



HAL
open science

Criblage virtuel sur grille de composés isolés au Vietnam

Quang Bui The, Trung-Tung Doan, van Son Nguyen, Vincent Breton, Hong Quang Nguyen, Quoc Long Pham, Mai Huong Le

► **To cite this version:**

Quang Bui The, Trung-Tung Doan, van Son Nguyen, Vincent Breton, Hong Quang Nguyen, et al.. Criblage virtuel sur grille de composés isolés au Vietnam. Rencontres Scientifiques France Grilles 2011, Sep 2011, Lyon, France. hal-00660142

HAL Id: hal-00660142

<https://hal.science/hal-00660142>

Submitted on 16 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Criblage virtuel sur grille de composés isolés au Vietnam

T. Quang BUI (1), T. Tung DOAN (2), V. Son NGUYEN (3), Vincent BRETON (4), H. Quang NGUYEN(5), Q. Long PHAM(6), M. Huong LE (7)

(1) buihequang@gmail.com, Laboratoire MSI, Institut de la Francophonie pour l'Informatique, 42 Ta Quang Buu, Hanoi, Vietnam

(2) dttung@gmail.com, Laboratoire MSI, Institut de la Francophonie pour l'Informatique, 42 Ta Quang Buu, Hanoi, Vietnam

(3) son.nguyen.pt@gmail.com, L'Institut de Chimie des Produits Naturels de l'Académie des Sciences du Vietnam (INPC), Hanoi, Vietnam

(4) bretun@clermont.in2p3.fr, Laboratoire de Physique Corpusculaire, CNRS/IN2P3, 24 avenue des Landais, BP10448, 63000 Clermont-Ferrand, France

(5) nguyen.hong.quang@auf.org, Laboratoire MSI, Institut de la Francophonie pour l'Informatique, 42 Ta Quang Buu, Hanoi, Vietnam

(6) phamquoclong1960@gmail.com, L'Institut de Chimie des Produits Naturels de l'Académie des Sciences du Vietnam (INPC), Hanoi, Vietnam

(7) lehuong00@gmail.com, L'Institut de Chimie des Produits Naturels de l'Académie des Sciences du Vietnam (INPC), Hanoi, Vietnam

Overview

The Institute of Natural Products Chemistry of Vietnam has collected data of Vietnamese natural products for discovery of medicine. Now, they need an information system for managing and processing these data of natural product of Vietnam. They want to use the computing power for virtual screening on data. In this project, we use the software Autodock[1] for virtual screening and the WISDOM Production Environment (WPE)[2] to manage jobs on the grid infrastructure. We present in this paper the goal of this project, deployment of system WPE on the grid computing for virtual screening and several performance tests of WPE to evaluate its capacity.

Enjeux scientifiques, besoin de la grille :

Au Vietnam, la médecine traditionnelle a une longue histoire de développement. Elle utilise des parties de plantes médicinales naturelles, par exemple, racine, fleur, tige d'un arbre, feuille... Il y a actuellement environ 4000 plantes médicinales qui ont été enregistrés au Vietnam. L'Institut de Chimie des Produits Naturels de l'Académie des Sciences du Vietnam (INPC) collecte des échantillons issus de la biodiversité locale et détermine la structure 3D des molécules isolées. L'enjeu pour cet institut est de constituer une base de données des échantillons et de mettre en place une chaîne de traitement des échantillons pour déterminer sur quelles cibles biologiques les produits isolés sont potentiellement actifs.

Dans la chaîne de traitement des échantillons, l'étape de "docking" a pour but de simuler l'interaction entre les molécules par ordinateur. On prend des résultats obtenus de cette étape pour prédire les propriétés des produits isolés. Nous avons besoin d'une grande puissance de calcul pour utiliser cette

technique parce que la donnée traitée est très grande et le même calcul effectué par 1 ordinateur prendrait des centaines d'années.

Développements, déploiement sur la grille :

Nous utilisons le logiciel Autodock [1] pour faire le “docking”. AutoDock est une suite d'outils de “docking” automatisé et libre de droit. Il est conçu pour prédire comment les petites molécules, telles que des substrats ou des médicaments candidats, se lient à un récepteur dont la structure 3D est connue.

Nous utilisons le système WISDOM Production Environment (WPE) [2] pour soumettre et gérer efficacement les tâches sur la grille. Le système WPE a été développé par le Laboratoire de Physique Corpusculaire – Clermont Ferrand (LPC). La conception du WPE est illustrée dans la figure suivante :

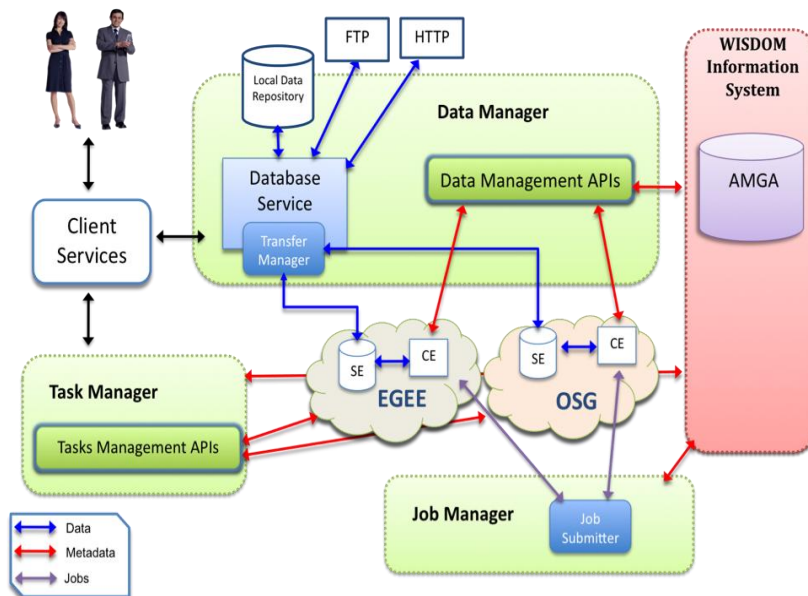


Figure 1 - La structure de WPE

Le WPE est composé de 4 éléments principaux. Le Job Manager soumet les jobs pilotes (ou agents) sur la grille pour réaliser les tâches dans le Task Manager (TM). Pour exécuter le Job Manager, on a besoin d'un certificat correspondant à l'organisation virtuelle où les jobs seront soumis. Les tâches sont enregistrées et gérées par le Task Manager. Un agent interagit avec le Task Manager pour récupérer une tâche puis l'exécute sur la grille. Les états des agents sont enregistrés dans le WIS (WISDOM Information Service). Le WIS est installé sur un serveur de base de données AMGA [6]. On peut surveiller le WPE et interroger ses états pendant son exécution. Le Data Manager a pour but de gérer des fichiers sur la grille en mode différé (batch). Le WPE est maintenant installé sur les serveurs de l'infrastructure de grille.

Résultats scientifiques :

Des tests ont été effectués sur les composants de WPE pour évaluer sa performance.

a. Test de performance du Job Manager

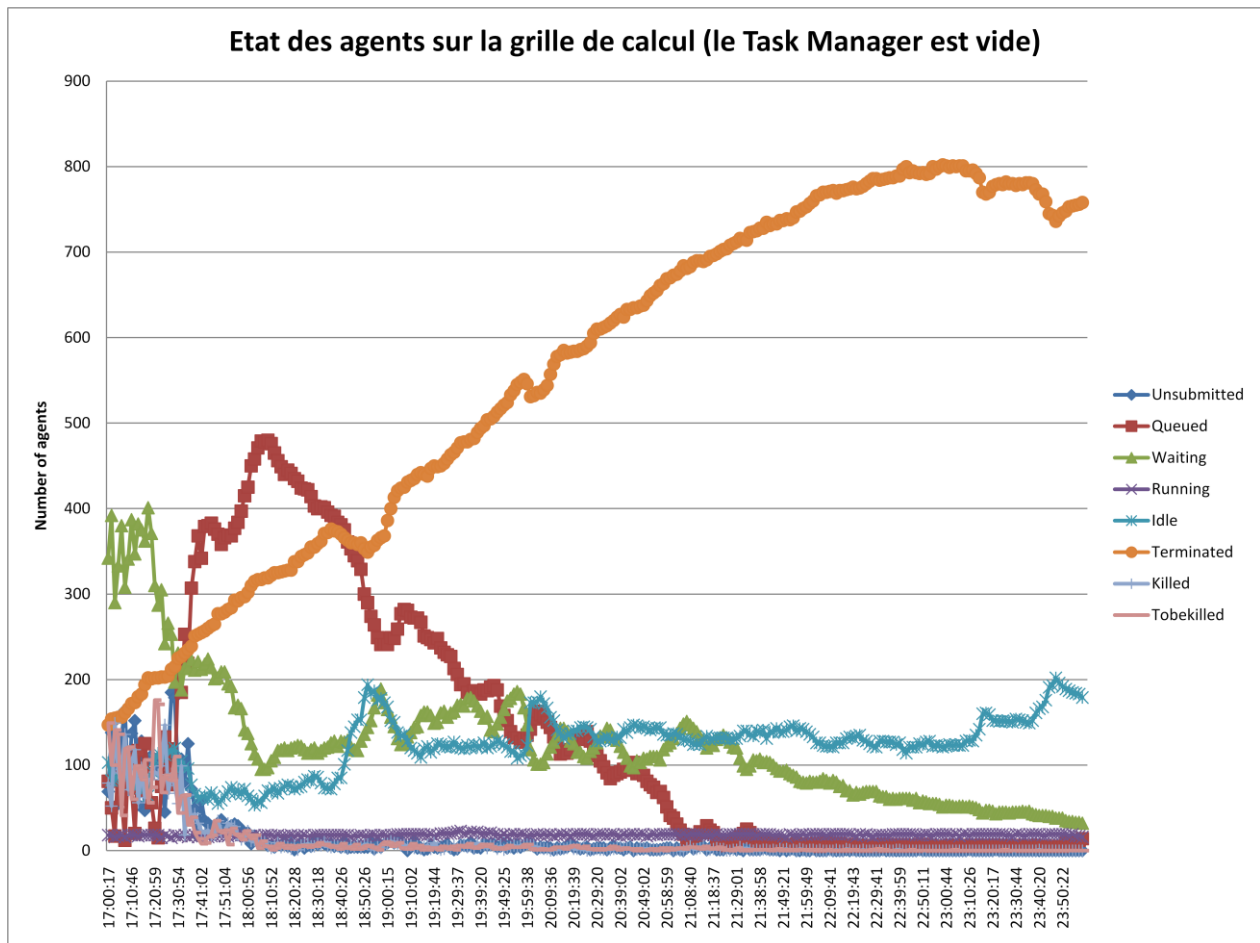


Figure 2 – L'évolution d'état des agents du WPE sur la grille de calcul dans le cas le Task Manager est vide

Dans le premier test, 1000 agents sont soumis sur la grille par le Job Manager. Il n'y a pas de tâche dans le Task Manager. Ensuite nous avons surveillé les états des agents du WPE. On peut voir l'état des agents du WPE dans la figure 2. Elle montre un problème de Job Manager : l'agent est terminé automatiquement dans le cas où il n'y a pas de tâches disponibles dans le Task Manager. De plus, le nombre d'agents d'état « idle » et « running » est faible. La performance du Job Manager n'est donc pas bonne dans ce cas parce que le pourcentage des agents disponibles pour exécuter des tâches est petit.

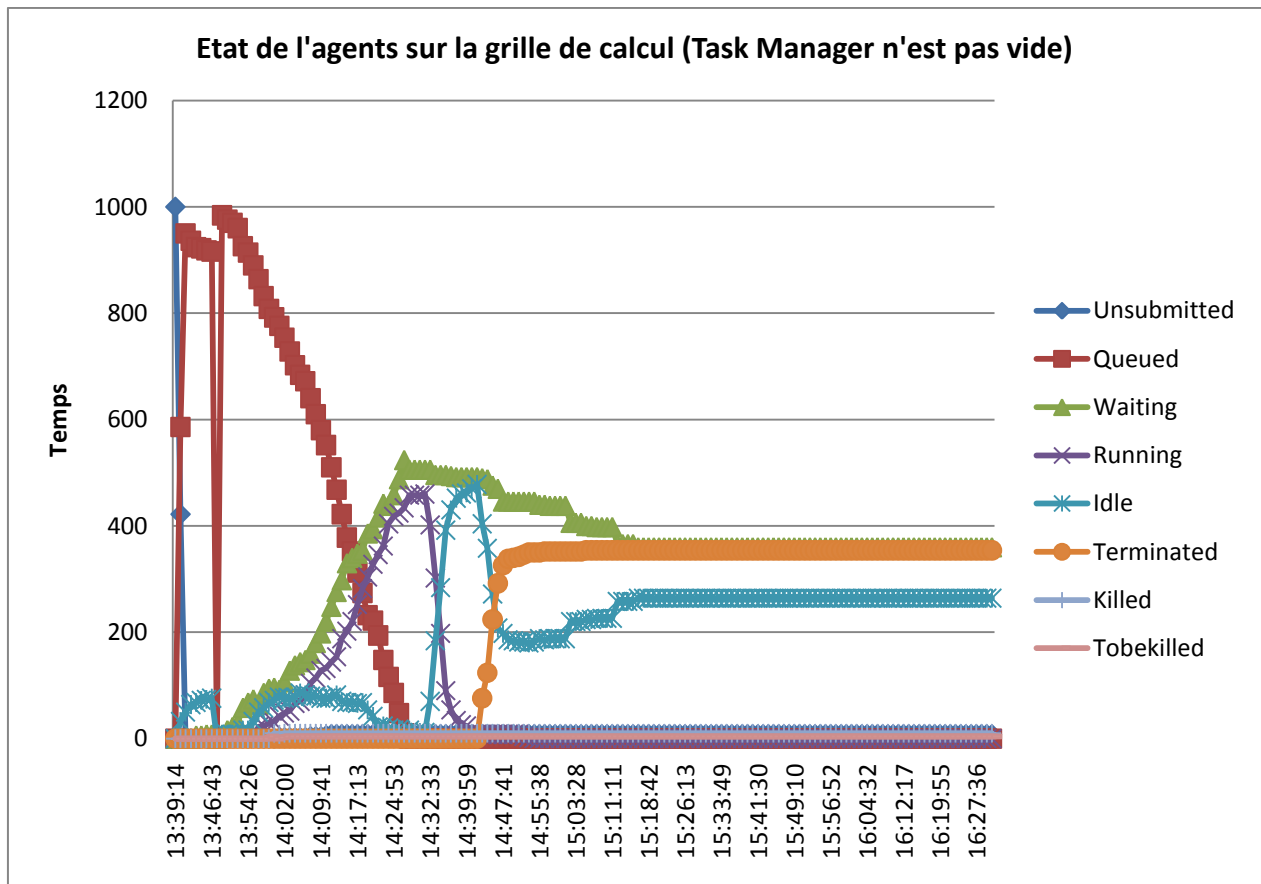


Figure 3 – L'évolution de l'état des agents du WPE sur la grille de calcul dans le cas le Task Manager n'est pas vide

Dans le deuxième test, le Job Manager a soumis 1000 agents sur la grille. Des tâches Autodock sont créées dans le Task Manager. La figure 3 représente l'évolution d'état des agents du WPE pendant son exécution. D'après le résultat obtenu, l'agent n'est pas terminé automatiquement et la performance du Job Manager est plus haute que cela dans le cas où le Task Manager est vide car dans ce cas là il y a plus d'agents d'état « idle » et « running ».

b. Test de performance du Task Manager

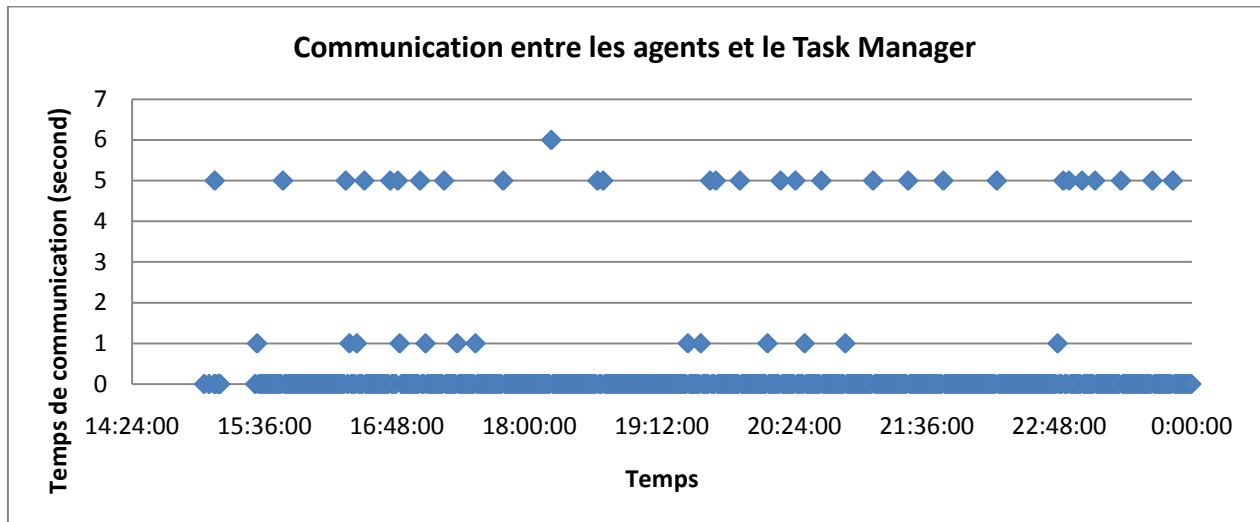


Figure 4 – Surveillance du temps de communication entre les agents et le Task Manager

Nous avons soumis 500 agents sur la grille de calcul. Ces agents interrogent le Task Manager pour savoir s’il y a des tâches prêtes dans le Task Manager. Le Task Manager va répondre à l’information de tâche. Dans ce test, nous mesurons le temps de la communication entre l’agent et le Task Manager. Nous avons surveillé la communication entre l’agent et le Task Manager pendant 9 heures. Dans ce test, le temps maximum est de 6 secondes, le temps minimum est presque nul, le temps moyen de 0,28 seconde.

c. Test de performance du WISDOM Information Service (WIS)

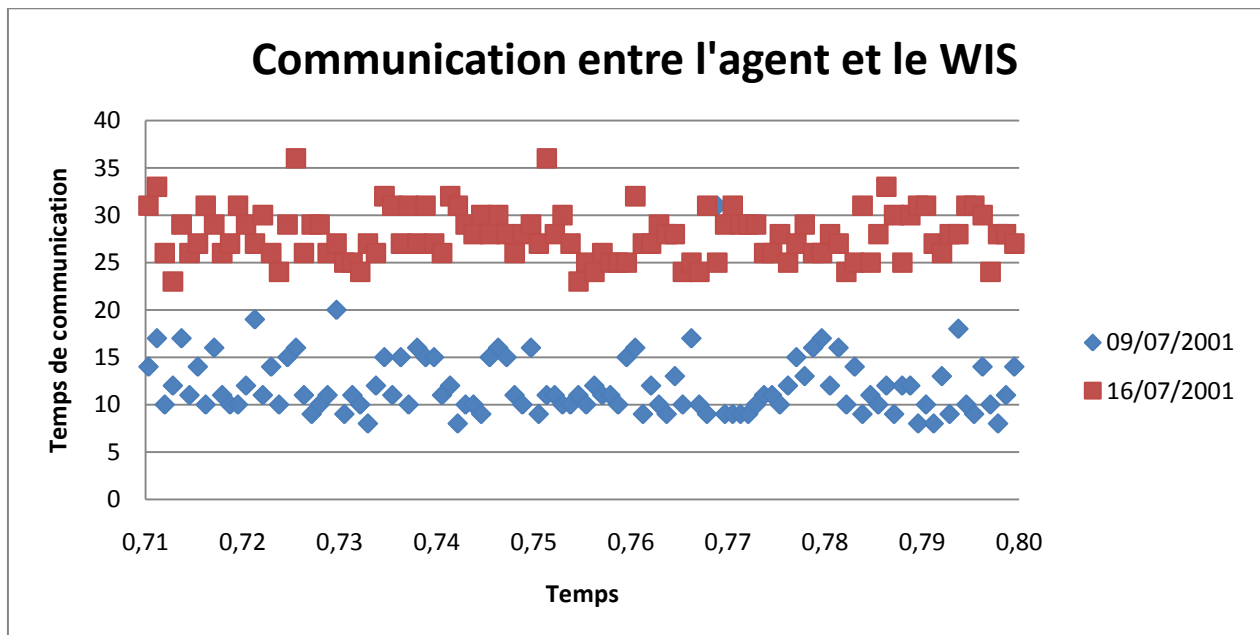


Figure 5 – Surveillance du temps de communication entre les agents et le WIS

Pour tester la performance du WIS, nous avons mesuré le temps de la communication entre les agents et le WIS pendant deux heures.

Dans le premier test, nous avons soumis 500 agents sur la grille de calcul le 9/7/2011. Pendant le cycle de la vie d'un agent, il enregistre toujours son état sur le WIS. Le temps maximum est de 31 secondes, le temps minimum de 7 secondes et le temps moyen de 11.8 secondes

Dans le deuxième test, nous avons soumis 500 agents le 16/7/2011. Le temps maximum de la communication entre les agents et le WIS est de 38 secondes, le temps minimum de 18 secondes et le temps moyen de 26,62 secondes. D'après les résultats obtenus, le temps de communication augmente avec le temps.

d. Test de performance du WPE avec l'Autodock.

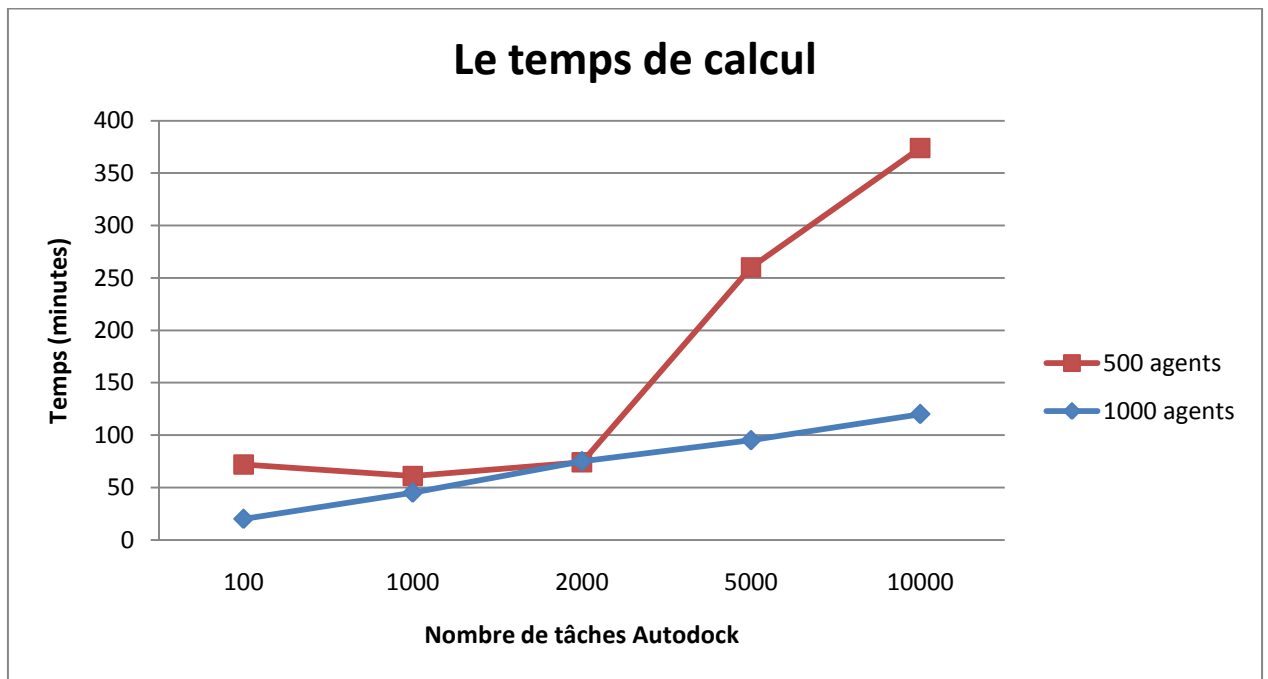


Figure 6 : Le temps de calcul des tâches Autodock sur la grille

Dans le premier test, 1000 agents sont soumis sur la grille par le Job Manager. Ensuite, nous avons créé 100, 1000, 2000, 5000, 10000 tâches Autodock dans le Task Manager. Nous mesurons le temps pour finir toutes les tâches. Le résultat obtenu est représenté dans la figure 5. Normalement, il faut environ 3 minutes pour calculer une tâche Autodock sur un seul ordinateur. D'après les résultats obtenus, le temps de calcul est sensiblement réduit quand on effectue des tâches Autodock sur la grille.

Dans le deuxième test, 500 agents sont soumis sur la grille par le Job Manager. Ensuite, nous avons créé aussi 100, 1000, 2000, 5000, 10000 tâches Autodock dans le Task Manager. D'après les résultats obtenus, si on utilise 500 agents, le temps de calcul est plus long que lorsqu'on en utilise 1000 agents.

De plus, l'exécution du WPE sur la grille n'est pas stable. Quelque fois, le système compte des problèmes, par exemple, on ne peut pas soumettre des agents sur la grille ou des agents qui ont soumis sur la grille n'exécutent pas des tâches dans le Task Manager ... Dans ces cas, nous devons resoumettre des agents sur la grille.

Perspectives :

Avec le résultat obtenu, on peut conclure que dans notre projet nous pouvons utiliser le WPE pour effectuer l'étape de « docking ». Cependant, il y a maintenant encore des problèmes dans l'exécution du WPE. Les travaux effectués doivent permettre de trouver des solutions pour augmenter la performance du WPE.

Aujourd'hui, le WPE est lancé sur l'infrastructure de la grille EGI. Nous essayons de transférer le WPE sur l'infrastructure du « cloud computing » de StratusLab [3].

Référence

- [1] D. S. Goodsell, G. M. Morris, A. J. Olson. *Automated docking of flexible ligands: applications of AutoDock*. Journal of Molecular Recognition, Volume 9 Issue 196, pages 1–5, January/February 1996
- [2] V. Kasam, J. Salzemann, M. Botha et al., *WISDOM-II: screening against multiple targets implicated in malaria using computational grid infrastructures*, *Malaria Journal*, vol. 8, no. 1, article 88, 2009
- [3] E. Huedo, R. Moreno-Vozmediano, R. S. Montero and I. M. Llorente. *Architectures for Enhancing Grid Infrastructures with Cloud Computing*. Computer Communications and Networks, pages 55-69, 2011.
- [4] P. Andretto et al., *The gLite workload management system*, Journal of Physics: Conference Series, Volume 119, part 6, 2008
- [5] P. Andretto et al. *Practical approaches to grid workload and resource management in the EGEE project*. Proceedings of CHEP'04, Interlaken, Switzerland, 27 Sept.–1 Oct. 2004.
- [6] N. Santos and B. Koblitz. *Metadata services on the grid*. In Proceedings of Advanced Computing and Analysis Techniques (ACAT'05), May 2005.