

하이퍼큐브상에서 큐브이분할에 의한 효율적인 다중전송기법

홍의석, 임화경, 김성천
서강대학교 전자계산학과

Efficient Multicasting Scheme by Cube-Bipartitioning Method on Hypercube Multicomputer Systems

Euiseok Hong, Hwakyung Rim, Sungchun Kim
Dept. of Computer Science, Sogang University

요 약

대규모 분산 병렬 컴퓨터는 분산된 메모리를 갖는 프로세스 노드들간의 결합으로 이루어진 시스템으로써 이러한 분산 다중 시스템에서의 효율적인 집합적 통신의 구현은 시스템 성능에 커다란 영향을 미치는 중요한 요소이다. 이 중 다중전송(multicasting)은 다양한 응용프로그램에서 이용되는 중요한 통신 패턴이다. 본 논문에서는 워홀라우팅을 사용하는 하이퍼큐브상에서 목적노드를 균등하게 분배할 수 있는 차원을 통해 큐브를 분할하고 이에 따라 다중전송을 수행함으로써 다중전송단계를 줄일 수 있는 기법을 개발하고 이에 대한 성능을 실험을 통하여 비교, 분석하였다. 제안한 기법은 도달가능집합이 아닌 목적노드를 균등하게 분배할 수 있는 차원으로 큐브를 분할하여 다중전송을 수행함으로써 목적노드가 일부 하위큐브로 편중되지 않게 하여 평균통신단계를 약 15% 감소 시킬 수 있었다.

1. 서 론

최근 슈퍼 컴퓨터의 동행은 프로세서의 수가 증가함에 따라 이에 상응하는 성능의 향상을 얻기 위한 확장가능 병렬 컴퓨터(scalable parallel computers)의 설계로 진행되고 있다 이러한 시스템들을 대규모 병렬 컴퓨터(massively parallel computers : MPCs)라 하는데, 이러한 시스템들은 분산된 메모리를 갖는 프로세스 노드들간의 결합으로 이루어 진다. 이러한 대규모 병렬 컴퓨터에서의 효율적인 집합적 통신(collective communication)은 시스템 성능에 커다란 영향을 미친다[1, 2]. 이 중 특히 동일한 메시지를 하나의 전송시작 노드에서 임의의 수의 목적 노드로 전송하는 다중전송은 데이터의 복제, 신호 처리와 같은 다양한 응용프로그램에서 이용되는 중요한 통신 패턴이다[3]. 워홀라우팅을 사용하는 대규모 병렬 컴퓨터에서 노드간의 통신은 각각의 분리된 라우터에 의하여 제어되고, 각 라우터는 외부 채널을 통하여 인접한 노드의 라우터와 연결되며, 내부 채널을 통하여 지역 프로세서와 통신한다 이러한 내부채널은 입력과 출력을 위한 한쌍 이상의 채널로 구성되는데 이 입력력 채널의 수에 따라 단일 포트 모델(one-port model)과 다중 포트 모델(all-port model)로 구분된다. 한 쌍 이상의 채널을 갖는 구조, 즉 다중 포트 모델은 내부채널과 매핑될 수 있는 모든 외부 채널로 메시지를 동시에 전송하거나 받을 수 있으므로 단일 포트 모델의 병목현상을 감소시킬 수 있다[2, 4]. 효율적인 다중전송을 위하여 최근 연구되고 있는 방법은 통신 처리 과정에 포함된 노드들의 수를 균등하게 나눌 수 있도록 네트워크 위상을 재귀적으로 분할하는 방법으로, 단계별채널경쟁회피(stepwise contention free)와 통신 단계(communication step)의 수를 줄일 수 있는 방법이다. 본 논문에서는 워홀 라우팅을 사용하는 하이퍼큐브 위상에 대하여 재귀적 위상 분할(recursive topology partition) 방법과 비슷한 방식으로, 먼저 메시지를 전송해야 하는 목적노드들의 수를 균등하게 분할할 수 있는 차원을 결정하고, 분할된 큐브를 통하여 메시지를 전송함으로써 메시지간의 채널경쟁을 피하고, 전체적인

통신 단계를 줄일수 있는 효율적인 다중전송기법을 제안한다. 본 논문의 나머지 부분은 2장에서 앞서 언급한 기존의 다중전송 방법들과 이 방법들의 문제점을 살펴보고, 3장에서는 이러한 문제점들을 극복할 수 있는 새로운 전송기법에 대하여 기술한다 1장에서는 제안한 기법을 기존의 방법들과 시뮬레이션을 통하여 비교분석하고, 5장에서 결론을 맺고자 한다

2. 다중전송기법

2.1 위상 분할 기반의 다중전송

단일전송에 기반한 다중전송기법은 그 특성상 다중전송트리가 한쪽으로 치우게 된다. 가장 좋은 성능을 보이는 W-sort 기법[1]의 경우 메시지 수신노드으로써 수신노드를 선택하기 때문에 아직 메시지를 수신하지 못한 나머지 목적 노드들의 수가 혼잡노드가 전송에 사용할 개별적인 외부 채널의 수보다 많으면 통신단계가 증가하게 된다. 보다 일반적으로 목적노드의 수가 증가할수록 이러한 결과가 나타날 가능성이 크다 그러나 혼잡노드가 아닌 메시지 수신노드들은 더 이상 전송해야 할 목적 노드가 없기 때문에 이 노드들에서 아직 메시지를 받지 못한 목적노드들로 혼잡노드가 자신의 하위 목적노드로 메시지를 수신하는 단계에서 메시지를 전송한다면 통신 단계를 줄일 수 있게 된다.

2.2 일반적인 다중전송 알고리즘

Haiwan과 Özgüner[7]는 재귀적인 큐브분할을 기반으로 한 다중전송기법을 제안하였다. 다중포트모델을 사용하는 하이퍼큐브상에서 목적노드들은 전송시작노드에서 각각의 서로다른 포트를 통하여 도달할 수 있는 도달가능집합으로 정의, 분류하고, 각 집합에서 최초로 메시지를 수신할 노드를 결정한다. 이 노드들을 수신 노드(receiver node)라하며 다음 통신단계에서 나머지 목적노드들로 메시지를 전송해야 한다. 다음으로 아직 메시지를 수신하지 못한 노드들을 이미 수신한 노드들에 균등하게 배분하여 통신 단계를 줄일 수 있도록 큐브를 분할 한다. 이 과정을 모든 목적노드들이 메시지를 수신할 때까지

반복수행한다 이러한 도달가능집합은 서로 겹치는 노드없이 분리되어 있고, 공통된 전송노드에서 이들 집합으로의 메시지 전송은 채널 경쟁이 없음을 알 수 있다. n-차원 하이퍼큐브에서 완전한 다중전송을 수행하는데 최악의 경우 n 단계가 소요된다. 이 알고리즘의 효율성은 단계 2에서의 수신노드 선택방법과 단계 3의 큐브분할 방법에 의하여 좌우된다. [7]의 저자는 단계 2에서의 수신노드 선택 기법(Receiver Selection -Based Method : RSB) 과 단계 3에서의 큐브 분할 기법(Cube Partition-Based Method : CPB)을 제안하였다.

- 알고리즘 . 일반적인 다중전송 알고리즘**
- 1 전송시작 노드에 대한 도달가능집합에 따라 목적노드들을 구분
 - 2 각 목적노드의 그룹에서 하나의 노드를 수신노드로 선택
 - 3 각 하위큐브에서 분할된 큐브의 크기를 균등하게 할 수 있는 하나의 수신노드만이 선택되도록 큐브를 분할
 - 4 큐브내의 목적노드들을 분할정보에 따라 그룹화
 - 5 각 하위큐브의 수신노드로 메시지를 전송
 - 6 현재 전송시작노드에 할당된 목적노드들이 남아 있다면 1에서 5의 과정을 반복

2.3 큐브분할기반 기법

이 휴리스틱 방법은 주어진 수신노드의 집합으로부터 효율적인 분할을 선택하여 통신 단계를 줄이고자 하는 방법이다. 알고리즘의 2단계에서 수신노드의 선택은 임의로 결정한다. 각 분할의 단계에서 큐브는 하나의 자위를 따라 분할하며 이 과정을 개귀적으로 반복한다. 이 과정에서는 분할의 결과 가장 큰 하위 큐브의 크기를 최소화 하는데 주력한다 이 때 사용되는 것은 균등 분할 차원(even-split dimension)[7]이란 개념으로 주어진 k-큐브와 이 큐브가 가진 목적노드들에 대하여 목적노드의 수를 균등하게 분할할 수 있는 차원을 의미한다. 이 기법은 앞절에서 언급한 W-sort 기법의 다중포트의 장점을 이용할 수 없는 단점을 개선하였지만, 수신노드를 임의로 선택하기 때문에 항상 목적노드들을 균등하게 분배할 수 있는 수신노드를 선택할 수 없다는 단점이 있다.

2.4 수신노드선택기반 기법

CPB 기법의 단점을 보완하여, 수신노드의 선택시 목적노드를 균등하게 분배할 수 있는 가능성이 높은 노드를 수신노드로 선택하고자 하는 방법이 수신노드 선택 기법(RSB method)이다. CPB 기법과는 달리 알고리즘의 3단계에서 큐브를 분할하기 위한 차원의 선택은 최저차원부터 순차적으로 이루어진다 큐브 분할 차원을 최저 차원부터 선택하는 것은 내림차순 e-cube 경로배정 알고리즘을 사용하는 하이퍼큐브의 경우 k 도달가능집합의 크기 때문이다 RSB 기법은 앞서 언급한 미리 정의된 분할 차원의 순서를 적용할 때 알고리즘의 2 단계에서 가장 큰 크기를 갖는 하위큐브의 크기를 최소화하기 위한 수신노드를 선택하고자 하는 휴리스틱한 방법에 기인한다. 이는 각 단계마다 분할된 두 개의 하위큐브에 균등한 수신노드가 존재하도록 함으로써 가능하다.

3. 균등한 큐브분할을 이용한 다중전송

이 장에서는 앞절에서 살펴본 위상분할 기반의 다중전송 개념을 이용하여 새로운 목적노드 분배기법을 적용함으로써 보다 효율적인 다중전송을 제공하는 기법을 제안하고자 한다. 이 기법은 개념적으로 임의의 목적노드들을 최대한 균등하게 분할할 수 있는 차원을 통하여 큐브를 분할함으로써 통신단계를 줄이고자 하였다.

3.1 효율적인 다중전송을 위한 큐브분할

[정의] 이중노드(Dual Node) : n-차원 하이퍼큐브에 대하여 임의의 k-차원에 대하여 이분할된 두 개의 하위큐브 $(c_{n-1}c_{n-2}...0c_k-1...c_1c_0)$ 와 $(c_{n-1}c_{n-2}...1c_k-1...c_1c_0)$, $c_k \in \{0, 1\}$, $k = 0, 1, 2, \dots, n-1$ 에 대하여 각각의 하위큐브에 존재하는 노드 중 분할한 k 차원의 비트만 다르고 나머지 비트들은 같은 두 개의 노드, $(a_{n-1}a_{n-2}...0a_k-1...a_1a_0)$ 와 $(a_{n-1}a_{n-2}...1a_k-1...a_1a_0)$ 를 이중노드라 한다.

하이퍼큐브 구조의 대칭성에 의하여 큐브내의 임의의 수를 갖는

노드들은 대부분의 경우 그 수를 거의 비슷한 수로 나눌 수 있는 차원이 존재한다. 즉 한 큐브는 거의 동일한 수의 노드를 갖는 두 개의 큐브로 양분될 수 있다는 뜻이다. 특히 전체노드수에 대한 목적노드의 비율이 클수록 이러한 가능성이 높아진다 하이퍼큐브는 차원이 증가함에 따라 이전차원의 큐브 두 개가 결합하여 형성되는 구조이므로 어느 방향, 즉 차원으로 큐브를 분할하여도 같은 결합형태를 갖는 두 개의 큐브로 분할 된다

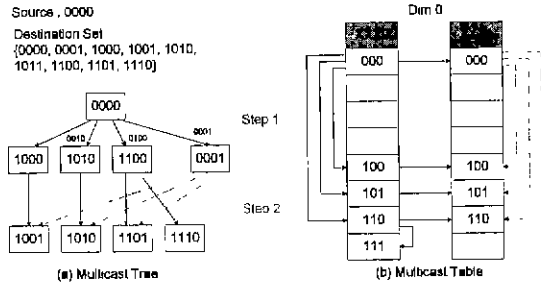
3.2 다중전송을 위한 경로배정

분할된 큐브내에서의 메시지전송 방법은 기본적으로 채널의 경쟁으로 인한 메시지전송시간 및 단계의 지연을 피하기 위하여 고정차원 순서 경로배정(dimension-ordered routing)[4]을 기반으로 한다. 기존 RSB 기법에서는 이러한 정해진 경로배정에 따라 도달할 수 있는 목적노드들을 분리하여 메시지 전송을 수행하였다. 이 경우 모든 도달가능집합에 목적노드가 존재할 경우에는 문제가 없으나 그렇지 않은 경우, 즉 목적노드가 존재하지 않은 도달가능집합이 존재하는 경우에는 이 차원을 통한 메시지 전송이 이루어지지 않게 된다. 따라서 임의의 노드에서 동시에 모든 채널을 통하여 메시지를 전송할 수 있는 다중포트 모델의 장점을 충분히 활용할 수 없게 되고 그리하여 다중전송단계가 증가하는 결과를 초래하게 된다. 이러한 문제점을 해결하기 위해서 경로배정 방법을 약간 수정하면 모든 포트를 사용하여 메시지를 전송할 수 있고 그리하여 다중전송단계를 줄일 수 있다. 즉 최초의 메시지 수신노드의 수가 이전 기법보다 증가하여 다음 전송단계에서 전송해야 하는 노드의 수가 감소하고 따라서 이전기법에 비하여 전송단계를 줄일 수 있게 된다 이러한 특징은 목적노드의 수가 많아질수록 더욱 유효하다. 이 기법에서의 경로배정은 기존의 차원 순서 경로배정을 약간 변형하여 도달가능집합에 목적노드가 없더라도 이 차원을 이용하여 메시지를 전송한다. 하이퍼큐브상에서 다중전송의 수행시 한 노드에서 동시에 메시지를 전송할 수 있는 차원은 한정되어 있으므로 어떠한 노드부터 먼저 메시지를 전송하는가가 다중전송단계에 영향을 미친다 제안하는 큐브 이분할기반의 다중전송 기법은 이중노드를 이용하여 다중전송의 효율성을 높이고자 하는 방법이므로, 가능한 이중노드를 수신노드로서 선택하는 것이 바람직하다.

3.3 큐브 이분할 기반의 다중전송

앞절에서의 단계를 거쳐 메시지 수신노드를 선택한 뒤 다중전송의 첫 번째 단계에서 수신노드로 메시지를 전송한다. 다중전송의 두 번째 단계에서는 현재 메시지를 가진 노드들로부터 두가지 경로를 통하여 다중전송을 수행한다. 즉 첫 번째 단계에서 메시지를 수신한 노드들로부터 각각의 하위큐브내에서 아직 메시지를 수신하지 못한 노드들의 경로배정 방법과 큐브분할 차원을 통한 큐브간 메시지 전송이 그것이다 이 두가지 독립적인 경로를 통하여 나머지 메시지를 전송받지 못한 노드들로 다중전송을 수행한다. 임의의 n-차원 하이퍼큐브는 그 위상의 특성상 동등한 두 개의 (n-1)-차원 하이퍼큐브로 분할될 수 있다. 따라서 n-차원 하이퍼큐브상에서의 다중전송은 2k 개의 (n-k)-차원 하이퍼큐브에 대한 다중전송을 병렬적으로 수행함으로써 이루어진다. 또한 각 큐브간의 이중노드간 메시지전송을 이용하여 다중전송 단계를 줄일 수 있다. n-차원 하이퍼큐브는 각 차원에 따라서 같은 위상을 갖는 (n-1)-차원 하위큐브로 분할 될 수 있다. 또한 균등분할 차원에 따라 큐브를 분할하면 그 목적노드의 수가 균등하게 분할되는 경우 m개의 목적노드를 갖는 n-차원 하이퍼큐브는 각각 m/2개의 목적노드를 갖는 (n-1)-차원 하이퍼큐브로 분할될 수 있다. m이 홀수일 경우에는 $\lceil m/2 \rceil$ 개 또는 $\lfloor m/2 \rfloor$ -1개의 목적노드를 갖는 두 개의 (n-1)-차원 하위큐브로 분할될 수 있다. 또한 각각의 하위큐브 역시 동등하게 (n-2)-차원 하위큐브로 분할되어 결국 전체 하이퍼큐브는 4개의 하위큐브로 분할될 수 있다. 따라서 개귀적으로 n-차원 하이퍼큐브는 각각 $\lceil m/2^k \rceil$ 또는 $\lfloor m/2^k \rfloor$ -1개의 목적노드를 갖는 (n-k)-차원 하위큐브로 분할될 수 있다. 이상과 같은 다중전송 기법은 다중전송단계를 줄일 수 있다는 장점뿐만 아니라 메시지 전송을 위한 다양하고 독립적인 경로를 제공함으로써 시스템내의 오류

발생시 효율적이고 적응적인 대체경로를 제공함으로써 보다 신뢰성 있는 다중전송을 제공할 수 있다.



[그림 1] CBPB 기법을 이용한 다중전송

4. 성능 평가 및 결과 분석

이 장에서는 시뮬레이션을 통하여 기존의 수신노드선택 기반(RSB) 다중전송기법과 제안하는 큐브 이분할 기반(CBPB) 다중전송 기법의 성능을 비교 분석한다. 수신노드선택 기반 다중전송기법은 기존에 제안되어 왔던 다중전송기법중 가장 좋은 성능을 보여주고 있는 기법이며, 시뮬레이션에 사용된 도구는 SLAM II이다. 시뮬레이션은 SunOS Release 4.1.3 환경의 워크스테이션에서 수행하였으며, 대상이 된 상호연결망은 4, 5, 6, 7 차원의 하이퍼큐브구조이다.

4.1 성능평가요소 및 기본가정

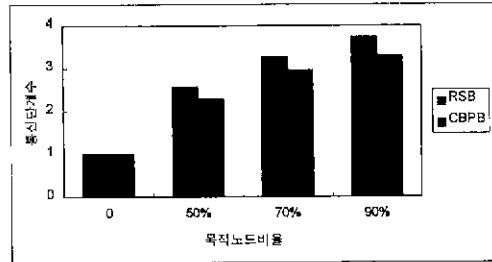
제안한 기법의 가장 큰 목적은 다중전송을 완료하는데 필요한 통신단계를 줄이는데 있다. 따라서 기존 RSB 다중전송기법과 CBPB 다중전송기법간에 평균 다중전송 통신단계를 비교한다. 임의의 주어진 목적노드들에 대하여 다중전송을 수행하는데 필요한 통신 단계의 평균값을 비교함으로써 기법간의 효율성을 검증한다. 실험은 4차원, 5차원, 6차원, 7차원 하이퍼큐브에서 임의로 선택된 다중전송 메시지를 수신해야 하는 목적노드의 수를 전체 노드수의 50%, 70%, 90%로 변화시키면서 다중전송수행이 완료되는데 필요한 통신단계수의 평균을 비교하였다. m개의 목적노드를 포함하는 n-차원 하이퍼큐브의 경우 다중전송에 필요한 이론적 최소값은 $\lceil \log_{m+1}(m+1) \rceil$ 이다. 각각에 대한 적용치는 다음과 같다.

	목적노드수 (최소 통신 단계)			
	4차원	5차원	6차원	7차원
50%	8 (2)	16 (2)	32 (2)	64 (2)
70%	12 (2)	22 (2)	45 (2)	100 (3)
90%	14 (2)	29 (2)	58 (3)	116 (3)

4.2 시뮬레이션 결과 분석

이 장에서는 앞에서 언급한 성능평가 요소들을 바탕으로 수행한 시뮬레이션의 결과를 분석한다. 전체적인 평균 통신 단계가 15%정도 감소함을 보이고 있다. 목적노드가 각 차원을 통하여 도달가능한 집합에 균등하게 분포할 때는 제안한 CBPB기법이 RSB기법과 유사하거나 보다 우수한 성능을 나타내었으며, 목적노드가 일부 차원에 편중되었을 경우에는 첫 번째 단계에서의 수신노드 선택의 방법과 그 수 때문에 CBPB기법이 보다 우수한 성능을 보이고 있다. 또한 제안하고 있는 CBPB기법은 목적노드의 수가 증가하더라도 통신단계에 큰 변화가 없음을 알 수 있다. 이 기법은 사용 가능 차원을 통하여 수신노드를 결정하므로 기존 기법에서와 같은 수신노드 수의 감소로 인한 통신단계 증가를 막을 수 있기 때문이다. 다음단계의 메시지전송을 책임지는 수신노드의 수가 감소한다는 것은 각 수신노드에 분배될 목적노드들의 수가 증가하는 결과를 초래하므로 다중전송통신단계가 증가할 가능성이 커진다. 따라서 많은 수의 수신노드를 가질수록 통신단계를 줄일 수 있다. n-차원 하이퍼큐브는 한 노드에 n개의 링크를 가지고 있고 이를 통하여 메시지를 전송하고 수신하는데, RSB 기법은 수신노드선택 방법의 특성에 의하여 도달가능집합에 목적노드들이 존재하지 않을 때는 이 차원에 대한 수신노드를 결정할 수 없기

때문에 수신노드의 수가 CBPB 기법에 비하여 줄어들게 된다. CBPB 기법은 변형된 경로배정방법을 사용하므로 사용가능한 차원을 통하여 선택할 수 있는 수신노드의 수를 극대화하여 다중전송단계를 감소시킬 수 있다.



[그림 2] 6차원 하이퍼큐브상에서의 평균통신단계

5. 결론

기존 다중전송기법은 다중전송트리 구성시 목적노드의 불균등한 분배에 의하여 다중전송단계가 증가하는 단점이 있었다. 본 논문에서는 이러한 단점을 해결하기 위하여 차원을 통해 큐브를 분할하고 이에 따라 다중전송을 수행함으로써 다중전송단계를 줄일 수 있는 기법을 제안하였다. 제안한 기법은 기존 기법의 통신단계를 약 15% 정도 감소시키므로써 보다 효율적인 다중전송을 제공한다. 기존의 기법은 도달가능집합을 기준으로 목적노드를 분할하기 때문에 도달가능집합의 크기 불균형으로 인하여 다중전송단계가 증가하였는데, 제안한 기법에서는 목적노드를 균등하게 분할할 수 있는 차원으로 큐브를 분할하여 다중전송을 수행함으로써 목적노드가 일부 하위큐브로 편중되지 않게 하였으며, 본문에서 고려하지 않았지만 오류발생시에도 큐브간의 통신을 지원하여 메시지 전송에 다양한 경로를 제공함으로써 효율적으로 대처할 수 있다. 제안한 기법은 하나의 전송단계내에 전송시작노드로부터 여러개의 수신노드로 메시지를 동시에 전송한다. 따라서 단계별 채널경쟁회피를 위하여 이전 단계에서의 모든 메시지들이 수신노드로 도착하고 난 후에야 시작되기 때문에 이러한 메시지 전송의 동기화를 고려한다면 보다 효율적인 다중전송기법을 제공할 수 있을 것이다.

참고 문헌

[1] D. F. Robinson, D Judd, P. K. McKinley, and B. H. C. Cheng, "Efficient Collective Data Distribution in All-Port Wormhole-Routed Hypercubes," Proc. Supercomputing '93, Nov. pp. 792-801, 1993.
 [2] P. K. McKinley and C. Treffetz, "Efficient Broadcast in All-Port Wormhole-Routed Hypercubes," Proc. of Int'l Conference Parallel Processing, vol. II, Aug. pp. 288-291, 1993.
 [3] P. K. McKinley, T. Yih-jia, and D. F. Robinson, "Collective Communication in Wormhole-Routed Massively Parallel Computers," IEEE Computer, vol. 28, no. 12, pp.39-50, Dec. 1995.
 [4] T. Yih-jia and P. K. McKinley, "A Broadcast Algorithm for All-Port Wormhole-Routed Torus Networks," IEEE Trans. on Parallel and Distributed Systems, vol. 7, no. 8, Aug. 1996.
 [5] L. Ni and P. K. McKinley, "A Survey of Wormhole Routing Techniques in Direct Networks," IEEE Computer, no. 2, pp. 62-76, 1993.
 [6] P. K. McKinley, A. Esfahanian, and L. Ni, "Unicast-Based Multicast Communication in Wormhole-Routed Networks," IEEE Trans on Parallel and Distributed Systems, vol. 5, no. 12, pp. 1252-1265. 1994.
 [7] V. Halwan, and F. Ozgüner, "Efficient Multicast Algorithms in All-Port Wormhole-Routed Hypercubes," Proc. of Int'l Conference Parallel Processing, Aug. pp. 84-91, 1997.