



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월01일
(11) 등록번호 10-2235622
(24) 등록일자 2021년03월29일

(51) 국제특허분류(Int. Cl.)
H04L 29/08 (2006.01)

(52) CPC특허분류
H04L 67/2842 (2013.01)
H04L 67/12 (2013.01)

(21) 출원번호 10-2020-0163933

(22) 출원일자 2020년11월30일

심사청구일자 2020년11월30일

(56) 선행기술조사문헌

KR1020110002808 A

(뒷면에 계속)

(73) 특허권자

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

이수경

서울특별시 서대문구 연세로 50, 제4공학관 D710호(신촌동, 연세대학교)

김창영

경기도 광주시 오포읍 문형새술길 7, 107동 1202호(오포문형 양우내안에 아파트)

김태영

경기도 고양시 일산서구 강선로 142, 1710동 305호(일산동, 후곡마을17단지아파트)

(74) 대리인

특허법인우인

전체 청구항 수 : 총 11 항

심사관 : 이준석

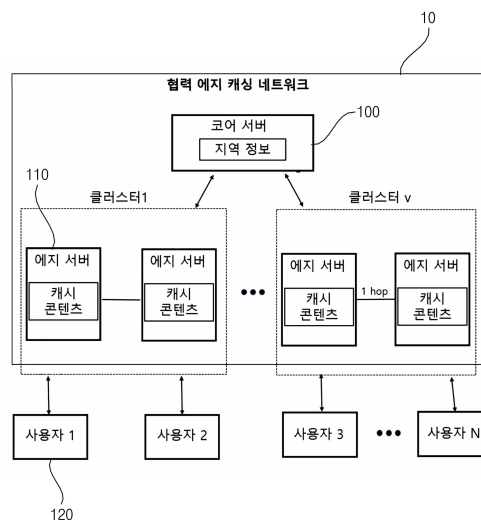
(54) 발명의 명칭 IoT 환경에서의 협력 에지 캐싱 방법 및 그를 위한 장치

(57) 요약

IoT 환경에서의 협력 에지 캐싱 방법 및 그를 위한 장치를 개시한다.

본 발명의 실시예에 따른 협력 에지 캐싱 방법은, 에지 서버로부터 요청 기록 정보를 획득하는 정보 획득 단계; 상기 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 클러스터링 처리 단계; 및 상기 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행하는 캐싱 처리 단계를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류

H04L 67/2823 (2013.01)

H04L 67/289 (2013.01)

H04L 67/325 (2013.01)

H04L 67/327 (2013.01)

(56) 선행기술조사문헌

KR101546199 B1

KR101529602 B1

KR102161161 B1

KR101924165 B1

이 발명을 지원한 국가연구개발사업

과제고유번호 1711110430

과제번호 2019R1A2C1086191

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 중견연구자지원사업

연구과제명 다중계층 이중 엣지에서 IoT 서비스에 대한 강화학습 기반의 자원 할당 기술
연구(2/3)

기 여 율 1/1

과제수행기관명 연세대학교 산학협력단

연구기간 2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

코어 서버에서 협력 에지를 캐싱하는 방법에 있어서,

에지 서버로부터 요청 기록 정보를 획득하는 정보 획득 단계;

상기 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 클러스터링 처리 단계; 및

상기 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행하는 캐싱 처리 단계

를 포함하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 2

제1항에 있어서,

상기 클러스터링 처리 단계는,

복수의 에지 서버 중 하나의 제1 에지 서버를 선택하여 클러스터를 생성하는 제1 에지 서버 선택 단계;

상기 제1 에지 서버와 연결된 제2 에지 서버를 선택하는 제2 에지 서버 선택단계; 및

상기 제2 에지 서버를 기 설정된 클러스터 추가 조건과 비교하고, 상기 클러스터 추가 조건을 만족하는 경우 상기 제2 에지 서버를 상기 클러스터에 추가하는 클러스터 추가 단계

를 포함하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 3

제2항에 있어서,

상기 제2 에지 서버 선택단계는,

상기 제1 에지 서버와 연결된 복수의 에지 서버 중 상기 제1 에지 서버와 한 개의 홉(Hop)으로 연결된 상기 제2 에지 서버를 선택하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 4

제2항에 있어서,

상기 클러스터 추가 단계는,

상기 제2 에지 서버가 다른 클러스터에 속하지 않아야 한다는 제1 조건 및 상기 제2 에지 서버에 캐시 히트(hit) 되었을 때의 시간이 캐시 미스(miss) 되었을 때의 시간보다 기 설정된 일정 시간(t_{cl}) 만큼의 차이가 발생하여야 한다는 제2 조건을 포함하는 상기 클러스터 추가 조건을 만족하는 경우에만 상기 제2 에지 서버를 상기 클러스터에 추가하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 5

제4항에 있어서,

상기 클러스터링 처리 단계는,

상기 복수의 에지 서버 각각이 소정의 클러스터에 속하게 될 때까지 상기 제1 에지 서버 선택 단계, 상기 제2 에지 서버 선택단계 및 상기 클러스터 추가 단계를 반복하여 수행하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 6

제1항에 있어서,

상기 캐싱 처리 단계는,

복수의 에지 서버 각각에 대하여 상기 콘텐츠를 저장했을 때의 변화되는 변화 대기시간(T_w^i)을 산출하는 변화 대기시간 산출 단계;

복수의 에지 서버 각각에 대한 상기 변화 대기시간을 기반으로 클러스터 각각의 평균 대기시간(T_w)을 산출하는 평균 대기시간 산출 단계;

상기 클러스터 중 가장 낮은 평균 대기시간(T_w)의 클러스터를 선택하고, 상기 클러스터 내에 포함된 복수의 에지 서버 중 하나의 에지 서버에 상기 콘텐츠의 조합을 선정하여 상기 콘텐츠가 저장되도록 캐싱을 수행하는 캐싱 수행 단계

를 포함하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 7

제6항에 있어서,

상기 변화 대기시간 산출 단계는,

상기 콘텐츠의 인기도(p_j), 클러스터 내에 저장되어 있는 콘텐츠의 수(Y_j) 및 클러스터의 도착률(λ)을 사용하여 상기 콘텐츠를 상기 에지 서버에 캐싱하였을 때 변경되는 요청 도착률(arrival rate, $\Delta \lambda_i$)을 산출하는 단계;

상기 콘텐츠의 크기(b_j), 사용자 단말기와 에지 서버 사이의 대역폭(r_j), 에지 서버에서의 상기 콘텐츠의 도착률($w_{i,j}$)을 사용해 에지 서버의 전체 서비스 시간(total service time)을 산출하는 단계; 및

변경되는 요청 도착률과 상기 에지 서버의 전체 서비스 시간을 사용하여 상기 변화되는 변화 대기시간(T_w^i)을 산출하는 단계

를 포함하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 8

제6항에 있어서,

상기 캐싱 수행 단계는,

상기 에지 서버의 상기 콘텐츠를 캐싱하기 위한 저장 공간을 확인하고, 상기 저장 공간이 있는 경우 상기 콘텐츠를 저장하되,

상기 저장공간으로 인해 상기 콘텐츠를 저장할 수 없는 경우, 상기 클러스터 내에서 상기 에지 서버 다음으로 상기 대기 시간이 낮은 다른 에지 서버와 상기 콘텐츠를 조합하여 상기 콘텐츠가 상기 다른 에지 서버에 저장되도록 캐싱을 수행하는 것을 특징으로 하는 협력 에지 캐싱 방법.

청구항 9

협력 에지 캐싱 네트워크에서 복수의 에지 서버와 연동하는 코어 서버에 있어서,

에지 서버로부터 요청 기록 정보를 획득하는 정보 획득부;

상기 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 클러스터링 처리부; 및

상기 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행하는 캐싱 처리부

를 포함하는 것을 특징으로 하는 코어 서버.

청구항 10

제9항에 있어서,

상기 클러스터링 처리부는,

복수의 에지 서버 중 하나의 제1 에지 서버를 선택하여 클러스터를 생성하고, 상기 제1 에지 서버와 연결된 제2 에지 서버를 선택하며,

상기 제2 에지 서버를 기 설정된 클러스터 추가 조건과 비교하고, 상기 클러스터 추가 조건을 만족하는 경우 상기 제2 에지 서버를 상기 클러스터에 추가하는 것을 특징으로 하는 코어 서버.

청구항 11

제9항에 있어서,

상기 캐싱 처리부는,

복수의 에지 서버 각각에 대하여 상기 콘텐츠를 저장했을 때의 변화되는 변화 대기시간(T_w^i)을 산출하고, 복수의 에지 서버 각각에 대한 상기 변화 대기시간을 기반으로 클러스터 각각의 평균 대기시간(T_w)을 산출하며,

상기 클러스터 중 가장 낮은 평균 대기시간(T_w)의 클러스터를 선택하고, 상기 클러스터 내에 포함된 복수의 에지 서버 중 하나의 에지 서버에 상기 콘텐츠의 조합을 선정하여 상기 콘텐츠가 저장되도록 캐싱을 수행하는 것을 특징으로 하는 코어 서버.

발명의 설명

기술 분야

[0001] 본 발명은 IoT 환경에서 대규모 캐싱 요청을 고려하여 콘텐츠 전송 시간을 최소화하기 위한 협력 에지 캐싱 방법 및 그를 위한 장치에 관한 것이다.

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 발명의 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 종래 기술 중 IoT 환경에서 Edge server를 활용하여 콘텐츠 전송 효율을 향상시키기 위하여 다양한 연구가 진행되고 있다. 이러한 에지 서버에 대한 기술은 한국등록특허 제10-1546199호에 기재되어 있다.

[0004] 종래의 IoT 환경 시스템에서는 최적화된 협력 캐싱 환경에서 콘텐츠 제공 서비스를 하기 위해 협력 캐싱 방법을 제안하였다. 제안한 기술은 콘텐츠 정보, 네트워크 정보 및 캐싱 서버 저장 용량 정보를 수집하는 단계, 상기 정보들을 활용하여, 협력 캐싱 설정 정보를 생성하는 단계 그리고 해당 정보를 캐싱 서버로 전송하는 단계로 이루어져 있다.

[0005] 종래의 IoT 환경 시스템에서는 네트워크 정보와 콘텐츠 정보에 따라 복수의 캐싱 서버가 콘텐츠를 사용자 장치로 전송함으로써 콘텐츠 전송 속도와 네트워크 트래픽을 감소시키는 것을 목표로 한다.

[0006] 종래의 IoT 환경 시스템에서는 저장 용량이 C인 캐싱 서버의 스토리지를 스토리지 분할 계수(λ)에 따라 두개의 영역으로 분할한다. 제1 영역에는 인기 있는 콘텐츠가 저장되고, 제2영역에는 인기가 낮거나 최근에 사용자로부터 액세스된 콘텐츠가 저장된다. 캐싱할 콘텐츠와 업데이트 주기는 네트워크 크기(M), 네트워크 링크 용량(r), 사용자 장치 및 캐싱 서버의 전송 패킷 크기(l_{req}), 캐싱 서버 및 콘텐츠 서비스 제공자의 응답 패킷 크기(l_{res}), 콘텐츠 전달 네트워크에 연결된 사용자 장치 개수(S), 콘텐츠의 개수(N_g)와 크기(l)에 따라 결정된다. 이러한 정보들을 사용하여 콘텐츠 다운로드 시간 이 최소값이 되는 협력 캐싱을 설정한다.

[0007] 다만, 종래의 IoT 환경 시스템에서는 협력 캐싱 환경에서 콘텐츠 다운로드 시간을 최소화하는 방법을 제시하였

으나, IoT 환경에서의 캐싱 요청이 크게 늘어났을 때 전송 효율이 현저히 감소된다는 문제점이 있다. 이에, 이러한 문제점을 해결하기 위한 기술이 필요하다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 대규모 캐싱 요청을 만족시키기 위해 협력 에지 서버(edge server)를 사용하는 기법에 관한 것으로 캐싱 콘텐츠 전송 효율을 향상시키기 위한 IoT 환경에서의 협력 에지 캐싱 방법 및 그를 위한 장치를 제공하는 데 주된 목적이 있다.

과제의 해결 수단

[0009] 본 발명의 일 측면에 의하면, 상기 목적을 달성하기 위한 협력 에지 캐싱 방법은, 에지 서버로부터 요청 기록 정보를 획득하는 정보 획득 단계; 상기 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 클러스터링 처리 단계; 및 상기 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행하는 캐싱 처리 단계를 포함할 수 있다.

[0010] 또한, 본 발명의 다른 측면에 의하면, 상기 목적을 달성하기 위한 코어 서버는, 에지 서버로부터 요청 기록 정보를 획득하는 정보 획득부; 상기 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 클러스터링 처리부; 및 상기 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행하는 캐싱 처리부를 포함할 수 있다.

발명의 효과

[0011] 이상에서 설명한 바와 같이, 본 발명은 대규모 콘텐츠 요청을 분산시키고, 최종적으로 캐싱 콘텐츠 가져오는 콘텐츠 전송 시간을 줄일 수 있다는 효과가 있다.

[0012] 또한, 본 발명은 IoT 기기와 유동 인구가 많은 도심 지역에서도 안정적인 콘텐츠 전송 시간을 보장함을 통해 시간 민감성 IoT-어플리케이션의 안전성과 사용자들의 경험 만족도를 상승시킬 수 있는 효과가 있다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 실시예에 따른 협력 에지 캐싱 네트워크 시스템을 개략적으로 나타낸 블록 구성도이다.

도 2는 본 발명의 실시예에 따른 협력 에지 처리장치를 개략적으로 나타낸 블록 구성도이다.

도 3은 본 발명의 실시예에 따른 클러스터링 처리 동작을 설명하기 위한 도면이다.

도 4는 본 발명의 실시예에 따른 협력 에지 처리 방법의 클러스터링 동작을 설명하기 위한 순서도이다.

도 5는 본 발명의 실시예에 따른 캐싱 처리 동작을 설명하기 위한 도면이다.

도 6은 본 발명의 실시예에 따른 협력 에지 처리 방법의 캐싱 동작을 설명하기 위한 순서도이다.

도 7은 본 발명의 실시예에 따른 협력 에지 처리에 따른 콘텐츠 전송 시간을 나타낸 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다. 이하에서는 도면들을 참조하여 본 발명에서 제안하는 IoT(Internet of Things) 환경에서의 협력 에지 캐싱 방법 및 그를 위한 장치에 대해 자세하게 설명하기로 한다.

[0015] 본 발명은 IoT 환경에서 발생하는 대규모 캐싱 요청을 만족시키기 위해 협력 에지 서버(edge server)를 사용하는 기법에 관한 것으로 캐싱 콘텐츠 전송 효율을 향상시키는 것을 목표로 한다. 특히, 본 발명은 에지 서버들이

가지고 있는 컴퓨터 자원의 한계를 극복하기 위한 클러스터링 기법과 그리고 콘텐츠의 인기도(popularity)와 요청 (request) 수요에 맞는 콘텐츠를 여러 에지 서버에 저장하는 캐싱 알고리즘을 사용함으로써 하나의 에지 서버에 집중되어 생기는 대기 시간 문제와 이로 인한 지연 시간 발생 문제를 해결하기 위한 것이다.

- [0016] 본 발명이 적용될 수 있는 기본환경은 IoT에서의 무선 통신일 수 있으며, 무선 통신 시스템에는 IoT 센서, 기기, 그리고 스마트폰/태블릿과 같은 사용자 디바이스 등을 포함할 수 있다. 본 발명은 특히, 이러한 디바이스들이 많이 몰려 있는 도심 지역과 같은 곳에서 적용될 경우 큰 효과를 나타낼 수 있다.
- [0017] 도 1은 본 발명의 실시예에 따른 협력 에지 캐싱 네트워크 시스템을 개략적으로 나타낸 블록 구성도이다.
- [0018] 본 실시예에 따른 협력 에지 캐싱 네트워크 시스템(10)은 코어 서버(100), 에지 서버(110) 및 사용자 단말기(120)를 포함하여 구성된다. 도 1의 협력 에지 캐싱 네트워크 시스템(10)은 일 실시예에 따른 것으로서, 도 1에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 협력 에지 캐싱 네트워크 시스템(10)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다.
- [0019] 협력 에지 캐싱 네트워크 시스템(10)의 협력 에지 캐싱의 콘텐츠 전송 시간 최적화 기술을 제안하며, IoT와 에지 서버 환경 시스템에 관한 것이다.
- [0020] 본 발명의 실시예에 따른 에지 서버(110)는 에지 서버 요소 중 하나로써 클라우드 보다 사용자 단말기(120)에 가까이 위치하여 캐싱을 수행하는 서버를 의미한다. 에지 서버(110)는 에지 서버는 클라우드보다는 현저히 낮은 컴퓨터 자원을 가지고 있다.
- [0021] 본 발명의 실시예에 따른 코어 서버(Core Sever, 100)는 이러한 에지 서버들을 지역적으로 관리하는 서버로서, 클러스터 처리 및 캐싱 처리에 대한 알고리즘을 수행한다.
- [0022] IoT 환경의 협력 에지 캐싱 네트워크 시스템(10)에서는 무수히 많은 사용자 단말기(120)로부터 캐싱 요청이 들어오게 된다. 그렇게 될 경우, 하나의 에지 서버에 많은 캐싱 요청이 쌓이게 되어 대기 지연시간을 증가시키고 최종적으로 콘텐츠 전송 시간이 느려지게 될 수 있다.
- [0023] 본 발명의 협력 에지 캐싱 네트워크 시스템(10)에서는 이러한 문제를 해결하기 위하여 다수의 에지 서버(110)가 서로 협력을 통해 여러 개의 콘텐츠를 캐싱하여 사용자 단말기(120)에 제공하도록 한다. 사용자 단말기(120)는 가장 인접한 에지 서버(110)가 속한 클러스터로부터 콘텐츠를 요청할 수 있다.
- [0024] 협력 에지 서버로 들어오는 무수히 많은 캐싱 요청에 대응하기 위하여 코어 서버(100)는 다음과 같은 알고리즘을 수행한다.
- [0025] 코어 서버(100)는 에지 서버(110)의 평균 전송시간에 따른 클러스터링 (Clustering)을 처리한다.
- [0026] 코어 서버(100)는 콘텐츠를 요청하는 사용자 단말기(120)가 과도하게 멀리 떨어져 있는 에지 서버(110)로부터 콘텐츠를 가져오는 것을 방지하고, 코어 서버(100)의 지역 내에서 에지 서버(110)의 저장공간을 효율적으로 사용하기 위해 지역 내의 에지 서버(110)들을 클러스터링 한다.
- [0027] 또한, 코어 서버(100)는 콘텐츠의 도착률에 따른 콘텐츠 캐싱을 처리한다.
- [0028] 코어 서버(100)는 지역 내의 에지 서버(110)들로부터의 캐싱 기록(Cache History)을 기초로 하여 콘텐츠(j)가 클러스터 내의 복수의 에지 서버(120)들 중 하나의 에지 서버(i)에 저장되었을 때 변화되는 대기 지연 시간(T_w^i)을 계산하고, 변화되는 대기 지연 시간(T_w^i)을 캐싱을 처리하는데 사용한다.
- [0029] 코어 서버(100)에서, 변화되는 대기 지연 시간(T_w^i)은 에지 서버(110)의 도착률, 콘텐츠의 인기도, 클러스터내에 저장된 콘텐츠의 수 등에 따라 결정될 수 있다. 코어 서버(100)에서, 에지 서버(110)들의 저장 공간 효율을 위하여 추가적인 캐싱이 콘텐츠 전송 시간을 일정시간 t_{cl} 이상만큼 감소시킬 수 없을 때 캐싱은 중단된다.
- [0030] 도 2는 본 발명의 실시예에 따른 협력 에지 처리장치를 개략적으로 나타낸 블록 구성도이다.
- [0031] 본 실시예에 따른 협력 에지 처리장치(200)는 정보 획득부(210), 클러스터링 처리부(220) 및 캐싱 처리부(230)를 포함한다. 도 2의 협력 에지 처리장치(200)는 일 실시예에 따른 것으로서, 도 2에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 협력 에지 처리장치(200)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다.

- [0032] 협력 에지 처리장치(200)는 협력 에지 서버의 동작을 제어하는 동작을 수행한다. 여기서, 협력 에지 처리장치(200)는 클러스터링 처리 및 캐싱 처리를 통해 협력 에지 서버의 동작을 제어한다.
- [0033] 도 2에서, 협력 에지 처리장치(200)는 코어 서버(100)와 별도의 장치인 것으로 기재하고 있으나 반드시 이에 한정되는 것은 아니며, 코어 서버(100)와 동일한 장치 또는 코어 서버(100)에 포함된 일부 구성요소로 구현될 수 있다. 이하, 협력 에지 처리장치(200)에 포함된 구성요소에 대해 설명하도록 한다.
- [0034] 정보 획득부(210)는 에지 서버로부터 요청 기록 정보를 획득한다.
- [0035] 정보 획득부(210)는 코어 서버의 바운더리(기 설정된 지역) 내의 에지 서버 들로부터 사용자 단말기의 요청 기록(Requests History), 캐싱 기록(Cache History) 등을 포함하는 요청 기록 정보를 수신한다.
- [0036] 요청 기록 정보에는 사용자 단말기와 에지 서버 간의 전송 지연 시간(Transmission Delay, $t_{c,bc}$)과 어떠한 콘텐츠가 요청되었는지 등에 대한 정보가 포함된다. 또한, 요청 기록 정보에는 콘텐츠의 인기도(p_i), 클러스터 내에 저장되어 있는 콘텐츠의 수(Y_j) 및 클러스터의 도착률(λ)을 등에 대한 정보가 포함될 수 있다.
- [0037] 클러스터링 처리부(220)는 요청 기록 정보를 기반으로 적어도 하나의 에지 서버를 포함하는 적어도 하나의 클러스터를 생성하는 동작을 수행한다.
- [0038] 클러스터링 처리부(220)는 복수의 에지 서버 중 하나의 제1 에지 서버를 선택하여 클러스터를 생성한다.
- [0039] 이후, 클러스터링 처리부(220)는 제1 에지 서버와 연결된 제2 에지 서버를 선택한다. 여기서, 클러스터링 처리부(220)는 제1 에지 서버와 연결된 복수의 에지 서버 중 상기 제1 에지 서버와 한 개의 홉(Hop)으로 연결된 제2 에지 서버를 선택한다.
- [0040] 이후, 클러스터링 처리부(220)는 제2 에지 서버를 기 설정된 클러스터 추가 조건과 비교하고, 클러스터 추가 조건을 만족하는 경우 제2 에지 서버를 클러스터에 추가한다.
- [0041] 클러스터링 처리부(220)는 제1 조건 및 제2 조건을 포함하는 클러스터 추가 조건을 기반으로 제2 에지 서버를 클러스터에 추가할지 여부를 결정한다. 구체적으로, 클러스터링 처리부(220)는 제2 에지 서버가 다른 클러스터에 속하지 않아야 한다는 제1 조건 및 제2 에지 서버에 캐시 히트(hit) 되었을 때의 시간이 캐시 미스(miss) 되었을 때의 시간보다 기 설정된 일정 시간(t_{cl}) 만큼의 차이가 발생하여야 한다는 제2 조건을 포함하는 클러스터 추가 조건을 만족하는 경우에만 제2 에지 서버를 클러스터에 추가한다.
- [0042] 한편, 클러스터링 처리부(220)는 복수의 에지 서버 각각이 소정의 클러스터에 속하게 될 때까지 제1 에지 서버를 선택하는 동작, 제2 에지 서버를 선택하는 동작, 클러스터 추가 조건에 따라 클러스터에 에지 서버를 추가하는 동작을 반복하여 수행한다.
- [0043] 이하, 클러스터링을 처리하는 과정을 자세히 설명하도록 한다.
- [0044] 각 에지 서버는 자신에게 들어오는 요청(requests)을 기록한다. 요청 기록에는 사용자와 에지 서버 간의 전송 지연 시간(Transmission Delay)과 어떠한 콘텐츠가 요청되었는지가 포함된다.
- [0045] 코어 서버는 에지 서버로부터 전달받은 요청 기록을 통해 각 에지 서버에서 히트(hit) 되었을 때의 지연 시간($T_{hit}=t_{c,bc}+t_{bc,c}$)을 계산할 수 있다. 코어 서버는 계산된 T_{hit} 들을 기반으로 다음과 같은 클러스터링 과정을 수행한다. 여기서, $t_{c,bc}$ 및 $t_{bc,c}$ 는 에지 서버와 사용자 단말기 간의 전송 지연 시간을 의미한다.
- [0046] (1) 코어 서버는 먼저 자신의 지역에 속하는 에지 서버 중 하나를 선택하여 클러스터를 생성한다.
- [0047] (2) 코어 서버는 그 다음 선택된 에지 서버와 한 개의 홉(hop)으로 연결된 에지 서버(i)를 선택하고 다음 두 가지 조건이 만족할 경우 해당 에지 서버를 클러스터에 추가한다.
- [0048] 조건 1: i는 다른 클러스터에 속하지 않아 있어야 한다.
- [0049] 조건 2: $T_{hit} + t_{cl} < T_{miss}$
- [0050] T_{miss} 는 캐시 미스가 발생하여 클라우드로부터 콘텐츠를 받아왔을 때의 전송 지연 시간을 의미하며 서로 너무 멀리 떨어진 에지 서버끼리 클러스터가 되는 것을 방지하기 위하여 캐시 히트 되었을 때의 시간이 캐시 미스 되었을 때 보다 일정 시간 t_{cl} 만큼의 차이가 발생하였을 때만 에지 서버(i)를 클러스터에 추가한다. 여기서, t_{cl} 은 클

러스터의 크기를 결정하기 위한 threshold일 수 있다.

[0051] (3) 코어 서버는 (1) 내지 (2) 과정에서 클러스터에 추가되는 에지 서버가 없을 때까지 반복한다. 코어 서버는 모든 에지 서버가 클러스터에 속 하게 될 때까지 (1)로 돌아가 클러스터링 과정을 반복한다.

[0052] 캐싱 처리부(230)는 요청 기록 정보를 기반으로 기 설정된 기준 이상의 캐싱 요청이 발생하는 콘텐츠를 적어도 하나의 에지 서버에 저장되도록 캐싱 처리를 수행한다.

[0053] 캐싱 처리부(230)는 복수의 에지 서버 각각에 대하여 상기 콘텐츠를 저장했을 때의 변화되는 변화 대기시간(T_w^i)을 산출한다.

[0054] 캐싱 처리부(230)는 요청 기록 정보에 포함된 콘텐츠의 인기도(p_j), 클러스터 내에 저장되어 있는 콘텐츠의 수(Y_j) 및 클러스터의 도착률(λ)을 사용하여 상기 콘텐츠를 상기 에지 서버에 캐싱하였을 때 변경되는 요청 도착률(arrival rate, $\Delta \lambda_i$)을 산출한다. 이후, 캐싱 처리부(230)는 콘텐츠의 크기(b_j), 사용자 단말기와 에지 서버 사이의 대역폭(r_j), 에지 서버에서의 상기 콘텐츠의 도착률($w_{i,j}$)을 사용해 에지 서버의 전체 서비스 시간(total service time)을 산출하고, 변경되는 요청 도착률과 상기 에지 서버의 전체 서비스 시간을 사용하여 상기 변화되는 변화 대기시간(T_w^i)을 산출한다.

[0055] 이후, 캐싱 처리부(230)는 복수의 에지 서버 각각에 대한 상기 변화 대기시간을 기반으로 클러스터 각각의 평균 대기시간(T_w)을 산출한다.

[0056] 이후, 캐싱 처리부(230)는 클러스터 중 가장 낮은 평균 대기시간(T_w)의 클러스터를 선택하고, 클러스터 내에 포함된 복수의 에지 서버 중 하나의 에지 서버에 콘텐츠의 조합을 선정하여 콘텐츠가 저장되도록 캐싱을 수행한다. 여기서, 캐싱 처리부(230)는 에지 서버의 콘텐츠를 캐싱하기 위한 저장 공간을 확인하고, 저장 공간이 있는 경우 해당 에지 서버에 콘텐츠가 저장되도록 캐싱을 수행한다.

[0057] 한편, 캐싱 처리부(230)는 저장공간으로 인해 콘텐츠를 저장할 수 없는 경우, 클러스터 내에서 에지 서버 다음으로 대기 시간이 낮은 다른 에지 서버와 콘텐츠를 조합하여 다른 에지 서버에 콘텐츠가 저장되도록 캐싱을 수행한다.

[0058] 이하, 캐싱을 처리하는 과정을 자세히 설명하도록 한다.

[0059] 코어 서버는 클러스터링 이후, 대규모 캐싱 요청으로 인해 발생하는 대기시간을 줄이기 위하여 수요가 많은 콘텐츠를 1 개 이상의 에지 서버에 저장하는 캐싱 알고리즘을 수행한다.

[0060] 이를 위해서는 코어 서버는 먼저 새로운 콘텐츠(j)를 에지 서버(i)에 캐싱 하였을때 변경되는 요청 도착률(arrival rate)을 계산한다. 여기서, 요청 도착률(arrival rate)은 콘텐츠(j)의 인기도(p_j), 클러스터 내에 저장되어 있는 콘텐츠(j)의 수(Y_j), 클러스터의 도착률(λ)을 사용하여 계산될 수 있다. 요청 도착률(arrival rate)은 [수학식 1]에 의해 계산될 수 있다. [수학식 1]의 변수들은 클러스터링 처리 이전 과정에서 받은 요청 기록을 통하여 얻을 수 있다.

수학식 1

$$\Delta \lambda_i = - \frac{\lambda p_j}{Y_j(Y_j + 1)}$$

[0061]

[0062] 또한, 코어 서버는 콘텐츠의 크기(b_j), 사용자 단말기와 에지 서버 사이의 대역폭(r_i), 에지 서버에서의 콘텐츠의 도착률($w_{i,j}$)을 사용하여 에지 서버(i)의 전체 서비스 시간(total service time)을 계산한다. 여기서, 서비스 시간(total service time)은 [수학식 2]를 통해 계산될 수 있다.

수학식 2

$$E \left[\frac{1}{\mu_i} \right] = \sum_{j \in J} \frac{1}{\mu_i}$$

[0063]

[0064]

마지막으로, 코어 서버는 [수학식 1]에서 유도된 변경된 도착률을 통해 변경된 후의 도착률(λ'_i)을 얻을 수 있으며, 변경된 후의 도착률(λ'_i)과 [수학식 2]를 사용하여 콘텐츠 복제본이 클러스터에 하나 추가되었을 때 변화되는 대기시간(waiting time, T'_w)을 산출할 수 있다. 여기서, 코어 서버는 [수학식 3]을 통해 변화되는 대기시간(T'_w)을 산출할 수 있다.

수학식 3

$$T'_w = \frac{\lambda'_i E \left[\left(\frac{1}{\mu'_i} \right)^2 \right]}{2 \left(1 - \lambda_i E \left[\frac{1}{\mu'_i} \right] \right)}$$

[0065]

[0066]

[수학식 3]에서 λ_i 는 변경 되기 전의 도착률을 의미한다.

[0067]

코어 서버는 [수학식 3]을 통해 얻은 변화되는 대기시간을 통하여, 대기시간의 변화가 가장 클 때의 콘텐츠(j)와 에지 서버(i)를 선택하여 캐싱을 수행한다.

[0068]

코어 서버는 [수학식 3]의 수식을 바탕으로, 캐싱 과정을 수행한다.

[0069]

우선 코어 서버는 모든 에지 서버(i)와 콘텐츠(j)에 대하여, 콘텐츠(j)를 에지 서버(i)에 저장했을 때의 각 에지 서버(i)들의 변화되는 대기시간(T'_w)들을 산출하고, 이에 따른 클러스터의 평균 대기시간(T_w)을 산출한다.

[0070]

이때, 코어 서버는 가장 낮은 평균 대기시간(T_w)을 유도하는 에지 서버(i) 및 콘텐츠(j)를 선택한 후에, 에지 서버(i)에 콘텐츠(j)를 저장할 수 있을 만큼의 공간이 있는지 여부를 확인한다.

[0071]

코어 서버는 에지 서버(i)에 콘텐츠(j)를 저장할 수 있을 만큼의 공간이 있을 경우 콘텐츠(j)를 에지 서버(i)에 저장한다. 코어 서버는 만약 에지 서버(i)에 콘텐츠(j)를 저장할 공간이 없을 경우, 차선의 다른 에지 서버 및 콘텐츠(j)의 조합을 선택한다. 차선의 다른 에지 서버는 클러스터 내에서 에지 서버(i) 다음으로 가장 낮은 평균 대기시간(T_w)을 유도하는 에지 서버일 수 있다.

[0072]

만약, 코어 서버는 모든 조합에 대해서 콘텐츠(j)를 에지 서버에 저장할 수 없을 경우, 캐싱 과정을 중단한다. 즉, 코어 서버는 클러스터 내의 모든 에지 서버의 저장 공간으로 인해 캐싱이 중단될 때까지 최선의 에지 서버(i) 및 콘텐츠(j)의 조합을 선택하여 캐싱을 수행한다.

[0073]

도 3은 본 발명의 실시예에 따른 클러스터링 처리 동작을 설명하기 위한 도면이다.

[0074]

도 3을 참고하면, 코어 서버(100)는 에지 서버(110)들로부터 에지 서버(110)와 사용자 단말기(120)들 간의 지연시간(T_{hit})의 정보를 받고 이를 기반으로 클러스터링을 수행한다.

[0075]

코어 서버(100)는 클러스터링 과정 시 하나의 에지 서버(120)를 선택하여 클러스터(310)를 생성한 후 클러스터

에 1-hop으로 연결된 에지 서버 중 $T_{hit} + t_{cl} < T_{miss}$ 를 만족하는 에지 서버(120)들을 클러스터(310)에 추가한다.

- [0076] 만약, 코어 서버(100)는 만족되는 에지 서버가 없을 경우 하나의 에지 서버를 선택하는 과정으로 돌아가 다른 에지 서버를 선택하고, 선택된 다른 에지 서버들을 클러스터(310)에 추가하는 동작을 반복하여 수행한다.
- [0077] 도 4는 본 발명의 실시예에 따른 협력 에지 처리 방법의 클러스터링 동작을 설명하기 위한 순서도이다.
- [0078] 코어 서버(100)는 자신의 지역에 속하는 에지 서버 중 하나를 선택하여 클러스터(v)를 생성한다(S410).
- [0079] 코어 서버(100)는 그 다음 선택된 에지 서버와 한 개의 홉(hop)으로 연결된 에지 서버(i)를 선택한다(S420).
- [0080] 코어 서버(100)는 에지 서버(i)가 다른 클러스터에 속하지 않는지 여부를 확인한다(S430).
- [0081] 코어 서버(100)는 에지 서버(i)가 다른 클러스터에 속하지 않으면, 캐시 미스가 발생하여 클라우드로부터 콘텐츠를 받아왔을 때의 전송 지연 시간과 캐시 히트 되었을 때의 지연 시간을 비교하여 일정 시간 t_{cl} 만큼의 차이가 발생하는지 여부를 확인한다(S440, S450).
- [0082] 코어 서버(100)는 캐시 미스가 발생하여 클라우드로부터 콘텐츠를 받아왔을 때의 전송 지연 시간과 캐시 히트 되었을 때의 지연 시간을 비교하여 일정 시간 t_{cl} 만큼의 차이가 발생하였을 때만 에지 서버(i)를 클러스터(v)에 추가한다(S460).
- [0083] 코어 서버(100)는 클러스터에 추가되는 에지 서버가 없을 때까지 단계 S420 내지 S460을 반복하여 수행한다(S470).
- [0084] 도 4에서는 각 단계를 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 4에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 4는 시계열적인 순서로 한정되는 것은 아니다.
- [0085] 도 4에 기재된 본 실시예에 따른 협력 에지 처리 방법은 애플리케이션(또는 프로그램)으로 구현되고 단말장치(또는 컴퓨터)로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 협력 에지 처리 방법을 구현하기 위한 애플리케이션(또는 프로그램)이 기록되고 단말장치(또는 컴퓨터)가 읽을 수 있는 기록매체는 컴퓨팅 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치 또는 매체를 포함한다.
- [0086] 도 5는 본 발명의 실시예에 따른 캐싱 처리 동작을 설명하기 위한 도면이다.
- [0087] 도 5을 참고하면, 각 에지 서버(110)는 사용자 단말기(120)로부터 자신들에게 들어온 캐싱 요청에 대한 캐싱 기록(Cache History)을 가지고 있다. 캐싱 기록에는 어떠한 데이터가 얼마나 요청되어 있는지 등에 대한 정보가 포함되어 있다.
- [0088] 코어 서버(100)는 각 에지 서버(110)로부터 캐싱 기록을 받아 각 에지 서버의 요청 도착률(λ_i), 서비스 시간($\frac{1}{\mu_i}$), 데이터의 인기도 등을 계산할 수 있다. 이 때, 코어 서버(100)는 $\frac{\text{콘텐츠 } j \text{ 에 대한 요청수}}{\text{전체 요청수}}$ 를 이용하여 콘텐츠(j)에 대한 인기를 계산할 수 있다.
- [0089] 코어 서버(100)는 계산된 도착률과, 서비스 시간, 인기도 등을 이용하여 [수학식 1] 및 [수학식 2]를 유도할 수 있으며, 최종적으로 [수학식 3]을 유도하여 [수학식 3]의 결과값(변화되는 대기시간)이 가장 클 때의 에지 서버(i) 및 콘텐츠(j)를 선택하여 캐싱을 수행한다.
- [0090] 도 6은 본 발명의 실시예에 따른 협력 에지 처리 방법의 캐싱 동작을 설명하기 위한 순서도이다.
- [0091] 코어 서버(100)는 모든 에지 서버(i)와 콘텐츠(j)에 대하여, 콘텐츠(j)를 에지 서버(i)에 저장했을 때의 각 에지 서버(i)들의 변화되는 대기시간(T_w^i)들을 산출한다(S610).
- [0092] 코어 서버(100)는 각 에지 서버(i)들의 변화되는 대기시간(T_w^i)들을 기반으로 클러스터의 평균 대기시간(T_w)을 산출하고, 가장 낮은 평균 대기시간(T_w)을 유도하는 에지 서버(i) 및 콘텐츠(j)를 선택한다(S620).
- [0093] 코어 서버(100)는 에지 서버(i)에 콘텐츠(j)를 저장할 수 있을 만큼의 공간이 있는지 여부를 확인한다(S630).
- [0094] 코어 서버(100)는 에지 서버(i)에 콘텐츠(j)를 저장할 수 있을 만큼의 공간이 있을 경우 콘텐츠(j)를 에지 서버

(i)에 저장하여 캐싱을 수행한다(S640).

코어 서버(100)는 에지 서버(i)에 콘텐츠(j)를 저장할 공간이 없을 경우, 차선의 다른 에지 서버 및 콘텐츠(j)의 조합을 선택하거나, 모든 조합에 대해서 콘텐츠(j)를 에지 서버에 저장할 수 없을 경우, 캐싱 과정을 중단한다(S650).

도 6에서는 각 단계를 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 6에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 6은 시계열적인 순서로 한정되는 것은 아니다.

도 6에 기재된 본 실시예에 따른 협력 예지 처리 방법은 애플리케이션(또는 프로그램)으로 구현되고 단말장치(또는 컴퓨터)로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 협력 예지 처리 방법을 구현하기 위한 애플리케이션(또는 프로그램)이 기록되고 단말장치(또는 컴퓨터)가 읽을 수 있는 기록매체는 컴퓨팅 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치 또는 매체를 포함한다.

도 7은 본 발명의 실시예에 따른 협력 에지 처리에 따른 콘텐츠 전송 시간을 나타낸 예시도이다.

본 발명에 따른 협력 에지 캐싱 네트워크 시스템(10)에서 제안한 에지 서버 클러스터링과 변화되는 대기시간에 따른 콘텐츠 캐싱 방법 효과를 확인하기 위하여 보이기 위하여 시뮬레이션을 수행하였고 그 결과로 대조군으로 사용된 콘텐츠를 하나의 에지 서버에 저장하는 캐싱 보다 콘텐츠 전송 지연 시간이 줄어드는 것을 확인하였다.

협력 에지 캐싱 네트워크 시스템(10)에서 제안한 협력 캐싱 알고리즘을 통해 대기시간을 최소화하는 콘텐츠 캐싱 방법을 평가하기 위하여 에지 서버에 도착하는 초당 캐싱 요청 횟수(λ)를 달리하여 콘텐츠 전송 지연 시간을 측정하였다.

도 7은 캐싱 요청 횟수에 따른 각 알고리즘 별 콘텐츠 전송 지연 시간을 나타낸 것이다.

총래의 original로 표기한 알고리즘에 비하여 본 발명에서 제안한 proposed로 표기한 알고리즘을 사용했을 때 모든 구간에서 더 낮은 콘텐츠 전송 지연시간을 보여주는 것을 확인할 수 있다.

도 7에서는 추가적으로 지프(Zipf) 변수의 정도를 0.3(도 7의 (a)), 0.7(도 7의 (b)), 1.0(도 7의 (c))으로 다르게 주어지는 상황에서 지연 시간을 비교하였다. 도 7의 결과에서 지프(zipf) 변수가 증가함에 따라 캐싱 콘텐츠 전송 시간은 증가하며 모든 지프(zipf) 변수 변화 구간에서 제안한 기술이 종래 기술보다 낮은 콘텐츠 전송 시간으로 콘텐츠를 전송하는 것을 확인할 수 있다.

이상의 설명은 본 발명의 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명의 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명의 실시예들은 본 발명의 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

10: 협력 에지 캐싱 네트워크 시스템

100: 코어 서버

110: 예지 서버

120: 사용자 단말기

200: 협력 에지 처리장치

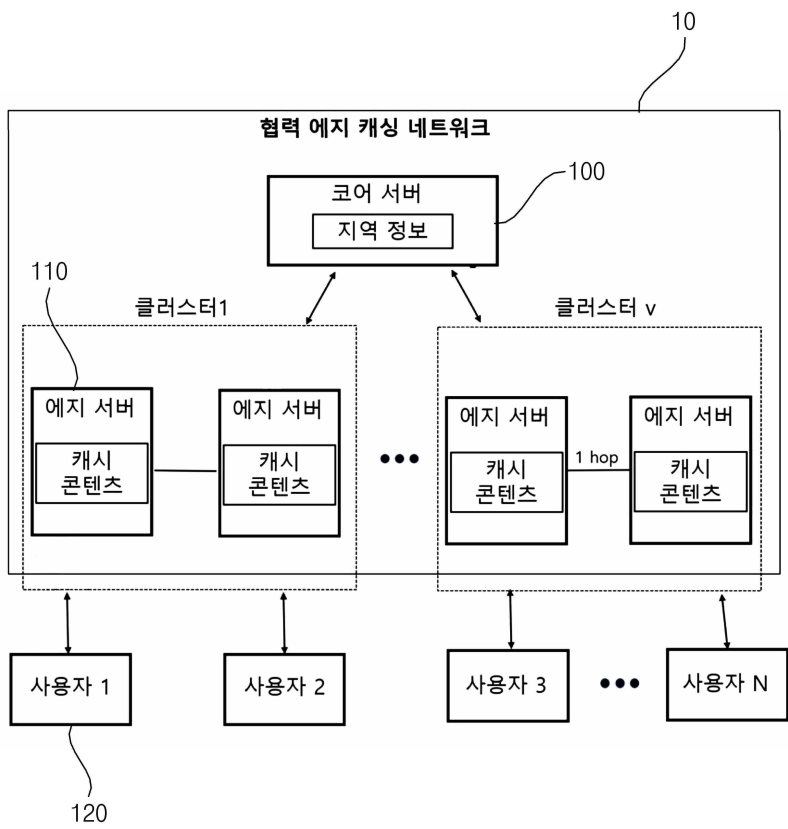
210: 정보 획득부

220: 클러스터링 처리부

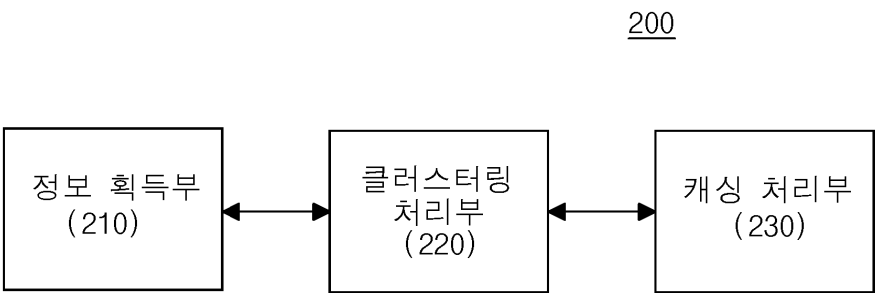
230: 캐싱 처리부

도면

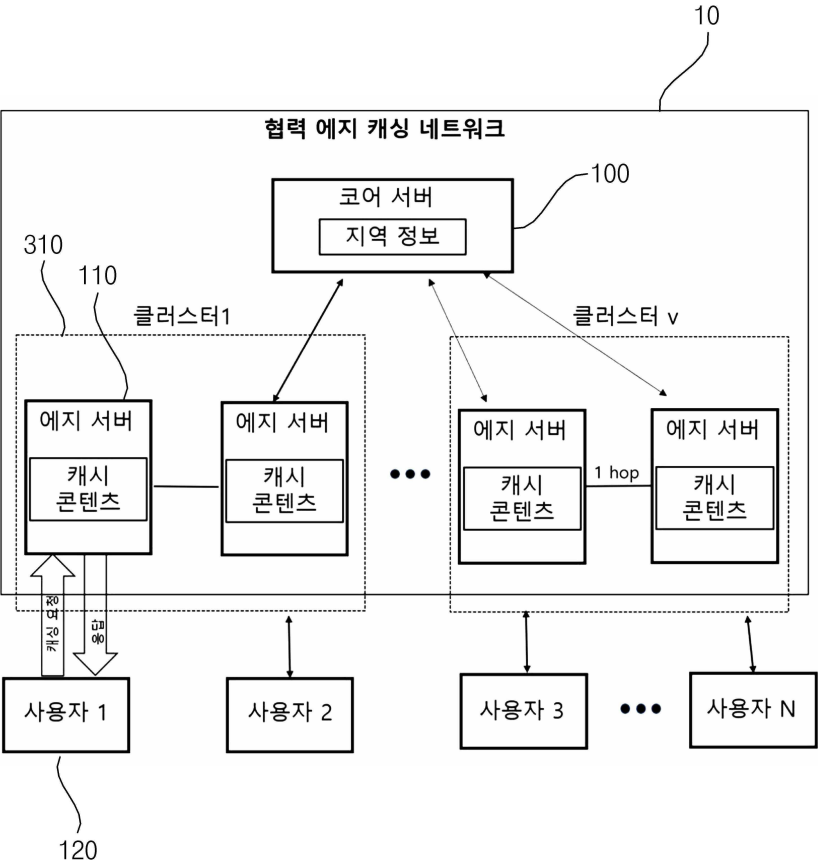
도면1



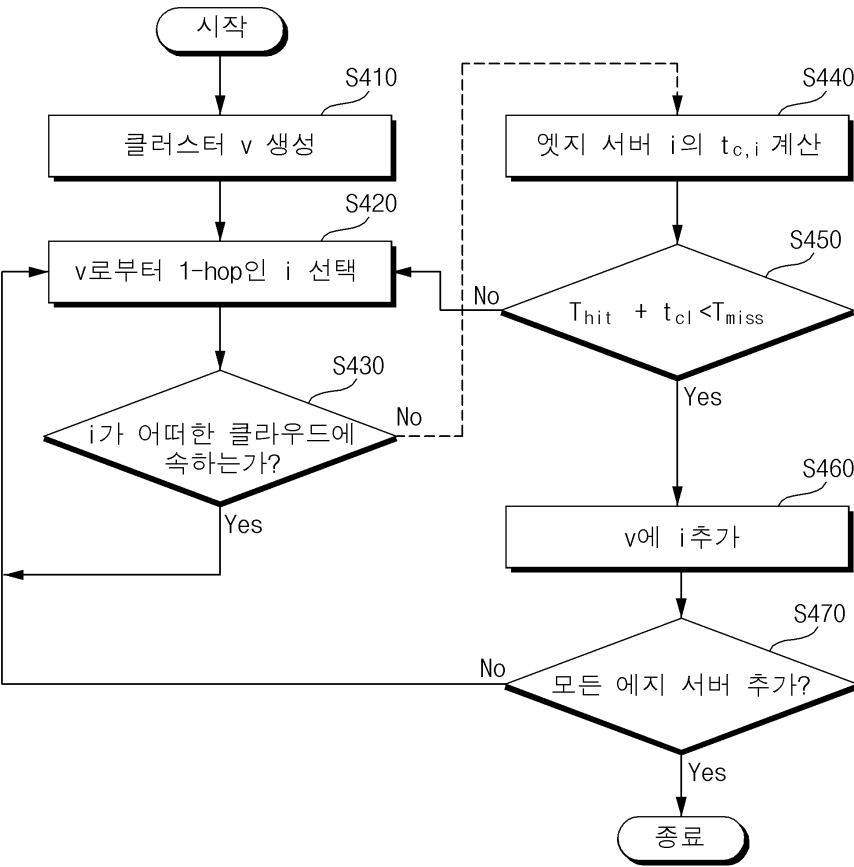
도면2



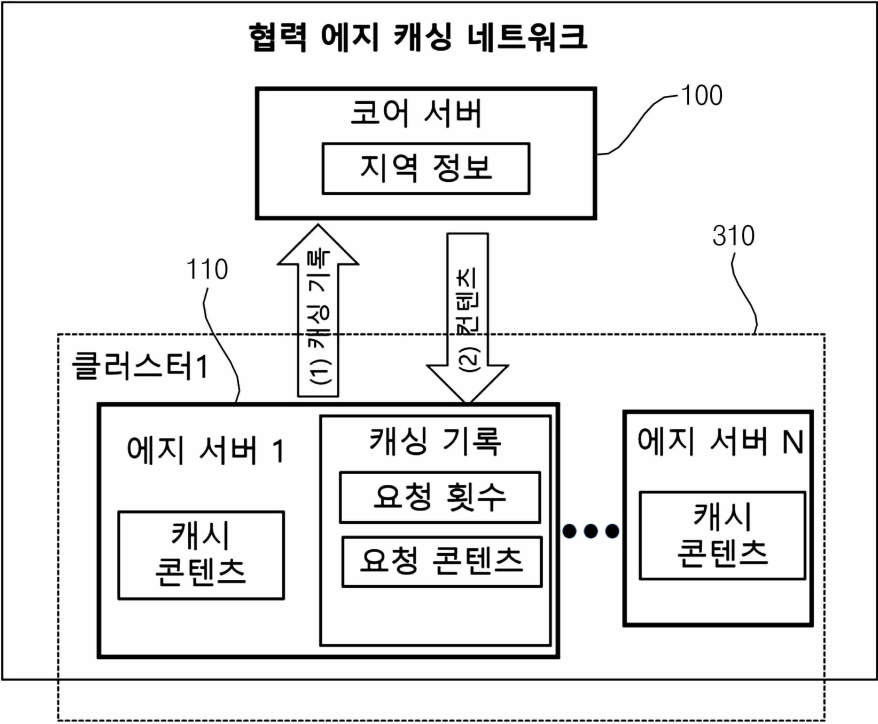
도면3



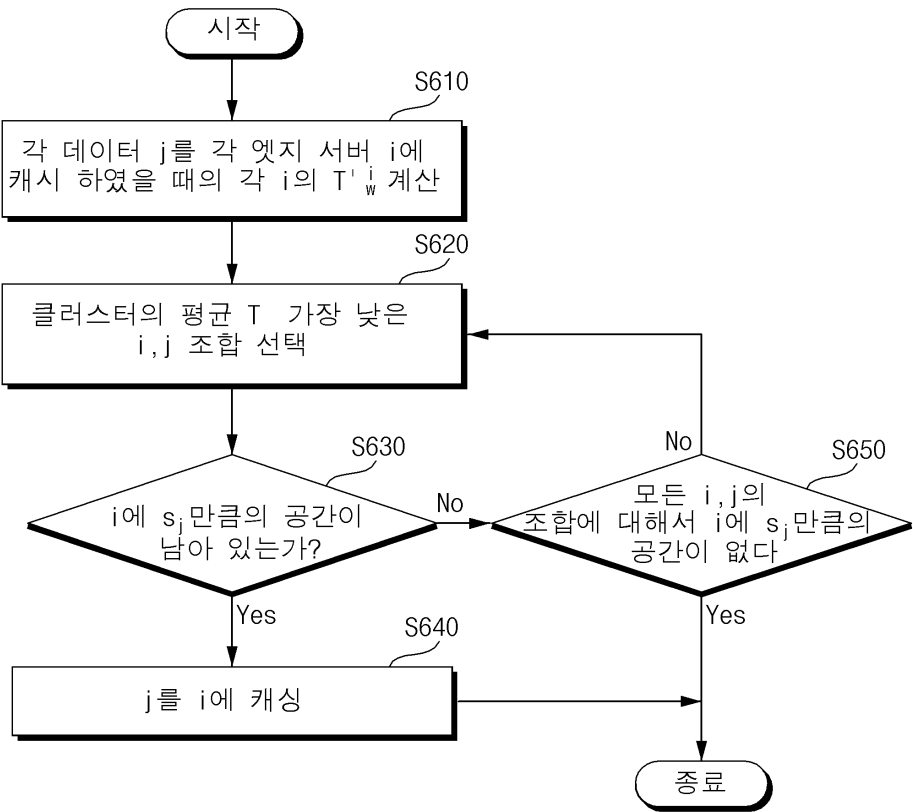
도면4



도면5



도면6



도면7

