(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0107281
(43) 공개일자 2013년10월01일

(51) 국제특허분류(Int. Cl.)
H04L 29/08 (2006.01)
(21) 출원번호 10-2013-7006739
(22) 출원일자(국제) 2011년09월20일
심사청구일자 없음
(85) 번역문제출일자 2013년03월15일
(86) 국제출원번호 PCT/EP2011/066256
(87) 국제공개번호 WO 2012/038385
국제공개일자 2012년03월29일
(30) 우선권주장
10305999.4 2010년09월20일
유럽특허청(EPO)(EP)

(71) 출원인
톰슨 라이센싱
프랑스 92130 이씨레뮬리노 루 잔다르크 1-5
(72) 발명자
반 켐펜 알렉산더
프랑스, 쎄쏜 세비뉴 에프-35576, 씨에스 17616, 쟈크 드 쌍 블랑, 아브뉘 드 쌍 블랑, 테크니컬러 알&디 프랑스 975
르 머러 에르완
프랑스, 쎄쏜 세비뉴 에프-35576, 씨에스 17616, 쟈크 드 쌍 블랑, 아브뉘 드 쌍 블랑, 테크니컬러 알&디 프랑스 975
(뒷면에 계속)
(74) 대리인
문경진, 김학수

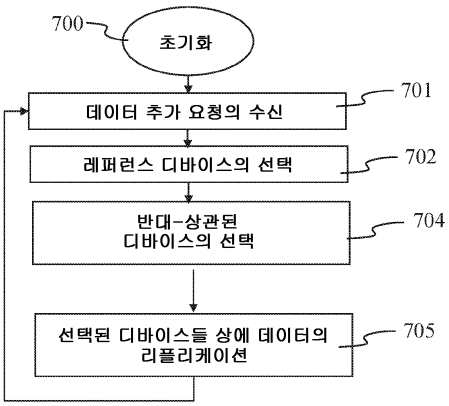
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법 및 대응하는 디바이스**

(57) 요 약

본 발명은 일반적으로 분산된 데이터 스토리지 시스템들에 관한 것이다. 특히, 본 발명은 네트워크 노드들 사이의 데이터의 교환을 위해 필요한 대역폭, 및 데이터의 항목을 저장하기 위해 필요한 네트워크 노드들의 개수에 따라 네트워크에 작은 영향을 주는 큰 데이터 이용도와 데이터 스토리지 리소스들을 결합시키는 분산된 데이터 스토리지 시스템 내에서의 데이터 배치의 방법에 관한 것이다.

대 표 도 - 도7



(72) 발명자

스트로브 질

프랑스, 쎄쏜 세비뉴 에프-35576, 씨에스 17616, 쟈크 드 쌍 블랑, 아브뉘 드 쌍 블랑, 테크니컬러 알&디 프랑스 975

커마렉 안네-마리

프랑스, 렌 세덱스 에프-35042, 캠퍼스 유니버시떼어 드 뷸리우, 아이엔알아이에이 - 인스티튯 내셔널 드 러셰어셰이 앙 인포매띠끄 에 앙 오토매띠끄

특허청구의 범위

청구항 1

스토리지 디바이스들로서 적어도 사용되는 네트워크 디바이스들을 포함하는 분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법에 있어서,

상기 분산된 데이터 스토리지 시스템에 데이터 항목을 저장하기 위한 요청의 수신 단계(701);

레퍼런스 디바이스로서 제1 네트워크 디바이스의 제1 선택 단계(40), 및 레퍼런스 디바이스의 시간에 따른 이용도(availability) 및 비이용도(unavailability)의 결정 단계;

적어도 하나의 제2 네트워크 디바이스의 제2 선택 단계(41, 42)로서, 상기 레퍼런스 디바이스의 시간에 따른 이용도에 대한 상기 적어도 하나의 제2 네트워크 디바이스의 시간에 따른 이용도의 대응(correspondence)의 함수로서 적어도 하나의 제2 네트워크 디바이스의 제2 선택 단계(41, 42);

적어도 하나의 제3 네트워크 디바이스의 제3 선택 단계(43, 44)로서, 레퍼런스 디바이스의 시간에 따른 비이용도에 대한 상기 적어도 하나의 제3 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제3 네트워크 디바이스의 제3 선택 단계(43, 44); 및

상기 제2 및 제3 선택 단계들에서 선택된 적어도 하나의 제2 및 적어도 하나의 제3 네트워크 디바이스에 상기 데이터 항목의 저장 단계(705);를 포함하는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 2

제1항에 있어서,

k는 상기 데이터 항목이 리플리케이팅되는(replicated) 네트워크 디바이스들의 개수인데, 여기서 k는 상기 데이터 항목을 저장하기 위한 상기 요청에 명시되고, 상기 제1, 제2, 및 제3 선택 단계들은 상기 데이터 항목이 적어도 k개의 네트워크 디바이스들을 통해 리플리케이팅될 때까지 반복되며, 레퍼런스 디바이스의 상기 제1 선택 단계는 상기 방법의 이전 반복 시 레퍼런스 디바이스로서 이미 선택된 네트워크 디바이스의 선택을 배제시키는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 3

제1항에 있어서,

k는 상기 데이터 항목이 리플리케이팅되는 네트워크 디바이스들의 개수인데, 여기서 k는 상기 데이터 항목을 저장하기 위한 상기 요청에 명시되고, 상기 제2 및 제3 선택 단계들은 상기 데이터 항목이 적어도 k개의 네트워크 디바이스들을 통해 리플리케이팅될 때까지 반복되는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 4

제2항 또는 제3항에 있어서,

상기 비이용도는 각 원소(element)가 시간의 단위(unit of time)를 나타내는 벡터(24)로서 저장되고, 상기 벡터(24)의 원소들의 개수가 시간 스팬(time span)을 나타내고, 상기 벡터(24)가 관련된 네트워크 디바이스의 상기 시간의 단위 동안의 이용도를 각 원소 값이 나타내는 경우, 제1 미리 결정된 원소 값은 이용도를 나타내고, 제2 미리 결정된 원소 값은 비이용도를 나타내는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 5

제4항에 있어서,

상기 제1 미리 결정된 원소 값은 양(positive)의 미리 결정된 값이고, 상기 제2 미리 결정된 원소 값은 음(negative)의 미리 결정된 값이며, 상기 레퍼런스 디바이스의 상기 비이용도의 함수로서 상기 제2 네트워크 디바이스의 상기 선택은 두 개의 상기 벡터들(24) 사이의 각도(angle)의 계산을 통해 결정된 반대-상관(anti-correlation)의 레벨에 따라 결정되고, 두 개의 벡터들(24)은 상기 각도가 0에 근접할 때 크게 상관되고, 두 개의 벡터들(24)은 상기 각도가 π에 근접할 때 크게 반대-상관되며, 상기 각도의 값은 상기 반대-상관의 레벨을 나타내는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 6

제5항에 있어서,

상기 각도는

$$\theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \cdot ||\vec{y}||}\right)$$

에 따라 계산되며, 여기서 Θ는 상기 각도이고, x와 y는 상기 두 개의 벡터들(24)인,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 7

제4항에 있어서,

상기 제1 미리 결정된 원소 값은 이진 수 1이고, 상기 제2 미리 결정된 값은 이진 수 0이며, 반대-상관의 레벨은 두 개의 상기 벡터들(24) 사이의 로직 부울리안(logical Boolean) XOR 연산의 계산을 통해 결정되고, 두 개의 벡터들(24)은 상기 로직 부울리안 XOR 연산의 결과가 상기 벡터들(24) 각각의 상기 원소들의 개수에 근접하는 1들의 큰 개수를 포함할 때, 크게 반대-상관되며, 두 개의 벡터들(24)은 상기 로직 부울리안 XOR 연산의 상기 결과가 0에 근접하는 1들의 개수를 포함할 때, 크게 상관되고, 상기 부울리안 XOR 연산의 상기 결과에서 1들의 개수는 반대-상관의 상기 레벨을 나타내는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 8

제2항 내지 제7항 중 어느 한 항에 있어서,

상기 레퍼런스 디바이스는 상기 네트워크 디바이스들로부터 랜덤하게 선택되는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 9

제2항 내지 제7항 중 어느 한 항에 있어서,

상기 레퍼런스 디바이스는 상기 네트워크 디바이스들로부터 결정론적으로 선택되는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법.

청구항 10

스토리지 디바이스들로서 적어도 사용되는 네트워크 디바이스들을 포함하는 분산된 데이터 스토리지 시스템에 데이터를 저장하기 위한 디바이스에 있어서,

상기 분산된 데이터 스토리지 시스템에 데이터 항목을 저장하기 위한 요청을 수신하기 위한 수신기(640);

레퍼런스 네트워크 디바이스로서 제1 네트워크 디바이스의 제1 선택(40)을 위한 선택기(620), 및 상기 레퍼런스

디바이스의 시간에 따른 이용도 및 비이용도의 결정을 위한 수단;

적어도 하나의 제2 네트워크 디바이스의 제2 선택(41, 42)을 위한 선택기(620)로서, 상기 레퍼런스 네트워크 디바이스의 시간에 따른 이용도에 대한 상기 제2 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제2 네트워크 디바이스의 제2 선택(41, 42)을 위한 선택기(620);

적어도 하나의 제3 네트워크 디바이스의 제3 선택(43, 44)을 위한 선택기(620)로서, 레퍼런스 네트워크 디바이스의 시간에 따른 비이용도에 대한 상기 적어도 하나의 제3 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제3 네트워크 디바이스의 제3 선택(43, 44)을 위한 선택기(620); 및

제2 선택을 위한 상기 선택기 및 상기 제3 선택을 위한 상기 선택기에 의해 선택된 네트워크 디바이스들에 상기 데이터 항목을 저장하기 위한 스토리지 수단(600, 610);을 포함하는 것을 특징으로 하는,

분산된 데이터 스토리지 시스템에 데이터를 저장하기 위한 디바이스.

명 세 서

기 술 분 야

[0001] 본 발명은 일반적으로 분산된 데이터 스토리지 시스템들에 관한 것이다. 특히, 본 발명은 네트워크 노드들 사이의 데이터의 교환을 위해 필요한 대역폭, 및 데이터의 항목을 저장하기 위해 필요한 네트워크 노드들의 개수에 따라 네트워크에 작은 영향을 주는 큰 데이터 이용도(availability)와 데이터 스토리지 리소스들을 결합시키는 분산된 데이터 스토리지 시스템 내에서의 데이터 배치(data placement)의 방법에 관한 것이다.

배 경 기 술

[0002] 비디오 및 이미지 처리 디바이스들과 같은 대용량 데이터 처리 디바이스들의 신속히 퍼지는 전개에 따라, 직접적인 스토리지를 위해, 또는 백업 스토리지의 부분으로서 막대한 양의 데이터에 대한 신뢰할 수 있는 스토리지가 요구된다. 보다 많은 디바이스들에 네트워크 연결성(network connectivity)이 제공됨에 따라, 네트워크 연결된 디바이스들 내에서의 데이터에 대한 분산된 스토리지는 비용 효율적인 해결책으로 간주된다. 인터넷과 같은 비-관리형 네트워크들(non-managed networks)을 통해 전개될 수 있는 분산된 데이터 스토리지 시스템들에서, 데이터 이용도 및 데이터 손실에 대한 복구력(resilience)을 보장하도록 데이터의 동일한 항목을 다수의 네트워크로 연결된 디바이스들에 복사하는 방법들이 개발되었다. 이는 데이터 리플리케이션(data replication)이라고 일컬어진다. 데이터 리플리케이션은 폭넓은 의미에서 받아들여져야 하며, 단순한 데이터 듀플리케이션(data duplication) 뿐만 아니라, 삭제(erasure) 및 (인코딩된 데이터가 복구를 위해 스토리지 디바이스들에 위치되는) 재생 코드들(regenerating codes)과 같은 코딩 기술들의 사용을 포괄한다. 디바이스 고장으로 인한 영구적인 데이터 손실 또는 일시적인 디바이스 비이용도(unavailability)에 따른 일시적인 데이터 손실의 위험에 대처하기 위해, 큰 리플리케이션 팩터{즉, 큰 수의 복사들(a high number of copies)}가 필요하다. 하지만, 통신 및 필요한 스토리지 사이즈에 관한 비용들(소위, 리플리케이션 비용들)을 감소시키기 위해, 작은 리플리케이션를 갖는 것이 오히려 바람직하다.

발명의 내용

해결하려는 과제

[0003] 필요한 것은 이용도 필수 사항들(availability requirements) 및 리플리케이션 비용들을 공동으로 고려하는 분산된 데이터 스토리지에 대한 높은 레벨의 데이터 이용도를 달성하는 해결책이다.

과제의 해결 수단

[0004] 본 발명은 종래의 기술의 불편한 점들의 일부를 경감시키는 것을 목적으로 한다.

[0005] 피어-투-피어(peer-to-peer) 네트워크들과 같은 대규모의 분산된 데이터 네트워크들에서, 디바이스들은 연속적으로 네트워크에 합류하고 이를 탈회한다. 각 디바이스는 자체적인 연결 해제 및 연결 동작을 가지고 있다. 예를 들어, 특정의 디바이스들은 항상 연결되지만, 다른 디바이스들은 낮 동안에만 연결되고 밤에는 연결이 해제되며, 그 밖의 디바이스들은 보다 랜덤한 연결 동작을 가지고 있다.

[0006] 인터넷 연결 게이트웨이들에 기초한 피어-투-피어 네트워크들에 대해, 텔레커뮤니케이션 작동자들에 의해 이들

서비스들의 가입자들에게 제공되는 게이트웨이들은 피어들(peers)로 간주되며, 가입자들은 상기 피어들을 밤중에 또는 부재 시에 마음대로 스위치 오프(swich off)한다.

[0007] 분산된 데이터 스토리지 시스템에서 피어 또는 디바이스가 스위치 오프되거나 또는 고장이 나면, 이들이 저장한 데이터는 더 이상 이용 가능하지 않다. 데이터의 특정 항목을 저장하는 모든 디바이스들이 스위치 오프되거나 또는 고장이 나는 경우, 데이터 항목은 더 이상 이용 가능하지 않으며, 데이터 항목은 적어도 스위치 오프된 디바이스들이 다시 스위치 온되거나, 또는 각각 수리되는 시간 동안 손실된 것으로 간주된다. 따라서, 상기 시간 동안, 저장된 데이터의 이용도는 보장되지 않는다. 데이터 리플리케이션 팩터를 증가시키는 것은 이러한 문제에 대한 가능한 해결책이지만, 데이터 통신 및 필요한 데이터 스토리지 사이즈에 관한 데이터 스토리지 비용들에 매우 영향을 끼친다.

[0008] 본 발명의 목적은 분산된 데이터 스토리지 시스템의 부분인 디바이스들을 통한 데이터의 분산의 특정한 형평(로드 밸런싱)을 고려하면서, 주어진(원하는) 이용도를 위한 리플리케이션 팩터와 스토리지 비용 사이의 균형(tradeoff)을 최적화하는 것이다.

[0009] 이를 위해, 본 발명은 디바이스들의 이용도에 대한 지식에 기초한 데이터 리플리케이션을 위한 디바이스 세트의 선택을 제안한다. 본 발명의 특정한 실시예에 따르면, 이러한 이용도 지식은 네트워크 디바이스들에 의해, 또는 일부 네트워크 디바이스들 자체에 의해서만 네트워크를 모니터링함으로써 획득된다. 변형 실시예에 따르면, 이러한 지식은 인터넷 서버와 같은 네트워크의 서버들 중 하나로부터 획득된 분산된 데이터 스토리지 네트워크로의 연결에 대한 측정(measurement)에 의해 획득된다. 이용도 지식을 획득하는 방법은 능동적(active) 또는 수동적(passive)이며, 능동적 방법의 한 예시는 '핑(ping)' 메시지들의 사용을 통한 것이고, 수동적 방법의 한 예시는 분산된 데이터 스토리지 시스템의 중앙 집중된 서버에 의해 저장된 연결 로그들(connection logs)을 사용하는 것이다. 디바이스는 고장으로 인해 일시적 또는 영구적으로 연결이 해제될 때, 이용 가능하지 않는 것으로 간주된다.

[0010] 분산된 데이터 스토리지 시스템에 데이터를 저장하는 것을 최적화시키기 위해, 본 발명은 스토리지 디바이스들로서 적어도 사용되는 네트워크 디바이스들을 포함하는 분산된 데이터 스토리지 시스템에 데이터를 저장하는 방법을 제안한다. 디바이스들이 스토리지 디바이스들로서 적어도 사용된다는 것은 디바이스들이 분산된 스토리지 네트워크를 위한 스토리지 디바이스들로서 사용되며, 또한 오디오 및 비디오 프로그램들의 수신과 같은 다른 목적들을 위해 동시에 사용될 수 있다는 것을 의미한다. 한 예시로서, 디바이스들은 오디오 및 비디오 프로그램들의 수신을 위한 셋톱 박스들이거나, 또는 외부 네트워크, 개인용 컴퓨터들, 또는 핸드헬드 모바일 디바이스들에 대한 액세스 권한을 제공하는 게이트웨이들일 수 있다. 본 방법은 분산된 데이터 스토리지 시스템에 데이터 항목을 저장하기 위한 요청의 수신 단계, 레퍼런스 디바이스로서 제1 네트워크 디바이스의 제1 선택 단계, 및 레퍼런스 디바이스의 시간에 따른 이용도 및 비이용도의 결정 단계, 레퍼런스 디바이스의 시간에 따른 이용도에 대한 적어도 하나의 제2 네트워크 디바이스의 시간에 따른 이용도의 대응(correspondence)의 함수로서 적어도 하나의 제2 네트워크 디바이스의 제2 선택 단계, 레퍼런스 디바이스의 시간에 따른 비이용도에 대한 적어도 하나의 제3 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제3 네트워크 디바이스의 제3 선택 단계, 및 제2 선택 단계 및 제3 선택 단계에서 선택된 적어도 하나의 제2 네트워크 디바이스 및 적어도 하나의 제3 네트워크 디바이스에 데이터 항목을 저장하는 단계를 포함한다.

[0011] 변형 실시예에 따르면, k는 데이터 항목이 리플리케이팅되는 네트워크 디바이스들의 개수인데, 여기서 k는 데이터 항목을 저장하기 위한 요청에 명시되고, 제1, 제2, 및 제3 선택 단계들은 데이터 항목이 적어도 k개의 네트워크 디바이스들을 통해 리플리케이팅될 때까지 반복되며, 레퍼런스 디바이스의 제1 선택 단계는 본 방법의 이전 반복 시 레퍼런스 디바이스로서 이미 선택된 네트워크 디바이스의 선택을 배제시킨다.

[0012] 변형 실시예에 따르면, k는 데이터 항목이 리플리케이팅되는 네트워크 디바이스들의 개수인데, 여기서 k는 데이터 항목을 저장하기 위한 상기 요청에 명시되고, 제2 및 제3 선택 단계들은 데이터 항목이 적어도 k개의 네트워크 디바이스들을 통해 리플리케이팅될 때까지 반복된다.

[0013] 변형 실시예에 따르면, 비이용도는 각 원소(element)가 시간의 단위(unit of time)를 나타내는 벡터(24)로서 저장되고, 벡터(24)의 원소들의 개수가 시간 스팬(time span)을 나타내고, 벡터(24)가 관련된 네트워크 디바이스의 시간의 단위 동안의 이용도를 각 원소 값이 나타내는 경우, 제1 미리 결정된 원소 값은 이용도를 나타내고, 제2 미리 결정된 원소 값은 비이용도를 나타낸다.

[0014] 변형 실시예에 따르면, 제1 미리 결정된 원소 값은 양(positive)의 미리 결정된 값이고, 제2 미리 결정된 원소

값은 음(negative)의 미리 결정된 값이며, 레퍼런스 디바이스의 비이용도의 함수로서 제2 네트워크 디바이스의 선택은 두 개의 벡터들(24) 사이의 각도(angle)의 계산을 통해 결정된 반대-상관(anti-correlation)의 레벨에 따라 결정되고, 두 개의 벡터들(24)은 각도가 0에 근접할 때 크게 상관되며, 두 개의 벡터들(24)은 각도가 π에 근접할 때 크게 반대-상관되고, 상기 각도의 값은 반대-상관의 레벨을 나타낸다.

$$\theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \cdot ||\vec{y}||}\right)$$

[0015] 변형 실시예에 따르면, 각도는 에 따라 계산되며, 여기서 Θ는 각도이고, x와 y는 두 개의 벡터들이다.

[0016] 변형 실시예에 따르면, 제1 미리 결정된 원소 값은 이진수 1이고, 제2 미리 결정된 값은 이진수 0이며, 반대-상관의 레벨은 두 개의 벡터들(24) 사이의 로직 부울리안(logical Boolean) XOR 연산의 계산을 통해 결정되고, 두 개의 벡터들(24)은 로직 부울리안 XOR 연산의 결과가 각각의 두 개의 벡터들(24)의 원소들의 개수에 근접하는 1들의 큰 개수를 포함할 때, 크게 반대-상관되며, 두 개의 벡터들(24)은 로직 부울리안 XOR 연산의 결과가 0에 근접하는 1들의 개수를 포함할 때, 크게 상관되고, 부울리안 XOR 연산의 결과 중 1들의 개수는 반대-상관의 레벨을 나타낸다.

[0017] 변형 실시예에 따르면, 레퍼런스 네트워크 디바이스는 네트워크 디바이스들로부터 랜덤하게 선택된다.

[0018] 변형 실시예에 따르면, 레퍼런스 네트워크 디바이스는 네트워크 디바이스들로부터 결정론적으로 선택된다. 한 예시로서, 라운드-로빈 선택(round-robin selection)은 결정론적인 선택(deterministic selection)이다.

[0019] 분산된 데이터 스토리지 시스템에 데이터를 저장하는 것을 최적화하기 위해, 본 발명은 스토리지 디바이스들로서 적어도 사용되는 네트워크 디바이스들을 포함하는 분산된 데이터 스토리지 시스템에 데이터를 저장하기 위한 디바이스를 포함하며, 본 디바이스는 다음의 수단들: 분산된 데이터 스토리지 시스템에 데이터 항목을 저장하기 위한 요청을 수신하기 위한 수신기; 레퍼런스 네트워크 디바이스로서 제1 네트워크 디바이스의 제1 선택을 위한 선택기, 및 레퍼런스 디바이스의 시간에 따른 이용도 및 비이용도의 결정을 위한 수단; 레퍼런스 네트워크 디바이스의 시간에 따른 이용도에 대한 제2 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제2 네트워크 디바이스의 제2 선택을 위한 선택기; 레퍼런스 네트워크 디바이스의 시간에 따른 비이용도에 대한 적어도 하나의 제3 네트워크 디바이스의 시간에 따른 이용도의 대응의 함수로서 적어도 하나의 제3 네트워크 디바이스의 제3 선택을 위한 선택기; 및 제2 선택을 위한 선택기 및 제3 선택을 위한 선택기에 의해 선택된 네트워크 디바이스들에 데이터 항목을 저장하기 위한 스토리지 수단;을 포함한다.

[0020] 본 발명의 보다 많은 장점들은 본 발명의 특정한 비-제한적인 실시예들에 대한 개시 사항을 통해 나타날 것이다.

[0021] 실시예들은 다음의 도면들을 참조하여 설명될 것이다.

발명의 효과

[0022] 본 발명은 분산된 데이터 스토리지 시스템의 부분인 디바이스들을 통한 데이터의 분산의 특정한 형평(로드 밸런싱)을 고려하면서, 주어진(원하는) 이용도를 위한 리플리케이션 팩터와 스토리지 비용 사이의 균형(tradeoff)을 최적화시킨다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 특정 변형을 구현하기 위해 적합한 분산된 스토리지 네트워크 구조를 도시하는 도면.

도 2는 도 1의 디바이스들(10 내지 16) 중 하나와 같은 디바이스 또는 분산된 스토리지 디바이스에 대한 이용도 데이터가 수학적인 벡터 구조를 나타내는 n-차원 배열에 저장되는 본 발명의 변형 실시예를 도시하는 도면.

도 3은 24시간 시간 선(35)에 관한 두 개의 디바이스들의 이용도를 도시하는 도면.

도 4는 4개의 리플리케이팅 디바이스들을 포함하는 본 발명의 특정 실시예를 도시하는 도면.

도 5는 리플리케이션 디바이스들이 랜덤하게 선택되는 종래의 기술의 해결책의 성능에 본 발명의 성능이 비교되는 그래프를 도시하는 도면.

도 6은 본 발명의 방법을 구현하는 디바이스의 예시적인 실시예를 도시하는 도면.

도 7은, 예를 들어 도 1의 디바이스들(10 내지 16) 중 하나에 의해 구현되는 본 발명의 방법에 대한 특정 실시예를 구현하는 알고리즘을 도시하는 도면.

발명을 실시하기 위한 구체적인 내용

[0024] 도 1은 본 발명의 특정 변형을 구현하기 위해 적합한 분산된 스토리지 네트워크 구조를 도시한다. 본 도면은 게이트웨이(12), 서버(17), 또 다른 게이트웨이(14), 및 네트워크 디바이스(13)에 연결된 네트워크(1003)를 도시하는데, 게이트웨이(12)는 네트워크 디바이스들(10 및 11)을 네트워크(1003)에 연결하고, 또 다른 게이트웨이(14)는 네트워크 디바이스들(15 및 16)을 네트워크(1003)에 연결하며, 네트워크 디바이스(13)는 네트워크(1003)에 직접 연결된다. 모든 또는 단지 일부의 디바이스들(10 내지 16)은 데이터를 저장하기 위한 이들의 용량에 따라, 그리고 분산된 스토리지 네트워크의 구성원에 따라 분산된 스토리지 네트워크 내의 분산된 스토리지 디바이스들인 것으로 고려될 수 있다.

[0025] 도 1에 도시된 것과는 다른 유형의 네트워크들은 본 발명과 호환될 수 있다. 예를 들어, 네트워크들은 하나 이상의 네트워크 스위칭 노드들을 포함하거나, 또는 하나 이상의 서브네트워크들을 포함하는 네트워크들이며, 무선 네트워크 디바이스들을 연결한다.

[0026] 본 발명은 홈 네트워크를 인터넷에 연결하는 홈 게이트웨이, 웹 서버, 비디오 서버, 또는 무선 디바이스, 핸드헬드와 같은 임의의 유형의 네트워크 디바이스에서 구현될 수 있다.

[0027] 도 2는 디바이스들(10 내지 16) 중 하나와 같은 피어 또는 분산된 스토리지 디바이스에 대한 이용도 데이터가 수학적인 벡터 구조를 나타내는 n-차원 배열에 저장되는 본 발명의 특정 실시예를 도시한다. 수평 선(20-21-22)은 시간 선(23)에 관한 분산된 스토리지 디바이스의 이용도를 나타낸다. 구조(24)는 n-차원 벡터를 나타낸다. 여기에 도시된 특정 실시예에 따르면, 24시간들에 대해 n=24는 또한 이후부터는 샘플 주기로서 언급된다. 점선들(20, 22)은, 예를 들어 디바이스가 스위치 오프되기 때문에, 즉 디바이스가 이용 가능하지 않기 때문에, 분산된 스토리지 디바이스가 액세스될 수 없을 때의 24시간 시간 선(23) 상의 순간들(moments) 또는 시간 랩들(time laps)을 나타낸다. 실선(21)은 분산된 스토리지 디바이스 내의 데이터가 액세스될 수 있을 때, 즉 디바이스가 이용 가능할 때의 시간 선(23) 상의 순간 또는 시간 랩을 나타낸다. 변형 실시예에 따르면, 디바이스들의 이용도에 대한 지식은 이후부터 샘플 시간으로 언급되는 한 시간 샘플들로 나누어진다. 각 샘플은 24시간 시간 스케일 상의 특정 시간에 대응한다. 본 발명의 특정 실시예에 따르면, 각 시간에 대해 벡터의 대응 색인은, 디바이스가 해당 시간 동안 이용 가능했던 경우(즉, 디바이스가 네트워크에 연결되었던 경우, 또는 변형에 따르면, 이것의 데이터가 액세스될 수 있었던 경우), 양(positive)의 미리 결정된 값으로 설정된다. 대응 색인은 디바이스가 샘플 주기에 걸쳐 이용 가능하지 않았을 경우, -1과 같은 음(negative)의 미리 결정된 값으로 설정된다. 물론, 1시간의 샘플 시간 및 24의 샘플 주기는 본 발명의 원리들을 예증하는 예시들일 뿐이며; 다른 샘플 시간들 및 주기들이 가능하고, 비교해보면 이들의 상이한 장점들과 단점들을 갖는다. 예를 들어, 벡터는 24의 차원을 갖는데, 이는 1일 동안의, 또는 7일 동안의 24*7의 이용도 데이터를 저장하기에 충분하다. 샘플 시간 및 주기의 결정은 (계산 및 통신을 위해) 사용된 네트워크 리소스들과 데이터의 정확도 사이의 균형의 부분이다.

[0028] 변형 실시예에 따르면, 샘플 시간 및/또는 샘플 주기는 미리 결정된다. 보다 유리한 실시예에 따르면, 샘플 시간 및 주기는 분산된 데이터 스토리지 디바이스들의 연결 동작(connectivity behavior)에 대한 측정(measurements) 이후에 결정된다. 이후, 최상의 샘플 시간 및 샘플 주기는 디바이스들에 대한 비이용도의 최상의 순환 주기들을 획득하는 상기 측정들로부터 결정될 수 있다. 예를 들어, 순환(recurrency)은 순환하는 낮/밤 및 주말 이용도/비이용도 교대(recurrent day/night and weekend availability/unavailability alternations)를 포함하도록 샘플 주기가 한 주로 설정될 때 준수될 수 있다. 변형 실시예에 따르면, 분산된 스토리지 디바이스들은 협력적인 방식으로 양호한 샘플 시간 및 주기를 결정하는 임무를 수행하여, 이로써 중앙 집중된 서버에게 상기 임무를 경감시킬 수 있다.

[0029] 따라서, 각 디바이스는 이전에 설명된 포맷과 같은 보편적인 포맷에 따라 자체의 이용도 벡터를 갖는다. 변형 실시예에 따르면, 각 디바이스는 자체의 이용도 벡터를 저장한다. 이러한 변형은 디바이스 이용도에 대한 중앙 집중된 모니터링을 피하는 장점을 가지며, 디바이스들 사이의 통신을 위한 필요성과 중앙 집중된 이용도 스토리지를 감소시킨다.

[0030] 본 목적은 이제 데이터가 리플리케이션 세트를 통해 리플리케이팅될 때, 전체의 샘플 주기에 걸쳐 세트의 적어

도 하나의 디바이스의 이용도를 통해 데이터의 이용도를 보장하는 디바이스들 세트를 구축하는 것이다. 한 예시는 도 3에 도시된다. 본 도면은 24시간 시간 선(35)에 관한 디바이스 1(30, 31) 및 디바이스 2(32, 33)의 이용도를 도시한다. 점선들(30 및 33)은 디바이스들이 이용 가능하지 않다는 것을 의미한다. 실선들(31 및 32)은 디바이스들이 이용 가능하다는 것을 의미한다. 선(34)은, 디바이스 1과 디바이스 2를 포함하는 디바이스 세트가 선택될 때, 데이터 항목의 이용도는 100%이며, 따라서 디바이스 세트는 양호한 리플리케이션 세트임을 도시한다. 본 발명의 변형 실시예에 따르면, 양호한 리플리케이션 세트를 결정하기 위해, 본 발명은 다음의 공식에 따라 리플리케이션 세트가 이로부터 선택될 디바이스들에 대한 이용도 벡터들 사이의 각도의 측정을 포함한다:

[0031] $$\theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \cdot ||\vec{y}||}\right)$$

[0032] 이러한 공식은 벡터 x와 벡터 y 사이의 각도 Θ에 대한 값을 제공한다. 각도가 0에 근접할 때, 벡터들은 상관된다. 각도가 π에 근접할 때, 벡터들은 반대-상관된다. 두 개의 반대-상관된 벡터들은 반대의 동작을 포함하는 두 개의 디바이스들, 예를 들어 낮 동안 유일하게 이용 가능한 디바이스, 및 반대로 밤 동안 유일하게 이용 가능한 다른 디바이스를 나타낸다. 이후, 주어진 이용도 벡터를 갖는 주어진 디바이스에 대해, 주어진 디바이스와 가장 상관되는 이용도 벡터를 갖는 적어도 하나의 디바이스, 및 주어진 디바이스와 가장 반대-상관되는 이용도 벡터를 갖는 적어도 하나의 디바이스를 리플리케이션 세트로 선택하면, 샘플 주기는 리플리케이션 세트를 포함하기 위해 디바이스들을 랜덤하게 선택하는 종래의 기술의 방법들 보다 양호한 성능으로 커버된다. 종래의 기술에 따르면, 등가적인 또는 보다 낮은 레벨의 이용도는 리플리케이션 세트 내의 디바이스들의 개수를 증가시킴으로써 획득된다.

[0033] 본 발명의 변형 실시예에 따르면, 상관 및 반대-상관된 디바이스들은 벡터들 사이의 각도로 앞서 설명된 바와 같이 검출되지 않고, 오히려 로직 이진 연산을 통해 검출된다. 앞서 설명된 변형 실시예에서와 같이, 디바이스의 이용도 데이터는 각 색인이 샘플 시간, 예를 들어 1시간을 나타내며, 벡터 길이가 샘플 주기, 예를 들어 24시간을 나타내는 벡터로 변환된다. 하지만, 벡터가 이전 변형 실시예에 따라 +1 또는 -1로 채워지는 경우에, 현재 설명되는 변형 실시예에서 벡터는 0들 및 1들로 채워지는데, 0은 샘플 시간 동안의 비이용도를 나타내고, 1은 샘플 시간 동안의 이용도를 나타낸다. 예를 들어:

[0034] 디바이스 i : 010110010010010100100100

[0035] 완전히 반대-상관된 디바이스 : 101001101101101011011011

[0036] 이러한 변형 실시예에 따르면, 디바이스 i의 완전히 반대-상관된 디바이스는, 디바이스 i가 0들을 갖는 경우에는 로직 1들을 갖고, 디바이스 i가 1들을 갖는 경우에는 로직 0들을 갖는 디바이스이다. 이러한 변형 실시예에 따르면, 반대-상관된 디바이스들 중에서의 선택은 디바이스 i의 이용도 벡터와 반대-상관을 위한 후보들의 이용도 벡터들 사이의 로직 XOR 연산의 결과들을 비교함으로써 수행된다. XOR 연산의 결과들이 1들을 많이 포함하면 할수록, 디바이스들은 보다 더 반대-상관된다. 이러한 방법은 디바이스 i와의 이용도에 있어서 이들의 오버랩(overlap)을 최소화시키는 디바이스들을 통해 레퍼런스 디바이스 i의 이용도에 있어서의 갭들(gaps)을 채우도록 한다. 벡터들 사이의 각도에 대한 측정에 기초한 앞서 설명된 변형 실시예로서, 이러한 변형 실시예는 방해(interruption)가 없는 샘플 주기에 걸쳐 이용 가능한 제한된 선택의 디바이스들 상에 데이터를 부족하게 분산시키지 않고, 분산된 데이터 스토리지 시스템 상에 데이터를 분산시키는데, 이는 상이한 데이터 항목들에 관련된 리플리케이션 세트들에 포함된 디바이스들의 차별화(differentiation)를 통해 그룹화된 데이터 손실의 위험들을 보다 양호하게 흩뿌리는 장점을 갖는다. 측정된 각도가 반대-상관의 레벨을 나타내는 앞서 설명된 '각도' 변형 실시예에서와 같이, 현재 설명된 변형에서 두 개의 디바이스들의 이용도 벡터들 사이의 XOR 연산의 결과에 있어서 1들의 개수는 반대-상관의 레벨에 대한 측정이다. 예를 들어, 벡터들이 24개의 원소들을 포함하는 경우, 24개의 1들은 반대-상관의 가장 크게 가능한(완전한) 레벨을 나타내고, 0개의 1들은 반대-상관의 가장 작게 가능한 레벨을 나타낸다. 중간의 수들은 반대-상관의 중간 레벨들을 나타낸다. 물론, 반대-상관의 레벨들은 임의의 주어진 스케일, 예를 들어 간단한 산술 연산들을 적용함으로써 1부터 10까지의 스케일로 스케일링될 수 있다.

[0037] 물론, 본 발명의 상기 변형 실시예들은 디바이스들의 이용도 동작의 특정한 예측 가능성(predictability)을 가

정하는데, 그 이유는 실시예들이 과거에 측정된 이용도 데이터를 사용하기 때문이다. 하지만, 테스트들은 실제로 디바이스들의 이용도 동작이, 예를 들어 24시간 주기에 걸쳐서, 또는 7일 주기에 걸쳐서 순환적임을 입증했다. 디바이스들의 이용도 동작이 예측 가능하지 않다면, 본 발명의 방법은 여전히 스토리지 디바이스들의 순전한 랜덤한 선택 방법들만큼 양호한 이용도를 획득하는데, 그 이유는 언급된 바와 같이 각 레퍼런스 디바이스 i가 랜덤하게 선택되기 때문이다.

[0038] 도 4는 4개의 리플리케이팅 디바이스들에 대한 예시의 도움으로 본 발명의 특정 실시예를 도시한다. 이러한 변형 실시예에 따르면, 레퍼런스로서 사용된 주어진 디바이스 i(40)에 대해, 두 개의 디바이스들 1(41) 및 2(42)는 레퍼런스 디바이스 i(40)의 이용도 벡터와 가장 상관되는 이용도 벡터들을 통해 선택되고, 두 개의 디바이스들 3(43) 및 4(44)는 레퍼런스 디바이스 i(40)의 이용도 벡터와 가장 반대-상관되는 이용도 벡터들을 통해 선택되며, 즉 다른 말로, 디바이스들 1 및 2는 레퍼런스 디바이스의 시간에 따른 이용도에 대한 이들의 시간에 따른 이용도의 대응의 함수로서 선택되지만, 디바이스들 3 및 4는 레퍼런스 디바이스의 시간에 따른 비이용도에 대한 이들의 시간에 따른 이용도의 대응의 함수로서 선택된다. 이러한 반대-상관된 이용도 동작을 통해 디바이스들을 선택하는 것의 이점은 반대-상관된 디바이스들이, 디바이스 i(40), 및 이에 따라 i에 상관된 디바이스들 1 및 2가 이용 가능하지 않는 동안 일시적인 구역들(temporal zones)로 남겨진 갭들을 채운다는 사실에 있다. 데이터는 선택된 디바이스들 1 내지 4에 리플리케이팅된다. 본 발명의 이러한 변형 실시예는 이후부터 '그룹' 변형으로 언급되는데: 디바이스 j는 리플리케이션 팩터 k를 갖는 스토리지 네트워크에 진입하는 데이터의 항목을 저장하기를 원하고; 디바이스 n은 디바이스 i(레퍼런스 디바이스)를 랜덤하게 선택한다. 또 다른 변형 실시예의 디바이스 j에 따라, 또는 변형 실시예의 레퍼런스 디바이스 i에 따라, 서버는 디바이스 i와 상관된 특정 수의 디바이스들, 및 디바이스 i와 반대-상관된 특정 수의 디바이스들('특정 수'는 원하는 리플리케이션 팩터 k에 의존하며; 이러한 디바이스들의 선택은 레퍼런스 디바이스 i의 선택 단계를 배제시킨 디바이스들의 선택 단계들을 결국 반복함)을 선택하고, 원하는 리플리케이션 팩터 k를 달성하기 위해 이러한 디바이스들에 데이터 항목을 복사한다. 상기 예시에 따르면, 리플리케이션 팩터 k=4이고; 데이터 항목은 각각의 4개의 스토리지 디바이스들에 복사된다.

[0039] 또 다른 변형 실시예에 따르면, 특정 수의 상관된 디바이스들 및 특정 수의 반대-상관된 디바이스들을 선택하는 것 대신에, i와 반대-상관된 단일의 디바이스가 선택되는데, 이는 '패칭(patching)' 변형으로 더 언급된다. 데이터 항목은 이러한 디바이스들에 복사되는데; K>2이면, 프로세스는 '새로운' 레퍼런스 디바이스 i의 선택, 및 i와 반대-상관된 새로운 디바이스를 위한 검색으로 반복되며, 데이터 항목을 이러한 디바이스들에 복사한다. 프로세스는 k가 달성될 때까지 반복된다. 이러한 디바이스들의 선택은 결국 레퍼런스 디바이스 i의 선택을 포함하는 디바이스들의 선택을 반복하게 한다.

[0040] 변형들은 모두 종래의 기술에 대한 개선이며, '그룹' 변형은 특히 n개 중에서(n의 적절한 값들 중에서) 항상 t개의 인코딩된 블록들을 갖기 위해 적합하고, 이는, 예를 들어 고장에 대처하기 위해 n개 중에서 t개의 삭제 코드 블록들의 이용도를 요구하는 삭제 코드들(erasure codes)에 매우 잘 적응된다. '패칭' 변형은 데이터 항목 중 적어도 하나의 사본을 계속해서 이용 가능하게 유지하기에 보다 효율적인데, 그 이유는 '그룹' 변형이 때때로 일부 일시적인 갭들을 야기할 수 있기 때문이며, 즉 패칭 변형에 비해, 데이터의 주어진 항목을 저장하는 적어도 하나의 이용 가능한 디바이스가 항상 존재한다는 것을 덜 보장하기 때문이다.

[0041] 앞서 언급된 바와 같이, 본 발명의 변형 실시예에 따르면, 디바이스 "j"가 데이터를 저장하기 위해 분산된 데이터 스토리지 시스템에 진입하면, 본 시스템은 상관 및 반대-상관된 디바이스들을 포함하는 리플리케이션 세트의 결정에 대한 기본적인 역할을 할 레퍼런스 디바이스 "i"를 결정한다. 앞서 논의된 변형 실시예에 따르면, 레퍼런스 디바이스 "i"는 네트워크 내의 중앙 서버에 의해 랜덤하게 선택된다. 또 다른 변형 실시예에 따르면, 레퍼런스 디바이스 "i"는 한번 이상 동일한 레퍼런스 디바이스를 확실히 않도록 결정론적인 방식으로 결정되며, 이로써 네트워크상에서 데이터의 보다 양호한 분산을 획득하게 한다. 본 발명의 또 다른 특정 실시예에 따르면, 이전에 선택된 레퍼런스 디바이스들이 배제된 디바이스들의 세트로부터 레퍼런스 디바이스 i를 랜덤하게 선택하도록 모든 변형들은 결합되어서, 분산된 데이터 스토리지 네트워크상에서 데이터의 보다 양호한 분산을 획득하게 한다.

[0042] 본 발명의 변형 실시예에 따르면, 리플리커들(replicas)의 개수는 전체 샘플 주기를 '커버(cover)'하도록 리플리케이션 세트의 용량에 의존한다. 이러한 변형에 따르면, 이미 선택된 디바이스 세트의 결과적인 이용도를 통해 반대-상관된 디바이스들을 찾는 단계들은 이용도에 있어서 갭이 더 이상 존재하지 않을 때까지 반복된다.

[0043] 도 5는 본 발명의 성능이 데이터 항목의 리플리케이션을 위한 디바이스들이 랜덤하게 선택되는 종래의 기술의

해결책의 성능에 비교되는 그래프를 도시한다. 본 그래프는 시뮬레이션 데이터에 기초한다. 확인될 수 있는 바와 같이, 98%의 이용도를 획득하기 위해, 종래의 기술의 랜덤한 리플리케이션 세트는 8개의 디바이스들{그래프(54), 포인트(52)}을 사용하지만, 본 발명의 방법에서는 5개의 디바이스들만을 필요로 한다{그래프(53), 포인트(51)}.

[0044] 도 6은 특정 실시예에 따라 본 발명을 구현할 수 있는 디바이스를 도시한다. 본 디바이스는 본 발명의 특정 변형을 구현하는 디바이스, 전용의 서버와 같은 본 발명의 특정 변형을 구현하는 특정의 중앙 집중된 디바이스, 또는 본 발명의 특정 변형을 구현하지 않는 디바이스일 수 있다.

[0045] 디바이스(60)는 판독-전용 메모리(ROM, 600), 랜덤 액세스 메모리(RAM, 610), 중앙 프로세싱 유닛(CPU, 620), 클럭(630), 네트워크 인터페이스(640), 그래픽 인터페이스(650), 및 사용자 인터페이스(660)를 포함한다. 모든 이들 구성 요소들은 데이터- 및 통신 버스(670)를 통해 상호 연결된다. CPU(620)는 메모리 구역(601) 내의 ROM(600)에 저장된 프로그램에 따라 디바이스(60)를 제어한다. 클럭 디바이스(630)는 디바이스(60)의 구성 요소들에 공통의 타이밍을 제공하여, 이들의 연산을 순차 배열화 및 동기화시킨다. 네트워크 인터페이스(640)는 연결(6000)을 통해 외부 디바이스들과 데이터를 수신 및 송신한다. 그래픽 인터페이스(650)는 연결(6001)을 통해 외부의 렌더링 디바이스에 연결된다. 사용자 인터페이스(660)는 연결(6002)을 통해 사용자로부터 입력 명령들을 수신한다. ROM 메모리 구역(601)은 본 발명의 방법을 구현하는 명령어들을 포함한다. 디바이스(60)의 전원이 켜지면, CPU(620)는 ROM 메모리 구역(601)으로부터 RAM 메모리 구역(611)에 프로그램 '프로그(Prog)'를 복사하고, 복사된 프로그램을 실행한다. 복사된 프로그램이 실행되면, 프로그램은 RAM 메모리 구역(615) 내에서의 실행을 위해 필요한 변수들을 위한 메모리 공간을 할당하고, 디바이스 어드레스들의 목록(602), 디바이스 이용도 벡터들(603), 및 디바이스 세트들(604)을 각각 RAM 메모리들(612 내지 614)에 복사한다. 디바이스(60)는 이제 연산 가능하며, 데이터 항목을 저장하기 위한 요청의 수신 시에 본 발명의 방법은 활성화되며, 본 방법은 ROM(600)의 메모리 구역(601)에 저장된 프로그램의 부분이다. 네트워크 인터페이스(640)는 데이터 항목을 저장하기 위한 요청들의 수신을 위한 수신기의 역할을 한다. 그 중에서도 특히, CPU(620)는 레퍼런스 디바이스로서 네트워크 디바이스들 중 하나를 제1 선택하기 위한 선택기의 역할을 한다. CPU(620)는 또한 적어도 하나의 네트워크 디바이스를 제2 선택하기 위한 선택기의 역할을 하는데, 상기 적어도 하나의 네트워크 디바이스는 레퍼런스 네트워크 디바이스를 위해 결정된 이용도에 반대-상관된 미리 결정된 시간 주기에 걸친 이용도를 갖고, CPU(620)는 적어도 하나의 선택된 네트워크 디바이스를 데이터 항목을 저장하기 위한 리플리케이션 세트에 추가하기 위한 수단의 역할을 한다. ROM 메모리(600)와 RAM 메모리(610)는 리플리케이션 세트 내의 네트워크 디바이스들 상에 데이터 항목을 저장하기 위한 스토리지 수단이다.

[0046] 변형 실시예에 따르면, 본 발명은 하드웨어로, 예를 들어 전용의 구성 요소로서(예를 들어, ASIC, FPGA 또는 VLSI로서){각각 << 주문형 반도체(Application Specific Integrated Circuit)>>, <<필드-프로그래머블 게이트 어레이(Field-Programmable Gate Array)>>, 및 <<베리 라지 스케일 인티그레이션(Very Large Scale Integration)>>}, 또는 디바이스 내에 집적된 독특한 전자 구성 요소들로서 전체적으로 구현되거나, 또는 하드웨어와 소프트웨어의 혼합의 형태로 전체적으로 구현된다.

[0047] 도 7은, 예를 들어 도 1의 디바이스들(10 내지 16) 중 하나에 의해, 또는 도 6의 디바이스(60)에 의해 구현되는 본 발명의 방법의 특정 실시예에 대한 알고리즘을 도시한다. 본 알고리즘은 초기화 단계(700)로 시작하며, 상기 초기화 단계에서 사용되는 변수들이 초기화된다. 이후, 단계(701)에서 분산된 데이터 스토리지 시스템에 데이터를 저장하기 위한 요청이 수신된다. 특정 변형 실시예에 따르면, 상기 요청은 도 1의 서버(17)와 같은, 본 발명의 단계들을 구현하는 중앙 서버에 의해 수신된다. 이러한 변형 실시예는 '중앙 집중된' 변형 실시예로서 더 상술된다. 또 다른 변형 실시예에 따르면, 상기 요청은 본 발명을 구현하는 네트워크 디바이스들 중 하나에 의해, 예를 들어 도 1의 네트워크 디바이스들(10 내지 16) 중 어느 한 디바이스에 의해, 또는 도 6의 디바이스(60)에 의해 수신된다. 이러한 변형 실시예는 '분산된(decentralized)' 변형으로 더 상술된다. 이후, 단계(702)에서 선택은 도 1의 네트워크 디바이스들(10 내지 16) 중 하나의 선택과 같은 레퍼런스 디바이스로서의 네트워크 디바이스에 의해, 또는 도 6의 디바이스(60)에 의해 수행된다. 중앙 집중된 또는 분산된 변형 실시예에 따르면, 이러한 선택은 중앙 집중된 서버(중앙 집중된 변형)에 의해, 또는 도 1의 디바이스들(10 내지 16)과 같은 다른 네트워크 디바이스들 중 하나(분산된 변형)에 의해 수행된다. 분산된 변형의 변형 실시예에 따르면, 단계(701)의 요청을 수신하는 디바이스는 단계(702)의 레퍼런스 디바이스를 선택하는 디바이스와 동일한 디바이스가 아닌데, 이때는 이 단계가 도 1의 서버(17)와 같은 중앙 집중된 서버에 남겨질(left over) 때이다. 레퍼런스 디바이스는 랜덤하게 선택되거나, 또는 변형 실시예에 따라 이용 가능한 디바이스들의 결정론적인 라운드 로빈 선택을 통해 선택된다. 이후, 단계(703)에서, 이용도가 레퍼런스 디바이스의 이용도와 반대-상관된 디바이스가 나머지 디바

이스들, 즉 레퍼런스 디바이스를 배제시킨 네트워크 디바이스들로부터 선택된다. 예를 들어, 도 1의 디바이스(10)가 레퍼런스 디바이스이면, 나머지 디바이스들은 디바이스들(11 내지 16)이다. 이러한 반대-상관된 디바이스(또는 '완전한' 반대-상관된 디바이스를 찾는 것이 어렵거나 또는 불가능할 수 있으므로, 네트워크상에서 발견될 수 있는 보다 더 반대-상관된 디바이스)는 네트워크 디바이스들을 위해 알려진 이용도 데이터에 따라 선택된다. 하나 이상의 반대-상관된 디바이스들이 선택된다면, 이들은 레퍼런스 디바이스 i와 보다 더 반대-상관된 디바이스들의 목록의 랭크(rank)에 따라 반복적으로 선택된다. 이러한 이용도 데이터는 특정 시간 주기에 걸쳐 네트워크 디바이스를 위해 측정된 이용도를 나타낸다. 본 발명의 특정 실시예에 따라, 이용도 데이터는 각각의 네트워크 디바이스들 자체에 의해 획득된다. 본 발명의 변형 실시예에 따르면, 이용도 데이터는 네트워크 디바이스들의 이용도 동작을 모니터링하는 중앙 서버에 의해 획득된다. 본 발명의 또 다른 변형 실시예에 따르면, 이용도 데이터는 디바이스들로부터 이용도 데이터를 수집하는 도 1의 서버(17)와 같은 중앙 집중된 서버로부터 획득되며, 디바이스들 그 자체에 의해 측정된다. 반대-상관의 개념은 이미 앞서 설명되었다. 선택된 반대-상관된 디바이스는 단계(701)의 데이터 추가 요청의 데이터 항목을 저장하기 위해 사용될 디바이스들의 세트에 추가되며, 또한 리플리케이션 세트라고도 언급된다. 이러한 리플리케이션 세트는 IP/port 어드레스들과 같은 각각의 개별 디바이스를 어드레싱하는 것을 가능하게 하는 식별자들로 나타내어지며, 특정 실시예에 따라 중앙 집중된 서버에 저장되는데, 예를 들어 분산된 데이터 스토리지 네트워크에 저장된 각각의 특정 데이터 항목에 대해, 관련된 리플리케이션 세트가 저장되거나, 또는 변형 실시예에 따라 레퍼런스 디바이스와 같은 다른 네트워크 디바이스들 중 하나에 저장된다. 리플리케이션 세트는 어떤 디바이스들이 주어진 데이터 항목을 저장하는지를 결정할 수 있도록 주어진 데이터 항목과 관련이 있다. 본 발명의 특정 실시예에 따르면, 이용도 데이터는 도 2의 벡터(24)와 같은 벡터의 형태로 저장되며, 이의 각 원소는 1시간과 같은 시간의 단위를 나타내고, 벡터의 원소들의 개수는 24시간과 같은 시간 스팬을 나타낸다. 벡터 원소들은 대응하는 디바이스가 대응하는 시간의 단위 동안 이용 가능하지 않았는지를 나타내는 값들로 채워지는데, 예를 들어 벡터[5]=+1은 대응하는 네트워크 디바이스가 오전 05:00부터 오전 05:59까지 이용 가능했다는 것을 나타내고, 벡터[11]=-1은 대응하는 네트워크 디바이스가 오전 11:00부터 오전 11:59까지 이용 가능하지 않았다는 것을 나타내며, 또는 오히려 앞서 설명된 변형 실시예에 따르면, 이진 값 1은 이용도를 나타내고, 이진 값 0은 비이용도를 나타낸다. 특정 실시예에 따르면, 벡터의 값을 설정하는 것을 가능하게 하는 임계값이 설정된다. 예를 들어, 임계값은 단위 시간의 절반으로 설정되고, 단 20분만 이용 가능했던 디바이스에 대해, 대응하는 벡터의 원소는 (선택된 변형 실시예 'xor' 또는 '각도'에 따라) 이진 값 0 또는 이진 값 -1로 설정될 것이다.

[0048] 이후, 단계(704)에서, 네트워크 디바이스가 선택된다. 이러한 선택된 디바이스는 이후 리플리케이션 세트에 추가된다. 상관의 개념은 이미 앞서 설명되었다. 다음 단계(705)에서, 데이터 항목은 리플리케이션 세트의 디바이스들에 리플리케이팅(복사)되고, 알고리즘은 단계(701)의 반복으로 계속된다. 각 단계(704)에서, k/2 디바이스들은 리플리케이션 세트로 선택되지만, 다시 k/2 상관된 디바이스들은 나중에 선택될 것이다(이는 앞서 설명된 '그룹' 변형이다). 변형 실시예(앞서 설명된 '패칭' 변형)에 따르면, 한 디바이스는 각 단계(704)에서 선택되며, 단계들(702 내지 704)은 k개의 디바이스들이 선택될 때까지 반복되어서, 모든 변형들에서, 하지만 상이한 방식으로 원하는 리플리케이션 팩터가 획득된다. 각각의 '그룹' 및 '패칭' 변형은 앞서 설명된 바와 같이 자체적인 장점들을 갖는다.

부호의 설명

[0049] 10, 11, 13, 15, 16 : 네트워크 디바이스

12, 14 : 게이트웨이 17 : 서버

1003 : 네트워크 600 : ROM

601, 611 : 프로그 602, 612 : 디바이스 어드레스들의 목록

603 : 피어 이용도 벡터들 613 : 디바이스 이용도 벡터들

604, 614 : 리플리케이션 세트들 610 : RAM

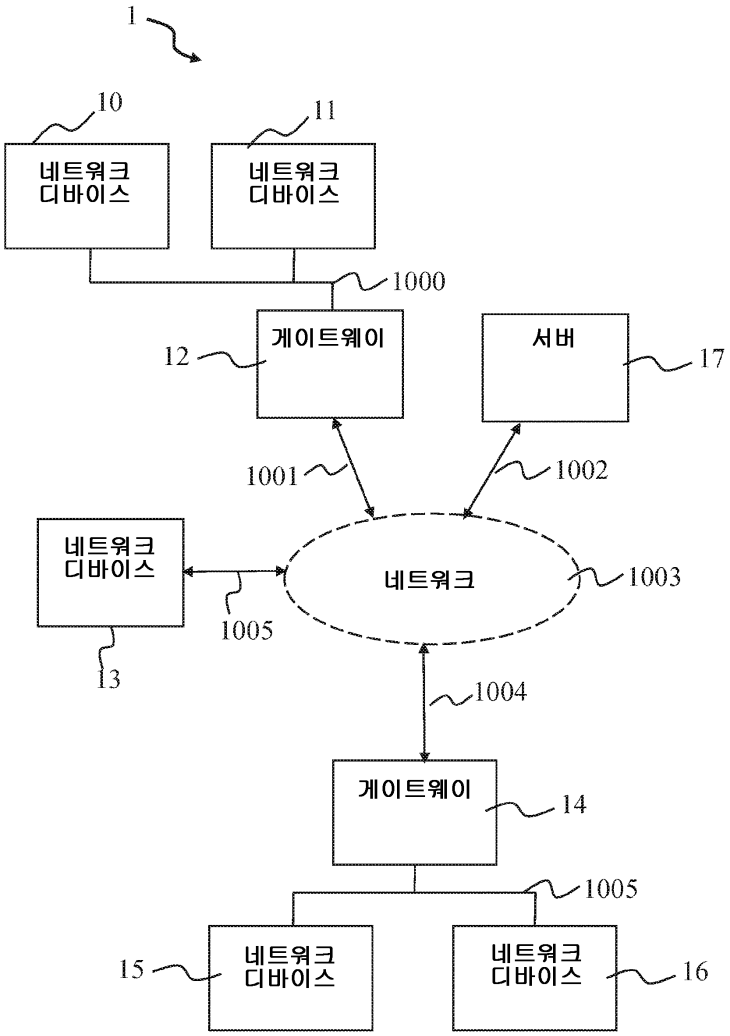
615 : 데이터 620 : CPU

630 : 클럭 640 : 네트워크 인터페이스

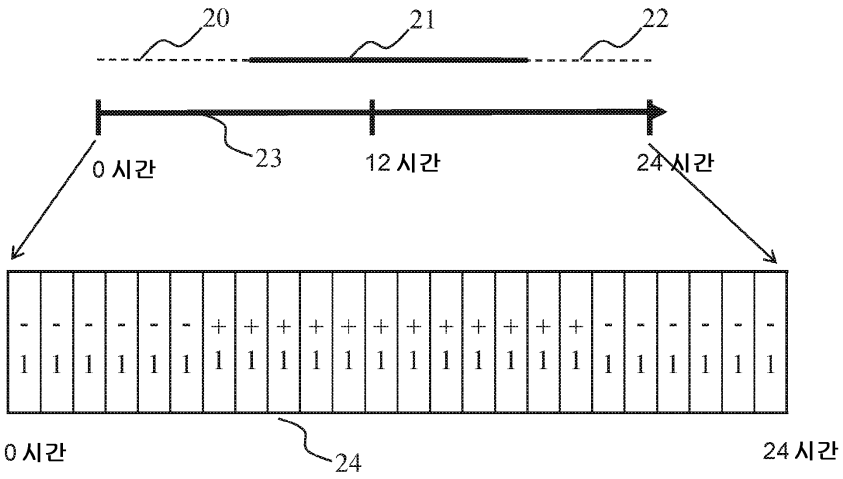
650 : 그래픽 인터페이스 660 : 사용자 인터페이스

도면

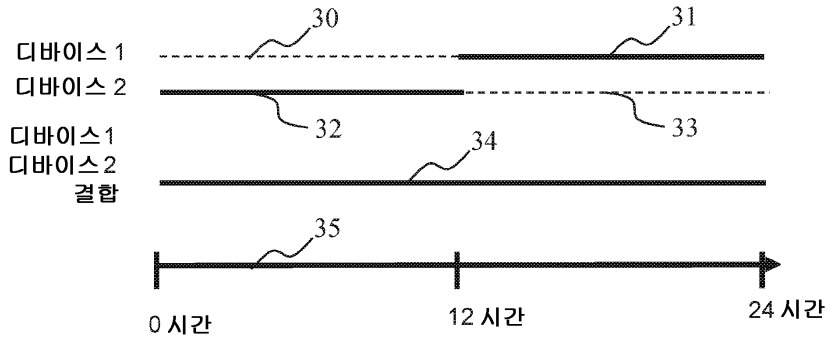
도면1


1
10
11
네트워크 디바이스
네트워크 디바이스
1000
12
게이트웨이
서버
17
1001
1002
네트워크 디바이스
네트워크
1003
1005
13
1004
게이트웨이
14
1005
15
네트워크 디바이스
네트워크 디바이스
16

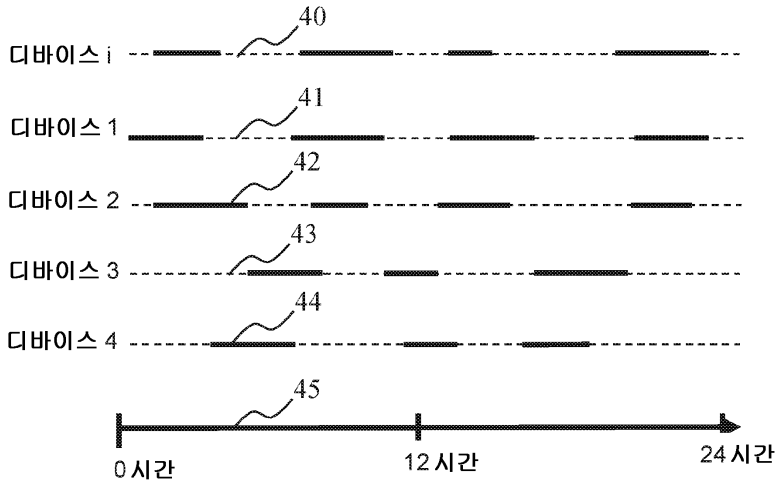
도면2


20
21
22
0 시간
23
12 시간
24 시간
0 시간
24
24 시간

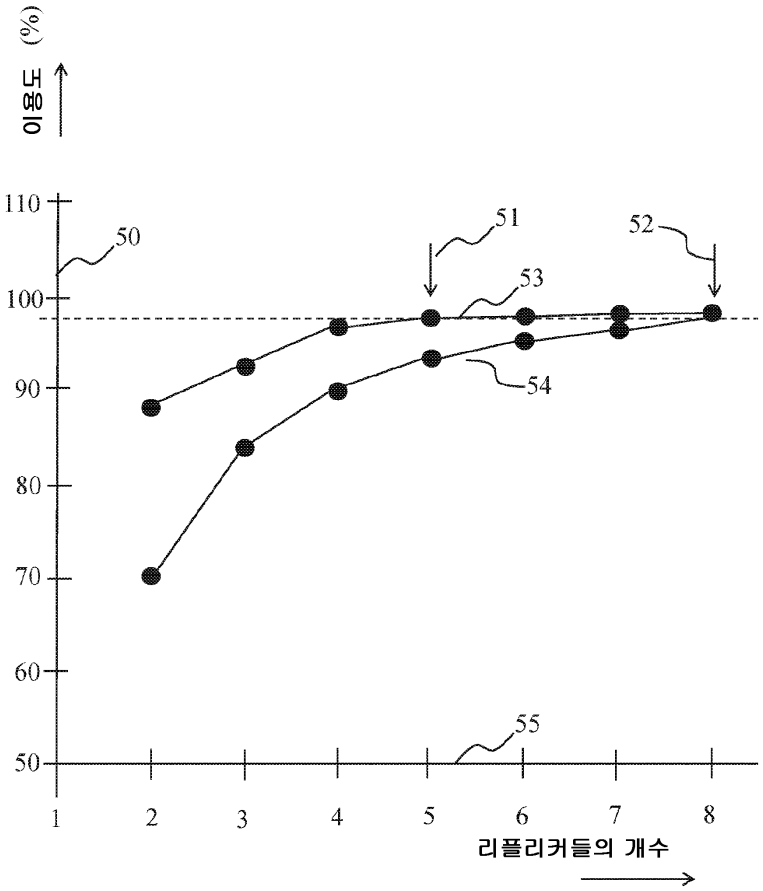
도면3


디바이스 1
디바이스 2
디바이스1
디바이스2
결합
30
31
32
33
34
35
0 시간
12 시간
24 시간

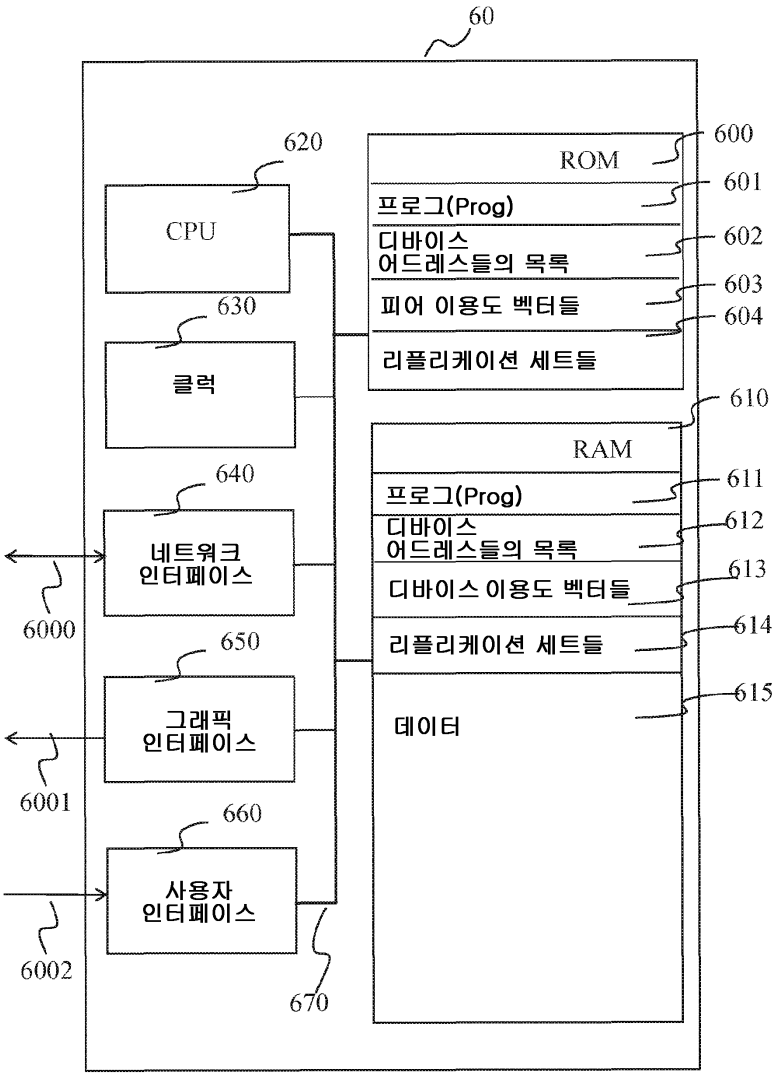
도면4


디바이스 i
디바이스 1
디바이스 2
디바이스 3
디바이스 4
40
41
42
43
44
45
0 시간
12 시간
24 시간

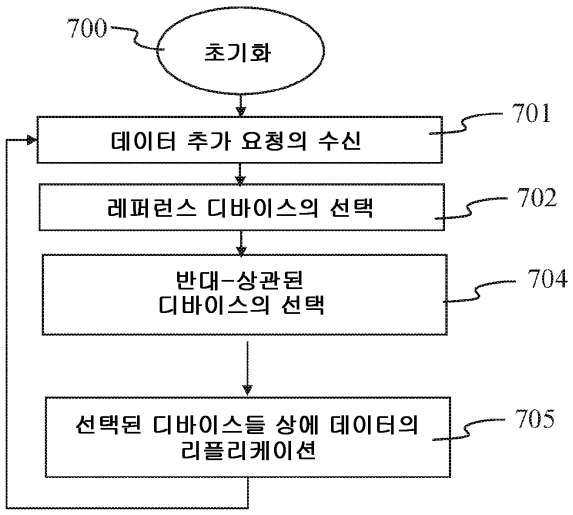
도면5



도면6


60
620
CPU
630
클럭
640
네트워크
인터페이스
6000
650
그래픽
인터페이스
6001
660
사용자
인터페이스
6002
670
ROM
600
프로그(Prog)
601
디바이스
어드레스들의 목록
602
피어 이용도 벡터들
603
604
리플리케이션 세트들
RAM
610
프로그(Prog)
611
디바이스
어드레스들의 목록
612
디바이스 이용도 벡터들
613
리플리케이션 세트들
614
데이터
615

도면7


700
초기화
데이터 추가 요청의 수신
701
레퍼런스 디바이스의 선택
702
반대-상관된
디바이스의 선택
704
선택된 디바이스들 상에 데이터의
리플리케이션
705