(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0070930
(43) 공개일자 2015년06월25일

(51) 국제특허분류(Int. Cl.)
G06F 19/00 (2011.01)
(21) 출원번호 10-2014-0072361
(22) 출원일자 2014년06월13일
심사청구일자 2014년06월13일
(30) 우선권주장
1020130157162 2013년12월17일 대한민국(KR)

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
허의남
경기도 용인시 기흥구 사은로126번길 46, 316동 1802호 (보라동, 현대모닝사이드1차아파트)
손아영
서울특별시 동대문구 경희대로 26, 전자정보대학 331호 (회기동)
(74) 대리인
특허법인 신지

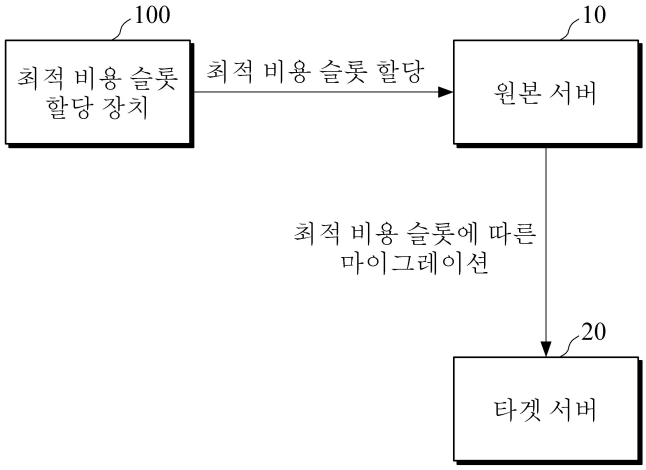
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **최적 비용을 고려한 마이그레이션 자원 할당 시스템 및 할당 방법**

(57) 요 약

본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템은 하나 이상의 가상 단말을 할당하고, 마이그레이션된 서비스를 가상 단말에서 제공하는 타겟 서버, 하나 이상의 가상 단말을 할당하고, 마이그레이션 슬롯이 할당되며, 할당된 마이그레이션 슬롯을 통해 가상 단말에서 제공하는 서비스를 상기 타겟 서버로 마이그레이션하는 원본 서버 및 거절 가능성 및 대기 시간을 고려하여 예측 대기 시간을 산출하고, 산출된 예측 대기 시간 및 비용을 고려하여 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하여 원본 서버에 추가적인 슬롯 및 마이그레이션 슬롯을 할당하는 최적 비용 슬롯 할당 장치를 포함한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업
과제고유번호 17110-01260
부처명 미래창조과학부
연구관리전문기관 한국정보통신산업진흥원
연구사업명 대학IT 연구센터육성 지원사업
연구과제명 실시간 모바일 클라우드 서비스 플랫폼 연구개발
기 여 율 1/1
주관기관 경희대학교 산학협력단
연구기간 2013.06.01 ~ 2013.12.31

명 세 서

청구범위

청구항 1

하나 이상의 가상 단말을 할당하고, 마이그레이션된 서비스를 가상 단말에서 제공하는 타겟 서버;

하나 이상의 가상 단말을 할당하고, 상기 할당된 하나 이상의 가상 단말 중에서 하나 이상의 가상 단말에 마이그레이션 슬롯이 할당되며, 상기 할당된 마이그레이션 슬롯을 통해 가상 단말에서 제공하는 상기 서비스를 상기 타겟 서버로 마이그레이션하는 원본 서버; 및

거절 가능성 및 대기 시간을 고려하여 예측 대기 시간을 산출하고, 상기 산출된 예측 대기 시간 및 비용을 고려하여 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하여 상기 원본 서버에 추가적인 슬롯 및 마이그레이션 슬롯을 할당하는 최적 비용 슬롯 할당 장치;

를 포함하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 2

제1항에 있어서,

상기 최적 비용 슬롯 할당 장치는,

상기 거절 가능성을 산출하고, 상기 산출된 거절 가능성 및 대기 시간을 고려하여 예측 대기 시간을 산출하고, 상기 산출된 예측 대기 시간의 임계값을 설정하며, 상기 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 3

제2항에 있어서,

상기 거절 가능성은,

$$P_r(R) = \sum_{j=0}^{K} P_{N,j}$$

에 의해 산출되며,

$P_r(R)$는 전체 거절 가능성을 나타내고, $P_{N,j}$은 가용 전체 슬롯 개수가 N 일 때, 마이그레이션 개수가 K인 경우의 거절 가능성을 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 4

제2항에 있어서,

상기 최적 비용 슬롯 할당 장치는,

동시에 하나의 마이그레이션 슬롯만 허용된 제1 케이스 및 동시에 둘 이상의 마이그레이션 슬롯이 허용된 제2 케이스 각각에 대해 대기 시간을 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 5

제4항에 있어서,

상기 최적 비용 슬롯 할당 장치는,

상기 거절 가능성의 임계값 범위 내에서 상기 제1 케이스 및 상기 제2 케이스의 대기 시간 및 거절 가능성을 고

려하여 예측 대기 시간을 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 6

제5항에 있어서,

상기 예측 대기 시간은

$$E[T^D] = \sum_{i=2K+1}^{N} P_{i,K} T_{i,K}^D + \sum_{i=0}^{K-1} P_{N,j} T_{N,j}^D$$

에 의해 산출되며,

상기 $P_{i,K}$ 는 제1 케이스의 각 상태의 거절 가능성을 나타내고, $T_{i,K}^D$는 제1 케이스의 대기 시간을 나타내며, $P_{N,j}$ 는 제2 케이스의 각 상태의 거절 가능성을 나타내고, 상기 $T_{N,j}^D$는 제2 케이스의 대기 시간을 나타내며, 상기 $P_{N,i}$ 은 거절 가능성을 나타내고, 상기 Th_R 은 거절 가능성의 임계값을 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 7

제6항에 있어서,

상기 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수는

$$\min_{\Delta N, K'} \Delta N \tau \Omega + E'\left[T^D\right] \delta \Omega$$

;

$$E'\left[T^D\right] \leq Th_W$$

및

$$E'[T^D] = \sum_{i=2K'+1}^{N'} P_{i,K'} T_{i,K'}^D + \sum_{j=0}^{K'-1} P_{N',j} T_{N',j}^D.$$

에 의해 산출되며,

상기 Th_W는 상기 설정된 예측 대기 시간의 임계값을 나타내고, 상기 N`은 상기 추가적인 슬롯 $\Delta N$과 현재 사용되는 전체 자원 N의 합을 나타내고, 상기 K`은 상기 마이그레이션 슬롯 개수를 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 시스템.

청구항 8

거절 가능성을 산출하는 단계;

대기 시간을 산출하는 단계;

상기 거절 가능성의 임계값을 설정하는 단계;

상기 대기 시간 및 상기 거절 가능성을 고려하여 예측 대기 시간을 산출하는 단계;

상기 예측 대기 시간의 임계값을 설정하는 단계; 및

추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하는 단계;

를 포함하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 9

제8항에 있어서,

상기 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하는 단계는,

상기 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 10

제9항에 있어서,

상기 거절 가능성은,

$P_r(R) = \sum_{j=0}^{K} P_{N,j}$ 에 의해 산출되며,

$P_r(R)$ 는 전체 거절 가능성을 나타내고, $P_{N,j}$ 은 가용 전체 슬롯 개수가 N 일 때, 마이그레이션 개수가 K인 경우의 거절 가능성을 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 11

제9항에 있어서,

상기 대기 시간 및 상기 거절 가능성을 고려하여 예측 대기 시간을 산출하는 단계는 동시에 하나의 마이그레이션 슬롯만 허용된 제1 케이스 및 동시에 둘 이상의 마이그레이션 슬롯이 허용된 제2 케이스 각각에 대해 대기 시간을 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 12

제11항에 있어서,

상기 대기 시간 및 상기 거절 가능성을 고려하여 예측 대기 시간을 산출하는 단계는 상기 거절 가능성의 임계값 범위 내에서 상기 제1 케이스 및 상기 제2 케이스의 대기 시간 및 거절 가능성을 고려하여 예측 대기 시간을 산출하는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 13

제11항에 있어서,

상기 예측 대기 시간은

$E[T^D] = \sum_{i=2K+1}^{N} P_{i,K} T_{i,K}^D + \sum_{i=0}^{K-1} P_{N,j} T_{N,j}^D$ 에 의해 산출되며,

상기 $P_{i,K}$ 는 제1 케이스의 각 상태의 거절 가능성을 나타내고, $T_{i,K}^D$는 제1 케이스의 대기 시간을 나타내며, $P_{N,j}$ 는 제2 케이스의 각 상태의 거절 가능성을 나타내고, 상기 $T_{N,j}^D$는 제2 케이스의 대기 시간을 나타내며, 상기 $P_{N,i}$ 은 거절 가능성을 나타내고, 상기 Th_R 은 거절 가능성의 임계값을 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

청구항 14

제13항에 있어서,

상기 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수는

$$\min_{\Delta N, K'} \quad \Delta N \tau \Omega + E'\left[T^D\right] \delta \Omega$$

;

$$E'\left[T^D\right] \leq Th_W$$

및

$$E'[T^D] = \sum_{i=2K'+1}^{N'} P_{i,K'} T^D_{i,K'} + \sum_{j=0}^{K'-1} P_{N',j} T^D_{N',j}$$

에 의해 산출되며,

상기 Th_W는 상기 설정된 예측 대기 시간의 임계값을 나타내고, 상기 N`은 상기 추가적인 슬롯 $\Delta N$과 현재 사용되는 전체 자원 N의 합을 나타내고, 상기 K`은 상기 마이그레이션 슬롯 개수를 나타내는 것을 특징으로 하는 최적 비용을 고려한 마이그레이션 자원 할당 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 클라우드 컴퓨팅 환경에 관한 기술로, 보다 상세하게는 클라우드 컴퓨팅 환경에서의 자원할당에 관한 기술이다.

배 경 기 술

[0002] 클라우드 환경은 가상 물리서버의 새로운 패러다임이다. 특히, 클라우드 환경과 연계된 가상 머신은 하나의 물리적 장치를 둘 이상의 가상머신으로 분할함으로써, 자원 활용 및 공유의 효율성을 극대화 할 수 있다. 가상 머신(Virtual Machine)을 할당하는 과정은 일반적으로 가상 머신 할당 시 발생하는 작업부하의 특성을 파악하여 가장 효율적인 작업부하를 가질 수 있도록 자원을 할당한다. 그리고, 필요한 경우 성능 최적화 및 저전력 운용을 위해 가상머신 마이그레이션(Migration)을 활용하여 작업 효율을 높이고 있다. 가상 머신에서 실시간 마이그레이션(Live Migration)은 클라우드 컴퓨팅 환경에서 나타나는 중요한 컴포넌트이며, 데이터 센터에서 가상 머신 마이그레이션은 성능 및 비용을 향상시킬 수 있는 방법이 될 수 있다. 가상 머신을 마이그레이션 하는 것은 두 개의 물리 호스트를 필요로 하는데 하나는 가상 머신이 본래 존재하던 호스트이고, 다른 하나는 가상 머신이 마이그레이션할 타겟 호스트이다. 그러나, 가상 머신 마이그레이션은 추가적인 자원(slot)을 필요로 하며, 자원(Slot)을 모두 사용하고 있을 경우 새로운 서비스 요청에 대한 거절 가능성이 높아진다.

[0003] 데이터 센터 내에서의 마이그레이션은 가상 머신의 메모리에 상주하는 콘텐츠가 하나의 물리적 호스트에서 다른 하나의 물리적 타겟 호스트로 이동하는 과정이다. 마이그레이션은 크게 오프라인 마이그레이션(Offline Migration) 및 실시간 마이그레이션(Live Migration)으로 구분할 수 있다. 오프라인 마이그레이션은 원본 호스트에서 실행중인 가상 머신이 중지되고 중지된 가상 머신 메모리의 데이터(콘텐츠)가 타겟 호스트로 이동된 후, 타겟 호스트에서 가상 머신이 다시 시작되는 마이그레이션이다. 오프라인 마이그레이션에서, 가상 머신이 타겟 호스트로 이동하면, 원본 호스트의 메모리는 프리 상태가 된다. 반면에, 실시간 마이그레이션은 원본 호스트의 가상 머신이 동작하는 상태에서 타겟 호스트로 메모리의 데이터가 이동한다. 처음에 모든 페이지들이 원본 호스트에서 타겟 호스트로 움직이며, 동시에 이전에 전송된 단계가 복사되는 동안 CPU 상태와 다른 남아있는 메모리 페이지들이 전송되는데 이 페이지들은 더티 페이지(Dirty Page)가 된다. 이후, 원본 호스트는 동작을 중지하고, 타겟 호스트가 가상 머신이 동작하는 주요 호스트가 된다. 실시간 마이그레이션은 타겟 호스트로 가상 머신이 이동하는 과정에서 실행중인 가상 머신이 종료되지 않고, 끊임없이 변화되는 메모리 페이지가 전송되기 때문에, 더 작은 서비스 다운타임(Downtime)을 가지게 된다.

[0004] 실시간 마이그레이션은 원본 호스트에서 타겟 호스트로 가상 머신이 마이그레이션 하는 과정에서, 원본 가상 머신 메모리의 데이터를 원본 가상 머신 실행 중에도 끊김 없이 이동시키기 위해 마이그레이션 슬롯을 필요로 한다. 만약 할당된 이동 슬롯의 최대 개수보다 많은 마이그레이션 요청이 수신되면, 수용 가능한 이동 슬롯의 개수가 부족하게 되어 가상 머신의 마이그레이션 요청이 거절되거나 대기 시간(Waiting)이 길어질 수 있다. 대기 시간을 줄이기 위해서는 마이그레이션 슬롯의 개수를 늘리는 것이 바람직하지만, 서버는 가용한 가상 단말의 개

수가 한정적이기 때문에, 마이그레이션 슬롯의 개수가 증가할수록 가용한 가상 단말의 개수가 줄어들게 된다. 이 경우, 새로운 서비스 제공에 대한 거절가능성이 높아지게 되기 때문에, 서비스 제공자의 이익이 감소하게 될 것이다. 따라서, 물리적 자원의 효과적인 사용을 위해서는 한정된 물리적 자원 내에서 최적의 자원을 할당하고, 이동 슬롯을 선택하는 것이 필요하다.

[0005] 대한민국 공개특허 제10-2013-0074952호는 클라우드 컴퓨팅 시스템의 자원을 효율적으로 사용하기 위한 가이드를 제공하는 장치 및 방법에 대한 발명으로, 가상 단말 할당 자원의 비용을 고려하여 최적화하는 기술을 포함하고 있다. 하지만, 이는 실행시간에 기초하여 추정된 메모리 사용량만을 고려할 뿐, 대기 시간 및 비용 문제는 고려되지 않았다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 공개특허 제10-2013-0074952호

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 과제는 가상 단말의 실시간 마이그레이션 과정에서 한정된 자원 내에서 대기 시간을 최소화하면서도 비용을 최소화 시킬수 있는 최적의 자원 할당 및 마이그레이션 슬롯을 선택하기 위한 자원 할당 방법에 관한 기술을 제공하는 것이다.

과제의 해결 수단

[0008] 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템은 하나 이상의 가상 단말을 할당하고, 마이그레이션된 서비스를 가상 단말에서 제공하는 타겟 서버, 하나 이상의 가상 단말을 할당하고, 마이그레이션 슬롯이 할당되며, 할당된 마이그레이션 슬롯을 통해 가상 단말에서 제공하는 서비스를 상기 타겟 서버로 마이그레이션하는 원본 서버 및 거절 가능성 및 대기 시간을 고려하여 예측 대기 시간을 산출하고, 산출된 예측 대기 시간 및 비용을 고려하여 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하여 원본 서버에 추가적인 슬롯 및 마이그레이션 슬롯을 할당하는 최적 비용 슬롯 할당 장치를 포함한다.

[0009] 그리고, 최적 비용 슬롯 할당 장치는 거절 가능성을 산출하고, 산출된 거절 가능성 및 대기 시간을 고려하여 예측 대기 시간을 산출하고, 산출된 예측 대기 시간의 임계값을 설정하며, 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출한다. 또한, 최적 비용 슬롯 할당 장치는 동시에 하나의 마이그레이션 슬롯만 허용된 제1 케이스 및 동시에 둘 이상의 마이그레이션 슬롯이 허용된 제2 케이스 각각에 대해 대기 시간을 산출하고, 거절 가능성의 임계값 범위 내에서 제1 케이스 및 제2 케이스의 대기 시간 및 거절 가능성을 고려하여 예측 대기 시간을 산출한다.

[0010] 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 방법은 먼저, 거절 가능성 및 대기 시간을 산출하고, 거절 가능성의 임계값을 설정한다. 그리고, 대기 시간 및 거절 가능성을 고려하여 예측 대기 시간을 산출한다. 예측 대기 시간이 산출되면, 예측 대기 시간의 임계값을 설정하고, 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출한다. 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출하는 단계는 설정된 예측 대기 시간의 임계값을 만족하며 최적 비용을 최소화하는 추가적인 슬롯 및 마이그레이션 슬롯 개수를 산출한다.

발명의 효과

[0011] 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템 및 할당 방법은 예측된 대기 시간과 비용을 함께 고려하여, 설정된 예측 대기 시간의 임계값을 만족하면서, 최적 비용을 최소화하기 때문에, 사용자에게 제공되는 서비스 품질을 만족시키면서, 동시에 소요 비용을 최소화할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 일 실시예를 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 최적 비용 슬롯 할당 장치의 서비스 전이도를 나타내는 도면이다.

도 3a는 본 발명에 따른 최적 비용 슬롯 할당 장치의 제1 케이스의 대기 시간을 설명하기 위한 도면이다.

도 3b는 본 발명에 따른 최적 비용 슬롯 할당 장치의 제2 케이스의 대기 시간을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 실험 결과를 나타내는 도면이다.

도 5a 및 도 5b는 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템 및 종래의 포아송 분포에 따른 추가 슬롯의 개수 및 비용을 비교한 결과를 나타내는 도면이다.

도 6은 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 방법을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0013] 이하, 본 발명의 실시예를 첨부된 도면들을 참조하여 상세하게 설명한다. 본 명세서에서 사용되는 용어 및 단어들은 실시예에서의 기능을 고려하여 선택된 용어들로서, 그 용어의 의미는 발명의 의도 또는 관례 등에 따라 달라질 수 있다. 따라서 후술하는 실시예에서 사용된 용어는, 본 명세서에 구체적으로 정의된 경우에는 그 정의에 따르며, 구체적인 정의가 없는 경우는 당업자들이 일반적으로 인식하는 의미로 해석되어야 할 것이다.

[0014] 도 1은 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 일 실시예를 나타내는 도면이다.

[0015] 도 1을 참조하면, 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템은 최적 비용 슬롯 할당 장치(100), 원본 서버(10) 및 타겟 서버(20)를 포함할 수 있다. 본 발명에 따른 최적 비용 슬롯 할당 장치(100)는 원본 서버(10)에서 서비스를 제공하고 있는 하나 이상의 가상 머신(Virtual Machine)을 타겟 서버(20)로 마이그레이션(Migration)한다. 원본 서버(10) 및 타겟 서버(20)는 특정한 서버로 한정되는 것은 아니며, 서비스를 마이그레이션하기 위한 가상 단말의 슬롯을 보유하고 있는 서버가 원본 서버(10)가 되고, 서비스를 제공하는 가상 단말의 슬롯이 마이그레이션하는 대상 서버가 타겟 서버(20)가 되며, 원본 서버(10) 및 타겟 서버(20)는 서로 교체될 수 있다.

[0016] 원본 서버(10)에서 한정된 연산 자원(Resource) 내에서 다수의 가상 머신의 슬롯(Slot)을 할당하여 사용자에게 서비스를 제공한다. 하지만, 원본 서버(10)의 연산 자원에 여유가 부족하거나, 사용자에게 보다 효과적으로 서비스를 제공할 수 있는 서버가 있는 경우, 해당 서버를 타겟 서버(20)로 하여 서비스를 마이그레이션한다. 최적 비용 슬롯 할당 장치(100)가 원본 서버(10)에서 타겟 서버(20)로 서비스를 마이그레이션하는 방법은 오프라인 마이그레이션과 실시간 마이그레이션으로 분류할 수 있다. 오프라인 마이그레이션은 마이그레이션에 소요되는 시간이 짧으나, 서비스를 제공하는 가상 머신이 정지된 상태로 마이그레이션이 수행되기 때문에 서비스가 끊기는 다운타임이 존재하게 된다. 반면에, 실시간 마이그레이션은 가상 머신이 실행중인 상태에서 마이그레이션을 수행하기 때문에 다운타임은 없으나, 소요시간이 증가하게 된다.

[0017] 최적 비용 슬롯 할당 장치(100)가 실시간 마이그레이션을 수행할 때, 마이그레이션을 수행할 가상 머신의 슬롯 개수가 늘어나면, 한번에 마이그레이션을 수행할 수 있는 서비스 요청이 늘어날 수 있으나, 타겟 서버(20)로 새로운 서비스 요청이 수신될 시 가용한 가상 머신의 슬롯이 부족하여 거절(Rejection)이 발생할 수 있다. 예를 들어, 원본 서버(10) 및 타겟 서버(20)의 슬롯 개수가 4개 할당되어 있고, 원본 서버(10)에서 총 네 개의 서비스를 제공하고 있을 때, 원본 서버(10)의 슬롯 4개 모두를 동시에 타겟 서버(20)로 마이그레이션 할 경우 타겟 서버(20)의 슬롯 4개 모두가 마이그레이션 상태에 놓이기 때문에, 타겟 서버(20)로 새로운 서비스 요청이 수신되면 거절이 발생할 수 있다. 반면에, 원본 서버(10)의 슬롯 중에서 2개씩 마이그레이션을 수행한다면, 마이그레이션 소요 시간 동안 타겟 서버(20)에 두 개의 슬롯에 여유가 있기 때문에, 새로운 서비스 요청이 타겟 서버(20)로 수신되어도 정상적으로 서비스가 가능하다. 하지만, 이 경우, 원본 서버(10)에 남아있는 두 개의 슬롯은 처음 두 개의 마이그레이션이 종료될 때까지 마이그레이션을 수행하지 못하고 대기 시간이 발생하게 된다.

[0018] 이와 같은 마이그레이션 과정에서의 대기 시간을 낮추기 위해서는 원본 서버(10)에 추가적인 가상 단말 슬롯을 할당하는 것이 바람직하다. 가상 단말 슬롯의 수가 많아지게 되면 대기 시간이 줄어들게 된다. 하지만, 동시에 추가 비용이 발생하게 된다. 추가 비용은 추가적인 슬롯에 대한 비용 및 재할당하는 과정 동안에 사용하는 원본 서버(10)의 비용이 될 수 있다. 따라서, 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 장치(100)는 대기 시간 및 거절 가능성을 고려하여 마이그레이션 슬롯(K)를 설정하는 과정에서 추가적인 슬롯에 대한

비용을 최소화할 수 있도록 거절가능성 및 예측 대기 시간과 최적 비용을 고려하여 마이그레이션 슬롯의 개수(K`)와 추가적인 가상 단말 슬롯의 개수 즉, 추가 슬롯(ΔN)을 설정한다.

[0019] 따라서, 본 발명에 따른 최적 비용 슬롯 할당 장치(100)는 동시에 마이그레이션을 수행할 수 있는 최적의 마이그레이션 슬롯 개수(K)를 설정한다. 최적 비용 슬롯 할당 장치(100)는 거절 가능성 및 대기 시간을 포아송 프로세스(Possion Process) 및 지수분포를 이용해 산출하고, 설정된 예측 대기 시간 이내에서 최소값을 가지도록 추가 슬롯(ΔN) 및 마이그레이션 슬롯 개수(K`)를 설정한다. 최적 비용 슬롯 할당 장치(100)가 추가 슬롯(ΔN) 및 최적의 마이그레이션 슬롯 개수 K`를 설정하는 과정은 후술하는 도 2 내지 도 3a/도 3b에서 도면과 함께 구체적으로 설명하도록 한다.

[0020] 도 2는 본 발명의 일 실시예에 따른 최적 비용 슬롯 할당 장치의 서비스 전이도를 나타내는 도면이다.

[0021] 도 2를 참조하면, 본 발명의 일 실시예에 따른 최적 비용 슬롯 할당 장치(100)의 서비스 전이도(Transition Diagram, 200)를 통해 서비스 마이그레이션 과정에서 거절 가능성을 산출한다. 서비스 전이도(200)는 데이터 센터 내의 가능한 흐름을 사용한 것이며, 이를 사용하여 거절 가능성과 대기 시간에 각 상태의 가능성을 기반으로 공식화한다. 서비스 전이도(200)에서, 가상 머신 서비스가 끝나면 전체 가상머신의 수는 1 감소하고, 새로운 요청이 수신되면 전체 가상머신 슬롯의 수는 1증가하게 된다. 마이그레이션이 끝나게 되면, 전체 가상머신과 마이그레이션 슬롯의 수는 1 감소하게 된다. 그리고, 마이그레이션 요청이 수신되면 전체 가상머신 슬롯과 마이그레이션 슬롯의 수는 1 증가하게 된다.

[0022] 서비스 전이도(200)를 구성하는 각각의 구성요소 (I,J)에서 I는 사용하고 있는 전체 슬롯 즉, 서비스를 제공중인 가상 단말 슬롯을 나타내며, J는 마이그레이션 하고 있는 슬롯을 나타낸다.

표 1

서비스 전이도(200)의 균형방정식

State	Rate that the process leaves = rate that it enters	Category
$(0,0)$	$\lambda_s P_{0,0} = \mu_s P_{1,0}$	1
$(N,0)$	$N\mu_s P_{N,0} = \lambda_s P_{N-1,0}$	2
$(i,0);$ $0 < i < N$	$(i\mu_s + \lambda_s + i\lambda_m)P_{i,0} =$ $\lambda_s P_{i-1,0} + (i+1)\mu_s P_{i+1,0} + \mu_m P_{i+1,1}$	3
$(2K,K);$ $2K = N$	$K\mu_m P_{N,K} = \lambda_m P_{N-1,K-1}$	4
$(N,j);$ $0 < j < \frac{N}{2}$	$((N-2j)\mu_s + j\mu_m)\,P_{N,j} = \lambda_s P_{N-1,j} + (N-2j+1)\lambda_m P_{N-1,j-1}$	5
$(i,j);$ $i \neq N, j < K,$ $0 < j < \frac{i}{2}$	$((i-2j)\mu_s + j\mu_m + \lambda_s + (i-2j)\lambda_m)\,P_{i,j} = (i-2j+1)\lambda_m P_{i-1,j-1}$ $+\lambda_s P_{i-1,j} + (i-2j+1)\mu_s P_{i+1,j} + (j+1)\mu_m P_{i+1,j+1}$	6
$(i,K);$ $0 < K < \frac{i}{2}$	$((i-2K)\mu_s + K\mu_m + \lambda_s)\,P_{i,K} =$ $(i-2K+1)\lambda_m P_{i-1,K-1} + \lambda_s P_{i-1,K} + (i-2K+1)\mu_s P_{i+1,K}$	7

[0023]

표 2

[0024] 변수 정의

변수명	설명	
K	마이그레이션 과정동안 제한할 VM 수(마이그레이션 슬롯 개수)	
N	현재 사용되는 전체 자원(가용 전체 슬롯 개수)	
λ_s	서비스 발생 비율(새로운 사용자 도달률)	포아송 프로세스
λ_m	마이그레이션 서비스 발생 비율	
μ_s	평균서비스 시간	지수분포

μ_m	평균 마이그레이션 시간
Th_R	거절 가능성이 만족해야할 임계값
$E[T^D]$	예측 대기 시간(지수분포에 근거)

[0025] 표 1은 서비스 전이도(200)의 균형방정식을 나타낸다. 안정 상태를 가정하여 rate in = rate out(안정 상태: 시스템이 한 상태를 떠나면 언젠가는 돌아온다는 안정 상태를 가정) 균형 방정식을 만들 수 있다. 상태 (0,0)을 나타내는 카테고리(category) 1(201)에서 서비스 발생 비율(λ_s)과 (0,0)에서의 거절가능성($P_{0,0}$)의 곱은 평균 서비스 시간(μ_s)과 (1,0)에서의 거절가능성의 곱과 같다. 카테고리 1(201)이 나타내는 상태 (0,0)은 서비스를 제공중인 가상 단말 슬롯 및 마이그레이션 하고 있는 슬롯 모두 0인 상태, 즉 어떤 서비스 및 마이그레이션도 수행하고 있지 않은 상태를 나타낸다. 그리고, 상태 (1,0)은 서비스를 제공중인 가상 단말 슬롯이 1개이고 마이그레이션 하고 있는 슬롯은 0인 상태로서, 상태 (0,0)에서 (1,0)으로 변화된 상태는 하나의 서비스가 발생한 것을 나타낸다. 그리고, 카테고리 2의 상태(N,0)에서 N λ_s $P_{N,0}$은 μ_s $P_{N-1,0}$과 동일하다.

[0026] 상태 (1,0)에서 상태 (2,1)로 변화되는 것은 하나의 가상 단말 슬롯을 통해 서비스를 제공하는 중에 마이그레이션이 발생한 것으로, 마이그레이션 또한 서비스이므로 서비스가 1 증가한다. 이를 기반으로 상태 (3,1)의 거절 가능성을 계산하면, 수학식 1과 같다.

수학식 1

[0027]
$$
\begin{aligned}
&(0,0)\ \lambda_s P_{0,0} = \mu_s P_{1,0},\\
&(1,0)\ (\mu_s + \lambda_s + \lambda_m) P_{1,0} = \lambda_s P_{0,0} + 2\mu_s P_{2,0} + \mu_m P_{2,1},\\
&(2,0)\ (2\mu_s + \lambda_s + 2\lambda_m) P_{2,0} = \lambda_s P_{1,0} + 3\mu_s P_{3,0} + \mu_m P_{3,1},\\
&(3,0)\ 3\mu_s P_{3,0} = \lambda_s P_{2,0},\\
&(2,1)\ (\mu_m + \lambda_s) P_{2,1} = \lambda_m P_{1,0} + \mu_s P_{3,1},\\
&(3,1)\ (\mu_s + \mu_m) P_{3,1} = \lambda_s P_{2,1} + 2\lambda_m P_{2,0},
\end{aligned}
$$

[0028] 수학식 1은 도 1의 전이도 및 표 1에 기초하여 상태 (3,1)의 거절가능성을 산출하는 수식을 나타낸다. 수학식 1에서, λ_s=0.5, λ_m=1, μ_s=0.2, μ_m=0.1로 주어질 때, (3,1) 상태의 거절 가능성을 개산하면, $P_{0,0}$=0.0103, $P_{1,0}$=0.0258, $P_{2,0}$=0.0323, $P_{3,0}$=0.0269, $P_{2,1}$=0.2585 및 $P_{3,1}$=0.6461로 산출될 수 있다. 상술한 각 상태의 거절 가능성에 기초하여 거절 가능성은 수학식 2와 같이 나타낼 수 있다.

수학식 2

[0029]
$$
P_r(R) = \sum_{j=0}^{K} P_{N,j}
$$

[0030] 수학식 2에서, 거절 가능성 Pr(R)은 각 상태의 거절 가능성의 합산 형태로 나타내어 진다. 예를 들어, 서비스를 제공하는 가상 머신 슬롯이 3(N=3)이고, 마이그레이션 슬롯 개수가 1(K=1)이라면, 거절 가능성 Pr(R)은 $P_{3,0}+P_{3,1}$=0.0269+0.6461=0.673으로 산출될 수 있다.

[0031] 도 3a는 본 발명에 따른 최적 비용 슬롯 할당 장치의 제1 케이스의 대기 시간을 설명하기 위한 도면이다.

[0032] 도 3a를 참조하면, 본 발명에 따른 최적 비용 슬롯 할당 장치는 마이그레이션 과정 동안 충분한 슬롯을 가지고 있지 않으면 대기 시간이 발생한다. 대기 시간 발생은 할당된 마이그레이션 슬롯이 모두 사용중이거나, 원본 서

버(10)의 가용 가상 머신 슬롯이 모두 사용 중일 경우 발생하게 된다. 할당된 마이그레이션 슬롯이 모두 사용중인 경우, 상태는 (2K+1,K), (2K+2,K),…,(N,K)가 된다. 그리고, 원본 서버(10)의 가용 가상 머신 슬롯이 모두 사용 중일 경우 상태는 (N,0), (N,1),…,(N,K-1)이 된다.

[0033] 제1 케이스는 동시에 하나의 마이그레이션 슬롯만 허용된 경우로, 총 4명의 사용자(User)가 원본 서버(10)에 접속하여 있으며, 원본 서버(10)는 가용 가상 머신 슬롯이 5개(N=5)이고, 마이그레이션 슬롯 개수가 1개(K=1) 일 때, 사용자 3(User 3) 및 사용자 4(User 4)가 마이그레이션을 요청한 경우를 나타낸다. 도 3a에서 ts는 서비스 시간, tm은 마이그레이션 시간을 나타내며, S^s는 각 사용자의 서비스 요청 시간을 나타낸다. 사용자 3이 마이그레이션 요청 시간(S_1^m)이후 마이그레이션을 시작하여 마이그레이션 시간(301) 동안 마이그레이션을 수행한다.

[0034] 사용자 4는 마이그레이션 요청 시간(S_2^m)이후 마이그레이션을 요청하지만, 제1 케이스에서는 K=1이기 때문에 마이그레이션을 바로 수행하지 못하고 대기 시간(T, 302)가 발생하게 된다. 또한, 사용자 3이 마이그레이션을 수행하는 도중에 사용자 1의 서비스가 종료되지만, 마이그레이션 슬롯의 개수가 1개이기 때문에, 사용자 4에 대한 마이그레이션이 시작될 수 없다. 사용자 4는 대기 상태를 유지하다 사용자 3의 마이그레이션이 끝난 시점에 사용자 4에 대한 마이그레이션이 수행된다. 따라서, 제1 케이스(N=5,K=1)의 사용자 4의 대기 시간(302)은 수학식 3과 같이 산출된다.

수학식 3

[0035] $$T_{i,k}^{D} = E[S_1^m] + E[t_m] - E[S_{K+1}]$$

[0036] $$2K < i \leq N$$

수학식 4

[0037] $$E[S_i^s] = \frac{i}{\lambda_s}, E[S_i^m] = \frac{i}{\lambda_m}$$

[0038] 수학식 3에서 $T_{i,k}^{D}$는 제1 케이스의 대기 시간을 나타내고, $E[S_1^m]$는 첫 번째 마이그레이션 요청 사용자(사용자 3)의 마이그레이션 요청 시간을 나타내며, $E[t_m]$는 사용자 3의 마이그레이션 요청시간이며, 일반화하여 수학식 3을 도출하여 마이그레이션 요청 사용자 K의 요청 시간을 나타낼 수 있다. 마이그레이션 시간을 나타내고, $E[S_{K+1}]$는 두 번째 마이그레이션 요청 사용자(사용자 4)의 마이그레이션 요청 시간을 나타내며, 일반화를 통해 K+1번째 요청자에 의한 요청 서비스 시간을 나타낼 수 있다. 즉, 제1 케이스의 대기 시간은 사용자 3의 마이그레이션 요청 시간과 마이그레이션 시간에서 사용자 4의 마이그레이션 요청 시간을 제거하여 산출될 수 있다. 수학식 4는 지수 분포를 이용하여 산출된 사용자 i에 대한 평균 서비스 시간 및 마이그레이션 시간을 나타낸다. 지수 분포를 갖는 동일한 두 사건 중에서 어느 것이든지 먼저 사건이 일어날 때까지의 시간 분포도 역시 지수시간분포를 가지며 평균 서비스 시간은 지수 시간의 1/2이 되기 때문에 지수분포에 의해 사용자 i에 대한 평균 서비스 시간 및 마이그레이션 시간은 위와 같다. 수학식 3 및 수학식 4를 정리하면, 수학식 5와 같이 나타난다.

수학식 5

[0039]
$$T_{i,K}^{D} = \frac{1}{\lambda_m} + \frac{1}{\mu_m} - \frac{K+1}{\lambda_m} = \frac{\lambda_m - K\mu_m}{\lambda_m \mu_m}$$

[0040]
$$T_{i,K}^{D} = \begin{cases} \dfrac{\lambda_m - K\mu_m}{\lambda_m \mu_m}, & 0 \le K \le \rho_m, \\ 0, & \rho_m < K \le N \end{cases}$$

[0041] $\rho_m = \lambda_m / \mu_m$ 수학식 5는 제1 케이스일 경우 대기 시간을 산출하기 위한 수학식이다. 제1 케이스의 경우 최종적으로 정리된 수학식 5를 통해 대기 시간을 산출할 수 있다. 수학식 5에서 ρ_m는 서비스 이용률을 나타낸다.

[0042] 도 3b는 본 발명에 따른 최적 비용 슬롯 할당 장치의 제2 케이스의 대기 시간을 설명하기 위한 도면이다.

[0043] 도 3b를 참조하면, 제2 케이스는 총 4명의 사용자(User)가 원본 서버(10)에 접속하여 있으며, 원본 서버(10)는 가용 가상 머신 슬롯이 5개(N=5)이고, 마이그레이션 슬롯 개수가 2개(K=2)일 때, 사용자 3(User 3) 및 사용자 4(User 4)가 마이그레이션을 요청한 경우를 나타낸다.

[0044] 사용자 3이 마이그레이션 요청 시간(S_1^m)이후, 마이그레이션을 시작하여 마이그레이션 시간(tm) 동안 마이그레이션을 수행한다. 사용자 4는 마이그레이션 요청 시간(S_2^m)이후 마이그레이션을 요청하지만, 총 5개의 가용 가상 머신 슬롯이 5개이고, 사용자가 4명이기 때문에, 마이그레이션을 수행할 수 있는 가상 머신 슬롯은 한 개이다. 하지만, 한 개의 가상 머신 슬롯은 사용자 3이 마이그레이션하고 있기 때문에, 사용자 4는 마이그레이션을 수행하지 못하고 대기 시간을 가지게 된다.

[0045] 303에서는 사용자 3의 마이그레이션이 끝난 이후, 사용자 4의 마이그레이션이 수행된다. 반면에, 304에서는 사용자 3의 마이그레이션이 끝나기 전에, 사용자 1의 서비스가 종료된다. 따라서, 사용자 1이 사용하던 가상 머신 슬롯에 여유가 생겨 사용자 3의 마이그레이션이 종료 여부와 관계없이 사용자 1의 서비스 종료 이후 사용자 4의 마이그레이션을 수행할 수 있다. 이를 통해, 제2 케이스의 대기 시간을 산출할 수 있다.

수학식 6

[0046]
$$\begin{aligned} T_{N,j}^{D} &= \Pr(Y_m \cdot (E[S_1^m] + t_m - E[S_{j+1}^m]) + \Pr(Y_s) \cdot E[S_1^s] + t_s - E[S_{j+1}^m] \\ &= P(t_s \ge t_m + E[S_1^m] + E[S_{j+1}^m] \cdot (\frac{1}{\lambda_s} + \frac{1}{\mu_s} - \frac{j+1}{\lambda m}) \\ &= e^{\mu_s(\lambda_m - \mu_m)/\mu_m\lambda_m}(\frac{\lambda_m - j\mu_m}{\mu_m\lambda_m}) + (1 - e^{-\mu_s(\lambda_m - \mu_m)/\mu_m\lambda_m})(\frac{1}{\lambda_s} + \frac{1}{\mu} - \frac{j+1}{\lambda_m}) \end{aligned}$$

[0047] 수학식 6은 제2 케이스의 대기 시간을 산출하기 위한 수학식이다. 수학식 6에서 j는 사용자를 나타낸다. 제1 케이스의 대기 시간을 산출하는 수학식 5 및 제2 케이스의 대기 시간을 산출하는 수학식 6과 거절 가능성을 산출하는 수학식 2를 고려하여 본 발명에 따른 최적 비용을 고려한 최적 비용을 고려한 마이그레이션 자원 할당 장치(100)의 예측 대기 시간을 산출할 수 있다.

[0048] 제 2 케이스의 대기시간의 경우는 4명의 유저가 요청을 했을 때 현재 총 슬롯은 5개이고 마이그레이션 할 수 있는 슬롯은 2개이다. 요청은 유저에 따른 서비스가 4개이고 마이그레이션 서비스는 2개이다. 요청에 따른 총 필요한 슬롯의 수는 6개이나 주어진 총 슬롯은 5개이므로 이 사이에 대기시간이 발생하게 된다. 303의 경우, 서비

스 4개의 요청이 슬롯에 할당되고 마이그레이션 서비스 요청슬롯은 하나만 시작하게 된다. 마이그레이션 요청하나가 먼저 끝나게 되어 기다렸다가 마이그레이션을 시작하게 된다. 304의 경우는 서비스 4개 요청이 슬롯에 할당되고 마이그레이션 서비스 요청슬롯은 하나만 시작하게 된다. 이 경우에는 사용자 1의 서비스가 먼저 끝나게 되어 끝날때까지 기다리다가 마이그레이션을 시작하게 된다. 제2 케이스의 경우에는 수학식 7과 같이 303(a) 및 304(b)의 경우를 합하여 대기 시간을 산출할 수 있다.

수학식 7

[0049]

$$\underbrace{\Pr(Y_{\mathrm{m}}).(E[S_1^{\mathrm{m}}]+t_{\mathrm{m}}-E[S_{j+1}^{m}])}_{(a)} + \underbrace{\Pr(Y_{\mathrm{s}}).(E[S_1^{s}]+t_{\mathrm{s}}-E[S_{j+1}^{m}])}_{(b)}$$

수학식 8

[0050]

$$E[T^D]=\underbrace{\sum_{i=2K+1}^{N} P_{i,K}T_{i,K}^{D}}_{case\,1}+\underbrace{\sum_{i=0}^{K-1} P_{N,j}T_{N,j}^{D}}_{case2}$$

[0051] 수학식 5 및 수학식 7을 정리하여, 제1 케이스 및 제2 케이스를 모두 포함하는 최적 비용 슬롯 할당 장치(100)의 예측 대기 시간을 수학식 8의 $E[T^D]$ 와 같이 산출할 수 있다.

[0052] 본 발명에 따른 최적 비용 슬롯 할당 장치(100)에서는 예측 대기 시간뿐만 아니라, 마이그레이션 과정 동안의 비용을 고려한다. 마이그레이션 과정 동안의 비용은 추가적인 슬롯에 대한 비용 및 재할당하는 과정 동안에 사용하는 서버의 비용을 포함할 수 있다. 재할당하는 과정 동안에 사용하는 서버의 비용은 재할당하는 과정에서도 비용이 소모되고 원래의 원본 서버의 슬롯도 동작하고 있기 때문에, 비용이 소모되며 재할당하는 과정 동안에도 서버의 비용이 소모된다. 따라서, 서버의 비용을 출이기 위해서는 마이그레이션 과정의 대기 시간을 줄여야 한다. 마이그레이션 슬롯의 수가 많아지게 되면 대기 시간 및 거절가능성이 줄어들게 되지만, 비용이 커지게 된다. 따라서, 둘 사이의 균형을 맞출 필요가 있다. 수학식 9는 이와 같은 비용과 대기 시간 사이의 균형을 고려하여 최적 비용을 산출하기 위한 수학식이다.

수학식 9

[0053]

$$\min_{\Delta N,K'} \quad \Delta N\tau\Omega+E'\left[T^D\right]\delta\Omega$$

[0054]

$$E'\left[T^D\right]\le Th_W$$

[0055]

$$E'[T^D]=\sum_{i=2K'+1}^{N'} P_{i,K'}T_{i,K'}^{D}+\sum_{j=0}^{K'-1} P_{N',j}T_{N',j}^{D}.$$

[0056] 수학식 9에서 Th_W는 대기 시간의 임계값을 나타내고, $E'[T^D]$는 수학식 8에 따른 예측 대기 시간을 나타내며, ΔN은 추가되는 마이그레이션 슬롯 개수를 나타내며, Ω은 하나의 슬롯당 비용을 나타내고, τ는 마이그레이션 과정 동안 추가적인 슬롯을 사용하는데 필요한 시간을 나타내며, K'는 비용을 고려하여 마이그레이션 과정 동안 제안할 마이그레이션 슬롯 개수를 나타내고, N`은 추가적인 슬롯 ΔN과 현재 사용되는 전체 자원의 합을 나타낸다. 먼저, 예측 대기 시간의 임계값을 설정하고, 설정된 예측 대기 시간의 임계값을 만족하는 결과를 산출한다.

표 3

[0057] 최적비용 결정 알고리즘

알고리즘 : 최적비용

Require: $N, \lambda_s, \lambda_m, \mu_s, \mu_m, \tau, \delta, Th_R, Th_W, \Omega$
  $N' = N + 1$;
  $N'^* = N'$;
  Determine K';
  $K'^* = K'$;
  Find $E'[T^D]$;
  OptimalCost = $\Delta N \tau \Omega + E'[T^D]\,\delta\Omega$;
  while ($E'[T^D] \simeq 0$) **do**
    $N' = N' + 1$;
    Determine K';
    Find $E'[T^D]$;
    Cost = $\Delta N \tau \Omega + E'[T^D]\,\delta\Omega$;
    if Cost < OptimalCost **then**
      $N'^* = N'$;
      $\Delta N^* = N' - N$;
      $K'^* = K'$;
      OptimalCost = Cost;
    end if
  end while
  return $\Delta N^*, K'^*$,OptimalCost.

[0058] 추가되는 마이그레이션 슬롯 개수인 ΔN을 증가시킨다면, 예측 대기시간이 감소하지만, 동시에 비용이 증가하게 된다. 따라서, 표 3의 최적비용 결정 알고리즘을 이용하여 전체 비용을 최소화하는 ΔN 및 K'을 산출한다. 표 3의 최적비용 결정 알고리즘은 먼저 ΔN의 값을 1부터 반복적으로 증가($N'=N+1$)시키면서, 수학식 9를 산출하여, 대기 시간의 임계값을 만족하는 최적비용을 결정하고, 최적비용에 따른 ΔN 및 K'을 산출한다. 즉, 대기 시간의 임계값을 만족하면서, 동시에 수학식 9를 최소화하는 추가되는 마이그레이션 슬롯 개수 및 비용을 고려하여 마이그레이션 과정 동안 제안할 마이그레이션 슬롯 개수를 산출한다.

[0059] 이 과정에서, $E'[T^D] \simeq 0$는 정지 상태(Stop Condition)를 나타낸다. $E'[T^D] \simeq 0$는 대기 시간이 0 또는 0에 근접한 상태이기 때문에, 추가되는 마이그레이션 슬롯이 불필요하여 더 적은 비용을 얻을 수 있다. 그리고, ΔN은 전체 슬롯의 개수 N'과 초기 전체 슬롯의 개수 N의 차를 통해 산출될 수 있다.

[0060] 도 4는 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 실험 결과를 나타내는 도면이다.

[0061] 도 4를 참조하면, 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 슬롯 할당 시스템에 표 4의 실험 조건을 적용하여 실험 결과를 산출한다.

표 4

[0062]

현재 사용되는 전체 자원	N=100
서비스 발생 비율(새로운 사용자 도달율)	λ_s=0.1
마이그레이션 서비스 발생 비율	λ_m=0.05
평균 서비스 시간	μ_s=0.001(17mins)
평균 마이그레이션 시간	μ_m=0.025(40second)
거절 가능성의 임계값	Th_R=15%

[0063] 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 슬롯 할당 시스템의 실험 결과는 산출된 마이그레이션 대기 시간 그래프(410) 및 산출된 비용 그래프(420)를 포함한다. λ_s=0.1일때, ThR ≤ 15%, ThW ≤ 3 s를 만족시키는 경우, ΔN이 6이면서 최적 비용 3700을 얻을 수 있다. 전체 슬롯수(N`)가 증가함에 따라 비용이 증가하는 반면에, 전체 대기 시간(410)이 감소하는 것을 확인할 수 있다. ΔN이 10일 경우에서 비교해 봤을 때, 비용은 4300이므로 본 발명은 비용을 20%정도 절약할 수 있는 것으로 나타났다.

[0064] 도 5a 및 도 5b는 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템 및 종래의 포아송 분포에 따른 추가 슬롯의 개수 및 비용을 비교한 결과를 나타내는 도면이다.

[0065] 도 5a 및 도 5b를 참조하면, 포아송(Poisson)dl 10일 때, 예측 대기 시간의 임계값을 증가시켜, 예측 대기 시간의 임계값의 변화에 대한 슬롯의 증가 및 비용의 효과 사이의 관계를 실험 결과로 확인하였다. 예측 대기 시간의 임계값 Th_W 및 추가 슬롯의 개수 ΔN의 관계를 나타내는 그래프(510)를 살펴보면, 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 추가 슬롯 개수 결과(511)는 예측 대기 시간의 임계값이 10초 일 때, 추가 슬롯 개수가 10에서 2로 변경된다. 예측 대기 시간의 임계값 Th_W 및 비용 Ω의 관계를 나타내는 그래프(520)를 살펴보면, 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 시스템의 비용 결과(521)는 예측 대기 시간의 임계값이 10초 일 때, 비용이 4250에서 2100으로 약 50%가 줄어들었다. 이와 같이 불필요한 슬롯을 줄여 비용감소 효과를 볼 수 있다. 넓은 의미에서 알고리즘은 서버 내에서의 요구를 만족하는 슬롯의 개수의 증가를 결정할 수 있으며, 여분의 슬롯에 필요한 비용과 에너지를 절감할 수 있다.

[0066] 도 6은 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 방법을 나타내는 흐름도이다.

[0067] 도 6을 참조하면, 본 발명의 일 실시예에 따른 최적 비용을 고려한 마이그레이션 자원 할당 방법은 먼저, 마이그레이션 과정에서의 거절 가능성을 산출한다(601). 도 2의 서비스 전이도를 통해 서비스 마이그레이션 과정에서의 거절 가능성을 산출할 수 있다. 표 1의 서비스 전이도의 균형 방정식에서 상태 (0,0)을 나타내는 카테고리(category) 1(201)에서 서비스 발생 비율(λ_s)과 (0,0)에서의 거절가능성($P_{0,0}$)의 곱은 평균서비스 시간(μ_s)과 (1,0)에서의 거절가능성의 곱과 같다. 카테고리 1(201)이 나타내는 상태 (0,0)은 서비스를 제공중인 가상 단말 슬롯 및 마이그레이션 하고 있는 슬롯 모두 0인 상태, 즉 어떤 서비스 및 마이그레이션도 수행하고 있지 않은 상태를 나타낸다. 그리고, 상태 (1,0)은 서비스를 제공중인 가상 단말 슬롯이 1개이고 마이그레이션 하고 있는 슬롯은 0인 상태로서, 상태 (0,0)에서 (1,0)으로 변화된 상태는 하나의 서비스가 발생한 것을 나타낸다. 그리고, 카테고리 2의 상태(N,0)에서 $N\lambda_s P_{N,0}$은 $\mu_s P_{N-1,0}$과 동일하다.

[0068] 상태 (1,0)에서 상태 (2,1)로 변화되는 것은 하나의 가상 단말 슬롯을 통해 서비스를 제공하는 중에 마이그레이션이 발생한 것으로, 마이그레이션 또한 서비스이므로 서비스가 1 증가한다. 이를 통해, 거절 가능성은 수학식 2와 같이 나타낼 수 있다. 수학식 2에서, 거절 가능성 Pr(R)은 각 상태의 거절 가능성의 합산 형태로 나타내어진다.

[0069] 다음으로 제1 케이스 및 제2 케이스의 대기 시간을 산출한다(602). 본 발명에 따른 최적 비용 슬롯 할당 장치(100)는 마이그레이션 과정 동안 충분한 슬롯을 가지고 있지 않으면 대기 시간이 발생한다. 대기 시간 발생은 할당된 마이그레이션 슬롯이 모두 사용중이거나, 원본 서버(10)의 가용 가상 머신 슬롯이 모두 사용 중일 경우 발생하게 된다.

[0070] 제1 케이스는 한번에 하나의 마이그레이션 슬롯 만을 허용하는 경우이다. 하나의 실시예로서, 제1 케이스는 총 4명의 사용자(User)가 원본 서버(10)에 접속하여 있으며, 원본 서버(10)는 가용 가상 머신 슬롯이 5개(N=5)이고, 마이그레이션 슬롯 개수가 1개(K=1) 일때, 사용자 3(User 3) 및 사용자 4(User 4)가 마이그레이션을 요청한 경우를 나타낸다. 도 3a에서 ts는 서비스 시간, tm은 마이그레이션 시간을 나타내며, S^s는 각 사용자의 서비스 요청 시간을 나타낸다. 사용자 3이 마이그레이션 요청 시간(S_1^m)이후 마이그레이션을 시작하여 마이그레이션 시간(tm) 동안 마이그레이션을 수행한다.

[0071] 사용자 4는 마이그레이션 요청 시간(S_2^m)이후 마이그레이션을 요청하지만, 제1 케이스에서는 K=1이기 때문에 마이그레이션을 바로 수행하지 못하고 대기 시간(T)가 발생하게 된다. 또한, 사용자 3이 마이그레이션을 수행하는 도중에 사용자 1의 서비스가 종료되지만, 마이그레이션 슬롯의 개수가 1개이기 때문에, 사용자 4에 대한 마이그레이션이 시작될 수 없다. 사용자 4는 대기 상태를 유지하다 사용자 3의 마이그레이션이 끝난 시점에 사용자 4에 대한 마이그레이션이 수행된다. 따라서, 제1 케이스의 사용자 4의 대기 시간은 수학식 3과 같이 산출된다. 제1 케이스의 대기 시간은 사용자 3의 마이그레이션 요청 시간과 마이그레이션 시간에서 사용자 4의 마이그레이션 요청 시간을 제거하여 산출될 수 있다. 수학식 4는 지수 분포를 이용하여 산출된 사용자 i에 대한 평균 서비스 시간 및 마이그레이션 시간을 나타낸다. 지수 분포를 갖는 동일한 두 사건 중에서 어느 것이든지 먼저 사건이 일어날 때까지의 시간 분포도 역시 지수시간분포를 가지며 평균 서비스 시간은 지수 시간의 1/2이 되기 때문에 지수분포에 의해 사용자 i에 대한 평균 서비스 시간 및 마이그레이션 시간은 위와 같다. 수학식 3 및 수학식 4를 정리하면, 수학식 5와 같이 나타난다.

[0072] 제2 케이스는 동시에 둘 이상의 마이그레이션 슬롯을 허용하는 경우이다. 하나의 실시예로서, 제2 케이스는 총 4명의 사용자(User)가 원본 서버(10)에 접속하여 있으며, 원본 서버(10)는 가용 가상 머신 슬롯이 5개(N=5)이고, 마이그레이션 슬롯 개수가 2개(K=2)일 때, 사용자 3(User 3) 및 사용자 4(User 4)가 마이그레이션을 요청한 경우를 나타낸다. 사용자 3이 마이그레이션 요청 시간(S_1^m)이후, 마이그레이션을 시작하여 마이그레이션 시간(tm) 동안 마이그레이션을 수행한다. 사용자 4는 마이그레이션 요청 시간(S_2^m)이후 마이그레이션을 요청하지만, 총 5개의 가용 가상 머신 슬롯이 5개이고, 사용자가 4명이기 때문에, 마이그레이션을 수행할 수 있는 가상 머신 슬롯은 한 개이다. 하지만, 한 개의 가상 머신 슬롯은 사용자 3이 마이그레이션하고 있기 때문에, 사용자 4는 마이그레이션을 수행하지 못하고 대기 시간을 가지게 된다. 301에서는 사용자 3의 마이그레이션이 끝난 이후, 사용자 4의 마이그레이션이 수행된다. 반면에, 302에서는 사용자 3의 마이그레이션이 끝나기 전에, 사용자 1의 서비스가 종료된다. 따라서, 사용자 1이 사용하던 가상 머신 슬롯에 여유가 생겨 사용자 3의 마이그레이션이 종료 여부와 관계없이 사용자 1의 서비스 종료 이후 사용자 4의 마이그레이션을 수행할 수 있다. 이를 통해, 제2 케이스의 대기 시간을 산출할 수 있다. 이는 상술한 수학식 6으로 표현될 수 있다.

[0073] 다음으로, 거절 가능성의 임계값을 설정한다(603). 서비스를 마이그레이션 하는 과정에서 마이그레이션 슬롯의 개수가 늘어날수록 대기 시간은 줄어들 수 있으나, 동시에 거절 가능성이 증가하게 된다. 따라서, 본 발명에 따른 최적 비용을 고려한 마이그레이션 자원 할당 방법에서는 대기 시간과 거절 가능성 사이의 협상을 통해 최적의 마이그레이션 슬롯의 개수를 산출한다. 따라서, 협상 조건으로 거절 가능성의 임계값을 설정한다. 거절 가능성의 임계값을 낮게 설정한다면 거절 가능성은 낮아지지만 대기 시간이 증가하게 되고, 거절 가능성의 임계값을 높게 설정한다면 거절 가능성은 높아지지만 대기 시간은 감소하게 된다. 따라서, 서비스 품질(QoS)이나, 제공하는 서비스의 종류에 따라 거절 가능성의 임계값을 설정하게 된다.

[0074] 다음으로, 대기 시간 및 거절 가능성을 고려하여 예측 대기 시간을 산출한다(604). 상술한 601 단계 내지 603 단계에서 산출된 제1 케이스 및 제2 케이스의 대기 시간과 거절 가능성을 고려하여 예측 대기 시간을 산출한다. 예측 대기 시간을 산출하는 공식은 상술한 수학식 7로 표현될 수 있다. 수학식 7에서, $E[T^D]$ 는 제1 케이스 및 제2 케이스를 모두 포함하는 최적 비용 슬롯 할당 장치(100)의 대기 시간을 나타낸다.

[0075] 다음으로, 예측 대기 시간의 임계값을 설정(605)하고, 설정된 예측 대기 시간의 임계값 및 비용을 고려하여 추가적인 슬롯 개수 및 마이그레이션 슬롯 개수를 결정한다(606). 본 발명에 따른 최적 비용 슬롯 할당 장치(100)에서는 예측 대기 시간뿐만 아니라, 마이그레이션 과정 동안의 비용을 고려한다. 마이그레이션 과정 동안의 비용은 추가적인 슬롯에 대한 비용 및 재할당하는 과정 동안에 사용하는 서버의 비용을 포함할 수 있다. 재할당하는 과정 동안에 사용하는 서버의 비용은 재할당하는 과정에서도 비용이 소모되고 원래의 원본 서버의 슬롯도 동작하고 있기 때문에, 비용이 소모되며 재할당하는 과정 동안에도 서버의 비용이 소모된다. 따라서, 서버의 비용을 줄이기 위해서는 마이그레이션 과정의 대기 시간을 줄여야 한다. 마이그레이션 슬롯의 수가 많아지게 되면 대기 시간 및 거절가능성이 줄어들게 되지만, 비용이 커지게 된다. 따라서, 둘 사이의 균형을 맞출 필요가 있다. 이와 같은 비용과 대기 시간 사이의 균형을 고려하여 최적 비용을 산출하기 위한 수학식은 상술한 수학식 9와 같이 산출될 수 있다.

[0076] 추가되는 마이그레이션 슬롯 개수인 ΔN 을 증가시킨다면, 예측 대기시간이 감소하지만, 동시에 비용이 증가하게 된다. 따라서, 상술한 표 3의 최적비용 결정 알고리즘을 이용하여 전체 비용을 최소화하는 ΔN 및 K' 을 산출한다. 표 3의 최적비용 결정 알고리즘은 먼저 ΔN 의 값을 1부터 반복적으로 증가($N'=N+1$)시키면서, 수학식 9를 산출하여, 대기 시간의 임계값을 만족하는 최적비용을 결정하고, 최적비용에 따른 ΔN 및 K' 을 산출한다. 즉, 대기 시간의 임계값을 만족하면서, 동시에 수학식 9를 최소화하는 추가되는 마이그레이션 슬롯 개수 및 비용을 고려하여 마이그레이션 과정 동안 제안할 마이그레이션 슬롯 개수를 산출한다.

[0077] 이 과정에서, $(\mathrm{E}'[T^D] \cong 0)$ 는 정지 상태(Stop Condition)를 나타낸다. $(\mathrm{E}'[T^D] \cong 0)$ 는 대기 시간이 0 또는 0에 근접한 상태이기 때문에, 추가되는 마이그레이션 슬롯이 불필요하여 더 적은 비용을 얻을 수 있다. 그리고, ΔN 은 전체 슬롯의 개수 N'과 초기 전체 슬롯의 개수 N의 차를 통해 산출될 수 있다.

[0078] 상술한 내용을 포함하는 본 발명은 컴퓨터 프로그램으로 작성이 가능하다. 그리고 상기 프로그램을 구성하는 코드 및 코드 세그먼트는 당분야의 컴퓨터 프로그래머에 의하여 용이하게 추론될 수 있다. 또한, 상기 작성된 프로그램은 컴퓨터가 읽을 수 있는 기록매체 또는 정보저장매체에 저장되고, 컴퓨터에 의하여 판독되고 실행함으로써 본 발명의 방법을 구현할 수 있다. 그리고 상기 기록매체는 컴퓨터가 판독할 수 있는 모든 형태의 기록매체를 포함한다.

[0079] 이상 바람직한 실시예를 들어 본 발명을 상세하게 설명하였으나, 본 발명은 전술한 실시예에 한정되지 않고, 본 발명의 기술적 사상의 범위 내에서 당분야에서 통상의 지식을 가진자에 의하여 여러 가지 변형이 가능하다.

부호의 설명

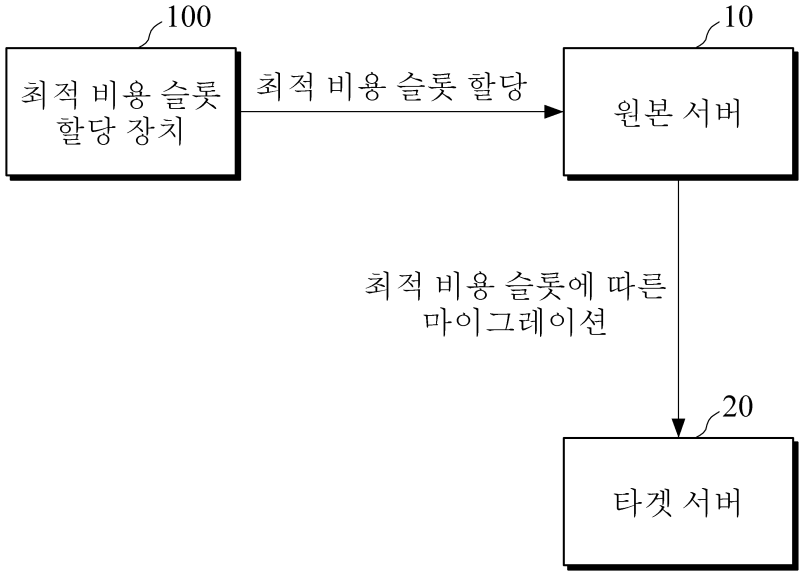
[0080] 10: 원본 서버

20: 타겟 서버

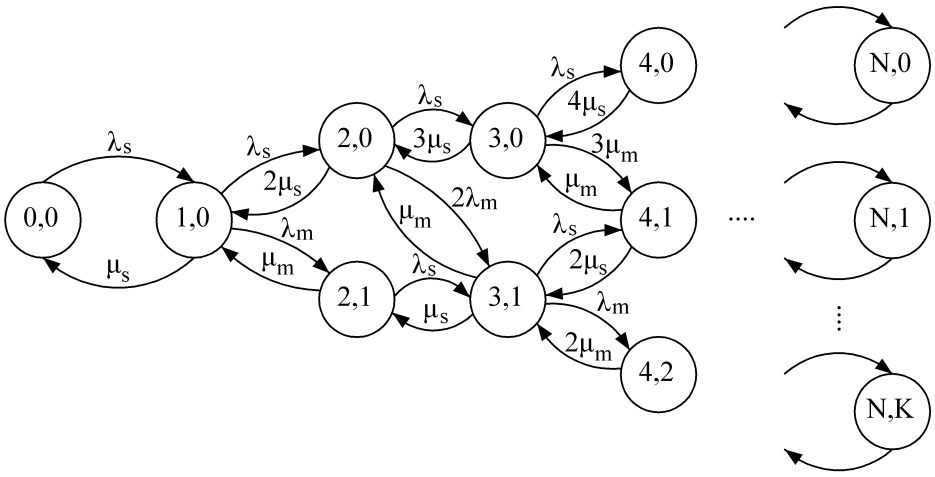
100: 최적 비용 슬롯 할당 장치

도면

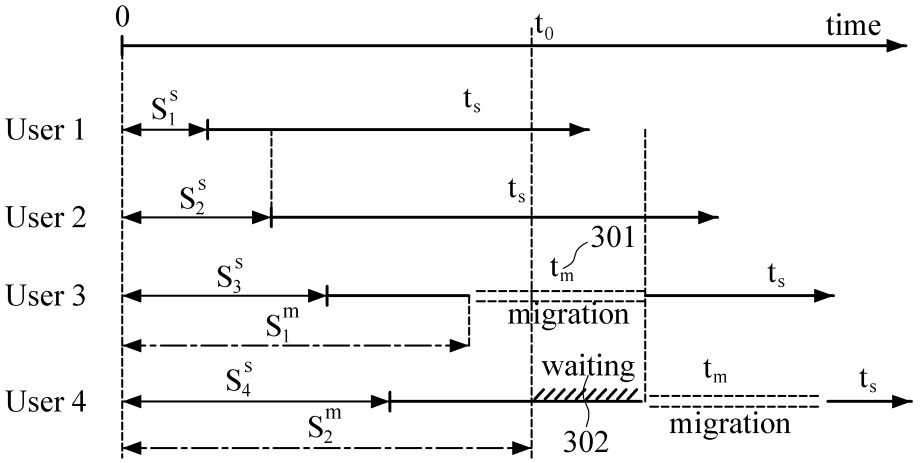
도면1


100
최적 비용 슬롯 할당 장치
최적 비용 슬롯 할당
10
원본 서버
최적 비용 슬롯에 따른 마이그레이션
20
타겟 서버

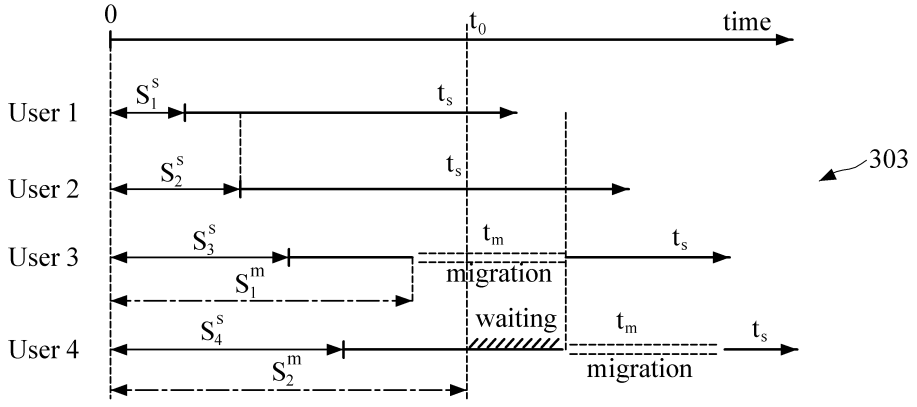
도면2

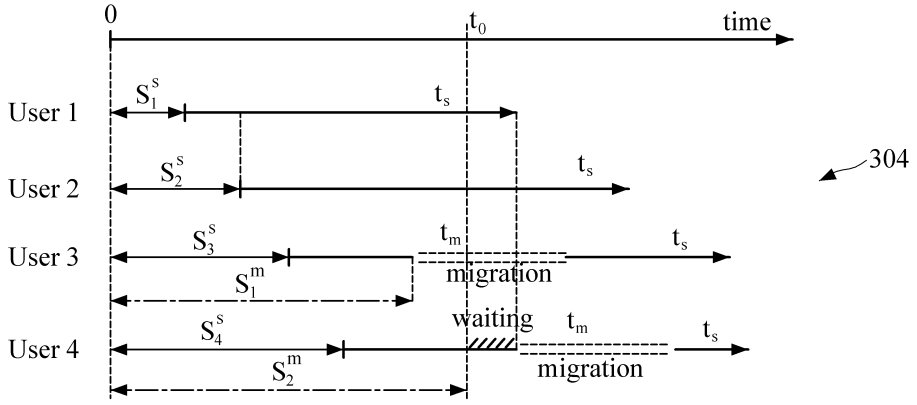

0,0
1,0
2,0
2,1
3,0
3,1
4,0
4,1
4,2
N,0
N,1
N,K
λs
μs
λm
μm
2μs
3μs
4μs
2λm
3μm
2μm

도면3a

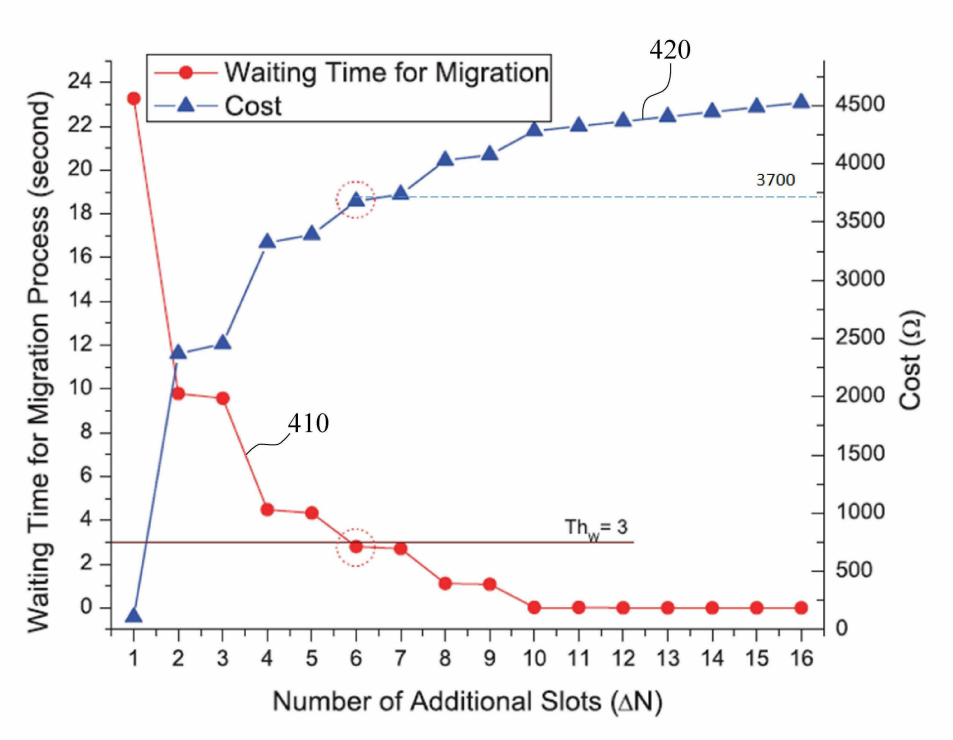

0
t_0
time
User 1
User 2
User 3
User 4
S_1^s
S_2^s
S_3^s
S_4^s
S_1^m
S_2^m
t_s
t_m
301
migration
waiting
302

도면3b

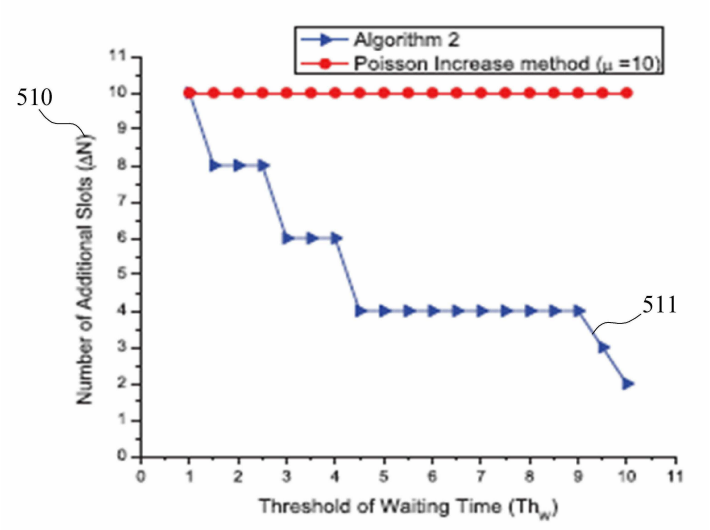

0
t_0
time
User 1
User 2
User 3
User 4
S_1^s
S_2^s
S_3^s
S_4^s
S_1^m
S_2^m
t_s
t_m
migration
waiting
303


0
t_0
time
User 1
User 2
User 3
User 4
S_1^s
S_2^s
S_3^s
S_4^s
S_1^m
S_2^m
t_s
t_m
migration
waiting
304

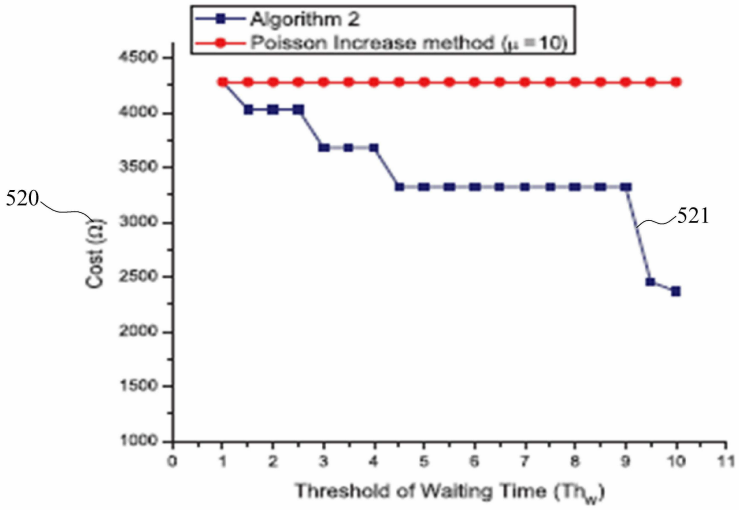
도면4


Waiting Time for Migration
Cost
420
410
3700
Th_W = 3
Waiting Time for Migration Process (second)
Cost (Ω)
Number of Additional Slots (ΔN)

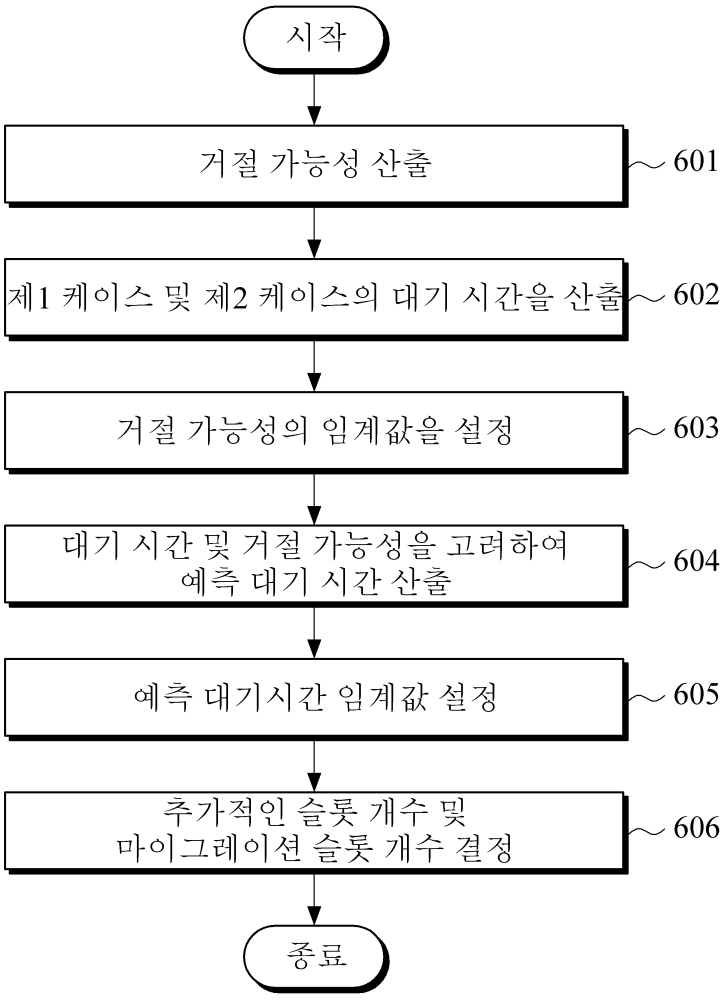
도면5a


Algorithm 2
Poisson Increase method (μ =10)
510
511
Number of Additional Slots (ΔN)
Threshold of Waiting Time (Th_W)

도면5b


Algorithm 2
Poisson Increase method (μ=10)
520
521
Cost (Ω)
4500
4000
3500
3000
2500
2000
1500
1000
0
1
2
3
4
5
6
7
8
9
10
11
Threshold of Waiting Time (Th_w)

도면6


시작
거절 가능성 산출
601
제1 케이스 및 제2 케이스의 대기 시간을 산출
602
거절 가능성의 임계값을 설정
603
대기 시간 및 거절 가능성을 고려하여
예측 대기 시간 산출
604
예측 대기시간 임계값 설정
605
추가적인 슬롯 개수 및
마이그레이션 슬롯 개수 결정
606
종료