

# Mise en œuvre et évaluation d'un système de buffer à la demande

Encadrement :

Daniel HAGIMONT - Professeur Toulouse INP

## Description du contexte du projet

L'optimisation de la gestion des ressources dans les datacenters a été très étudiée pour les ressources CPU et mémoire, mais relativement peu pour la ressource réseau.

Les applications déployées dans les datacenters sont très souvent distribuées et composées de plusieurs tiers communicant avec des protocoles de communication standards comme TCP. On observe qu'une proportion significative de ces communications véhiculent des données qui ne sont pas ni lues ni modifiées par des serveurs intermédiaires. Elles sont simplement retransmises. De façon simplifiée, un serveur S1 envoie des données à un serveur S3 en passant par un serveur intermédiaire S2 qui ne fait que relayer les données. Une optimisation consiste alors à transférer les données directement de S1 vers S3, dans le but de réduire la charge de traitement sur S2 et la charge sur l'infrastructure réseau. Le défi est de permettre une telle optimisation sans modification des applications exécutées sur les serveurs.

Dans le but d'implanter un telle optimisation, nous proposons d'implanter un mécanisme de tampon à la demande (on demand buffers). Les données envoyées sont envoyées virtuellement (pas réellement) et sont conservées sur le serveur émetteur S1. Le serveur intermédiaire S2 reçoit les données virtuellement dans un tampon protégé en lecture écriture. Si le serveur S2 accède aux données, il y a défaut de protection et les données sont récupérées depuis S1. Si le serveur renvoie les données vers S3 sans y accéder, les données sont envoyées virtuellement et n'auront pas réellement transité par S2.

## Objectifs :

Ce travail a fait l'objet de précédentes investigations, au cours desquelles le fonctionnement général a été prouvé. Cependant, les solutions proposées comportaient des modifications dans le code des applications. L'objectif de ce projet est d'évaluer une solution indépendante des applications s'exécutant sur les serveurs. Pour cela, une architecture multi-tiers, constituée de plusieurs niveaux de serveurs standards tels que Nginx, sera mise en place.

## Étapes du projet :

Les étapes à suivre pour la réalisation de ce travail sont les suivantes :

1. Prise en main et finalisation de la solution existante ;
2. Conception d'un protocole expérimental :
  - a. Définir les architectures sur lesquelles effectuer les tests ;
  - b. Déterminer les métriques à mesurer ;
  - c. Définir des scénarios d'exécution ;

- d. Exécuter le protocole expérimental ;
- e. Analyser les résultats.

## **Ressources :**

- Chapitre 4 de la thèse de Brice EKANE APAH, *Tampon mémoire à la demande*.
- Codes fournis.

## **Intérêts pour les étudiants :**

Outre le développement de compétences en programmation système, ce projet permettra aux étudiants de découvrir la démarche de travail orientée recherche : ils participeront à la conception d'une nouvelle solution et conduiront l'exécution d'un protocole expérimental.