(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년01월09일
(11) 등록번호 10-1936033
(24) 등록일자 2019년01월02일

(51) 국제특허분류(Int. Cl.)
H04L 9/08 (2006.01) *H04L 29/06* (2006.01)
(21) 출원번호 10-2012-0010669
(22) 출원일자 2012년02월02일
심사청구일자 2017년02월01일
(65) 공개번호 10-2012-0090812
(43) 공개일자 2012년08월17일
(30) 우선권주장
13/021,538 2011년02월04일 미국(US)
(56) 선행기술조사문헌
Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption(2010.06.08.)*
Public Key Based Cryptoschemes for Data Concealment in Wireless Sensor Networks(2006년)
A New Approach to Secure Aggregation of Private Data in Wireless Sensor Networks(2009년)
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
팔로 알토 리서치 센터 인코포레이티드
미국 캘리포니아주 94304 팔로 알토 코요테 힐 로드 3333
(72) 발명자
시 런팅
미국 캘리포니아 95130 새너 제이 베인 플레이스 4340
초우 리처드
미국 캘리포니아 94087 서니베일 그로스빅 아벤누브 1674
찬 츠 홍 휴버트
홍콩 포크풀람 로드 더 유니버시티 오브 홍콩 초우 예이 칭 빌딩 디파트먼트 오브 컴퓨터 사이언스
(74) 대리인
장훈

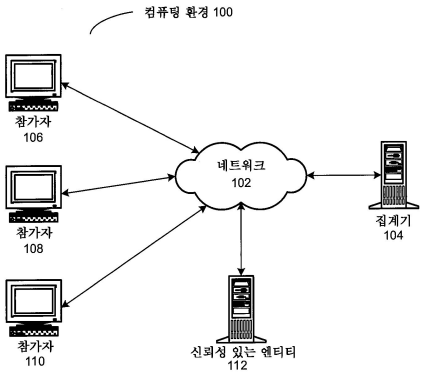
전체 청구항 수 : 총 7 항 심사관 : 남기영

(54) 발명의 명칭 **시계열 데이터의 프라이버시-보존 집계**

(57) 요 약

개인 스트림 집계(PSA) 시스템은 사용자의 프라이버시를 손상시키지 않고 사용자 데이터를 데이터 집계기에 제공한다. 시스템은 사용자들의 세트에서 로컬 사용자에 대한 비밀 키를 결정함으로써 시작할 수 있고, 사용자들의 세트 및 데이터 집계기와 연관된 비밀 키들의 합은 0과 같다. 시스템은 또한 로컬 사용자와 연관된 데이터 값들의 세트를 선택한다. 이후, 시스템은 암호화된 데이터 값들의 세트를 생성하기 위해 비밀 키에 부분적으로 기초하여 상기 세트에서 개별 데이터 값들을 암호화하여, 그에 의해 사용자들의 세트와 연관된 개별 데이터 값들을 복호화하지 않고, 집계값을 복호화하는 동안 사용자들의 세트와 상호작용하지 않고 데이터 집계기가 사용자들의 세트에 걸쳐 집계 값을 복호화하게 한다. 시스템은 또한 암호화된 값들의 세트를 데이터 집계기에 보낸다.

대 표 도



명 세 서

청구범위

청구항 1

데이터 집계기(data aggregator)와 중요 데이터(sensitive data)를 공유하기 위한 방법에 있어서,

사용자들의 세트내의 로컬 사용자를 위해 비밀 키를 결정하는 단계로서, 상기 사용자들의 세트 및 상기 데이터 집계기와 연관된 상기 비밀 키들의 합은 0과 같은, 상기 비밀 키를 결정하는 단계;

클라이언트 컴퓨터에서, 상기 로컬 사용자와 연관된 데이터 값들의 세트를 선택하는 단계;

암호화된 데이터 값들의 세트를 생성하기 위해 상기 비밀 키에 부분적으로 기초하여 상기 세트내의 개개의 데이터 값들을 암호화하고, 그에 의해 상기 데이터 집계기가 상기 사용자들의 세트와 연관된 개개의 데이터 값들을 복호화하지 않고, 집계 값(aggregate value)을 복호화하는 동안 상기 사용자들의 세트와 상호 작용하지 않고, 상기 사용자들의 세트에 걸쳐 상기 집계 값을 복호화하게 하는 단계; 및

상기 암호화된 데이터 값들의 세트를 상기 데이터 집계기에 전송하는 단계를 포함하고,

상기 데이터 값들의 세트는 시계열(time series)을 포함하고, 상기 비밀 키는 상기 시계열들내의 상기 개개의 데이터 값들을 암호화하기 전에 결정되고, 상기 집계 값은 상기 사용자들의 세트와 연관된 상기 개개의 데이터 값들의 합을 포함하고,

사용자 i 및 시간 기간 t에 대한 개개의 데이터 값 $x_{i,t}$의 암호화는 다음 식:

$$c_{i,t} = g^{x_{i,t}} \cdot H(t)^{sk_i}$$

을 계산하는 것을 포함하고,

$c_{i,t}$는 사용자 i 및 시간 기간 t에 연관된 암호화된 값이고, g는 생성자이고, sk_i는 사용자 i와 연관된 비밀 키이며, H(t)는 해쉬 함수인, 중요 데이터 공유 방법.

청구항 2

제 1 항에 있어서,

상기 로컬 사용자를 위해 비밀 키를 결정하는 단계는 신뢰성 있는 소스로부터 상기 비밀 키를 수신하는 단계를 포함하는, 중요 데이터 공유 방법.

청구항 3

제 1 항에 있어서,

상기 로컬 사용자를 위해 비밀 키를 결정하는 단계는 안전한 다자간 프로토콜(secure multi-party protocol)을 이용하는 단계를 포함하는, 중요 데이터 공유 방법.

청구항 4

삭제

청구항 5

제 1 항에 있어서,

상기 개개의 데이터 값들을 암호화하는 단계는 랜덤 노이즈(random noise)를 갖는 수정된 데이터 값들의 세트를 생성하기 위해 적어도 상기 데이터 값들의 서브세트에 랜덤 값들을 부가하는 단계를 포함하는, 중요 데이터 공유 방법.

청구항 6

제 1 항에 있어서,

상기 개개의 데이터 값들을 암호화하는 단계는 또한 상기 개개의 데이터 값들의 고위 모멘트들(higher moments)을 암호화하여, 상기 데이터 집계기가 상기 사용자들의 세트에 걸쳐 상기 데이터 값들에 대한 분포를 결정하게 하는 단계를 포함하는, 중요 데이터 공유 방법.

청구항 7

컴퓨터에 의해 실행될 때, 상기 컴퓨터가 데이터 집계기와 중요 데이터를 공유하기 위한 방법을 수행하도록 하는 명령들을 저장하는 비-일시적인 컴퓨터 판독가능한 저장 매체에 있어서,

사용자들의 세트내의 로컬 사용자를 위해 비밀 키를 결정하는 단계로서, 상기 사용자들의 세트 및 상기 데이터 집계기와 연관된 상기 비밀 키들의 합은 0과 같은, 상기 비밀 키를 결정하는 단계;

상기 로컬 사용자와 연관된 데이터 값들의 세트를 선택하는 단계;

암호화된 데이터 값들의 세트를 생성하기 위해 상기 비밀 키에 부분적으로 기초하여 상기 세트내의 개개의 데이터 값들을 암호화하고, 그에 의해 상기 데이터 집계기가 상기 사용자들의 세트와 연관된 개개의 데이터 값들을 복호화하지 않고, 집계 값을 복호화하는 동안 상기 사용자들의 세트와 상호 작용하지 않고, 상기 사용자들의 세트에 걸쳐 상기 집계 값을 복호화하게 하는 단계; 및

상기 암호화된 데이터 값들의 세트를 상기 데이터 집계기에 전송하는 단계를 포함하고,

상기 데이터 값들의 세트는 시계열(time series)을 포함하고, 상기 비밀 키는 상기 시계열들내의 상기 개개의 데이터 값들을 암호화하기 전에 결정되고, 상기 집계 값은 상기 사용자들의 세트와 연관된 상기 개개의 데이터 값들의 곱(product)을 포함하고,

사용자 i 및 시간 기간 t에 대한 개개의 데이터 값 $x_{i,t}$의 암호화는 다음 식:

$$c_{i,t} = x_{i,t} \cdot H(t)^{sk_i}$$

을 계산하는 것을 포함하고,

$c_{i,t}$는 사용자 i 및 시간 기간 t에 연관된 암호화된 값이고, sk_i는 사용자 i와 연관된 비밀 키이며, H(t)는 해쉬 함수인, 비-일시적인 컴퓨터 판독가능한 저장 매체.

청구항 8

데이터 집계기와 중요 데이터를 공유하기 위한 장치에 있어서,

사용자들의 세트내의 로컬 사용자를 위해 비밀 키를 결정하도록 구성된 셋업 메커니즘으로서, 상기 사용자들의 세트 및 상기 데이터 집계기와 연관된 상기 비밀 키들의 합은 0과 같은, 상기 셋업 메커니즘;

암호화 메커니즘으로서,

상기 로컬 사용자와 연관된 데이터 값들의 세트를 선택하고,

암호화된 데이터 값들의 세트를 생성하기 위해 상기 비밀 키에 부분적으로 기초하여 상기 세트내의 개개의 데이터 값들을 암호화하고, 그에 의해 상기 데이터 집계기가 상기 사용자들의 세트와 연관된 개개의 데이터 값들을 복호화하지 않고, 집계 값을 복호화하는 동안 상기 사용자들의 세트와 상호 작용하지 않고, 상기 사용자들의 세트에 걸쳐 상기 집계 값을 복호화하게 하도록 구성된, 상기 암호화 메커니즘; 및

상기 암호화된 데이터 값들의 세트를 상기 데이터 집계기에 전송하도록 구성된 통신 메커니즘을 포함하고,

상기 데이터 값들의 세트는 시계열(time series)을 포함하고, 상기 비밀 키는 상기 시계열들내의 상기 개개의 데이터 값들을 암호화하기 전에 결정되고, 상기 집계 값은 상기 사용자들의 세트와 연관된 상기 개개의 데이터 값들의 합을 포함하고,

사용자 i 및 시간 기간 t에 대한 개개의 데이터 값 $x_{i,t}$를 암호화할 때, 상기 암호화 메커니즘은 또한 다음 식:

$$c_{i,t} = g^{x_{i,t}} \cdot H(t)^{sk_i}$$

을 계산하도록 구성되고,

$c_{i,t}$는 사용자 i 및 시간 기간 t에 연관된 암호화된 값이고, g는 생성자이고, sk_i는 사용자 i와 연관된 비밀 키이며, H(t)는 해쉬 함수인, 중요 데이터 공유 장치.

발명의 설명

발명의 내용

해결하려는 과제

과제의 해결 수단

[0001] 시스템은 사용자의 프라이버시를 손상시키지 않고 데이터 집계기(data aggregatpr)에 사용자의 데이터를 제공한다. 시스템은 사용자들의 세트내의 로컬 사용자를 위해 비밀 키를 결정하고, 사용자들의 세트 및 데이터 집계기와 연관된 비밀 키들의 합은 0과 같다. 또한, 시스템은 로컬 사용자와 연관된 데이터 값들의 세트를 선택한다. 그 후, 시스템은 암호화된 데이터 값들의 세트를 생성하기 위해 비밀 키에 부분적으로 기초하여 세트내의 개개의 데이터 값들을 암호화하고, 그에 의해 데이터 집계기가 사용자들의 세트와 연관된 개개의 데이터 값들을 복호화하지 않고, 집계 값을 복호화하는 동안 사용자들의 세트와 상호 작용하지 않고, 상기 사용자들의 세트에 걸쳐 집계 값을 복호화하게 한다. 또한, 시스템은 암호화된 데이터 값들의 세트를 데이터 집계기에 전송한다.

[0002] 일부 실시예들에서, 시스템은 신뢰성 있는 소스로부터 비밀 키를 수신함으로써 로컬 사용자를 위해 비밀 키를 결정한다.

[0003] 일부 실시예들에서, 시스템은 안전한 다자간 프로토콜(secure multi-party protocol)을 이용함으로써 로컬 사용자를 위해 비밀 키를 결정한다.

[0004] 일부 실시예들에서, 데이터 값들의 세트는 시계열의 데이터 값들을 포함한다. 또한, 시스템은 시계열내의 개개의 데이터 값들을 암호화하기 전에 비밀 키를 결정할 수 있다.

[0005] 일부 실시예들에서, 개개의 데이터 값들을 암호화하는 것은 랜덤 노이즈(random noise)를 갖는 수정된 데이터 값들의 세트를 생성하기 위해 적어도 데이터 값들의 서브세트에 랜덤 값들을 부가하는 것을 포함한다.

[0006] 여기서, △는 집계 값의 민감도이고, n은 사용자들의 수이다.

[0007] 일부 실시예들에서, 상기 개개의 데이터 값들을 암호화하는 것은 상기 개개의 데이터 값들의 고위 모멘트들(higher moments)을 또한 암호화하는 것을 포함하고, 그에 의해 데이터 집계기가 사용자들의 세트에 걸쳐 데이터 값들을 위한 분포를 결정하게 한다.

[0008] 다른 실시예에서, 시스템은 사용자들의 세트와 연관된 데이터 값들의 세트에 대한 집계 값을 계산한다. 시스템은 데이터 집계기를 위한 비밀 키를 결정하고, 사용자들의 세트 및 데이터 집계기와 연관된 비밀 키들의 합은 0과 같다. 또한, 시스템은 대응하는 사용자들의 세트로부터 암호화된 데이터 값들의 세트를 수신한다. 그 후, 시스템은 사용자들의 세트와 연관된 개개의 데이터 값들을 복호화하지 않고 사용자들의 세트에 걸쳐 집계 값들을 결정하기 위해 비밀 키를 사용하고, 집계 값을 데이터 집계기에 제공한다.

[0009] 일부 실시예들에서, 집계 값은 사용자들의 세트와 연관된 개개의 데이터 값들의 곱(product)을 포함한다.

도면의 간단한 설명

[0010] 도 1은 실시예에 따른 개인 스트림 집계를 위한 예시적인 컴퓨팅 환경을 도시하는 도면.

도 2는 실시예에 따른 신뢰성 없는 데이터 집계기와 데이터를 안전하게 공유하기 위한 프로세스를 도시하는 흐름도.

도 3은 실시예에 따른 참가자들의 세트에 의해 제공되는 암호화된 데이터로부터 집계 값을 결정하기 위해 데이터 집계기에 의해 수행되는 프로세스를 도시하는 흐름도.

도 4는 실시예에 따른 참가자들의 세트와 데이터 집계기 사이에 정보의 예시적인 흐름을 도시하는 도면.

도 5는 실시예에 따른 프라이버시 보존 데이터 집계를 용이하게 하는 컴퓨터 시스템을 도시하는 도면.

도 6은 실시예에 따른 프라이버시 보존 데이터 집계를 용이하게 하는 장치를 도시하는 도면.

발명을 실시하기 위한 구체적인 내용

[0011] 본 발명의 실시예들은 사용자들이 암호화된 데이터의 스트림을 신뢰성 없는 집계기에 업로드하게 하고, 집계기가 각각의 시간 간격 동안 집계 통계들을 복호화하기 위해 비밀 키를 이용하게 하는 개인 스트림 집계(PSA; Private Stream Aggregation) 시스템을 제공한다. PSA 기술은 집계기가 원하는 통계들 및 그의 보조 지식으로부터 추론할 수 있는 것 이외에 어떤 의도치 않은 정보에 대해서는 습득할 수 없다는 것을 의미하는, 불확실한(oblivious) 집계기이다. PSA 기술은 집계기에 공개된 통계가 특정 개인이 참가하는지 여부에 의해 큰 영향을 받지 않도록, 개개의 참가자들에 대하여 분포된 차등의 프라이버시(distributed differential privacy)를 보장한다.

[0012] 신뢰성 없는 데이터 집계기는 개개의 데이터 포인트들에 액세스하지 않고 시계열 데이터에 대한 집계 통계를 계산할 수 있다(예를 들어, 참가자들의 데이터의 합 또는 곱). 예를 들어, 데이터 집계기가 매주 n 개의 회사들의 총 매출액의 추적을 유지하고자 할 수 있다. 본 발명의 실시예들에 의해 제공된 프라이버시 모델은 각 개개의 회사가 매주 그들의 소득에 대한 노이지 암호화(noisy encryption)를 데이터 집계기에 업로드하게 한다. 또한, 데이터 집계기는 참가하는 회사들의 소득들에 대한 노이지 합(noisy sum)을 복호화할 수 있지만, 개개의 회사와 연관된 매주의 소득 데이터를 얻을 수는 없다. 따라서, 데이터 집계기는 개개의 회사에 대하여 특정한 부가의 정보를 추정할 수 없다.

[0013] 도 1은 실시예에 따라 개인 스트림 집계에 대한 예시적인 컴퓨팅 환경을 도시한다. 구체적으로, 컴퓨팅 환경(100)은 컴퓨터 네트워크(102), 데이터 집계기(104), 및 참가자들(106, 108, 110)을 포함할 수 있다. 일부 실시예들에서, 컴퓨팅 환경(100)은 또한 컴퓨팅 환경(100)을 위해 신뢰성 있는 셋업을 수행하는 신뢰성 있는 엔티티(112)를 포함할 수 있다. 집계기(104), 참가자들(106 내지 110), 및 신뢰성 있는 엔티티(112)는 서버 컴퓨터, 데스크탑, 랩탑, 휴대용 컴퓨팅 장치, 또는 무선 센서와 같은 컴퓨터 네트워크(102)에 연결된 컴퓨팅 장치들을 포함할 수 있다. 또한, 컴퓨터 네트워크(102)는 유선 네트워크, 무선 네트워크, 또는 그의 조합을 포함할 수 있다.

[0014] 컴퓨팅 시스템(100)은 적어도 3개의 키 동작들, 즉 셋업, 노이지 암호화, 집계 복호화를 거친다. 초기의 셋업 동작 동안, 신뢰성 있는 엔티티(112)는 참가자들(106 내지 110) 각각에 비밀 키를 할당하고, 디코딩 기능을 갖는 집계기(104)에 제공하는 비밀 키를 집계기(104)에 할당할 수 있다, 구체적으로, 집계기(104) 및 참가자들(106 내지 110)에 연관된 비밀 키들에 대한 값들의 합은 0과 같다. 이 특징은 참가자들(106 내지 110)이 그들의 데이터를 암호화하게 하여, 집계기의 비밀 키만이 집계 값을 복호화할 수 있도록 한다(예를 들어, 개개의 참가자들로부터 수신된 데이터 값들의 합을 복호화한다).

[0015] 집계기(104) 및 참가자들(106 내지 110)은 안전한 다자간 프로토콜을 이용하여 초기의 셋업 동작을 수행할 수 있고, 그에 의해 신뢰성 있는 엔티티(112)의 도움을 필요로 하지 않는다. 예를 들어, 집계기(104) 및 참가자들(106 내지 110)은, 그들의 비밀 키들을 공개하지 않고, 비밀 키들의 합이 0인 비밀 키들의 세트를 결정하기 위해 서로 협상할 수 있다.

[0016] 노이지 암호화 동작 동안, 참가자(예를 들어, 참가자(106))는 그의 비밀 키를 이용하여 데이터 값을 암호화할 수 있다. 일부 실시예들에서, 참가자는 데이터 값들을 암호하기 전에 데이터 값에 노이즈를 부가할 수 있다. 예를 들어, 참가자가 시간 기간에 걸쳐 데이터 시계열의 데이터 값들의 세트를 암호화하기 때문에, 암호화된 데이터 값들의 서브세트는 랜덤 노이즈를 포함할 수 있지만, 암호화된 데이터 값들의 남은 서브세트는 랜덤 노이즈를 포함하지 않는다.

[0017] 또한, 집계 복호화 동작 동안, 집계기(104)는 그의 비밀 키, 및 집계기(104)가 참가자들(106 내지 110)로부터 수신한 암호화된 값들의 세트에 부분적으로 기초하여 집계 값을 결정한다. 집계기(104)가 임의의 참가자들(106 내지 110)에 대한 비밀 키들을 알고 있지 않기 때문에, 집계기(104)는 임의의 참가자들(106 내지 110)로부터 얻은 개개의 데이터 값을 복호화할 수 없다는 것에 유의한다. 그러나, 본 발명의 실시예들에 의해 제공된 개인 스

트림 집계 기술은, 집계기(104)가 그의 비밀 키를 이용하여 암호화된 데이터 값들의 세트를 위한 집계 값을 결정하는 것을 허용한다.

[0018] 하기의 단락들은 컴퓨팅 환경(100)과 연관된 개인 스트림 집계 시스템에 대한 다양한 예시적인 애플리케이션들을 제공한다. 각각의 예시적인 애플리케이션에서, 개인 스트림 집계 기술은 참가자들(예를 들어, 개개의 사용자들 또는 조직들)이 그들의 개개의 데이터 값들을 데이터 집계기에 공개하지 않고 그들의 개개의 데이터를 데이터 집계기에 제공하는 것을 허용한다.

[0019] 센서 네트워크들은 일반적으로 빌딩들의 안전을 감시하거나, 트래픽 흐름들을 측정하거나, 또는 환경 오염물질들을 추적하는 등의 다양한 애플리케이션들을 위해 전개된다. 전형적인 애플리케이션에서, 전개된 센서 노드들은 주기적으로 그들의 지식들을 중앙국에 전송하고, 상기 중앙국은 패턴을 식별하거나 통계를 결정하기 위해 데이터를 조사한다. 여러 상황들에서, 개개의 센서들에 의해 형성된 지식들은, 특별히 센서들이 다수의 참가중인 조직들에 걸쳐 전개되는 경우에 프라이버시가 중요해질 수 있다. 본 발명의 실시예들에 의해 제공된 개인 스트림 집계 기술은, 중앙국(예를 들어, 집계기(104))이 임의의 특정 센서 노드(예를 들어, 참가자들(106 내지 110))로부터 상세한 정보를 얻을 수 없게 될 프라이버시 보장을 참가하는 회사들에 제공한다. 따라서, 이 프라이버시 보장은 그 조직에 걸쳐 센서 노드들을 전개시킴으로써 중요한 리서치 프로젝트들에 기여하도록 참가하는 회사들을 독려할 수 있다.

[0020] 다른 예시는 전자 “스마트 그리드(smart grid)” 및 “스마트 미터링(smart metering)" 기술들의 출현에 있다. 스마트 미터들은 매달 한번 마주하는 매 15분간 유틸리티 사용을 판독하는 것과 같이, 전통적인 미터들보다 훨씬 더 미세한 입도로 유틸리티 사용을 판독한다. 유틸리티들의 이러한 세립 샘플링(fine-grained sampling)은 중요 정보를 추론하기 위해 충분히 상세한 정보를 유틸리티 회사에 제공할 수 있다. 예를 들어, 유틸리티 회사는 가정내의 개인들의 수, 그들의 특정 가전제품들의 사용뿐만 아니라, 그들의 수면/업무 습관들을 추론할 수도 있다. 본 발명의 실시예들에 의해 제공된 개인 스트림 집계 기술은 어떤 특정 가정으로부터의 실제 세립의 유틸리티 사용을 공개하지 않고 한 세트의 가정들(예를 들어, 참가자들(106 내지 110))에 걸쳐 유틸리티 사용의 세립 샘플링을 얻기 위한 기능을 유틸리티 회사(예를 들어, 도 1의 집계기 (140))에 제공할 수 있다. 이로 인해 가정들이 스마트 그리드 기술들을 향해 나아가고, 집계 통계들이 여전히 스마트 그리드 오퍼레이터들을 위해 그의 감시 성과들 및 가격 최적화들을 달성하기에 충분한 것인지에 대한 걱정 및 의구심이 완화될 수 있다.

[0021] 의학 연구는 의학 데이터로부터 크게 혜택을 받지만, 프라이버시 걱정이 이 데이터가 수집되어 유포되는 정도를 제한한다. 본 발명의 실시예들에 의해 제공된 개인 스트림 집계 기술은 연구자들이 개인들 또는 조직들의 그룹에 걸쳐 높은 수준의 통계들을 얻고, 케어기버들(caregivers) 또는 원격 측정 장치들에 의해 지속적으로 업로드되는 데이터로부터 높은 수준의 통계들을 얻게 할 수 있다.

[0022] 많은 연구 프로젝트들 또는 소프트웨어 기술들은 개인들에 걸쳐 프라이버시 걱정들을 몰아낼 수 있는 인구 투표, 감지 및 감시의 정도를 수행한다. 예를 들어, 특정 법인 소프트웨어는 참가자의 유용성을 추정하기 위해 카메라, 와이파이, 컴퓨터 활동에서 얻은 참가자의 데이터를 사용하고, 동료들이 주어진 시점에 그 참가자와 통신하기 위한 최고의 수단을 식별하는데 도움이 될 수 있다. 하지만, 참가자의 업무 습관들에 대한 상세한 정보는 이 유용성 정보로부터 추론될 수 있다. 따라서, 일부 사용자들은 그들의 유용성에 관한 이러한 정보가 회사 관리자들에 의해 오용될 수 있는 두려움에 참가하기를 꺼려할 수 있다. 본 발명의 실시예들에 의해 제공된 개인 스트림 집계 기술은, 참가자가 선택된 개인들과 상세한 유용성 정보를 공유하는 것을 허용하는 동안에만, 기업이 참가자들의 그룹에 걸쳐 통계 정보를 얻게 할 수 있다.

[0023] 클라우드 컴퓨팅(cloud computing)이 인기를 얻음에 따라, 개인들 및 조직들은 제3자 클라우드 서비스에 대한 데이터의 증가량을 저장하고 있다. 클라우드 서비스 제공자들은 다양한 사회적 및 경제적 목표들을 실현하기 위해, 이 데이터를 통해 유용한 통계들을 계산하기를 원한다. 하지만, 참가하는 회사들은 클라우드 서비스들을 많이 사용하지 않는 주요한 이유로서 그 데이터의 보안 및 프라이버시에 대한 걱정을 나타낸다. 본 발명의 실시예들에 의해 제공된 개인 스트림 집계 기술은, 클라우드 서비스 제공자가 개개의 참가자들로부터 중요 정보를 얻지 않고, 시간이 지남에 따라 다수의 참가자들로부터 특정한 집계 통계들을 추적할 수 있게 한다.

[0024] 도 2는 신뢰성 없는 데이터 집계기와 데이터를 안전하게 공유하기 위한 프로세스를 도시하는 플로우 차트를 나타낸다.

[0025] 시스템은 참가자들의 세트 및 데이터 집계기와 연된된 비밀 키들의 합이 0과 같도록, 참가자들의 세트내의 로컬 참가자를 위한 비밀 키를 결정(동작 202)함으로써 시작할 수 있다. 예를 들어, 동작(202)은 신뢰성 있는 소스로

부터 비밀 키를 수신하거나, 또는 데이터 집계기 및 참가자들의 세트를 위한 비밀 키를 결정하기 위해 안전한 다자간 프로토콜(secure multi-party protocol)을 이용하는 것을 포함할 수 있다.

[0026] 다음에, 시스템은 데이터 집계기와 공유하기 위해 데이터 값들의 세트를 선택할 수 있고(동작 204), 랜덤 노이즈를 가진 수정된 데이터 값들의 세트를 생성하기 위해 적어도 데이터 값들의 서브세트에 랜덤 값을 부가할 수 있다(동작 206, 선택적). 데이터 값들이 시계열을 포함할 수 있다는 것에 유의한다. 그 후, 시스템은 세트내의 개개의 데이터 값들을 암호화하고(동작 208), 암호화된 데이터 값들을 데이터 집계기에 전송한다(동작 210).

[0027] 집계 값은 특정 시간기간 동안 참가자들의 세트와 연관된 개개의 값들의 합산을 포함할 수 있다. 이 경우, 참가자 i 및 시간 기간 t에 대한 개개의 데이터 값 $x_{i,t}$의 암호화는 이하의 식을 계산하는 것을 포함한다.

[0028] $$c_{i,t} = g^{x_{i,t}} \cdot H(t)^{sk_i}$$

[0029] 특별히, $c_{i,t}$는 참가자 i와 연관된 암호화된 값이고, 시간 기간 t,g는 생성자이고, sk_i는 참가자 i와 연관된 비밀 키이며, H(t)는 해쉬 함수이다.

[0030] 집계 값은 참가자들의 세트와 연관된 개개의 값들의 곱을 포함한다. 이 경우, 참가자 i에 대한 개개의 데이터 값 $x_{i,t}$의 암호화는 하기의 식을 계산하는 것을 포함한다.

[0031] $$c_{i,t} = x_{i,t} \cdot H(t)^{sk_i}$$

[0032] 도 3은 참가자들의 세트에 의해 제공된 암호화된 데이터로부터 집계 값을 결정하기 위해 데이터 집계기에 의해 수행된 프로세스를 도시하는 플로우 차트를 나타낸다.

[0033] 시스템은 참가자들의 세트 및 데이터 집계기와 연된된 비밀 키들의 합이 0과 같도록, 데이터 집계기를 위한 비밀 키를 결정(동작 302)함으로써 시작할 수 있다. 예를 들어, 동작(302)은 신뢰성 있는 소스로부터 비밀 키를 수신하는 것을 포함하거나, 또는 데이터 집계기 및 참가자들의 세트를 위한 비밀 키들을 결정하기 위해 안전한 다자간 프로토콜을 이용하는 것을 포함할 수 있다.

[0034] 다음에, 시스템은 대응하는 참가자들의 세트로부터 암호화된 데이터값들의 세트를 수신할 수 있다(동작 304). 시스템은 암호화된 데이터 값들의 세트를 위한 집계 값을 결정하기 위해 데이터 집계기와 연관된 비밀 키를 이용한다(동작 306). 그 후, 시스템은 집계 값을 데이터 집계기에 제공한다(동작 308).

[0035] 집계 값의 결정은 이하의 식을 계산하는 것을 포함한다:

[0036] $$V = H(t)^{sk_0} \prod_{i=1}^{n} c_{i,t}$$

[0037] 특별히, $c_{i,t}$는 참가자 i와 연관된 암호화된 값이고, 시간 기간 t,n은 전체 참가자들의 수이고, sk_0는 데이터 집계기와 연관된 비밀 키이고, H(t)는 해쉬 함수이다.

[0038] 집계 값이 V의 이산 대수(discrete log)를 계산하는 것을 더 포함하는 것을 결정하도록, 집계 값은 참가자들의 세트와 연관된 개개의 데이터 값들의 합산을 포함한다. 일부 실시예들에서, 집계 값은 참가자들의 세트와 연관된 개개의 데이터 값들의 곱을 포함한다.

[0039] 일반화된 애플리케이션에서는, 하나의 데이터 집계기와 n 개의 참가자들이 있을 수 있다. 표기상의 편리함을 위해, 참가자들은 1,...,n으로 번호 매겨지고, 데이터 집계기는 0으로 번호 매겨진다. 예를 들어, 비밀 키 $sk_{i\neq 0}$는 참가자와 연관되고, 비밀 키 sk_0는 데이터 집계기와 연관된다. 또한, [n]:={1,2,...,n}으로 하고, D는 참가자의 데이터를 위해 허용가능한 값들의 특정 도메인을 나타낸다고 하자. 따라서, 시간 기간 t∈N인 경우, 참가자 $i \in [n]$과 연관된 데이터는 값 $x_{i,t} \in D$를 가진다.

[0040] 간단함을 위해, 참가자의 데이터 값에 대한 표기법은 항상 서브스크립트 t를 포함할 수 없고, i∈[n]에 대한 데이터 값들의 세트 x_i는 공통의 시간 기간에 대응한다고 추측될 수 있다. 따라서, $x=(x_1,\ldots, x_n)\in D^n$는 일부 시

간 기간에서 모든 참가자들로부터의 값들의 벡터를 나타낸다고 하자. 또한, 집계기는 범위 O에 속하는 원하는 통계값을 계산하기 위해 함수 f(x)를 이용한다. 따라서, 집계기는 함수 $f:D^n \rightarrow O$에 의해 나타내어지는 집계 통계들을 계산한다.

[0041] 각각의 참가자는 집계기가 사용자의 입력들에 대한 임의의 보조 정보를 가질 때조차도 강하게 프라이버시 보증들을 달성하도록 독립적인 랜덤 노이즈를 생성할 수 있다. 구체적으로, 각각의 참가자는 일부 샘플 공간 Ω로부터의 노이즈 r_i가 $\mathbf{r} := (r_1,...,r_n) \in \Omega^n$에 의해 나타내지도록, 다른 참가자들과는 관계없이 랜덤 노이즈 r_i를 생성할 수 있다. 표기의 목적들을 위해, 해트 변수(hatted variable)는 (예를 들어, 일부 랜덤 값 r 및 랜덤화 함수 X와 연관된) 참가자 데이터의 랜덤화된 버전들을 표시하고, 비-해트 변수들(non-hatted variables)은 원래의 참가자 데이터를 표시한다. 따라서, X:D×Ω→D는 값을 암호화 및 집계기에 업로드하기 전에 데이터의 노이지 버전 $x_i := X(x_i,r_i)$를 계산하기 위해 참가자들에 의해 이용될 수 있는 랜덤화 함수를 표시한다. 그 후, 집계기는 원하는 통계 f(x)에 도입된 노이즈가 미리 결정된 레벨 내에 있도록, $\hat{\mathbf{x}} := (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$의 암호화된 값들로부터 노이지 통계 $f(\hat{x})$을 계산한다.

[0042] 참가자 i는 그 참가자에 맞춰진 원하는 데이터 분포에 따라 노이즈 r_i를 생성할수 있다. 또한, 각각의 참가자 i는 다른 랜덤화 함수 $X_i(x_i,r_i)$를 데이터 x_i에 적용할 수 있다. 보다 더 일반적인 시나리오에서, 각각의 참가자는 x_i 및 r_i를 집계기에 전송하기 전에 별도로 데이터 x_i 및 랜덤니스 r_i를 암호화할 수 있고, 상기 집계기는 암호화된 입력들에 대한 랜덤화된 집계 함수 $f:D^n \times \Omega^n \rightarrow O$를 계산할 수 있다. 하지만, 간단함을 위해, 하기의 단락들은 $\hat{f}(x,r) = f(\hat{x})$인 특별한 경우로 고려한다. 또한, 하기의 단락들은 참가자들이 데이터를 암호화하기 전에 같은 랜덤화 함수 X를 적용하는 시나리오를 커버한다.

[0043] 본 발명의 실시예들에 의해 제공된 개인 스트림 집계(PSA) 메커니즘은 몇몇의 키 함수들 즉, Setup(), NoisyEnc(), 및 AggrDec() 함수로 이루어진다.

[0044] $Setup(1^{\lambda})$: setup() 함수는 보안 파라미터 λ에 이용되고, 공개 파라미터들 param, 각각의 참가자를 위한 비밀 키 sk_i 뿐만 아니라, 각각의 시간 기간에서 집계 통계들을 복호화하기 위해 집계기에 의해 사용되는 집계기 sk_o를 위한 비밀 키를 출력한다. 각각의 참가자 i는 비밀 키 sk_i를 얻고, 데이터 집계기는 기능 sk_0을 얻는다. 신뢰성 있는 셋업 단계 후에, 암호화된 데이터 값들을 데이터 집계기에 업로드하는 것을 제외하고, 참가자들과 데이터 집계기 사이에는 더 이상의 상호작용이 필요로 되지 않는다.

[0045] NoisyEnc(param,sk_i,t,x,r,): 시간 셋업 t 동안, 각각의 참가자는 노이즈 r를 갖는 데이터 x을 인코딩하기 위해 NoisyEnc 함수를 호출한다. 그 결과는 노이즈 r로 랜덤화된 데이터 포인트 x의 노이지 암호화이다. 때때로, NoisyEnc 함수는 NoisyEnc(param, sk_i, t, $\hat{x}$)로서 기록되며, 여기서 $\hat{x} := X(x,r)$은 참가자의 데이터의 노이지 버전이고, X는 근본적인 랜덤화 함수이다.

[0046] AggrDec(param,sk_o,t,c_1,c_2,...,c_n): 복호화 알고리즘은 같은 시간 기간 t 동안 공개 파라미터들 param, 기능 sk_o, 및 암호문들 c_1, c_2, ..., c_n에 소용된다. 각각의 i∈[n]인 경우, c_i=NoisyEnc(sk_i,t, $\hat{x}_i$)로 하고, 여기서 각각의 $\hat{x}_i := \chi(x_i,r_i)$이다. $x:=(x_1,...,x_n)$ 및 $\hat{\mathbf{x}} := (\hat{x}_1,...,\hat{x}_n)$ 라고 하자. 복호화 알고리즘은 타겟된 통계들 f(x)의 노이지 버전인 $f(\hat{x})$을 출력한다.

[0047] 일부 실시예들에서, 집계 함수 $f(\hat{x})$는 참가자들의 세트를 위해 데이터 값들의 노이지 합산(noisy summation)을 생성한다. 이러한 상황에서, 참가자 데이터 x_i는 일부 소수 p에 대한 Z_p에 속하고, 집계 함수는

$sum(\hat{\mathbf{x}}) := \sum_{i=1}^{n} \hat{x}_i$ 로서 정의된다. 또한, 각각의 참가자는 정수들의 세트로부터 노이즈 r_i를 생성하여 랜덤화 함수 $X(x_i,r_i):=x_i+r_i \bmod p$를 적용할 수 있다(즉, 참가자는 데이터를 암호화하기 전에 부가의 노이즈를 통합한다).

[0048] 마찬가지로, 일부 다른 실시예들에서, 집계 함수 $f(\hat{x})$는 참가자들의 세트에 대한 데이터 값들의 노이지 곱을 생성한다. 이러한 상황에서, 집계 함수는 $product(\hat{\mathbf{x}}) := \prod_{i=1}^{n} \hat{x}_i$ 로서 정의된다.

[0049] 도 4는 본 발명의 실시예에 따른 참가자들의 세트와 데이터 집계기 사이의 정보의 예시적인 플로우를 도시하는 다이어그램을 나타낸다. 구체적으로, 참가자들(402 내지 406)은, 참가자가 비밀 키 sk_i를 이용하여 대응하는 암호화된 데이터 값 c_i를 생성하기 전에 노이즈 r_i를 데이터 값 x_i에 부가할 수 있도록, 대응하는 암호화된 데이터 값들의 세트를 생성할 수 있다. 또한, 집계기(408)는 참가자들(402 내지 406)로부터 암호화된 데이터 값들을 수신하고, 그의 비밀 키 sk_0을 이용하여 집계 값(410)을 생성한다.

[0050] 집계기 불확실 요건(aggregator obliviousness rquirement)은 집계기가 공개된 통계들 $f(\hat{x})$ 및 집계기가 이미 알고 있는 임의의 보조 데이터로부터 유추할 수 있는 것 이외에 아무것도 알지 못할 것이라는 보장을 제공한다. 또한, 이러한 요건의 달성은 적절한 집계기 기능(예를 들어, 집계기의 비밀 키 sk_0)이 없는 자가 아무것도 습득하지 못할 것이라는 것을 보장한다.

[0051] 직관적으로, 집계기 불확실 요건은 하기의 보안 개념들을 충족한다.

[0052] · 집계기는 각각의 시간 기간 동안 노이지 합만을 습득할 수 있고, 그 이상은 아무것도 습득할 수 없다. 예를 들어, 집계기는 모든 참가자들의 암호문들의 적당한 서브세트로부터 임의의 부분적인 정보를 습득할 수 없다.

[0053] · 집계기의 비밀 키를 알지 못하고, 여러 참가자들이 남아있는 사용자들에 대해 연합을 형성한다 하더라도, 상대방(adversary)은 암호화된 데이터에 대해 아무것도 습득하지 않는다.

[0054] · 집계기가 참가자들의 서브세트와 공모하거나, 또는 암호화된 데이터의 서브세트가 누설되는 경우, 집계기는 남아있는 참가자들의 합을 필연적으로 습득할 수 있다. 하지만, 집계기는 남아있는 참가자들에 대한 개개의 데이터 값들을 습득할 수는 없다.

[0055] 집계기 불확실 보안의 개념은 합산 이외의 일반적인 통계들까지도 확장될 수 있다. 하지만, 상대방이 성공적인 공격으로부터 얻을 수 있는 정보를 제한하기 위해 추가의 주의가 취해져야만 한다. 예를 들어, 상대방이 참가자들의 세트 $K \subseteq [n]$을 손상시켰다면, 그후 상대방은 이들 참가자들을 대신하여 무엇이든 암호화할 수 있다. 따라서, 상대방은 손상된 참가자들의 세트 K를 위해 임의의 원하는 평문 벡터(plaintext vector) $\hat{\mathbf{x}}_K = \{\hat{x}_i \mid i \in K\}$를 플러그인하여, 평문 벡터를 암호화하고, 그 후 $\hat{\mathbf{x}}_K$에 기초하여 집합 통계들을 복호화하기 위해 AggrDec 함수를 호출할 수 있다. 따라서, 보안의 정의는 이것이 최선이고 상대방을 위한 전략일 뿐이라는 사실을 반영해야 한다(즉, 상대방은 이러한 공격으로부터 취한 정보 이외의 추가의 정보를 습득할 수 없다.) 이러한 요건은 하기의 단락에 의해 정의된다.

[0056] 정의 1(집계기 불확실 보안) : 평문 벡터(예를 들어, x 또는 x')가 챌린저(challenger)에 의해 암호화되어 있다는 구별에서 무시할 만한 이점 이상을 갖는 어떤 확률적인 다차-시간 상대방이 없는 경우, PSA는 불확실한 집계기이다. 이러한 요건은 2개의 평문 벡터들이 그의 집계 값들에 대해 동등할 때 일반적으로 충족된다. 합산 통계의 경우, 이 요건은 아래와 같이 정의된다:

[0057] $$\sum_{i \in U} x_i = \sum_{i \in U} \hat{x}_i'$$

[0058] 다른 통계들과 연관된 일반적인 질의들에 대하여 이 조건을 충족시키는 것이 보다 어려울 수 있다.

[0059] 정의 2(암호화-일회 보안) : 각각의 정직한 참가자는 시간 기간마다 일회만 암호화하도록 하는 것이 예상된다. PSA는:

[0060] 1) 상기 보안 게임에서 무시할 만한 이점 이상을 가진 어떤 확률적인 다차-시간 상대방이 없는 경우,

[0061] 2) 다음의 제약, 즉 $\forall i \in U,\ \forall (x,r) \in D \times \Omega,$ 을 유지하는 경우,

[0062] "암호화-일회" 모델내에서 불확실한 집계기일 거라고 하며,

[0063] 여기서, 투플(tuple)(i,t^{*},x,r)은 임의의 암호화 질의에서 나타나지 않는다.

[0064] 하기의 단락은 집계기 불확실 보안을 달성하기 위한 암호의 구성을 기술한다. G는 결정적 디피-헬만(Decisional Diffie-Hellman)이 어려운 프라임 차수 p의 순환 그룹을 나타낸다고 하자. 또한, H:Z→G는 수학적 그룹 G에 정수를 맵핑하는 해쉬 함수(랜덤 오라클로서 모델화됨)를 나타낸다고 하자.

[0065] Setup(1^{λ}): 신뢰성 있는 셋업 단계 동안, 신뢰성 있는 딜러는 $\sum_{i=0}^{n} sk_i = 0$ 이도록, 랜덤 생성자 $g \in G$, 및 n+1 랜덤 비밀들 $sk_i \in Z_p$를 선택한다. 공개 파라미터들은 param:=g로서 초기화된다는 것에 유의한다. 또한, 데이터 생성자는 기능 sk_0을 얻고, 참가자 i는 비밀 키 sk_i를 얻는다.

[0066] 따라서, 시간 기간 t 동안, 각각의 참가자는 $i \in [n]$에 대한 $R_{i,t}=H(t)^{ski}$를 계산하고, 집계기는 $R_{0,t}=H(t)^{sk0}$을 계산한다. sk_i에 대한 합이 0과 같기 때문에, $\prod_{i=0}^{n} R_{i,t} = 1$ 이 뒤따른다는 것을 유의한다. 이 속성은 참가자들이 신뢰성 있는 셋업 단계 후에 서로 통신하는 것을 필요로 하지 않고, NoisyEnc() 및 AggrDec() 오퍼레이션들을 독립적으로 기능하게 하는 것을 허용한다. 또한, 결정적 디피-헬만이 순환 그룹 G 에 대해 어렵다는 것을 고려할 때, 이것은 번호들 $R_{i,t}$이 랜덤 오라클 모델(random oracle model)하에서 외관상 랜덤이라는 것이 뒤따른다.

[0067] NoisyEnc(param,sk_i,t,$\hat{x}$): 참가자 i가 시간 셋업 t 동안 값 $\hat{x} \in Z_p$ 을 암호화하는 경우, 참가자는 하기의 암호문을 계산한다:

[0068] $$c \leftarrow g^{\hat{x}} \cdot H(t)^{sk_i}$$

[0069] 각각의 참가자가 데이터 값을 암호화하기 전에 데이터 값들에 노이즈를 부가할 것이 가정되는 것을 고려할 때, 랜덤화된 평문 값은 항 $\hat{x} = x + r \bmod p$ 로 표현된다.

[0070] AggrDec(param,sk_0,t,$c_1,c_2,\ldots,c_n$): 집계기는 다음과 같이 집계 값 V를 계산한다:

[0071] $$V = H(t)^{sk_0} \prod_{i=1}^{n} c_i$$

[0072] 이와 같이 할 때, $i \in [n]$에 대한 c_i=NoisyEnc(param,sk_0,t,$\hat{x}_i$)라는 것에 유의한다. 또한, 집계기를 위한 sk_i 및 참가자들의 세트가 0까지 부가하기 때문에, $\prod_{i=0}^{n} H(t)^{sk_i} = 1$ 이다. 따라서, V는 다음의 형태를 따릅니다.

[0073] $$V = g^{\sum_{i=1}^{n} \hat{x}_i}$$

[0074] 합 $\sum_{i=1}^{n}\hat{x}_i$ 을 복호화하기 위해, 집계기는 V 베이스 g의 이산 대수를 계산한다. 평문 공간이 작은 경우, 복호화는 부르트-포스 검색(brute-force search)을 통해 달성될 수 있다. 평문 공간에서 복호화 시간의 대략 제곱근을 필요로 하는 폴라드의 람다 방법(Pollard's lambda method)을 이용하는 보다 나은 방식이 있다. 예를 들어, 각각의 참가자들의 입력은 범위 {0, 1, ..., △}이라고 가정한다. 그후, 참가자들의 합은 범위 범위 {0, 1, ..., n△}내에 들어 있다. 이 경우, 복호화는 폴라드의 방법을 이용하여 $\sqrt{n\Delta}$ 사간을 필요할 것이다. 바꿔말하면, n△은 다차 시간(polynomial time)에서 성공적인 복호화를 보장하기 위해 비밀 파라미터 λ에서 다차임이 요구된다.작은 평문 공간의 한정은 부가적으로 준동형 암호화 방식들(homomorphic encryption schemes)(예를 들어, 엘 가말 암호화(El Gamal encryption) 및 BGN 준동형 암호화)로서 이용될 때 디피-헬만에 기초하는 암호화 방식들의 전형이라는 것에 유의한다.

[0075] 정리 1: 결정적 디피-헬만 문제가 그룹 G에서 어렵고, 해쉬 함수 H가 랜던 오라클이라고 가정하면, 상기 단락들에서 제공된 구성은 "일회 암호화" 모델에서의 집계기 불확실 보안을 충족한다.

[0076] 제안된 암호의 구성에서, 암호화는 적어도 디피-헬만 그룹내의 해쉬 동작(예를 들어, SHA-256), 2개의 모듈러 지수들(modular exponentiations) 및 하나의 곱셈을 포함한다. 해쉬 함수 및 그룹 곱셈을 계산하는 시간이 지수를 위한 시간에 비하여 훨씬 작기 때문에, 실행 시간은 2개의 모듈러 지수들에 의해 지배된다. eBACS 프로젝트에 의해 보고되는 번호들을 벤치마크하는 것에 따라, 최신의 64 비트 데스크탑 계산들에 대해, 전형적인 디피-헬만 그룹 모듈러의 1024 비트 프라임를 이용하여 모듈러 지수를 계산하는데 대략 3 ms가 걸린다. "curve25519"와 같은 고속의 타원 곡선들을 이용하면, 모듈러 지수를 계산하는데는 0.3 ms 만이 걸린다. 따라서, 암호화는 최신의 컴퓨터 상에서 대략 0.6 ms내에 행해질 수 있다. 집계 통계들의 암호화는 이산 대수를 취하는 것을 필요로 하고, 무작위 대입 방법(brute-force method)을 이용하는 경우, 각각의 가능한 평문을 시도하는데 0.3 ms를 필요로 하는 하나의 모듈러 지수를 취한다. 따라서, 본 발명의 실시예에 의해 제공된 PSA 기술은 평문 공간이 작은 상황들에서 유용하다. 예를 들어, 각각의 참가자의 평문이 통신을 위한 그녀의 이용가능성을 나타내는 하나의 비트를 포함하고, 대략 1000 개의 참가자들이 있을 때, 복호화는 무작위 대입 방식을 이용하여 약 0.3 s내에 행해질 수 있다. 복호화를 위해 폴라드의 람다 방법을 채택하는 것은 추가의 가속을 제공할 수 있어, 약 $\sqrt{n\Delta}$ 까지 실행 시간을 감소시킬 수 있으며, 여기서 n은 참가자들의 수이고, 각각의 참가자의 값은 세트 {0,1,...△}내에 라고 가정한다.

[0077] 본 발명의 실시예들은 집계기가 노이지 통계들만을 습득하지만, 각 개개의 값들을 습득하지 못할 것이라는 것을 보장한다. 따라서, 개개의 참가자들에 대한 프라이버시는 최종 통계 $f(\hat{x})$ 가 충분한 랜덤니스를 축적하는 동안은 보장된다. 이것은 각 개인이 개개의 데이터 값들에 덜 랜덤한 노이즈를 부가하는 것을 허용한다. 또한, 참가자들의 특정 함수가 손상되어 데이터 집계기와 공모하도록 결정되는 경우, 참가자들은 그들의 데이터 또는 랜덤니스를 집계기에 공개할 수 있다. 이 경우, 남아있는 손상되지 않는 참가자들의 랜덤니스가 그들의 프라이버시를 보호하기에 충분한 것이 바람직하다.

[0078] 프라이버시의 이러한 개념은 분포된 차등의 프라이버시(Distributed Differential Privacy)라고 지칭되고, 이는 공개된 통계내의 노이즈가 참가자들로부터 수집된다는 사실을 반영한다. 하기의 단락들은 분포된 차등의 프라이버시의 개념을 정의한다.

[0079] 집계기가 하기의 방식으로 생성된 n개의 참가자들의 랜덤화된 데이터 $\hat{\mathbf{x}}\in D^n$에 대해 함수 $f:D^n\to Q$를 평가한다는 것을 상기한다. 각각의 참가자 i는 일부 분포에 따른 독립적인 랜덤니스 $r_i\in\Omega$를 생성하고, 곱 $x_i:=X(x_i,r_i)$를 생성하기 위해 데이터 x_i에 랜덤화 함수 $\chi:D\times\Omega\to D$ 를 적용한다. $x\in D^n$ 이고, $r\in\Omega^n$이라고 고려할 때, 개념 $\hat{\mathbf{x}}=\hat{\mathbf{x}}(\mathbf{r}):=(\chi(x_1,r_1),\chi(x_2,r_2),\ldots,\chi(x_n,r_n))$ 은 r에 대한 $\hat{\mathbf{x}}$의 의존성이 얼마나 내포되어 있는지를 나타낸다. 또한, 참가자들의 서브세트 K를 고려할 때, $r_K:=\{r_i:i\in K\}$ 및 $\overline{K}$는 K(즉, $\overline{K}=\{1,2,\ldots,n\}\setminus K$)의 보수라고 하자.

[0080] 하기의 분포된 차등의 프라이버시 요건은 각각의 시간 기간 t∈N에 적용한다.

[0081] 정의 3((ε,δ)-DD-프라이버시): ε>0, 0≤δ<1, 및 0<ɣ≤1이라 가정한다. (즉, 결합 분포 r:=(r_1,...,r_n) 및 랜덤화 함수 X)에 의해 주어진)데이터 랜덤화 절차는, 하기의 조건이 유지되는 경우, 함수 f에 대하여 그리고 손상되지 않은 참가자들의 ɣ 함수하에서 (ε,δ)-분포된 차등의 프라이버시(DD-프라이버시)를 달성한다. 임의의 이웃하는 벡터들 $\mathbf{x},\mathbf{y} \in D^n$ 의 경우, 임의의 서브세트 S⊆O의 경우, 및 적어도 크기 ɣn 의 손상되지 않은 참가자들의 임의의 서브세트 $\overline{K}$ 의 경우,

[0082] $$\Pr[f(\hat{\mathbf{x}}) \in S \mid \mathbf{r}_K] \leq \exp(\varepsilon) \cdot \Pr[f(\hat{\mathbf{y}}) \in S \mid \mathbf{r}_K] + \delta.$$

[0083] 정의 3에서, 2개의 벡터들 $\mathbf{x},\mathbf{y} \in D^n$ 은, 그들이 정확히 하나의 좌표에서 차이가 있는 경우에 이웃들 또는 이웃하는 벡터들이라고 한다. 이것은 정확히 한 사용자가 데이터 값을 변경하는 경우의 시나리오에 대응한다.

[0084] K가 손상된 노드들의 세트인 경우, 상기한 정의는 남아있는 정직한 참가자들의 랜덤니스가 차등의 프라이버시를 보장하기에 충분할 것을 요구한다. 따라서, 확률은 손상된 참가자들로부터의 랜덤니스 r_K에 좌우된다. 바꿔말하면, 확률은 정직한 참가자들로부터의 랜덤니스 $\mathbf{r}_{\overline{K}}$ 를 통해 이루어진다. DD-프라이버시의 정의는 손상되지 않는 참가자들의 임의의 세트 $\overline{K}$ 에 대해, $|\overline{K}| \geq \gamma m$ 인 동안, 정의 3하에서 방정식을 유지할 것을 요구한다.

[0085] 하기의 단락들은 (ε,δ)-차등의 프라이버시 보장을 달성하는 암호의 구성을 구축하는 방법을 도시한다. 이 암호의 기술은 참가자들에게 데이터 값을 암호화하기 전에 데이터 값에 노이즈를 부가함으로써 그 자신의 데이터의 차등의 프라이버시를 보장하는 책임을 준다. 분포된 차등의 프라이버시 보장을 달성하기 위해, 하기의 2개의 시도들이 처리될 필요가 있다.

[0086] 참가하는 개인들에 대한 차등의 프라이버시를 보장하기 위해, 공개된 통계는 적절한 진폭의 랜덤 노이즈 r을 포함해야 한다. 실제 설정들에서, 참가자들은 서로 신뢰하지 않을 수 있어, 참가자들의 서브세트가 손상되고 데이터 집계기와 공모하는 것이 가능하다. 최악의 경우, 모든 참가자가 다른 n-1개의 참가자들이 손상되어 집계기와 공모할 수 있다고 생각하는 경우, 참가자는 그 자신의 데이터의 프라이버시를 보장하기 위해 충분한 노이즈를 부가할 필요가 있다. 하지만, 이것은 결과적으로 집계기가 원하는 값을 초과하는 에러를 축적하는 결과를 나을 것이다.

[0087] 참가자들의 적어도 ɣ함수가 정직하고 손상되지 않았다면, 노이즈 생성 업무은 이들 참가자들 사이에 분포될 수 있다. 각각의 참가자가 보다 적은 노이즈를 부가할수 있고, 최종 통계내의 노이즈가 충분히 큰 동안은, 개인 프라이버시가 보호된다. 따라서, 참가자들은 ɣ에 대한 하한(lower bound)에 대해 사전 추정치를 갖는다고 추정된다. 그러나, 참가자들은 참가자들이 어느 정도 손상되었는지를 정확히 알 필요가 없다. 각각의 참가자는 ɣ에 의존하는 분포로부터 노이즈를 생성하는 것이 가정된다. 정직한 참가자들은 이 프로토콜을 따를 것이지만, 손상된 참가자들은 데이터 집계기에 그들의 노이즈를 공개하거나 노이즈를 부가하지 않을 것을 선택할 수 있다. 이러한 구성은 수락가능한 레벨들내에 최종 통계의 에러를 유지하는 동안, 높은 확률로 공개된 통계가 정직한 참가자들로부터 충분한 노이즈를 축적할 것을 보장한다.

[0088] 암호의 구성과 연관된 대수 제약들 내에 다른 시도가 행해지는 것을 포함한다. 일반적인 암호화 방식들은 이산 값들의 그룹으로부터 픽업될 평문 값을 필요로 한다. 따라서, 암호의 구성은 이산 그룹내에서 데이터 및 노이즈 값들을 인코딩할 수 있어야 한다. 또한, 집계기 불확실을 달성하는 암호 구성 기술은 작은 평문 공간에서 작용해야 한다. 그러므로, 보다 일반적으로 사용되고 있는 라플라스 분포(Laplace distribution)를 이용하는 대신, 대칭적 기하 분포(symmetric geometric distribution)을 이용함으로써 그러한 이산 그룹들과 작업하는 것이 가능하다.

[0089] 본 발명의 실시예들은 이산 데이터 값들의 그룹과 작업하도록 대칭적 기하 분포를 이용한다. 대칭적 기하 분포가 한정되어 있지 않아, 그룹의 크기 또는 평문 공간의 크기가 오버플로우될 수 있다는 것에 유의한다. 따라서, 본 발명의 실시예들은 그러한 오버플로우의 확률이 충분히 작아, 집계기가 높은 성공 확률을 갖는 노이지 통계들을 성공적으로 복호화할 수 있게 한다.

[0090] 정의 4(기하 분포): α>1라 하고, Geom(α)는 k에서 확률 질량 함수가 $\frac{\alpha-1}{\alpha+1}\cdot\alpha^{-|k|}$ 이도록, 정수 값들을 취하는 대칭적 기하 분포를 나타낸다고 하자. 또한, $Geom^{+}(\alpha)$가 k에서 확률 질량 함수가 $(\alpha-1)\alpha^{-k}$이도록, 양의 정수값들을 취하는 일측의 기하 분포를 나타낸다고 하자.

[0091] 대칭적 기하 분포 Geom(α)는 라플라스 분포 Lap(b)(여기서, $\alpha\approx\exp(\frac{1}{b})$)의 이산 버전으로 보여질 수 있고, 그의 확률 밀도 함수는 $x\mapsto\frac{1}{2b}\exp(-\frac{|x|}{b})$ 이다. 하기의 Geom 분포의 속성은 정수값들을 출력하는 별도의 개인 매카니즘들을 설계하는데 유용하다.

[0092] 속성 1: ε>0이라 하고, u 및 v는, $|u-v|\le\Delta$이도록, 2개의 정수들이라고 가정한다. 또한, r은 분포 $Geom(\exp(\frac{\varepsilon}{\Delta}))$을 갖는 랜덤 변수라고 하자. 그후, 임의의 정수 k에 대해서는, $\Pr[u+r=k]\le\exp(\varepsilon)\cdot\Pr[v+r=k]$이다.

[0093] 속성 1은 타겟된 통계 f(x)가 민감도 △를 갖는 경우, △에 비례한 진폭을 갖는 기학 노이즈의 부가는 차등의 프라이버시를 달성하기에 충분함을 제시한다. 이전에 언급한 바와 같이, 참가자들은 집계기나 서로에 대해 신뢰하지 않는다. 따라서, 집계기에 진정한 통계를 공개하는 것이 차등의 프라이버시를 분명히 위반하는 것이기 때문에, 집계기는 노이즈 생성의 일로 인해 신뢰되지 않아야 한다. 또한, 그렇지 않으면, 이 지정된 참가자가 또한 진정한 통계를 습득할 수 있을 것이기 때문에, 개개의 참가자들은 이러한 어느 하나의 일로 인해 신뢰되지 않아야 한다.

[0094] 합산을 위한 DD-프라이버시를 달성하기 위해, $x=(x1,..,xn)\in D^{n}$이라 하고, $r=(r_1,...,r_n)=\Omega^{n}$는 특정 시간 기간의 모든 참가자들로부터의 각각의 데이터 및 노이즈 값들을 나타낸다고 하자. 따라서, $D=O=Z_p$(즉, 추가 모듈로(addition modulo) p를 구비한 순환 그룹) 이고, Ω=Z이다. 또한, 집계 함수를 $sum:D^{n}\to O$로 가정하고, $sum(\mathbf{x})=\sum_{i=1}^{n}x_i p$ 이다. 또한, 각각의 참가자는 동일한 랜덤화 함수 $X(x_i,r_i):=x_i+r_i p$를 이용한다.

[0095] 임의의 2개의 요소들 u,v∈Z에 대해서, |u-v|는 u=v+sp 또는 v=u+sp이도록 가장 작은 음이 아닌 정수 s라고 하자. 또한, Z_p의 요소에 정수를 부가하는 경우, 모듈로 p를 이용하여 부가가 수행된다고 가정할 수 있다.

[0096] 또한, 각각의 참가자의 원래 데이터가 도메인 {0,1,..,△}내에 들어있다고 가정한다. 그래서, 합의 중요도는 한 참가자의 변화에 대하여 △이다. 바꿔 말하면, 단일 참가자가 데이터를 변화시키는 경우, 합은 최대 △로 변화한다. 속성 1로부터, $Geom(\exp(\frac{\varepsilon}{\Delta}))$ 노이즈가 출력에 통합되는 경우, ε-차등의 프라이버시가 달성된다는 것을 상기한다. 현재의 경우, 참가자들은 공동으로 최종 출력 통계에 노이즈를 제공한다. 적어도 γn개의 참가자들이 정직하고 손상되지 않는 경우, 유사한 진폭의 노이즈가 축적되도록 보장하는 것이 목표이다. 이러한 방식으로, 차등의 프라이버시가 보장될 뿐만 아니라, 축적된 노이즈가 최종 출력에 한정되도록 보정되어, 에러가 작아지게 된다.

[0097] 따라서, 본 발명의 실시예들에 의해 제공된 암호화 메커니즘은 (ε,δ)-DD-프라이버시를 보장하고, 한편으로는 대략 $O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}})$ 진폭의 작은 에러를 보장한다. 참가자들의 상수 분수 γ가 정직한 동안은, 에러 항은 참가자들의 수 n과는 관계가 없다. 진폭 $\Theta(\frac{\Delta}{\varepsilon})$ 의 축적된 노이즈가 차등의 프라이버시를 보장할 필요가 있다고 고려할

때, 그 결과는 근사하게 최적이 된다. 또한, $\gamma = O(\frac{1}{n})$ (즉, 각각의 참가자는 모든 다른 참가자들이 손상될 수 있거나, 그들의 일정수만이 정직하다고 생각하는)때를 극단의 경우라고 고려할 때, 축적된 노이즈는 $O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}}) = O(\frac{\Delta}{\varepsilon}\sqrt{n})$ 일 것이다. 각각의 참가자가 프라이버시를 보장하기 위해 이 경우에 크기 $\Theta(\frac{\Delta}{\varepsilon})$ 의 대칭 노이즈를 부가해야 하기 때문에, 이것은 또한 직감적으로 감지된다. 따라서, 크기 $\Theta(\frac{\Delta}{\varepsilon})$ 의 n개의 독립적인 대칭 노이즈들의 합이 결과적으로 높은 확률을 갖는 크기 $O(\frac{\Delta}{\varepsilon}\sqrt{n})$ 의 최종 노이즈가 되는 것이 뒤따른다.

[0098] 정리 2(DD-낮은 에러를 갖는 프라이버시 절차) : $\varepsilon > 0$이고, $0 < \delta < 1$이라 하자. 각각의 참가자의 데이터가 Z_p에서 폭△의 간격내의 정수들로부터 나온다고 가정하고, 여기서 $\Delta \geq \frac{\varepsilon}{3}$ 이다. 또한, n개의 참가자들의 적어도 γ 함수가 $\gamma \geq \frac{1}{n}\log\frac{1}{\delta}$ 이도록 손상된다고 가정한다. 그후, 합에 대하여 (ε , δ)-DD-프라이버시인 $r=(r_1,\ldots,r_n)$을 생성하도록 랜덤화된 절차가 존재한다. 또한, $x \in (Z_p)^n$인 경우, 확률을 갖는 $\log\frac{2}{\eta} \leq \frac{1}{\gamma}\log\frac{1}{\delta}$ 이도록 모든 $0 < \eta < 1$인 경우, r의 랜덤 선택을 통해 확률 적어도 1-η은

[0099] $$|sum(\mathbf{x}) - sum(\hat{\mathbf{x}})| \leq \frac{4\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}}$$ 이고, 여기서

[0100] $$\hat{\mathbf{x}} := \hat{\mathbf{x}}(\mathbf{r}) := (x_1 + r_1, x_2 + r_2, \ldots, x_n + r_n)p$$ 이다.

[0101] 표 1은 정리 2에서의 보장을 달성하는 절차를 기술하고 있다. 부가하여, 보조정리(Lemma) 1 및 정리 3하에는 분석이 제공된다. 구체적으로, 정리 3은 노이지 통계 sum($\hat{\mathbf{x}}$)이 진정한 출력 sum(x)으로부터 얼마나 많이 벗어나는지를 분석한다. $\hat{\mathbf{x}}$ 때문에, 이것은 $Z := \sum_{i=1}^{n} r_i$ 의 크기를 제한하기에 충분하다.

[0102] 표1

[0103] ----------DD-개인 데이터 랜덤화 절차

[0104] ----------

[0105] $\alpha := \exp(\frac{\varepsilon}{\Delta})$ 이고, $\beta := \frac{1}{\gamma n}\log\frac{1}{\delta} \leq 1$ 이라 하자.

[0106] x=(x1,...,xn)은 특정 시간 기간내의 모든 참가자들의 데이터를 표시한다.

[0107] foreach 참가자 i ∈[n] do

[0108] 하기의 분포에 따라 노이즈 r_i를 샘플링화:

[0109] $$r_i \leftarrow \begin{cases} \text{Geom}(\alpha) & \text{with probability } \beta \\ 0 & \text{with probability } 1-\beta \end{cases}$$

[0110] $\hat{x}_i \leftarrow x_i + r_i \bmod p$ 를 계산함으로써 랜덤화

[0111] ---

[0112] 보조정리 1 : ε>0이고, 0<δ<1이라 하자. 참가자들의 적어도 γ 함수가 손상된다고 가정한다. 그 후, 상기 랜덤화 절차는 $\beta = \min\{\frac{1}{\gamma n}\log\frac{1}{\delta}, 1\}$ 에 대한, sum에 대하여 (ε, δ)-DD 프라이버시를 달성한다.

[0113] 정리 3(에러 한정): ε>0이고, 0<δ<1이라 하자. 각각의 참가자들의 데이터는 폭△의 간격내의 정수들로부터 나온다고 가정하고, 여기서 $\Delta \geq \frac{\varepsilon}{3}$ 이다. n개의 참가자들의 적어도 $\gamma \geq \frac{1}{n}\log\frac{1}{\delta}$ 함수가 손상된다고 가정한다. 표 1에서의 랜덤화된 절차가 $\alpha := \exp(\frac{\varepsilon}{\Delta})$ 및 $\beta := \frac{1}{\gamma n}\log\frac{1}{\delta} \leq 1$ 을 갖는 $r := (r_1, \ldots, r_n)$을 생성하도록 실행된다. 그 후, 모든 0<η<1인 경우, $\log\frac{2}{\eta} \leq \frac{1}{\gamma}\log\frac{1}{\delta}$ 이고, 적어도 1-η 확률은

[0114] $$|\sum_{i=1}^{n} r_i| \leq 4\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}} \cdot \frac{\sqrt{\alpha}}{\alpha - 1} \leq \frac{4\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}}.$$

[0115] 정리 3에 따라, 축적된 에러는 높은 확률을 갖는 $O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}})$ 에 의해 한정된다. 각각의 참가자의 값이 도메인 $D=\{0, \ldots, \Delta\}$으로부터 피크된다고 가정한다. 그 후, 집계기는 단순히 범위 $[-O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}}), n\Delta + O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}})]p$ 내에서 합을 복호화하려고 하며, 여기서 p는 사용중의 수학 그룹(mathematical group)의 크기이다. 복호화는 높은 확률로 성공할 것이다.

[0116] 정리 3은 측정 집중 결과(measure concentration result)이며, 이는 각각의 r_i의 모멘트 생성 함수를 분석함으로써 입증된다. 손상되지 않은 참가자들의 상수 분수 γ가 존재하는 한, 에러 한정은 n과 무관한 것이 관측된다. Geom(α) 분포의 분산이 높은 확률을 갖는 $\frac{2\alpha}{(\alpha-1)^2}$ 이기 때문에, 에러는 대부분 최종 대답에 Geom(α)의 하나의 복제를 부가하는 것보다 더 나쁜 상수 팩터이며, 이는 ε-차등의 프라이버시를 보장하는데 필요한 노이즈의 최소량이다.

[0117] 분석가들은 종종 인구를 통한 분포들을 연구하고 싶어한다. 일부 실시예들에서, PSA 시스템은, 집계기가 n개의 참가자들의 데이터의 근사적 분포를 주기적으로 평가하는 것을 허용하도록 확장될 수 있다. 예를 들어, 분포가 가우시안(Gaussian)이라고 알려져 있다고 가정하면, 각각의 참가자는 원래의 값뿐만 아니라 그의 제곱(square)을 암호화하기에 충분하다. 이들 암호화된 값들을 이용하여, 집계기는 평균 및 분산(또는 제 2 모멘트)을 통해 분포를 회복시킬 수 있다. 다른 분포들에 대해, 참가자들이 또한 고위 모멘트들을 암호화하는데 필요할 수 있다. 일반적으로, 각각의 참가자가 암호화하는 모멘트들이 많을수록, 집계기가 분포를 추정하는 것이 더 나아진다.

[0118] 일부 실시예들에서, PSA 시스템은 집계 값(예를 들어, 개개의 값들의 합)에의 공개 액세스를 가능하게 하지만, 개개의 값들에의 공개 액세스를 가능하게 하지 않는다. 예를 들어, 집계기 기능은 sk_0=0으로 설정될 수 있어, 집계기의 기능을 공개한다. 또한, n개의 참가자들은 0까지 부가하는 값들 $sk_1, \ldots, sk_n$을 수신한다. n개의 참가자들 및 임의의 공개 집계기는 일상적인 집계 통계들의 암호화 및 복호화를 수행한다. 집계 합을 얻기 위해서는 이산 대수가 계산되어야 하고, 그리하여 다시 평문 공간이 작아져야 한다.

[0119] 일부 실시예들에서, 액세스 제어 계층들을 제공하기 위해 PSA 시스템을 위한 프로토콜이 넣어질 수 있고, 여기서 높은 레벨들의 엔티티들은 그 아래의 모든 잎 노드들(leaf nodes)을 통해 수집된 통계들에의 액세스를 가진다. 셋업 단계에서, 레벨 j>1의 엔티티에는 그 아래 레벨의 엔티티들의 비밀들이 합이 주어진다. 또한, j=1인 경우(즉, 단 하나의 수집된 레벨), 잎 노드들보다 위의 각각의 엔티티에는 그보다 아래의 참가자들의 비밀들의 음의 합이 주어진다.

[0120] 일부 실시예들에서, PSA 메커니즘은 합 대신에 곱의 불확실한 계산을 지원한다. 이를 달성하기 위해, 참가자는 데이터 값 X를 $c \leftarrow \chi \cdot H(t)^{sk_i}$ 로서 암호화한다. 평문이 지수에 더 이상 존재하지 않기 때문에, 곱들을 위한 이 방식은 평문 공간이 작아지는 것을 요구하지 않는다.

[0121] 도 5는 본 발명의 실시예들에 따라 프라이버시-보존 데이터 집계를 용이하게 하는 컴퓨터 시스템을 도시한다. 컴퓨터 시스템(502)는 프로세서(504), 메모리(506), 및 저장 장치(508)를 포함한다. 또한, 컴퓨터 시스템(502)은 디스플레이 장치(510), 키보드(512), 및 포인팅 장치(513)에 연결될 수 있다. 저장 장치(508)는 운영 시스템(514), 애플리케이션들(516), 및 데이터(526)를 저장할 수 있다.

[0122] 애플리케이션들(516)은 컴퓨터 시스템(502)에 의해 실행될 때, 컴퓨터 시스템(502)으로 하여금 이 개시물에 기술되어 있는 방법들 및/또는 프로세스들을 수행하게 할 수 있는 명령들을 포함할 수 있다, 구체적으로, 애플리케이션들(516)은 비밀키을 결정하기 위한 명령들(셋업 모듈(520)), 평문 값을 암호화하기 위한 명령들(암호화 모듈(522)), 및 암호화된 데이터 값들의 세트로부터 집계 값을 복호화하기 위한 명령들(집계 복호화 모듈(524))을 포함할 수 있다.

[0123] 데이터(526)는 이 개시물에 기술되어 있는 방법들 및/또는 프로세스들에 의해 입력으로서 요구되거나 출력으로서 생성되는 임의의 데이터를 포함할 수 있다. 구체적으로, 데이터(526)는 적어도 비밀 키, 평문 데이터 값들, 랜덤 노이즈 값들, 암호화된 데이터 값들, 및 집계 값을 저장할 수 있다.

[0124] 도 6은 본 발명의 실시예에 따라 프라이버시-보존 데이터 집계를 용이하게 하는 장치를 도시한다. 장치(600)은 유선 또는 무선 통신 채널을 통해 서로 통신할 수 있는 복수의 메커니즘들을 포함할 수 있다. 장치(600)는 하나 이상의 집적 회로들을 이용하여 실현될 수 있고, 장치(600)는 도 6에 도시된 것보다 몇 개 또는 그 이상의 메커니즘들을 포함할 수 있다. 또한, 장치(600)는 컴퓨터 시스템에 통합되거나, 다른 컴퓨터 시스템들 및/또는 장치들과 통신할 수 있는 개별 장치로서 실현될 수 있다. 구체적으로, 장치(600)는 통신 메커니즘(602), 셋업 메커니즘(604), 암호화 메커니즘(606), 및 집계 복호화 메커니즘(608)을 포함할 수 있다.

[0125] 일부 실시예들에서, 통신 메커니즘(602)은 암호화된 데이터 값들을 데이터 집계기에 전송하고, 및/또는 참가자들의 세트로부터 암호화된 데이터 값들을 수신하도록 구성될 수 있다. 또한, 셋업 메커니즘(604)은 비밀 키를 결정하도록 구성될 수 있고, 암호화 메커니즘(606)은 평문 값을 암호화하도록 구성될 수 있고, 집계 복호화 메커니즘(608)은 암호화된 데이터 값들의 세트로부터 집계 값을 복호화하도록 구성될 수 있다.

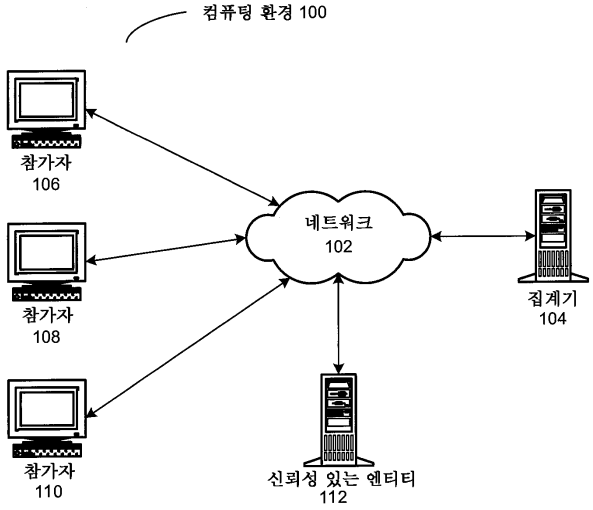
부호의 설명

[0126]

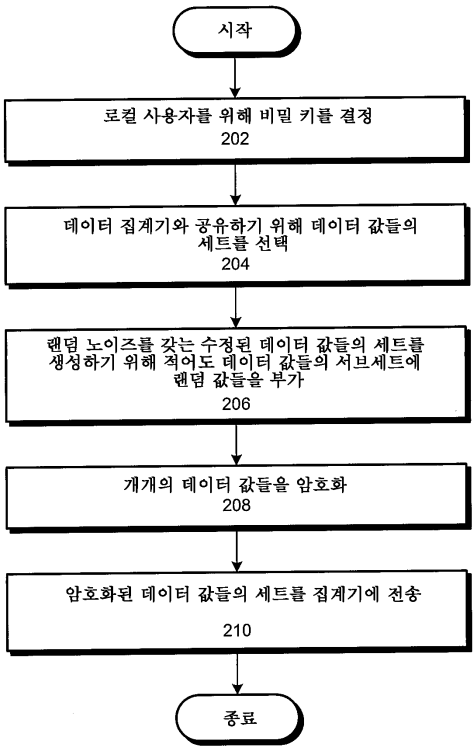
100 : 컴퓨팅 시스템	102 : 컴퓨터 네트워크
104 : 데이터 집계기	112 : 신뢰성 있는 엔티티
602 : 통신 메커니즘	604 : 셋업 메커니즘
606 : 암호화 메커니즘	608 : 집계 복호화 메커니즘

도면

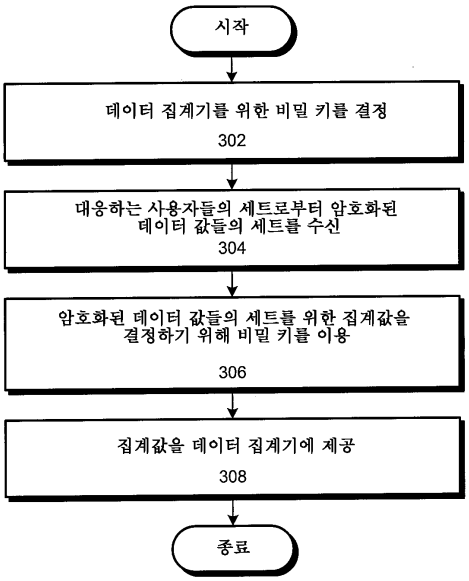
도면1


컴퓨팅 환경 100
참가자
106
참가자
108
참가자
110
네트워크
102
집계기
104
신뢰성 있는 엔티티
112

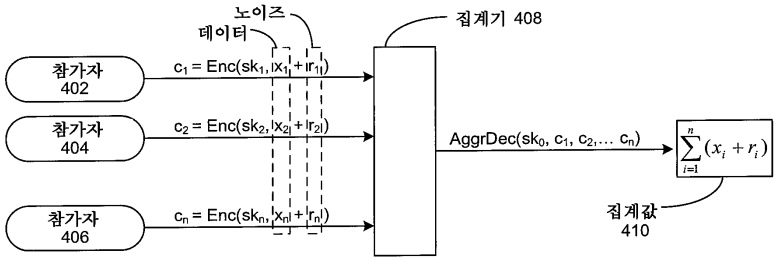
도면2


시작
로컬 사용자를 위해 비밀 키를 결정
202
데이터 집계기와 공유하기 위해 데이터 값들의 세트를 선택
204
랜덤 노이즈를 갖는 수정된 데이터 값들의 세트를 생성하기 위해 적어도 데이터 값들의 서브세트에 랜덤 값들을 부가
206
개개의 데이터 값들을 암호화
208
암호화된 데이터 값들의 세트를 집계기에 전송
210
종료

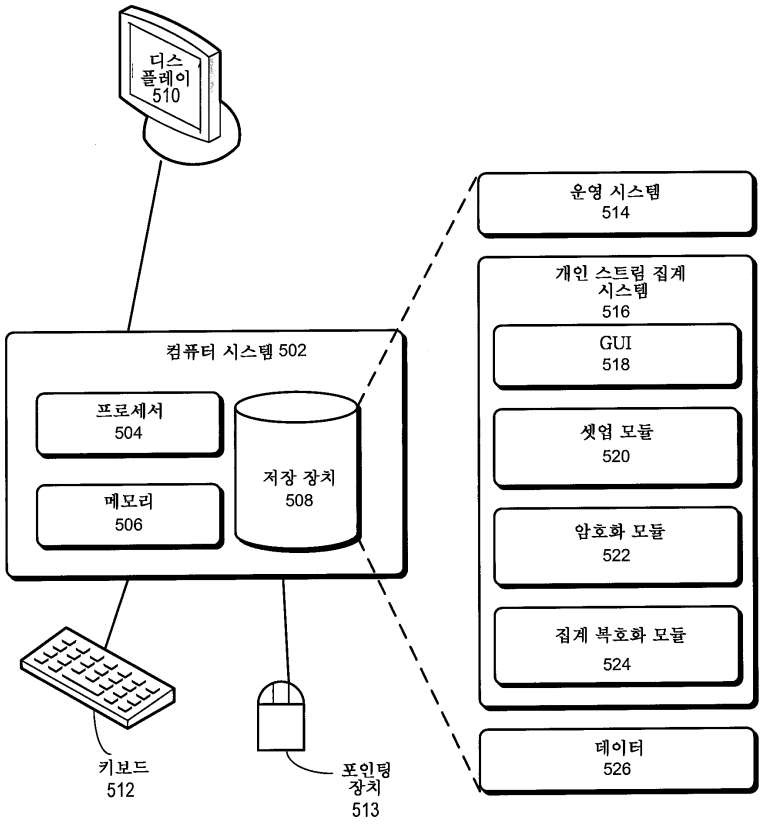
도면3


시작
데이터 집계기를 위한 비밀 키를 결정
302
대응하는 사용자들의 세트로부터 암호화된
데이터 값들의 세트를 수신
304
암호화된 데이터 값들의 세트를 위한 집계값을
결정하기 위해 비밀 키를 이용
306
집계값을 데이터 집계기에 제공
308
종료

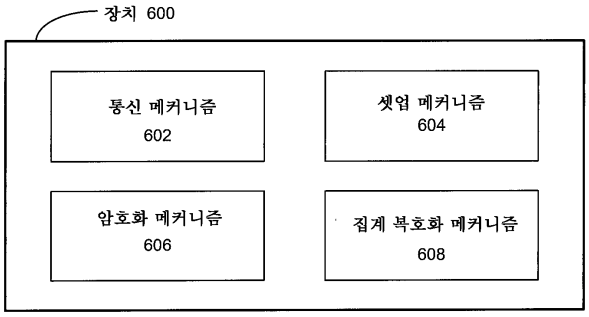
도면4


노이즈
데이터
집계기 408
참가자
402
참가자
404
참가자
406
$c_1 = Enc(sk_1, |x_1| + |r_1|)$
$c_2 = Enc(sk_2, |x_2| + |r_2|)$
$c_n = Enc(sk_n, |x_n| + |r_n|)$
$AggrDec(sk_0, c_1, c_2, \ldots c_n)$
$\sum_{i=1}^{n}(x_i + r_i)$
집계값
410

도면5


디스
플레이
510
운영 시스템
514
개인 스트림 집계
시스템
516
GUI
518
셋업 모듈
520
암호화 모듈
522
집계 복호화 모듈
524
데이터
526
컴퓨터 시스템 502
프로세서
504
메모리
506
저장 장치
508
키보드
512
포인팅
장치
513

도면6


장치 600
통신 메커니즘
602
셋업 메커니즘
604
암호화 메커니즘
606
집계 복호화 메커니즘
608