

2B1)

3차원 대기오염 모형에 그리드 컴퓨팅 기술 적용 Grid Computing Application to a three dimensional comprehensive air pollution model

조석연 · 조금원¹⁾

인하대학교 환경토목 공학부, ¹⁾KISTI 슈퍼컴퓨터 센터

1. 서론

과거 50여년간 전산기는 성능과 효율면에서 획기적인 발전을 거듭하여 왔다. 더욱이 최근 20여년간 microcomputer와 PC 분야의 발전으로 저렴한 가격으로 전산자원을 활용할 수 있게 되었다. 그러나, 1개의 CPU로 도달할 수 있는 속도의 한계에 근접함에 따라서, CPU 성능의 향상 속도가 느려지고 있다. 현재 초고성능 전산기로도 많은 시간과 비용이 소요되는 문제가 과학 및 공학 분야에 아직 많이 있는 상황에서 이러한 전산속도 발전의 둔화는 심각한 문제가 되고 있다.

이러한 CPU 속도의 문제와는 달리, 최근 10여년간 전산기간 통신이 활발하여지면서, 통신기술의 크게 발달하고 통신기반도 확충됨에 따라서 통신속도는 빠르게 향상되고 있다. 통신기술의 발전을 자료의 공유로 연결하려는 이공학자들의 노력이 world wide web을 탄생시켰고 world wide web은 이공계 분야뿐만 아니라 인문사회과학과 일반 생활에도 큰 영향을 미쳤다. 최근에는 그리드(Grid)라는 새로운 기술이 나와서 world wide web으로는 달성할 수 없었던 분산된 전산자원의 집적화를 선도하고 있다. 여기서, 분산계산의 기반구조를 언급하는 그리드는 1990년대에 만들어진 단어로 일반가정에서 전기줄만 연결하면 전기를 자유로이 사용할 수 있듯이, 전산망만 연결하면 사용자가 세계에 분산되어 있는 전산자원을 자유로이 사용할 수 있음을 의미한다.

대기환경분야는 수천에 달하는 화학종이 복잡한 기상장내에서 광화학반응에 의한 생성 및 소멸과 건성 및 습성에 의한 침착을 주 대상으로 한다. 대기환경 DB 분야는, tera byte에서 penta byte에 달하는 기상과 대기오염도측정 자료 및 배출량 자료를 대상으로 하며 대기환경 모형분야는 삼차원 이동/확산/광화학반응/침착을 대상으로 한다. 특히, 대기환경 모형분야는 방대한 저장용량과 막대한 전산시간을 요구하여 대표적인 슈퍼컴퓨터영역으로 알려져 있다. 물론, 최근 전산기의 발전으로 적지않은 부분의 계산을 데스크탑 워크스테이션으로 처리할 수 있지만, 막대한 전산비용때에 아직도 상세 광화학 반응기작, 에어로졸과 기상 물질과 상호작용과 1년 이상의 장기간 모사는 제한적으로만 수행되고 있다. 이러한, 대기환경분야에 그리드 기술을 적용시키는 연구가 미국과 유럽등지에서 활발하게 일어나고 있다. 본 연구에서는 그리드 컴퓨팅을 적용하여 삼차원 종합 대기오염 모형을 병렬화시켰으며, 병렬화에 따른 수치효율의 증대를 분석하였다.

2. 연구방법

종합대기오염 모형과 같은 화학반응과 유체역학 복합방정식(coupled chemistry and transport equation)이 요구하는 대규모 전산용량을 소화하기 위한 방법으로 분산컴퓨팅이 시도되어 왔다. 분산컴퓨팅 방법은 대상 모형의 수치기법은 물론 사용될 전산기의 구조 및 통신 구조에 의하여 결정된다. 그리드 기반하에 분산 컴퓨팅은 리눅스 네트워크 클러스터의 한 형태인 Beowulf cluster와 유사한 형태를 갖는다. Beowulf cluster는 1994년 NASA에서 Sterling과 Don Becker가 ESS(Earth and Space Sciences) project로 16개 PC를 연결하여 리눅스로 구동하도록 개발한 클러스터를 지칭하며, 이후에 "Beowulf Project"가 발족하면서 Beowulf cluster의 하드웨어와 소프트웨어가 빠르게 발전하였고, 적용분야도 이공학 전분야로 확산되었다.

Beowulf cluster는 한 곳에 있는 여러개의 PC/workstation이 고정적으로 묶어서 작동하는데 반하여, 그리드에서는 Globus라는 software를 사용하여 여러 장소에 분산되어 다수의 PC/workstation를 유동적으로 연결하여 사용하는 장점이 있다. 즉, Globus는 network에 연결된 전산기중에서 사용하지 않는 전

산기를 선정하여 탄력적으로 cluster를 구성함으로써 유휴자원을 최대한으로 활용할 수 있다. 이와 같이 여러개의 PC/workstation을 연결하여 한 개의 계산을 하기 위해서는 MPI(Message Passing Interface) 소프트웨어를 사용하여야 한다. 본 연구에서는 Globus를 지원하는 MPI인 MPCH-G2를 사용하여 FORTRAN 프로그램을 구동하였다.

중합대기오염 모형과 같은 삼차원 모형에 병렬처리하는 기법으로는 sub-domain방법이 널리 사용된다. Sub-domain 방법에서는 대상구간을 수십에서 수백개의 sub-domain으로 구분하고, 각 sub-domain에서 해를 구하고 경계조건 부합을 이용하여 각 sub-domain해를 조율한다. 본 방법은 일반 유체역학 모형과 같이 정상상태 해를 구하는 경우에 기존 해법의 반복과정에 경계조건의 부합성을 자연스럽게 포함시킬 수 있다는 점에서 널리 사용된다. 그러나, 대기오염 종합모형과 같이 시간에 따른 변화를 모사하는 경우에는 부가적인 반복을 통하여 각 시간단계별 해를 조율하여야 하기 때문에 전산시간 요구량이 크게 증가하게 된다. 반면에, 대기오염 종합모형은 time splitting기법을 이용해서 이동과 화학 방정식을 분리할 뿐만아니라, 이동방정식도 1차원방정식으로 분리시킨다. 따라서, 아래와 같이 각 분리된 방정식을 병렬처리하는 방법을 개발하여 병렬화를 이룩하였다. 본 연구에서는 KISTI 슈퍼컴퓨터 센터가 최근에 개발하여 운영하고 있는 test bed를 이용하였는데, 본 test bed는 16개 PC를 가진 cluster 3개를 globus를 통하여 연동하고 있다.

3. 결과 및 고찰

본 연구에서는 국내외에서 널리 사용되고 있는 종합대기오염 모형인 multiscale STEM 모형을 MPI-fortran을 이용하여 병렬화한 STEM-MPI code를 개발하여 PC수에 따른 전산효율을 검토하였다. 아래그림은 2개부터 24개의 PC를 연동할 때에 PC수에 따른 전산효율의 증대(speed-up)을 나타낸 것인데, "grid"라고 표기된 것은 그리드 기반에서 전산효율의 증대를 표시하며, "ideal"은 단순히 PC 수 증가한 것 만큼 전산효율이 증가하는 이상적인 경우를 표시한다. 본 그림에 나타난 바와 같이 PC수가 적을 때에는 이상적 효율증가와 유사한 값을 보이다가 PC수가 증가하면 이상적 효율보다 크게 낮아지는 것을 알 수 있다. 이는 본 연구에서 사용된 병렬기법이 계산과정은 완전히 병렬화하지만 입력은 부분적으로만 병렬화되기 때문으로 사료된다.

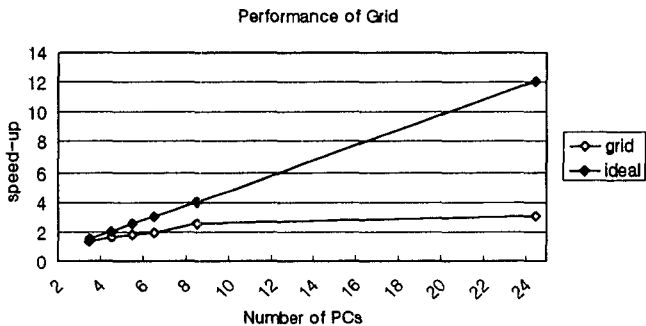


Fig. 1. STEM-MPI의 병렬 효율