(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년01월04일
(11) 등록번호 10-1935528
(24) 등록일자 2018년12월28일

(51) 국제특허분류(Int. Cl.)
G06Q 50/30 (2012.01) *G06F 21/62* (2013.01)
(52) CPC특허분류
G06Q 50/30 (2015.01)
G06F 21/6245 (2013.01)
(21) 출원번호 10-2017-0160491
(22) 출원일자 2017년11월28일
심사청구일자 2017년11월28일
(56) 선행기술조사문헌
JP2008102664 A*
JP2016012074 A
JP2003196784 A
박철희 외 2인_데이터의 범용적인 활용을 위한 차분 프라이버시를 만족하는 데이터 공개 기법_한국정보과학회 2017년 한국컴퓨터종합학술대회 논문집_2017.06_pp.1036~1038
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
서강대학교 산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박석
서울특별시 종로구 평창12길 16-6, 2동 202호 (평창동, 평창삼호빌라)
조건형
경기도 성남시 수정구 위례순환로 17, 3110동 1002호 (창곡동, 위례 31단지)
(74) 대리인
유미특허법인

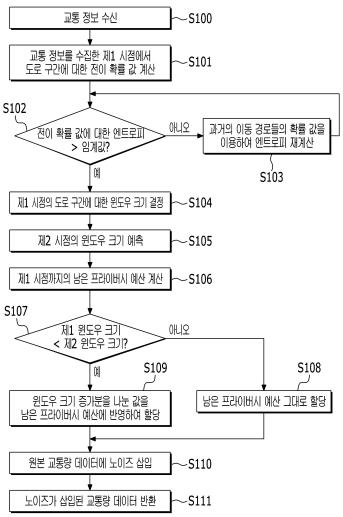
전체 청구항 수 : 총 13 항 심사관 : 박장환

(54) 발명의 명칭 **차분 프라이버시를 적용한 교통량 배포 시스템 및 방법**

(57) 요 약

교통량 배포 시스템이 도로 교통망에서 교통량을 배포하는 방법으로서, 제1 시점에 복수의 도로 구간 정보와 해당 도로 구간의 원본 교통량 데이터를 포함하는 교통 정보를 수신하고, 제1 시점의 도로 구간에 대한 제1 윈도우 크기를 계산한다.제1 시점에 계산된 제1 윈도우 크기 또는 제1 윈도우 크기와 제1 시점 이전의 제2 시점에 이미 구해진 윈도우 크기 중 어느 하나를 토대로, 제1 시점 이후 시간인 제3 시점에 대한 제2 윈도우 크기를 예측한다. 제1 윈도우 크기와 제2 윈도우 크기를 토대로 제1 시점에 할당할 프라이버시 예산을 결정하고, 결정한 프라이버시 예산을 토대로 원본 교통량 데이터에 노이즈가 삽입된 노이즈 삽입 교통량 데이터를 반환한다.

대 표 도 - 도3



이 발명을 지원한 국가연구개발사업
과제고유번호 2017-0-00498
부처명 미래창조과학부
연구관리전문기관 IITP(정보통신기술진흥센터)
연구사업명 정보통신·방송 연구개발 사업 계획서
연구과제명 차분 프라이버시 기반의 비식별화 기술 개발
기 여 율 1/1
주관기관 서강대학교 산학협력단
연구기간 2017.04.01~2017.12.31

명 세 서

청구범위

청구항 1

교통량 배포 시스템이 도로 교통망에서 교통량을 배포하는 방법으로서,

제1 시점에 복수의 도로 구간 정보와 해당 도로 구간의 원본 교통량 데이터를 포함하는 교통 정보를 수신하고, 상기 제1 시점에서 복수의 도로 구간에 대하여 계산한 전이 확률 값을 미리 설정한 임계값과 비교한 결과에 따라 상기 제1 시점의 도로 구간에 대한 제1 윈도우 크기를 계산하는 단계,

상기 제1 시점에 계산된 제1 윈도우 크기 또는 상기 제1 윈도우 크기와 제1 시점 이전의 제2 시점에 이미 구해진 윈도우 크기 중 어느 하나를 토대로, 상기 제1 시점 이후 시간인 제3 시점에 대한 제2 윈도우 크기를 예측하는 단계,

상기 제1 윈도우 크기와 제2 윈도우 크기를 토대로 상기 제1 시점에 할당할 프라이버시 예산을 결정하는 단계, 그리고

상기 결정한 프라이버시 예산을 토대로 상기 원본 교통량 데이터에 노이즈가 삽입된 노이즈 삽입 교통량 데이터를 반환하는 단계

를 포함하는 교통량 배포 방법.

청구항 2

제1항에 있어서,

상기 제1 윈도우 크기를 계산하는 단계는,

상기 계산한 전이 확률 값에 대한 엔트로피가 미리 설정한 임계값 보다 큰지 확인하는 단계, 그리고

상기 전이 확률 값이 임계값 보다 크면, 상기 제1 윈도우 크기를 미리 설정된 값으로 결정하는 단계

를 포함하는 교통량 배포 방법.

청구항 3

제2항에 있어서,

상기 전이 확률 값은 상기 제1 시점에서 복수의 도로 구간 각각의 전이 행렬로 계산되고,

상기 전이 확률 값에 대한 엔트로피는 상기 전이 확률 값과 전이 확률 값의 로그(log) 값으로 계산되는 교통량 배포 방법.

청구항 4

제2항에 있어서,

상기 전이 확률 값이 임계값 보다 작으면,

상기 전이 확률 값이 임계값 보다 작은 도로 구간의 상기 제1 시점 이전의 과거 이동 경로들의 확률 값을 이용하여 엔트로피를 재계산하는 단계

를 포함하는 교통량 배포 방법.

청구항 5

제1항에 있어서,

상기 프라이버시 예산을 결정하는 단계는,

상기 제1 시점까지의 남은 프라이버시 예산을 계산하는 단계

를 더 포함하는 교통량 배포 방법.

청구항 6

제5항에 있어서,

상기 프라이버시 예산을 결정하는 단계는,

상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 크면, 상기 남은 프라이버시 예산 전부를 상기 제1 시점의 프라이버시 예산으로 할당하는 단계

를 포함하는 교통량 배포 방법.

청구항 7

제6항에 있어서,

상기 프라이버시 예산을 결정하는 단계는,

상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 작으면, 상기 제1 윈도우 크기에서 제2 윈도우 크기의 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여, 제1 시점의 프라이버시 예산으로 할당하는 단계

를 포함하는 교통량 배포 방법.

청구항 8

제7항에 있어서,

상기 노이즈 삽입 교통량 데이터를 반환하는 단계는,

상기 제1 시점에 할당된 프라이버시 예산을 토대로, 상기 원본 교통량 데이터에 삽입할 노이즈 량을 계산하는 단계, 그리고

상기 계산된 노이즈량만큼 상기 원본 교통량 데이터에 노이즈를 삽입하는 단계

를 포함하는 교통량 배포 방법.

청구항 9

도로 교통망에서 교통량을 배포하는 교통량 배포 시스템으로서,

제1 시점에 수집되어 복수의 도로 구간 정보와 해당 도로 구간의 원본 교통량 데이터를 포함하는 교통 정보를 토대로, 상기 제1 시점에서 복수의 도로 구간에 대한 전이 확률 값을 계산하고, 상기 계산한 전이 확률 값에 대한 엔트로피가 미리 설정한 임계값과 비교하여 미리 설정된 값 또는 상기 제1 시점 이전의 과거 이동 경로들의 확률 값을 토대로 계산된 엔트로피를 이용하여, 상기 제1 시점의 도로 구간에 대한 제1 윈도우 크기를 계산하는 윈도우 크기 계산 모듈,

상기 제1 시점에 계산된 제1 윈도우 크기 또는 상기 제1 윈도우 크기와 제1 시점 이전의 제2 시점에 이미 구해진 윈도우 크기 중 어느 하나를 토대로, 상기 제1 시점 이후 시간인 제3 시점에 대한 제2 윈도우 크기를 예측하는 윈도우 크기 예측 모듈, 그리고

상기 계산한 제1 윈도우 크기와 예측한 제2 윈도우 크기를 이용하여, 상기 제1 시점에 상기 원본 교통량 데이터에 삽입할 노이즈 량을 계산하여 노이즈가 삽입된 교통량 데이터를 배포하는 교통량 배포 모듈

을 포함하는 교통량 배포 시스템.

청구항 10

삭제

청구항 11

제9항에 있어서,

상기 교통량 배포 모듈은,

상기 제1 윈도우 크기와 제2 윈도우 크기를 토대로, 상기 제1 시점에 각 도로 구간에 대한 윈도우에서 사용하고 남은 프라이버시 예산을 구하는 교통량 배포 시스템.

청구항 12

제11항에 있어서,

상기 교통량 배포 모듈은,

상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 크면, 상기 남은 프라이버시 예산 전부를 상기 제1 시점의 프라이버시 예산으로 할당하는 교통량 배포 시스템.

청구항 13

제12항에 있어서,

상기 교통량 배포 모듈은,

상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 작으면, 상기 제1 윈도우 크기에서 제2 윈도우 크기의 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여, 제1 시점의 프라이버시 예산으로 할당하는 교통량 배포 시스템.

청구항 14

제13항에 있어서,

상기 교통량 배포 모듈은,

상기 제1 시점에 할당된 프라이버시 예산을 토대로, 상기 원본 교통량 데이터에 삽입할 노이즈 량을 계산하는 교통량 배포 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 차분 프라이버시를 적용한 교통량 배포 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 최근 교통 분석, 트렌드 분석과 같이 많은 분야에서 주기적으로 배포되는 스트림 데이터를 필요로 하고 있다. 이러한 스트림 데이터에는 민감한 개인 정보가 내포되어 있기 때문에, 개인의 식별 문제와 같은 프라이버시 침해가 발생할 수 있다. 따라서 개인 정보를 보호하면서도 데이터의 유용성을 노이즈 삽입을 통해 조절할 수 있는 차분 프라이버시(differential privacy)가 사용되고 있다.

[0003] 그러나, 스트림 데이터는 무한대의 시간 크기를 갖기 때문에, 배포된 모든 데이터를 보호하려면 많은 노이즈가 데이터에 삽입되어, 데이터의 유용성이 떨어진다는 단점이 있다. 이러한 단점을 극복하기 위하여 최근 윈도우 내에 들어있는 교통량 데이터에 대해서만 차분 프라이버시를 적용하는 연구가 진행되고 있다.

[0004] 그러나, 종래의 차분 프라이버시를 적용하더라도 윈도우의 크기가 고정되어 있기 때문에 데이터가 수집된 시간대의 특징이 고려되지 않고, 도로 구간간의 상관 관계를 간과하여 불필요한 노이즈가 많이 삽입되는 단점이 있다.

선행기술문헌

특허문헌

(특허문헌 0001) 한국등록특허공보 제10-1753097호 (등록일: 2017.06.27.)

(특허문헌 0002) 한국등록특허공보 제10-0414356호 (등록일: 2003.12.23.)

발명의 내용

해결하려는 과제

[0005] 따라서, 본 발명은 도로 구간의 상관 관계와 시간대를 고려한 가변 길이 윈도우 기법을 이용하는 차분 프라이버시를 적용한 교통량 배포 시스템 및 방법을 제공한다.

과제의 해결 수단

[0006] 상기 본 발명의 기술적 과제를 달성하기 위한 본 발명의 하나의 특징인 교통량 배포 시스템이 도로 교통망에서 교통량을 배포하는 방법으로서,

[0007] 제1 시점에 복수의 도로 구간 정보와 해당 도로 구간의 원본 교통량 데이터를 포함하는 교통 정보를 수신하고, 상기 제1 시점의 도로 구간에 대한 제1 윈도우 크기를 계산하는 단계, 상기 제1 시점에 계산된 제1 윈도우 크기 또는 상기 제1 윈도우 크기와 제1 시점 이전의 제2 시점에 이미 구해진 윈도우 크기 중 어느 하나를 토대로, 상기 제1 시점 이후 시간인 제3 시점에 대한 제2 윈도우 크기를 예측하는 단계, 상기 제1 윈도우 크기와 제2 윈도우 크기를 토대로 상기 제1 시점에 할당할 프라이버시 예산을 결정하는 단계, 그리고 상기 결정한 프라이버시 예산을 토대로 상기 원본 교통량 데이터에 노이즈가 삽입된 노이즈 삽입 교통량 데이터를 반환하는 단계를 포함한다.

[0008] 상기 제1 윈도우 크기를 계산하는 단계는, 상기 제1 시점에서 복수의 도로 구간에 대한 전이 확률 값을 계산하는 단계, 상기 계산한 전이 확률 값에 대한 엔트로피가 미리 설정한 임계값 보다 큰지 확인하는 단계, 그리고 상기 전이 확률 값이 임계값 보다 크면, 상기 제1 윈도우 크기를 미리 설정된 값으로 결정하는 단계를 포함할 수 있다.

[0009] 상기 전이 확률 값은 상기 제1 시점에서 복수의 도로 구간 각각의 전이 행렬로 계산되고, 상기 전이 확률 값에 대한 엔트로피는 상기 전이 확률 값과 전이 확률 값의 로그(log) 값으로 계산될 수 있다.

[0010] 상기 전이 확률 값이 임계값 보다 작으면, 상기 전이 확률 값이 임계값 보다 작은 도로 구간의 상기 제1 시점 이전의 과거 이동 경로들의 확률 값을 이용하여 엔트로피를 재계산하는 단계를 포함할 수 있다.

[0011] 상기 프라이버시 예산을 결정하는 단계는, 상기 제1 시점까지의 남은 프라이버시 예산을 계산하는 단계를 포함할 수 있다.

[0012] 상기 프라이버시 예산을 결정하는 단계는, 상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 크면, 상기 남은 프라이버시 예산 전부를 상기 제1 시점의 프라이버시 예산으로 할당하는 단계를 포함할 수 있다.

[0013] 상기 프라이버시 예산을 결정하는 단계는, 상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 작으면, 상기 제1 윈도우 크기에서 제2 윈도우 크기의 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여, 제1 시점의 프라이버시 예산으로 할당하는 단계를 포함할 수 있다.

[0014] 상기 노이즈 삽입 교통량 데이터를 반환하는 단계는, 상기 제1 시점에 할당된 프라이버시 예산을 토대로, 상기 원본 교통량 데이터에 삽입할 노이즈 량을 계산하는 단계, 그리고 상기 계산된 노이즈량만큼 상기 원본 교통량 데이터에 노이즈를 삽입하는 단계를 포함할 수 있다.

[0015] 상기 본 발명의 기술적 과제를 달성하기 위한 본 발명의 또 다른 특징인 도로 교통망에서 교통량을 배포하는 교통량 배포 시스템으로서,

[0016] 제1 시점에 수집되어 복수의 도로 구간 정보와 해당 도로 구간의 원본 교통량 데이터를 포함하는 교통 정보를 토대로, 상기 제1 시점의 도로 구간에 대한 제1 윈도우 크기를 계산하는 윈도우 크기 계산 모듈, 상기 제1 시점에 계산된 제1 윈도우 크기 또는 상기 제1 윈도우 크기와 제1 시점 이전의 제2 시점에 이미 구해진 윈도우 크기 중 어느 하나를 토대로, 상기 제1 시점 이후 시간인 제3 시점에 대한 제2 윈도우 크기를 예측하는 윈도우 크기 예측 모듈, 그리고 상기 계산한 제1 윈도우 크기와 예측한 제2 윈도우 크기를 이용하여, 상기 제1 시점에 상기

원본 교통량 데이터에 삽입할 노이즈 량을 계산하여 노이즈가 삽입된 교통량 데이터를 배포하는 교통량 배포 모듈을 포함한다.

[0017] 상기 윈도우 크기 계산 모듈은, 상기 제1 시점에서 복수의 도로 구간에 대한 전이 확률 값을 계산하고, 상기 계산한 전이 확률 값에 대한 엔트로피가 미리 설정한 임계값과 비교하여 미리 설정된 값 또는 상기 제1 시점 이전의 과거 이동 경로들의 확률 값을 토대로 계산된 엔트로피를 이용하여 상기 제1 윈도우 크기를 계산할 수 있다.

[0018] 상기 교통량 배포 모듈은, 상기 제1 윈도우 크기와 제2 윈도우 크기를 토대로, 상기 제1 시점에 각 도로 구간에 대한 윈도우에서 사용하고 남은 프라이버시 예산을 구할 수 있다.

[0019] 상기 교통량 배포 모듈은, 상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 크면, 상기 남은 프라이버시 예산 전부를 상기 제1 시점의 프라이버시 예산으로 할당할 수 있다.

[0020] 상기 교통량 배포 모듈은, 상기 제1 윈도우 크기가 상기 제2 윈도우 크기보다 작으면, 상기 제1 윈도우 크기에서 제2 윈도우 크기의 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여, 제1 시점의 프라이버시 예산으로 할당할 수 있다.

[0021] 상기 교통량 배포 모듈은, 상기 제1 시점에 할당된 프라이버시 예산을 토대로, 상기 원본 교통량 데이터에 삽입할 노이즈 량을 계산할 수 있다.

발명의 효과

[0022] 본 발명에 따르면 도로 구간 간의 상관 관계와 시간대를 고려하여 윈도우 길이를 가변으로 하여 적용하기 때문에, 원본 교통 데이터에 대한 오차가 적어진다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 실시예에 따른 교통량 배포 시스템이 적용된 환경의 예시도이다.

도 2는 본 발명의 실시예에 따른 교통량 배포 시스템의 구조도이다.

도 3은 본 발명의 실시예에 따른 교통량 배포 방법에 대한 흐름도이다.

도 4는 본 발명의 실시예에 따른 도로 교통망과 원본 교통량 데이터를 나타낸 예시도이다.

도 5는 본 발명의 실시예에 따른 윈도우 크기 계산 알고리즘을 나타낸 예시도이다.

도 6은 본 발명의 실시예에 따른 교통량 배포 알고리즘을 나타낸 예시도이다.

도 7은 본 발명의 실시예에 따른 오차 값을 나타낸 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 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.

[0025] 이하 도면을 참조로 하여, 본 발명의 실시예에 따른 교통량 배포 시스템 및 이를 이용한 교통량 배포 방법에 대해 상세히 설명한다.

[0026] 본 발명의 실시예에 대해 설명하기 앞서, 발명의 실시예에 대한 설명에 등장하는 용어인 '윈도우'를 현재 도로 구간으로 차량이 도착할 때, 다양한 경로들이 포함될 수 있는 최소한의 시간으로 정의한다.

[0027] 도 1은 본 발명의 실시예에 따른 교통량 배포 시스템이 적용된 환경의 예시도이다.

[0028] 도 1에 도시된 바와 같이, 교통량 배포 시스템(100)은 교통량을 수집하여 제공하는 교통 정보 제공 시스템(200)과, 교통 정보를 토대로 계산되어 배포되는 교통량을 수신하는 단말(300)과 연동한다. 여기서, 교통 정보 제공 시스템(200)과 단말(300)의 형태는 어느 하나로 한정하지 않는다. 그리고 본 발명의 실시예에서는 교통량 배포 시스템(100)과 교통 정보 제공 시스템(200)이 물리적으로 별도의 시스템으로 구현되어 있는 것을 예로 하여 설명하나, 하나의 시스템에 통합된 형태로 구현될 수도 있다.

[0029] 교통 정보 제공 시스템(200)은 교통 정보를 수집하여 교통량 배포 시스템(100)에 제공한다. 교통 정보 제공 시스템(200)이 수집하는 교통 정보에는 교통 정보가 수집된 시간 정보, 해당 시간에 도로 구간에 존재하는 차량 정보, 해당 차량이 이동하는 이동 경로 정보, 도로 구간 정보, 교통량 데이터 등을 포함한다.

[0030] 여기서, 실제 교통 정보가 수집된 도로 교통망과 교통량 배포 시스템(100)이 수신한 교통 정보에 포함된 교통량 데이터에 대해 도 4를 참조로 먼저 설명한다.

[0031] 도 4는 본 발명의 실시예에 따른 도로 교통망과 원본 교통량 데이터를 나타낸 예시도이다.

[0032] 도 4의 (a)는 교통 데이터가 수집된 도로 교통망을 간략하게 나타낸 예시도이고, 도 4의 (b)는 원본 교통량 데이터를 나타낸 예시도이다.

[0033] 도 4의 (a)에 나타낸 바와 같이 각 도로 구간에는 식별 정보(R1~R5)가 부여되어 있고, 임의의 시점 t에서 배포되는 원본 교통량 데이터는 도 4의 (b)의 열에 나타낸 바와 같다. 배포된 데이터가 누적되면 될수록, 도 4의 (b)와 같이 테이블 형태가 된다.

[0034] 그리고 임의의 자연수 w에 대해, 길이가 t인 제1 이동 경로(S_t)와 제2 이동 경로(S_t')가 w-이웃(w-neighboring)인 것은 다음 두 조건을 만족해야 한다.

[0035] 제1 조건은 $1 \le i_1 < i_2 \le t$인 i_1과 i_2에 대해, $i_2 - i_1 + 1 \le w$을 만족해야 하고,

[0036] 제2 조건은 $S_t[i_1] \neq S_t'[i_1]$이고, $S_t[i_2] \neq S_t'[i_2]$인 i_1과 i_2가 존재해야 한다. 여기서, i는 i 번째 인덱스를 의미한다. 즉, S_t는 길이가 t인 이동 경로 St에서 차량이 i번째에 방문한 장소 또는 도로를 의미한다.

[0037] 예를 들어, t = 5이고 w = 3이라면, S_t={R_1, R_3, R_5, R_2, R_7}와 S_t'={R_1, R_4, R_6, R_3, R_7}인 것을 예로 하여 설명한다. 여기서, S_t와 S_t'은 각각 길이가 5인 이동 경로를 나타내는 것이고, R_k는 해당 차량이 방문한 장소 혹은 도로를 나타낸다. S_t 와 S_t'은 집합으로 표시되어 있는데, 각 집합의 원소의 순서대로 차량이 방문한 장소 혹은 도로가 나열된 것이다.

[0038] 그리고, w=3인 경우 i_1=2에 대해 $S_t[i_1]$은 R_3이고, $S_t[i_1]$은 R_4가 된다. 이와 마찬가지로 w=3인 경우 와 i_2=4에 대해 대해 $S_t[i_2]$는 R_2이고, $S_t[i_2]$은 R_3가 된다. 따라서 $S_t[i_1] \neq S_t'[i_1]$, $S_t[i_2] \neq S_t'[i_2]$, $i_2 - i_1 + 1 \le 3$이므로, 3-이웃이 된다.

[0039] 여기서, 임의의 알고리즘 A가 w-event를 만족한다는 것은, w-이웃인 두 경로 S_t, S_t'를 입력으로 갖고, 모든 출력 값 r에 대해 다음 수학식 1을 만족한다는 것이다.

수학식 1

[0040]
$$\log\left(\frac{\Pr(\mathcal{A}(S_t) = r)}{\Pr(\mathcal{A}(S_t') = r)}\right) \le \epsilon$$

[0041] 예를 들어, 도 4의 (b)의 교통량 데이터에 3-event ϵ -차분 프라이버시가 만족된다고 가정한다. 그러면 $t=t_3$에서 교통량 데이터가 배포될 때, t3에 해당하는 윈도우는 t1, t2, t3 열을 감싸며 표시된 W_{t_3}이고, $\epsilon_1 + \epsilon_2 + \epsilon_3 \le \epsilon$를 만족한다. 이는 교통량 데이터가 배포되는 시점 t부터 이전의 t-w-1까지 각 차량의 이동 경로가 보호된다는 것을 의미한다.

[0042] 한편, 상기 도 1을 이어 설명하면 교통량 배포 시스템(100)은 교통 정보 제공 시스템(200)으로부터 교통 정보를 전달 받아, 교통 정보에 포함된 시간 정보와 도로 구간 정보를 토대로 윈도우 길이를 가변한다. 그리고, 가변한 윈도우 길이가 적용된 교통 정보를 가공하여 교통량 정보로 생성한다. 교통량 배포 시스템(100)이 생성한 교통량 정보는 단말(300)에 배포된다.

[0043] 여기서, 교통 정보를 가공하여 교통량 정보로 생성하는 교통량 배포 시스템(100)의 구조에 대해 도 2를 참조로 설명한다.

[0044] 도 2는 본 발명의 실시예에 따른 교통량 배포 시스템의 구조도이다.

[0045] 도 2에 도시된 바와 같이, 교통량 배포 시스템(100)은 인터페이스(110), 프로세서(120), 그리고 메모리(130)를 포함한다. 그리고 프로세서(120)는 윈도우 크기 계산 모듈(121), 윈도우 크기 예측 모듈(122) 그리고 교통량 배포 모듈(123)을 포함한다.

[0046] 인터페이스(110)는 교통 정보 제공 시스템(200) 및 단말(300)과 연동한다. 그리고, 교통 정보 제공 시스템(200)으로부터 전송되는 교통 정보를 수신하고, 프로세서(120)가 가공한 교통량을 단말(300)에 전달한다. 여기서, 교통 정보에는 교통 정보가 수집된 시간 정보, 해당 시간에 도로 구간에 존재하는 차량 정보, 해당 차량이 이동하는 이동 경로 정보, 도로 구간 정보 등이 포함되어 있다.

[0047] 윈도우 크기 계산 모듈(121)은 인터페이스(110)를 통해 수신한 교통 정보에서 도로 구간을 확인한다. 그리고, 확인한 각 도로 구간에 대한 윈도우 크기를 계산한다. 여기서, 윈도우 크기 계산 모듈(121)이 각 도로 구간에 대하여 계산한 윈도우 크기를 '제1 윈도우 크기'라 지칭하며, 제1 윈도우 크기를 계산하는 알고리즘에 대해 도 5를 참조로 먼저 설명한다.

[0048] 도 5는 본 발명의 실시예에 따른 윈도우 크기 계산 알고리즘을 나타낸 예시도이다.

[0049] 도 5에 도시된 바와 같이, 윈도우 크기 계산 모듈(121)은 각 도로 구간에 대해 윈도우 크기를 구하기 위해, 각 도로 구간(R_i)에 대해 현재 시간(t)에서의 전이 확률(Base_p)를 구한다. 윈도우 크기 계산 모듈(121)은 구한 전이 확률 값에 대한 엔트로피를 토대로 해당 도로 구간이 교차점인지 여부를 확인한다.

[0050] 즉, 전이 확률 값에 대한 엔트로피가 미리 설정한 임계값(Θ)보다 큰 경우, 해당 도로 구간은 교차점으로부터 나온 도로 구간인 것으로 확인한다. 교차점으로 확인한 경우, 윈도우 크기 계산 모듈(121)은 해당 구간의 윈도우 크기를 2로 설정한다.

[0051] 그러나, 전이 확률 값에 대한 엔트로피가 임계값 보다 작은 경우, BFS 방식으로 해당 도로 구간과 연결되어 있는 과거의 이동 경로들의 확률 값을 이용하여 엔트로피를 재계산한다. 이 과정을 엔트로피가 임계값을 넘을 때까지 반복한다. 윈도우 크기 계산 모듈(121)은 많은 노이즈 삽입을 방지하기 위해, 윈도우 크기를 미리 설정한 크기(MaxW)로 제한하여 윈도우 크기를 계산한다.

[0052] 한편, 상기 도 2의 윈도우 크기 예측 모듈(122)은 윈도우 크기 계산 모듈(121)이 도로 구간의 윈도우 크기를 계산하면, 계산한 윈도우 크기를 토대로 다음 시점의 윈도우 크기를 예측한다.

[0053] 즉, 윈도우 크기는 시간대에 따른 전이 행렬을 사용하기 때문에 가변적이다. 반면 w-event 프라이버시는 각 시간 단위에 프라이버시 예산이 할당되기 때문에, 윈도우 크기의 증가는 프라이버시 예산 할당 시 남아있는 예산이 존재하지 않으므로 0의 예산이 할당되는 문제를 야기할 수 있다. 예를 들어, t 시점에 3의 크기를 갖는 윈도우가 t+1 시점에서 크기가 4로 증가한다고 가정한다. 그러면 t+1 시점에 할당되는 프라이버시 예산은 윈도우 내에 남은 프라이버시 예산이 없기 때문에 0이 할당된다. 이는 무한대의 노이즈 삽입을 의미한다.

[0054] 따라서, w-event 프라이버시 기법이 과거 w-1 개의 시간 단위 동안의 데이터만을 고려하기 때문에, 윈도우 크기 예측 모듈(122)은 tm_{t+1}은 Markov 가정에 따라 이전의 Maxw-1개의 전이 행렬들을 곱하여 구한다. 그리고 상기 도 5에서 설명한 알고리즘과 마찬가지 방식으로 윈도우 크기를 구한다.

[0055] 교통량 배포 모듈(123)은 윈도우 크기 계산 모듈(121)이 계산한 현재 시점의 제1 윈도우 크기와, 윈도우 크기 예측 모듈(122)이 예측한 다음 시점의 제2 윈도우 크기를 토대로, 현재 시점까지 각 도로 구간에 대한 윈도우에서 사용하고 남은 프라이버시 예산($\epsilon_{rm,i}$)을 구한다.

[0056] 그리고, 제1 윈도우 크기가 제2 윈도우 크기보다 작으면, 현재 시점에 할당되는 프라이버시 예산에 윈도우의 크기 증가분을 나눈 프라이버시 예산을 할당한다. 그러나, 제1 윈도우 크기가 제2 윈도우 크기보다 크거나 같으면, 현재 시점에 할당되는 프라이버시 예산을 그대로 할당한다.

[0057] 교통량 배포 모듈(123)이 프라이버시 예산을 할당하는 교통량 배포 알고리즘에 대해 도 6을 참조로 먼저 설명한다.

[0058] 도 6은 본 발명의 실시예에 따른 교통량 배포 알고리즘을 나타낸 예시도이다.

[0059] 도 6에 도시된 바와 같이, 제1 시점(t)에 각 도로 구간(R_i)에 대하여 제1 윈도우 크기와 제2 윈도우 크기가 구해지면, 제1 시점(t)까지 도로 구간(R_i)에 대한 윈도우에서 사용하고 남은 프라이버시 예산($\epsilon_{rm,i}$)을 구한다. 그리고 제1 윈도우 크기가 제2 윈도우 크기보다 작으면, 제1 시점에 할당되는 프라이버시 예산($\epsilon_{t,i}$)에 윈도우의 크기 증가분을 나눈 값을 남은 프라이버시 예산($\epsilon_{rm,i}$)에 반영하여 할당한다.

[0060] 반면, 윈도우의 크기가 증가하지 않는다면 제1 시점에 할당되는 프라이버시 예산($\epsilon_{t,i}$)에 남은 프라이버시 예산($\epsilon_{rm,i}$)을 그대로 할당한다. 교통량 배포 모듈(123)은 각 도로 구간(R_i)에 해당하는 원본 교통량 데이터에 분포 Lap($1/\epsilon_{t,i}$)를 따르는 노이즈를 삽입하고, 노이즈가 삽입된 교통량 데이터를 반환한다.

[0061] MaxW=4인 경우에 대한 제안 기법의 과정을 도 6을 통해 살펴보면, 먼저 각 도로 구간(R_i)에 대한 윈도우의 크기가 3으로 제1 시점(t1)에서 제3 시점(t3) 동안 유지되며, 제3 시점에서 제2 윈도우 크기가 4로 증가한다고 가정한다. 제3 시점(t3)에 윈도우에 남은 예산은 $\epsilon_{rm,i} = \epsilon - (\epsilon/3 + \epsilon/3)$ 이 된다. 그러면 교통량 배포 모듈(123)은 알고리즘에 따라, $\epsilon_{t_3,i}$ 에 남은 예산의 절반만 할당해 주어, 제4 시점(t_4)에서도 예산을 활용할 수 있도록 한다.

[0062] 그리고, 제6 시점(t6)에서 다음 시점의 윈도우의 크기가 1만큼 작아진다고 가정한다. 이런 경우엔 $\epsilon_{t_6,i}$에는 $\epsilon_{rm,i} = \epsilon/3$ 만큼의 예산이 그대로 할당된다. 제7 시점(t_7)에 해당하는 윈도우에 포함되는 시점들은 제5 시점(t_5), 제6 시점(t_6), 그리고 제7 시점(t_7)뿐이므로, 여전히 MaxW-event 프라이버시를 만족함을 알 수 있다.

[0063] 한편, 상기 도 2의 메모리(130)는 교통량 배포 시스템(100)을 구동하기 위한 다양한 프로그램이 저장되어 있다. 또한, 교통량 배포 모듈(123)에서 반환한 교통량 데이터를 저장, 관리한다. 저장된 교통량 데이터는 인터페이스(110)를 통해 단말(300)로 전송될 수 있다. 또한, 교통량 데이터가 요구되지 않을 경우, 미리 설정된 주기에 따라 메모리(130)에서 삭제될 수도 있다.

[0064] 이상에서 설명한 교통량 배포 시스템(100)을 이용하여 교통량 데이터를 배포하는 방법에 대해 도 3을 참조로 설명한다.

[0065] 도 3은 본 발명의 실시예에 따른 교통량 배포 방법에 대한 흐름도이다.

[0066] 도 3에 도시된 바와 같이, 교통량 배포 시스템(100)은 교통 정보 제공 시스템(200)으로부터 교통 정보를 수신한다(S100). 여기서, 교통 정보에는 교통 정보가 수집된 시간 정보, 해당 시간에 도로 구간에 존재하는 차량 정보, 해당 차량이 이동하는 이동 경로 정보, 도로 구간 정보, 교통량 데이터 등이 포함되어 있다.

[0067] 교통량 배포 시스템(100)은 S100 단계에서 수신한 교통 정보가 수집된 제1 시점에서 도로 구간에 대한 전이 확률 값을 계산한다(S101). 만약 제1 도로 구간에서 제2 도로 구간으로 이동한 차량이 없으면 전이 확률은 0이 된다. 그러나 제1 도로 구간에서 제2 도로 구간으로 이동한 차량이 있다면, 이동한 차량 수를 토대로 전이 확률 값을 계산한다. 여기서, 교통량 배포 시스템(100)은 상기에서 도 5를 참조로 먼저 설명한 알고리즘을 토대로, 전이 확률 값을 다음 수학식 2를 이용하여 계산한다.

수학식 2

[0068] 전이 확률 값$(Base_p) = tm_t(R_i)$

[0069] 여기서 tm_t는 t 시간에서의 전이 행렬을 의미한다. 전이 행렬의 각 행과 열은 각각의 도로 구간을 의미하고, 행렬의 원소 값 $tm_t[r][c]$의 의미는 행 r에 해당하는 도로 구간에 있는 차량들이 이전 시점(t-1)에서 열 c에 해당하는 도로 구간으로 왔을 확률을 의미한다.

[0070] 교통량 배포 시스템(100)이 전이 확률 값을 계산하면, 계산한 전이 확률 값에 대한 엔트로피가 미리 설정되어 있는 임계값 보다 큰지 확인한다(S102). 여기서, 전이 확률 값에 대한 엔트로피는 다음 수학식 3을 이용하여 계산된다.

수학식 3

[0071] $$\text{전이 확률 값에 대한 엔트로피}(E) = -Base_p * \log(Base_p)$$

[0072] 수학식 2를 통해 계산된 엔트로피가 미리 설정된 임계값보다 작으면, 해당 도로 구간의 과거 이동 경로들의 확률 값을 이용하여 엔트로피를 재계산한다(S103). 엔트로피 재계산 절차는 계산된 엔트로피가 임계값보다 클때까지 반복한다.

[0073] 그러나 계산된 엔트로피가 미리 설정된 임계값보다 크면, 교통량 배포 시스템(100)은 제1 시점의 도로 구간에 대한 윈도우 크기를 결정한다(S104). 여기서, 윈도우 크기는 2로 설정되는 것을 예로 하여 설명하며, S104 단계에서 결정된 윈도우 크기를 제1 윈도우 크기라 지칭한다.

[0074] 제1 윈도우 크기가 2로 설정되면, 교통량 배포 시스템(100)은 제2 시점 즉, 제1 윈도우 크기를 결정하기 위해 교통 정보가 수집된 제1 시점에서 시간상으로 다음에 오는 제2 시점의 윈도우 크기를 예측한다(S105). 제2 시점의 윈도우 크기는 시간대에 따른 전이 행렬을 사용하기 때문에, 가변적이다.

[0075] 따라서, 교통량 배포 시스템(100)은 제2 시점의 윈도우 크기는 제1 시점 직전의 제3 시점에서의 '최대 윈도우 크기 - 1개'의 전이 행렬들을 곱하여 구한다. 즉, $tm_{t+1} = tm_{t-MaxW+2} * tm_{t-MaxW+3} * \cdots * tm_t$ 의 식으로 제2 시점의 윈도우 크기를 구한다. 그리고 상기 도 5의 알고리즘을 적용하여 최종적으로 제2 시점의 윈도우 크기(이하, 제2 윈도우 크기라 지칭)를 예측한다.

[0076] 제1 윈도우 크기와 제2 윈도우 크기가 계산되거나 예측되면, 교통량 배포 시스템(100)은 제1 시점까지 남은 프라이버시 예산을 계산한다(S106). S106 단계에서 남은 프라이버시 예산($\epsilon_{rm,i}$)을 계산하기 위해, 교통량 배포 시스템(100)은 다음 수학식 4를 이용하여 계산한다.

수학식 4

[0077] $$\epsilon_{rm,i} = \epsilon - \sum_{k=t-w_i+1}^{t-1} \epsilon_{k,i}$$

[0078] 여기서 ϵ 는 윈도우에 할당된 프라이버시 예산을 의미하고, k는 R_i에 대한 윈도우에서 윈도우에 포함되어 있는 가장 오래된 시점을 의미한다. 즉, 시그마 항은 윈도우 안에 포함된 시간 단위들 각각에 할당된 프라이버시 예산의 합산을 의미한다.

[0079] 수학식 3을 통해 남은 프라이버시 예산을 계산한 후, 교통량 배포 시스템(100)은 제1 윈도우 크기가 제2 윈도우 크기보다 작은지 확인한다(S107). 만약 제2 윈도우 크기가 작다면, S106 단계에서 계산한 남은 프라이버시 예산을 제1 시점에 그대로 할당한다(S108).

[0080] 그러나, 제1 윈도우 크기가 제2 윈도우 크기보다 작다면, 제1 윈도우 크기에서 제2 윈도우 크기의 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여, 제1 시점에 할당한다(S109). 그리고, 해당 도로 구간에 해당하는 원본 교통량 데이터에 노이즈를 삽입한 후(S110), 노이즈가 삽입된 교통량 데이터를 반환한다(S111). 여기서, S110 단계에서 삽입되는 노이즈는 분포 Lap($1/\epsilon_{t,i}$)만큼 삽입된다.

[0081] 이상의 절차에 따라 배포된 교통량 데이터와 종래 교통량 데이터 배포 방법에 따른 교통량 데이터를 비교한 예에 대해 도 7을 참조로 설명한다.

[0082] 도 7은 본 발명의 실시예에 따른 오차 값을 나타낸 예시도이다.

[0083] 도 7에 도시된 바와 같이, 각 도로 구간에 삽입되는 노이즈의 양을 평균 절대 오차(1)와 평균 상대 오차(2)를 기준으로 평가하고, 평가는 다음 수학식 5를 이용한다. 각 실험의 결과는 10번씩 반복한 결과의 평균이며, ϵ는 1로 고정하였다. 그리고 각 윈도우 크기에 대한 예측 실패 확률은 3% 내외로 하되, 노이즈 삽입은 고려하지 않는다.

수학식 5

[0084]
$$\text{평균절대오차}(MAE)=\sum_{r\in R}(|D(r)-D'(r)|/|R|)$$
$$\text{평균상대오차}(MRE)=\sum_{r\in R}(|D(r)-D'(r)|/|R|*D(r))$$

[0085] 여기서 D(r)은 도로 구간 r에 대한 원본 교통량이고, D'(r)은 노이즈가 삽입된 교통량, R은 전체 도로 구간의 집합을 의미한다.

[0086] 임계값은 도로 구간의 윈도우 크기를 결정하는데, 임계값이 커질수록 평균적인 윈도우 크기는 증가한다. 이는, 각 시간 단위에 할당되는 프라이버시 예산의 감소와 더 많은 노이즈 삽입을 의미한다.

[0087] 도 7의 (a)에 나타낸 바와 같이 MaxW = 5일 때, 임계값의 변화에 따른 각 도로 구간의 교통량의 평균 절대, 상대 오차를 확인하면, 본 발명의 실시예에 따른 방법을 이용하면 임계값이 증가할수록 기존의 방법과 비슷한 평균 절대, 상대 오차를 갖는 것을 알 수 있다.

[0088] 그리고 도 7의 (b)에 나타낸 바와 같이 MaxW는 각 도로 구간의 윈도우 크기의 최대값을 나타내는 것이다. MaxW가 커질수록 각 시간 단위에 할당되는 프라이버시 예산은 감소하고 더 많은 노이즈가 삽입된다.

[0089] 그리고 임계값이 0일 때 MaxW의 변화에 따른 각 도로 구간의 교통량의 평균 절대, 상대 오차를 살펴보면, MaxW가 증가할수록 기존의 방법으로는 모든 도로 구간에 동일한 MaxW가 적용되어 더 많은 노이즈가 삽입된다. 반면, 본 발명의 실시예에 따른 방법을 이용하면, MaxW보다 작은 크기의 윈도우를 갖는 도로 구간이 많아 상대적으로 일정한 오차를 갖는 것을 알 수 있다.

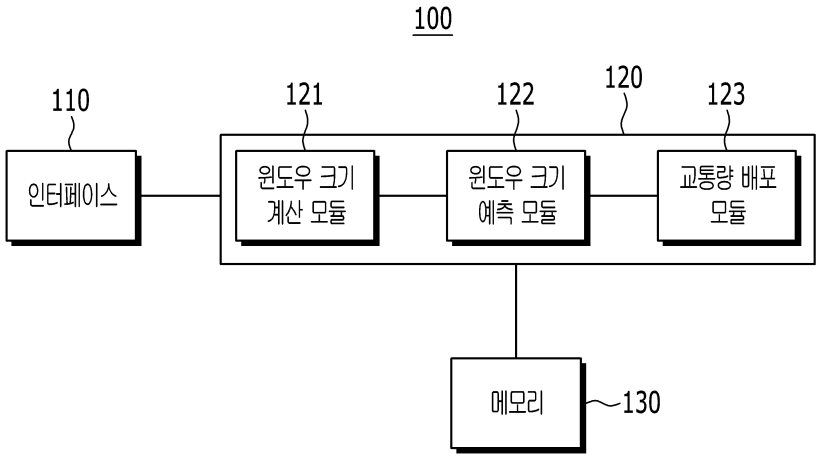
[0090] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

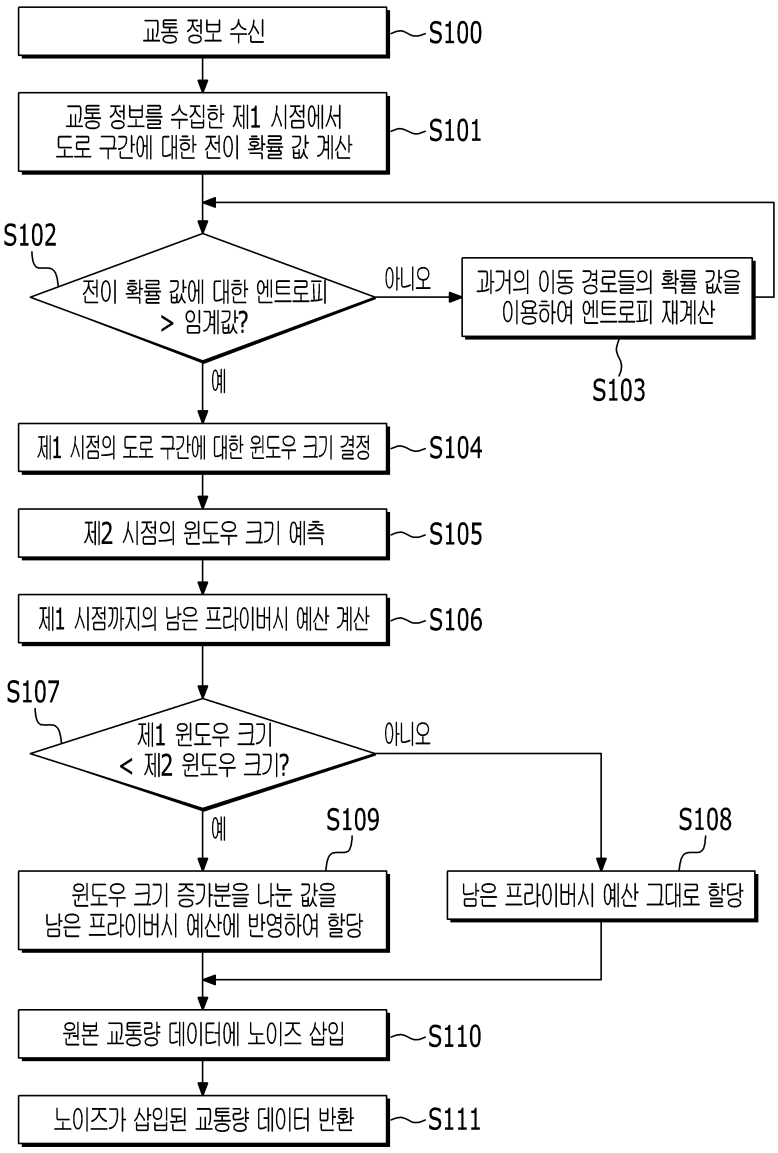
도면1



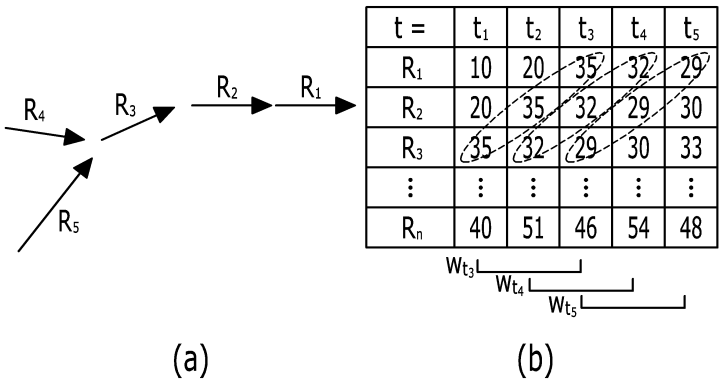
도면2


100
110
121
122
120
123
인터페이스
윈도우 크기 계산 모듈
윈도우 크기 예측 모듈
교통량 배포 모듈
메모리
130

도면3


교통 정보 수신
S100
교통 정보를 수집한 제1 시점에서 도로 구간에 대한 전이 확률 값 계산
S101
S102
전이 확률 값에 대한 엔트로피 > 임계값?
아니오
과거의 이동 경로들의 확률 값을 이용하여 엔트로피 재계산
S103
예
제1 시점의 도로 구간에 대한 윈도우 크기 결정
S104
제2 시점의 윈도우 크기 예측
S105
제1 시점까지의 남은 프라이버시 예산 계산
S106
S107
제1 윈도우 크기 < 제2 윈도우 크기?
아니오
예
S109
S108
윈도우 크기 증가분을 나눈 값을 남은 프라이버시 예산에 반영하여 할당
남은 프라이버시 예산 그대로 할당
원본 교통량 데이터에 노이즈 삽입
S110
노이즈가 삽입된 교통량 데이터 반환
S111

도면4



도면5

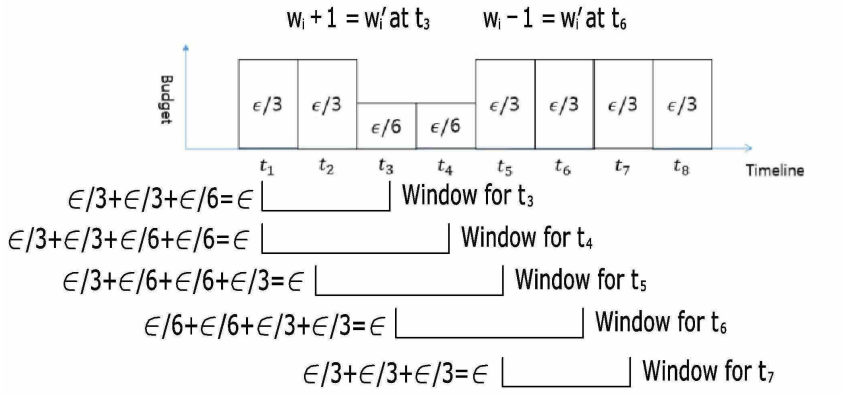
Algorithm 1. GetWindowSize	
Input:	Transition matrices $tm_1, tm_2, \cdots, tm_t$ Timestamp t, Threshold of entropy θ Maximum window size $MaxW$
Output:	Window size for each road: $\{w_1, w_2, \cdots, w_n\}$
1	For each R_i do
2	Base_p = $tm_t(R_i)$; E = −Base_p * log(Base_p)
3	If E > θ then Set $w_i = 2$
4	Else
5	Enqueue R_i to $Prev_road$, Base_p to $Prev_prob$
6	while $k = t - 1$ to $t - MaxW + 1$ and E > θ do
7	Set E = 0
8	For each R_j incident with $Prev_road.pop()$ do
9	$P = tm_k(R_j) * Prev_prob.pop()$
10	$E \mathrel{+}= -P * log(P)$
11	If E > θ the Set $w_i = t - k + 2$ and break
12	Else Enqueue R_j to $Prev_road$, P to $Prev_prob$

도면6

Algorithm 2. Publishing

Input: Transition matrices $tm_1, tm_2, \cdots, tm_t$
Timestamp t
Desired privacy leakage ϵ, Raw data at t D_t
Output: Noisy data at t D_t'

1 Get window size w_i for each R_i using Algorithm 1.
2 Predict window size w_i' using Algorithm 1 with tm_{t+1}.
3 For each R_i do
4 $\epsilon_{rm,i} = \epsilon - \sum_{k=t-w_i+1}^{t-1} \epsilon_{k,i}$
5 If $w_i < w_i'$ then $\epsilon_{t,i} = \epsilon_{rm,i}/(w_i' - w_i + 1)$
6 Else $\epsilon_{t,i} = \epsilon_{rm,i}$
7 Set $\lambda = 1/\epsilon_{t,i}$ and $D_t'[i] = D_t'[i] + Lap(\lambda)$
8 Return D_t'



도면7

