ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ξηρουχάκης Παντελής
Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Μ. Κατεβαίνης

Τετάρτη, 27/03/2019, 18:00
Αίθουσα Α121, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

" Σχεδίαση και κατασκευή του κομματιού αποστολής, μιας προηγμένης μηχανής Άμεσης Προσπέλασης Μνήμης (RDMA) "

ΠΕΡΙΛΗΨΗ

Στις εφαρμογές που απαιτούνε υπολογιστές υψηλών επιδόσεων (HPC) η χαμηλή καθυστέρηση επικοινωνίας ανάμεσα σε απομακρυσμένους κόμβους είναι καίριας σημασίας για την απόδοση της εφαρμογής. Το InfiniBand και άλλες έτοιμες επιλογές δικτύων μπορούν να μειώσουν αυτή τη καθυστέρηση αλλά απαιτούνε εξειδικευμένες και ακριβές κάρτες δικτύου που στη γενική περίπτωση είναι απομακρυσμένες από τον επεξεργαστή. Σε αυτή την εργασία θα περιγράψουμε της σχεδίαση και κατασκευή μιας προηγμένης Μηχανής Άμεσης Προσπέλασης Μνήμης(RDMA) που κατασκευάσαμε στο ΙΤΕ μέσα στα πλαίσια του ευρωπαϊκού έργου ExaNeSt η οποία υπερτερεί σε πολλά σημεία ως προς το InfiniBand. i) Κόβουμε τις μεγάλες μεταφορές σε πολλές μικρές υπό-μεταφορές, επιτρέποντας μας να χρησιμοποιούμε ταυτόχρονα πολλές διαδρομές μέσα στο δίκτυο (multi-pathing) σε επίπεδο υπό-μεταφοράς. ii) Υποστηρίζουμε επαναλήψεις μεταφοράς, σε επίπεδο υπό-μεταφοράς, εάν οι συνθήκες το απαιτούν. iii) Χρησιμοποιώντας την

μονάδα εικονικής μετάφρασης περιφερικών (smmu) της ARM, δεν χρειάζεται να καρφιτσώνουμε περιοχές μνήμης ενώ παράλληλα έχουμε πρόσβαση σε όλο την εικονική μνήμη του συστήματος. Επιπρόσθετα παρέχουμε έναν αριθμό από εικονικά κανάλια τα οποία είναι σε θέση να δουλευθούν ταυτόχρονα έχοντας χιλιάδες εκκρεμείς μεταφορές. Η προηγμένη Μηχανή Άμεσης Προσπέλασης Μνήμης μας έχει σχεδιαστεί ώστε να μπορεί να υποστηρίξει μεταφορές από πολλαπλά μονοπάτια έτσι ώστε να είναι σε θέση να εκμεταλλευτεί τα πλούσια σε παράλληλα μονοπάτια δίκτυα, από τα οποία αποτελούνται οι μοντέρνοι Υπολογιστές Υψηλών επιδόσεων. Σε αυτή την εργασία παρουσιάζουμε την υλική κατασκευή αυτής της RDMA στο Zynq Ultrascale+ MPSoC. Το υλικό έχει σχεδιαστεί έτσι ώστε να μπορεί να δουλέψει σε ταχύτητες τόσο υψηλές όσο 200MHz έχοντας καθυστερήσεις τόσο μικρές όσο ένα Μίκρο-δευτερόλεπτο ενώ παράλληλα καταναλώνει ελάχιστους πόρους, αφήνοντας αρκετό χώρο να χρησιμοποιηθεί από άλλες μορφές επιταχυντών. Επίσης σχεδιάσαμε και ενώσαμε τον μεταγωγέα πακέτων καθώς και την διεπαφή δικτύου που χρειάζεται έτσι ώστε να εκμεταλλευτούμε το μεγάλη εικονική μνήμη που παρέχεται από το πρωτότυπο μας. Εφαρμόσαμε την RDMA μας σε πολλαπλές συνδεδεμένες μεταξύ τους FPGAs και τρέξαμε διάφορα προγράμματα αναφοράς, έτσι ώστε να μπορέσουμε να αξιολογήσουμε τις επιδόσεις μας. Τα αποτελέσματα δείχνουνε τεράστια βελτίωση έναντι στο κλασικό 10G ethernet καθώς και προηγούμενες RDMA μηχανές μας.

Pantelis Xirouchakis
M.Sc. Thesis

Computer Science Department
University of Crete
Master's Thesis Supervisor: Professor, M. Katevenis

Wednesday 27/03/2019, 18:00
Room A121, Computer Science Dept., University of Crete

“Design and Implementation of the Send Part of an Advanced RDMA Engine”

ABSTRACT

In High Performance Computing (HPC), low latency communication between remote processes is crucial to application performance. InfiniBand and other off-the-shelf networks can reduce the latency but require special and costly network interface cards, which are loosely coupled with CPU. In this work, we describe the design and implementation of an advanced RDMA engine

developed within the ExaNeSt EU project, which has a number of advantages over Infinibad: i) We segment RDMA transfers in blocks, and support block-level multipathing of RDMA transfers on a per-block basis. ii) We perform selective end-to-end retransmissions. iii) We do not need to pin the regions of RDMA transfers in memory, while at the same time we support accessing the full virtual address space of processes, using ARM SMMU. Additionally, we provide a number of virtual channels able to work simultaneously with many outstanding transfers. Our advanced RMDA engine is designed to support multi-pathing in order to be able to utilize the rich parallel links found in HPC networks. In this work, we describe the hardware implementation of the RDMA engine on the Zynq Ultrascale+. The hardware design has been optimized to meet timing requirements of up to 200Mhz while consuming little resources, leaving plenty of space to be used by i.e accelerators. We have also designed and integrated the interconnect required, as well as the Network Interface (NI) in order to utilize the large Global Virtual Address Space (GVAS) provided by our hardware prototype. We have implemented our advanced RDMA on multiple interconnected FPGAs and have run HPC benchmarks and applications in order to verify and evaluate our design. The results show great improvement over 10G ethernet, as well as our previous RDMA implementations. Finally, our RDMA has been designed to easily accommodate many more features with little to no change, such as congestion management.