Multi-node Power/Performance Modeling for HPC System

Sangwoo Han, Tae Yang Jeong, Eui-Young Chung
Dept. of Electrical and Electronic Engineering
Yonsei University
Seoul, Korea
Swhan0330@yonsei.ac.kr, drthvbfg@yonsei.ac.kr, eychung@yonsei.ac.kr

Abstract

High Performance Computing (HPC) System refers to using a cluster of hundreds or more processing nodes for application which requires a lot of computation. The need for HPC systems has increased in recent years as more applications require a lot of computations such as Deep learning and AI. Various tools and techniques are being developed and researched to efficiently operate HPC systems in line with this need. In this paper, we develop multi-node power/performance modeling. This will help HPC system users to predict power and performance before they configure the system so that they can implement the optimal HPC system for programmers. Power and performance prediction modeling achieves 90% accuracy and takes less than an hour to predict.

Keywords: Multi-node, Power/Performance Simulation model

1. Introduction

HPC systems are widely used in various areas that require a lot of computations, such as molecular mechanics simulation, weather forecasting, and artificial neuron simulation. Not only hardware for HPC, but also the market for software to optimize HPC system is increasing. As the need for HPC systems grows, profiling and optimization tools, and performance modelling to help HPC programmers are consistently studied and developed.

In this paper, we develop multi-node power and performance modeling to support the HPC system. The *prediction modeling achieves 90% accuracy and takes less than an hour to predict.*

2. Background & Motivation

As mentioned earlier, as the need for HPC systems increases, profiling and optimizing tools [1][2] and performance simulation modelling [3] are consistently being studied and developed to support HPC programmers.

However, most performance modeling is too slow to predict performance for high-volume applications which used in HPC system. Also, studies for power prediction models which consider HPC multi-node systems, are insufficient. HPC systems are difficult to configure diverse systems in terms of price and performance realistically. The performance of the application varies depending on the configuration of the system. So it is important to configure a multi-node system optimized for the application. By using the multi-node power and performance modeling developed in this paper, it will make it easier and simpler to configure multi-node system.

Message Passing Interface (MPI) is library which is standard for the exchange of information for parallel processing on multiple nodes. When an application containing MPI is run, the number of process you set up is created and runs in parallel. Each process executes an application, exchanging data and information with different processes. The advantage of this method is that processors can send and receive data between multiple hardware-separated nodes. Most of the multi-node system uses MPIs to run applications.

3. Implement

Figure 1 illustrates the overall methodology of our multi-node power and performance modeling.

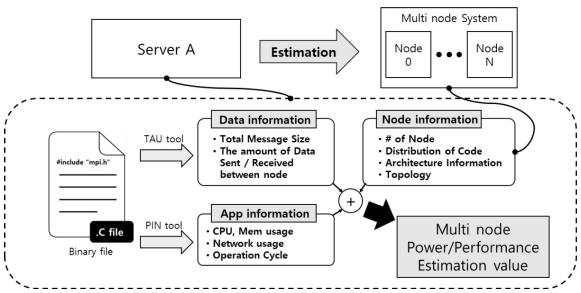

Figure 1. Power/performance estimation flow

Table 1: Recommended margins

Application	Mem access instruction / Total instruction	Data Movement Between Node	DRAM Usage	Sever execution time	Multi-node execution time
CoMD	0.41	1,800,770 K	249.31	40.934	43.072
AMG	0.43	419,435 K	3487	39.203	33.858
MiniFE	0.36	427,547 K	1428	25.864	21.022
SW4lite	0.56	190,109 K	777	20.945	21.518

Power and performance prediction of multiple nodes proceeds on server A. First, we analyze the application characteristics by using TAU [2] and PIN [4] tools for the binary file containing the MPI code, and to predict power and performance, we calculate with information about the multi-node system (node 0 to node n) users want to predict.

TAU is a profiling and optimization tool developed by Oregon University to help programmers. TAU is the tool which can analyze how applications behave on multiple nodes through compiler-level analysis of source code. Using this tool, we analyze the amount of data movement between multi nodes. PIN developed by intel is a framework which can create dynamic program analysis tools. Using this framework, we can make tool which can extract trace or analyze various data dynamically during program execution. With PIN, we develop a tool that can analyze the amount of memory access, the number of instructions during execution, and the percentage of memory access instructions relative to all instructions.

The following is a methodology of power and performance modeling based on analysis data. Table 1 shows the results of the analysis using the TAU tool and the PIN tool when the application is executed with two nodes and a single node. For example, As shown in Table 1, when the amount of data movement between the nodes (3rd column) increases, the performance decreases when running on multi-nodes. Also, the performance of a memory bounded application (SW4lite) which has more instructions to access memory among the whole instructions (the value of 2nd column in SW4lite is more than 0.5) is determined by the performance of memory rather than other characteristics. In addition to these analyses, various application characteristics such as Memory Usage and CPU Usage were analyzed to predict performance.

Finally, we analyze and parameterize the configuration environment of the multi-node system which user wants to predict and the configuration environment of the server which we have. By applying this parameter, we can predict performance more accurately. For power modeling: we predict the power of the multi-node system based on the power values measured by the server, parameter, value of predicted performance.

4. Evaluation

To verify our power and performance modeling, we constructed the experiment environment as shown in figure 2

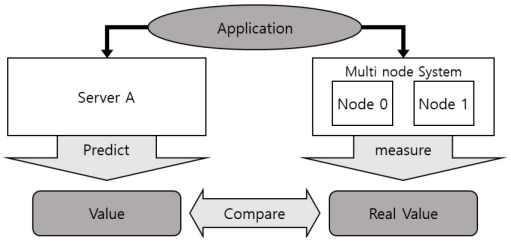

Figure 2 Experiment configuration

We use our power and performance simulation model to predict power and performance about multi-node system which are combined with two computers, in server A. And we compare predict value and real value that measured in Multi node system. The hardware configurations of the server and the multi-node system are shown in the table 2 below.

Table 2: Recommended margins

	Specification
Server A	
- CPU	Intel Xeon E3-1231@ 3.40GHz
- GPU	GTX 980 Ti
- Memory	16 GB
- Disk	2 TB
Node 0	
- CPU	Intel Core i5-4570 @ 3.20GHz
- GPU	GTX 750 Ti
- Memory	16 GB
- Disk	3 TB
Node 1	
- CPU	Intel Core i5-4690 @ 3.50GHz
- GPU	GTX 750 Ti
- Memory	16 GB
- Disk	3 TB

Figure 3 below shows a comparison of the predicted and the actual values. The accuracy of the test results was achieved within 90 per cent.

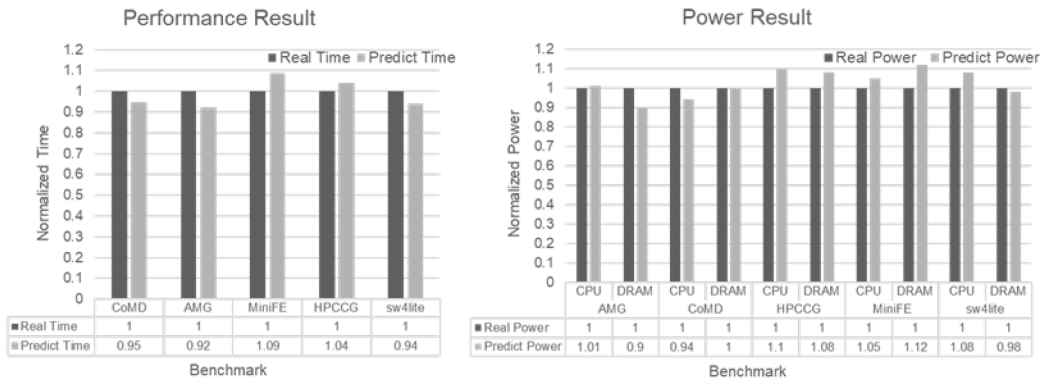

Figure 3 Experiment Result

5. Conclusion

In this Paper, we develop multi-node power and performance model for HPC system. We predicted the performance and power of the application through the amount of data movement between nodes, the internal operation characteristic of the application, and hardware analysis of each node.

Compare with real power/performance and predicted power/performance, the accuracy was 90% and predicted time was fast within one hour.

Currently, we have developed and verified performance and power models for two nodes. But we plan to extend the performance / power model by increasing the number of nodes in the future

6. Acknowledgment

This research was supported by the MOTIE (Ministry of Trade, Industry & Energy) (No.10080590, Technology Development of Unified Memory System for Heterogeneous System Architecture) and KSRC (Korea Semiconductor Research Consortium) support program for the development of the future semiconductor device and by the Graduated School of YONSEI University Research Scholarship Grants in 2017 and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (2016R1A2B4011799) and by the Institute of BioMed-IT, Energy-IT, and Smart-IT Technology (BEST), a Brain Korea 21 plus program, Yonsei University.

References

[1] Adhianto, Laksono, et al. "HPCToolkit: Tools for performance analysis of optimized parallel programs." Concurrency and Computation: Practice and Experience 22.6 (2010): 685-701..

[2] Shende, Sameer S., and Allen D. Malony. "The TAU parallel performance system." The International Journal of High Performance Computing Applications 20.2 (2006): 287-311.

[3] Deelman, Ewa, et al. "PANORAMA: An approach to performance modeling and diagnosis of extreme-scale workflows." The International Journal of High Performance Computing Applications 31.1 (2017): 4-18.

[4] Luk, Chi-Keung, et al. "Pin: building customized program analysis tools with dynamic instrumentation." Acm sigplan notices. Vol. 40. No. 6. ACM, 2005.