Multi-User Concurrent Job Scheduling Method of Network Analysis Application Based on CPU/GPU Cluster

Yadi Luo¹,a, Jing Li¹,*, Jun Lu²,b, Ziming Guo³,c, Bo Yan³,d and Wenyuan Ning³,e

¹China Electric Power Research Institute, Beijing, China
²State Grid Anshan Electric Power Supply Company, Anshan, China
³State Grid Jibei Electric Power Company Limited, Beijing, China

*Corresponding author e-mail: xihuan_89@163.com, aluoyadi_cau@163.com,
bas_luj@163.com, cguo.ziming@jibei.sgcc.com.cn, dyan.bo.c@jibei.sgcc.com.cn,
ening.wenyuan@jibei.sgcc.com.cn

Abstract. Multi-channel multi-core CPU parallel and CPU+GPU heterogeneous parallel are effective means to enhance Network Analysis Application computing. To solve the job scheduling problem of Network Analysis Application with multi-level dispatching, multi-user, multi-task in CPU/GPU heterogeneous cluster environment, a multi-user concurrent job scheduling method for Network Analysis Application in CPU/GPU cluster is proposed. According to the characteristics of Network Analysis Application, it is suggested that state estimation and dispatcher power flow should be accelerated in parallel with CPU, while static security analysis, perturbation power flow calculation and interruption capacity scanning should be accelerated in parallel with CPU+GPU. The job scheduling method proposed in this paper can satisfy the high concurrent requests of multi-level Dispatching & Control Center in the isomorphic and heterogeneous computing environments.

1. Introduction

With the rapid development of UHV AC/DC hybrid power grid, the inter-regional and inter-provincial power grids are increasingly closely linked. The problems of AC-DC coupling, cross-regional resource optimization, global impact of local faults, large-scale interconnection of intermittent new energy generation and clean energy absorption require that the Network Analysis Application of Dispatching & Control Center at state, branches of state and Provincial have the ability of unified analysis and high performance computing. Network-wide unified Network Analysis has the characteristics of diversified services, complex simulation models and huge computing scale. In order to improve the computation speed of Network Analysis, CPU/GPU computing cluster can be built [1]. The method of CPU parallel acceleration or CPU+GPU parallel acceleration can be used to improve the performance of the Network Analysis. Because CPU is good at dealing with complex logic and GPU has powerful floating-point computing ability, the Network Analysis Applications should choose suitable hardware according to computing type and computing scale. At the same time, there are dependencies or priority scheduling problems among different network analysis applications, different users and different Dispatching &
Control Centers. Therefore, multi-user concurrent job scheduling based on CPU/GPU cluster is required to solve the job scheduling problems of multi-level dispatching, multi-user, multi-task Network Analysis.

This paper presents a method of multi-user concurrent job scheduling based on CPU/GPU cluster for the Network Analysis Application. It is responsible for the whole process of job submission from the user to the user's return of execution results, as well as the collection, organization, monitoring and access permission control of computing resources, so as to meet the high concurrent request of multi-level Dispatching & Control Centers for the whole grid unified Network Analysis under isomorphic and heterogeneous computing environments.

2. The current research status of Network Analysis Application
The Network Analysis Application is to analyze and evaluate the operation of power grid, analyze the impact of fault on the safe operation of power grid, and provide accurate real-time data sections for other on-line analysis applications. The Network Analysis Application mainly includes state estimation, dispatcher power flow, sensitivity analysis, and static security analysis, short-circuit current calculation, probability power flow and other functions. In recent years, the research on the Network Analysis Application mainly focuses on improving the computational speed and accuracy.

In order to improve the computational speed of Network Analysis Application, there are two main traditional methods: one is to use boundary equivalence, model simplification or algorithm simplification, which has the problem of losing the accuracy of Network Analysis Application; the other is to improve the speed of Network Analysis Application based on multi-channel and multi-core CPU parallel computing method, which generally adopts coarse-grained parallel computing method of sub-task or network model partition. Limited by the parallelism of the method and the parallel processing capability of CPU, there are bottlenecks in its acceleration performance for large-scale power grid network analysis.

Since 2007, GPU has been gradually applied to general scientific computing. With the introduction of Compute Unified Device Architecture (CUDA) and the third generation of dedicated computing cards by NVIDIA, GPU has become one of the most important high performance parallel computing (HPC) technologies. Compared with multi-core CPU, GPU has strong floating-point computing power and high memory bandwidth. It is a hardware architecture oriented to throughput optimization, and is good at performing the same operation on a large number of data at the same time. GPU has preliminary application research in power system analysis and calculation, but it has not been widely popularized at present. With the development of GPU technology, the further development of related parallel algorithm research and the introduction of artificial intelligence technology, GPU will have a wide application prospect in power system analysis and calculation. One of the research directions is to optimize application programs and to study high performance computing methods for network analysis in heterogeneous multi-core and heterogeneous multi-core cluster environments, in which new high performance computing software and hardware technology is used, and combined with applicable computing technology [2].

3. CPU/GPU heterogeneous cluster
The Dispatching & Control Centers at all levels in China have their own computer clusters, and the main clusters are CPU clusters. In order to make full use of the existing resources, the construction of GPU clusters will connect GPU computing cards as peripherals to CPUs within nodes through high-speed PCI bus, and interconnect them through high-speed Ethernet or high-speed switching network. According to the different types of servers or workstations, there are situations where single GPU card can be installed, multi-GPU card can be installed and no GPU card can be installed. Therefore, the single node in final CPU/GPU heterogeneous cluster can not only contain single-core CPU, multi-core CPU or even multi-CPU computing resources, but also include single GPU or multi-GPU [3].

In view of this situation, this paper proposes a multi-user concurrent job scheduling method for network analysis and application [4-7].
4. The job scheduling process

The job scheduling process is shown in Figure 1. The specific steps are as follows.

1. Client submitting job computing requirements;
2. The network analysis application job requests of Multi-level dispatching and multi-user are sent to the job pool, and jobs are sorted in the job pool.
3. Job pool initiates job calculation request to job scheduler according to the job sequencing.
4. The Job Scheduler initiates a request for computing resources to the resource manager, and issues job assignment instructions to the job scheduling assistance service according to the available computing resources and job categories returned by the resource manager.
5. Job Scheduling Collaborative Services issue job computing commands to available computing nodes;
6. Computing nodes calculate jobs, and through uploading heartbeat, the computing nodes report computing resources to Resource Managers and report job operation to Job Scheduler.
7. Job Scheduler sends job operation to job pool to provide basis for job sequencing.

5. Job Scheduling Strategy

5.1. Job Sorting
Job sequencing should follow the following principles:

1. State estimation is the basic application of Network Analysis. Other network analysis applications need to be calculated on the basis of the results of state estimation. Therefore, in real-time, when the application of state estimation is being calculated in the system, the calculation of other network analysis applications should wait for the end of state estimation calculation.
2. The start-up mode of network analysis application includes periodic operation, event triggering operation and manual start-up. Therefore, there may be an application triggering or running at the same time.
time at a certain time. At this time, it is also necessary to run the next job after the operation of an application operation has been completed.

5.2. Job Assignment
The following principles should be followed when assigning jobs:

1. Because of the simultaneous operation of multi-system, multi-user and multi-task for the network analysis jobs, when there are too many jobs, the server CPU may occupy too much, even the CPU occupies up to 100%, thus affecting the normal operation of the system, so the maximum number of jobs that can run simultaneously can be allocated according to the situation of computing resources;

2. Because of the difference between GPU and CPU’s architecture and programming model, different network analysis applications have clearly adopted CPU parallel or CPU+GPU parallel before software programming. For this reason, the suitable computing nodes for network analysis applications are set up through configuration files. It is suggested that CPU parallel acceleration be used for state estimation and dispatcher power flow, and CPU + GPU parallel acceleration for static security analysis, perturbation power flow calculation and interruption capacity scanning.

3. When the network analysis job fails or timeouts, the abnormal job is forced to end, and the job situation is reported to the job scheduling.

5.3. Job Scheduling Strategy
To sum up, the job scheduling strategy adopted in the Network Analysis Application of multi-user concurrent based on CPU/GPU cluster is first-come-first-served+backfill algorithm priority scheduling strategy, as follows:

1. The overall scheduling strategy serves first come, first served and meets the principle of fairness;

2. For periodic start-up jobs and dependent jobs, backfilling algorithm is used for scheduling.

3. Setting higher priority for emergency start-up jobs, priority scheduling strategy is adopted.

6. Computing resource allocation

6.1. Applicability Analysis of CPU/GPU in Network Analysis Application
Tests and studies show that the acceleration effect of single power flow is not particularly ideal. With the increase of the size of the matrix, the acceleration effect gradually appears, and the acceleration ratio is 2 times at 1000 nodes. Considering the cost-benefit ratio, state estimation and dispatcher power flow application are more suitable for CPU operation.

Batch power flow has a good acceleration effect. For small-node systems, it presents an astonishing acceleration ratio. It is very suitable for probabilistic power flow (Monte Carlo simulation) of small-scale systems. For large-node systems, it can also achieve acceleration effect of more than 100 times, and it is suitable for sensitivity analysis, static security analysis and other applications.

In addition, setting up multi-card GPU and GPU cluster can effectively avoid the problem that all batch power flow calculations cannot be completed at one time due to insufficient memory, and can further improve the computing efficiency.

6.2. Computational Resource Allocation Principle

1. For network analysis jobs using CPU parallel, priority should be given to computing nodes with CPU only. If CPU computing nodes are not idle, the idle CPU + GPU computing nodes should be selected for computing.

2. For network analysis jobs with CPU+GPU parallel acceleration, CPU+GPU heterogeneous computing nodes are selected for calculation.

3. In order to reduce communication delay and communication time, the same network analysis job should be calculated at the same computing node as far as possible.
7. Conclusion
The proposed method can provide job scheduling services for network analysis applications in homogeneous and heterogeneous cluster computing environments, satisfy the high concurrent computing requests of multi-level scheduling, multi-user and multi-task network analysis jobs, and provide technical support for fast computing of network analysis applications.

Acknowledgments
This work was financially supported by Project of State Grid Corporation of China (Research on Key Computing Services for High Performance Network Analysis Applicable to Unified Analysis of the Whole Power Grid, DZB17201800023).

References
[1] CHEN Xuan, KAN BoWen, LIU GuangYi. The Latest Development of GPU and Its Prospective Application in Power System [J]. Electric Power Information and Communication Technology, 2018, 16 (03): 16 - 25.
[2] Tang Can. High Performance Basic Algorithm of Integrated Grid Based on GPU-Acceleration [D]. Zhejiang University, 2017.
[3] Yuling Fang, Qingkui Chen. A real-time and reliable dynamic migration model for concurrent taskflow in a GPU cluster [J]. Cluster Computing, 2019, Vol.22 (2), pp.585 – 599.
[4] WANG XiaoNing, XIAO Haili, CAO Rongqiang. Design and implementation of an optimal job scheduling model in the high performance computing environment [J]. Computer Engineering & Science. 2017, 39 (04): 619 - 626.
[5] LU FengShun, SONG JunQiang, YIN FuKang, ZHANG LiLun. Survey of CPU/GPU Synergetic Parallel Computing [J]. Computer Science. 2011, 38 (03): 5 - 9+46.
[6] YAN Jiankang, CHEN Gengsheng. Improved H-Storm Platform Based Heterogeneous Resource [J]. Computer Engineering, 2018, 44 (4): 1 - 11.
[7] CHEN Wenbin, Yang RuiRui, YU JunQing. Multi-granularity partition and scheduling for stream programs based on multi-CPU and multi-GPU heterogeneous architectures [J]. Computer Engineering & Science. 2017, 39 (01): 15 - 26.