Implémentation d'une couche transport fidèle pour l'émulateur QEMU BXIv3

Niveau: dernière année de cycle Ingénieur ou Master 2 (Bac+5)
Durée: 5 à 6 mois
Lieu: division R&D d'Eviden à Échirolles (Grenoble) - Isère
Équipe: BXI Low Level
Contact: dl-bxi-sw-ll@eviden.com

Contexte

Le calcul haute performance

Eviden, à travers sa filiale Bull, est l'un des acteurs majeurs dans la course vers le calcul haute performance Exascale. Le supercalculateur Leonardo fabriqué par Bull se hisse à la quatrième place du Top500 dans son classement de septembre 2023. Certains de ces supercalculateurs ont la chance de pouvoir embarquer le réseau d'interconnexion haute performance BXI, également conçu et fabriquée par Eviden/Bull. Ces réseaux d'interconnexions sont composés de plusieurs centaines de cartes réseau appelées Network Interface Controller (NIC) ainsi que de switchs à plusieurs niveaux qui forment ensemble la topologie réseau. Les objectifs de ces réseaux d'interconnexion haute performance sont double :

- Pouvoir traiter les tâches de communication réseau rapidement et en parallèle des phases de calcul ;
- La mise à l'échelle de la performance sur des milliers de nœuds communicants entre eux.

La dernière génération du réseau BXI permet d'obtenir un débit utile de 100Gb/s avec une latence pouvant descendre sous la micro seconde. Avec les nouvelles générations de processeur et l'arrivée de la cinquième génération du bus PCIe, ce débit ne suffit plus pour rivaliser avec la vitesse des processeurs. Dans l'optique d'offrir une solution à la hauteur des derniers calculateurs, la troisième génération de la technologie BXI est en cours de conception.

Émulateur BXI sous Qemu

Le logiciel libre Qemu permet d'exécuter un ou plusieurs systèmes d'exploitation (et leurs applications) isolés dans des machines virtuelles sur une même machine physique. Il embarque des versions émulées de la plupart des périphériques PCI courants : son, USB et réseau. Les systèmes d'exploitation invités partagent ainsi les ressources de la machine physique de façon relativement invisible. Qemu peut également être utilisé pour des besoins de recherche et développement sur des composants matériel. L'équipe BXI Low Level, qui s'occupe du pilote Linux pour le projet BXI, a utilisé cette technologie pour développer un émulateur de la carte réseau. Cet émulateur permet de travailler sur les couches logicielles (driver, bibliothèques exposées aux utilisateurs) sans attendre la disponibilité du matériel et donc de prototyper rapidement de nouvelles idées.

Cet émulateur se base actuellement sur une couche réseau simple, privilégiant la facilité de développement, de débogage et de maintenance.

Architecture de la carte BXIv3

La carte réseau BXIv3 implémente la spécification Portals et est architecturée en trois grandes couches : la gestion des commandes et ressources Portals, la gestion des opérations Portals et la couche transport. L'émulateur actuel se divise similairement selon ces couches.

La couche transport est chargée de transmettre des messages vers des nœuds distants. Afin de transmettre des messages cette couche doit implémenter un certain nombre de fonctionnalités, telles que le découpage de message en paquets, le multiplexage, la fiabilité ou le contrôle de congestion.

Objectif du stage

Le futur stagiaire se verra proposer la ré-implémentation de la couche transport de l'émulateur BXIv3 pour être fidèle à l'architecture de la carte réelle. La méthode de travail sera basée sur des cycles itératifs constitués d'étapes de design, de développement et de validation. Une fois les développements suffisamment avancés, il sera possible de faire communiquer l'émulateur et un NIC BXIv3. Pour arriver à l'objectif proposé le stagiaire devra étudier le document d'architecture de BXIv3.

Le fruit du travail du stagiaire pourra alors être utilisé pour permettre de tester des cas d'erreur du NIC, ou de prototyper de nouvelles fonctionnalités.

Profil recherché

Le profil idéal doit avoir de bonnes connaissances sur les points suivants :

- Bases de réseau (Ethernet)
- Langage C
- Gestion de version de code (Git)
- Debug de protocole réseau (gdb, wireshark, ...)

Des connaissances sur les points suivants sont appréciées :

- Virtualisation (Qemu)

Le stagiaire devra faire preuve de persévérance dans la compréhension du document d'architecture.

Mots clés

QEMU | PCIe | Ethernet | Réseau | Portals

Bibliographie

Eviden / High-Performance Computing Solutions
https://eviden.com/solutions/advanced-computing/high-performance-computing/

Spécification Portals
https://www.sandia.gov/app/uploads/sites/144/2023/03/portals43.pdf