

통합된 컴퓨팅 자원기반 분할된 작업의 총소요시간 최소화를 위한 다중 작업 스케줄링

한석현*, 유기성*, 김호용*, 전주은*, 정영식*

*동국대학교 멀티미디어공학과

e-mail:shhan@dongguk.edu

Multi-Job Scheduling for Minimum Makespan of Decomposed Job based on Integrated Computing Resources

Seok-Hyeon Han*, GiSung Yu*, Hoyong Kim*, Jueun Jeon*,
Young-Sik Jeong*

*Dept of Multimedia Engineering, Dongguk University

요 약

모바일, IoT, 데스크탑의 컴퓨팅 자원을 통합한 환경에서 다중 작업을 처리하는 연구가 진행되고 있다. 통합된 컴퓨팅 자원(Integrated Computing Resources)에서 다중 작업(Multi-Job)을 처리할 경우에는 실시간으로 발생하는 작업 부하 및 대규모의 컴퓨팅 능력이 요구된다. 또한 사용자에게는 단일 작업 처리 시간과 유사한 작업 처리 속도를 제공해야한다. 기존 클라우드 컴퓨팅의 작업 처리 연구에서는 고성능의 컴퓨팅 자원을 이용하여 단일 작업 처리 속도를 향상시키는 연구는 진행되었으나 다중 작업 처리에 대한 연구는 미흡하다. 본 논문에서는 통합된 컴퓨팅 자원에서 두 개 이상의 작업을 수행하여 작업 처리량을 향상 시키는 다중 작업 스케줄링(MJS-MM)을 제안한다. MJS-MM은 서브미션된 작업을 분할(Decomposition)하고 가용 컴퓨팅의 성능기반 작업을 수행하여 총소요시간(Makespan)을 최소화 할 수 있도록 한다.

1. 서론

통합된 컴퓨팅 자원 환경은 모바일, IoT, 데스크탑 등의 동종 및 이기종의 컴퓨팅 자원으로 구성되고 초고속 컴퓨팅 및 다양한 컴퓨팅 분야에 활용이 가능하다. 다중 작업 컴퓨팅은 둘 이상의 사용자가 요청한 다중 작업들을 컴퓨팅 자원을 활용하여 수행하는 것을 말한다. 다양한 컴퓨팅 서비스 제공이 가능한 통합된 컴퓨팅 자원은 컴퓨팅 자원을 추가하여 컴퓨팅 능력을 확장하거나 대용량의 작업 처리가 가능하다. 그러나 통합된 컴퓨팅 자원 환경에서 여러 작업을 동시에 처리하는 연구는 단일 작업의 처리 능력을 극대화 하는 고성능 컴퓨팅 연구에 비해 미흡하다. 또한 통합된 컴퓨팅 자원기반 다중 작업 컴퓨팅의 작업 측면에서는 사용자에게 단일 작업을 처리하는 것과 같은 작업 처리 속도를 제공하기 어렵고, 자원 측면에서는 지속적으로 요청되는 작업을 여러 컴퓨팅 자원으로 분산하여 작업 부하 발생을 줄이는 고가용성 고려가 필요하다[1, 2, 3, 4].

본 논문에서는 통합된 컴퓨팅 자원 환경에서 컴퓨팅

자원 성능을 기반으로 다중 작업을 수행하여 작업의 총소요시간을 최소화 하는 MJS-MM(Multi-Job Scheduling for Minimum Makespan)을 제안한다. MJS-MM은 실시간으로 서브미션되는 작업을 분할하고, 컴퓨팅 성능기반으로 작업의 총소요시간이 최소화할 수 있도록 가용 컴퓨팅 자원에 할당한다.

2. 관련연구

기존 연구에서는 클라우드 컴퓨팅에서 고도의 계산 능력을 요구하는 초고속 컴퓨팅 연구들이 진행되었다[1, 2]. pipsCloud[1]는 원격을 통한 실시간으로 대규모 작업을 위한 동적 자원 관리 제안하였고, Dazhong[2]은 자원 확장에 따른 고성능 컴퓨팅의 성능을 평가하였지만 실시간으로 발생하는 다중 작업 처리에 대한 스케줄링 방법은 미흡하다. 본 논문에서는 다중 작업이 발생할 경우 작업의 총소요시간을 단축시키는 스케줄링을 제공한다.

또한 클라우드 컴퓨팅에서의 컴퓨팅 자원 관리 및 작업 처리 방법에 대한 연구들이 진행되었다[3, 4]. Syed[3]은 클라우드 컴퓨팅에 필요한 컴퓨팅 자원의 공급 방법을 제안 하였으나 이기종의 자원을 통합된 자원으로 구성 고려가 필요하다. Shamsollah[4]은 우선순위에 따른 작업 처리 스케줄링을 제안하였으나 동적으로 변화하는 가용한

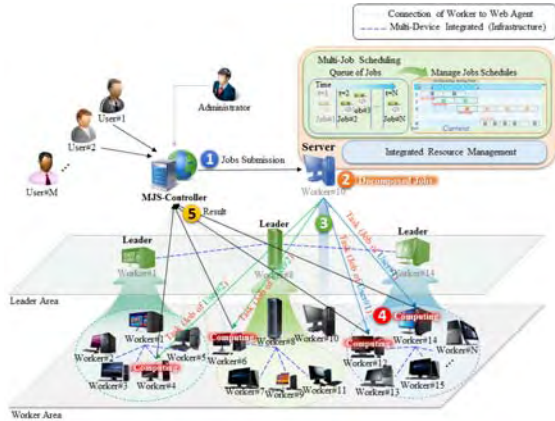
* 이 논문은 교육부와 한국연구재단의 BK21플러스 사업의 장학 지원을 받아 수행된 연구결과임 또한 이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2017R1D1A1A09000631).

컴퓨팅 자원 확인이 어렵다.

이에 통합된 컴퓨팅 자원 환경에서 컴퓨팅 자원 성능을 고려하고 다중 작업을 분할하여 처리하는 다중 작업 처리 스케줄링 연구가 필요하다.

3. MJS-MM 설계

통합된 컴퓨팅 자원 환경에서의 총소요시간 최소화를 위한 다중 작업 스케줄링의 개요는 (그림 1)과 같다.

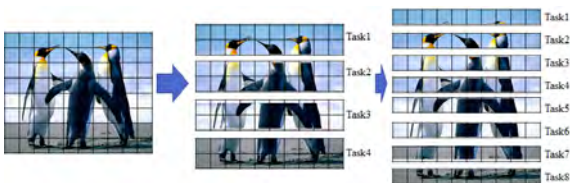


(그림 1) 통합 환경에서의 MJS-MM 개요

(그림 1)의 ①은 사용자가 컴퓨팅 서비스가 가능한 통합된 자원 환경으로 분할이 가능한 작업을 서브미션 한다. ②는 서버가 서브미션된 작업을 실시간으로 작업 큐에 저장하고 워커에 분산하여 처리가 가능하도록 작업을 분할한다. ③은 서버가 워커의 컴퓨팅 자원 성능을 기준으로 순서에 따라 분할된 자원을 할당한다. ④는 워커가 태스크로 분할된 작업을 수행 한다. ⑤는 수행이 완료된 태스크는 사용자에게 전송한다.

통합된 컴퓨팅 자원은 퍼블릭 클라우드에 연결되지 않은 연구실 또는 대학 규모에서 사용되는 PC, 노트북, 서버, 클러스터 등의 컴퓨팅 자원을 통합한 환경이다. MJS-MM에서의 총소요시간은 컴퓨팅 서비스가 가능한 통합된 컴퓨팅 자원 환경에 서브미션된 작업이 분할되고 수행 완료까지의 시간을 뜻한다.

MJS-MM은 작업의 크기를 기준으로 태스크로 분할한다. 가능한 모든 워커에 태스크를 분산 한다. 이미지 프로세싱의 작업 분할의 예는 (그림 2)와 같다.



(그림 2) 작업 분할의 예

MJS-MM의 서버와 워커의 통합된 컴퓨팅 자원 구성 구조도는 (그림 3)와 같다. MJS-MM 통합된 컴퓨팅 자원 구성의 서버 측면은 Resource Manager, Leader Manager,

Job Manager와 워커 측면은 Task Manager, Resource Manager으로 관리된다. 통합 컴퓨팅 자원의 역할은 Viewer를 통해 구성된다.



(그림 3) MJS-MM 구조도

4. 결론

본 논문에서는 통합된 컴퓨팅 자원 환경에서 컴퓨팅 자원 성능을 기반으로 분할된 작업의 총소요시간을 최소화하는 MJS-MM(Multi-Job Scheduling for Minimum Makespan)을 제안하였다. MJS-MM을 통해 다양한 사용자가 요청한 작업의 처리량이 증가하였다.

향후에는 다양한 스케줄링 알고리즘을 적용하여 작업별 성능을 확인할 수 있도록 연구하고자 한다.

참고문헌

[1] Lizhe Wang, Yan Ma, Jining Yan, Victor Chang, Albert Y. Zomayad, "pipsCloud: High performance cloud computing for remote sensing big data management and processing," Future Generation Computer Systems, Vol. 78, No. 1, pp.353-368, Jan. 2018.

[2] Dazhong Wu, Xi Liu, Steve Hebert, Wolfgang Gentzsch, Janis Terpenney, "Democratizing digital design and manufacturing using high performance cloud computing: Performance evaluation and benchmarking," Journal of Manufacturing Systems, Vol. 43, No. 2, pp.316-326, Apr. 2017.

[3] Syed Hamid Hussain Madni, Muhammad Shafie Abd Latiff, Yahaya Coulibaly, Shafi'i Muhammad Abdulhamid, "Resource scheduling for infrastructure as a service(IaaS) in cloud computing: Challenges and opportunities," Journal of Network and Computer Applications, Vol. 68, pp. 173-200, Jun. 2016.

[4] Shamsollah Ghanbari, Mohamed Othmana, "A Priority based Job Scheduling Algorithm in Cloud Computing," International Conference on Advances Science and Contemporary Engineering 2012, Jakarta, Indonesia, 24-25, Oct. 2012, pp. 778 - 785.