^b\right)\left(\prod_{i'=1}^{j}\frac{\lambda(\alpha+\mu)}{(s+i')\alpha\mu+(s+i'-1)\mu^2}\right)\left(\prod_{\kappa=0}^{j-1}\frac{1}{(s+\kappa)^\kappa}\right)\left(\frac{1}{(s+j)^{i-L_j}}\right)\left(\frac{\lambda}{\mu}\right)^{i-b-j}\pi_0 \\ \text{pour } j=1,\ i=L_1\text{, où } L_1=b+1 \text{ et pour } j=\overline{1,r-1},\ L_j+1\le i\le L_{j+1}-1; \\ \left(\frac{1}{s^{b-s}s!}\left(\frac{\lambda}{\mu}\right)^b\right)\left(\prod_{i'=1}^{j}\frac{\lambda(\alpha+\mu)}{(s+i')\alpha\mu+(s+i'-1)\mu^2}\right)\left(\prod_{\kappa=0}^{j-2}\frac{1}{(s+\kappa)^\kappa}\right)\left(\frac{1}{(s+j-1)^{(i-L_{j-1})-1}}\right)\left(\frac{\lambda}{\mu}\right)^{i-b-j}\pi_0 \\ \text{pour } j=\overline{2,r},\quad i=L_j; \\ \left(\frac{1}{s^{b-s}s!}\left(\frac{\lambda}{\mu}\right)^b\right)\left(\prod_{i'=1}^{r}\frac{\lambda(\alpha+\mu)}{(s+i')\alpha\mu+(s+i'-1)\mu^2}\right)\left(\prod_{\kappa=0}^{r-2}\frac{1}{(s+\kappa)^\kappa}\right)\left(\frac{1}{(s+r-1)^{(L_r-1)-L_{r-1}}}\right)\left(\frac{1}{(s+r)^{i-L_r}}\right) \\ \left(\frac{\lambda}{\mu}\right)^{i-b-r}\pi_0 \text{pour } L_r+1\le i\le k. \end{cases} \quad (9$$

La probabilité que le système se trouve dans un état de sur-approvisionnement est:

$$p_{\text{over}} = \sum_{i=0}^{a} \pi_i. \tag{10}$$

La probabilité que le système soit à l'état normal est:

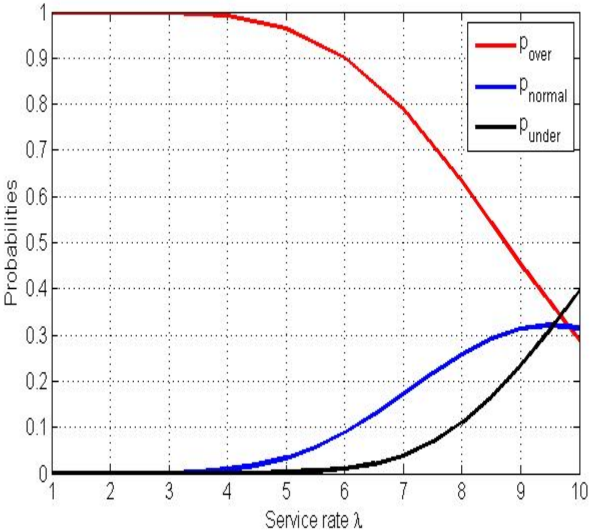
$$p_{\text{normal}} = \sum_{i=a+1}^{b} \pi_i. \tag{11}$$

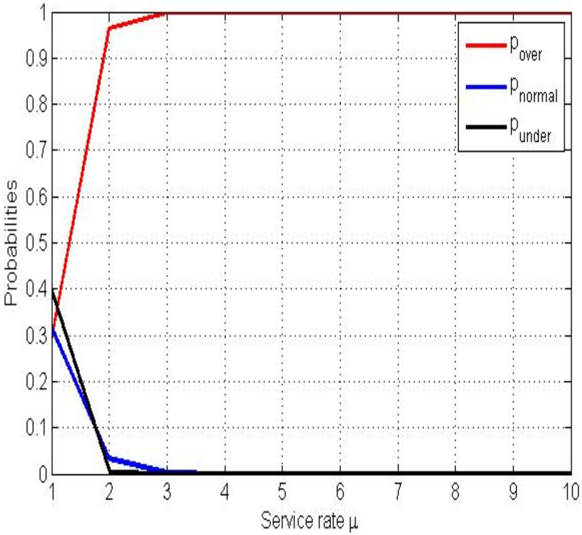
La probabilité que le système se trouve dans un état de sous-approvisionnement est:

$$p_{\text{under}} = \sum_{i=b+1}^{k} \pi_i. \tag{12}$$

En utilisant les probabilités 10, 13 et 12, la valeur de l'élasticité peut être obtenue:

$$\text{Élasticité} = \sum_{i=a+1}^{b} \pi_i = 1 - \left(\sum_{i=0}^{a} \pi_i + \sum_{i=b+1}^{k} \pi_i\right). \tag{13}$$


Probabilities
Service rate λ
p_{over}
p_{normal}
p_{under}
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10


Probabilities
Service rate μ
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
p_{over}
p_{normal}
p_{under}

Conclusion et Perspectives

- Nous nous sommes intéressés à la modélisation analytique des systèmes Cloud élastiques.
- En considérant une définition quantitative et formelle de l'élasticité dans le Cloud Computing, nous avons développé un modèle analytique pour étudier l'élasticité en traitant un système Cloud élastique (une Cloud plate-forme) comme un modèle de files d'attente $M/M/s+r/k$ où le nombre de serveurs actifs dépend du nombre de demandes de service présentes dans le système.
- Pour analyser et calculer la valeur de l'élasticité d'une manière précise, nous avons effectué une étude quantitative d'analyse de l'état stationnaire de notre modèle.

Conclusion et perspectives

Dans la continuité de ce travail, nous envisageons les perspectives suivantes:

- Extension du modèle $M/M/s+r/k$ en considérant le **processus MMPP** comme processus d'arrivée (Afin de tenir compte de la variation des taux d'arrivées des demandes de service Cloud dans le temps).
- Analyse mathématique du modèle $MMPP/G/c/k$ en considérant le cas des **taux de services variables** pour tenir compte des demandes des utilisateurs de service Cloud aux différents nombres de ressources pour différentes durées.
- Analyse mathématique du modèle d'attente $MMPP/G/c/k$ avec **priorité et dépendance entre les arrivées** des demandes de service Cloud.

Merci de votre attention.