



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년09월27일
(11) 등록번호 10-1782342
(24) 등록일자 2017년09월21일

- (51) 국제특허분류(Int. Cl.)
G06F 9/44 (2006.01) **G06F 12/10** (2016.01)
G06F 15/17 (2006.01)
- (21) 출원번호 10-2012-7018504
- (22) 출원일자(국제) 2010년11월23일
 심사청구일자 2015년10월26일
- (85) 번역문제출일자 2012년07월16일
- (65) 공개번호 10-2012-0115519
- (43) 공개일자 2012년10월18일
- (86) 국제출원번호 PCT/US2010/057871
- (87) 국제공개번호 WO 2011/084257
 국제공개일자 2011년07월14일
- (30) 우선권주장
 12/640,272 2009년12월17일 미국(US)
- (56) 선행기술조사문현
 US20070050764 A1*
 US20090025007 A1*
 US20090094603 A1*
 US20080263553 A1*

*는 심사관에 의하여 인용된 문현

- (73) 특허권자
마이크로소프트 테크놀로지 라이센싱, 엘엘씨
 미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
- (72) 발명자
오션스 제이콥
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
- 그린 더스틴 엘**
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
- (74) 대리인
제일특허법인

전체 청구항 수 : 총 20 항

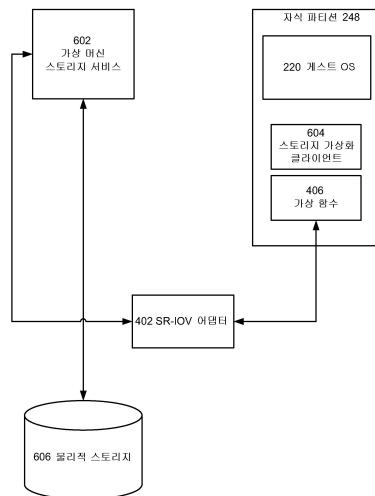
심사관 : 김경완

(54) 발명의 명칭 **가상 스토리지 타겟 오프로드 기법**

(57) 요 약

가상 머신 스토리지 서비스는 고유 네트워크 식별자를 사용할 수 있고, SR-IOV 규격 장치는 가상 머신과 가상 머신 스토리지 서비스 사이에 I/O를 이송시키는 데 사용될 수 있다. 가상 머신 스토리지 서비스는 자식 파티션에 오프로드되거나, 고유 네트워크 식별자와 함께 다른 물리적 머신으로 이전될 수 있다.

대 표 도



명세서

청구범위

청구항 1

이전가능한 스토리지 서비스(migratable storage service)를 실현하기 위한 회로 -상기 이전가능한 스토리지 서비스는 자식 파티션에 대한 가상 하드 디스크로의 입/출력 요청들을 관리하도록 구성되며, 상기 이전가능한 스토리지 서비스는 네트워크에 대한 고유 네트워크 식별자가 할당됨- 와,

상기 가상 하드 디스크를 이전할 필요 없이 상기 이전가능한 스토리지 서비스를 적어도 다른 파티션으로 이전하기 위한 회로와,

상기 이전가능한 스토리지 서비스를 네트워크 내의 스토리지 타겟으로서 구성하기 위한 회로를 포함하는 이전가능한 스토리지 서비스를 위한 시스템.

청구항 2

제 1 항에 있어서,

상기 이전가능한 스토리지 서비스를 적어도 다른 파티션으로 이전하기 위한 회로는

상기 이전가능한 스토리지 서비스를 원격 컴퓨터 시스템 상의 파티션에 이전하기 위한 회로를 더 포함하는 이전가능한 스토리지 서비스를 위한 시스템.

청구항 3

제 1 항에 있어서,

상기 자식 파티션에 대한 입/출력 요청들에 연관된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환시키도록 입/출력 메모리 관리 유닛을 구성하기 위한 회로를 더 포함하는

이전가능한 스토리지 서비스를 위한 시스템.

청구항 4

제 1 항에 있어서,

상기 자식 파티션으로부터 입/출력 작업 요청을 수신하기 위한 회로 -상기 자식 파티션은 상기 네트워크에 대한 제 2 고유 네트워크 식별자를 포함하는 가상 기능에 부착됨(attached)- 를 더 포함하는

이전가능한 스토리지 서비스를 위한 시스템.

청구항 5

제 1 항에 있어서,

제 1 파티션에서 상기 이전가능한 스토리지 서비스를 실행하고, 제 2 파티션에서 가상 머신들을 관리하도록 구성된 관리 서비스를 실행하기 위한 회로 -상기 자식 파티션은 제 3 파티션임- 를 더 포함하는

이전가능한 스토리지 서비스를 위한 시스템.

청구항 6

제 1 항에 있어서,

상기 이전가능한 스토리지 서비스를 상기 고유 네트워크 식별자를 포함하는 네트워크 어댑터의 가상 기능과 연계시키고, 상기 자식 파티션을 상기 네트워크 어댑터의 제 2 가상 기능과 부착하기 위한 회로를 더 포함하는 이전가능한 스토리지 서비스를 위한 시스템.

청구항 7

제 1 항에 있어서,

상기 자식 파티션으로부터의 입/출력 작업 요청의 수신에 대한 응답으로 논리 프로세서에 통지를 전송하고, 상기 논리 프로세서가 상기 이전가능한 스토리지 서비스를 실행하고 있는지를 결정하기 위한 회로를 더 포함하는 이전가능한 스토리지 서비스를 위한 시스템.

청구항 8

제 1 항에 있어서,

입/출력 트래픽이 네트워크 어댑터를 통해 상기 고유 네트워크 식별자와 적어도 하나의 다른 고유 네트워크 식별자 사이에서 이동할 때 상기 입/출력 트래픽이 보안 대책을 따르는지를 결정하기 위한 회로를 더 포함하는 이전가능한 스토리지 서비스를 위한 시스템.

청구항 9

컴퓨터 시스템에 의해 실행되는, 이전가능한 스토리지 서비스를 위한 컴퓨터 방법으로서,

네트워크 어댑터에 대한 제 1 고유 네트워크 식별자를, 파티션 내에서 인스턴스화되고(instantiated) 자식 파티션 및 연관된 가상 하드 드라이브 디스크에 대한 입/출력 요청들을 관리하도록 구성된 스토리지 서비스에 부착하는 단계와,

상기 네트워크 어댑터에 의해 실현되는 가상 기능을 상기 자식 파티션에 부착하는 단계 -상기 가상 기능은 제 2 고유 네트워크 식별자를 포함하며, 상기 스토리지 서비스는 네트워크 내에 스토리지 타겟으로서 사용되도록 구성됨- 를 포함하는

이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 10

제 9 항에 있어서,

상기 제 1 고유 네트워크 식별자를 포함하도록 제 2 가상 기능을 구성하기 위한 요청을 제 2 네트워크 어댑터를 포함하는 원격 컴퓨터 시스템에 전송하는 단계를 더 포함하는

이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 11

제 9 항에 있어서,

상기 스토리지 서비스를 자식 파티션에 이전하고, 상기 자식 파티션에 할당된 제 2 가상 기능이 상기 제 1 고유

네트워크 식별자를 사용하도록 구성하는 단계를 더 포함하는
이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 12

제 9 항에 있어서,

입/출력 메모리 관리 유닛이 상기 자식 파티션으로부터의 상기 입/출력 요청들에 관련된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는 단계를 더 포함하는

이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 13

제 9 항에 있어서,

입/출력 트래픽이 상기 고유 네트워크 식별자와 상기 제 2 고유 네트워크 식별자 사이에서 이동할 때, 상기 입/출력 트래픽의 보안 대책 준수 여부를 모니터링하도록 상기 네트워크 어댑터를 구성하는 단계를 더 포함하는

이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 14

제 9 항에 있어서,

원격 컴퓨터 시스템으로부터 사전 결정된 임계치를 넘는 입/출력 요청량이 수신되었다는 판단에 응답하여, 상기 스토리지 서비스를 인스턴스화하고 상기 제 1 고유 네트워크 식별자를 상기 스토리지 서비스에 할당하기 위한 요청을 상기 원격 컴퓨터 시스템으로 전송하는 단계를 더 포함하는

이전가능한 스토리지 서비스를 위한 컴퓨터 방법.

청구항 15

명령어가 저장된 컴퓨터 관독가능 저장 매체 장치로서, 상기 명령어는 컴퓨팅 장치 상에서 실행 시 상기 컴퓨팅 장치로 하여금 적어도,

스토리지 서비스를 자식 파티션에서 실현하고 -상기 스토리지 서비스는 제 2 자식 파티션에 대한 가상 하드 드라이브 디스크 입/출력 요청들을 관리하도록 구성되며, 상기 스토리지 서비스에는 네트워크 내의 제 1 고유 네트워크 식별자가 할당되며, 상기 네트워크 내에서 스토리지 타겟으로 사용되도록 또한 구성됨-,

상기 스토리지 서비스를 다른 파티션으로 이전하도록 하는

컴퓨터 관독가능 저장 매체 장치.

청구항 16

제 15 항에 있어서,

상기 컴퓨팅 장치 상에서 실행 시 상기 컴퓨팅 장치로 하여금 적어도,

원격 컴퓨터 시스템으로부터 사전 결정된 임계치를 넘는 입/출력 요청량이 수신되었다는 판단에 응답하여, 상기 스토리지 서비스를 인스턴스화하고 상기 제 1 고유 네트워크 식별자를 상기 스토리지 서비스에 할당하기 위한 요청을 상기 원격 컴퓨터 시스템으로 전송하도록 하는 명령어가 또한 저장됨

컴퓨터 관독가능 저장 매체 장치.

청구항 17

제 15 항에 있어서,

상기 컴퓨팅 장치 상에서 실행 시 상기 컴퓨팅 장치로 하여금 적어도 상기 스토리지 서비스를 다른 파티션으로 이전하도록 하는 명령어는 상기 컴퓨팅 장치로 하여금 상기 스토리지 서비스를 하이퍼바이저(hypervisor)로 이전하도록 하는

컴퓨터 관독가능 저장 매체 장치.

청구항 18

제 15 항에 있어서,

상기 컴퓨팅 장치 상에서 실행 시 상기 컴퓨팅 장치로 하여금 적어도 상기 스토리지 서비스를 다른 파티션으로 이전하도록 하는 명령어는 상기 컴퓨팅 장치로 하여금 상기 스토리지 서비스를 부모 파티션으로 이전하도록 하는

컴퓨터 관독가능 저장 매체 장치.

청구항 19

제 15 항에 있어서,

제 2 고유 네트워크 식별자와 연관된 제 2 가상 기능에 상기 제 2 자식 파티션을 할당 -상기 제 2 자식 파티션 및 상기 자식 파티션은 상기 제 1 고유 네트워크 식별자 및 상기 제 2 고유 네트워크 식별자에 할당됨- 하는 명령어를 더 포함하는

컴퓨터 관독가능 저장 매체 장치.

청구항 20

제 15 항에 있어서,

상기 자식 파티션에 대한 입/출력 요청들에 연관된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환시키도록 입/출력 메모리 관리 유닛을 구성하는 명령어를 더 포함하는

컴퓨터 관독가능 저장 매체 장치.

발명의 설명

배경 기술

[0001] 작업 부하를 끌어서 데이터센터로 옮기기 위해 가상 머신 기술(virtual machine technology)을 사용할 수 있다. 하나의 물리적 호스트로부터 다른 호스트로 작업 부하를 옮기는 이런 기능은 하드웨어 및 관리 비용이 훨씬 적게 드는 동적 머신 통합(dynamic machine consolidation)을 가능하게 하기 때문에, 사용자들에게 굉장히 유익하다. 가상 머신은 통상적으로 하이퍼바이저(hypervisor) 내에 있는 스토리지 가상화(storage virtualization)를 처리하는 모듈을 통해 저장 장치에 액세스한다. 이러한 모델에서, 가상 머신은 통상적으로 “파티션 버스 (Partition Bus)”라는 제목의 미국 특허 출원 제11/128,647호에 설명된 예시적인 파티션 버스와 같은 인터-파티션(inter-partition) 통신 버스 등의 소프트웨어 통신 경로를 통해 모듈에 스토리지 I/O 요청을 전송하며, 상기 출원의 내용은 그 전체가 본원에 참조로서 통합된다. 가상 머신과 하이퍼바이저(또는 관리 파티션) 간의 통신에는 통신 경로와 메시지 전송 시 발생할 수 있는 임의의 컨택스트 스위치 관리로 인해 CPU 사이클 비용이 들

게 된다. 따라서, CPU 비용을 줄임으로써 I/O 요청을 관리하는 효율을 증가시키는 기법이 요구된다.

발명의 내용

과제의 해결 수단

- [0002] 본 발명의 예시적인 실시예는 방법을 설명한다. 이 예에서, 상기 방법은 이전가능한(migratable) 스토리지 서비스를 실현하는(effectuating) 단계 - 이전가능한 스토리지 서비스는 자식 파티션(child partition)에 대한 가상 하드 디스크 입/출력 요청을 관리하도록 구성되며, 이전가능한 스토리지 서비스는 네트워크에 대한 고유 네트워크 식별자(unique network identifier)가 할당됨 -, 및 이전가능한 스토리지 서비스를 네트워크 상의 스토리지 타겟으로서 구성하는 단계를 포함하지만, 이에 한정되지 않는다. 전술한 내용에 추가하여, 다른 양태들이 본 발명의 일부를 구성하는 특허청구범위, 도면 및 텍스트에서 설명된다.
- [0003] 본 발명의 예시적인 실시예는 방법을 설명한다. 이 예에서, 상기 방법은 네트워크 어댑터에 대한 제 1 고유 네트워크 식별자를 자식 파티션에 대한 가상 하드 드라이브 디스크 입/출력 요청을 관리하도록 구성된 스토리지 서비스에 부착하는 단계 및, 네트워크 어댑터에 의해 실현된 가상 기능(virtual function)을 자식 파티션에 연관짓는 단계 - 가상 기능은 제 2 고유 네트워크 식별자를 포함함 - 를 포함하지만, 이에 한정되지 않는다. 전술한 내용에 추가하여, 다른 양태들이 본 발명의 일부를 구성하는 특허청구범위, 도면 및 텍스트에서 설명된다.
- [0004] 본 발명의 예시적인 실시예는 방법을 설명한다. 이 예에서, 상기 방법은 자식 파티션에서 스토리지 서비스를 실행하는 단계 - 스토리지 서비스는 제 2 자식 파티션에 대한 가상 하드 드라이브 디스크 입/출력 요청을 관리하도록 구성되며, 저장 장치는 네트워크의 고유 네트워크 식별자가 할당됨 - 를 포함하지만, 이에 한정되지 않는다. 전술한 내용에 추가하여, 다른 양태들이 본 발명의 일부를 구성하는 특허청구범위, 도면 및 텍스트에서 설명된다.
- [0005] 본 발명의 하나 이상의 다양한 양태들은 본원에서 설명되는 본 발명의 양태들을 실행하기 위한 회로 및/또는 프로그래밍을 포함할 수 있지만, 이에 한정되지 않으며, 이러한 회로 및/또는 프로그래밍은 시스템 설계자의 선택들에 따라 본원에서 설명되는 양태들을 실행하도록 구성되는 하드웨어, 소프트웨어 및/또는 펌웨어의 사실상 임의의 조합일 수 있다는 것을 당업자라면 이해할 것이다.
- [0006] 전술한 것은 요약이며, 따라서 필요에 의해서 간소화, 일반화 및 세부 사항의 생략을 포함한다. 당업자라면 이 요약은 예시적일 뿐 어떠한 식으로든 한정하려는 의도가 없음을 이해할 것이다.

도면의 간단한 설명

- [0007] 도 1은 본 발명의 양태들을 구현할 수 있는 예시적인 컴퓨터 시스템을 도시한다.
- 도 2는 본 발명의 양태들을 실시하기 위한 운영 환경을 도시한다.
- 도 3은 본 발명의 양태들을 실시하기 위한 운영 환경을 도시한다.
- 도 4는 SR-IOV 규격 네트워크 장치를 포함하는 컴퓨터 시스템을 도시한다.
- 도 5는 가상 환경에서의 메모리 간의 관계를 도시한다.
- 도 6은 본 발명의 일 실시예를 도시한다.
- 도 7은 본 발명의 양태들을 설명하기 위한 운영 환경을 도시한다.
- 도 8은 본 발명의 양태들을 실시하기 위한 동작 절차를 도시한다.
- 도 9는 도 8의 동작 절차의 대안적인 실시예를 도시한다.
- 도 10은 본 발명의 양태들을 실시하기 위한 동작 절차를 도시한다.
- 도 11은 도 10의 동작 절차의 대안적인 실시예를 도시한다.
- 도 12는 본 발명의 양태들을 실시하기 위한 동작 절차를 도시한다.

도 13은 도 12의 동작 절차의 대안적인 실시예를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0008] 실시예들은 하나 이상의 컴퓨터에서 실행될 수 있다. 도 1 및 아래의 설명은 본 발명을 구현할 수 있는 적절한 컴퓨팅 환경의 간략한 일반 설명을 제공하기 위한 것이다.
- [0009] 본원에서 사용되는 회로라는 용어는 하드웨어 인터럽트 제어기, 하드 드라이브, 네트워크 어댑터, 그래픽 프로세서, 하드웨어 기반 비디오/오디오 코덱과 같은 하드웨어 컴포넌트, 및 이러한 하드웨어를 동작시키는 데 사용되는 펌웨어를 포함할 수 있다. 회로라는 용어는 마이크로프로세서, 주문형 집적회로(application specific integrated circuits), 및/또는 하나 이상의 논리 프로세서, 예컨대, 펌웨어 및/또는 소프트웨어에 의해 구성되는 멀티- 코어 범용 처리 장치의 하나 이상의 코어를 포함할 수도 있다. 논리 프로세서(들)는 메모리, 예컨대 RAM, ROM, 펌웨어 및/또는 대용량 기억 장치로부터 로딩되는 기능(들)을 수행하도록 동작할 수 있는 논리를 구현하는 명령어에 의해 구성될 수 있다. 회로가 하드웨어와 소프트웨어의 조합을 포함하는 예시적인 실시예에서, 구현자는 논리를 구현하는 소스 코드를 작성할 수 있으며, 이어서 소스 코드는 논리 프로세서에 의해 실행될 수 있는 기계 판독 가능 코드로 컴파일된다. 당업자는 현재의 기술 수준이 하드웨어로 구현된 기능이나 소프트웨어로 구현된 기능 간에 거의 차이가 없을 정도로 진화한 것을 알 수 있으므로, 본원에 기술된 기능들을 실현하기 위한 하드웨어 대 소프트웨어의 선택은 단지 설계 선택일 뿐이다. 달리 말하자면, 당업자는 소프트웨어 프로세스가 등가의 하드웨어 구조로 변환될 수 있고, 하드웨어 구조 자체도 등가의 소프트웨어 프로세스로 변환될 수 있다는 것을 알 수 있으므로, 하드웨어 구현 대 소프트웨어 구현의 선택은 구현자에게 일임된다.
- [0010] 이제 도 1을 참조하면, 예시적인 컴퓨팅 시스템(100)이 도시되어 있다. 컴퓨팅 시스템(100)은 논리 프로세서(102), 예컨대, 실행 코어의 하이퍼쓰레드(hyperthread)를 포함할 수 있다. 하나의 논리 프로세서(102)가 도시되지만, 다른 실시예들에서 컴퓨터 시스템(100)은 다수의 논리 프로세서, 예컨대, 프로세서 기판마다 다수의 실행 코어들 및/또는 다수의 실행 코어들을 각자 가질 수 있는 다수의 프로세서 기판을 구비할 수 있다. 도면에 도시된 바와 같이, 각종 컴퓨터 판독가능 저장 매체(110)는 각종 시스템 컴포넌트들을 논리 프로세서(102)에 연결하는 하나 이상의 시스템 버스에 의해 상호 접속될 수 있다. 시스템 버스는 임의의 다양한 버스 아키텍처를 사용하는 메모리 버스 또는 메모리 제어기, 주변 장치 버스 및 로컬 버스를 포함하는 임의의 다양한 유형의 버스 구조일 수 있다. 예시적인 실시예에서, 컴퓨터 판독가능 저장 매체(110)는 예컨대, RAM(random access memory)(104), 저장 장치(106), 예컨대, 전기 기계식 하드 드라이브, 고체 상태 하드 드라이브 등, 펌웨어(108), 예컨대, 플래시 RAM 또는 ROM, 및 이동식 저장 장치(118), 예컨대, CD-ROM, 플로피 디스크, DVD, 플래시 드라이브, 외부 저장 장치 등을 포함할 수 있다. 당업자라면 자기 카세트, 플래시 메모리 카드, 디지털 비디오 디스크, 베르누이 카트리지(Bernoulli cartridges)와 같은 다른 유형의 컴퓨터 판독가능 저장 매체들도 사용될 수 있음을 알아야 한다.
- [0011] 컴퓨터 판독가능 저장 매체(110)는 컴퓨터(100)를 위해 프로세서 실행가능 명령어(122), 데이터 구조, 프로그램 모듈 및 기타 데이터의 비휘발성 및 휘발성 저장을 제공할 수 있다. 시동 중에 컴퓨터(100) 내의 구성 요소들 사이의 정보 전송을 돋는 기본 루틴을 포함하는 BIOS(basic input/output system)(120)는 펌웨어(108)에 저장될 수 있다. 다수의 프로그램이 펌웨어(108), 저장 장치(106), RAM(104), 및/또는 이동식 저장 장치(118)에 저장될 수 있고, 운영 체제 및/또는 애플리케이션을 포함하는 논리 프로세서(102)에 의해 실행될 수 있다.
- [0012] 컴퓨터(100)는 키보드 및 포인팅 장치를 포함하는, 그러나 이에 제한되지는 않는, 입력 장치들(116)을 통해 명령어들 및 정보를 수신할 수 있다. 다른 입력 장치들로는 마이크로폰, 조이스틱, 게임 패드 또는 스캐너 등을 포함할 수 있다. 이를 및 다른 입력 장치들은 시스템 버스에 연결되는 직렬 포트 인터페이스를 통해 논리 프로세서(102)에 접속될 수 있으며, USB(universal serial bus) 포트와 같은 다른 인터페이스에 의해 종종 접속되기도 한다. 디스플레이 또는 다른 유형의 디스플레이 장치도 그래픽 프로세서(112)의 일부이거나 그에 연결될 수 있는 비디오 어댑터와 같은 인터페이스를 통해 시스템 버스에 접속될 수 있다. 디스플레이뿐만 아니라, 컴퓨터는 통상적으로 스피커 및 프린터와 같은 다른 주변 출력 장치들(도시되지 않음)을 포함한다. 도 1의 예시적인 시스템은 호스트 어댑터, SCSI(Small Computer System Interface) 버스, 및 SCSI 버스에 접속된 외부 저장 장치도 포함한다.
- [0013] 컴퓨터 시스템(100)은 원격 컴퓨터에 대한 논리적 접속들을 이용하여 네트워킹된 환경에서 동작할 수 있다. 원격 컴퓨터는 다른 컴퓨터, 서버, 라우터, 네트워크 PC, 피어(peer) 장치 또는 다른 공통 네트워크 노드일 수 있으며, 통상적으로는 컴퓨터 시스템(100)과 관련하여 전술한 구성 요소들 중 다수 또는 모두를 포함할 수 있다.

[0014] LAN 또는 WAN 네트워킹 환경에서 사용될 때, 컴퓨터 시스템(100)은 네트워크 인터페이스 카드(114)를 통해 LAN 또는 WAN에 접속될 수 있다. 내장형 또는 외장형일 수 있는 NIC(114)는 논리 프로세서에 접속될 수 있다. 네트워킹된 환경에서, 컴퓨터 시스템(100)와 관련하여 도시된 프로그램 모듈들 또는 이들의 부분들은 원격 메모리 저장 장치에 저장될 수 있다. 기술된 네트워크 접속들은 예시적이며, 컴퓨터들 간에 통신 링크를 설정하기 위한 다른 수단이 사용될 수 있음을 이해할 것이다. 더욱이, 본 발명의 다양한 실시예들은 컴퓨터화된 시스템들에 특히 적합한 것으로 생각되지만, 본원의 어떤 것도 본 발명을 그러한 실시예들로 한정하는 것을 의도하지 않는다.

[0015] 이제, 도 2 및 3을 참조하면, 이들은 가상 머신을 실현하기 위해 구성된 컴퓨터 시스템들(200 및 300)의 상위 레벨 블록도들을 도시한다. 본 발명의 예시적인 실시예들에서, 컴퓨터 시스템(200 및 300)은 도 1에서 설명된 구성 요소들 및 가상 머신을 실현하도록 동작하는 컴포넌트들을 포함할 수 있다. 다시 도 2에서, 이런 컴포넌트는 이 분야에서 가상 머신 모니터라고도 불릴 수 있는 하이퍼바이저(202)이다. 도시된 실시예에서 하이퍼바이저(202)는 컴퓨터 시스템(100)의 하드웨어에 대한 액세스를 제어하고 중재하도록 구성될 수 있다. 대체로, 하이퍼바이저(202)는 파티션이라고 불리는 실행 환경, 예컨대, 가상 머신을 생성할 수 있다. 실시예들에서, 자식 파티션은 하이퍼바이저(202)에 의해 지원되는 격리(isolation)의 기본 단위로서 간주될 수 있다. 즉, 각각의 자식 파티션(246 및 248)은 하이퍼바이저(202) 및/또는 부모 파티션의 제어 하에 있는 일련의 하드웨어 자원들, 예컨대 메모리, 장치, 논리 프로세서 사이클 등에 맵핑될 수 있고, 하이퍼바이저(202)는 한 파티션의 프로세스들을 다른 파티션의 자원에 대한 액세스로부터 격리할 수 있으며, 예컨대, 한 파티션의 게스트 운영 체제가 다른 파티션의 메모리로부터 격리될 수 있다. 실시예들에서, 하이퍼바이저(202)는 독립형 소프트웨어 제품이거나, 운영 체제의 일부이거나, 마더보드의 펌웨어 내에 내장되거나, 특수화된 집적 회로들 또는 이들의 결합일수 있다.

[0016] 도시된 예에서, 컴퓨터 시스템(100)은 오픈 소스 커뮤니티에서 도메인 0과 마찬가지인 것으로 간주될 수 있는 부모 파티션(204)을 포함한다. 부모 파티션(204)은 오픈 소스 커뮤니티에서 백엔드 드라이버(back-end driver)라고 통상적으로 불리는 가상화 서비스 제공자들(VSPs)(228)을 이용함으로써, 자식 파티션들에서 실행되는 게스트 운영 체제들에 자원들을 제공하도록 구성될 수 있다. 이러한 예시적인 아키텍처에서, 부모 파티션(204)은 기반 하드웨어(underlying hardware)에 대한 액세스를 게이트할 수 있다. 대체로, VSP들(228)은 오픈 소스 커뮤니티에서 프론트엔드 드라이버(front-end drivers)라고 통상적으로 불리는 가상화 서비스 클라이언트들(VSC들)을 통해 하드웨어 자원들로 인터페이스들을 다중화하는 데 사용될 수 있다. 각각의 자식 파티션은 가상 프로세서들(230-232)과 같은 하나 이상의 가상 프로세서들을 포함할 수 있으며, 게스트 운영 체제들(220-222)은 가상 프로세서들 상에서 실행할 쓰레드들을 관리하고 스케줄링 할 수 있다. 일반적으로, 가상 프로세서들(230-232)은 특정 아키텍처를 갖는 물리적 프로세서의 표현을 제공하는 실행 가능 명령어들 및 관련 상태 정보이다. 예컨대, 하나의 자식 파티션은 인텔 x86 프로세서의 특성들을 갖는 가상 프로세서를 구비할 수 있는 반면, 다른 가상 프로세서는 PowerPC 프로세서의 특성들을 가질 수 있다. 이 예에서, 가상 프로세서들은 컴퓨터 시스템의 논리 프로세서들에 맵핑될 수 있으며, 따라서 명령어의 가상 프로세서 실행은 논리 프로세서들에 의해 백킹(backing)될 것이다. 따라서, 이러한 실시예들에서는, 예를 들어 다른 논리 프로세서가 하이퍼바이저 명령어들을 실행하고 있는 동안에, 다수의 가상 프로세서가 동시에 실행될 수 있다. 파티션 내의 가상 프로세서들, 다양한 VSC들 및 메모리의 결합은 가상 머신으로 간주될 수 있다.

[0017] 게스트 운영 체제들(220-222)은 예를 들어, Microsoft[®], Apple[®], 오픈 소스 커뮤니티 등으로부터의 운영 체제들과 같은 임의의 운영 체제를 포함할 수 있다. 게스트 운영 체제들은 사용자/커널 동작 모드들을 사용할 수 있으며, 스케줄러들, 메모리 관리자들 등을 포함할 수 있는 커널들을 구비할 수 있다. 각각의 게스트 운영 체제(220-222)는 단말 서버, 전자 상거래 서버, 이메일 서버 등 및 게스트 운영 체제들 그 자체와 같은 애플리케이션들을 저장하고 있을 수 있는 관련 파일 시스템들을 구비할 수 있다. 게스트 운영 체제들(220-222)은 가상 프로세서들(230-232) 상에서 실행될 쓰레드들을 스케줄할 수 있으며, 그러한 애플리케이션들의 인스턴스들(instances)이 실행될 수 있다.

[0018] 이제 도 3을 참조하면, 이 도면은 앞서 도 2에서 설명한 것의 대안적인 아키텍처를 도시한다. 도 3은 도 2의 것들과 유사한 컴포넌트들을 도시하지만, 이 실시예에서 하이퍼바이저(202)는 가상화 서비스 제공자(228) 및 장치 드라이버들(224)을 포함할 수 있으며, 부모 파티션(204)은 구성 유ти리티들(236)을 포함할 수 있다. 이러한 아키텍처에서, 하이퍼바이저(202)는 도 2의 하이퍼바이저(202)와 동일 또는 유사한 기능을 수행할 수 있다. 도 3의 하이퍼바이저(202)는 독립형 소프트웨어 제품이거나, 운영체제의 일부이거나, 마더보드의 펌웨어 내에 내장되거나 또는 하이퍼바이저(202)의 일부가 특수화된 집적 회로들에 의해 실현될 수 있다. 이 예에서, 부모 파티션(204)은 하이퍼바이저(202)를 구성하는 데 사용될 수 있는 명령어들을 구비할 수 있지만, 하드웨어 액세스 요

청들은 부모 파티션(204)으로 전달되는 것이 아니라 하이퍼바이저(202)에 의해 처리될 수 있다.

[0019] 본 발명의 실시예들에서, 그 전체가 본원에 분명하게 참조로서 통합된 “SR-IOV 스펙(Single Root Input/Output Virtualization specification)” 개정 1.0에 따르는 네트워크 어댑터가 도면에서 설명된 것들과 같은 컴퓨터 시스템에 설치될 수 있다. 예시적인 어댑터는 Intel®의 “기가비트 ET 듀얼 포트 서버 어댑터(Gigabit ET Dual Port Server Adapter)”일 수 있다. SR-IOV 가능 네트워크 장치들은 예를 들어, 물리적 기능으로의 인터페이스를 가상화함으로써 가상 머신들 간에 I/O 어댑터 또는 기타 임의의 프로세스를 공유할 수 있는 하드웨어 장치이다. VF(가상 기능, virtual function)으로도 알려진 가상화된 각각의 인터페이스는 대개 컴퓨터 시스템의 PCI-익스프레스 버스 상의 개별 네트워크 인터페이스이다. 예를 들어, 각 가상 기능은 에뮬레이션된 PCI 구성 공간 및 고유 네트워크 식별자, 예컨대, MAC 어드레스(media access control address), 월드 와이드 네임(world wide name) 등을 가질 수 있다. 따라서, 각 가상 기능은 물리적 기능에 액세스하기 위한 고유한 어드레스를 갖고 확실히 구분된 개별 경로를 지원할 수 있다.

[0020] 도 4를 참조하면, 이 도면은 SR-IOV 규격 어댑터(402)(“어댑터”)를 포함하는 컴퓨터 시스템(400)을 도시한다. 전술한 바와 마찬가지로, 컴퓨터 시스템(400)은 도 1-3에 관한 상기의 컴포넌트들과 유사한 컴포넌트들을 포함할 수 있다. 어댑터(402)는 포트에 해당할 수 있는 물리적 기능(410)을 포함할 수 있으며, 이는 네트워크 및 내부 라우터(412)에 접속될 수 있다. 내부 라우터(412)는 예컨대, 가상 기능(404 또는 406)이 할당된 각각의 가상 포트를 갖고 있는 가상 어댑터와 같은, 어댑터(402)의 네트워크 식별자(420-424)로 또는 그로부터 데이터를 라우팅하도록 구성될 수 있다.

[0021] 예시적인 실시예에서, 네트워크 어댑터(402)는 이더넷(Ethernet) 어댑터일 수 있고, 가상 기능은 가상 이더넷 어댑터일 수 있다. 이런 예에서, 가상 기능의 고유 식별자는 이더넷 MAC 어드레스일 것이다. 광 채널(Fibre channel) 예에서, 어댑터(402)는 광 채널 호스트 버스 어댑터일 수 있으며, 가상 기능은 월드 와이드 노드 이름 및 월드 와이드 포트 이름을 포함하는 월드 와이드 네임을 갖고 있는 가상 광 채널 호스트 버스 어댑터일 수 있다. 인피니밴드(Infiniband) 예에서, 가상 기능은 글로벌 식별자를 갖고 있는 가상 인피니밴드 말단일 수 있다.

[0022] 네트워크 어댑터(424)는, 광 채널 호스트 버스 어댑터 또는 이더넷 어댑터와 같은 특정 네트워크 어댑터들이 다수의 고유 식별자들이 단일 물리적 포트를 공유할 수 있게 함을 나타내는 점선으로 도시된다. 광 채널에서, 이러한 기능은 N_Port ID 가상화 또는 NPIV라고 하며, 이더넷에서, 어댑터는 이른바 무차별 모드(promiscuous mode)에서 동작하고, 내장된 가상 스위치를 포함하거나, 메모리 버퍼를 분리하기 위해 특정 MAC 어드레스로 보내지는 데이터를 필터링하고 라우팅할 수 있다.

[0023] 각 네트워크 식별자는 네트워크 상에서 전송될 수 있도록 정보를 포맷팅하도록 구성된 소프트웨어 프로토콜 스택(414-418)과 연계될 수 있다. 특정 TCP/IP 예에서, 프로세스는 애플리케이션 층 포트를 통해 TCP/IP 스택의 애플리케이션 층의 인스턴스에 결합될 수 있다. 결국, 프로토콜 스택의 다른 기능들에 의해 처리된 정보는 패브릭(fabric)을 통해 전송될 수 있는 데이터 프레임을 조립하는 역할을 하는 MAC 층(media access control layer)으로 알려진 층 안의 기능 그룹에 의해 처리될 수 있다. 프로토콜 스택의 이런 층은 네트워크에서 전송된 프레임들에 가상 기능에 관한 MAC 어드레스를 추가한다. 그 후, 프로토콜 스택은 프레임의 정보를 전기 신호로 변환하여 프레임들을 네트워크로 보내도록 구성된 물리적 층에 조립된 프레임들을 전달한다.

[0024] I/O-MMU(input/output memory management)(426)를 사용하여 PCI-익스프레스 상호 접속과 같이 직접적인 메모리 액세스 동작을 실행할 수 있는 I/O 상호 접속을 RAM에 연결할 수 있다. 본 발명의 일 실시예에서, I/O-MMU(426)는 파티션들로부터의 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는, 하이퍼바이저(202)의 페이지 테이블(page tables)을 포함할 수 있다. I/O-MMU(426)는 컴퓨터 시스템(400)의 여러 장소에 있을 수 있음을 나타내는 점선으로 표시된다. 예를 들어, I/O-MMU는 마더보드 상의 칩이나 논리 프로세서의 컴포넌트일 수 있다.

[0025] 도 5는 본 발명의 실시예에서 게스트 물리적 어드레스 및 시스템 물리적 어드레스 간의 관계를 도시한다. 게스트 메모리는 하이퍼바이저(202)에 의해 제어되는 메모리의 도면이다. 게스트 물리적 어드레스는 하이퍼바이저(202)에 의해 관리되는 SPA(system physical address), 예컨대, 물리적 컴퓨터 시스템의 메모리에 의해 백킹될 수 있다. 도면에 도시된 바와 같이, 일 실시예에서, GPA들 및 SPA들은 메모리 블록들, 예컨대, 메모리의 하나 이상의 페이지들로 배열될 수 있다. GPA들과 SPA들 사이의 관계는 “강화된 새도우 페이지 테이블 알고리즘(Enhanced Shadow Page Table Algorithms)”이라는 제목의 미국 특허 출원 제11/128,665호에 설명된 것들과 같은 새도우 페이지 테이블에 의해 유지될 수 있으며, 상기 출원의 내용은 그 전체가 본원에 참조로서 통합된다. 동작 시에, 게스트 운영 체제가 GPA 블록 1에 데이터를 저장할 때, 데이터는 실제로는 시스템 상의 블록 6과 같

은 상이한 SPA에 저장될 수 있다. 본 발명의 실시예에서, I/O-MMU(426)는 I/O 동작 중에 변환을 수행하여, 하나의 GPA 공간으로부터 다른 GPA 공간으로 직접 스토리지 데이터를 옮길 수 있다. 이와 같은 실시예에서, 이런 변환을 실현하기 위해 하이퍼바이저 명령어를 실행하지 않아도 되기 때문에 논리 프로세서 사이클이 절약될 수 있다.

[0026] 도 6은 가상 스토리지 타겟 오프로드 기법을 설명하기 위한 상위 레벨의 운영 환경을 도시한다. 도 6은 SR-IOV 네트워크 어댑터(402) 및 그 가상 기능(406)을 통해 스토리지 가상화 클라이언트(604)와 통신하는 가상 머신 스토리지 서비스(602)를 도시한다. 도면에 도시된 바와 같이, 본 발명의 이런 실시예에서, SR-IOV 네트워크 어댑터(402)는 소프트웨어 통신 경로를 우회하여 가상 머신들과 가상 머신 스토리지 서비스들 간에 I/O를 전송하는데 사용될 수 있다. 이는 결국 가상 머신을 위한 I/O를 실행하는 데 사용되는 CPU 사이클의 양을 감소시키고, 스토리지 서비스(602)를 이전하는 기능을 향상시키고, 부모 파티션에서 실행되는 호스트 운영 체제 및/또는 하이퍼바이저(202)의 부담을 줄일 가능성이 있다.

[0027] 가상 머신 스토리지 서비스(602)는 자식 파티션을 대신하여, SAN에서 제공되는 LUN(logical unit numbers)과 같은 물리적 저장 장치, 예컨대, 다른 스토리지 가상화 기법에 의해 이미 가상화된 디스크와 통신하도록 구성될 수 있다. 일례에서, 이는 가상 머신들로부터 I/O 요청을 수신하고 이 요청을 LUN으로 라우팅하도록 가상 머신 스토리지 서비스(602)를 구성하는 것을 포함할 수 있다. 다른 예에서, LUN이 서브-얼로케이션된(sub-allocated)된 경우, 가상 머신 스토리지 서비스(602)는 가상 하드 드라이브들을 생성하고, 이들을 가상 머신에 노출시키고, LUN 또는 물리적 드라이브 상의 VHD(virtual hard drive) 파일로써 이들을 저장하도록 구성될 수 있다. VHD 파일은 단일 파일 안에 캡슐화될 수 있는 가상 머신 하드 디스크를 나타낸다. 가상 머신 스토리지 서비스(602)는 이 파일을 파싱하고(parse), 마치 물리적 스토리지인 것처럼 게스트 운영 체제(220)에 노출될 수 있는 디스크를 실현할 수 있다. 가상 머신 스토리지 서비스(602)에 의해 생성되는 가상 하드 디스크는 로컬인 것처럼 보이는 게스트 운영 체제에 액세스 가능한 버스에게 보여진다.

[0028] 본 발명의 일 실시예에서, 가상 머신 스토리지 서비스(602)는 고유 네트워크 식별자를 가상 머신 스토리지 서비스(602)에 부착하고, 예컨대, 가상 머신 스토리지 서비스(602)를 데이터센터에서의 스토리지 타겟으로 광고하기 위해 사용되는 스토리지 타겟 파라미터들을 구성함으로써, 네트워크의 광 채널 타겟이나 iSCSI(internet small computer system interface) 타겟과 같은 스토리지 타겟으로 구성될 수 있다. iSCSI 예시 환경에서, 가상 머신 스토리지 서비스(602)는 인터넷 프로토콜을 통해 자식 파티션들에 액세스하는 LUN을 실현시킴으로써 iSCSI 타겟을 구현할 수 있다. 가상 머신 스토리지 클라이언트(604) 또는 게스트 운영 체제는 가상 머신 스토리지 서비스(602)의 어드레스를 얻을 수 있고, SCSI 하드 디스크로의 접속을 에뮬레이션하는 접속이 설정될 수 있다. 가상 머신 스토리지 클라이언트(604)는 SCSI 또는 하드 드라이브를 다루는 방식과 동일한 방식으로 가상 머신 스토리지 서비스(602)를 다룰 수 있으며, 가상 머신 스토리지 서비스(602)는 자식 파티션들에게 가상 하드 드라이브를 제공할 수 있다. 이 예시에서, 가상 머신 스토리지 클라이언트(604)는 네트워킹된 파일 시스템 환경에서처럼 원격 디렉토리를 마운팅(mounting)하지 않고도 가상 머신 스토리지 서비스(602)가 제공하는 가상 디스크 상에서 직접 파일 시스템을 생성 및 관리할 수 있다. 게스트 OS(220)의 관점에서 보면, 이는 하드 드라이브와 유사한 방식으로 동작하는 하나 이상의 논리 유닛에 접속하는 네트워크에 접속된 네트워크 어댑터를 가지고 있다.

[0029] 도 7은 본 발명의 양태들을 실시하기 위한 예시적인 운영 환경을 도시한다. 도 6과 유사하게, 소프트웨어 통신 경로를 사용하여 I/O를 전송할 필요를 없앰으로써 가상 머신들과 가상 머신 스토리지 서비스들 간에 I/O를 전송하는 데 하나 이상의 SR-IOV 네트워크 어댑터가 사용될 수 있다. 이는 가상 머신을 위한 I/O를 실행하는 데 사용되는 CPU 사이클의 양을 감소시키고, 스토리지 서비스(602)를 이전하는 기능을 향상시키고, 호스트 운영 체제 및/또는 하이퍼바이저(202)의 부담을 잠재적으로 줄일 가능성이 있다.

[0030] 이런 예시적인 환경에서, 두 컴퓨터 시스템(700 및 702)을 포함하는 데이터센터가 스위치(704)에 접속된 것으로 도시된다(두 개의 컴퓨터 시스템이 도시되었지만, 당업자라면 데이터센터가 훨씬 많은 컴퓨터 시스템을 가질 수 있음을 이해할 수 있다). 컴퓨터 시스템(700 및 702)은 도 1-4에서 설명된 것들과 유사한 컴포넌트들을 가지고 있으며, 스위치(704)는 상호접속되는 스위치 및 라우터들의 전체 인프라일 수 있다. 나아가, 컴퓨터 시스템(700 및 702)은 본원에 개시된 기법을 보다 명확하게 설명하기 위해 특정 특징들을 포함하는 것으로 도시되며, 본 발명은 도시된 토플로지에서 구현되는 것에 한정되지는 않는다.

[0031] 컴퓨터 시스템(700)은 본원에 개시된 기법을 따라 스토리지 서비스(602)를 이전하도록 구성된 관리자(250)를 포함할 수 있고, 따라서 가상 머신 스토리지 서비스(602)는 동일하거나 상이한 컴퓨터 시스템에서 한 파티션으로부터 다른 파티션으로 이전될 수 있음을 나타내기 위해 점선으로 도시된다. 가상 기능(706 및 708)은 특정 실시

예에서 가상 머신 스토리지 서비스(602)가 가상 기능을 통해 액세스할 필요 없이 직접 SR-IOV 어댑터(402)와 접속할 수 있음을 나타내기 위해 점선으로 표시된다. 이런 예시적인 실시예에서, 부모 파티션(204 및 712)은 물리적 하드웨어를 제어할 수 있고, 가상 기능을 필요로 하지 않게 된다.

[0032] 도면의 일반적인 개관을 계속 살펴보면, 가상 머신 스토리지 서비스(602)는 본 발명의 실시예에서 자신에게 할당된 고유 식별자를 추출하고 이 식별자를 임의의 필요한 상태 정보와 함께 다른 파티션으로 이동시킴으로써 이전될 수 있다. 일례로, 이 프로세스는 논리 프로세서 실행 관리자(250)에 의해 고유 식별자를 추출하고, 논리 프로세서 실행 관리자(250)에 의해 그 고유 식별자를 다른 파티션의 가상 기능에 부착하도록 어댑터(402 또는 718)에 지시하며, 논리 프로세서 실행 관리자(250)에 의해 가상 머신 스토리지 서비스(602)의 인스턴스가 가상 기능에 부착되도록 인스턴스에 지시하는 단계를 포함할 수 있다. 다른 예로, 이 프로세스는 논리 프로세서 실행 관리자(250)에 의해 고유 식별자를 추출하고, 논리 프로세서 실행 관리자(250)에 의해 그 고유 식별자를 어댑터(402 또는 718)에 부착하도록 어댑터(402 또는 718)에 지시하며, 논리 프로세서 실행 관리자(250)에 의해 고유 식별자를 사용하여 패브릭에서 통신하도록 다른 파티션에 인스턴스화된(instantiated) 가상 머신 스토리지 서비스(602)의 인스턴스에 지시하는 단계를 포함할 수 있다.

[0033] 다음은 동작 절차를 나타내는 일련의 순서도들이다. 이해의 편의를 위해, 처음의 순서도들이 전반적인 "큰 그림"의 관점을 통해 구현들을 설명하고, 후속 순서도들이 추가 사항들 및/또는 세부 사항을 더 제공하도록 순서도들이 조직된다. 더욱이 당업자라면 점선에 의해 도시되는 동작들은 옵션으로 간주된다는 것을 알 수 있다.

[0034] 이제, 도 8을 참조하면, 이 도면은 본 발명의 양태들을 실시하기 위한 동작 절차를 나타낸다. 도면에 도시된 바와 같이, 동작 절차는 동작 800에서 시작하고, 동작 802는 이전가능한 스토리지 서비스를 실현하는 단계를 나타내며, 이전가능한 스토리지 서비스는 자식 파티션에 대한 가상 하드 디스크 입/출력 요청을 관리하도록 구성되고, 이전가능한 스토리지 서비스는 네트워크에 대한 고유 네트워크 식별자가 할당된다. 예를 들어, 도 6을 참조하면, 가상 머신 스토리지 서비스(602)와 같은 이전가능한 스토리지 서비스는 컴퓨터 시스템에 의해 실현될 수 있다. 즉, 가상 머신 스토리지 서비스(602)를 나타내는 명령어들이 논리 프로세서에 의해 실행될 수 있다. 가상 머신 스토리지 서비스(602)는 고유 네트워크 식별자가 부착되어 있으며, 한 파티션에서 다른 파티션으로 저절로, 즉, 다른 관리 모듈을 옮기지 않고도 옮겨질 수 있기 때문에, 가상 머신 스토리지 서비스(602)는 이전 가능하다고 간주된다.

[0035] 예시적인 실시예에서, 가상 머신 스토리지 서비스(602)는 네트워크에 대한 고유 식별자를 배타적으로 사용할 수 있고, 예를 들어, 이는 데이터센터의 고유 네트워크 어드레스를 사용하여 통신하는 유일한 프로세스일 수 있다. 이런 예시에서, 가상 머신 스토리지 서비스(602)는, 상태 정보가 다른 파티션으로 전송되어 가상 머신 스토리지 서비스(602)의 다른 인스턴스를 구성하는 데 사용될 수 있도록 자신의 상태를 직렬화하도록 구성될 수 있다. 다른 예시적인 실시예에서, 가상 머신 스토리지 서비스(602)는 가상 기능에 부착된 가상 머신에서 실행될 수 있다. 이런 예에서, 가상 머신 스토리지 서비스(602)는 또한 고유 식별자를 사용하여 네트워크와 배타적으로 통신할 수도 있다. 가상 머신 스토리지 서비스(602)의 이전은 가상 머신 스토리지 서비스(602)를 포함하는 가상 머신의 상태를 직렬화하여 이를 다른 파티션에 전송하는 것을 포함할 수 있다.

[0036] 구체적인 예를 들어, 도 7를 참조하면, 가상 머신 스토리지 서비스(602)는 부모 파티션(204)로부터 자식 파티션(246)으로 이전될 수 있다. 이런 구체적인 예시에서, 논리 프로세서는 관리자(250)를 실행할 수 있고, 즉, 논리 프로세서는 관리자(250)를 나타내는 명령어를 실행할 수 있고, 데이터센터에서의 통신을 위해 가상 머신 스토리지 서비스(602)에 의해 사용되는 고유 식별자를 추출할 수 있다. 그 후, 고유 식별자는 자식 파티션(246)으로 전송되고, 가상 머신 스토리지 서비스(602)의 인스턴스가 시작될 수 있다. 어댑터(402)의 라우팅 테이블이 업데이트될 수 있고, I/O 요청은 어댑터(402)에 의해 부모 파티션(204) 대신 자식 파티션(246)으로 라우팅될 수 있다. 이런 예에서, 자식 파티션(246)은 이미 사용 중인 임의의 다른 고유 식별자들에 대해 이 고유 식별자를 사용하도록 구성될 수 있다.

[0037] 도 8의 설명을 계속하면, 동작 804는 이전가능한 스토리지 서비스를 데이터센터에서의 스토리지 타겟으로 구성하는 단계를 나타낸다. 예를 들어, 본 발명의 일 실시예에서, 가상 머신 스토리지 서비스(602)가 데이터센터의 스토리지 타겟이 되도록 구성될 수 있다. 전술한 바와 유사하게, 가상 머신 스토리지 서비스(602)는 네트워크의 고유 네트워크 식별자가 부착될 수 있고, 게스트 OS(220)에 의해 스토리지 타겟으로 탐지될 수 있다. 게스트 OS(220)와 가상 머신 스토리지 서비스(602) 사이에 통신 세션이 열릴 수 있고, 게스트 OS(220)는 가상 머신 스토리지 서비스(602)에 의해 노출된 가상 하드 드라이브(들)를 탐지하여, 마치 로컬 하드 드라이브인 것처럼 가상 디스크들을 사용할 수 있다. 구체적인 예를 들면, 가상 머신 스토리지 서비스(602)는 전술한 바와 같이

iSCSI 타겟을 에뮬레이션할 수 있다. 이런 예시에서, 가상 머신 스토리지 서비스(602)는 물리적 디스크 대신 가상 디스크를 노출시키고, LUN 또는 물리적 디스크에 읽고 쓰기를 함으로써 가상 머신으로부터의 I/O를 처리할 수 있다.

[0038] 도 9를 참조하면, 이 도면은 도 8의 동작 절차의 대안적인 실시예를 도시한다. 동작 906은 이전가능한 스토리지 서비스를 원격 컴퓨터 시스템에 이전하는 단계를 나타낸다. 예를 들어, 도 6을 참조하면, 일 실시예에서 이전가능한 스토리지 서비스, 예컨대, 가상 머신 스토리지 서비스(602)는 데이터센터의 원격 컴퓨터 시스템에 이전될 수 있다. 예를 들어, 일 실시예에서, 원격 컴퓨터 시스템은 현재 가상 머신 스토리지 서비스(602)를 호스팅하는 컴퓨터 시스템보다 넓은 I/O 대역폭을 이용가능하여, 스토리지 서비스(602)를 옮기려는 결정을 할 수 있다. 이런 예에서, 논리 프로세서는 관리자(250)를 실행하고, 스토리지 서비스(602)에 부착된 고유 식별자를 추출하여, 이를 원격 컴퓨터에 전송할 수 있다. 그 후에, 원격 컴퓨터의 관리자(250)는 스토리지 서비스(602)의 인스턴스에 그 고유 식별자를 부착할 수 있다.

[0039] 구체적인 예를 들어, 도 7을 참조하면, 가상 머신 스토리지 서비스(602)는 자식 파티션(246)으로부터 부모 파티션(712)으로 이전될 수 있다. 이런 구체적인 예에서, 컴퓨터 시스템(700)의 관리자(250)는 가상 머신 스토리지 서비스(602)에 부착된 고유 식별자를 추출하여 이를 컴퓨터 시스템(702)으로 전송할 수 있다. 컴퓨터 시스템(702)의 관리자(250)가 논리 프로세서 상에서 실행되어, 부모 파티션(712)에서 실행되고 있는 가상 머신 스토리지 서비스(602)의 인스턴스에 그 고유 식별자를 부착할 수 있다. 이런 예에서, 가상 스토리지 서비스(602)는, 가상 기능(708) 유무와는 관계 없이 자식 파티션(246)에서 가상 머신 스토리지 서비스(602)에 의해 서비스되었던 클라이언트들로부터의 I/O를 전송/수신할 때, 고유 식별자를 사용할 수 있다.

[0040] 이런 구체적인 예시에서, 가상 머신 스토리지 서비스(602)에 대한 상태 정보 및 프로토콜 스택은 컴퓨터 시스템(702)에 전송되어 I/O 서비스가 중단되지 않을 수 있다. 예를 들어, 컴퓨터 시스템(702)의 관리자(250)가 컴퓨터 시스템(700)의 프로토콜 스택의 기능상 동일한 상태를 최소한 반영하도록 프로토콜 스택을 구성하기에 충분한 정보가 컴퓨터 시스템(702)에 전송될 수 있다. 상태 정보는 전송될 다음 패킷의 수, 사용된 소켓(socket) 번호, 최대 버퍼 크기, 서버의 포트 번호, 클라이언트의 포트 번호 등을 포함할 수 있다. 상태 정보는 상위 레벨의 프로토콜 정보와 같은 정보도 포함할 수 있다. 다른 예로 사용된 암호화 프로토콜에 관련된 정보를 들 수 있다.

[0041] 이런 예시적인 실시예에서, 클라이언트의 관점에서는 접속이 끊어지는 대신 잠시 멈추었기 때문에, 클라이언트들에 대한 서비스가 중단되지 않고 작동될 것이다. 예를 들어, 가상 머신 스토리지 서비스(602)가 이전될 때, 프로토콜 스택은 실행하고 있던 현재 동작들을 예컨대, 완료하거나 취소함으로써 끝낼 수 있고, 프로토콜이 짧은 시간 동안 정보 전송을 중지하도록 요청하는 백오프 메시지(back off message)를 가상 머신 스토리지 클라이언트(604)에 둑인 프로토콜에 선택적으로 전송할 수 있다. 컴퓨터 시스템(702) 상의 프로토콜 스택이 인스턴스화되었을 때, 이 스택은 컴퓨터 시스템(700) 상의 프로토콜 스택과 동일한 상태를 가질 수 있고, 네트워크 상에서 컴퓨터 시스템(700)과 이전에 연계되었던 고유 식별자와 통신할 수 있다. 컴퓨터 시스템(702) 상에 새롭게 구성된 프로토콜 스택은 재개 메시지(resume message)를 선택적으로 전송하도록 구성될 수 있고, 가상 머신 스토리지 클라이언트(604)를 제공하는 프로토콜은 I/O의 전송을 재개할 수 있다. 스위치(704)는 프로토콜 메시지들이 컴퓨터 시스템(702)의 가상 머신 스토리지 서비스(602)에 전송되도록 라우팅을 해결할 수 있다.

[0042] 도 9의 설명을 계속하면, 동작 908은 입/출력 메모리 관리 유닛이 자식 파티션에 대한 입/출력 요청과 관련된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하도록 구성하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 본 발명의 실시예에서, 컴퓨터 시스템(700)의 입/출력 메모리 관리 유닛(426)은 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는 데 사용될 수 있다. 예를 들어, 게스트 운영 체제(220)가 I/O 동작을 시작할 때, 예컨대, 읽기나 쓰기를 할 때, 게스트 운영 체제(220)는 시스템 물리적 어드레스로 변환될 필요가 있을 수 있는 게스트 물리적 어드레스를 포함하는 명령어를 생성할 수 있다. 예시적인 실시예에서, 이를 변환은 MMU가 아닌 I/O-MMU(426)에서 일어날 수 있다. 메모리 변환을 I/O-MMU(426)에 오프로드시킴으로써, 하이퍼바이저(202) 및/또는 부모 파티션(204)에 대한 부담이 줄어든다. 예를 들어, 게스트 OS(220)는 임의의 디스크 오프셋을 게스트 메모리 어드레스로 읽기 위한 요청을 포함하는 읽기 명령을 발행할 수 있다. 이런 예시에서, 입/출력 메모리 관리 유닛(426)은 자식 파티션(248)의 게스트 메모리 어드레스를 시스템 어드레스에 매핑시키는 테이블을 사용하여, 게스트 메모리 어드레스를 게스트가 읽어들이기를 원하는 게스트 메모리 어드레스를 물리적으로 백킹하는 시스템 어드레스로 변환할 수 있다. 가상 머신 스토리지 서비스(602)는 이 요청을 수신하고, 클라이언트가 요청하고 있는 정보를 얻고, 이전에 요청된 데이터를 포함하는 응답 메시지를 제공할 수 있다. 응답은 게스트 메모리 어드레스로써 지정된 버퍼에서 제공될 수 있으며, 이 경우 어댑터(402) 및 I/O-MMU(426)는 제

공된 게스트 메모리 어드레스를 시스템 물리적 어드레스로 변환할 수 있고, 이후에 어댑터(402)가 응답 베퍼로 부터의 응답 데이터를 요청 베퍼로 복사하여 사용자의 요청을 충족시킬 수 있다.

[0043] 이런 기법은 클라이언트가 가상 머신 스토리지 서비스(602)와 동일한 물리적 컴퓨터에 있을 때, 주변 장치에 의해 실행되는 메모리-투-메모리 DMA(direct memory access) 동작과 유사하다. 이런 예시적인 실시예에서, I/O 동작은 네트워크 어댑터(402)가 시스템 물리적 어드레스의 하나의 블록으로부터 정보를 검색하고, 가상 머신 스토리지 클라이언트(604)나 가상 머신 스토리지 서비스(602)를 대신하여 이를 시스템 물리적 어드레스의 다른 블록으로 옮기기 때문에, 메모리-투-메모리 DMA 동작과 유사할 수 있다. 구체적인 예로 가상 머신 스토리지 클라이언트(604)에서 발행한 읽기 명령을 들 수 있다. 이런 예에서, 가상 머신 스토리지 클라이언트(604)는 자신이 제어하는 메모리 페이지들로 읽기를 원하는 스토리지 데이터의 페이지들을 지정하는 읽기 명령을 발행할 수 있다. 이런 예에서, 데이터 페이지들은 요청을 충족시키기 위해 가상 머신 스토리지 서비스(602)에 의해 사용되는 페이지들로 복사되고, 가상 머신 스토리지 클라이언트(604)에 의해 지정되는 메모리 페이지들에 데이터가 복사된다.

[0044] 도 9의 설명을 계속하면, 동작 910은 자식 파티션으로부터 입/출력 작업(job) 요청을 수신하는 단계를 나타내며, 여기서 자식 파티션은 네트워크에 대한 제 2 고유 네트워크 식별자를 포함하는 가상 기능에 부착된다. 예를 들어, 도 6에 도시된 바와 같이, 일 실시예에서 자식 파티션(248)은 가상 기능(406)을 포함할 수 있다. 이런 예에서, 자식 파티션(248)은 배타적으로 가상 기능(406)을 통해 SR-IOV 어댑터(402)에 접속할 수 있고, I/O 요청을 전송할 수 있다. 어댑터(402)는 명령어가 가상 머신 스토리지 서비스(602)와 관련된 고유 식별자에 대한 것인지를 결정하고, 그 명령어를 가상 머신 스토리지 서비스(602)에 전송할 수 있다. 이 경우에, 자식 파티션(248)으로부터의 I/O 명령어는 하이퍼바이저(202)를 통해 전송된 요청 없이도 또는 파티션-투-파티션 통신 인터페이스를 통하여 않고도 가상 머신 스토리지 서비스(602)에 전송될 수 있다. 또한, 어댑터(402)는 클라이언트(604) 및 가상 머신 스토리지 서비스(603)의 고유 식별자들을 사용하여 어떤 메모리 페이지들을 베퍼로써 사용할지, 그리고 그에 따라, 어떤 어드레스 공간 사이에 데이터를 복사할지를 결정할 수 있다.

[0045] 구체적인 예시에서, I/O 요청은 (게스트 물리적 어드레스에서) 데이터의 위치, 및 데이터가 기록되어야 하는 가상 하드 드라이브 상의 위치를 지정하는 쓰기 명령일 수 있다. 이런 예에서, 스토리지 가상화 클라이언트(604)는 가상 머신 스토리지 서비스(602)의 고유 식별자를 겨냥한 하나 이상의 정보 패킷에 요청을 넣을 수 있다. 이런 예에서, 어댑터(402)는 이 요청을 수신하고 이를 가상 머신 스토리지 서비스(602)에 전송할 수 있다. 게다가, 어댑터(402)는 자식 파티션 게스트 물리적 어드레스로부터 가상 머신 스토리지 서비스(602)에 할당된 시스템 물리적 어드레스로 데이터를 옮길 수 있다. 즉, 어댑터(402) 및 I/O MMU(426)는 게스트 물리적 어드레스로부터 시스템 물리적 어드레스로 전송 및 수신 베퍼를 모두를 변환하도록 구성될 수 있고, 어댑터(402)는 시스템 물리적 어드레스에 있어서 내부 전송 베퍼로부터 수신 베퍼로 데이터를 내부적으로 복사할 수 있다. 이후에, 가상 머신 스토리지 서비스(602)는 그 가상 하드 드라이브 구현에 부합한 적절한 장소에 그 데이터를 저장할 수 있다. 이것이 가상 하드 드라이브 파일의 사용을 포함하거나, LUN에 데이터를 저장하거나, 또는 데이터를 저장하기 위한 다른 기법 및 장소를 부가적으로 포함할 수 있음을 당업자라면 이해할 수 있다.

[0046] 도 9의 설명을 계속하면, 동작 912는 제 1 파티션에서 이전가능한 스토리지 서비스를 실행하고, 제 2 파티션에서 가상 머신을 관리하도록 구성된 관리 서비스를 실행하는 단계를 나타내고, 여기서 자식 파티션이 제 3 파티션이 된다. 예를 들어, 일 실시예에서 가상 머신 스토리지 서비스(602)는 자식 파티션(246)과 같은 제 1 파티션에서 실행될 수 있고, 부모 파티션(204)은 관리 서비스를 실행할 수 있으며, 가상 머신 스토리지 클라이언트(604)가 파티션(248)에서 실행될 수 있다. 이런 예시적인 실시예에서, 가상 머신 스토리지 서비스(602)는 관리 프로세스와는 별개의 파티션에 있다. 이런 구성에서, 자식 파티션(246)은 SAN 타겟처럼 기능하는 전용 스토리지 파티션으로 효과적으로 동작할 수 있다. 이런 구성은 하이퍼바이저(202) 및 부모 파티션의 부담을 덜어줄 수 있다. 예를 들어, 부모 파티션으로부터 스토리지 서비스를 분리시킴으로써, 운영 체제 내부 락(intra-operating system locking)이 줄어들 수 있다. 나아가, 이런 방식으로 컴퓨터 시스템을 구성함으로써, 파티션들 간에 전송되어야 할 메시지 수가 감소되어 하이퍼바이저 스케줄러에 대한 부담이 줄어든다.

[0047] 도 9의 설명을 계속하면, 동작 914는 이전가능한 스토리지 서비스를 고유 네트워크 식별자를 포함하는 네트워크 어댑터의 가상 기능과 연계시키고, 자식 파티션을 네트워크 어댑터의 제 2 가상 기능에 부착하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 일 실시예에서 가상 머신 스토리지 서비스(602)는 가상 기능, 예컨대, 가상 기능(404)과 연계될 수 있다. 가상 머신 스토리지 서비스(602)가 자식 파티션(246)에서 실행되는 일례에서, 가상 기능(404)은 제어된 방식으로, 즉, 자식 파티션(246)의 임의의 프로세스가 그 파티션 외부의 데이터에는 액세스하지 않도록 보장하는 방식으로 어댑터(402)에 액세스하도록 사용될 수 있다. 또한, 가상 머신 스냅샷 동작

이 가상 머신 스토리지 서비스(602)를 이전하는 데 사용될 수 있다.

[0048] 도 9의 설명을 계속하면, 동작 916은 자식 파티션으로부터의 입력/출력 작업 요청의 수신에 대한 응답으로 논리 프로세서에 통지를 전송하고, 논리 프로세서가 이전 가능한 스토리지 서비스를 실행하고 있음을 결정하는 단계를 나타낸다. 예를 들어, 일 실시예에서 I/O 작업이 소프트웨어 처리를 필요로 할 때, 하이퍼바이저(202)는 중단(interrupt)을 수신하고 실행할 수 있다. 하이퍼바이저(202)는 가상 머신 스토리지 서비스(602)를 실행하고 있거나 실행 예정인 논리 프로세서를 식별하고, 중단 또는 가벼운 통지를 전송함으로써 논리 프로세서에게 통지할 수 있다. 가상 머신 스토리지 서비스(602)가 자식 파티션에 위치하고 있으면, 메시지를 제공하기 위해 관리 파티션을 깨우지 않고도 논리 프로세서에 중단을 전송할 수 있다. 가상 머신 스토리지 서비스(602)가 현재 가상 머신 스토리지 서비스(602)에 대한 컨택스트 스위치를 실행하고 있으면, 중단이 일어날 필요가 없거나 일어나서는 안되고, 대신 가벼운 통지가 사용될 수 있다.

[0049] 도 9의 설명을 계속하면, 동작 918은 입/출력 트래픽이 네트워크 어댑터를 통해 고유 네트워크 식별자와 적어도 하나의 다른 고유 네트워크 식별자 사이에서 이동할 때 입/출력 트래픽이 보안 대책(security policy)을 따르는지를 결정하는 단계를 나타낸다. 예를 들어, 일 실시예에서, 어댑터(402)는 네트워크 트래픽에 관한 보안 대책을 포함할 수 있다. 이런 예시적인 실시예에서, 어댑터(402)는 가상 머신 스토리지 서비스(602)와 다른 고유 식별자, 예컨대, 가상 머신에 부착된 것 사이에서 전송되는 입/출력 트래픽이 보안 대책을 따르는지를 결정하도록 구성될 수 있다. 구체적인 예로, 보안 대책은 모든 입/출력 트래픽의 암호화를 요구할 수 있다. 이런 예에서, 어댑터(402)는 가상 하드 드라이브로의 쓰기가 평문(clear text)인지 또는 암호화되었는지를 결정하도록 구성될 수 있다. 다른 예를 들면, 보안 대책은 다른 가상 근거리 통신망들의 말단 사이에 어떤 데이터 트래픽도 허용되지 않게 가상 근거리 통신망들이 완전히 분리되도록 요구할 수도 있다.

[0050] 이제, 도 10을 참조하면, 이 도면은 본 발명의 양태들을 실시하기 위한, 동작들 1000, 1002, 및 1004를 포함하는 동작 절차를 도시한다. 동작 절차는 동작 1000에서 시작하고, 동작 1002는 네트워크 어댑터에 대한 제 1 고유 네트워크 식별자를 자식 파티션에 관한 가상 하드 드라이브 디스크 입/출력 요청을 관리하도록 구성되는 스토리지 서비스에 부착하는 단계를 나타낸다. 예를 들어, 도 6을 참조하면, 본 발명의 일 실시예에서 SR-IOV 어댑터(402)는 다수의 네트워크 식별자를 실현하고, 그들 중 하나를 가상 머신 스토리지 서비스(602)에 할당할 수 있다. 광 채널 예시에서, 광 채널 호스트 버스 어댑터는 NPIV(N-Port ID virtualization)을 사용하여 다수의 고유 식별자가 동일한 포트에 사용되도록 할 수 있다. 이런 광 채널 예시에서, 가상 머신 스토리지 서비스(602)는 패브릭 상에서 통신하기 위해 할당된 NPIV를 배타적으로 사용할 수 있다.

[0051] 도 10의 설명을 계속하면, 동작 1004는 네트워크 어댑터에 의해 실현되는 가상 기능을 자식 파티션에 부착하는 단계를 나타내며, 여기서 가상 기능은 제 2 고유 네트워크 식별자를 포함한다. 예를 들어, 다시 도 6을 참조하면, SR-IOV 어댑터(402)는 고유 네트워크 식별자를 포함하는 가상 기능(406)을 인스턴스화하고, 이를 가상 머신에 부착할 수 있다. 이런 예시적인 실시예에서, 어댑터(402)는 하이퍼바이저(202) 또는 별개의 파티션-투-파티션 통신 메커니즘을 우회함으로써, I/O 요청을 어댑터를 통해서 스토리지 서비스(602)로 라우팅하는 스위치로서 기능하도록 구성된다. 이는 결국 파티션들에 통지하고 이들을 스위칭하기 위해 논리 프로세서에서 명령어를 실행하는 데 걸리는 시간을 줄이게 된다.

[0052] 이제 도 11을 참조하면, 도면은 추가 동작들 1106, 0018, 1110, 1112 및 1114를 포함하는, 도 10의 동작 절차의 대안적인 실시예를 도시한다. 동작 1106은 제 2 가상 기능이 제 1 고유 네트워크 식별자를 포함하도록 구성하기 위한 요청을 제 2 네트워크 어댑터를 포함하는 원격 컴퓨터 시스템에 전송하는 단계를 나타낸다. 예를 들어, 일 실시예에서 논리 프로세서가 관리자(250)에서 명령어를 실행하고 있을 수 있으며, 다른 어댑터를 가지고 있는 원격 시스템에서의 가상 기능이 가상 머신 스토리지 서비스(602)에 부착된 고유 네트워크 식별자를 포함하도록 구성하기 위한 요청을 생성할 수 있다. 도 7을 참조하면, 구체적인 예에서, 컴퓨터 시스템(700)의 관리자(250)는 어댑터(718)를 가지고 있는 컴퓨터 시스템(702)으로 생성된 요청을 전송할 수 있다. 이런 예시에서 컴퓨터 시스템(702)의 관리자(250)는 이 요청을 사용하여, 가상 기능(710)을 인스턴스화하고 가상 기능(710)이 가상 머신 스토리지 서비스(602)의 인스턴스에 연관된 고유 식별자를 포함하도록 지시할 수 있다.

[0053] 도 11의 설명을 계속하면, 동작 1108은 스토리지 서비스를 자식 파티션에 이전하고, 자식 파티션에 할당된 제 2 가상 기능이 제 1 고유 네트워크 식별자를 사용하도록 구성하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 논리 프로세서는 관리자(250)를 실행하고, 예컨대, 부모 파티션(204)로부터 자식 파티션(246)으로 가상 머신 스토리지 서비스(602)를 이전할 수 있다. 이런 예에서, 논리 프로세서는 관리자(250)를 실행하고, 가상 머신 스토리지 서비스(602)에 연관된 고유 식별자를 추출하여, 이를 어댑터(403)에 전송할 수 있다. 어댑터(402)는 가상

기능(404)을 인스턴스화할 수 있고, 여기에 고유 식별자를 부착할 수 있다. 그 이후에, 관리자(250)는 가상 머신 스토리지 서비스(602)의 인스턴스에 고유 식별자를 부착할 수 있다. 이런 예시적인 실시예에서, 가상 머신 스토리지 서비스(602)는 관리자 프로세스와는 별개의 파티션에 있고, 효과적으로 iSCSI 타겟처럼 기능하는 전용 스토리지 파티션이 될 수 있다.

[0054] 이제 동작 1110을 참조하면, 이 동작은 입/출력 메모리 관리 유닛이 자식 파티션으로부터의 입/출력 요청에 관련된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 본 발명의 일 실시예에서 컴퓨터 시스템(700)의 입/출력 메모리 관리 유닛(426)이 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는 데 사용될 수 있다. 예를 들어, 게스트 운영 체제(220)가 I/O 동작, 예컨대, 읽기 또는 쓰기를 시작할 때, 게스트 운영 시스템(220)은 게스트 물리적 어드레스를 포함하는 명령어를 생성한다. 이런 예에서, 입/출력 메모리 관리 유닛(426)은 자식 파티션(246)의 게스트 메모리 어드레스를 부모 파티션(204)이 사용하는 시스템 어드레스로 매핑시키는 테이블을 사용할 수 있다. 어댑터(402) 및 I/O MMU(426)는 게스트 물리적 어드레스로부터 시스템 물리적 어드레스로 전송 및 수신 버퍼를 모두를 변환하도록 구성될 수 있고, 어댑터(402)는 내부 전송 버퍼로부터 내부 수신 버퍼로 또는 그 역으로 데이터를 복사할 수 있다.

[0055] 이제 동작 1112를 참조하면, 이 동작은 입/출력 트래픽이 고유 네트워크 식별자와 제 2 고유 네트워크 식별자 사이에서 이동할 때, 네트워크 어댑터가 입/출력 트래픽의 보안 대책 준수 여부를 모니터링하도록 구성하는 단계를 나타낸다. 예를 들어, 일 실시예에서, 어댑터(402)는 네트워크 트래픽에 관한 보안 대책을 포함할 수 있다. 이런 예시적인 실시예에서, 어댑터(402)는 가상 머신 스토리지 서비스(602) 및 다른 고유 식별자, 예컨대, 가상 머신에 부착된 것과의 사이에서 전송되는 입/출력 트래픽이 보안 대책을 따르는지를 결정하도록 구성될 수 있다. 구체적인 예로, 특정 가상 머신들이 네트워크의 특정 고유 식별자를 사용하여 I/O를 전송하도록 요청하는 보안 대책을 들 수 있다. 이런 예에서, 어댑터(402)는 가상 머신들로부터의 정보 패킷을 모니터링하여, 이들이 보안 대책을 따르고 있는지를 결정할 수 있다.

[0056] 이제 동작 1114를 참조하면, 이 동작은 원격 컴퓨터 시스템으로부터 사전 결정된 임계치를 넘는 입/출력 요청량의 수신되었는지의 판단 여부에 따라, 스토리지 서비스를 인스턴스화하고 제 1 고유 네트워크 식별자를 스토리지 서비스에 할당하기 위한 요청을 원격 컴퓨터 시스템으로 전송하는 단계를 나타낸다. 예를 들어, 본 발명의 일 실시예에서 논리 프로세서는 관리자(250)를 나타내는 명령어를 실행하고, 컴퓨터 시스템(702)과 같은 원격 컴퓨터 시스템이 가상 머신 스토리지 서비스(602)의 인스턴스를 인스턴스화하고 이를 고유 식별자에 부착하게 하기 위한 요청을 전송할 수 있다. 논리 프로세서는 들어오는 I/O 요청에 관련된 고유 식별자들을 모니터링하고 컴퓨터 시스템(702)으로부터 임계치를 넘는 요청이 수신되었는지를 판단한 후에 이런 요청을 생성할 수 있다. 구체적인 예시에서, 관리자(250)는 지난 30분 동안 I/O 요청의 60%가 컴퓨터 시스템(702)과 현재 연관된 고유 식별자들로부터 수신되었다고 판단했을 수 있다. 이 경우, 관리자(250)는, 가상 머신 스토리지 서비스(602)가 컴퓨터 시스템(702)에서 지역적으로 실행된다면 데이터센터의 성능이 향상될 수 있다고 판단하고, 서비스를 이전할 수 있다.

[0057] 이제 도 12를 참조하면, 이 도면은 동작 1200 및 1202를 포함하는 동작 절차를 도시한다. 동작 절차는 동작 1200에서 시작하고, 동작 1202는 자식 파티션에서 스토리지 서비스를 실행하는 단계를 나타내며, 여기서 저장 장치는 제 2 자식 파티션에 대한 가상 하드 드라이브 디스크 입/출력 요청을 관리하도록 구성되고, 스토리지 서비스에 네트워크의 고유 네트워크 식별자가 할당된다. 예를 들어, 일 실시예에서, 가상 머신 스토리지 서비스(602)는 자식 파티션, 예컨대, 자식 파티션(246)에서 실현될 수 있고, 네트워크의 고유 식별자, 예컨대, 월드 와이드 네임이 할당될 수 있다. 이런 예시적인 실시예에서 자식 파티션(246)은 하이퍼바이저(202) 및/또는 부모 파티션(204)에 의해 제어될 수 있다. 이런 구성에서, 자식 파티션(246)은 효과적으로 iSCSI 타겟처럼 기능하는 전용 스토리지 파티션이 될 수 있다.

[0058] 이제 도 13을 참조하면, 이 도면은 동작들 1304, 1306, 1308, 1310 및 1312를 포함하는, 도 12의 동작 절차의 대안적인 실시예를 도시한다. 동작 1304를 참조하면, 이 동작은 원격 컴퓨터 시스템으로부터 사전 결정된 임계치를 넘는 입/출력 요청량의 수신되었는지의 판단 여부에 따라, 스토리지 서비스를 인스턴스화하고 제 1 고유 네트워크 식별자를 스토리지 서비스에 할당하기 위한 요청을 원격 컴퓨터 시스템으로 전송하는 단계를 나타낸다. 예를 들어, 본 발명의 일 실시예에서 논리 프로세서는 관리자(250)를 나타내는 명령어를 실행하고, 컴퓨터 시스템(702)과 같은 원격 컴퓨터 시스템이 가상 머신 스토리지 서비스(602)의 인스턴스를 인스턴스화하고 이를 고유 식별자에 부착하게 하기 위한 요청을 전송할 수 있다. 논리 프로세서는 들어오는 I/O 요청에 관련된 고유 식별자들을 모니터링하고 컴퓨터 시스템(702)으로부터 임계치를 넘는 요청이 수신되었는지를 판단한 후에 이런 요청을 생성할 수 있다. 구체적인 예시에서, 관리자(250)는 지난 30분 동안 I/O 요청의 60%가 컴퓨터

시스템(702)과 현재 연관된 고유 식별자들로부터 수신되었다고 판단했을 수 있다. 이 경우, 관리자(250)는, 가상 머신 스토리지 서비스(602)가 컴퓨터 시스템(702)에서 지역적으로 실행된다면 데이터센터의 성능이 향상될 수 있다고 판단하고, 서비스를 이전할 수 있다.

[0059] 도 13의 설명을 계속하면, 동작 1306은 스토리지 서비스를 하이퍼바이저에 이전하는 단계를 도시한다. 예를 들어, 도 7을 참조하면, 일 실시예에서 가상 머신 스토리지 서비스(602)가 하이퍼바이저(202)로 이전될 수 있다. 이런 예시적인 실시예에서, 컴퓨터 시스템(702)은 도 3에 도시된 것과 유사한 아키텍처를 가질 수 있으며, 자식 파티션(246)에서 하이퍼바이저(202)로 스토리지 서비스(602)를 이동시키는 결정이 내려질 수 있다. 이런 예에서, 논리 프로세서는 관리자(250)를 실행하고, 가상 머신 스토리지 서비스(602)와 연관된 고유 식별자를 추출할 수 있으며, 하이퍼바이저(202)는 이를 가상 머신 스토리지 서비스(602)의 인스턴스에 부착시킬 수 있다. 예시적인 실시예에서, 하이퍼바이저(202)는 하드웨어를 제어할 수 있으므로, 하이퍼바이저(202)는 어댑터(402)의 물리적 기능에 액세스하도록 구성될 수 있다. 광채널 구현예에서, 광채널 호스트 버스 제어기는 고유 식별자를 사용하여 어댑터(402)를 통해 I/O 명령어를 전송/수신하기 위해 NPIV를 사용할 수 있다.

[0060] 도 13의 설명을 계속하면, 동작 1308은 스토리지 서비스를 부모 파티션에 이전하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 일 실시예에서 가상 머신 스토리지 서비스(602)는 자식 파티션(246)에서 부모 파티션(204 또는 712)으로 이전될 수 있다. 이런 예에서, 논리 프로세서는 관리자(250)를 실행하고, 가상 머신 스토리지 서비스(602)와 연관된 고유 식별자를 추출하여, 이를 원격 컴퓨터 또는 로컬 컴퓨터 시스템 상의 부모 파티션(204) 둘 중 하나에 전송할 수 있다. 이후에, 고유 식별자가 스토리지 서비스(602)의 인스턴스에 부착될 수 있다.

[0061] 도 13의 설명을 계속하면, 동작 1310은 하이퍼바이저에 스토리지 서비스를 할당하는 단계를 도시한다. 이런 예시적인 실시예에서, 가상 기능(406)은 자식 파티션(248)에 부착될 수 있고, 네트워크 상의 제 2 고유 네트워크 식별자를 가질 수 있다. 도면에 도시된 바와 같이, 이런 예시적인 실시예에서, 자식 파티션들(246 및 248) 모두는 동일한 SR-IOV 어댑터(402)에 부착될 수 있다. 따라서, 이런 예시적인 실시예에서 I/O 요청들은 하이퍼바이저(202)나 파티션-투-파티션 통신 메커니즘을 통하는 대신 SR-IOV 어댑터(402)를 통해, 또한 스위치(704)에 I/O를 전송할 필요 없이 전달될 수 있다.

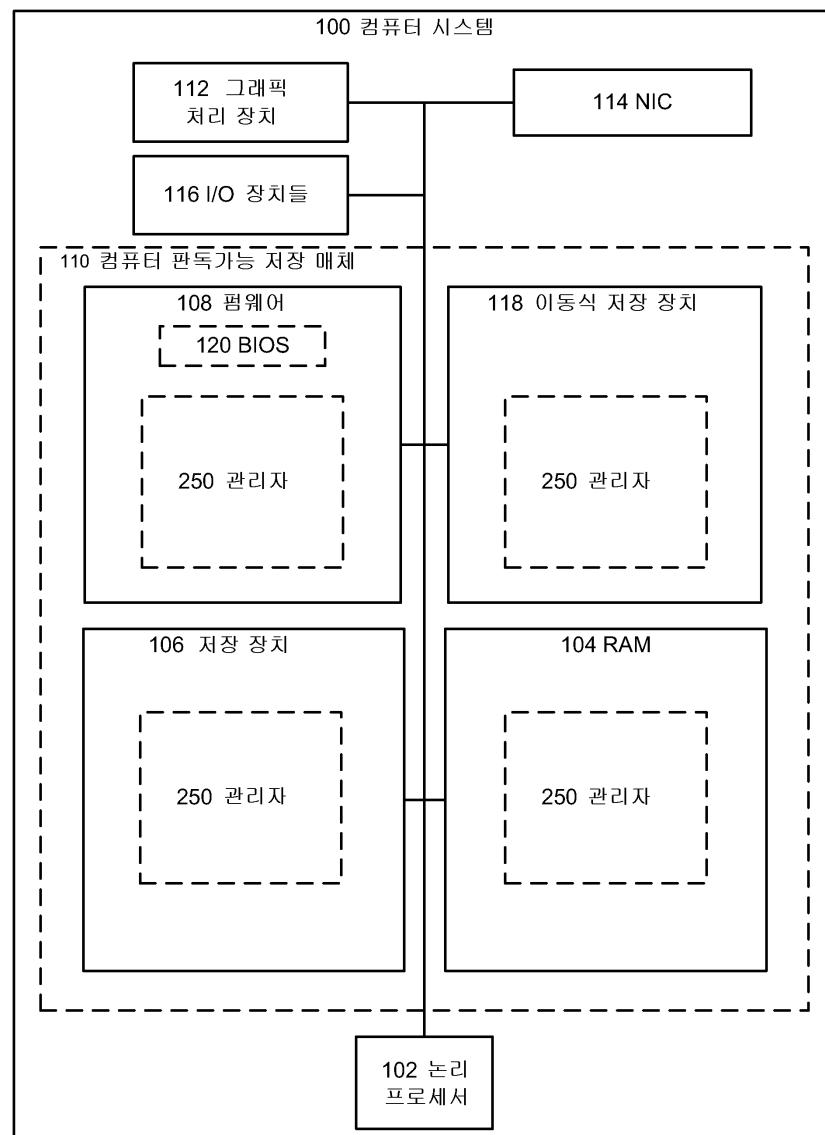
[0062] 도 13의 설명을 계속하면, 동작 1312는 자식 파티션에 대한 입/출력 요청에 관련된 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하도록 입/출력 메모리 관리 유닛을 구성하는 단계를 나타낸다. 예를 들어, 도 7을 참조하면, 본 발명의 일 실시예에서 컴퓨터 시스템(700)의 입/출력 메모리 관리 유닛(426)은 게스트 물리적 어드레스를 시스템 물리적 어드레스로 변환하는 데 사용될 수 있다. 예를 들어, 게스트 운영 체제(220)가 I/O 동작, 예컨대, 읽기 또는 쓰기를 시작할 때, 게스트 운영 체제(220)는 게스트 물리적 어드레스를 포함하는 명령어를 생성한다. 이런 예에서, 입/출력 메모리 관리 유닛(426)은 자식 파티션(248)의 게스트 메모리 어드레스를 부모 파티션이 사용하는 시스템 어드레스로 매핑시키는 테이블을 사용할 수 있다. 어댑터(402) 및 I/O MMU(426)는 게스트 물리적 어드레스로부터 시스템 물리적 어드레스로 전송 및 수신 버퍼를 모두를 변환하도록 구성될 수 있고, 어댑터(402)는 내부 전송 버퍼로부터 내부 수신 버퍼로 또는 그 역으로 데이터를 복사할 수 있다.

[0063] 전술한 상세한 설명은 예시 및/또는 동작 도면들을 통해 시스템 및/또는 프로세스의 다양한 실시예들을 설명하였다. 그러한 블록도들 및/또는 예들이 하나 이상의 기능 및/또는 동작을 포함하는 한, 이러한 블록도들 또는 예들 내의 각각의 기능 및/또는 동작은 광범위한 하드웨어, 소프트웨어, 편웨어 또는 사실상 이들의 임의의 조합에 의해 개별적으로 그리고/또는 공동으로 구현될 수 있다는 것을 당업자라면 이해할 것이다.

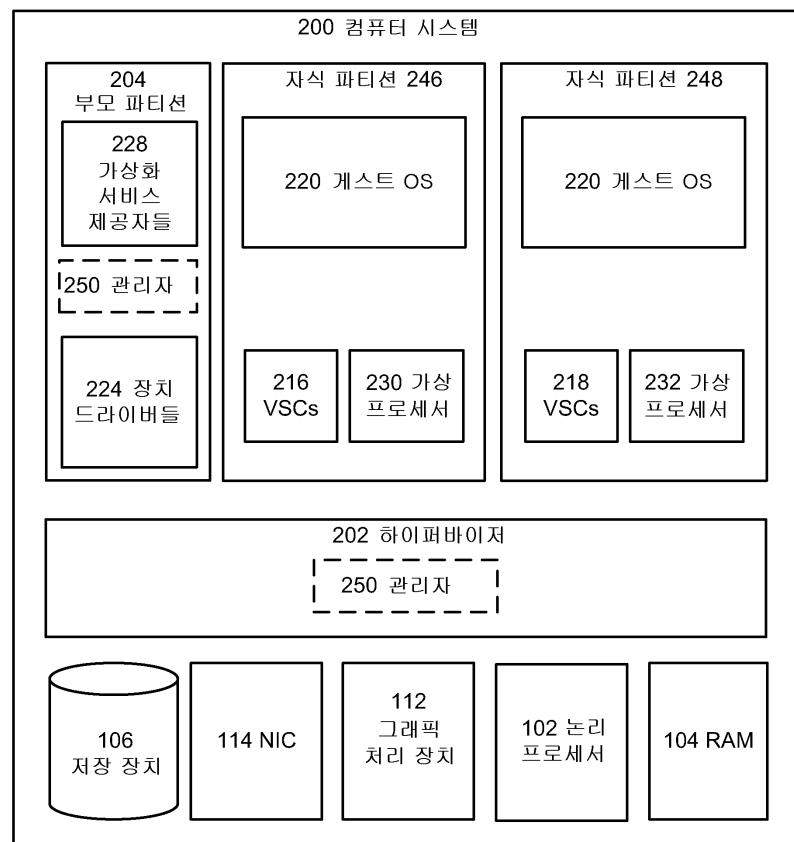
[0064] 본원에 기술된 본 발명의 특정 양태들이 도시되고 설명되었지만, 본원에서 설명된 본 발명 및 그보다 더 넓은 양태들로부터 벗어나지 않고, 본원의 개시에 기반하여, 변경들 및 수정이 이루어질 수 있으며, 따라서 첨부된 특허청구범위는 본원에서 설명된 본 발명의 진정한 사상 및 범위 내에 있는 바와 같은 모든 그러한 변경 및 수정을 그 범위 내에 포함하여야 한다는 것이 당업자에게는 자명할 것이다.

도면

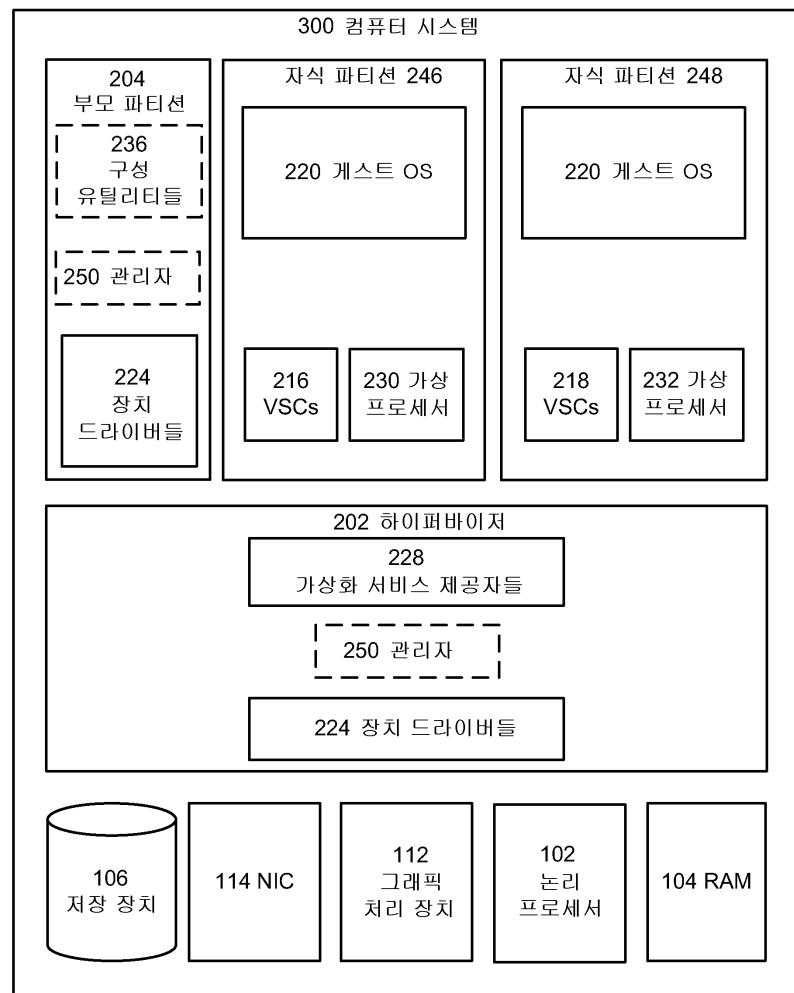
도면1



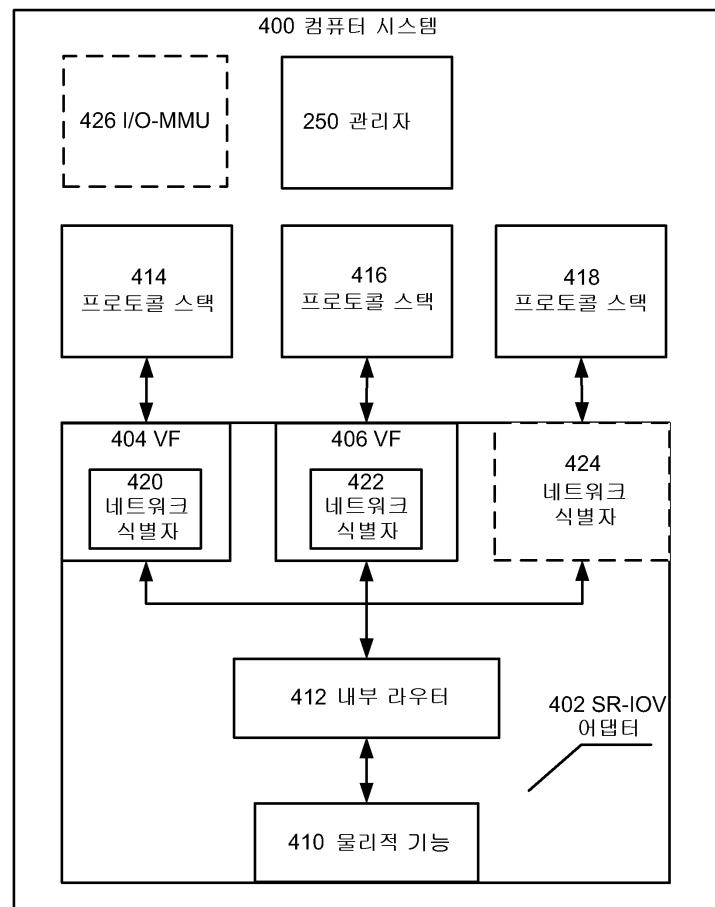
도면2



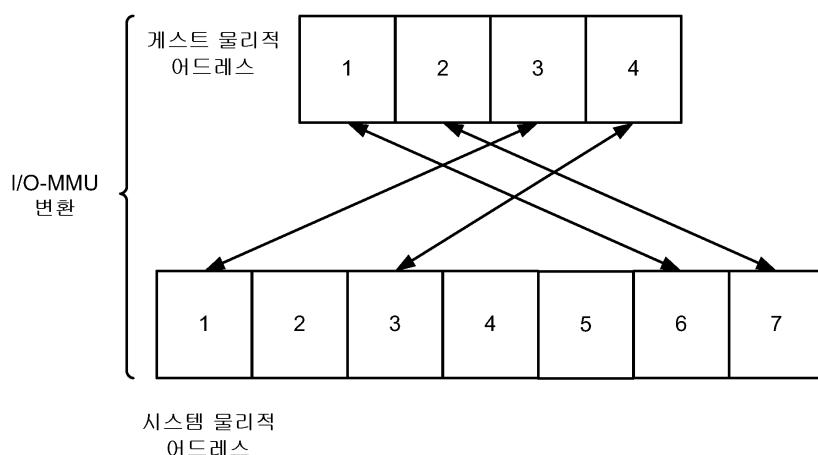
도면3



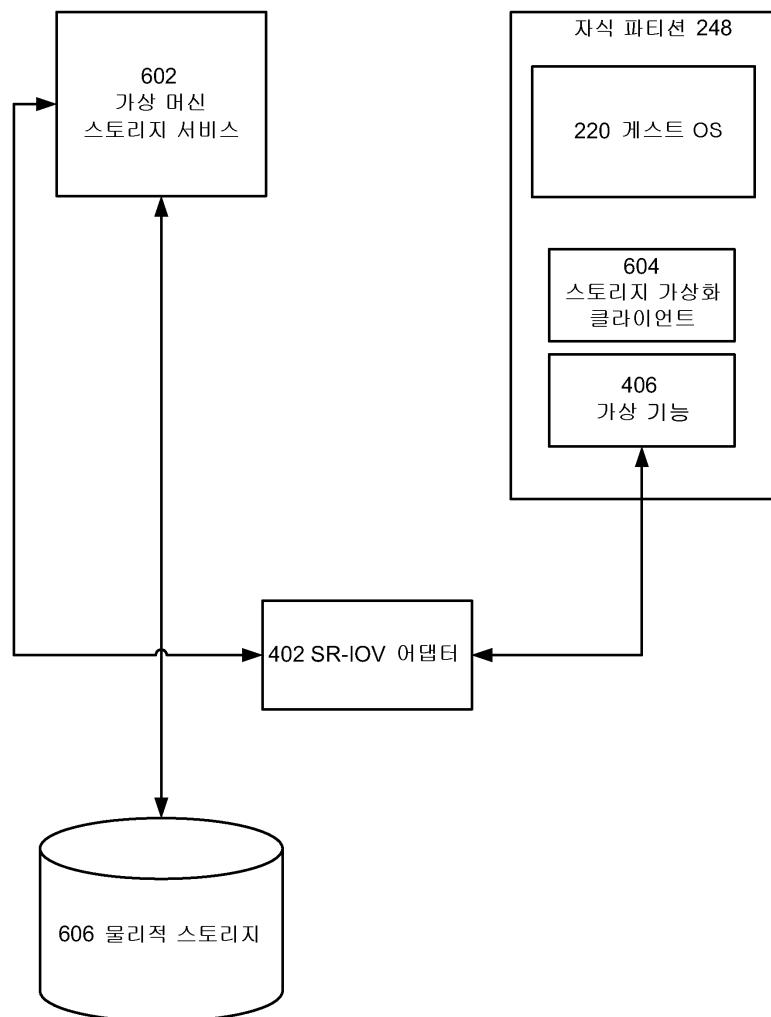
도면4



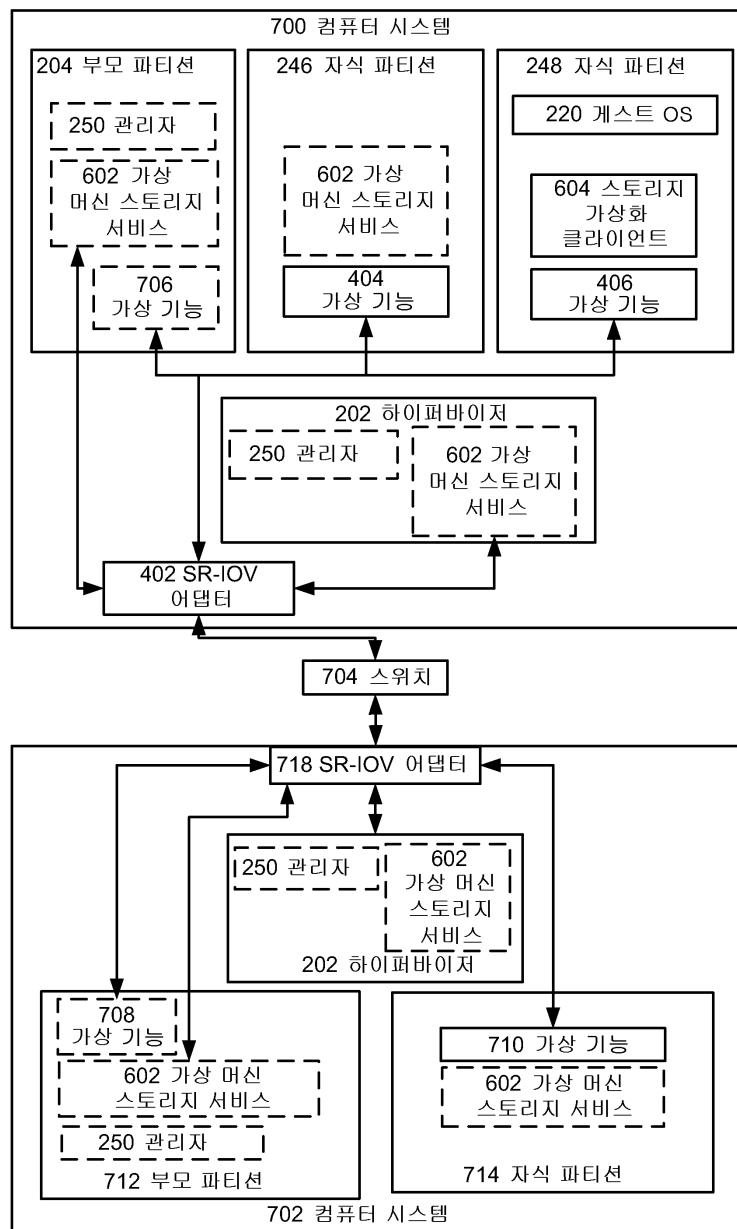
도면5



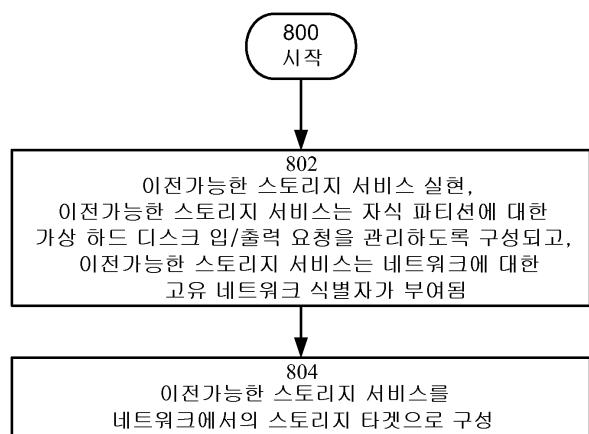
도면6



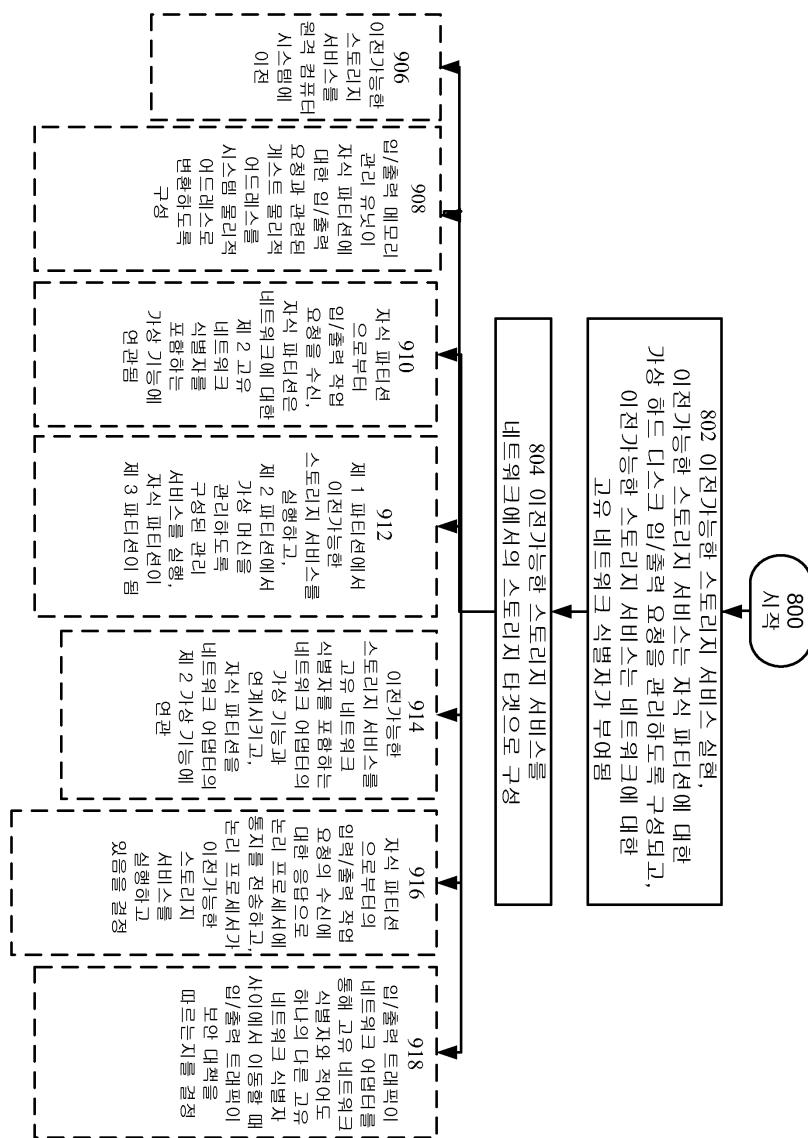
도면7



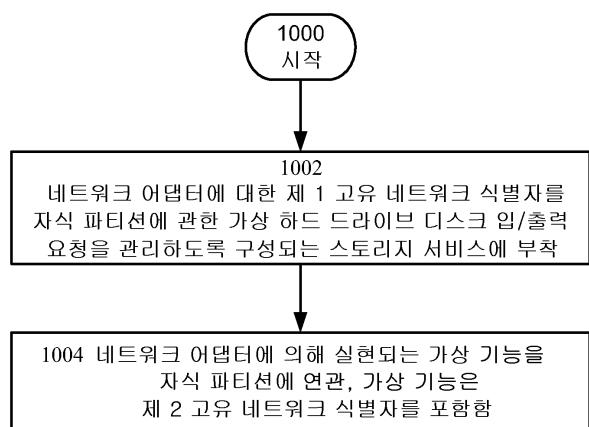
도면8



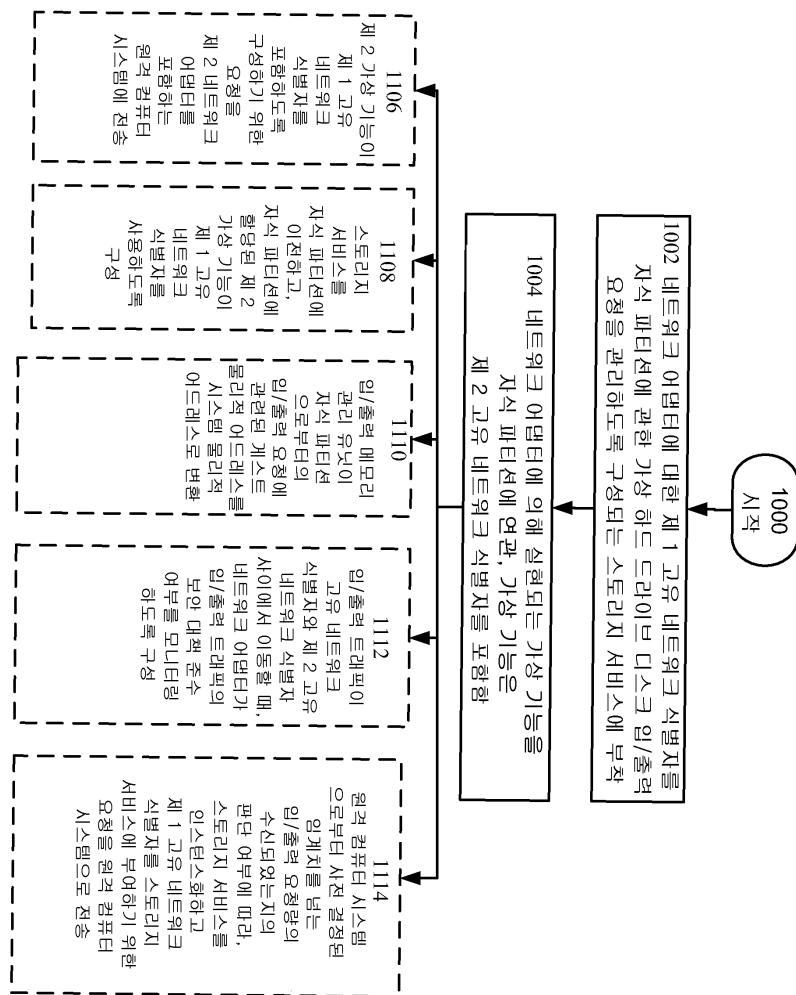
도면9



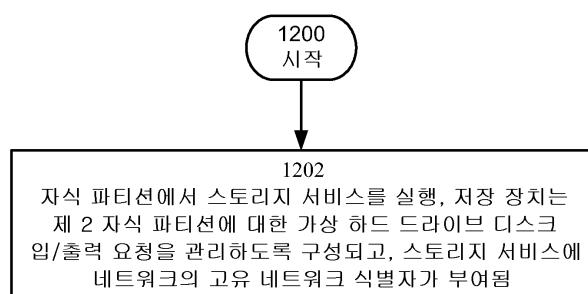
도면10



도면11



도면12



도면13

