PCIe 서브시스템 모니터링 도구 구현

차광호
한국과학기술정보연구원 슈퍼컴퓨팅기술개발센터
khocha@kisti.re.kr

Implementation of PCIe Subsystem Monitoring Tools

Kwangho CHA
Center for Supercomputing Technology Development,
Korea Institute of Science and Technology Information

요 약

다양한 AI 서비스들이 보급되면서 고성능 계산 자원에 대한 요구가 증대하고 있으며 이를 위해 여러 이기종 계산 자원들을 수용할 수 있는 시스템 아키텍처 기술들이 발표되고 있다. 특히 계산 자원들의 수용을 위한 시스템 버스 기술에 대한 관심이 높아지고 있다. 본 연구는 여러 새로운 시스템 버스들 속에서도 표준 시스템 버스라 할 수 있는 PCIe 버스를 대상으로 하고 있는데 PCIe 서브시스템의 상태를 실시간으로 모니터링하기 위한 도구를 제안하고 그 구현 결과를 소개한다.

1. 서론

PCIe 버스는 초기에는 PC나 서버에서 내부 디바이스 간의 연결을 지원하기 위한 IO 버스로 사용되었다. 그러나 다양한 이기종 계산 자원 수용의 및 시스템 버스와 인터커넥션 네트워크의 융합과 같은 새로운 요구 사항들이 생겨나면서 PCIe 버스 또한 발전된 추가 기능들을 제공하게 되었는데 장치간 연결뿐만 아니라 시스템 간 연결의 지원을 그 예로 들 수 있다. 본 연구에서는 이렇게 활용 범위를 넓히고 있는 PCIe 서브시스템의 상태를 모니터링하고 그 정보를 기록하기 위한 도구을 제안하고 그 구현 결과를 공유하고자 한다.

2. PCIe 버스와 PCIe 스위치

1992년에 발표된 PCI와 1999년에 발표된 PCI-X와의 호환성을 제공하는 후속 버스로 PCIe(Peripheral Component Interconnect Express) 1.0이 2002년에 발표되었다. 메모리, IO 공간 그리고 설정 공간(Configuration Space)에 대해 소프트웨어적인 접근이 가능하다는 특징이 있으며 2025년을 목표로 PCIe 7.0 규격이 정의되고 있다[1].

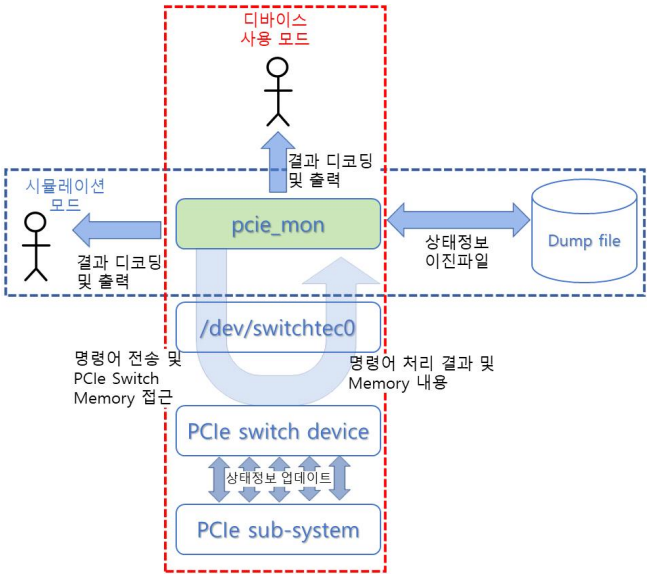
PCIe 버스의 주요 특징으로 양방향 점대점 시리얼 통신 수행, 분기 확대를 통한 확장성 제공 및 패킷 방식의 전송 프로토콜 사용을 생각해 볼 수 있다[1]. 또한 PCIe 스위치와 같은 특수 장치를 사용하면 PCIe 서브시스템의 적용 범위를 넓히고 안전성을 개선하며 PCIe 트래픽에 대해 지능화된 세부 설정을 적용할 수 있다. 이처럼 PCIe 스위치는 PCIe 전기 신호의 전송 거리를 연장하기 위한 목적으로 개발된 PCIe Bridge에 세부 설정 및 운영이 가능하도록 하는 기능이 추가된 지능화된 컴포넌트로 요사이에는 SoC 형태의 제품들이 사용되고 있다. 즉, 독자적인 임베디드 시스템이 운영에 필요한 설정값을 사용해서 PCIe 서브시스템을 제어하게 된다[2,3].

3. PCIe 서브시스템에 대한 모니터링 설계 및 구현

본 연구에서는 PCIe 스위치가 관리하는 PCIe 서브시스템의 상태 정보를 수집하고 관리하는 모니터링 도구를 설계하고 구현하고자 한다.

일반적으로 리눅스와 같은 운영체제는 /sys나 /proc 같은 시스템 파일들과 장치 드라이버의 정보를 조합하여 PCIe 장치나 버스에 대한 상태 정보를 사용자에게 제공하는 기능을 가지고 있다. 이때 관리자는 PCIe 장치나 버스에 대한 전체 정보를 바탕으로 모니터링 대상인 특정 장치의 정보를 찾아가는 과정을 거치게 되며 원시 정보 형태로 제공되기 때문에 사용자의 숙련도에 따라 상태 정보의 해석과 이해의 정도가 달라질 수 있다.

한편 앞서 언급한 바와 같이 최근의 PCIe 스위치는 자체적인 프로세서와 운영체제를 가지는 독립된



(그림 1) PCIe 모니터링 도구 개념도

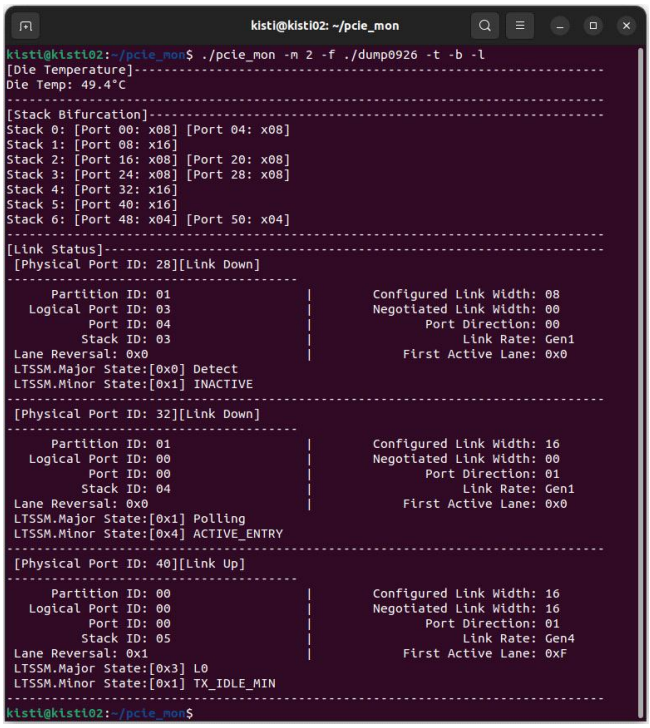
<표 1> 모니터링 대상 정보

분기 정보	PCIe 버스의 분기(Bifuraction) 설정 정보를 확인
링크 상태 정보	각 링크별 상태 (LTSSM상의 세부 상태 포함)
보드 온도 정보	PCIe 스위치가 포함된 보드의 현재 온도
PWM 정보	냉각팬과 관련된 PWM 정보

임베디드 시스템으로 볼 수 있으며 PCIe 스위치에서 동작하는 운영체제는 자체 메모리 공간을 사용해서 자신에게 부여된 PCIe 서브시스템의 관리를 수행하게 된다. 이때 각 버스와 분기와 상태 정보 또한 PCIe 스위치의 메모리 영역에 저장된다. 즉, 이러한 PCIe 스위치의 구조적인 특정을 활용한다면 PCIe 스위치가 관리하는 PCIe 서브시스템별 상태 정보의 획득과 관리를 좀 더 수월하게 진행할 수 있게 된다.

그림 1은 우리가 구현하고자 하는 PCIe 서브시스템 모니터링 도구의 개념도이다. 우선 PCIe 모니터링 도구는 PCIe 스위치를 의미하는 디바이스 파일을 통해 PCIe 스위치의 메모리 공간을 접근하여 정보를 획득하고 해석하여 사용자에게 제공할 수 있도록 하였다. 표 1에 현재까지 구현되어 확인 가능한 상태 정보를 표시하였다. 버스에 대해서는 설정된 분기(Bifurcation) 정보를 바탕으로 실제 어느 만큼의 레인을 할당받아 사용하고 있는지 그리고 Link Training and Status State Machine(LTSSM) 상에서 어떤 상태에 머물러 있는지를 확인할 수 있다.

상태 정보 모니터링 기능 외에 추가로 구현된



(그림 2) PCIe 모니터링 도구 실행 예

기능으로 상태 정보에 대한 덤프 기능이다. 즉 PCIe 스위치의 메모리에 존재하는 현재의 상태 정보를 그대로 파일에 기록한 뒤 이 파일을 읽어서 유용한 세부 정보로 해석할 수 있게 하였다. 또한 덤프 파일의 주기적인 저장이 가능하므로 PCIe 서브시스템의 특정 기간 동안의 동작 변화를 PCIe 스위치와 서브시스템이 존재하지 않더라고 분석할 수 있도록 하였다.

4. PCIe 모니터링 시스템 테스트

그림 2는 마이크로칩社의 PCIe 4.0 평가보드(PM42100-KIT)[4]를 대상으로 PCIe 모니터링 도구를 수행한 결과이다. 이 평가보드는 PCIe 4.0 100레인을 지원할 수 있으며 모두 7개의 PCIe 스택(스테이션)을 보유하고 있다. 개발된 PCIe 모니터링 도구는 텍스트 기반의 콘솔 프로그램으로 제작되어 있다. 그림처럼 PCIe 스위치 평가보드의 현재 온도를 시작으로 평가보드에 연결된 PCIe 버스들의 분기 정보 및 링크 상태 정보들을 표시하는데 문제가 없음을 알 수 있다.

5. 결론 및 향후 계획

PCIe 버스를 활용한 다양한 계산자원의 수용이 가능한 현 시점에서 PCIe 서브시스템의 상태를 모니터링하고 해석하는 것은 시스템의 안정성 향상을 위해서 중요한 기능이다. 본 연구에서는 PCIe 스위치가 활용하는 메모리 영역에 존재하는 PCIe 서브시스템의

상태 정보를 획득·분석하여 사용자에게 제공하는 PCIe 모니터링 시스템을 구현하고 그 실행 결과를 소개하였으며 PCIe 버스의 분기 설정 및 링크의 상태 정보를 문제없이 제공함을 확인하였다.

PCIe와 같은 시스템 버스 계통에 문제가 발생할 때는 시스템 이상 및 비정상 종료를 수반하는 경우가 많아서 문제의 원인 분석을 위한 관련 데이터의 확보가 쉽지 않다. 이에 본 모니터링 도구의 파일 저장 기능을 보완하여 획득 정보의 시간대별 저장 기능의 구현을 계획하고 있다. 즉, PCIe 상태에 대한 시계열 데이터를 바탕으로 PCIe 서브시스템 동작 패턴의 분석 가능성을 확인할 계획이다.

ACKNOWLEDGMENTS

이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: (KISTI)K24L2M1C6, (NTIS)2710018524)

참고문헌

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

[2] 차광호, 구경모, 오광진, “PCIe 기반 시스템 확장 기술 개발,” 978-89-294-1547-1 (95560), 2023,11.

[3] Jack Regula, “White paper: Using Non-transparent Bridging in PCI Express Systems,” Technical Report, 2004.

[4] PM42100-KIT, Switchtec™ Gen 4 PCIe® Switch Evaluation Kit, https://ww1.microchip.com/downloads/en/DeviceDoc/00003159.pdf