



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0135511
(43) 공개일자 2017년12월08일

(51) 국제특허분류(Int. Cl.)
H04L 12/841 (2013.01) H04L 12/805 (2013.01)
H04L 12/851 (2013.01)
(52) CPC특허분류
H04L 47/283 (2013.01)
H04L 47/245 (2013.01)
(21) 출원번호 10-2016-0067533
(22) 출원일자 2016년05월31일
심사청구일자 2016년05월31일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이원준
서울특별시 강남구 선릉로 221, 104동 2205호 (도곡동, 도곡테크아파트)
김규영
서울특별시 관악구 행운1가길 11, 202호 (봉천동)
(74) 대리인
특허법인엠에이피에스

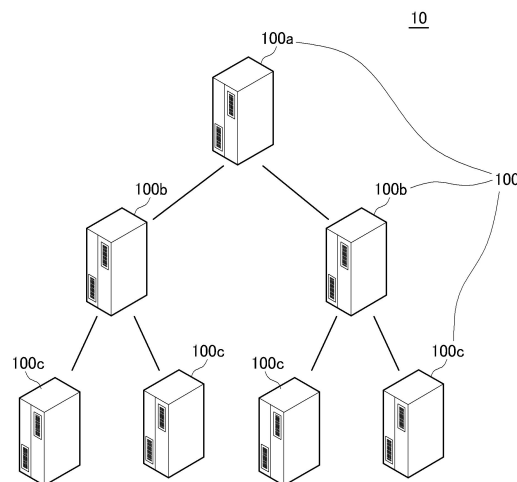
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법 및 장치

(57) 요약

본 발명에 따르는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법은, (a) 외부 객체로부터 쿼리를 수신하는 단계; (b) 상기 쿼리에 대한 응답을 위해 우선순위를 지정하는 단계; 및 (c) 상기 우선순위에 따라 패킷을 전송하는 단계;를 포함하며, 상기 우선순위를 지정하는 단계는, 상기 응답을 위한 패킷을 생성하여 다른 서버로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 상기 재전송 패킷에 대하여 높은 우선순위를 지정한다.

대표도 - 도1



(52) CPC특허분류

H04L 47/2483 (2013.01)

H04L 47/36 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711024057

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 핵심연구지원사업

연구과제명 무선 네트워킹 클라우드 설계 및 성능 최적화 기술

기 여 율 1/1

주관기관 고려대학교

연구기간 2015.06.01 ~ 2016.05.31

명세서

청구범위

청구항 1

서버에 의해 수행되는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법에 있어서,

- (a) 외부 객체로부터 쿼리를 수신하는 단계;
- (b) 상기 쿼리에 대한 응답을 위해 우선순위를 지정하는 단계; 및
- (c) 상기 우선순위에 따라 패킷을 전송하는 단계;를 포함하며,

상기 우선순위를 지정하는 단계는, 상기 응답을 위한 패킷을 생성하여 다른 서버로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 상기 재전송 패킷에 대하여 높은 우선순위를 지정하는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 2

제 1 항에 있어서,

상기 서버는,

사용자 단말의 쿼리에 응답하기 위하여 복수의 서버들이 분산 탐색을 수행하는 데이터 센터에 속하는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 3

제 2 항에 있어서,

상기 데이터 센터가, 사용자 단말로부터 상기 쿼리를 수신하는 중앙 서버 및 상기 중앙 서버의 작업을 분산 처리하기 위한 복수의 하위 서버로 구성되는 경우,

상기 스케줄링 방법을 수행하는 서버는 상기 중앙 서버에 대응하며, 상기 다른 서버는 상기 하위 서버에 대응하는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 4

제 3 항에 있어서,

상기 스케줄링 방법을 수행하는 서버는,

상기 쿼리에 대한 응답을 찾기 위한 복수의 패킷을 생성하여 복수의 다른 서버로 전송하고, 상기 복수의 다른 서버로부터 상기 전송된 패킷에 대한 응답을 모두 수신할 때까지 대기한 후, 상기 쿼리에 대한 응답을 상기 사용자 단말로 전송하는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 5

제 1 항에 있어서,

상기 (b) 단계는,

- (b-1) 상기 대상 패킷이 재전송 패킷인지 판단하는 단계;
- (b-2) 상기 대상 패킷이 재전송 패킷인 경우, 상기 대상 패킷의 플로우 크기와 미리 설정된 임계값과 비교하는 단계; 및
- (b-3) 상기 재전송 패킷인지 여부 및 상기 대상 패킷의 플로우 크기와 임계값과의 비교 결과에 따라 상기 대상 패킷에 대한 우선순위값을 지정하는 단계;를 포함하며,

상기 플로우를 상기 대상 패킷이 출발지에서 목적지로 전송되기까지 이동하는 경로에 관한 정보인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 6

제 5 항에 있어서,

상기 (b-1) 단계는,

상기 대상 패킷이 상기 다른 서버로 전송된 후, 미리 설정된 시간 이내에 상기 다른 서버로부터 수신확인 메시지를 수신하지 못하는 경우, 상기 전송된 대상 패킷을 재전송 패킷으로 판단하는 단계를 포함하는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 7

제 5 항에 있어서,

상기 (b-1) 단계는,

재전송 타이머 또는 TCP NewReno에서 제공되는 Fast Retransmit 기법을 통하여 상기 대상 패킷의 재전송 패킷 여부를 판단하는 단계를 포함하는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 8

제 5 항에 있어서,

상기 (b-2) 단계는,

상기 대상 패킷의 플로우에 대한 수신확인 메시지의 누적 데이터의 크기와 상기 임계값을 비교하는 단계를 포함하는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 9

제 8 항에 있어서,

상기 수신확인 메시지는 패킷을 수신한 서버가 패킷을 전송한 서버로 전송하는 확인 신호이며,

상기 수신확인 메시지가 발생될 때마다 이전부터 현재까지의 수신확인 메시지의 데이터의 크기가 누적 기록되는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 10

제 5 항에 있어서,

상기 (b-3) 단계는,

상기 대상 패킷의 플로우 크기가 상기 임계값 이하인 경우 상기 대상 패킷에 대하여 제 1 우선순위값을 지정하며,

상기 대상 패킷의 플로우 크기가 상기 임계값 초과이거나 상기 대상 패킷이 재전송 패킷이 아닌 경우, 상기 대상 패킷에 대하여 상기 제 1 우선순위값보다 낮은 제 2 우선순위값을 지정하는 단계;를 포함하는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 11

제 10 항에 있어서,

상기 제 1 및 제 2 우선순위값은 상기 대상 패킷의 IP 헤더의 DSCP(Differentiated Service Code Point) 필드 내에 기록되는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 12

제 1 항에 있어서,

상기 (c) 단계는,

분급기를 통하여 상기 우선순위 별로 대상 패킷들을 서로 다른 큐에 정렬하는 단계; 및

우선순위가 가장 높은 큐에 정렬되어 있는 대상 패킷들부터 전송하는 단계;를 포함하는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법.

청구항 13

패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 장치에 있어서,

패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법을 수행하기 위한 프로그램이 저장된 메모리; 및

상기 프로그램을 실행하기 위한 프로세서;를 포함하고,

상기 프로세서는, 상기 프로그램의 실행에 따라,

외부 객체로부터 쿼리를 수신하고,

상기 쿼리에 대한 응답을 위해 우선순위를 지정하고,

상기 우선순위에 따라 패킷을 전송하며,

상기 우선순위를 지정하는 과정에서, 상기 응답을 위한 패킷을 생성하여 다른 서버로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 상기 재전송 패킷에 대하여 높은 우선순위를 지정하는 것인, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 장치.

발명의 설명

기술 분야

[0001] 본 발명은 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법 및 장치에 관한 것으로서, 보다 상세하게는, 데이터센터 네트워크 상의 쿼리 완료 시간을 단축시키기 위한, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법 및 장치에 관한 것이다.

배경 기술

[0002] 데이터센터는 다수의 컴퓨터 서버와 스위치 등으로 구성되며 사회 관계망 서비스, 웹 검색 엔진, 빅데이터 분석 등에 폭 넓게 이용되고 있는 다양한 응용의 기반이 되는 시설이다.

[0003] 데이터센터는 사용자 단말로부터 쿼리를 요청받으며, 이 쿼리는 다수의 네트워크 플로우로 분할되어 데이터센터 내 다수의 서버에 할당되어 작업이 수행된다. 작업을 완료한 각 서버들은 작업 결과를 쿼리 작업 처리를 명령한 중앙 서버에 통보하며 이는 네트워크를 통해 전달된다. 모든 분배된 작업 결과가 취합되어야 쿼리가 완료되므로 쿼리 완료 시간은 분할된 네트워크 플로우들 중 가장 느린 플로우의 완료 시간에 의해 결정된다.

[0004] 여기서 쿼리 완료 시간은 중요한 요소이다. 만약 쿼리 완료 시간이 느리면 사용자는 서비스 품질에 만족하지 못하고 다른 서비스를 활용할 확률이 높아지며 이러한 사용자의 이탈은 데이터센터 활용 사업자의 수익 감소로 이어질 수 있다.

[0005] 데이터센터 네트워크 상의 쿼리 완료 시간 최소화를 위한 종래의 방법들은 다음과 같다. 먼저 데이터센터 네트워크 혼잡 제어 기법은 네트워크 혼잡도에 비례하여 송신 서버의 전송률을 감소시키는 기법이다. 데이터센터 네트워크 부하 분산 기법은 네트워크 트래픽을 특정 경로에 물리적으로 하는 것이 아닌, 네트워크 전체에 고르게 분산시키는 기법이다. 이러한 기법들은 모두 플로우의 패킷 손실 발생 확률을 줄이는 것을 원리로 한다. 플로우의 패킷 손실이 발생하는 경우 패킷 재전송으로 인해 플로우 완료 시간이 느려지므로, 가장 느린 플로우의 완료 시간에 의해 결정되는 쿼리 완료 시간 역시 느려지기 때문이다.

[0006] 그러나 위 종래의 방법들은 네트워크 장비 고장 등 소프트웨어 기술을 통해 방지가 불가능한 플로우의 손실이 발생하는 경우, 쿼리 완료 시간을 최소화시킬 수 없다는 문제가 생긴다. 또한 종래의 방법들은 ECN(Explicit

Congestion Notification)과 많은 수의 우선순위 큐 활용을 요구하므로 고가의 스위치에서만 적용 가능하기 때문에, 장비 교체비용이 제한적이고 주기가 긴 대학교와 중소기업 등의 데이터센터에선 적용이 어렵다. 또한 대규모 데이터센터는 다수의 사업자가 데이터센터 자원을 공유해서 사용하고 있으므로 우선순위 큐 활용시 요금액을 기반으로 예약 사용하는 경우가 많기 때문에, 많은 수의 우선순위 큐를 활용하는 기법을 적용하기가 어렵다.

[0007] 관련하여, 등록특허 제10-1350755호(공고일 : 2014.01.10)는 클라우드 컴퓨팅에서 다중 워크플로우를 위한 비용 기반 스케줄링 방법 및 그 시스템을 개시하고 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 패킷 재전송으로 인한 플로우의 지연시간을 최소화하는 플로우 스케줄링 방법 및 장치를 제공하는 것을 목적으로 한다.

[0009] 이를 통해, 본 발명은 큐리 완료 시간을 최소화함으로써 서비스 응답속도 향상을 제공하는 것을 목적으로 한다.

[0010] 또한, 본 발명은 저가의 스위치 장비에서도 플로우 지연시간 최소화 기술을 사용할 수 있도록 하는 것을 목적으로 한다.

과제의 해결 수단

[0011] 본 발명의 일 실시예에 따르는, 서버에 의해 수행되는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법은, (a) 외부 객체로부터 큐리를 수신하는 단계; (b) 상기 큐리에 대한 응답을 위해 우선순위를 지정하는 단계; 및 (c) 상기 우선순위에 따라 패킷을 전송하는 단계;를 포함하며, 상기 우선순위를 지정하는 단계는, 상기 응답을 위한 패킷을 생성하여 다른 서버로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 상기 재전송 패킷에 대하여 높은 우선순위를 지정한다.

[0012] 본 발명의 다른 실시예에 따르는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 장치는, 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법을 수행하기 위한 프로그램이 저장된 메모리; 및 상기 프로그램을 실행하기 위한 프로세서;를 포함하고, 상기 프로세서는, 상기 프로그램의 실행에 따라, 외부 객체로부터 큐리를 수신하고, 상기 큐리에 대한 응답을 위해 우선순위를 지정하고, 상기 우선순위에 따라 패킷을 전송하며, 상기 우선순위를 지정하는 과정에서, 상기 응답을 위한 패킷을 생성하여 다른 서버로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 상기 재전송 패킷에 대하여 높은 우선순위를 지정한다.

발명의 효과

[0013] 본 발명은 손실이 발생한 패킷이 재전송됨에 따라 발생하는 플로우의 지연 시간을 최소화함으로써 서비스 응답속도를 단축시켜, 사용자에게 보다 쾌적하고 신속한 서비스를 제공할 수 있다. 또한, 사업자는 이러한 데이터센터의 신속 서비스 효과를 바탕으로 사용자들로부터 보다 많은 수익을 창출할 수 있다. 또한 본 발명은 ECN(Explicit Congestion Notification) 적용을 요구하지 않으며 적은 개수의 우선순위 큐를 요구하기 때문에 저가의 스위치로 구성된 데이터센터에서도 적용 가능하여 데이터센터 구축 및 유지 비용을 절감할 수 있다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 일 실시예에 따르는 시스템에 대한 구조도이다.

도 2는 본 발명의 일 실시예에 따르는 서버의 구성에 대한 블록도이다.

도3은 도 2의 송신부의 구성에 대한 블록도이다.

도 4는 도 2의 송신부 및 스위치부의 동작과정을 설명하기 위한 개념도이다.

도 5 내지 도 7은 본 발명의 일 실시예와 종래기술의 성능을 비교하기 위한 실험결과 그래프이다.

도 8은 본 발명의 일 실시예에 따르는 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0015] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0016] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0017] 본 명세서에 있어서 '부(部)'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1 개의 유닛이 2 개 이상의 하드웨어를 이용하여 실현되어도 되고, 2 개 이상의 유닛이 1 개의 하드웨어에 의해 실현되어도 된다. 한편, '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, '~부'는 어드레싱 할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 '~부'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성요소들과 '~부'들로 더 분리될 수 있다. 뿐만 아니라, 구성요소들 및 '~부'들은 디바이스 또는 보안 멀티미디어카드 내의 하나 또는 그 이상의 CPU들을 재생시키도록 구현될 수도 있다.
- [0018] 이하에서 언급되는 "사용자 단말"은 네트워크를 통해 서버나 타 단말에 접속할 수 있는 컴퓨터나 휴대용 단말기로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop) 등을 포함하고, 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet), LTE(Long Term Evolution) 통신 기반 단말, 스마트폰, 태블릿 PC 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다. 또한, "네트워크"는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN) 또는 부가가치 통신망(Value Added Network; VAN) 등과 같은 유선 네트워크나 이동 통신망(mobile radio communication network) 또는 위성 통신망 등과 같은 모든 종류의 무선 네트워크로 구현될 수 있다.
- [0019] 이하, 도면을 참조하여 본 발명의 일 실시예에 대하여 구체적으로 설명한다.
- [0020] 도 1을 참조하면, 본 발명의 일 실시예에 따르는 시스템은 데이터센터(10)일 수 있다.
- [0021] 데이터센터(10)는 네트워크를 통해 계층적으로 연결된 복수의 서버(100)들을 포함한다.
- [0022] 구체적으로, 사용자 단말로부터 쿼리를 수신하는 중앙 서버(100a) 및 중앙 서버(100)의 작업을 분산 수행하기 위한 복수의 하위 서버들(100b, 100c)로 구성될 수 있다.
- [0023] 각 온라인 서비스들의 접속 사용자 수 증가로 인한 트래픽 증가, 처리해야할 데이터 용량의 증가 등으로 인해 많은 서비스들(예를 들어, 포털 검색 서비스, SNS 서비스)이 데이터센터(10)를 활용하고 있다.
- [0024] 예를 들어, 특정 주제에 대한 검색어를 포함하는 쿼리를 사용자 단말로부터 전송되어 중앙 서버(100a)가 수신한 경우, 중앙 서버(100a)는 복수의 하위 서버들(100b, 100c)을 통해 분산 처리 한다. 즉, 각 하위 서버(100b)들로 패킷을 전송하여 쿼리에 대응하는 응답값이 있는지 검색한다.
- [0025] 이때, 각 최하위 서버(100c)와 중앙 서버(100a) 간의 경로마다 플로우가 정의되고, 중앙 서버(100a)에서 요청패킷이 플로우에 따라 전송된 후, 응답값을 포함하는 응답패킷이 최하위 서버(100c)에서 중앙 서버(100a)로 전송됨으로써 플로우가 완료된다. 즉, 플로우는 특정 송신지에서 특정 수신지로 전달되는 모든 패킷들의 집합에 관한 것으로서, 패킷들의 연속적인 흐름을 나타내는 정보이며, 하나 이상의 패킷으로 구성될 수 있다. 이때, 최초 패킷이 송신된 시점부터 최종의 패킷이 수신된 시점까지의 소요 시간을 플로우 완료 시간이라 할 수 있다. 그런데, 데이터센터(10)에서는 각 플로우마다 하위 서버(100b, 100c)로 패킷을 전송하고 모든 응답을 수신한 경우에만 쿼리에 대한 응답값을 사용자 단말로 전송한다. 여기서 응답값과 플로우는 각각의 최하위 서버(100c)가 중앙

서버(100a)로 전송하는 것으로서, 여러 플로우에서 발생된 응답값이 수신된 상태라고 하더라도 어느 한 플로우의 응답값이 수신되지 못한 상태라면, 쿼리는 종료되지 않는다. 즉, 어느 하나의 플로우에서만 지연이 발생하더라도 쿼리에서 지연이 발생하는 것이다.

- [0026] 한편, 플로우의 지연은 전송된 패킷에 손실이 발생되어, 재전송해야 하는 상황이 발생하는 경우 쿼리의 지연에 큰 영향을 미칠 수 있다.
- [0027] 이에 따라, 본 발명의 일 실시예는 재전송 패킷에 대한 스케줄링을 통해 전체 지연시간을 단축시킨다.
- [0028] 이하, 도 2 내지 도 4를 통하여 구체적으로 설명하도록 한다.
- [0029] 본 발명의 일 실시예에 따르는 서버(100)는 데이터센터(10)의 중앙 서버(100a)일 수 있으나, 본 발명의 일 실시예는 중앙 서버(100a)에만 한정되어 적용될 수 있는 기술은 아니기 때문에, 각각의 모든 서버(100)에서도 적용될 수 있다.
- [0030] 본 발명의 일 실시예에 따르는 서버(100)는 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법을 수행하는 프로그램(또는 애플리케이션)이 저장된 메모리와 위 프로그램을 실행하는 프로세서를 포함하여 구성될 수 있다. 여기서 프로세서는 메모리에 저장된 프로그램의 실행에 따라 다양한 기능을 수행할 수 있는데, 각 기능에 따라 프로세서에 포함되는 세부 구성요소들을 수신부(110), 송신부(120) 및 스위치부(130)로 나타낼 수 있다.
- [0031] 수신부(110)는 외부 객체로부터 쿼리를 수신할 수 있다. 예를 들어, 사용자 단말로부터 쿼리를 수신할 수 있다. 이때, 쿼리는 다양한 정보를 포함할 수 있다. 예를 들어, 특정 이슈에 대한 검색어를 포함하거나, SNS 상의 친구를 검색하는 요청정보를 포함할 수도 있다. 또는, 수신부(110)는 외부 서버로부터 특정 응답값을 요청하는 패킷을 수신할 수도 있다.
- [0032] 송신부(120)는 쿼리에 대한 응답을 위해, 패킷을 생성하여 다른 서버 (또는 하위 서버)로 전송하되, 대상 패킷이 재전송 패킷인 경우 다른 패킷보다 재전송 패킷에 대하여 높은 우선순위를 지정한다.
- [0033] 송신부(120)는 패킷의 플로우 크기 판단부(121)와 우선순위 지정부(122)를 포함한다.
- [0034] 패킷의 플로우 크기 판단부(121)는 전송해야 할 대상 패킷이 재전송 패킷인지 판단한다. 이때, 재전송 타이머 또는 TCP NewReno에서 제공되는 Fast Retransmit 기법을 통하여 대상 패킷의 재전송 패킷 여부를 판단할 수 있다. 재전송 타이머를 이용하는 경우, 대상 패킷이 하위 서버(100b 또는 100c)로 전송된 후, 미리 설정된 시간 이내에 하위 서버(100b 또는 100c)로부터 수신확인(ACK : acknowledgement) 메시지를 수신하지 못하는 경우, 전송된 대상 패킷을 재전송 패킷으로 판단한다.
- [0035] 이어서, 패킷의 플로우 크기 판단부(121)는 대상 패킷이 재전송 패킷인 경우, 대상 패킷의 플로우 크기와 미리 설정된 임계값을 비교한다. 이때, 대상 패킷의 플로우의 크기는 대상 패킷의 플로우에 대한 수신확인 메시지의 누적 데이터의 크기를 통해 판단될 수 있다. 수신확인 메시지는 패킷을 수신한 서버(100)가 패킷을 전송한 서버(100)로 전송하는 확인 신호로서, 중앙서버(100a)는 수신확인 메시지가 발생될 때마다 이전부터 현재까지의 수신확인 메시지의 데이터의 크기를 누적 기록한다.
- [0036] 우선순위 지정부(122)는 재전송 패킷인지 여부 및 대상 패킷의 플로우 크기와 임계값과의 비교 결과에 따라 대상 패킷에 대한 우선순위값을 지정한다. 도 4를 참조하면, 재전송 패킷인 것으로 판단된 대상 패킷들 중 수신확인 메시지의 누적 데이터의 크기가 임계값 이하인 대상 패킷들에 대해서는 제 1 우선순위값을 지정한다. 제 1 우선순위값은 0일 수 있다. 한편, 누적 데이터의 크기가 임계값 초과인 대상 패킷들에 대해서는 제 2 우선순위값을 지정한다. 제 2 우선순위값은 1일 수 있다. 통상적으로 일반적인 패킷의 수신확인 메시지의 누적 데이터 크기는 일정 크기 이하이기 때문에, 임계값 초과인 대상 패킷들에 대해서는 1이 지정된다. 또한, 대상 패킷이 재전송 패킷이 아닌 경우에도 일반 패킷으로 간주하여 제 2 우선순위값이 지정될 수 있다.
- [0037] 이때, 제 1 및 제 2 우선순위값은 상기 대상 패킷의 IP 헤더의 DSCP(Differentiated Service Code Point) 필드 내에 기록될 수 있다.
- [0038] 스위치부(130)는 지정된 우선순위에 따라 순차적으로 패킷을 전송한다. 스위치부(130)는 도 4를 참조하면, 입력 포트, 스위치 패브릭 및 출력포트를 포함할 수 있다. 이때, 입력포트 내에는 분류기가 포함되어 있어, 분류기를 통해 우선순위값에 따라 패킷의 전송순서가 결정될 수 있다. 즉, 0과 1의 우선순위값을 갖는 패킷들을 별도로 구분하여 우선순위 큐에 정렬한 뒤, 0의 우선순위값이 지정된 패킷들을 먼저 전송하고, 1의 우선순위값이 지정

된 패킷들을 후차적으로 전송한다. 그에 따라, 재전송 패킷이 먼저 전송된다.

- [0039] 이와 같이, 본 발명의 일 실시예는 두 개의 우선순위 큐만 보유하고 있는 스위치 장치를 통해서도 지연시간 단축의 효과를 제공할 수 있기 때문에 저가의 스위치 장비 활용이 가능하다. 이를 통해 데이터센터(10) 구축시 비용을 절감시킬 수 있다.
- [0040] 한편, 도 5 내지 도 7은 ns-3 네트워크 시뮬레이터를 통해 시뮬레이션을 실시하여 본 발명과 종래기술들의 네트워크 트래픽 로드 (Load) 변화에 따른 쿼리 종료 시간 (Query Completion Times)을 나타낸 그래프이다.
- [0041] 본 실험예에서는 1Gbps 네트워크 인터페이스를 장착한 54 대의 서버(100)가 존재하는 3계층 Fat-tree 네트워크 토폴로지가 사용되었으며, 각 스위치 포트는 100KB 크기의 버퍼를 가진다. 또한 최소 재전송 타이머 값 (RTOmin)은 10ms로 설정되었다. 라우팅 및 로드밸런싱 방법으로는 표준 ECMP 라우팅 알고리즘이 사용되었으며, 네트워크 스위치로는 두 개의 우선순위 큐를 가진 스위치가 사용되었다.
- [0042] 그래프에서 TCP는 TCP NewReno를 의미하며, TFS는 TCP NewReno를 TCP 프로토콜로 사용한 본 발명의 일 실시예를 의미하며, DCTCP(Data Center TCP)는 ECN을 사용하는 TCP 프로토콜을 의미하며, TFS DCTCP는 DCTCP를 TCP 프로토콜로 사용하는 본 발명의 다른 실시예를 의미한다.
- [0043] 그래프의 y축값과 관련하여, 도 5 는 평균 쿼리 종료 시간, 도6은 쿼리 종료 시간의 중간값, 도 7은 99.9 백분위수에 해당하는 쿼리 종료 시간을 나타내고 있다. 도 5 내지 도 7을 참조하면, 대부분의 트래픽 로드에서 본 발명의 일 실시예가 TCP보다 더 나은 쿼리 종료 시간을 보임을 알 수 있으며 본 발명의 다른 실시예는 DCTCP보다 더 나은 쿼리 종료 시간을 보임을 알 수 있다. 최종적으로, 본 발명의 다른 실시예가 가장 우수한 성능을 나타냄을 알 수 있다.
- [0044] 이하, 도 8을 참조하여, 본 발명의 일 실시예에 따르는 패킷의 재전송으로 인한 지연을 단축시키기 위한 패킷의 플로우 스케줄링 방법에 대하여 구체적으로 설명하도록 한다.
- [0045] 이하의 방법은 상술한 서버(100)에 의해 수행되는 것이므로, 생략된 내용이 있다고 하더라도 상술한 내용으로 같음하도록 한다.
- [0046] 먼저, 본 발명의 일 실시예에 따르는 서버(100)는 대상 패킷이 재전송 패킷인지 판단한다(S110).
- [0047] 이어서, 서버(100)는 패킷의 플로우 크기가 미리 설정된 임계값 이하인지 비교 판단한다(S120).
- [0048] 만약 패킷의 플로우 크기가 미리 설정된 임계값 이하인 경우, 대상 패킷의 IP 헤더의 DSCP 필드 내에 높은 우선순위값을 기록한다(S140).
- [0049] 그러나, 패킷의 플로우 크기가 미리 설정된 임계값 초과인 경우, 대상 패킷의 IP 헤더의 DSCP 필드 내에 낮은 우선순위값을 기록한다(S130). 또한, 대상 패킷이 재전송 패킷이 아닌 경우에도, DSCP 필드 내에 낮은 우선순위값을 기록한다.
- [0050] 서버(100)는 지정된 우선순위에 따라 스케줄링하여 패킷을 전송한다(S150). 그 결과 재전송 패킷 중에서도 플로우의 크기가 큰 패킷들이 먼저 전송되고, 일반 패킷과 플로우 크기가 작은 패킷들이 후순위로 전송되면서, 패킷 손실에 의한 플로우 지연시간을 최소화할 수 있다.
- [0051] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0052] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.
- [0053] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해

할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

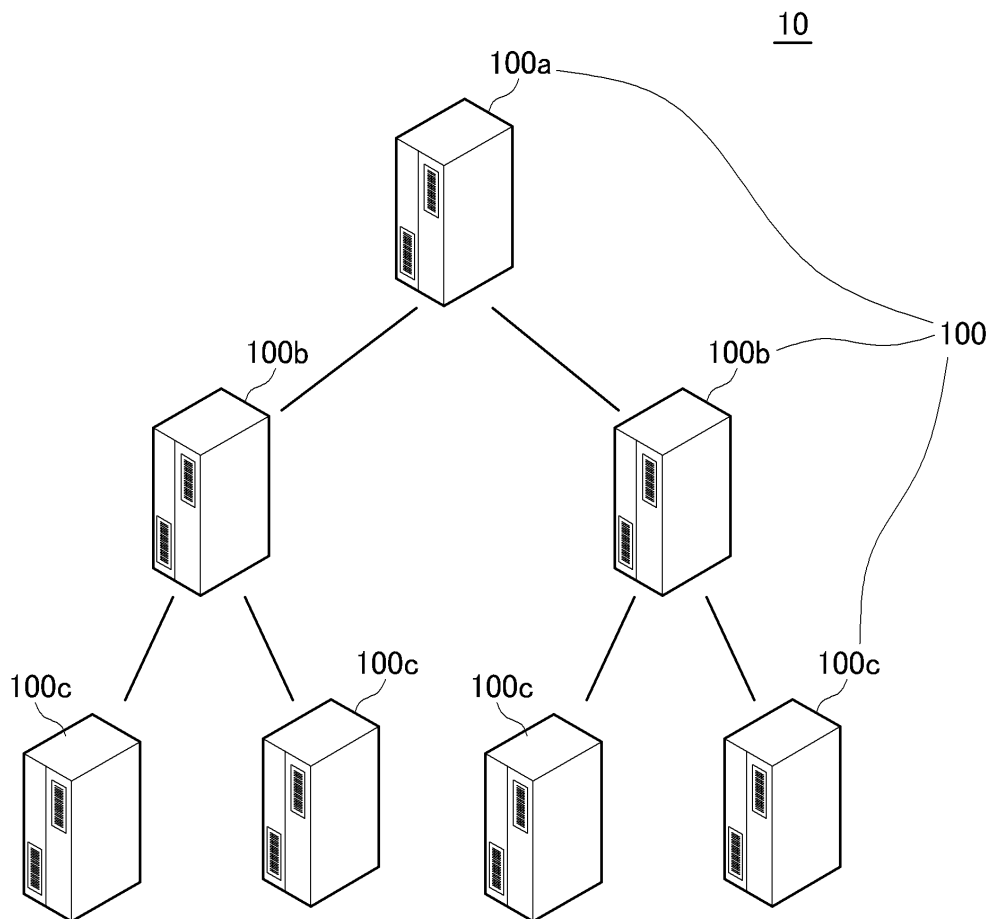
[0054] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

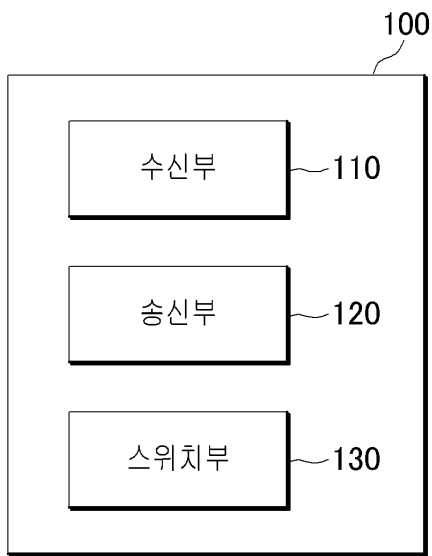
[0055] 100: 서버 110 : 수신부
120 : 송신부 130 : 스위치부

도면

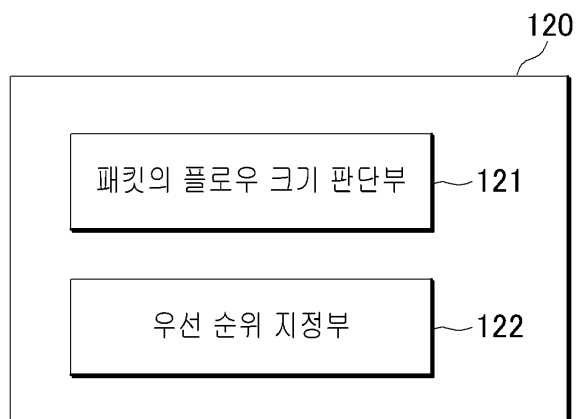
도면1



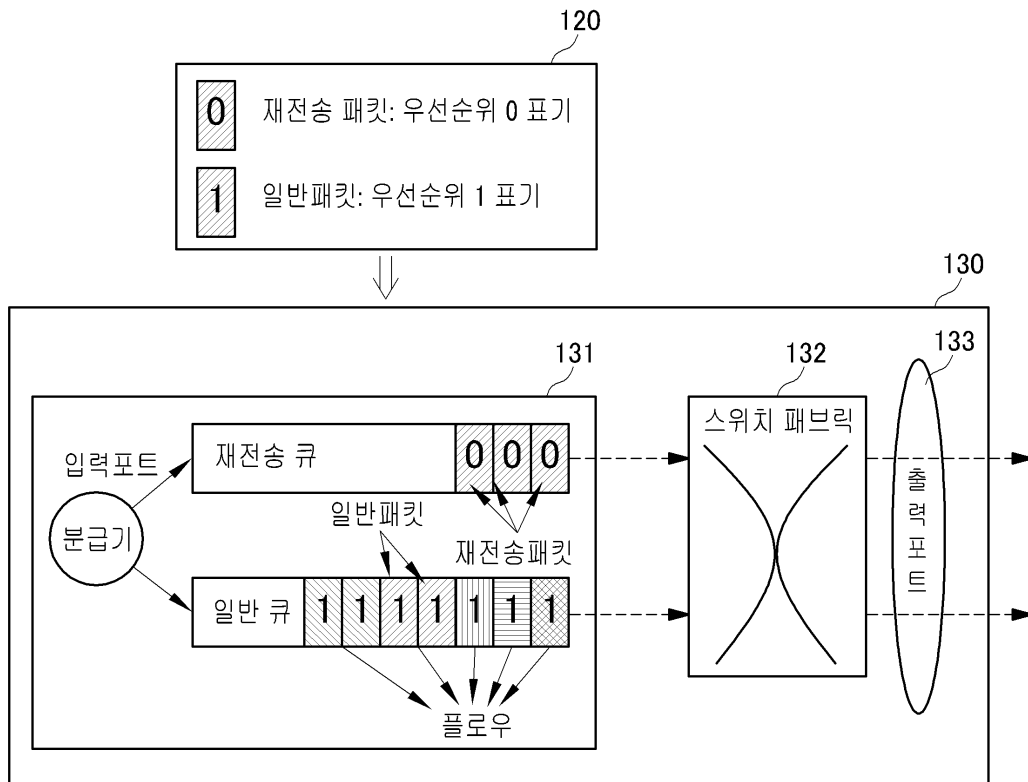
도면2



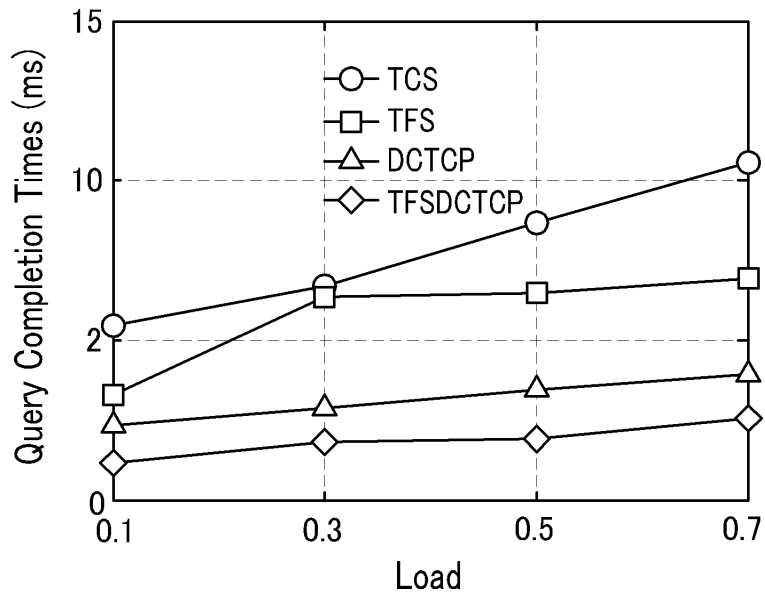
도면3



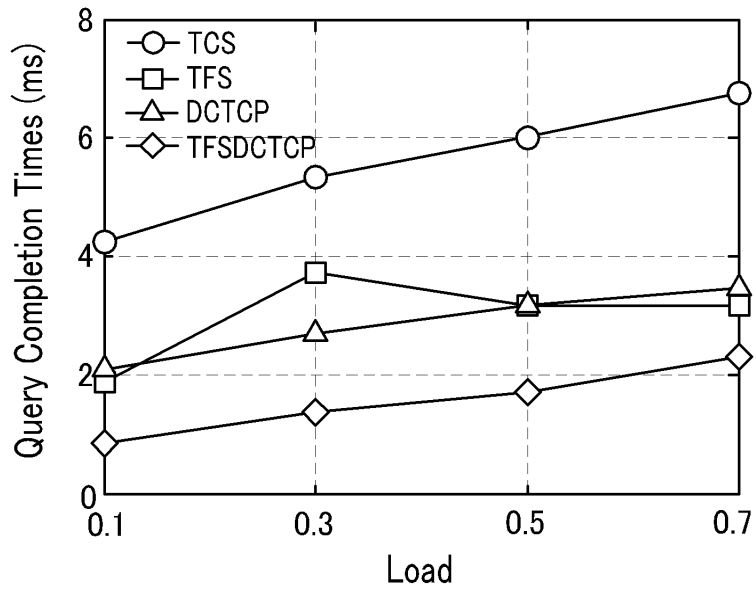
도면4



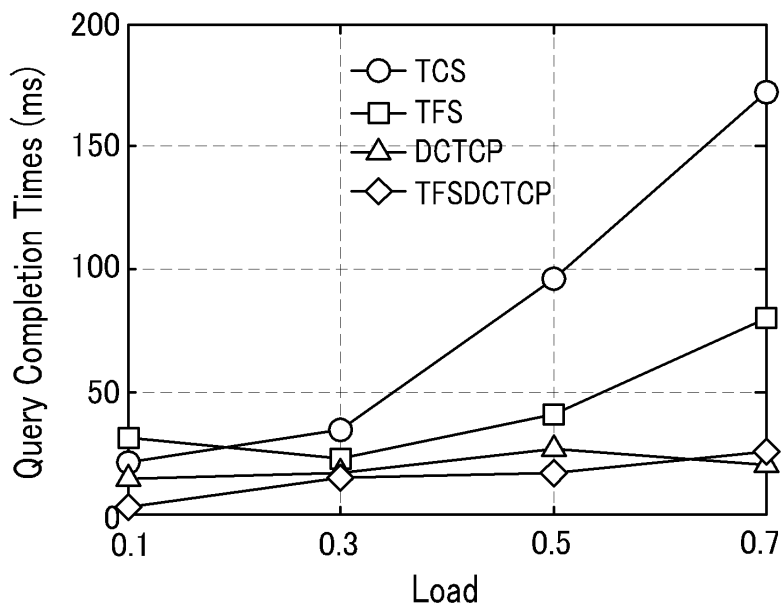
도면5



도면6



도면7



도면8

