Journal of The Korea Society of Computer and Information
Vol. 24 No. 9, pp. 43-49, September 2019
www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.09.043

Privacy-Preserving IoT Data Collection in Fog-Cloud Computing Environment

Jong-Hyun Lim*, Jong Wook Kim*

Abstract

Today, with the development of the internet of things, wearable devices related to personal health care have become widespread. Various global information and communication technology companies are developing various wearable health devices, which can collect personal health information such as heart rate, steps, and calories, using sensors built into the device. However, since individual health data includes sensitive information, the collection of irrelevant health data can lead to personal privacy issue. Therefore, there is a growing need to develop technology for collecting sensitive health data from wearable health devices, while preserving privacy. In recent years, local differential privacy (LDP), which enables sensitive data collection while preserving privacy, has attracted much attention. In this paper, we develop a technology for collecting vast amount of health data from a smartwatch device, which is one of popular wearable health devices, using local difference privacy. Experiment results with real data show that the proposed method is able to effectively collect sensitive health data from smartwatch users, while preserving privacy.

▸Keyword: e-Health, IoT, Data Collection, Local Differential Privacy, Fog-Cloud Computing

I. Introduction

사물 인터넷 기술은 스마트 홈(smart home), 스마트 그리드(smart grid), 스마트 헬스(smart health)와 같은 다양한 분야에서 활용되고 있다. 그중 스마트 헬스 분야는 개인 건강에 대한 관심의 증가로 인하여, 사물 인터넷 기술이 가장 활발히 적용되는 있는 분야이다 [1]. 가령, 헬스 관련 웨어러블 디바이스 시장이 최근 들어 급성장을 하고 있으며, 샤오미, 핏빗, 소니 같은 글로벌 정보통신기술 기업들은 다양한 제품을 출시하고 있다. 각종 센서를 장착한 헬스 웨어러블 디바이스는 사용자로부터 심장 박동 수, 걸음 수, 칼로리 등과 같은 개인 건강 정보를 쉽게 수집할 수 있게 한다.

최근 미국식품의약국(FDA)는 정보통신기술 기업들이 미국식품의약국의 승인 절차 없이 헬스 관련 웨어러블 디바이스 제품을 개발하는 것을 허가 하였다. 이로 인하여, 사용자의 건강 상태를 측정할 수 있는 다양한 웨어러블 디바이스 제품들이 시장에 출시되고 있다. 이러한 헬스 웨어러블 디바이스 기술 발전으로 인하여, 다양한 사용자로부터 방대한 양의 건강 데이터를 수집하고 분석하여 서비스에 활용하려는 시도가 활발히 이루어지고 있다. 하지만, 개인의 건강 데이터는 민감한 정보를 포함하고 있으므로, 무분별한 건강 데이터 수집은 개인의 프라이버시 침해 문제를 야기 시킬 수 있다 [2-3]. 그러므로 개인의 프라이버시를 보존하면서 헬스 웨어러블 디바이스로부터 민감한 건강 데이터를 수집하기 위한 기술 개발에 대한 필요성이 높아지고 있다.

최근 들어 프라이버시를 보존하면서 민감한 데이터 수집을 가능하게 하는 지역 차분 프라이버시(Local Differential Privacy, LDP) 기술이 주목 받고 있다 [4-6]. 본 논문에서는 지역 차분 프라이버시를 활용하여 포그-클라우드(Fog-Cloud) 컴퓨팅 환경에서 대중적인 헬스 웨어러블 디바이스 제품에 해당하는 스마트워치 사용자로부터 방대한 건강 데이터를 수집하기 위한 기술을 개발 한다. 그림 1에 나타나듯이 사용자의 스마트워치 내에서 지역 차분 프라이버시를 만족하도록 건강 데이터에 대한 변조가 이루어지고, 변조된 건강 데이터를 데이터 수집가에게 전송한다. 그러므로 스마트워치

• First Author: Jong-Hyun Lim, Corresponding Author: Jong Wook Kim
*Jong-Hyun Lim (dlsrks2019@naver.com), School of Computer Science, Sangmyung University
*Jong Wook Kim (jkim@smu.ac.kr), Dept. of Computer Science, Sangmyung University
• Received: 2019. 06. 24, Revised: 2019. 08. 24, Accepted: 2019. 08. 26.

사용자의 원본 데이터가 외부로 노출되는 것을 방지할 수 있다. 데이터 수집가는 다수의 스마트워치 사용자로부터 변조된 건강 데이터를 수집 후, 수집된 데이터를 데이터 사용자 (예, 스마트 헬스 케어 서비스 제공자, 데이터 분석가)에게 배포한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 본 논문의 배경지식에 대해 설명하고 3장에서는 지역 차분 프라이버시를 이용하여 스마트워치 사용자로부터 건강 데이터를 수집하기 위한 기술을 제안한다. 4장에서 본 논문에서 제안하는 기법의 성능평가를 수행한 후, 5장에서 결론을 맺는다.

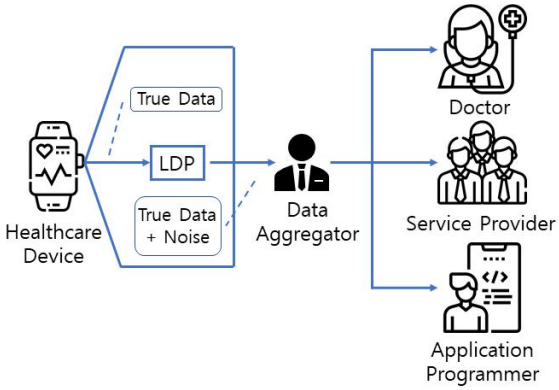


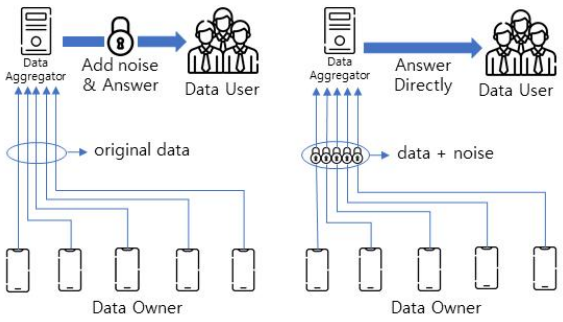
Fig. 1. Motivational Example

II. Background

1. Differential Privacy vs. Local Differential Privacy

차분 프라이버시 모델은 신뢰할 수 있는 데이터 수집가가 존재한다고 가정한다. 각각의 데이터 소유자(data owner)들은 원본 데이터를 데이터 수집가(data aggregator)에서 전송하고, 데이터 수집가는 각각의 데이터 소유자로부터 수집된 데이터를 변조한 후, 변조된 데이터를 데이터 사용자(data user)에게 제공 한다 (그림 2-(a)). 데이터 수집가는 변조된 데이터를 데이터 사용자에 제공하므로, 데이터 소유자의 프라이버시를 보호할 수 있다. 변조 과정에서 매개변수인 프라이버시 예산(ε)을 사용하여 프라이버시의 보호 수준을 결정한다. 이 모델은 공격자가 어떤 배경지식을 가지고 있든 데이터베이스에서 도출된 통계 결과에서 특정한 개인의 민감 속성을 알아내지 못하게 하는 것을 목표로 한다 [6].

하지만 실제 데이터 수집 환경에서 신뢰할 수 있는 데이터 수집가가 반드시 존재하지는 않는다. 지역 차분 프라이버시(LDP)는 위에 설명한 차분 프라이버시(DP)와는 달리 신뢰할 수 있는 데이터 수집가가 없는 환경을 가정하여 사용하는 모델이다 [6-7]. 이때 데이터 소유자가 직접 원본 데이터에 대해 차분 프라이버시를 만족하도록 변조를 수행하고, 변조된 데이터를 데이터 수집가에게 전송한다. 데이터 수집가는 변조된 데이터를 데이터 사용자에게 공유 한다 (2-(b)).



(a). Differential Privacy (b) Local Differential Privacy

Fig. 2. (a) DP vs. (b) LDP

ϵ-지역 차분 프라이버시를 만족하는 알고리즘 A는 다음과 같이 정의 된다.

$$\frac{\Pr[A(v_a)=O]}{\Pr[A(v_b)=O]} \leq e^{\epsilon} \qquad (1)$$

위의 식 (1)은 다음을 의미한다. 데이터 소유자의 데이터 v_a와 v_b를 알고리즘 A를 통해 변조한 결과가 O로 같을 확률을 e^{ϵ} 이하로 제한하여 데이터 수집가 (혹은, 데이터 사용자)로 하여금 데이터 소유자가 보낸 데이터가 v_a 인지 v_b 인지 높은 확률로 확신할 수 없게 한다. 이처럼 지역 차분 프라이버시에서 데이터를 변조하는 방식으로는 라플라스 기법과 임의화 응답 기법(randomized response)이 사용 된다 [8-11].

또한, 지역 차분 프라이버시는 다음과 같은 순차구성 특징을 만족한다. n개의 알고리즘이 각각 ϵ-차분 프라이버시를 만족할 때 각 메커니즘을 동일한 데이터에 적용하여 $\sum_{i=1}^{n}\epsilon_i = \epsilon$일 때, 알고리즘 A는 ϵ-지역 차분 프라이버시를 만족하고 이를 순차 구성이라고 한다. 그리고 각 메커니즘에서 사용된 프라이버시 비용의 합은 ϵ을 넘지 않아야 한다.

2. Fog-Computing

포그 컴퓨팅은 사용자 기기와 클라우드 컴퓨팅 사이에 컴퓨팅, 저장소, 네트워킹 서비스를 제공하는 플랫폼이다 [12-13]. 클라우드 컴퓨팅은 각 개인이나 기업이 별도의 시스템과 소프트웨어를 구입하지 않고, 클라우드 서비스 업체가 제공하는 컴퓨팅 자원 (예, 상용 소프트웨어, 데이터 저장 공간)을 사용료를 지불하고 필요한 만큼 빌려 쓰는 형태의 서비스를 의미한다. 이로 인하여, 각 개인이나 기업은 시스템을 구축하고 상용 소프트웨어를 구입하는데 드는 초기 비용을 절감 할 수 있으며, 시스템을 유지/보수하는데 드는 인적 자원 절감의 효과도 얻을 수 있다 [14].

그러나 다양한 분야에서 방대한 데이터가 생성·수집됨에 따라 클라우드 컴퓨팅 환경만을 이용한 데이터 처리 기법은 한계점을 드러내었다. 최근 들어 이러한 문제를 해결하기 위한 기술

www.kci.go.kr

로서 포그-클라우드 컴퓨팅 환경이 크게 주목받고 있다. 사용자에서 포그, 클라우드까지 이르는 전체 구조는 그림 3과 같이 구성된다. 첫 번째, 사용자 계층에서는 스마트 기기 및 사용자의 컴퓨팅 기기들은 물리적 거리 차이에 의해 소속된 지역에 배당된 포그 서버와 무선 인터넷 환경으로 커뮤니케이션을 수행한다. 두 번째, 포그 계층에서는 최종 사용자로부터 전송된 데이터의 저장소, 컴퓨팅 등 역할을 수행하며 이처럼 분산된 포그 서버는 사물 인터넷 애플리케이션의 새로운 적용을 수월케 하여 이동성을 보장 한다 [15]. 세 번째, 클라우드 계층 또한 각 포그 서버로부터 수집된 데이터들에 대한 처리, 저장 및 전송을 수행한다. 현재 포그-클라우드 컴퓨팅은 실시간 데이터 분석, 스마트 전력망 구축 등 다양한 분야에서 활용되고 있다.

본 논문에서는 포그-클라우드 컴퓨팅 환경에서 개인의 건강 데이터 수집하고자 한다. 이를 통해 폭발적으로 증가하는 데이터 크기에 의한 저장소 부하 및 네트워크 환경 문제를 해결하고자 한다.

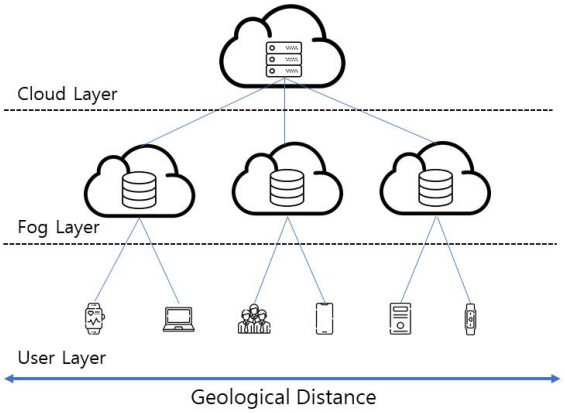


Fig. 3. Fog and Cloud Computing Architecture

III. e-Health Data Collection

본 장에서는 본 논문에서 제안하는 프라이버시를 보호하면서 스마트워치 사용자로부터 민감한 건강 데이터를 수집하기 위한 기법을 설명한다. 그림 4에서처럼 스마트워치에서 수집되는 데이터를 스마트워치 사용자가 지역 차분 프라이버시를 만족하도록 직접 변조를 한 후, 포그 서버로 전송을 한다. 포그 서버에서는 변조된 데이터를 수집하고 중간 통계 결과를 생성 후, 클라우드 서버로 전송한다. 다음 클라우드 서버는 수집된 중간 통계 결과를 집계하여 최종 통계 결과를 도출한다. 그 과정은 그림 4에서처럼 스마트워치 사용자 계층, 포그 계층, 클라우드 계층으로 나뉘며, 상세한 내용은 다음과 같다.

- 스마트워치 사용자 계층: 먼저 스마트워치 사용자의 기기에 내장된 센서를 사용하여 사용자의 건강 데이터를 수집하여, 파일에 저장한다. 24시간 주기로 파일에 저장된 건강 데이터를 이용하여 건강 데이터에 대한 통계적 데이터(예, 히스토그램)을 생성한 후, 지역 차분 프라이버시를 만족하도록 통계적 데이터를 변조한 후, 변조된 통계적 데이터를 포그 서버에 전송한다.

- 포그 계층: 각 포그 서버는 지역별로 분류된 스마트워치 사용자로부터 변조된 히스토그램을 수집 후, 중간 집계 결과를 구한 후, 이를 클라우드 서버로 전송한다.
- 클라우드 계층: 클라우드 서버는 모든 포그 서버들로부터 전송 받은 중간 집계 결과들을 수집한 후, 이를 이용하여 최종 집계 결과를 생성한다.

본 논문에서는 스마트워치 사용자로부터 건강 데이터를 수집 시, 히스토그램 형태의 데이터 수집을 가정하여, 제안하는 기법을 설명한다. 그러나 본 논문에서 제안하는 기법은 히스토그램 이외의 다양한 통계적 데이터 수집에도 동일하게 적용가능하다.

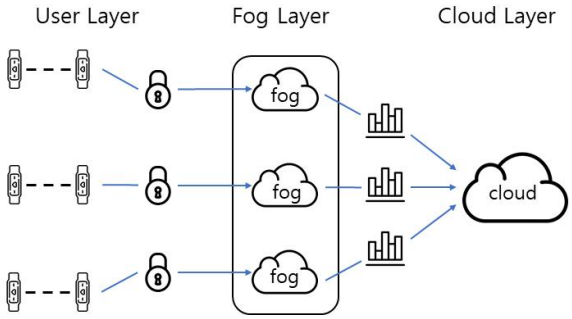


Fig. 4. System Architecture

1. User Layer

본 장에서 제안하는 지역 차분 프라이버시 기반의 건강 데이터 수집 기법에서는 r개의 포그 서버를 사용한다고 가정한다. 또한 각 포그 서버는 s명의 사용자로부터 하루에 k시간 동안 데이터를 수집한다고 가정한다.

스마트워치 사용자의 건강 데이터는 스마트워치에 내장된 각종 센서를 사용하여 수집된 후, 파일에 저장된다. 24시간 주기로 파일에 저장된 건강 데이터를 이용하여 건강 데이터에 대한 히스토그램을 다음과 같이 생성한다. i번째 포그 서버 (즉, fog_i)에 건강 데이터를 전송하는 j번째 스마트 사용자를 u_{ij}라 가정하자. 스마트워치 사용자 u_{ij}의 건강 데이터에 대한 히스토그램 H_{ij}^{Ori}은 다음 식 (2)와 같이 표현된다.

$$H_{ij}^{Ori} = \left\{((t_1, t_2), d_{1,j}^{Ori}), ((t_2, t_3), d_{2,j}^{Ori}), \cdots, ((t_k, t_{k+1}), d_{k,j}^{Ori})\right\} \quad (2)$$

이때, $((t_c, t_{c+1}), d_{c,j}^{Ori})$는 히스토그램의 c-번째 구간을 나타내며, 이는 t_c와 t_{c+1} 시간 사이에 수집된 건강 데이터에 대한 평균값이 $d_{c,j}^{Ori}$임을 의미한다.

히스토그램 생성 후, 차분 프라이버시를 만족하도록 스마트워치 기기 내에서 데이터 변조가 다음과 같이 진행된다. 라플라

www.kci.go.kr

스 메커니즘을 통해 생성된 잡음을 이용하여, 각 구간별 평균값에 잡음을 추가한다. 추가되는 잡음은 스케일(scale)이 $\frac{\triangle s}{\epsilon / k}$이고, 평균이 0인 라플라스 분포에서 추출한 임의의 실수 값이며, 본 논문에서는 이를 라플라스 노이즈라 한다. 수식 (3)에서 $\triangle s$는 차분 프라이버시의 전역 민감도로 임의의 레코드 하나가 특정한 통계 함수의 결과에 미칠 수 있는 최대한의 영향력이며, 다음과 같이 정의 된다.

$$\triangle s = d_{\max} - d_{\min} \quad (3)$$

이때, $d_{\max}$와 $d_{\min}$은 히스토그램의 각 구간별 평균값의 최대, 최솟값을 의미한다. ϵ은 프라이버시 비용에 해당한다. 라플라스 노이즈가 추가된 c-번째 구간의 데이터 변조는 다음의 수식 (4)와 같이 정의된다.

$$d_{c,j}^{Est} = d_{c,j}^{Ori} + Lap(\frac{\triangle s}{\epsilon / k}) \quad (4)$$

그러므로 스마트워치 사용자 u_{ij}의 건강 데이터에 대한 변조된 히스토그램 H_{ij}^{Est}은 다음과 같은 식 (5)로 나타낼 수 있다.

$$H_{ij}^{Est} = \{((t_1, t_2), d_{1,j}^{Est}), ((t_2, t_3), d_{2,j}^{Est}), \cdots, ((t_k, t_{k+1}), d_{k,j}^{Est})\} \quad (5)$$

마지막으로 변조된 히스토그램 H_{ij}^{Est}는 i번째 포그 서버 fog_i로 전송된다. 본 논문에서는 스마트워치와 포그 서버 사이에 데이터 전송은 TCP/IP 프로토콜을 이용하여 이루어진다고 가정한다.

2. Fog Layer

각 포그 서버에서는 각각의 스마트워치 사용자로부터 전송받은 히스토그램을 통합하여 중간 집계 결과를 생성한 후, 이를 클라우드 서버에 전송한다. i번째 포그 서버 fog_i에서 s명의 스마트워치 사용자로부터 전송받은 변조된 히스토그램을 $H_{i1}^{Est}, H_{i2}^{Est}, ..., H_{is}^{Est}$라 가정하자. 포그 서버 fog_i에서는 사용자로부터 전송받은 히스토그램을 이용하여 다음의 식 (6)과 같은 중간 집계 결과물을 생성한다.

$$H_{fog}^{i} = \{((t_1, t_2), d_1^i), ((t_2, t_3), d_2^i) \cdots, ((t_k, t_{k+1}), d_k^i)\} \quad (6)$$

이때, c-번째 구간의 집계 값 d_c^i는 다음 수식 (7)과 같다.

$$d_c^i = \frac{1}{s} \times \sum_{x=1}^{s} d_{c,x}^{Est} \quad (7)$$

위 과정의 결과로 각 포그 서버는 s명의 스마트워치 사용자로부터 전송받은 히스토그램 $H_{i1}^{Est}, H_{i2}^{Est}, ..., H_{is}^{Est}$를 이용하여 중간 집계 결과에 해당하는 히스토그램 H_{fog}^{i}를 생성한 후, 해당 포그 서버의 데이터베이스에 저장한다. 또한, 각 포그 서버는 통합된 중간 집계 결과를 중앙에 위치한 클라우드 서버로 전송한다. 이때 포그 서버와 클라우드 서버 사이 데이터 전송 또한 TCP/IP 프로토콜을 이용하여 이루어진다.

3. Cloud Layer

최종적으로 클라우드 서버는 각각의 포그 서버로부터 전송받은 중간 집계 결과에 대하여 다음의 과정을 거쳐 최종 집계 결과를 생성한다. $H_{fog}^{1}, H_{fog}^{2}, ..., H_{fog}^{r}$를 r개의 포그 서버로부터 전송받은 스마트워치 사용자의 건강 데이터에 대한 중간 집계 결과라 가정하자. 이를 이용하여 클라우드 서버에서는 $s \times r$명의 스마트워치 사용자의 건강 데이터에 대한 최종 집계 결과 H_{cloud}^{Est}를 다음의 식 (8)과 같이 생성한다.

$$H_{cloud}^{Est} = \{((t_1, t_2), d_1^{Est}), ((t_2, t_3), d_2^{Est}) \cdots, ((t_k, t_{k+1}), d_k^{Est})\} \quad (8)$$

이때, c-번째 구간의 집계 값 d_c^i는 다음의 수식 (9)와 같다.

$$d_{c,x}^{Est} = \frac{1}{s} \times \sum_{x=1}^{s} d_{c,x}^{Est} \quad (9)$$

즉, 클라우드 서버는 지역별로 나누어진 포그 서버의 중간 집계 결과물을 모두 통합한 최종 집계 결과를 위처럼 생성하고, 이를 데이터 사용자(예, 데이터 분석가)에게 배포한다.

IV. Experimental Evaluation

1. Experimental Setup

본 논문에서 제안된 방법을 검증하기 위해 프라이버시 비용 ε을 독립변수로 두고 원본 데이터와의 평균 절대 편차를 종속변수로 활용하여 프라이버시의 보호 정도를 측정하였다. 사용자는 샤오미의 mi band 2 기기에 내장된 심박 수 센서를 통해 하루 동안 심박 수 데이터를 수집한다. 수집된 심박 수 데이터는 일정 간격으로 측정되어 사용자의 기기에 저장이 된다. 일정 기간 수집된 데이터를 추출하고 일련의 과정을 진행한 후 포그 서버로 보내게 된다. 본 실험에서 사용된 데이터는 Oracle Virtual Box에서 운용되는 3개의 리눅스 가상머신 3대를 포그 서버로 사용하여 저장하였다. 이는 실제 포그 서버를 구성하기 위하여 물리 공간을 차지하는 실제 머신을 두기 어려운 환경을 보완하고자 가상 머신으로 대체하였다. 다음 실제 서버 컴퓨터를 중앙 클라우드 서버로 활용하여 3대의 포그 서버에서 전송된 데이터를 수집하였다. 각 서버는 12시부터 21시 사이에 스마트워치에서 수집된 104명, 118명, 80명의 심박 수를 하루 동안 기록하고 그 평균을 내어 히스토그램의 형태로 만든다. 생성된 히스토그램을 각 3개의 포그 서버로 보내며, 중앙에 있는 클라우드 서버는 302명의 히스토그램을 모두 수집한다. 수집 기간은 2018년 10월부터 2019년 2월까지의 데이터로 충분한 크기로 수집하기에 어려운 한계점을 극복하기 위해 5배 크기로 복제하여 520명,

www.kci.go.kr

590명, 400명, 총 1510명의 데이터로 실험하였다. 아래의 표는 실제 실험에 사용된 서버의 사양을 나타낸다.

Table. 1. Fog and Cloud server specs used in the system

Fog Server	HDD size	10 GB
	Memory size	2 GB
	CPU cores	1
Cloud Server	HDD size	4 TB
	Memory Size	64 GB
	CPU cores	8

본 논문에서 제안하는 시스템에서 통계 결과 변조가 활용도에 어떤 영향을 미치는지 확인하기 위해 본 논문에서는 통계 결과의 원본과 변조본 간의 절대 평균 편차(Mean Absolute Deviation, MAD)를 계산한다. 절대 평균편차의 크기를 측정하여 최종적으로 데이터 사용자에게 주어지는 변조된 통계 결과의 활용도를 확인하고자 다음과 같은 과정을 수행하였다. 해당 분석 과정에서는 위 3장의 포그, 클라우드 계층에서 사용된 원본 히스토그램과 변조된 히스토그램을 각각 H^{Ori}, H^{Est}으로 가정하며 이를 간략하게 표현한 형태는 식 (10-11)과 같다.

$$H^{Ori} = \{d_1^{Ori}, d_2^{Ori}, d_3^{Ori}, \cdots, d_k^{Ori}\} \quad (10)$$

$$H^{Est} = \{d_1^{Est}, d_2^{Est}, d_3^{Est}, \cdots, d_k^{Est}\} \quad (11)$$

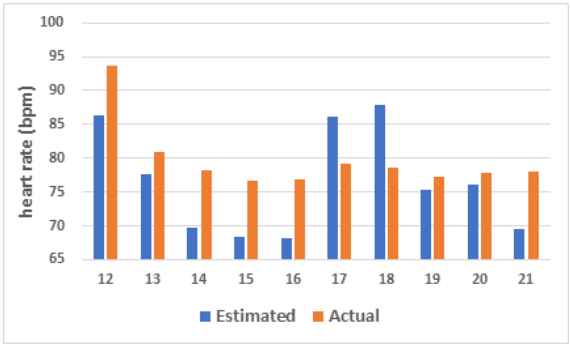
위 통계결과의 원본과 추정된 통계 결과 간의 절대 평균 편차(MAD)를 산출하기 위한 식은 다음과 같다.

$$MAD = \frac{1}{z} \times \sum_{z=1}^{k} |d_z^{Ori} - d_z^{Est}| \quad (12)$$

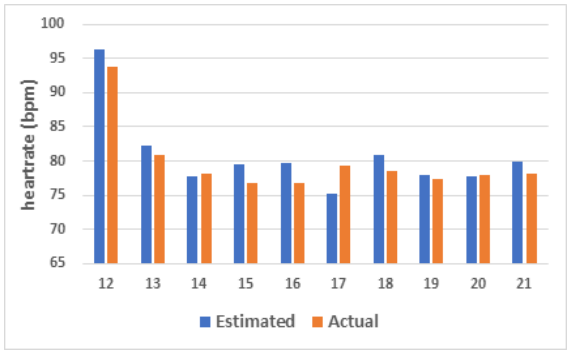
원본 통계 결과가 훼손된 정도를 측정하고 수식 (12)를 통해 도출한 평균 절대 편차는 데이터 사용자에게 주어지는 변조된 통계 결과의 활용도에 직결되는 정보이다. 결과적으로 절대 평균 편차의 크기가 작을수록 통계 결과의 활용도가 높다고 할 수 있다. 또한 데이터 사용자의 권한에 따라 프라이버시 비용을 달리함으로써 활용도를 조절 할 수 있어 본 논문에서 제안한 시스템이 프라이버시를 보존하는 환경에서 스마트워치 사용자의 건강 데이터 수집에 적절함을 보여준다.

2. Results and Discussion

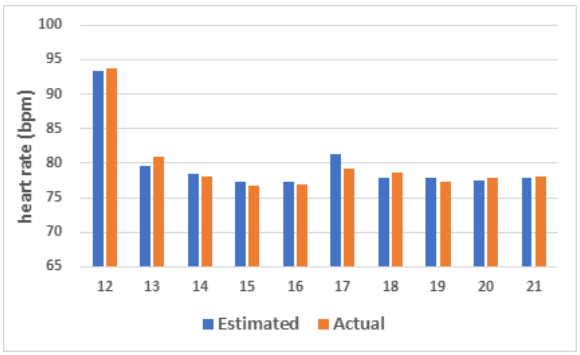
스마트워치 사용자의 데이터가 포그 서버로 보내질 때 해당 논문에서 제안된 방법을 통하여, 하루치에 해당하는 심박 수의 히스토그램이 변조된다. 이때 라플라스 메커니즘을 통해 ε-차분 프라이버시를 만족하는 잡음을 추가한다. 사용자는 해당 방법론을 통해 직접 변조를 하여 지역별로 소속된 포그 서버로 데이터를 전송한다. 이처럼 지역별로 분산된 서버로 전송하는 이유는 본 논문의 실험 당시 조건보다 사용자의 수가 월등히 커질 때 전송의 부하를 줄이기 위함이다. 각 포그 서버에서 프라이버시 비용을 1.0, 2.0, 4.0의 3가지 조건으로 추정한 데이터와 원본 데이터 사이 편차의 크기를 측정하여 변조된 데이터의 활용도를 측정하기 위해 실험을 진행하였으며 그 결과는 다음과 같다.



(a) ε = 1.0



(b) ε = 2.0



(c) ε = 4.0

Fig. 5. Estimated histogram for varying privacy budget in fog server

위 그림 5에서 확인할 수 있듯, 프라이버시 비용 ε을 독립변수로써 변화시키며 실험하였으며 위 결과는 그 예시로 3개의 포그 서버 중 한 곳에서 수집된 결과이다. 첫 번째, 프라이버시 비용이 1.0일 때는 절대 평균 편차가 6.45로 계산되었다. 두 번째로 프라이버시 비용을 2.0으로 책정하였을 때 1.89, 마지막으로 4.0의 프라이버시 비용이 주어졌을 때는 0.69로 계산되었다. 위 수치에서 확인할 수 있듯이 프라이버시 비용이 증가함에 따라 추정된 데이터는 원본 데이터와 유사한 값을 보였으며 절대 평균 편차의 크기가 작아졌다. 이는 프라이버시 비용이 증가함에 따라 원본 데이터에 대한 추정의 정확도가 증가함을 알

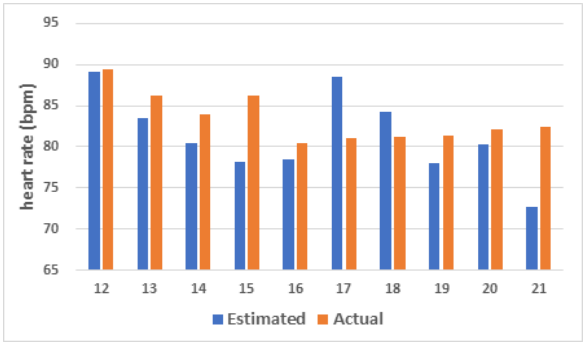
www.kci.go.kr

수 있다. 아래의 표 2는 각 포그 서버에서 여러 조건의 프라이버시 비용를 통해 계산된 3개의 절대 평균 편차를 나타낸다.

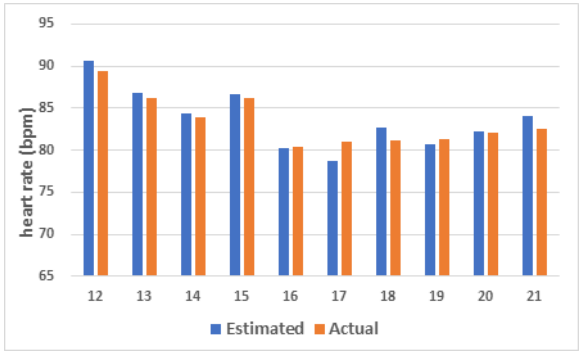
Table 2. Mean Absolute deviation for varying privacy budget in fog server

Privacy Budget	1.0	2.0	4.0
Mean Absolute Deviation	6.45	1.89	0.69

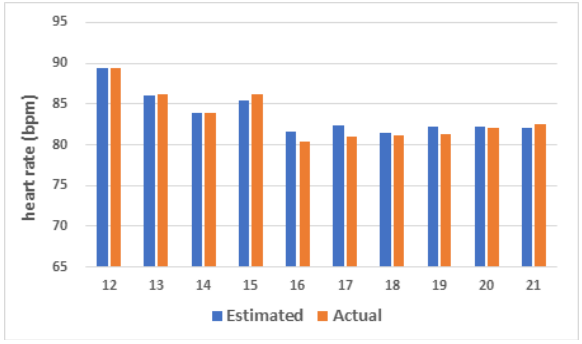
다음은 클라우드 서버가 각 포그 서버에서 전송된 중간 집계 결과를 모두 종합한 최종 집계 결과에 대한 실험 분석 결과이다. 위의 과정과 같이 최종 집계 결과의 원본과 추정된 데이터의 절대 평균 편차를 통해 변조된 데이터의 활용도를 측정하였으며 그 결과는 다음과 같다.



(a) ε = 1.0



(b) ε = 2.0



(c) ε = 4.0

Fig. 6. Estimated histogram for varying privacy budget in cloud server

위 그림 6은 프라이버시 비용 별로 추정된 데이터와 원본 데이터를 비교한 결과이다. 이때 아래의 표 3에서 나타나듯 프라이버시 비용이 1.0일 때 절대 평균 편차는 4.20, 2.0일 때 0.91, 3.0일 때 0.53으로 계산되었다. 앞선 포그 서버에서의 실험보다는 다소 작은 편차를 보였으며, 이를 통해 데이터의 양과 프라이버시 비용이 늘어날수록 원본 데이터와의 편차가 줄어듦을 확인할 수 있다.

Table 3. Mean Absolute deviation for varying privacy budget in cloud server

Privacy Budget	1.0	2.0	4.0
Mean Absolute Deviation	4.20	0.91	0.53

V. Conclusions and Future Work

본 논문에서는 포그-클라우드 컴퓨팅 환경에서 프라이버시를 보존하며 스마트워치 사용자의 건강 데이터를 수집하기 위한 방법을 제안하였다. 본 논문에서 제안한 기법은 지역 차분 프라이버시를 기반으로 사용자의 건강 데이터에 대한 변조가 사용자의 스마트워치 내에서 이루어지고, 변조된 데이터를 포그 서버에 전송함으로써, 사용자의 민감한 건강 데이터가 외부에 노출되는 것을 방지할 수 있다. 또한, 포그 서버에서의 데이터 처리 및 집계 과정을 추가함으로써, 클라우드 서버의 부하를 분산시킬 수 있다. 실 데이터를 통한 실험 결과는 본 논문에서 제안한 기법이 스마트워치 사용자의 프라이버시를 보존하면서, 효과적으로 스마트워치 사용자로부터 건강 데이터를 수집할 수 있음을 증명한다. 향후 연구에서는 본 논문에서 제안한 방법에 걸음 수, 칼로리 등 더욱 다양한 종류와 충분한 양의 데이터를 수집하고, 수집되는 데이터의 형태를 달리하여 더욱 정확도 높고 효율적인 수집 방법을 고안하고자 한다.

REFERENCES

[1] Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., & Mankodiya, K. (2018). Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare. (pp.661). Future Generation Computer Systems, 78, 659-676.

[2] AbuKhousa, E., Mohamed, N., & Al-Jaroodi, J. (2012). e-Health cloud: opportunities and challenges. (pp.623). Future internet, 4(3), 621-645.

[3] Samani, A., Ghenniwa, H. H., & Wahaishi, A. (2015). Privacy

www.kci.go.kr

in Internet of Things: A model and protection framework. (pp.606). Procedia Computer Science, 52, 606-613.
[4] Lu, R., Heung, K., Lashkari, A. H., & Ghorbani, A. A. (2017). A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access, 5, 3302-3312.
[5] Dwork, C. (2011). Differential privacy. Encyclopedia of Cryptography and Security, 338-340.
[6] 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.
[7] Wang, T., Blocki, J., Li, N., & Jha, S. (2017). Locally differentially private protocols for frequency estimation. pp. 729. In 26th {USENIX} Security Symposium ({USENIX} Security 17) (pp. 729-745).
[8] Du, W., & Zhan, Z. (2003, August). Using randomized response techniques for privacy-preserving data mining. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 505-510). ACM.
[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] Erlingsson, Ú., Pihur, V., & Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067). ACM.
[11] Du, W., & Zhan, Z. (2003, August). Using randomized response techniques for privacy-preserving data mining. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 505-510). ACM.
[12] Bonomi, F., Milito, R., Natarajan, P., & Zhu, J. (2014). Fog computing: A platform for internet of things and analytics. In Big data and internet of things: A roadmap for smart environments (pp. 169-186). Springer, Cham.
[13] Stojmenovic, I., & Wen, S. (2014, September). The fog computing paradigm: Scenarios and security issues. In 2014 Federated Conference on Computer Science and Information Systems (pp. 1-8). IEEE.
[14] Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012, August). Fog computing and its role in the internet of things. In Proceedings of the first edition of the MCC workshop on Mobile cloud computing (pp. 13-16). ACM.
[15] Masip-Bruin, X., Marín-Tordera, E., Alonso, A., & Garcia, J. (2016, June). Fog-to-cloud computing (F2C): The key technology enabler for dependable e-health services deployment. In 2016 Mediterranean ad hoc networking workshop (Med-Hoc-Net) (pp. 1-5). IEEE.

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 big data in distributed system. Such as Cloud-Computing, Fog-Computing, and system like Hadoop. And he is studying distributed databases.

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

www.kci.go.kr