

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Σκορδαλάκης Εμμανουήλ
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Επικ. Καθηγητής, Π. Πρατικάκης**

Δευτέρα, 30/09/2019, 17:00

Αίθουσα Β106, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**“ Υποστήριξη διαφορετικών επιπέδων υπηρεσιών μέσω διάφανης μετακίνησης
σελίδων σε κατανεμημένα συστήματα μνήμης”**

ΠΕΡΙΛΗΨΗ

Με την συνεχή εξέλιξη εφαρμογών υψηλής επίδοσης, αυξάνεται ραγδαία και η ανάγκη τους για περισσότερη μνήμη. Ως αποτέλεσμα, γίνεται όλο πιο επιτακτική η απαίτηση για περισσότερη μνήμη στους υπολογιστικούς κόμβους ίδιας συστάδας που εκτελούν τις παραπάνω εφαρμογές. Ωστόσο, ένας μεμονωμένος υπολογιστικός κόμβος έχει όρια ως προς τη χωρητικότητα μνήμης που μπορεί να διαθέτει. Τυπικά, με την εκτέλεση πολλαπλών εφαρμογών οι οποίες διαφέρουν σε υπολογιστικές και χωρικές απαιτήσεις, δημιουργείται μεταβαλλόμενο φορτίο εργασίας μεταξύ των υπολογιστικών κόμβων. Επομένως, κάποιοι από αυτούς τους κόμβους χρησιμοποιούν το μεγαλύτερο ποσοστό μνήμης τους ενώ κάποιοι άλλοι έχουν ανεκμετάλλευτη χωρητικότητα μνήμης η οποία θα μπορούσε δυναμικά να χρησιμοποιηθεί από εκείνους με μεγαλύτερο φορτίο μνήμης.

Συνεπώς η έννοια της διαχείρισης απομακρυσμένης μνήμης, είναι αντικείμενο έρευνας για πολλούς οργανισμούς, οι οποίοι έχουν υλοποιήσει ποικίλες τεχνικές για ανάγνωση

και εγγραφή δεδομένων σε απομακρυσμένη μνήμη. Παρόλο που η χρήση απομακρυσμένης μνήμης πρακτικά αυξάνει την συνολικά διαθέσιμη μνήμη ενός υπολογιστικού κόμβου, η πρόσβαση σε απομακρυσμένα δεδομένα μπορεί να μειώσει δραματικά την απόδοση μιας εφαρμογής, εξαιτίας της μεταφοράς αυτών των δεδομένων μέσω της ενδοσύνδεσης δικτύου της συστάδας. Επιπλέον, οι διεπαφές εφαρμογών λογισμικού που υλοποιούνται για να δώσουν πρόσβαση σε απομακρυσμένη μνήμη, αρχικά μπορεί να είναι περίπλοκες ως προς τη χρήση τους, και έπειτα η ευθύνη για την δέσμευση απομακρυσμένης μνήμης και την προτεραιότητα σε αυτήν μεταξύ των εφαρμογών, ανάγεται στις ίδιες τις εφαρμογές, παρόλο που έχει αρκετή πολυπλοκότητα, ειδικά όταν ένας μεγάλος αριθμός εφαρμογών εκτελείται ταυτόχρονα στον ίδιο υπολογιστικό κόμβο.

Σε αυτή τη μεταπτυχιακή εργασία, παρουσιάζουμε το σύστημα μετανάστευσης σελίδων (PMS), το οποίο επιβλέπει την χρήση κύριας μνήμης των υπολογιστικών κόμβων σε μία συστάδα, και μετακινεί δεδομένα αραιής πρόσβασης μιας διεργασίας, από την μνήμη ενός υπολογιστικού κόμβου με αυξημένο φορτίο μνήμης, σε ελεύθερη μνήμη ενός απομακρυσμένου υπολογιστικού κόμβου της ίδιας συστάδας με χαμηλό φορτίο μνήμης. Τα βασικά χαρακτηριστικά του PMS είναι η διάφανη μεταφορά LRU σελίδων μιας διεργασίας σε απομακρυσμένη μνήμη ενώ παράλληλα χρησιμοποιείται ένας αλγόριθμος δικαιοσύνης για την επιλογή σελίδων μνήμης μεταξύ πολλών διεργασιών που εκτελούνται στον ίδιο υπολογιστικό κόμβο. Επιπλέον, η απομακρυσμένη μνήμη χαρτογραφείται στον τοπικό κόμβο από το λειτουργικό σύστημα και άρα τα απομακρυσμένα δεδομένα μπορούν να βρίσκονται στην κρυφή μνήμη. Για την ακρίβεια, μια ανάγνωση και/ή εγγραφή στην απομακρυσμένη μνήμη συμβαίνει όταν έχουμε αστοχία στην κρυφή μνήμη. Αυτή η δυνατότητα αυξάνει την απόδοση μιας διεργασίας όταν δεν υπάρχουν πολλές αστοχίες, λόγω της μείωσης μεταφοράς δεδομένων στο δίκτυο. Τέλος, Το σύστημα επιτρέπει την επιστροφή σελίδων μνήμης τοπικά όταν ελευθερωθεί μνήμη στον κόμβο, ή αυξηθεί η συχνότητα πρόσβασης σε αυτές τις σελίδες.

Αξιολογούμε το PMS εκτελώντας εφαρμογές οι οποίες επικεντρώνεται στην προσπέλαση μεγάλου όγκου δεδομένων από τη μνήμη. Χρησιμοποιούμε εφαρμογές που κάνουν σειριακή προσπέλαση σε συνεχόμενες θέσεις μνήμης σε πίνακες της τάξης του 1 Gigabyte και έτσι βγάζουμε συχνά δεδομένα από την κρυφή μνήμη με αποτέλεσμα την συνεχή μεταφορά δεδομένων μέσω δικτύου. Με αυτόν τον τρόπο μπορούμε να μετρήσουμε την μείωση απόδοσης της εφαρμογής λόγω της προσπέλασης σε απομακρυσμένη μνήμη στη χειρότερη περίπτωση. Επίσης εκτελούμε εφαρμογές που έχουν αυξημένη χωρική τοπικότητα στην κρυφή μνήμη, και έτσι δείχνουμε ότι το ποσοστό απόδοσης της εφαρμογής αυξάνεται λόγω της μειωμένης πρόσβασης στην απομακρυσμένη μνήμη. Τέλος, παρατηρούμε την συμπεριφορά και απόδοση σε πραγματικές εφαρμογές υψηλής επίδοσης, χρησιμοποιώντας το PMS.

Skordalakis Emmanouil

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Assistant Professor, P. Pratikakis

Monday, 30/09/2019, 17:00

Room B106, Computer Science Dept., University of Crete

“Support for different service levels through transparent migration of pages in distributed memory systems”

ABSTRACT

With the constant evolution of high performance applications, their memory requirement is rapidly increasing. As a result, the demand for more memory on computer nodes of large clusters running those applications, continuously rises. However, an individual computer node has limits in terms of memory capacity. Typically, by running several processes of different computational and memory requirements on a cluster, creates fluctuating workloads among the computer nodes. Hence, several nodes use most of their memory while others are left with unused memory which could potentially be exploited by nodes with a heavy memory workload.

Consequently, the concept of remote memory management has become the subject for research by many organizations, which have implemented varying techniques for reading and writing data on remote memory. Although using remote memory practically increases the total available memory of a computer node, accessing data remotely, can critically minimize performance due to the data travelling through the network interconnection of the cluster. Furthermore, software APIs that are implemented to give processes access to remote memory, primarily can be complex, and secondly the responsibility for remote memory allocation and fair remote memory sharing among processes, is assigned to processes themselves, which can be quite complicated, especially when many processes are running simultaneously on the same computer node.

In this MSc thesis, we present the Page Migration System(PMS), which monitors main memory usage of the computer nodes on a cluster, and moves infrequently accessed data of a process from the memory of a computer node with heavy memory workload, to the unused memory of a remote computer node of the same cluster node with a lighter

memory workload. The key features of the PMS is that it transparently moves LRU pages of processes to remote memory while using a fairness algorithm when choosing memory pages among many processes running on the same computer node. What's more, remote memory is mapped on the local node, allowing the OS to cache remote data. To be precise, a read and/or write on remote memory happens when we get a cache miss. Cacheability offers better performance when there are less misses, by reducing network transfers. Finally the system is able to return memory pages locally if the overall node memory usage drops, or if the access frequency of those memory pages increases.

We evaluate the PMS using several benchmarks that stress the CPU in terms of memory access. We use benchmarks that perform raw serial access on arrays of around a Gigabyte in size and thus cause cache eviction frequently, essentially moving more data through the network. That way we can measure the performance drop of a process due to memory access in the worst case scenario. We also run cache blocking benchmarks that exploit temporal locality, and we show that we get a better performance that way by reducing operations on remote memory. Finally we observe the behaviour and performance on real HPC applications using the PMS.