Analysis of the Resource Affinity in NUMA Architecture for High Performance Network

Zhiyuan Bo[1, a]

[1] School of Electronic Information and Electronic Engineering, Shanghai Jiao Tong University, China

[a]bozhiyuan@sjtu.edu.cn

Keywords: NUMA, Virtualization, High Performance, Affinity.

Abstract. Multicore system is processors and the high speed network adapter is very common in cloud computing. So it becomes more and more important to make full use of recourses such as CPU, memory and I/O devices. However, traditional system architecture just consider about CPU and memory, cannot meet the need of high performance network. In this paper, we measure and analyze the performance of different mapping strategies for the NUMA architecture in virtual environment.

Introduction

Today, more and more applications are running on the cloud. With the development of cloud computing, how to make full use of the multi-core resources and provide efficient services in cloud environment, has become an important research area. One of the most important part is how to transfer data between high performance servers and cloud applications. Especially in the high performance network, the traditional system architecture can not completely meet the needs of high performance network, which will cause a great decline in performance.

The traditional multi-core system shares the same physical memory, but with the improvement of the computer processing performance and the increase of the number of cores, it brings a lot of memory access delay and bandwidth competition[1]. In order to enhance the utilization of CPU, Researchers has come up with an Independent Memory Controller (IMC) for each CPU to reduce the remote memory access. For each CPU has an independent memory, this architecture is named Non-Uniform Memory Access (NUMA) architecture.

The present study is to enhance the high performance network architecture of NUMA resources based on the affinity of memory and CPU [2], but failed to present a perfect solution for the demand of performance network cloud environment. The existing NUMA affinity model is a two-dimensional model, mainly focus on memory and CPU under the framework of NUMA performance optimization, which ignore the huge impact on the performance of high speed I/O in high performance network.

This paper presents the performance of different mapping situations in NUMA Architecture for high performance and analyzes the importance of I/O devices in the virtual environment. then we give the performance evaluation to the different strategies, and propose the future works.

Related Work

NUMA system has become more and more widely used, many manufacturers such as Intel, Dell have successfully launched NUMA based framework on their servers. As the development of high performance and the improvement of the high speed I/O device, traditional virtualization solutions cannot match the needs.

Xen is a virtual machine monitor developed by University of Cambridge. It intends to run as many as 100 full operating systems on a single computer. Xen supports a form of virtualization known as para virtualization, in which guests run a modified operating system [3]. This makes Xen can achieve high performance virtualization without special hardware support. XEN give a solution of IONUMA by binding the Virtual Machine (VM) on some NUMA nodes which have I/O devices, in order to reduce remote access to I/O devices. Openstack [4] is studying a kind of NUMA scheduling method based on I/O device, which provides high performance and low delay solution, named Network

Function Virtualization (NFV). VMware has also carried out the research and experiment of related I/O scheduling system in the virtual environment [5].

Proposed Approach

We present the architecture of non-uniform I/O access in virtual environment and measure the performance decline by different mapping strategies based on the KVM virtual machine monitor. And analyze the difference of performance.

Virtual Environment

In the cloud environment, it is very important to study the asymmetric I/O access, the existing virtualization technology such as Xen, KVM and other relatively perfect and mature solution have already give good support on traditional NUMA architecture, together with the hardware-assisted virtualization technology (such as IOMMU), make it possible for the research of NUMA architecture in virtual environment [6].

The virtual machine monitor through the method of binding memory and vCPU as much as possible to the same NUMA node to improve the affinity, but the host is insensible of the applications running in the VMs.

SR-IOV technology

SR-IOV technology is a kind of hardware solution based on virtualization, which can improve the performance and scalability. The SR-IOV standard allows efficient sharing of PCIe(Peripheral Component Interconnect Express) between virtual machines. It is implemented in hardware, so it allows comparable I/O performance to the host. SR-IOV technology also defines a new standard. In this way, it can allow the virtual machine directly be connected to the I/O device.

Through SR-IOV technology, I/O performance in the virtual environment has been greatly improved. At the same time, single I/O resource can be shared by many virtual machines. Shared devices will provide dedicated resources, and can also use shared generic resources. In this way, each virtual machine can access the unique resources.

Non-uniform Access

NUMA architecture not only brings the memory access to the asymmetric, but also brings the non consistency of other peripheral devices such as I/O devices. I/O devices are usually connected to one or more NUMA nodes under this architecture, so it also brings remote access to other nodes. Especially the high performance network, the data transmission rate is higher and higher, the bottleneck by the memory access asymmetry gradually shifted to the impact of I/O access.
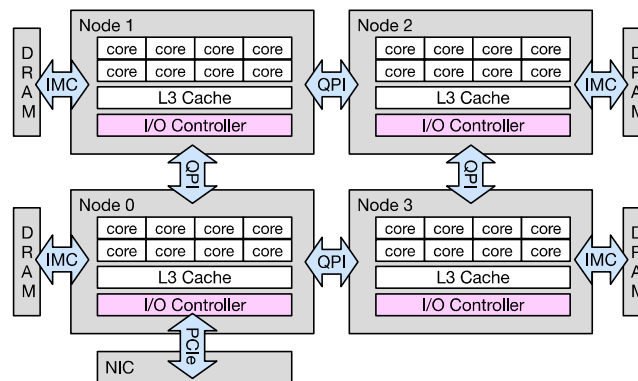

Figure 1. NUMA architecture

We can see the architecture clearly in Figure 1. The IMC is connected to the memory, which is connected with the network card through the I/O controller. One core is connected to another by the Intel Quick Interconnect Path Interconnect (QPI) technology [7]. With the application of high speed network cards, the bandwidth is already more than the fast channel interconnection provided by QPI technology, so the delay and remote access of access to the i/o device can not be ignored.

Experiment and Performance Evaluation

Experimental Environment

We use 4-socket server, and each socket contains a Xeon E5 Intel 4610 V2 processor. Both servers have a ConnectX-3 Mellanox 40 GigE network card. We use the SR-IOV network card virtualization technology to provide dedicated resource. We use KVM virtualization technology to set a VM and calculate the throughput in the process of data transmission, by running the Netperf [8] benchmark. The following Table 1 shows the experiment configuration details.

Table 1. Experiment configuration details

Item	Configuration
CPU	Intel Xeon 4-sockets processor each with 8 cores (2.3GHz)
Memory	4 sockets, each with 32GB
Network Adapter	Mellanox ConnectX-3 40 Gigabit Ethernet adapter
Hypervisor	KVM 2.0.0

Experimental Content

We mainly divided into four different situations as Figure 2 shows to study the effect of different placement issues on performance, to find the performance difference of memory, CPU, I/O devices, in order to analyze the network environment to pave the way for resource affinity.
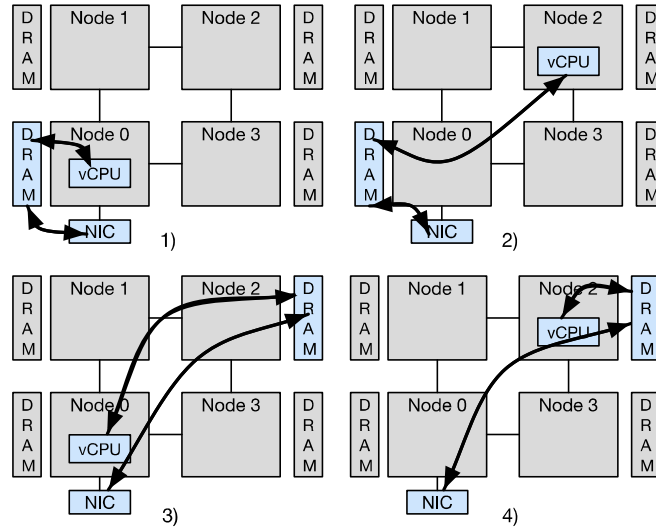

Figure 2. Four Different Situations

First, we run a VM in Node 0. We bind the virtual machine memory and the vCPU of the VM on node0 connected to the network card. Then we transfer the vCPU to The Node 2，which is the remotest node to the network card. In the third case, we change the vCPU back to Node 0 but bind the memory to the remotest Node. In the last situation, we bind the memory and vCPU to Node 2. We mainly analyze the different performance for the four strategies to measure the influence for NUMA Architecture in virtual environment

Performance Evaluation

We use the TCP_STREAM mode of Netperf for the four situations by different message size. We can see from the Table 2 and get the following three conclusions. 1)As the message size improving, the throughput also improves, but when the size become large enough, the throughput improvement less because of the bandwidth. 2)The throughput is larger when the memory is local with vCPU. 3) While the VM(vCPU) together with its memory is local to the network card, throughput is the highest among all the situations.

Table 2. Throughput in different situation

Throughput in different situation(10^6 bits/sec)	VM(vCPU) is local		VM(vCPU) is remote	
	Memory is local for the VM	Memory is remote for the VM	Memory is local for the VM	Memory is remote for the VM
1024 bytes	7643.12	7309.81	7236.61	7063.15
5120 bytes	15287.09	13934.70	13471.57	11659.12
16384 bytes	19127.12	16820.03	15936.44	13715.67
51200 bytes	22171.63	21392.34	19874.65	19231.74
1024000 bytes	22342.19	22087.35	219439.49	20865.28

Summary

The research aims at the affinity in high performance network, mainly about the NUMA architecture in virtual environment. Through our experiment, we quantify the change because of the remote access to I/O devices. For future work, we want to optimize the Non-Uniform I/O Access and propose a scheduler for virtual environment in NUMA architecture.

References

[1] Majo Z, Gross T R. Memory system performance in a NUMA multicore multiprocessor[C]//Proceedings of the 4th Annual International Conference on Systems and Storage. ACM, 2011: 12.

[2] Schroeder, Bianca, and Garth Gibson. "A large-scale study of failures in high-performance computing systems." IEEE Transactions on Dependable and Secure Computing 7.4 (2010): 337-350.

[3] Information on https://en.wikipedia.org/wiki/Xen

[4] Information on http://docs.openstack.org

[5] Banerjee A, Mehta R, Shen Z. NUMA Aware I/O in Virtualized Systems[C]//2015 IEEE 23rd Annual Symposium on High-Performance Interconnects. IEEE, 2015: 10-17.

[6] Kim C, Park K H. Credit-Based Runtime Placement of Virtual Machines on a Single NUMA System for QoS of Data Access Performance[J]. IEEE Transactions on Computers, 2015, 64(6): 1633-1646.

[7] Zhao R. Intel enterprise server processor packaging challenge and future trend[C]//Electronic Packaging Technology (ICEPT), 2014 15th International Conference on. IEEE, 2014: 1516-1521.

[8] Information on http://www.netperf.org/netperf/