JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 25 No. 6, pp. 65-72, June 2020
https://doi.org/10.9708/jksci.2020.25.06.065

Privacy-Preserving Aggregation of IoT Data with Distributed Differential Privacy

Jong-Hyun Lim*, Jong-Wook Kim*

*Student, Dept. of Computer Science, Sangmyung University, Seoul, Korea
*Professor, Dept. of Computer Science, Sangmyung University, Seoul, Korea

[Abstract]

Today, the Internet of Things is used in many places, including homes, industrial sites, and hospitals, to give us convenience. Many services generate new value through real-time data collection, storage and analysis as devices are connected to the network. Many of these fields are creating services and applications that utilize sensors and communication functions within IoT devices. However, since everything can be hacked, it causes a huge privacy threat to users who provide data. For example, a variety of sensitive information, such as personal information, lifestyle patters and the existence of diseases, will be leaked if data generated by smarwatches are abused. Development of IoT must be accompanied by the development of security. Recently, Differential Privacy(DP) was adopted to privacy-preserving data processing. So we propose the method that can aggregate health data safely on smartwatch platform, based on DP.

▸ **Key words**: Internet of Things, Privacy-Preserving Aggregation, Distributed Differential Privacy, Smart Healthcare, Homomorphic Encryption, Paillier Cryptosystem

[요 약]

오늘날 사물 인터넷은 우리에게 편의를 제공하기 위해 가정, 산업 현장 및 병원을 포함한 많은 장소에서 사용된다. 다양한 장치가 네트워크에 연결됨에 따라 많은 서비스들이 실시간 데이터 수집, 저장 및 분석을 통해 새로운 가치를 창출하고 있다. 이처럼 많은 분야에서 IoT 장치 내의 센서 및 통신 기능을 활용하는 서비스 및 애플리케이션을 개발하고 있다. 예시로 산업 분야에서 Samsung과 LG는 자사의 IoT 애플리케이션을 통해 가전과 IoT 기기를 연결하여 스마트 홈을 구축하는 서비스를 제공하며, 의료 및 건강 분야에서 Samsung과 Xioami와 같은 기업들은 피트니스 워치 및 앱을 통해 심전도를 확인하거나 운동량을 기록, 관리한다. 위 같은 사례에서 스마트 홈을 구축하는 서비스의 경우에 수집한 데이터를 통해 해당 가정의 생활 패턴이나 출퇴근 여부 등의 민감정보를 유출할 수 있다. 또한 의료 데이터로 사용하기 위해 측정한 데이터를 통해 개인 정보와 질병의 존재와 같은 민감정보를 유출할 수 있다. 따라서 이를 보호하기 위해 해당 논문이 제안하는 방법에 따라 데이터를 수집, 배포한다면 데이터를 제공하는 사용자의 개인 정보 보호에 위협을 막을 수 있다. 이를 해결하기 위해 최근에는 프라이버시 보호 데이터 처리에 차분 프라이버시(DP)가 채택되어왔다. 따라서 DP를 기반으로 스마트워치 플랫폼에서 건강 데이터를 안전하게 수집할 수 있는 방법을 제안하며, 이를 통해 위와 같이 다양한 분야에서 프라이버시를 보호하는 환경에서의 데이터 수집 및 배포를 가능케 할 수 있다.

▸ **주제어**: 사물 인터넷, 프라이버시 보호 수집, 분산 차분 프라이버시, 스마트 헬스케어, 동형암호

• First Author: Jong-Hyun Lim, Corresponding Author: Jong-Wook Kim
*Jong-Hyun Lim (dlsrks2019@naver.com), Dept. of Computer Science, Sangmyung University
*Jong-Wook Kim (jkim@smu.ac.kr), Dept. of Computer Science, Sangmyung University
• Received: 2020. 03. 31, Revised: 2020. 06. 03, Accepted: 2020. 06. 03.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

4차산업혁명위원회에 따르면 세계 헬스케어 산업 시장은 지난 2015년 기준 9조 1,000억 달러에서 2020년 11조 5,000억 달러 규모의 성장이 전망되었다. 이처럼 산업의 핵심 분야 중 하나인 헬스케어는 현재 인공지능(AI), 빅데이터, 사물 인터넷(Internet of Things, IoT) 등의 신기술과 융합하여 새로운 산업의 태동을 이루고 있다.

이처럼 사물 인터넷의 성장은 현대인의 삶 속 깊숙이 자리 잡아 많은 변화를 야기하고 있다. IoT를 통해 사물들을 연결함으로 인해, 외부에서 가정 내의 가전 기기를 원격으로 관리하며 산업 현장에서도 작업 인원과 진행 상황을 실시간으로 확인하는 등 수많은 이점을 만들어 내고 있다. 그중에서도 IoT 장비를 활용한 스마트 헬스케어 관련 산업이 화두가 되고 있다 [1-4]. IoT가 결합한 헬스케어 산업의 예시로 의료계에서는 애플과 같은 글로벌 기업과 손잡고 병원의 환자 전자건강기록(Electronic Health Record, EHR)을 헬스케어 애플리케이션에 연동하고 IoT 장비를 통해 환자가 직접 건강 데이터를 관리하고 병원과 소통하는 등 디지털 혁신을 이뤄내고 있다.

또한 IoT가 접목된 헬스케어 산업은 기기 내의 센서를 적극적으로 활용하여 실시간으로 데이터를 수집하고 기록한다. 예시로 3차원 가속기 센서를 통한 운동량 측정, 혈당 센서와 심혈관계 건강과 밀접한 관련이 있는 심전도 센서를 통한 환자 모니터링 등이 있다. 이처럼 센서를 활용한 IoT 장비는 정보통신 기술과 융합하여 데이터의 처리뿐만 아니라 스스로 연산이 가능한 컴퓨팅 장비의 역할을 수행한다.

하지만 위와 같은 산업의 발전은 개인의 프라이버시에 대한 침해 문제를 발생시켜 해결해야 하는 새로운 과제를 생성하였다. IoT 장비에서 생성되고 수집되는 수많은 데이터는 개인이 활동하는 시간, 장소, 질병의 유무 등 여러 데이터를 포함하고 이를 통한 개인에 대한 식별을 가능케 하여 심각한 사생활 침해 문제를 발생시킨다. 이러한 프라이버시 문제를 해결하기 위하여 앞서 많은 연구가 진행되어 왔다. 위 연구에서는 사용자의 프라이버시를 지키기 위해 네트워크 정책과 데이터를 주고받는 프로토콜, 그리고 접근 제어 방식의 신중한 선택과 암호화, 비식별화를 통한 데이터 변조 등 다양한 방법이 시도되었다 [5-7].

본 논문은 이전에 수행한 연구를 바탕으로 발전시킨 것으로, 이전의 논문은 개인의 건강 데이터를 안전하게 수집할 수 있는 차분 프라이버시 모델을 바탕으로 심박 수 데이터에 대해 히스토그램 형태로 수집하고 가공하는 데 집중하였다 [8]. 하지만 지역 차분 프라이버시(Local Differential Privacy, LDP)의 특징으로 인해 활용도 높은 데이터를 생성하기 위해서 많은 수의 사용자를 필요로 하는 한계점을 가졌고, 이를 해결하기 위해 본 논문에서는 분산 차분 프라이버시(Distributed Differential Privacy, DDP)를 적용하였다. DDP는 LDP와는 달리 매개변수를 분산하는 방식의 데이터 변조를 수행하고 데이터를 암호화하는 단계가 추가되어 앞선 방식보다 적절한 수준의 변조를 통해 더욱 활용도 높은 데이터 수집과 배포를 가능케 한다. 본 논문에서는 이를 통해 신뢰할 수 있는 데이터 수집가가 없는 환경에서 원본 데이터에 대한 보호를 수행하고 안전한 수집과 배포를 가능케 하였다.

II. Background

1. Differential Privacy

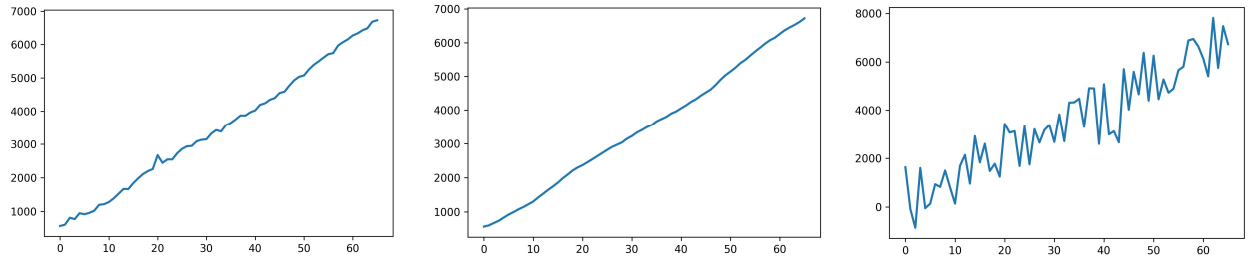
차분 프라이버시(Differential Privacy, DP)는 공격자가 어떤 배경지식을 가지고 있다 해도 해당 데이터베이스에서 도출된 통계 결과로부터 특정 개인을 유추할 수 없도록 하는 수학적 모델이다. 이 모델은 신뢰할 수 있는 데이터 수집가가 존재하며 데이터 수집가가 원본 데이터를 수집하여 변조하고 이를 데이터 사용자에게 전달한다. 이때 프라이버시 보호의 수준을 결정하는 매개변수 ε를 사용하고, ε-DP를 만족하는 알고리즘 A는 다음과 같이 정의된다.

$$P[A(D_a)\in S] \le e^{\varepsilon} \times P[A(D_b)\in S] \quad (1)$$

식 (1)에서 S는 A가 도출할 수 있는 결괏값의 집합이며 A를 거친 데이터셋 D_a와 D_b로부터 도출된 통계 결과가 같을 확률을 e^{ε}이하로 제한함을 의미한다 [9]. 이때 ε가 커질수록 원본과 가까운 값을 유추할 수 있다. ε-DP를 만족하는 알고리즘에는 라플라스, 지수 메커니즘 등이 있으며, n개의 알고리즘이 각각 ε-DP를 만족한다면 다음 식 (2)와 같이 순차 구성 정리를 만족한다.

$$\sum_{k=1}^{n} \varepsilon^k = \varepsilon \quad (2)$$

식 (2)의 순차 구성 정리는 n번 만큼 반복된 메커니즘에 사용된 프라이버시 비용의 합이 ε을 넘지 않음을 의미한다 [10].



(a) DP (b) DDP (c) LDP

Fig. 1. Estimated Data with DP, DDP, LDP based approach (ε=0.5, user count=443)

2. Distributed Differential Privacy

하지만 실제 환경에서는 신뢰할 수 있는 데이터 수집가가 반드시 존재하지는 않는다. 이를 위해 사용자가 직접 데이터를 변조하는 지역 차분 프라이버시(Local Differential Privacy, LDP)와 분산 차분 프라이버시(Distributed Differential Privacy, DDP) 모델이 존재한다. 각각의 사용자가 ε-DP를 만족하는 LDP와는 달리, DDP는 모든 사용자의 데이터를 통합하였을 때 비로소 ε-DP를 만족한다. 다시 말해 각각의 사용자는 ε-DP를 만족할 만큼 충분한 변조가 이루어지지 않기 때문에 추가로 암호화의 과정이 필요하다. DDP는 DP에서 사용되는 라플라스 분포의 분할성을 바탕으로 변조를 수행하며 다음과 같은 식으로 정의된다.

$$L(\lambda) = \sum_{i=1}^{N} [G^1(N,\lambda) - G^2(N,\lambda)] \quad (3)$$

식 (3)에서 λ을 척도 모수(scale parameter)로 가지는 라플라스 분포 L은 N과 λ을 각각 형태 모수(shape parameter)와 척도 모수로 가지는 N개의 감마 분포 G로 분할되며, N개의 감마 분포에서 독립적으로 추출된 실수의 차를 모두 합친 것이 라플라스 분포에서 생성된 잡음의 크기와 같음을 의미한다 [11-13].

위 그림 1에서 DDP를 통해 변조를 수행한 결과는 LDP와는 다소 차이를 보이지만, DP를 적용한 데이터와는 차이가 거의 없음을 알 수 있다. 이처럼 신뢰할 수 있는 데이터 수집가가 없는 환경이지만 신뢰할 수 있는 데이터 수집가가 있는 환경과 유사한 수준의 프라이버시를 보장하는 데 그 의의가 있다.

또한 아래 그림 2는 DP와 DDP 프로토콜의 전체적인 개념도를 비교하여 나타낸 것으로, 신뢰할 수 있는 수집가의 존재 여부와 잡음을 추가하는 주체, 그리고 추가적인 암호화의 여부 차이를 나타낸다.

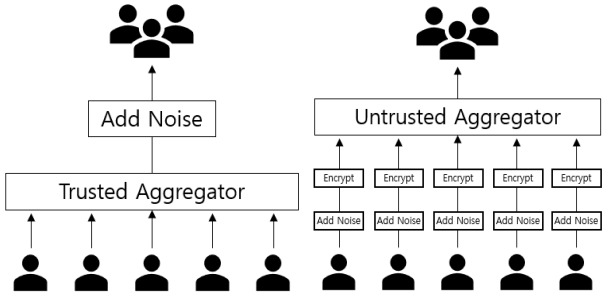


Fig. 2. Differential Privacy vs Distributed Differential Privacy

3. Homomorphic Encryption

DDP 환경에서는 위 2의 과정에서 부족한 데이터의 보안성을 확보하기 위하여 암호화 과정을 추가한다. 하지만 기존의 암호화 기술은 복호화의 과정 없이는 데이터를 활용할 수 없었다. 이를 극복하기 위해 동형 암호(Homomorphic Encryption)를 활용한다. 동형암호는 평문 m_1, m_2를 연산한 결과와 암호문 $E(m_1)$과 $E(m_2)$를 연산하였을 때 동일한 결과를 가지는 암호화 방식이다. 하지만 일정 횟수 이상 연산을 수행하면 노이즈가 발생하여 더 연산을 수행할 수 없으며, 암호화 과정 자체의 연산이 무거워 비교적 시간이 소요되는 단점이 있다. 동형암호에는 복호화 키 없이 무한히 연산을 수행할 수 있는 완전 동형 암호 방식, 덧셈과 곱셈만 가능한 부분 동형 암호 방식과 제한된 횟수의 연산만 가능한 제한 동형 암호 방식이 있다. 해당 논문에서는 부분 동형 암호의 일종인 Paillier Encryption을 사용하였다. Paillier Encryption은 암호화된 숫자에 덧셈하거나 암호화되지 않은 스칼라 수를 곱할 수 있다 [14-16].

III. DDP-based Data Aggregation

본 장에서는 신뢰할 수 있는 데이터 수집가가 없을 때, 개인의 건강 데이터를 프라이버시를 보호하는 데이터 수집(Privacy-Preserving Data Collection) 환경을 만족하

는 DDP를 통해 안전하게 수집하고 집계하는 방법에 대해 두 단계로 나누어 상세히 논의한다. 아래의 그림 3은 본 논문에서 제안하는 시스템의 구성도로, 각 사용자가 스마트워치를 통해 직접 DDP 환경에서 데이터를 처리하여 보낸 것을 데이터 수집가가 집계하여 분석하고 데이터가 있어야 하는 집단에 제공함을 나타낸다.

본 논문의 데이터 수집 시나리오에서 사용자, 즉 데이터 소유자 N명은 각각의 스마트워치의 센서를 통해 데이터를 수집하고자 한다. 또한 사용자는 하루 동안 활동하면서 걸음 수 데이터를 일정 간격에 따라 시계열 형태로 수집하고자 한다. 먼저 스마트워치는 센서를 통해 수집된 누적 걸음 수를 기기 내에 저장한다.

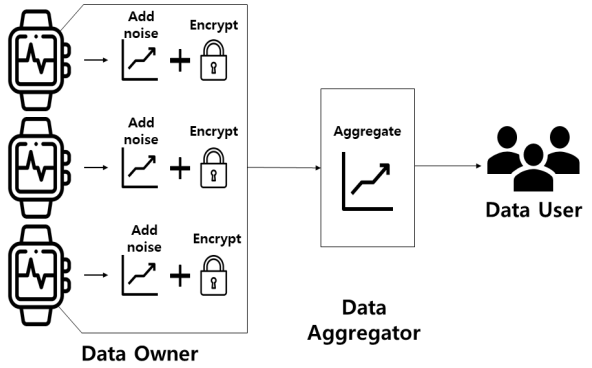


Fig. 3. System Architecture

1. Data Owner's Smartwatch-Side Processing

사용자의 로컬 기기 내에 저장된 데이터는 타임 스탬프와 해당 시간에 스마트워치의 센서에서 측정된 값으로 이루어진 쌍의 형태이다. 이처럼 하루 동안 수집된 데이터는 매일 측정을 마치면, DDP를 통해 변조와 암호화가 되어 데이터 수집가에게 보내진다. 각각의 사용자 데이터는 ε-DP를 만족하지 못한다. 하지만 모든 사용자의 데이터를 집계하였을 때 비로소 ε-DP를 만족하게 된다. 정해진 시간 동안 수집된 i번째 사람의 누적 걸음 수 원본 데이터 k개는 시계열 형태이며 다음 식 (4)와 같이 표현할 수 있다.

$$S_i^O = \{(t_1, d_{i1}), (t_2, d_{i2}), \cdots, (t_k, d_{ik})\} \tag{4}$$

N명 중 i번째 사용자의 원본 데이터 S_i^O가 생성된 후 기기 내에서 DDP를 적용한 데이터 변조가 다음과 같이 수행된다. 생성되는 잡음은 기존의 연구에서 사용된 LDP 환경의 라플라스 메커니즘에서 발전된 형식이다. 기존의 라플라스 메커니즘은 척도 모수 λ가 $\frac{\triangle s}{\varepsilon/k}$이고 평균이 0인 라플라스 분포 L에서 추출한 임의의 실수를 활용해 변조하였다. 이때 $\triangle s$는 전역 민감도로 측정값의 최댓값에서 최솟값을 제한 크기와 같으며, 하나의 레코드가 전체 통계 결과에 미칠 수 있는 최대의 영향력이다. 이와 달리 DDP는 라플라스 분포 L이 N개의 감마 분포 G로 분할되며, 형태 모수를 N, 척도 모수를 $\frac{\triangle s/N}{\varepsilon/k}$로 가지는 동일한 감마 분포로부터 추출한 임의의 실수 두 개의 차를 잡음으로 추가한다. 이와 같은 과정을 통해 생성된 잡음이 추가된 i번째 사용자의 j번째 데이터 d_{ij}^E는 다음과 같은 형태로 표현된다.

$$d_{ij}^E = d_{ij} + \{G_i^1(N,\lambda) - G_i^2(N,\lambda)\} \tag{5}$$

위 식 (5)에서 G^1과 G^2는 동일한 감마 분포 G이며 임의의 실수를 독립적으로 2회 추출함을 의미한다. 추출된 두 실수의 차를 이용해 변조한 i번째 사용자의 데이터 k개는 다음식 (6)과 같이 표현된다.

$$S_i^E = \{(t_1, d_{i1}^E), (t_2, d_{i2}^E), \cdots, (t_k, d_{ik}^E)\} \tag{6}$$

이처럼 변조된 데이터 S_{ij}^E는 LDP의 라플라스 메커니즘 방식과는 달리 척도 모수가 사용자별로 나뉘어 사용되어 추가되는 잡음의 크기가 매우 작아져 4장의 실험결과 그림 5의 (d), (e)에서 볼 수 있듯 그 차이가 크지 않음을 알 수 있다.

이를 보완하기 위해 우리는 동형 암호 방식 중 Paillier 암호화 방식을 사용하였다. Paillier 방식에 따라 먼저 정해진 bit 수에 맞게 Public(암호화) Key와 Private(복호화) Key의 쌍을 생성한다. 다음 위 식 6의 데이터 k를 평문(Plain text)으로 지정하여 생성된 암호화 키를 통해 k개의 암호문(Cipher text)을 생성한다. 이는 신뢰할 수 없는 데이터 수집가에게 전송하는 도중 개인의 데이터에 접근함으로써 발생할 수 있는 프라이버시 침해를 막기 위함이다. 이렇게 데이터 소유자가 각각의 스마트워치를 통해 변조하고 암호화한 결과는 하루에 한 번 TCP/IP 프로토콜을 통해 서버로 전송된다.

2. Data Aggregator's Server-Side Processing

클라이언트 단에서 전송된 데이터는 데이터 수집가에 의해 집계되어 통계 결과로 만들어진다. 아래 그림 4는 각 사용자가 데이터를 직접 변조하고 암호화하여 데이터 수집가에게 보내고 데이터 수집가가 이를 집계하여 필요로 하는 집단에 제공하는 과정을 나타낸다. 그림 4의 사용자

N명의 암호화 된 데이터 j개를 합친 통계 결과는 $\sum_{i,j=1}^{N,k} Enc(d_{ij}+\sigma_i)$로 표현되며 이 통계 결과를 필요로 하는 집단에 제공할 시 위 결과를 Private 키를 통해 복호화하여 제공한다.

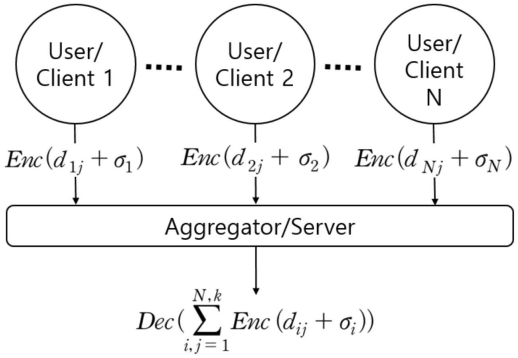


Fig. 4. Simplified process of encryption and decryption

또한 위 그림에서 N개의 평문 $P_i = d_{ij} + \sigma_i$으로부터 각각 암호문 C_i을 생성하고 이를 집계하여 통계 결과를 생성하는데 이때 평문을 연산한 결과와 암호문을 연산한 결과가 같은 부분 동형 암호의 성질을 사용한다. 이를 간단히 설명하기 위하여 위의 사용자 N명 중 i번째 사람의 평문과 암호문을 연산한 결과를 다음 수식과 같이 나타낼 수 있다.

$$\frac{1}{N}\times\sum_{i=1}^{N}P_i = \frac{1}{N}\times\sum_{i=1}^{N}C_i \qquad (7)$$

식 (7)에서 $C_i = Enc(P_i)$는 암호문을 나타내며 평문에 대해 통계 결과를 연산한 결과와 암호문에 대해 연산한 통계 결과가 같음을 알 수 있다. 이러한 과정을 통해 데이터 수집가는 개인의 데이터가 아닌 암호화된 통계 데이터에만 접근 가능하므로 프라이버시를 보장하는데 그 의의가 있다.

IV. Experimental Evaluation

1. Experimental Setup

본 논문에서 제안한 프라이버시 보호 환경에서 안전한 데이터 수집 방법을 검증하기 위해 독립 변인, 종속 변인을 각각 프라이버시 보호 매개변수 ε와 평균 절대 오차(Mean Absolute Error, MAE)를 두고 실험하였다. 본 논문의 시나리오에 따라 안전하게 데이터를 수집하고자 하는 사용자는 Samsung Gear S3 워치를 사용하여 실험에 참여하였다. 각각 사용자들은 기기 내의 보수계(Pedometer) 센서를 통해 하루 동안 자유롭게 활동하며 10분의 시간 간격으로 측정된 걸음 수를 기기 내 로컬 환경에 저장한다. 활동을 마치고 정해진 시간이 지났을 경우 저장된 데이터를 추출하고 본문에서 제안한 과정을 통해 데이터를 처리하여 서버로 보내게 된다. 해당 데이터를 처리하기 위해 본 실험 당시 아래 표 1과 같은 사양의 Ubuntu 16.04 버전의 서버를 사용하였다.

Table 1. Server specs used in the system

Server	HDD size	4 TB
	Memory size	64 GB
	CPU	4
	CPU model name	Intel(R) Xeon CPU E3-1230 v5 @ 3.40GHz

실험에 참여한 사용자 443명의 데이터를 하루 동안 기록하며 모든 사용자의 데이터로 통계 결과를 내어 이를 활용한다.

해당 실험 과정에서는 3장의 과정에서 암호화된 데이터를 복호화한 것을 바탕으로 원본 데이터와 비교하여 평균 절대 오차(Mean Absolute Error, MAE)를 계산하여 데이터의 활용성을 측정하였다. 본 논문에 사용된 동형암호 방식에 따르면 복호화 없이도 통계 결과를 낼 수 있으며 이를 활용하기 위해 통계 결과를 제공할 때 암호문을 해독하여도 모든 사용자의 변조된 데이터의 집계 결과만을 얻을 수 있다. 이처럼 어떤 상황에서도 사용자 데이터의 프라이버시를 지킬 수 있음을 보장하기 위해 복호화된 데이터를 활용하여 데이터의 활용도를 측정하였다. 이때 평균 절대 오차를 계산하는 식은 다음과 같다.

$$MAE = \frac{1}{k}\times\sum_{j=1}^{k}|d_j^O - d_j^E| \qquad (8)$$

위 식 (8)은 원본과 변조된 데이터 간 오차의 절댓값을 평균 낸 결과로 그 수치가 작을수록 원본에 가깝다. 또한 평균 절대 오차가 작을수록 데이터를 필요로 하는 집단에 제공될 때 그 활용도가 높다고 할 수 있다. 이를 통해 프라이버시 보호 매개변수 ε는 데이터 활용도와 이율배반적 관계에 있음을 알 수 있다. 추가로 평균 상대 오차(Mean Relative Error, MRE)를 계산하여 활용성을 측정하려 했으나 DDP로 인해 추가되는 잡음의 크기가 충분하지 않아 유의미한 결과를 낼 수 없어 제외하였다.

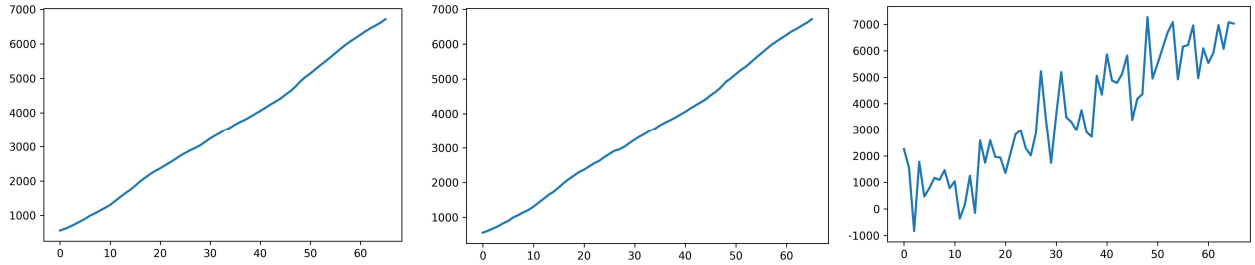


Fig. 5. Estimated Data (data size=443, ε=0.5) (a) original data (b) DDP based approach (c) LDP based approach

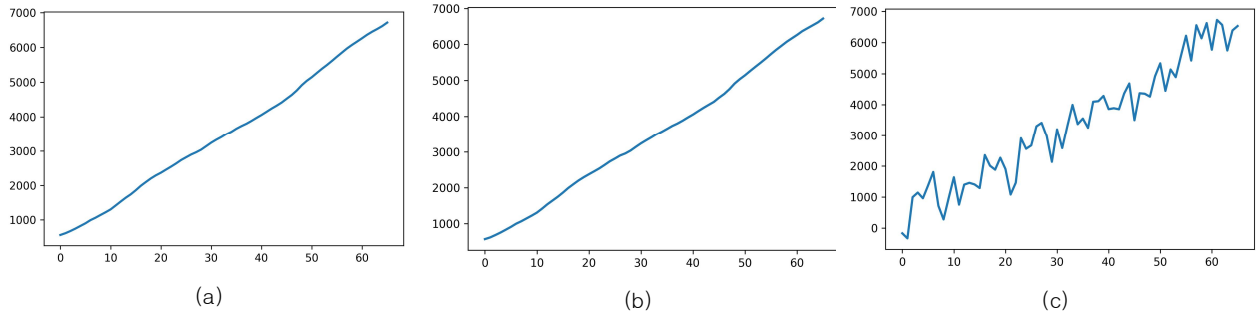


Fig. 6. Estimated Data (data size=443, ε=1.0) (a) original data (b) DDP based approach (c) LDP based approach

2. Results and Discussion

하루 동안 스마트워치와 함께 활동한 사용자들의 데이터가 서버로 전송될 때, 3장에서 제안한 방법으로 변조되고 암호화된다. 위 그림 5는 443명에 대한 실험결과로 프라이버시 보호 매개변수 ε를 0.5일 때 (a)는 원본, (b)는 DDP 방식, (c)는 LDP 방식으로 추정한 443명의 평균 데이터이며, 그림 6 또한 같은 크기의 데이터에 ε를 1.0으로 달리 책정하여 추정한 443명의 평균 데이터이다. 또 그림 5와 6의 (b)는 DDP 방식의 추정으로 변조와 암호화의 과정을 거친 데이터를 복호화한 결과이다. 이 두 그림에서 확인할 수 있듯, DDP 방식으로 추정된 데이터는 원본과 크게 차이가 나지 않으며 ε이 0.5에서 1.0으로 주어질 때도 차이를 체감하기 어렵다. 반면에 LDP 방식에서 ε을 증가시켰을 때 인지할 만한 차이가 나며, DDP 방식 추정 결과와도 큰 차이를 보인다.

위 두 그림에 사용된 데이터는 사용자가 주어진 프라이버시 보호 매개변수 ε를 통해 직접 데이터를 변조한 것이다.

또한 서버는 프라이버시 비용을 0.1, 0.5, 1.0의 3가지 수준으로 사용하여 데이터를 추정하고 원본 데이터와의 편차를 측정하여 아래 표 2와 같은 결과를 보였다. 아래 표 2에서 확인할 수 있듯, 프라이버시 보호 매개변수 ε을독립변인으로 두고 변화시켜 실험을 진행하였으며, 443명의 사용자에 대한 데이터를 수집하여 DDP와 LDP 방법에 따라 추정한 데이터의 평균 절대 오차를 나타낸다. ε을 0.1, 0.5, 1.0으로 책정하였을 때 LDP의 경우, 각각 평균 절대 오차가 162.21, 32.76, 14.72로 계산되었으며 DDP의 경우, 위 그림 5와 6에서는 ε에 따른 오차의 차이를 체감하기 어렵지만, 각각 44.96, 9.12, 4.82로 계산되어 ε의 증가에 따른 오차 감소를 확인할 수 있었으며 DDP가 LDP보다 오차의 크기가 작음을 알 수 있다. 이러한 수치에서 확인할 수 있듯 프라이버시 보호 매개변수 ε가 많이 주어질수록 평균 절대 오차는 작아지며 원본 데이터와의 편차가 작다. 다시 말해 ε가 크면 그만큼 원본과 유사한 데이터를 추정할 수 있어 실제 통계 데이터를 활용하면서 보다 더 유의미한 결과를 끌어낼 수 있다. 하지만 ε을 사용자에게 배분하면서, 그 크기에 따라 얼마나 원본에 가까운 데이터를 추정하는 정확도의 차이를 발생시키기에 데이터를 필요로 하는 집단의 성격에 따라 세심하게 책정할 필요가 있다.

Table 2. Mean Absolute Error for varying privacy budget with DDP, LDP basesd approach

Privacy Budget	0.1	0.5	1.0
Mean Absolute Error(DDP)	44.96	9.12	4.82
Mean Absolute Error(LDP)	162.21	32.76	14.72

V. Conclusions and Future Work

이번 연구를 통해, 우리는 프라이버시를 보호하는 환경에서 DDP를 활용해 IoT 기기의 건강 데이터를 안전하게 수집하는 방법을 제안했다. 본 논문에서 제안한 기법은 우리의 기존 연구에 전제된 LDP 환경의 단점을 보완하여 적은 사용자 수로도 유의미한 결과를 끌어낼 수 있었다. 본 연구는 신뢰할 수 있는 수집가가 없는 환경에서도 DP와 유사한 수준의 보안성을 확보 가능하다는 점에서 의의가 있다. 또한 라플라스 분포의 분할성을 이용한 감마 분포를 통한 세밀한 잡음의 생성 및 동형 암호를 활용한 보안성 확보를 이루어냈으며, 최악의 상황으로 암호 키가 유출되어 암호문을 해독하여도 변조된 통계 데이터만 접근하여 개인의 민감 데이터에 대한 프라이버시를 보장하였다. 추후 연구에서는 보수계 센서뿐만 아니라 더욱 다양한 센서를 활용하여 적용할 수 있는 데이터의 범위를 넓히고자 한다. 또한, 한정된 연산이 가능한 부분 동형 암호 방식을 보완할 추가적인 연구가 필요하다.

REFERENCES

[1] A ligtweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access, 5, 3302-3312.

[2] Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. International journal of communication systems, 25(9), 1101.

[3] Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., & Tarricone, L. (2015). An IoT-aware architecture for smart healthcare systems. IEEE Internet of Things Journal, 2(6), 515-526.

[4] Mahmud, R., Koch, F. L., & Buyya, R. (2018, January). Cloud-fog interoperability in IoT-enabled healthcare solutions. In Proceedings of the 19th international conference on distributed computing and networking (pp. 1-10).

[5] Samani, A., Ghenniwa, H. H., & Wahaishi, A. (2015). Privacy in Internet of Things: A model and protection framework. (pp.606). Procedia Computer Science, 52, 606-613.

[6] Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. (2015). Big data privacy in the internet of things era. IT Professional, 17(3), 32-39.

[7] Abomhara, M., & Køien, G. M. (2014, May). Security and privacy in the Internet of Things: Current status and open issues. In 2014 international conference on privacy and security in mobile systems (PRISMS) (pp. 1-8). IEEE.

[8] Kim, J. W., Lim, J. H., Moon, S. M., Yoo, H., & Jang, B. (2019, January). Privacy-Preserving Data Collection Scheme on Smartwatch Platform. In 2019 IEEE International Conference on Consumer Electronics (ICCE) pp. 2. IEEE.

[9] Kim, J. W., Kim, D. H., & Jang, B. (2018). Application of local differential privacy to collection of indoor positioning data. pp. 1. IEEE Access, 6, 4276-4286.

[10] McSherry, F. D. (2009, June). Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data (pp. 19-30).

[11] Goryczka, S., Xiong, L., & Sunderam, V. (2013, March). Secure multiparty aggregation with differential privacy: A comparative study. In Proceedings of the Joint EDBT/ICDT 2013 Workshops (pp. 155-163).

[12] Ács, G., & Castelluccia, C. (2011, May). I have a dream!(dif ferentially private smart metering). In International Workshop on Information Hiding (pp. 118-132). Springer, Berlin, Heidelberg.

[13] Shi, E., Chan, T. H., Rieffel, E., Chow, R., & Song, D. (2011). Privacy-preserving aggregation of time-series data. In Proc. NDSS (Vol. 2, pp. 1-17).

[14] Paillier, P. (1999, May). Public-key cryptosystems based on composite degree residuosity classes. In International conference on the theory and applications of cryptographic techniques (pp. 223-238). Springer, Berlin, Heidelberg.

[15] Boneh, D., Gentry, C., Halevi, S., Wang, F., & Wu, D. J. (2013, June). Private database queries using somewhat homomorphic encryption. In International Conference on Applied Cryptography and Network Security (pp. 102-118). Springer, Berlin, Heidelberg.

[16] Tebaa, M., El Hajji, S., & El Ghazi, A. (2012, July). Homomor phic encryption applied to the cloud computing security. In Proceedings of the World Congress on Engineering (Vol. 1, No. 2012,pp. 4-6).

Authors

Jong-Hyun Lim received the B.S. and M.S. student in Department of Computer Science from Sangmyung University, Seoul, Korea, in 2018, 2019 respectively. He is interested of dealing with system infrastructure. Such as Cloud-Computing, Fog-Computing, and cloud system like AWS and Azure. And he is studying system infrastructure.

Jong-Wook Kim received the Ph.D. degree from the Computer Science Department, Arizona State University, in 2009. Dr. Kim was a Software Engineer with the Query Optimization Group at Teradata, from 2010 to 2013. He is currently an Assistant Professor of computer science with Sangmyung University. His primary research interest is in the area of data privacy, distributed databases, and query optimization. He is a member of the ACM.