

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Καλυβιαννάκης Γεώργιος

Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτυχιακής Εργασίας: Αναπλ. Καθηγητής, Π. Πρατικάκης

Δρ. Ν. Καλλιμάνης(επιβλέπων)

Παρασκευή, 1 Απριλίου 2022, ώρα 10:00 π.μ.

Join Zoom Meeting

<https://zoom.us/j/96751731778>

“Μία Αποδοτική και Ελαφριά OpenSHMEM Υλοποίηση”

ΠΕΡΙΛΗΨΗ

Οι Υπολογισμοί Υψηλής Απόδοσης (HPC) αποκτούν ολοένα και περισσότερη δυναμική, αξιοποιώντας τα προνόμια που προσφέρει ο Τμηματοποιημένος Χώρος Καθολικών Διευθύνσεων (PGAS) για βέλτιστα αποτελέσματα. Μεγάλη πληθώρα προγραμματιστικών γλωσσών και βιβλιοθηκών έχουν δημιουργηθεί που επωφελούνται από τον PGAS, με το πιο ευρέως γνωστό να είναι το OpenSHMEM. Το OpenSHMEM ορίζει μια διεπαφή προγραμματιστικών εφαρμογών (API) η οποία αξιοποιεί μονόπλευρες απομακρυσμένες άμεσες προσπελάσεις μνήμης (RDMA) για εκτεταμένη χρήση σε HPC.

Σε αυτήν την μεταπτυχιακή εργασία παρουσιάζουμε το Gmem, μία υλοποίηση του OpenSHMEM πρότυπου, η οποία υποστηρίζει τις αρχιτεκτονικές δικτύου TCP/IP, RoCE και Infiniband, βασιζόμενη στο GSAS. Το GSAS είναι ένα ελαφρύ και απλό PGAS προγραμματιστικό API, το οποίο δίνει την δυνατότητα σε διεργασίες που τρέχουν σε έναν αριθμό από κόμβους, να επικοινωνούν μεταξύ τους με τρόπο παρόμοιο στο μοντέλο κοινόχρηστη μνήμης. Το Gmem αξιοποιεί την κοινόχρηστη μνήμη στις επικοινωνίες που λαμβάνουν χώρα εντός ενός κόμβου, δίνοντας έτσι την ικανότητα στους χρήστες να εκμεταλλευτούν πλήρως την τοπικότητα χώρου (Spatial Locality), χωρίς την παρέμβαση του λειτουργικού συστήματος ή της κάρτας δικτύου. Επιπλέον, χάρη στο RDMA καταφέρνουμε να διεξάγουμε λειτουργίες σε απομακρυσμένους

κόμβους, με εξαιρετικά χαμηλό χρόνο καθυστέρησης μεταφοράς δεδομένων (Latency) και υψηλή ροή μεταφορών (Throughput).

Το Gmem αξιολογήθηκε σε σύγκριση με τις OpenSHMEM υλοποιήσεις του OpenMPI και του MPICH, οι οποίες βασίζονται στο Unified Communication X (UCX) πρωτόκολλο, για τις αρχιτεκτονικές δικτύου TCP/IP και Infiniband. Στα πειράματά μας μετρήθηκε η απόδοση των λειτουργιών Αποστολής/Λήψης (PUT/GET) όταν πραγματοποιούνται σε απομακρυσμένους κόμβους, καθώς και των ατομικών λειτουργιών μνήμης και συλλογικών λειτουργιών. Παρατηρούμε ότι το Gmem δεν είναι μόνο ισάξιο με των ανταγωνιστικών υλοποιήσεων, αλλά σε πολλές περιπτώσεις επιτυγχάνονται καλύτερα αποτελέσματα. Συγκεκριμένα, στις λειτουργίες λήψης (GET) μεγάλου όγκου δεδομένων, επιτυγχάνουμε χρόνους καθυστέρησης μεταφοράς δεδομένων (Latency) 6x χαμηλότερους από το OpenMPI, και στις ατομικές λειτουργίες 1.25x καλύτερο χρόνο καθυστέρησης μεταφοράς (Latency).

University of Crete

Computer Science Department

M.Sc. Thesis

Kalyvianakis Georgios

**Master's Thesis Supervisor: Associate Professor P. Pratikakis
Dr. N.Kallimanis(Thesis Co-Advisor)**

Friday, 1 April 2022, 10:00 a.m.

Join Zoom Meeting

<https://zoom.us/j/96751731778>

“An Efficient and Lightweight OpenSHMEM Implementation”

ABSTRACT

The High Performance Computing (HPC) is rapidly gaining momentum, relying on the benefits of the Partitioned Global Address Space (PGAS) model for optimal results. Numerous languages and libraries have been introduced that leverage the PGAS model, with the most widely known being OpenSHMEM. OpenSHMEM is a standard specification that defines a one-sided, RDMA capable API for extensive use in HPC.

In this thesis we design and implement Gmem, an OpenSHMEM implementation supporting TCP/IP, RoCE and Infiniband networking protocols, backed by GSAS, a very

lightweight PGAS API, allowing processes spawning on a number of nodes to communicate in very similar way to shared memory schemantics. Gmem leverages shared memory for intra-node communications, enabling users to fully utilize spacial locality without involvement of the OS or the network adapter. With RDMA we are also able to perform operations on remote nodes with extremely low latency and high throughput.

We evaluate Gmem with the OpenSHMEM implementation of OpenMPI and MPICH that rely on the Unified Communication X (UCX) framework, for TCP/IP and Infiniband. In our tests we assess the performance of PUT/GET remote memory operations, several atomic memory operations and collectives operations. During our experiments, we observe that our implementation is not only on par with our competitors, but in some cases we even achieve greater results. In GET operations, for large size transfers we achieve 6x lower latency than OpenMPI, and in Atomic operations 1.25x better latency than OpenMPI.