PCIe 트랜잭션을 위한 주소 변환 성능 평가

차광호, 정기문
한국과학기술정보연구원
khocha@kisti.re.kr, kmjeong@kisti.re.kr

Evaluation of Address Translation Performance in PCIe Transactions

Kwangho CHA, Gi-mun JEONG

Korea Institute of Science and Technology Information

요 약

AI 기술의 급속한 확산은 고성능 시스템에 대한 수요를 가속화하고 있으며, 이는 시스템 직접도 향상과 같은 확장 기술의 발전에도 중요한 영향을 미치고 있다. 본 연구에서는 PCIe 버스를 확장하는 과정에서 필수적으로 요구되는 브리지의 주소 변환 기법을 심층적으로 고찰하고, 그 성능을 분석하였다. 그 결과, 변환 기법 간의 구조적 차이에도 불구하고 변환 기법과 성능 저하도 간의 유의미한 연관 관계를 찾을 수 없었다. 이는 예상과 달리 브리지에서의 주소 변환 기법이 통신 성능에 미치는 영향이 매우 제한적이라는 사실을 보여준다. 즉, 브리지의 주소 변환 방식을 결정할 때, 성능 최적화보다는 시스템 구성의 편의성이나 메모리 관리의 효율성과 같은 다른 설계 요소들이 더 중요한 고려사항으로 간주 될 수 있음을 시사한다.

1. 서론

시스템의 활용 분야 확대 또는 가용성 증대와 같은 목적을 위해서 시스템 버스의 확장을 고려하게 된다. 특히, 시스템 버스로 오랜 기간 사용되어 온 PCIe 버스를 확장하기 위해서는 브리지의 사용이 필수적이며 이 브리지는 다시 TB(Transparent Bridge)와 NTB(Non-Transparent Bridge)로 구분된다. 일반적으로 TB는 PCIe 버스에 연결되는 장치의 수가 증가될 때 사용된다. 즉 단일 호스트가 여러 장치들을 식별하고 PCIe 트랜잭션이 대상 장치로 라우팅될 수 있게 한다. 반면 NTB는 단일 호스트가 아닌 복수의 호스트가 PCIe 버스에 연결되는 상황을 가정하므로 TB의 원래 기능에 호스트(OS)간의 충돌을 방지하는 기능을 추가로 제공하고 있다[1].

이와 같은 NTB의 추가 기능으로 주소 변환 기능을 들 수 있다. 현재 트랜잭션이 가지고 있는 주소 정보를 상대측 서브 시스템에서 라우팅 가능한 주소로 바꾸어 주는 역할을 수행하며 직접 주소 변환과 룩업(Lookup) 테이블을 사용하는 방식 등이 제공된다.

본 연구에서는 NTB 기반 PCIe 트랜잭션의 주소 변환 기법을 기능과 성능면에서 분석하고, 이를 토대로 NTB 브리지에 대한 운영 정책 수립에 필요한 기초적 근거를 마련하고자 한다.

2. 시스템 확장용 브리지

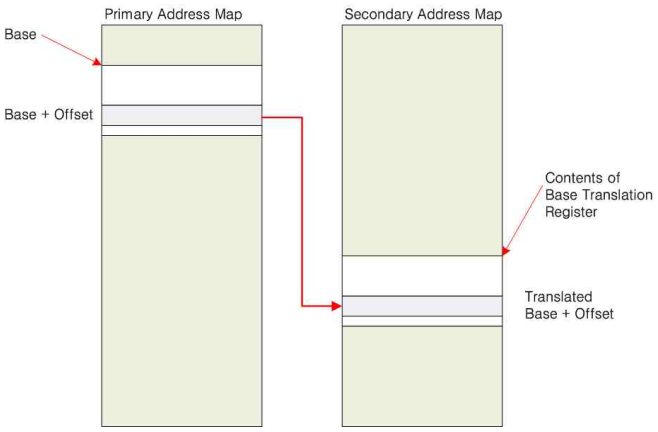
투명 브리지는 기본적으로 주소 변환없이 트랜잭션을 통과시키게 된다. 이를 위해 브리지는 각 메모리와 I/O 공간에 대한 베이스(base)와 리밋(limit) 레지스터의 설정을 통해 접근 구간을 정하고 그 안의 공간을 접근하는 유효한 트랜잭션인 경우, 이를 전달하게 된다.

반면 불투명 브리지는 단순 전달이 아니라 주소 변환 과정을 포함하고 있다. 로컬 프로세서는 BAR(Base Address Register)를 미리 설정하여 NTB로 전달된 트랜잭션의 주소 변환에 사용한다. 이렇게 트랜잭션은 BAR를 통해 변환된 주소로 매핑되어 반대편 서브시스템으로 전달된다.

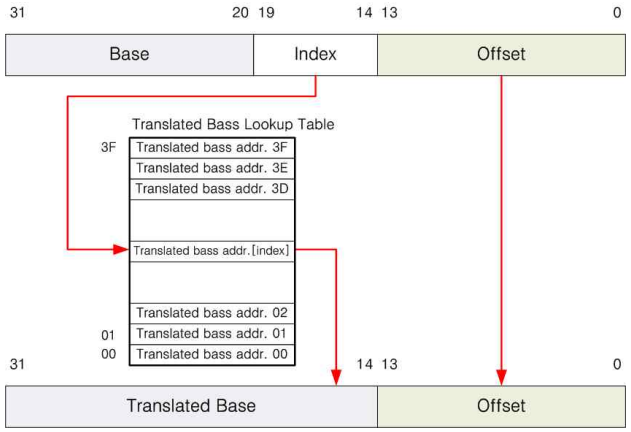
3. PCIe 트랜잭션을 위한 주소 변환 기법

3.1. 직접 주소 변환

직접 주소 변환은 NTB를 사용하는 주소 변환 중 가장 간단하고 직관적인 변환 방법으로 그림 1과 같은 구조이다. 오프셋(offset)을 유지하되 베이스(base)



(그림 1) 직접 주소 변환 예[2]



(그림 2) 룩업 테이블 기반 주소 변환 예[2]

주소를 변경하여 반대편의 유효한 주소를 구하게 된다. 즉, BAR를 통해 들어오는 트랜잭션의 경우 해당 BAR에서의 오프셋에 미리 설정한 베이스 주소를 더해서 유효 주소를 구하게 된다. 이때 미리 정한 베이스 주소는 BAR Base Translation 레지스터를 통해 확인 가능하며 각 BAR마다 독립적으로 설정된다.

이처럼 직접 주소 변환은 단순한 구조로 인한 신속한 처리가 장점이나 주소 공간의 매핑에서는 유연성이 떨어진다는 단점이 지적되고 있다.

3.2. 룩업 테이블 기반 주소 변환

Lookup 테이블 기반 주소 변환 방식은 그림 2와 같이 미리 설정한 단일 베이스 주소 대신 룩업 테이블을 사용하는 방법이다. 이 방식은 단순한 덧셈 방식보다 유연한 매핑이 가능하고 필요한 경우 주소 영역을 조각내어 다른 공간에 연결할 수도 있다. 또한, 주소 버스에서 어떤 비트 구간을 테이블의 인덱스로 사용할지를 프로그래밍할 수 있기 때문에, 변환 단위(즉, 조리개의 크기) 역시 자유롭게 조정이 가능하다.

룩업 테이블 기반 주소 변환은 주소 공간을 매핑할 때 좀 더 유연한 설정이 가능하다는 장점이 있으나 테이블 유지를 위한 하드웨어의 추가 설정 및 자원 소모, 테이블 접근으로 인한 시간 지연 등이 부담으로 거론된다.

3.3. 실험 환경 구성

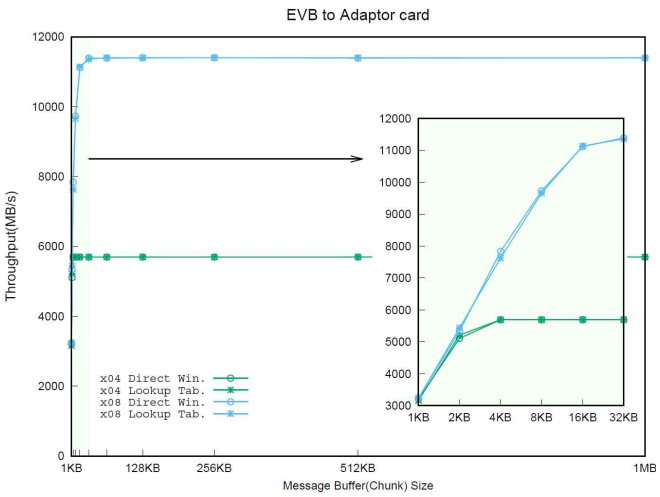
위에 설명한 주소 변환 방법이 통신 성능에 어떤 영향을 주었는지 확인하는 실험을 진행하였다. 표 1은 실험 환경의 구성을 보여준다. NTB 통신을 수행하기 위한 테스트 베드로는 Microchip사의 PCIe 4.0 평가 보드를 사용하였고 소프트웨어는 오프소스를 활용하였다. 즉, NTB 통신에 필요한 switchtec 하드웨어 드라이버와 ntb_perf.ko와 같은 테스트용 커널 모듈은 리눅스 배포판[3]에 포함된 소스 코드를 사용하여 준비하였다.

4. 성능 평가 및 분석

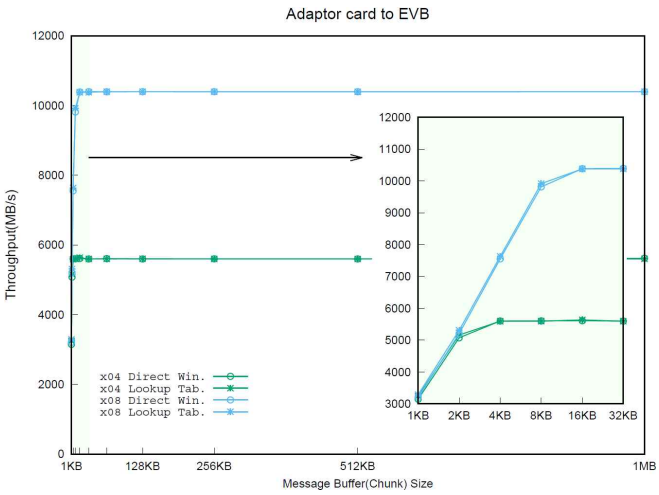
그림 3과 4에 실험 결과를 정리하였다. 실험은 크게 2가지로 구분되어 진행되었다. PCIe 버스는 방향성을 가지고 있기 때문에 각 방향에 대한 성능 측정을 진행하였는데 PCIe 평가 보드를 가지고 있는 PC에서 상대방 PC로 전송하는 그림 3의 경우와 어댑터 카드만 설치된 PC에서 EVB가 설치된 PC쪽으로 데이터를 전송하는 그림 4의 실험으로 구분된다. 어댑터 카드쪽 PC에서 평가 보드측 PC로 데이터를 전송할 때 4배속(x4)의 경우에는 이론 성능 대비 약 69%의 성능을 8배속(x8)의 경우에는 약 64%의 성능을 보인 반면, 평가 보드쪽 PC에서 어댑터 카드쪽 PC로 데이터를 보내는 경우에는 4배속(x4)과 8배속(x8) 모두 이론 성능 대비 약 70%의 성능을 보였다. 이는 평가 보드에 위치하는 PCIe 스위치 칩을 직접 접근할 수 있는 경우의 성능이 추가적인 장치들을 거쳐서 접근하는 것 보다 우수함을 보여 주는 결과이다.

<표 1> 실험 환경 세부 구성

서버 노드	CPU	Intel Xeon w5-3423
	운영체제	Ubuntu 22.04.4 LTS linux kernel 5.15.0-43-generic
	드라이버	switchtec driver(5.15.0-43-generic)
PCIe 4.0 EVB		Microchip PM42100-KIT



(그림 3) PCIe 4.0의 NTB 통신 성능;
EVB측 PC에서 데이터 전송



(그림 4) PCIe 4.0의 NTB 통신 성능;
어댑터 카드측 PC에서 데이터 전송

이와 같은 NTB를 사용하는 통신 성능은 앞서 언급한 ntb_perf.ko를 사용하여 측정하였는데 이 성능 측정용 커널 모듈을 실행하기 전에 NTB 커널 드라이버를 적재하면서 NTB 주소 변환 방법을 변경하도록 하였다. 그림 3과 4에 나타난 결과에서 보듯이 두 주소 변환 기법으로 인한 성능 차이는 크게 나타나지 않았다. 즉, 테이블을 유지하고 접근하는 과정의 오버헤드가 예상과 달리 크지 않음을 보여주는 결과이다. 특히 통신 성능이 거의 포화상태에 들어선 메시지 크기가 256 KB 이상인 구간에서는 성능 차이를 확인하기 어려웠다. 메시지의 크기가 작은 구간에서는 최대 약 3% 정도의 성능 차이를 보였지만 특정 변환 방식이 항상 우위의 성능을 보이는 것이 아니라서 주소 변환 방식과 통신 성능 간의 직접적인 연관성은 미비하다는 것을 확인할 수 있었다.

5. 결론

AI 기술의 확대 보급에 따라 고성능 시스템에 대한 수요가 증가하고 있으며 이는 시스템 직접도를 높이기 위한 시스템 확장 기술에도 영향을 미치고 있다. 본 연구에서는 PCIe 버스를 확장할 때 필요로 하는 브리지에서의 주소 변환 기술을 살펴보고 그 성능을 함께 평가하였다. PCIe 서브 시스템을 구성하는데 사용되는 상용 PCIe 스위치가 제공하는 주소 변환 기법들의 성능을 비교한 결과 변환 기법 자체의 오버헤드에는 큰 차이가 없는 것이 확인되었다. 즉, 변환 기법의 성능보다는 구현하고자 하는 시스템에서의 PCIe 트랜잭션의 특징이나 용도에 부합하는 성격의 변환 기법을 사용하여야 할 것으로 예상되며 향후 PCIe 서브 시스템의 운영 정책과 주소 변환 기법 간의 연관 관계를 분석할 계획이다.

ACKNOWLEDGMENTS

이 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다. (과제번호: (KISTI)K25L1M2C2)

참고문헌

[1] PCI-SIG, "PCI Express® Base Specification Revision 6.0.1," 29 August 2022.

[2] Jack Regula, "Using Non-transparent Bridging in PCI Express Systems," Technical Report, PLX Technology, Inc., 2004

[3] The Linux Kernel Archives, https://www.kernel. org