홈기반 분산공유메모리 상에서 결함허용시스템의 설계

김용국[0] 이성우 유기영
경 북 대 학 교 컴 퓨 터 공 학 과
{ykkim, swlee}@purple.knu.ac.kr, yook@knu.ac.kr

Fault-Tolerance on Home-based Distributed Shared Memory

Yong-Kuk Kim[0] Sung-Woo Lee Kee-Young Yoo
Dept. of Computer Engineering, Kyungpook National University

요 약

분산공유메모리의 성능향상을 위한 연구가 많이 진행 되고 있다. 홈 기반의 프로토콜은 분산공유메모리의 가장 큰 성능 저하 요인인 외부통신비용을 획기적으로 줄임으로써 어느정도의 괄목할 만한 결과를 보여 주고 있다. 본 논문은 홈기반 프로토콜의 적은 통신량과 신속한 자료의 저장 및 삭제능력에 기반한 결함 허용 시스템을 제안하며, 이러한 구현이 전체 시스템에 어느정도의 영향을 미치는지 실험한다.

1. 서론

고성능 범용 CPU의 사용이 점차 일반화 됨으로써 이미 슈퍼컴 등의 용도로 쓰이던 특수용도 CPU나 장비등은 가격경쟁력을 상실해가고 있다. NOW나 Linux-cluster등은 이러한 고성능 범용CPU와 우리주변에 흔히 쓰이는 이더넷 등을 매개체로 하여 보다 저렴한 가격으로 Multi-computing 환경을 꾸미고 있으며 이러한 추세는 점차 확산되고 있다. 하지만 아직도 이러한 시스템들이 메시지 기반의 분산 모델을 채택하고 있어서 일반 프로그래머들이 이러한 분산환경에서 동작하는 프로그램을 작성하기가 쉽지 않다. 그래서 제안된 모델중의 하나가 DSM(Distributed Shared Memory)모델이다. 이 DSM은 분산환경을 병렬 환경처럼 보이게 함으로써 프로그래머들이 보다 프로그램을 용이하게 작성할 수 있게 해주는 모델이다. DSM에 쓰이는 프로토콜은 여러 가지가 있는데 그 중 홈기반(Home-based)프로토콜은 최근 가장 활발히 연구되고 있는 프로토콜중 하나이다.

상업용 시스템에서는 아직 DSM이 많이 사용되고 있지 못하고 있다. 이러한 이유중의 하나는 DSM이 HA(High Available)시스템 보다는 HP(High Performance)시스템에 중점을 두고 있기 때문이다. 상업적으로 사용되는 시스템은 HP보다는 HA적인 요소가 더 많이 작용한다. 본 논문에서는 홈기반 DSM상에서 결함허용(Fault-Tolerance)을 구현하고 이러한 구현이 전체 시스템에 미치는 영향을 분석한다. 제안된 시스템에서는 경쟁상황(race condition)이 발생하지 않는다는 가정하에 Barrier와 Recoverpoint를 이용한다.

2. 관련연구

Princeton대학의 홈기반 프로토콜[2,3]은 P.Keleher가 제안한 LMW프로토콜[1]과 유사하지만 프로그램에 의해 각각의 공유페이지에 정적으로 홈 프로세서가 할당된다는 특성이 있다.[2]

홈기반 프로토콜에서 해제(release)작업을 할 때, 한 프로세서는 자신의 직전 해제이후에 수정된 모든 페이지들에 대해 차이본들을 생성하고 이들을 해당페이지의 홈 프로세서에게 전송한다. 홈기반 프로토콜은 일관성을 유지하기 위해 페이지 버전 시스템을 사용한다. 홈에 있는 페이지 버전은 홈이 다른 쓰기 프로세서에 의해 홈 복원이 이루어 질 때 마다 1씩 증가 한다. 홈은 다른 프로세서의 홈 복원 메시지나 페이지 요구 메시지에 대한 응답메시지에 버전을 실어 보낸다. 또한 모든 프로세서들의 락이나 barrier 메시지에 포함되어지는 쓰기 통보에도 페이지 버전이 포함된다. 요구 프로세서는 전송된 쓰기 통보들을 통합시킬 때, 자신의 페이지 버전이 전송된 쓰기통보의 버전보다 작다면 그 페이지를 무효화 시킨다. 만약 자신의 버전이 크거나 같다면 현재 페이지는 해당 쓰기 통보의 복원내용이 반영되었으므로 무효화 시킬 필요 없다. 여기서, 홈 프로세서의 해당 페이지들은 쓰기 보호 상태는 될 수 있지만 무효화 상태는 될 수 없다.

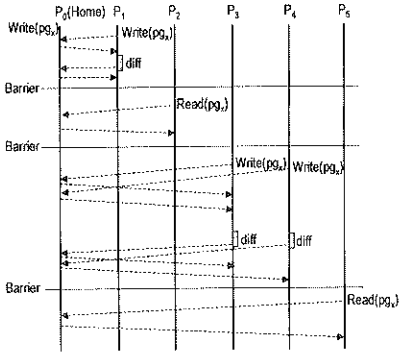

그림.1 홈기반 프로토콜의 작동

홈 프로토콜[4]은 홈을 가짐으로 생기는 장점과 많은 양의 자료구조가 빨리 삭제될 수 있다는 장점이 있다. 차이본 생성과 차이본 적용에 대한 비용을 고려하지 않더라도 차이본들은 프로토콜이 사용하는 메모리 오버헤드의 대부분을 차지한다. 홈 노드가 자신의 페이지를 접근할 때 외부로부터 차이본을 받아 올 필요도 없고 홈의 차이본은 다른 프로세서에게 전송될 필요도 없다. 따라서 외부통신을 발생시키지 않는다. 결과적으로 홈기반의 DSM시스템은 자료의 삭제와 저장측면에서 적은 외부통신요소를 가지므로 외부상황(Network stats)에 보다 강하기(robustness) 때문에 결함허용 시스템을 구성하는데 상당히 유리한 환경이다.

3.홈기반 LRC상에서 결함허용(Fault-Tolerance) 모델구현

홈 기반 프로토콜의 결함허용을 위해 본 논문에서는 recoverpoint 서버에 기반한 모델을 제시한다. 여기서 또 recoverpoint는 지역적 recoverpoint와 전역적 recoverpoint로 나뉘는데 이렇게 나누는 것은 단일 오류와 다중오류에 모두 대처할 수 있게 하기 위함이다.

결함허용을 위한 시나리오는 그림2와 같다. Recoverpoint를 이용한 결함허용은 항상 barrier의 동기시점부터 결함에 대한 복구가 진행된다. 모든 프로세서는 barrier에서 동기화 되며, 이 때 각 프로세서는 각자의 페이지 복사본을 갱신한다. 각 프로세서는 홈에 위치한 recoverpoint 서버에게 각자의 recoverpoint를 저장하도록 요청(rec_request)메시지를 보내고 응답(rec_reply)메시지를 받는다. 모든 프로세서가 recoverpoint를 서버로 보내 저장한 후 recoverpoint 서버는 승낙(rec_ack)메시지를 모든 프로세서에게 보내며, 이후 각 프로세서는 자신의 주어진 작업을 계속해 나간다.

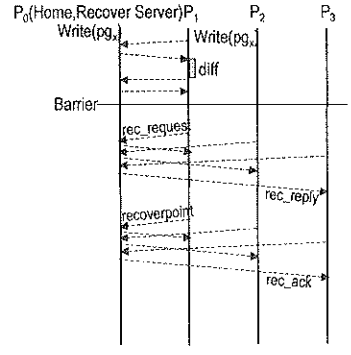

그림.2 recoverpoint의 작동

홈기반 프로토콜은 각각의 공유페이지에 있는 프로그램에 의해 정적으로 프로세서가 할당되므로 각 프로세서들은 독립적으로 recoverpoint를 가질 수 있다. Recoverpoint의 최적화를 위해 새로운 recoverpoint가 생성되면 이전의 recoveroint는 폐기된다. 이 특성은 메모리 사용 효율성을 증가시킨다. 홈기반 프로토콜에 의한 전역적인 불필요정보처리(garbage collection)로 일관된 전역적 recoverpoint의 생성이 가능해진다.이 전역적 recoverpoint는 1개 이상의 다중결함(fault)이 발생했을 때 사용된다.

Recoverpoint의 상태는 동일 차이본(diff)에 의해 복구가 가능하므로 기본적으로 모든 공유페이지들은 recoverpoint에서부터 실행할 수 있다.

Recoverpoint에 의한 단일 프로세서의 복구시 한 프로세서의 오류가 감지되면 오류가 난 프로세서는 자신의 가장 최신의 recoverpoint로부터 복구를 시작하며, 복구중인 이 프로세서는 복구하고 있다는 메시지를 기타 다른 모든 프로세서들에게 알린다. 이 메시지는 프로세서의 현재 시간(T_{recpnt})과 마지막으로 지역적 자료가 생성된 시간(T_{local})으로 구성된다. 모든 다른 프로세서들은 자신의 Vector Time과 T_{local}을 비교하여 자기의 Vector Time이 T_{local}보다 크다면 자신의 Vector Time을 홈으로 보내서 현재 복구중인 프로세서의 차이본과 비교한후 차이본을 재생성하여 전달받고 지역 메모리에 저장한다.복구중인 프로세서는 가장 큰 Vector Time을 가진 프로세서에게 T_{local}보다 큰 모든 interval을 요구한다.홈은 모든 프로세서로부터 T_{local}을 받아서 차이본을 생성하고 차이본과 T_{local}을 해당 프로세서에게 보낸다. 차이본과 T_{local}을 받은 프로세서들은 그 정보를 지역메모리에 저장한다. 홈은 마지막 Recoverpoint 기간부터 오류가 발생한 기간까지의 모든 interval을 리스트화 하여 RecIntList[p]에 저장한다.다중오류가 발생하면 모든 프로세서는 마지막으로 변경된 전역recoverpoint로 복귀(rollback)하며, 전역 recoverpoint 역시 개별recoverpoint와 마찬가지로 새로 생성될때나 마지막 복구작업이 끝난뒤에는 메모리 효율을 위해 폐기된다.

4. 성능평가

본 논문에서 제안한 홈기반 DSM 결함허용시스템의 성능평가를 위해 홈프로토콜기반의 변형된 CVM[1]을 사용하였고, IBM-SP2 슈퍼컴퓨터상에서 4노드에 4개의 응용프로그램을 대상으로 실험하였다.

응용 프로그램	프로세서 갯수	홈기반 DSM (msec)	결함허용홈기반 DSM (msec)
Water	1	10802	10900
	2	5826	5984
	4	3276	3482
Barnes	1	14132	14150
	2	7401	7301
	4	4025	4055
Tsp	1	9420	9435
	2	4752	4915
	4	2605	2754
SOR	1	2950	3017
	2	3274	3152
	4	2072	2702

표.1 실행시간 비교

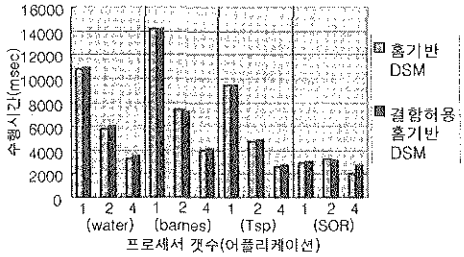

그림.3 수행시간 비교

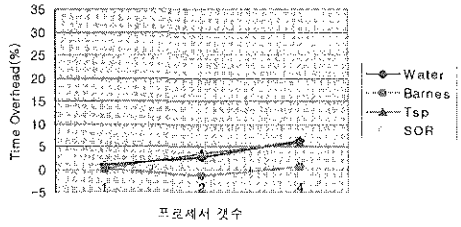

그림.4 Time Overhead 비교

5. 결론

본 논문에서는 홈의 특성에 따라 홈을 recoverpoint 서버로 하는 결함허용 시스템을 제안 하였다. 제안된 시스템은 시스템의 부하없이 단일 프로세서의 결함복구뿐만 아니라 다중 결함복구기능도 가능하다. 본 논문에서는 경쟁상황이 배제된 환경을 가정하였다. 만약 checkpoint를 저장하는 도중에 하나의 프로세서가 오류를 일으킨다면 전체 프로세서들이 recoverpoint로 복귀(rollback)해야 하고 이것으로 인해서 전체적인 성능감소가 있을 수 있다. 이문제는 저장장치의 분리와 RAID화를 통해 해결할 수 있다.

참고 문헌

1. P.Keleher," Distributed Shared Memory Using Lazy Release Consistency" ,PhD dissortation,Rice Univ.1994

2. L.Iftode, " Home-based Shared Virtual Memory" ,PhD thesis, Rice Univ.1998

3. R.Samanta, A.Bilas, L.Iftode, and J.P Singh," Home-based SVM Protocols for SMP clusters Design and Performance" In Proceedings of the 4th IEEE Symposium on High-Performance Computer Architecture, 1998

4. P.Keleher," Symmetry and Performance in consistency Protocols" In the 13rd Int. conference on supercomputing,999

5. M.Stumm, S.Zhou," Fault Tolerent Distributed Shared Memory Algorithms" ,IEEE,1990