

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Βελεγράκης Ιωάννης

Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Μ. Κατεβαίνης

Τρίτη, 31 Μαρτίου 2015, 16:30

Αίθουσα K206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**“Μηχανισμοί Λειτουργικού Συστήματος για Χρήση Απομακρυσμένων Πόρων σε ARM
Microservers ”**

ΠΕΡΙΛΗΨΗ

Οι πρόσφατες προσπάθειες για βελτιστοποίηση της απόδοσης αλλά και της κατανάλωσης ισχύος των μεγάλης κλίμακας Data Centers έφερε τους λεγόμενους Microservers στο προσκήνιο, οι οποίοι αποτελούνται από μικρότερης κατανάλωσης Υπολογιστικές Μονάδες (ΥΜ) συγκρινόμενοι με τις παραδοσιακές αρχιτεκτονικές. Η κεντρική ιδέα είναι ότι χρησιμοποιώντας μεγάλο αριθμό από τέτοιες ΥΜ μπορούμε να κατασκευάσουμε μία πολυπύρνη μηχανή που θα έχει υψηλή απόδοση στις πολυνηματικές εφαρμογές και παράλληλα θα έχει χαμηλή ενεργειακή κατανάλωση. Σε τέτοιες μηχανές, οι ακριβοί πόροι αναγκαστικά διαμοιράζονται μεταξύ των ΥΜ, καθώς δεν γίνεται να αποδοθούν πόροι για κάθε ΥΜ ξεχωριστά. Ωστόσο, η διαμοίραση των πόρων αυτών απαιτεί την υλοποίηση κατάλληλων μηχανισμών στο Λειτουργικό Σύστημα (ΛΣ) και γενικότερα στο επίπεδο του λογισμικού.

Σε αυτήν την εργασία, ερευνήσαμε και υλοποιήσαμε μηχανισμούς στο ΛΣ και στο επίπεδο διεργασιών, οι οποίοι είναι απαραίτητοι για την χρησιμοποίηση ΥΜ βασισμένων σε ARM αρχιτεκτονικές σε μεγάλης κλίμακας συστήματα. Υλοποιήσαμε την διαμοίραση πόρων δύο ειδών, χρησιμοποιώντας τους διαθέσιμους μηχανισμούς του hardware, όπως η Remote DMA και

Remote Load/Store. Συγκεκριμένα, υλοποιήσαμε μηχανισμούς για την απομακρυσμένη προσπέλαση μνήμης και την χρησιμοποίηση μιας κοινής και virtualized 10 Gbps διεπαφής δικτύου η οποία μπορεί να χρησιμοποιείται από πολλές ΥΜ ταυτόχρονα.

Η χρησιμοποίηση της απομακρυσμένη μνήμης υλοποιήθηκε με τους εξής ακόλουθους τρόπους: (1) Σαν επέκταση της τοπικής DRAM μίας ΥΜ, (2) Σαν απομακρυσμένη συσκευή για Swarp, και (3) Σαν απομακρυσμένη συσκευή χαρακτήρων που χρησιμοποιείται απευθείας από τις διεργασίες. Δείχνουμε ότι η χρήση της απομακρυσμένης μνήμης σε ένα σύστημα με ΛΣ δεν επιφέρει μείωση της απόδοσης.

Η από κοινού χρήση της διεπαφής δικτύου 10Gbps επιτυγχάνεται με ένα οδηγό στο ΛΣ, τον οποίο υλοποιήσαμε και ο οποίος επιτρέπει στο ΛΣ να βλέπει την διεπαφή αυτή σαν κλασική διεπαφή Ethernet. Αυτό είναι απαραίτητο για να μπορούν να τρέξουν διεργασίες που χρησιμοποιούν τα Berkeley Sockets, χωρίς να χρειάζεται να τροποποιηθούν. Ο οδηγός χρησιμοποιεί scatter-gather DMA για γρήγορη και zero-copy μετάδοση και παραλαβή πακέτων και λειτουργεί σε Full-Duplex χρησιμοποιώντας δύο ξεχωριστά κανάλια της DMA μηχανής. Επιπλέον, υποστηρίζει μηχανισμούς Interrupt Coalescing και μπορεί να διαχειριστεί τα MAC και PHY μέρη του hardware μέσω του πρωτοκόλλου Management Data Input/Output (MDIO).

Η εργασία αυτή δείχνει ότι πράγματι μπορούμε να φτιάξουμε ένα σύστημα βασισμένο σε ΥΜ ARM αρχιτεκτονικής, οι οποίες δεν είναι σχεδιασμένες να λειτουργούν σε ένα τέτοιο σύστημα. Πιστεύουμε ότι αυτή η εργασία θα γίνει ακόμα πιο σημαντική στο μέλλον, καθώς νέες 64bit ARM πλατφόρμες εμφανίζονται, οι οποίες στοχεύουν να χρησιμοποιηθούν σε μεγάλης κλίμακας Data Centers.

Velegrakis Ioannis

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Professor M. Katevenis

Tuesday, 31/03/2015, 16:30

Room K206, Computer Science dept., University of Crete

“Operating System Mechanisms for Remote Resource Utilization in ARM Microservers”

ABSTRACT

Recent efforts towards performance and power optimization in large-scale Data Centers have brought the use of Microservers in the forefront. Compared to traditional architectures, Microservers consist of smaller, less power-hungry Compute Units (CUs) compared to traditional architectures. The key concept is that by integrating such smaller CUs in high numbers, the resulting many-core system can achieve high multi-threaded performance, while maintaining a

low power profile. In such an environment, expensive resources must be shared among CUs, since it is costly and impractical to dedicate one to each CU. However, the management of these resources requires to implement sharing mechanisms in the Operating System (OS) and the software stack.

In this work, we investigated and implemented OS and user space software mechanisms that are necessary for the deployment of ARM-based CUs in such large-scale systems. We address the sharing of two kinds of remote resources by fully exploiting the underlying hardware features, such as Remote Direct Memory Access (DMA) and Remote Load/Store. In particular, we have implemented software mechanisms enabling access to remote memory and usage of a shared virtualized 10 Gbps Network Interface Card (NIC) by several CUs simultaneously.

Remote memory access is implemented in three different ways: (1) As an extension of the local DRAM of the CUs, (2) As a remote Swap Device, and (3) as an I/O character device accessed directly from user space. We demonstrate that remote memory can be used effectively without performance penalty in a system running a full OS.

The sharing of a virtualized 10Gbps NIC is achieved by a kernel network driver, that we have implemented, which enables utilization of the customized hardware as a standard Ethernet Device. This allows legacy applications that use Berkeley Sockets to run unmodified. The network driver makes use of scatter-gather DMA for fast zero-copy packet transmission and reception and operates in Full-Duplex mode by using two independent DMA channels. Additionally, it supports Interrupt Coalescing and management of the MAC and PHY hardware blocks using the Management Data Input/Output (MDIO) interface.

In conclusion, this work shows that we can indeed utilize a system built upon ARM-based CUs that were not originally designed to operate in such an environment, by the sharing of remote and shared resources by the Linux OS and its user space environment. We expect this work to become even more relevant with upcoming 64-bit ARM-based platforms, targeting large-scale servers for Data Centers.