높은 집적도를 가지는 Windows XP PC 클러스터 구축

이 성 기[*1], 신 재 렬[*2], 최 정 열[*3]

Construction of Highly Integrated PC Cluster based on Windows XP

S.-K. Lee[*1] and J.-R. Shin[*2] and J.-Y. Choi[*3]

ABSTRACT

A new PC cluster was designed and constructed based on Windows XP Operating system. Primary target of the present design was the high node density per rack by using the general PC parts those are cost-effective and readily available in the market. Other major design points were system cooling and the convenient maintenance using standard PC parts. Presently 24 nodes per rack seems to be optimum considering the specification of the network switching device, system cooling and power supply, but 40 nodes can be accommodated within a single rack at maximum. Windows XP was selected as a high-performance computing environment considering the cost and the convenience in acquisition, maintenance and education. Both Fast-Ethernet and Gigabit Ethernet network connection were tested and compared with previous data, especially for Linux cluter using Myrinet. The result shows that there is no significant difference between the operating systems and the Fast-Ethernet and/or Gigabit Ethernet are good solution for the high-performance PC cluster considering the cost and performance.

Key Words: PC 클러스터(PC cluster), 병렬처리 (Parallel processing), 윈도우즈(Windows)

1. 서 론

병렬처리기법의 개발로 인해 슈퍼컴퓨팅의 양상과 규모는 현저한 변화를 가져왔으며, 인터넷의 확산으로 인하여 저가의 고성능 네트워크 장비가 보급됨으로서, 슈퍼컴퓨터에 비해 월등한 가격 대 성능 비를 보이는 PC 클러스터의 구축이 실험실 수준에서 가능하게 되었으며, 이를 이용한 병렬 슈퍼컴퓨팅이 널리 확산되고 있다.

90년대 후반부터 MPP, SMP등의 슈퍼컴퓨터는 감소 추세를 보였고, 반대로 병렬연결을 이용한 랙마운트형 컴퓨터가 가격대비 높은 성능을 장점으로 하여 2000년대 들어서면서 강세를 보이고 있다. Fig. 1은 성능에 따른 순위 500이내 슈퍼컴퓨터의 종류별 개체수를 나타낸 그래프로, 유형별 분포를 보면 70, 80년대 시장에서 강세를 보이던 벡터형 슈퍼컴퓨터는 90년대 이후 병렬형 슈퍼컴퓨터에 밀려 점차 감소해 2002년 말 기준으로 7.4% 수준에 그치고 있다. 반면, 2000년도를 기점으로 하여 클러스터가 차지하는 비율은 아주 빠른 속도로 증가하고 있음을 보여준다.[1]

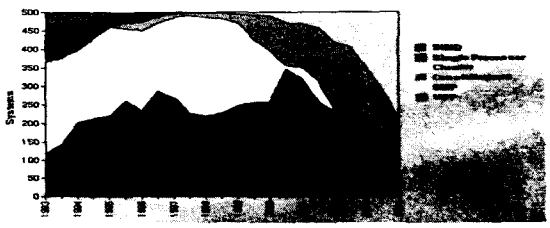


Fig. 1 슈퍼컴퓨터의 추세

*1 부산대학교 학부 항공우주공학과
*2 부산대학교 대학원 항공우주공학과
*3 부산대학교 항공우주공학과
*E-mail : aerochoi@pusan.ac.kr

우리나라의 슈퍼컴퓨터 시장의 전망 역시 병렬컴퓨터의 사용이 확대될 것으로 예상되며, 이러한 경향에 가장 큰 영향을 미치는 요인은 역시 가격으로 생각된다. 따라서 본 연구실에서는 일반적으로 시장에서 매우 저렴한 가격에 구할 수 있는 범용의 PC 구성 품을 이용하여, 이용 및 유지 보수의 편의성과 시스템의 낸각의 측면에서도 우수한 고집적도의 PC 클러스터를 설계 제작하였다. 시스템 구축, 유지 보수, 사용자 교육의 비용과 편의성을 고려하여 네트워크 장치와 운영체제는 각각 Fast-Ethernet 및 Gigabit Ethernet 기반의 네트워크와 Windows XP 운영체제를 이용하여 병렬처리 환경을 구축하였으며, 기존의 리눅스 및 Myrinet 기반 환경과 비교하여 시스템의 성능을 평가 하였다.

2. Windows XP 설치 및 네트워크 구성

2.1 Windows XP 설치

2.1.1 Windows XP설치 가능성

클러스터는 네트워크로 묶여져 하나의 자원처럼 사용되어지는 PC들의 집합으로서, 최근의 다양한 운용시스템을 가지는 PC클러스터들이 선보고 있다.[2] 이중 가장 널리 이용되고 있는 운용체제는 표준화된 네트워크 구성에 강점을 가지는 유닉스 기반의 리눅스 PC 클러스터로써, 운용체제가 공개되어 있어 저렴하게 PC클러스터를 구축할 수 있는 강점이 있으나, PC 운영체제의 주류인 Windows 환경에 익숙한 사용자들이 새로운 환경을 익혀야 한다는 단점을 가지며, 다양한 응용프로그램의 부재와 서로 다른 운용체제 환경을 병행 운용하여야하는 번거로움이 있다.

슈퍼컴퓨팅 부문에는 Windows 환경을 이용하고자 하는 시도가 과거 수년간 있었으나, 현재로서는 거의 이용되고 있지 않다. 그러나 네트워크 성능이 보강된 Windows 2000의 개발에 따라 병렬 슈퍼 컴퓨팅에도 유용하게 이용될 수 있는 여지를 보이고 있다. 따라서 본 연구에서는 Windows XP 기반의 PC 클러스터를 Fast-Ethernet, Gigabit-Ethernet으로 구축하여 보았고, 기존의 리눅스 기반의 Myrinet-Cluster와 성능을 비교 평가 하였다.[3]

2.1.2 Windows XP 설치

Windows XP의 설치는 Windows의 다른 운영체제의 설치와 다르지 않고, 단지 Cluster를 구성하고 운영하기 위한 서버의 특수 기능만이 다르다. 따라서 서버에는 Windows 2003 Server를 설치하였으며, 각 노드에는 Windows XP Professional을 설치하였다. 각 사용자 계정 및 네트워크 자원에 대한 정보를 관리하기 위하여 서버에는 Netsupport[4]란 원격지원 프로그램을 설치하였으며, 이를 통해서 클라이언트에 접속한 사용자나 수행하고 있는 작업을 확인하고 관리할 수 있다. 한편 외부에서는 서버에 사용자계정을 통해 원격으로 클러스터를 이용할 수 있도록 Terminal Server의 기능을 수행하도록 설정하였으며, MS 네트워크 클라이언트를 통하여 공유자원에 대한 접근이 가능하며, Telnet이나 FTP와 같은 원격지 서비스도 지원하도록 하였다.

2. 2 네트워크 구성

한편 네트워크 구성의 경우 병렬처리 도구들은 주로 TCP(Transport Control Protocol), VIA(Virtual Interface Architecture)같은 통신 매체를 이용하므로, 본 연구에서는 TCP와 DNS(Domain Name Service) 또는 work group을 이용하여 네트워크를 구축하였다.

Fig. 2는 클러스터의 네트워크 구성을 나타낸다. 여기서 서버는 외부 네트워크에 접속하기 위한 공인 IP를 가지고 있으며, 내부 네트워크는 사설 IP를 이용하여 구성하였다. 따라서 서버에는 두개의 NIC가 필요하며, 네트워크는 한 개의 도메인에 클라이언트들이 존재하는 형태이다. 따라서 외부에서는 서버 이외에 다른 클라이언트에는 직접 접속할 수 없으며 외부에서는 하나의 도메인 네임을 가지는 단일 시스템으로 인식된다. 이것이 전형적인 네트워크 설정이나 현재는 workgroup을 이용하여 네트워크를 구현했다.

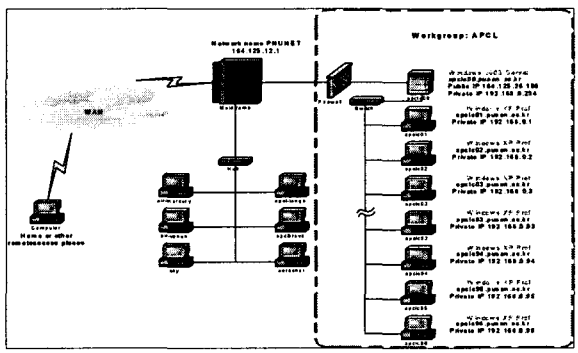


Fig. 2 Network Design

3. 병렬처리 도구 및 설치

병렬처리 클러스터는 분산 메모리 체계이므로 기본적으로 메시지 패싱을 통하여 병렬처리를 구현한다. 메시지 패싱 규약에는 PVM, MPI 등이 있지만 본 연구에서는 가장 널리 이용되는 MPI를 이용하여 구성하였다. Windows 시스템에 이용될 수 있는 공개된 MPI 패키지로는 MPICH.NT, MP-MPICH 등이 있으며, 상용으로는 WMPI, MPI-Pro 그리고 PaTENT이 있다.[5-9] 이들 중에 여러 패키지들이 MPICH.NT를 기본으로 개발되었으므로 본 연구에서는 MPICH.NT.1.2.5를 이용하여 병렬성능을 계산하였다. 한편, 컴파일러는 Windows 환경에서 널리 이용되고 있으며, 좋은 성능을 보이는 것으로 알려진 Compaq Visual Fortran 6.5를 이용하였다.

4. Windows XP PC Cluster의 구성

4.1 Hardware 구성

이번에 제작한 PC클러스터는 케이스가 없는 경우로, PC 부품을 이용하는 클러스터의 최대 장점은 가격 당 성능 비율이다. 그러나 이전의 PC클러스터는 병렬로 연결하였을 경우 부피를 많이 차지하게 되어 수 십대를 연결하기 위해서는 많은 공간을 필요로 한다는 단점을 가졌다. 이를 극복하기 위해서 랙 당 고 노드 집적도의 클러스터를 구축하여 공간 활용의 효율성을 가지도록 하였으며, 또한 수 십대의 CPU가 많은 열을 발생시키므로 케이스가 없는 블레이드 타입의 클러스터를 제작하여 냉각의 효율성도 높아지도록 하였다. 이렇게 함으로서, 랙마운트에 16대 구성 하였던 예전과 비교하여, 24대의 블레이드를 구성함으로서 공간 이용효율을 50%증가 시킬 수 있었다.

Table 1은 본 연구에서 구축한 2개의 Windows 클러스터들의 사양으로 100Mbps의 Fast Ethernet을 구성하고, 이를 바탕으로 1Gbps Gigabit Ethernet으로 업그레이드 시켜 보았다.

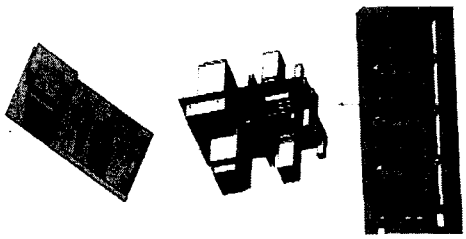
CPU는 AMD 3000+를 선택하여, 랙당 24노드-24프로세서, 전체 48개의 프로세서로 24포트의 허브에 연결되어 2개의 네트워크로 구성되어있다.

Table 1. Specification of Each Node of Clusters

	Fast-Ethernet Cluster	Gigabit Cluster
CPU	1x AMD Athlon XP 3000+	
M/B	ASUS A7N8K-VM/400	
Chipset	nForce2 IGP +nForce2 MCP	
Memory	512MB DDR-SDRAM	
HDD	IDE 40GB	
NIC	Fast Ethernet (on board)	Gigabit Ethernet Adapter
Network switch	Dell Power Connect 2124	Cisco-Linksys SR2024
OS	Windows XP	
MPI .lib	MPICH.NT.1.2.5	
Complier	Compaq Visual Fortran 6.5	

4. 2 설계 및 제작

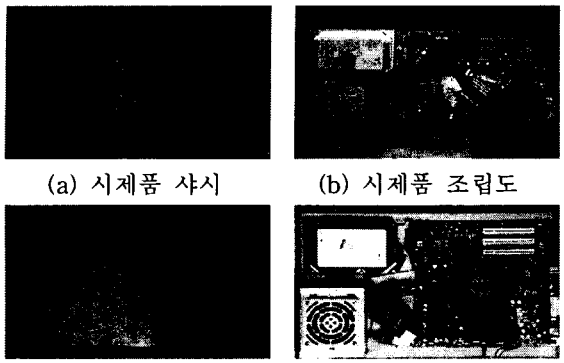
많은 양의 블레이드를 효율적으로 배치하기 위해 CATIA를 이용해서 Fig. 3과 같이 설계를 해보았다. 랙의 크기(0.6×0.75×1.8m)가 정해져 있기 때문에 최대한 많은 블레이드(0.43×0.28m)를 넣기 위해서 판의 양면과 한 면으로 된 블레이드를 간섭체크를 통해 설계를 하였다.



(a) 조립도 1 (b) 조립도 2 (c) 조립도 3

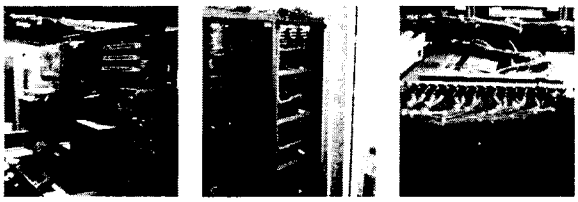
Fig. 3 CATIA를 이용한 설계도면

Fig. 4는 시제품을 만들어서 각 부품이 조립될 위치를 정하고, 설계한 결과가 적절한지를 판단 후, 제작의 효율성을 위해서 외부 용역을 통해 판을 제작하였다. 이 판을 블레이드라 하고 이러한 판에 CPU와 주변장치들을 장착하여 랙을 구성한 것을 블레이드 형 병렬 컴퓨터라 명하였다. Fig. 5는 시스템을 구성해 놓은 전체 사진이다.



(a) 시제품 샤시 (b) 시제품 조립도
(c) 양산품 샤시 (d) 양산품 조립도

Fig. 4 블레이드의 제작



(a) 내부 (b) 외부 (c) 네트워크연결

Fig. 5 시스템 구성

5. 성능 결과 분석

5.1 성능 평가 모델

병렬처리 성능 평가는 이전의 연구에서와 일관성을 유지하기 위하여 이차원 압축성/점성 유동의 예조건화 해석 코드를 이용하였다.[10] 낮은 마하수의 압축기 익렬 유동해석 문제로서, 계산 및 통신 부하에 따른 영향을 살펴보기 위하여 적은수의 211× 71 격자(coarse)와 많은 수의 631x211격자(fine)에 대한 해석을 각각 1000번과 100번의 반복수행에 따른 계산 시간을 평가하였다.

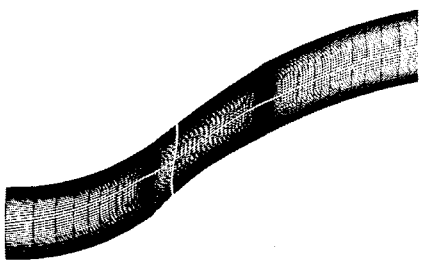


Fig. 6 Grid System of Parallel Computing

해석에 이용된 수치 기법으로는 공간이산화에 Roe의 풍상 차분법과 TVD-MUSCL을 이용한 고차정확도 기법이, 시간적분에는 Gauss-Seidel 내재적 해법이 이용되었으며, 난류모델로는 Menter의 k-ω SST 모델을 적용하였다.[10] 한편 병렬처리는 영역분할을 통하여 구현하였으며, 각 분할 영역에 동일한 수의 격자수를 배분함으로써 연산 부하를 동일하게 배분하였다. Fig. 6은 압축기 익렬의 격자를 구성한 것이다.

5.2 결과 분석

병렬처리 성능을 비교 평가하기 위하여 Windows Cluster와 Linux Cluster의 처리 성능을 비교하였다. 유사한 사양을 가지지만, 구성 시점에 따라 프로세서 및 네트워크 사양에는 약간씩 차이가 있다.

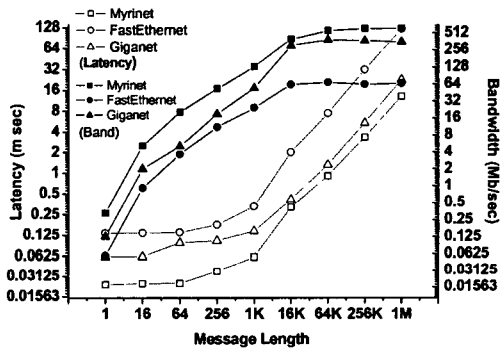


Fig. 7 Latency and Bandwidth

Fig. 7은 세 가지 클러스터의 데이터 크기에 따른 지연시간과 대역폭을 나타낸 것이다. 네크워크의 경우 이들의 이론상의 대역폭은 Fast-Ethernet 100 Mbps와 Gigabit Ethernet 1Gbps이며, 비교를 위한 Myrinet은 1Gbps지만 실험실상에 네트워크를 구축했을 때는 이보다 작은 대역폭을 갖는다. 지체시간의 측정 결과, Fast-Ethernet은 수 십 밀리 초, Myrinet과 Giga bit-Ethernet은 수 십 마이크로초의 지체시간을 가진다. Myrinet을 이용한 리눅스 클러스터와 Fast-Ethernet과 Gigabit-Ethernet으로 구축한 윈도우즈 클러스터의 경우 장비의 성능 차이를 빼고는 같은 경향을 띈다. Fast-Ethernet은 데이터의 양이 16KB 이상 전송 시 Myrinet 리눅스 클러스터보다 네트워크 성능이 떨어짐을 보이나 Gigabit-Etherent 윈도우즈 클러스터의 경우 16KB이상 전송 시 Myrinet 리눅스 클러스터와 유사한 네트워크 성능을 보이고 있다. 따라서 OS에 의한 네트워크 성능의 차이는 없다고 할 수 있으며 또한, 클러스터는 대용량의 계산에 쓰일 것

이기 때문에 저용량에서의 대여폭 손실을 고려할 필요가 없을 것이다. 한편 MPI Library의 성능도 네트워크의 성능에 영향을 미치므로 병렬처리 시스템의 성능은 종합적으로 평가하여야 한다.

병렬 성능을 평가하기 위해, Speed Up과 절대시간으로 나타내 비교 평가하였다. Speed Up은 계산시간을 CPU 1개를 사용하였을 경우 계산시간과 비교하여 나타낸 것으로 병렬연결의 효율을 평가하기 위한 것이고, 절대시간은 CPU 개수의 증가에 따른 계산 처리시간의 역수를 취하여 나타내어 시스템의 전체적인 성능을 비교 평가하기 위한 것이다.

Fig. 8은 프로세서의 증가에 따른 처리 속도를 측정한 것으로 격자수에 따라 구분해서 측정해 보았다. 격자수를 적게 하여 계산을 수행할 경우 일정한 수준에 이르러서는 CPU가 증가해도 계산시간이 줄어들지 않았다. 이것은 하드웨어 및 주변 장치들의 한계와 계산처리 데이터 크기가 작기 때문에 Fig. 7의 결과에서 알 수 있듯이 데이터 크기가 작을 경우 지연시간이 길고 대여폭의 손실이 커지기 때문이라 할 수 있다. 따라서 CPU가 더 증가하게 되더라도 계산속도의 한계를 가지게 된다. 반면, 격자수가 많아질수록 데이터의 크기가 커지기 때문에 대여폭의 손실이 줄어들어 fine grid의 그래프와 같이 병렬처리에 의한 계산속도의 향상을 볼 수 있다.

$$speed\,up = \frac{Single\ CPU\ Computing\ time}{Computing\ time}$$

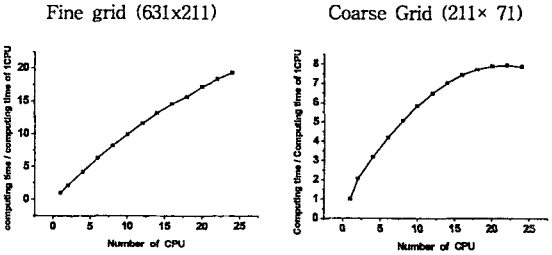


Fig. 8. Speed up of Gigabit Blade type Cluster

Fig. 9는 Fig. 8의 결과와 Fast-Ethernet 클러스터와 Myrinet 클러스터를 비교한 것이다. 우선, Fast-Ethernet 클러스터와 비교해 보면 CPU와 기타 하드웨어가 같음을 고려할 때 대여폭이 늘어날 경우 병렬 효율이 더 좋아짐을 알 수 있고, Myrient과 비교해 보면 네트워크 장비차이에 의한 대여폭의 손실이 더 크더라도 CPU 및 기타 하드웨어의 성능차이에 의해 병렬효율이 더 좋아졌음을 알 수 있다.

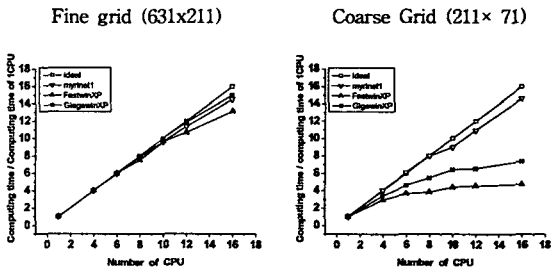


Fig. 9 Speed-Up Ratio

Fig. 10은 절대시간을 나타낸 것으로 계산처리시간의 역수를 그래프로 나타낸 것으로 계산 처리 성능을 비교해 놓은 것이다. CPU 및 기타하드웨어와 네트워크의 대여폭이 커질수록 클러스터의 성능이 우수해지고 있음을 볼 수 있다.

$$Absolute\ Time = \frac{1}{Computing\ time}$$

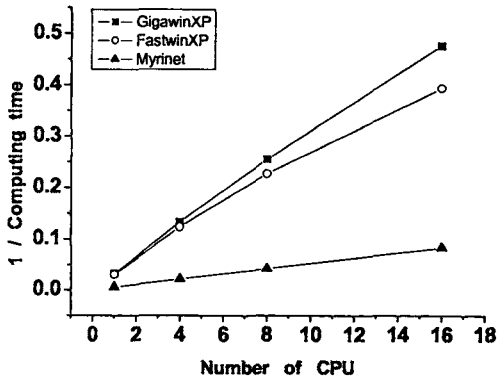


Fig. 10 Absolute Time of Fine grid

6. 발생 열의 비교

PC의 선택에 있어서 가장 먼저 고려되어야 할 것이 CPU의 종류이다. 본 연구에서는 부가부품이 많이 필요하지 않고 저비용으로 고성능을 지양하고 있으므로 AMD CPU를 선택하였다. 그러나 앞에서도 언급했듯이 AMD의 단점이 열이 많이 발생한다는 것이다. 발생하는 열에 대한 처리가 보완이 된다면 AMD CPU의 단점을 극복하고 가격대 성능 비를 향상시킬 수 있다. 따라서 블레이드 형으로 제작함으로

서 이를 보완해 보았다.

Fig. 11은 기존 보유하고 있던 병렬컴퓨터(Intel 1GHz, AMD 1600+)들과 새로 제작한 블레이드 형 병렬컴퓨터(AMD 3000+)에서 발생하는 열을 비교해 놓은 것이다. 기존의 것은 Intel은 랙당 8대를 구성한 것이고, AMD 1600은 랙당 16대를 구성한 것이며, 블레이드 형 병렬컴퓨터는 랙당 24대를 구성하였다. 병렬컴퓨터가 위치한 실내는 에어컨디셔너를 이용해 냉각을 하고 있는데 비교를 위해 냉각이 없을 경우와 냉각을 할 경우로 구분해 측정을 해보았다. 냉각이 없는 경우는 에어컨디셔너 작동을 멈춘 후 1시간이 지난 뒤 측정한 것이다.

냉각을 하고 있을 경우, Intel 1GHz와 블레이드 형(AMD3000+) 보다 AMD 1600의 경우가 온도가 높게 측정되었다. 이는 블레이드 형으로 하였을 경우 공기가 순환하기 쉽기 때문에 냉각도 용이함을 알 수 있다. AMD의 경우 CPU의 성능이 좋아 질수록 전력 소비가 많아지고 이에 따라 열이 많이 발생하게 되는데 이 실험에서 알 수 있듯이 블레이드 형의 클러스터가 냉각이 용이하고 시스템의 안정성을 가짐을 확인할 수 있다.

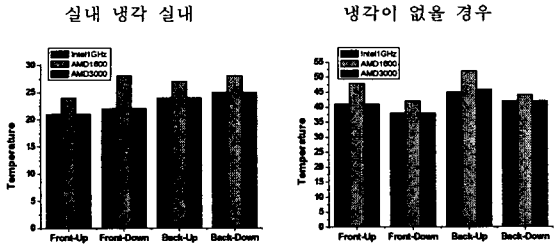


Fig. 11 발생하는 온도의 비교

냉각을 하지 않았을 경우, 온도가 50도 가까이 올라갔다. 특히 AMD 1600의 경우 50도가 넘었고, AMD 3000+는 46도로 측정되었기 때문에, 역시 블레이드 형으로 제작한 경우가 열 발생에 의한 손상을 줄일 수 있다는 것을 알 수 있다.

7. 결 론

본 연구에서는 Windows XP PC 클러스터 환경에 기초한 병렬 슈퍼컴퓨팅 환경을 구성하였으며, 기존의 리눅스 클러스터와 병렬 라이브러리 성능을 비교 평가하였다. 이 결과, Windows 환경은 저가의 보편적인 하드웨어 구성에도 불구하고 기존의 리눅스 클러스터와 비교 시 보다 우수하고 안정적인 병렬처리 성능을 보여 주었고, 또한 시스템 구성을 블레이드 형으로 제작함으로서 공간 이용의 효율성과 열 발생에 의한 손상의 위험을 줄일 수 있었다. 본 연구에서 구성한 시스템은 더 많은 수의 노드로의 확장에 제한이 없으므로, 이용자의 편의와, 다양한 소프트웨어와 높은 호환성, 단일 운영체제에 의한 컴퓨팅 환경 구성 등의 장점을 고려한다면 슈퍼컴퓨팅 환경의 새로운 대안으로 이용될 수 있을 것이다. 더욱이 Fig. 9의 절대 계산시간을 비교한다면 Windows 환경의 우수한 컴파일러를 이용할 수 있음에 기인하여 매우 우월한 성능을 보임을 알 수 있다. 또한 널리 확산된 인터넷과 Windows 운영체제를 기반으로 병렬처리를 위한 네트워크를 구성하는데 이론적인 제한이 없으므로, 향후 연구에서는 블레이드형 병렬컴퓨터의 문제점을 보완하고 인터넷 슈퍼컴퓨팅 환경으로서 Windows 운영체제의 이용가능성을 모색할 것이다.

참고문헌

[1] 'Super computer top 500' http://www.top500.org
[2] Baker. M, 2000, "Cluster Computing White Paper", Version2.0, University of Portsmouth, UK, Sept. http://www.dcs.port.ac.uk
[3] Pabst. T and Völkel. F, Spet, 2001, "Hot Spot - How Mo dern Processors Cope With Heat Emerg encies", http://www.tomshardware. com /cpu/ 01q3/010917/index.html.
[4] 'Netsupport.' http://www.syter.com/nsm/nsm/ f eatures.shtml
[5] "MPICH.NT.1.2.2", http://www-unix.mcs.anl.g o v/mpi/index.html
[6] "MP-MPICH.1.2",http://www.lfbs.rwth-aachen.d e/ ~karsten/projects/nt-mpich/index.html
[7] "Wmpi1.5", http://www.criticalsoftware.com/ w mpi/home/index.html
[8] "MPI-Pro", http://www.mpi-softtech.com/def au lt.asp
[9] "PaTENT",http://www.genias.net/geniasde.html
[10] 이기수, 김명호, 최정열 외, 1992, "Myrinet과 Fast-Ethernet PC Cluster에서 예조건화 Navier -Stokes 코드의 병렬화 효율 비교," 2001년 항공우주공학회 춘계학술발표회 논문, pp.18.