Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.5, Oct. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.5.1187

IoT환경에서 프라이버시를 보장하는 의료데이터 이상치 탐색 기법*

이 보 영,† 최 원 석, 이 동 훈‡
고려대학교 정보보호대학원

Privacy-Preserving Outlier Detection in Healthcare Services*

Bo Young Lee,† Wonsuk Choi, Dong Hoon Lee‡
Graduate School of Information Security, Korea University

요 약

최근 다양한 기능을 가진 센서가 개발됨에 따라 여러 종류의 데이터를 간편하게 측정할 수 있게 되었다. 특히, 센서들이 인터넷에 연결되는 사물인터넷(Internet of Things: IoT)환경과 헬스 케어 서비스가 결합하면서 원격에서 심박수, 혈중 산소 농도, 체온, 혈압 등의 사용자 데이터를 수집하는 어플리케이션이 등장하고 있다. 사용자의 유전정보를 이용하여 이상형을 찾거나 환자의 질병유무를 알려주는 어플리케이션 등이 대표적이 예이다. 이 때에 수집되는 사용자 데이터는 사용자의 프라이버시와 매우 밀접하기 때문에 이러한 정보는 반드시 보호되어야 한다. 즉, 사용자의 프라이버시를 보장하면서 서비스제공자는 적절한 서비스를 제공하여야 한다. 본 논문에서는 PhysioNet에서 제공하는 생체정보를 활용하여 헬스 케어 서비스를 제공하는 환경에서 프라이버시를 보장하며 서비스 제공자가 서비스를 제공할 수 있는 있는 기법을 제안한다.

ABSTRACT

Recently, as high-quality sensors are being developed, it is available to conveniently measure any kind of data. Healthcare services are being combined with Internet of things (IoTs). And applications that use user's data which are remotely measured, such as heart rate, blood oxygen level, temperature are emerging. The typical example is applications that find ideal spouse by using a user's genetic information, or indicate the presence or absence of a disease. Such information is closely related to the user's privacy, so biometric information must be protected. That is, service provider must provide the service while preserving user's privacy. In this paper, we propose a scheme which enables privacy-preserving outlier detection in Healthcare Service.

Keywords: Privacy-preserving Analysis, IoT, Emergency Medical Service, Security

I. 서 론

모든 사물이 지능화되고 네트워크로 연결되는 IoT환경은 생활 속에서 점점 그 영역이 확대되고 있다. 최근에 각종 생체 정보를 수집할 수 있는 센서가 부착된 웨어러블 디바이스를 이용한 다양한 어플리케이션이 개발되고 있다. 이는 의료행위의 공간이 병원에 국한 되는 게 아니라 개인이나 가정에서 직접 관리가 가능해지고 있음을 의미 한다[1]. 예를 들어, 결혼정보회사는 고객의 DNA를 분석하여 최적의 배우자를 찾아주는 서비스를 제공하고 있다[2]. IoT를 이루는 구성요소는 인간, 사물, 서비스로 인간과 사물로부터 수집되는 다양한 정보를 통해 사용자는 서비스를 제공받을 수 있다[3]. 개인과 사물에 부착된

접수일(2015년 8월 20일), 게재확정일(2015년 9월 16일)
* 이 논문은 삼성전자 미래기술육성센터의 지원을 받아 수행된 연구임 (과제번호 SRFC-TB1403-00)
† 주저자, lucy250@korea.ac.kr
‡ 교신저자, donghlee@korea.ac.kr(Corresponding author)

센서로 부터 데이터가 추출되며 이것은 스마트폰과 같은 단말기에 전송된다. 단말기로 모아진 데이터는 인터넷을 통해 특정 서비스공급자에게 실시간으로 전송되고, 서비스 공급자는 이 데이터를 가공하여 사용자에게 서비스를 제공할 것이다. IoT환경에서는 센서를 통해 수집된 데이터가 서비스로 변환되는 과정이 중요시 된다. 현재 많은 공급자들이 생체정보를 이용하여 상용화된 서비스를 제공하고 있으며, 스마트 디바이스의 제조사들은 헬스케어서비스 제공을 위한 플랫폼 구축에 힘쓰고 있다[4,5].

통신사와 지역단체 등 여러 단체가 협력하여 고혈압, 당뇨병 등 만성질환자를 위한 서비스를 제공하고 있다. 혈당 측정기에 검사를 위한 피를 떨어트리면 스마트폰에 혈당수치가 표시되고, 정기적으로 서버에 전송된다. 데이터가 축적, 분석되면서 사용자의 주치의가 측정결과를 실시간으로 확인해 건강 상담에 활용할 수 있다. 또한 병원에서는 환자를 원격으로 모니터링 할 수 있는 서비스를 개발 중에 있다. 활동량, 심박 수, 몸무게, 혈압, 혈당 등 여러 가지 개인건강정보(Personal Health Information, PHI)를 수집하여 사용자의 건강 상태에 위험이 감지될 때 의료진에게 알릴 수 있다[4]. 웨어러블 디바이스의 시장을 이끌고 있는 기업에서는 이와 같은 서비스를 제공하기 위한 헬스 케어 플랫폼을 발표하였고, 새로운 디지털 헬스 생태계를 구축하려는 움직임을 보이고 있다[5].

이와 함께 사용자의 의료 데이터를 보호하기 위한 법률들이 재정되고 있다. 미국은 의료정보 보안법(Health Insurance Portability and Accountability Act, HIPAA)을 통해 건강 정보를 보호하고 있으며, 국내에는 원격 의료와 관련된 법안을 의료법 제 30조의2에서 다루고 있다. 국내 현행 의료법에 의하면 전자의무기록의 경우에는 규정된 시설을 두어야 하고, 이를 관리·보존토록 규정하는 등 전자의무기록에 대해 제약사항을 두어 개인의료 기록을 보호하고 있다[6].이처럼 국내/외에서 의료법을 통해 사용자의 의료 기록에 대하여 보호하고 있으며 센서를 통해 추출되는 생체정보 역시 의료기록과 마찬가지로 사용자의 질병유무, 생활 패턴과 같이 프라이버시와 관련된 정보가 포함되기 때문에 기술적인 보호조치가 필요하다. 헬스 케어뿐만 아니라 다양한 분야에서도 서비스 공급자의 권한으로 데이터를 복호화하기 때문에 사고들이 빈번히 일어나고 있으며 이러한 정보 유출 사고들은 서비스 공급자에게 많은 권한이 주어지면서 사용자의 데이터를 모두 복호화 할 수 있기 때문에 문제가 되고 있다[12].

본 논문에서는 Emergency Medical Services(EMS) 환경에서 활용되어질 수 있는 프라이버시 보장 이상치 탐색 기법을 설계하였다. EMS 서비스는 환자가 응급 상황으로 판단될 때 보호자 혹은 구급대원에게 해당 상황을 알려줌으로써 적절한 대응을 할 수 있도록 하는 서비스를 말한다. IoT환경에서는 바이오센서를 이용하여 실시간으로 환자의 건강 상태를 모니터링하고 이상치를 탐색하여 병원 밖에서도 위급상황에 대비할 수 있는 서비스를 제공할 것이다. 본 논문은 바이오센서로부터 수집된 정보를 이용하여 병원 밖에서 EMS 서비스를 제공하는 환경을 가정한다. 이때에 모니터링을 위하여 실시간으로 전송되는 사용자의 생체정보는 생활패턴 혹은 사용자의 질병 유무 등 개인적인 정보를 포함하고 있기 때문에 프라이버시의 보호가 필요하다. 따라서 EMS 서비스를 위해 수집되는 생체정보는 제 3자에게 노출되어선 안되며, 서비스 공급자라 할지라도 사용자의 생체정보를 노출시키지 않고 적절한 서비스를 제공할 수 있어야 한다.

2장에서는 관련연구를 설명하고, 시스템에 필요한 배경지식은 3장에 요약되어있다. 4장은 본 논문이 제안하는 프라이버시를 보장하는 EMS 서비스를 위한 시스템모델을 제시하고 있으며, 5장은 EMS 환경에서 이상치를 탐지하는 기법설계에 대하여 자세히 다룬다. 6장은 시스템을 평가하고 있으며, 7장은 실험에 대하여 분석하였다.

II. 관련 연구

본 논문에서 병원 밖에서 EMS 서비스를 제공받을 수 있는 환경을 제시하고 있으며, 사용자의 프라이버시를 보장하기 위하여 생체데이터를 복호화 하지 않고 특정 연산 결과를 얻는 시스템이 필요하다. 복호화 하지 않고 암호화된 상태로 특정 연산이 가능한 Peter등의 기법[10]과 생체정보를 통하여 환자의 상태를 체크하는 Osman 등의 기법[11]을 설명한다.

2.1 Peter 등의 연구[10]

Peter 등은 여러 명의 사용자가 각자의 개인키/공개키 쌍을 이용하여 연산을 요청할 수 있는 아웃소싱 연산 기법을 설계하였다. 기존의 아웃소싱 논문들

은 연산 시 사용자와 서버 간 인터랙션이 필요하거나 사용자가 같은 키로 암호화 하여 전송하였을 때에만 서버가 연산을 수행 할 수 있는 한계가 있다. Peter 등이 제안한 아웃소싱 기법은 사용자가 데이터를 서버로 전송하면 서버는 사용자와 인터랙션 없이 연산이 가능하다. 또한 모든 사용자들이 같은 키가 아닌 각자의 개인키/공개키 쌍을 사용하여 전송할 수 있기 때문에 기존의 아웃소싱 기법보다 더 IoT 환경에 적합하다. 이 기법은 2개의 semi-trusted한 서버를 두고 덧셈연산이 가능한 동형암호의 성질을 이용하여 연산을 수행한다. 첫 번째 서버는 동형암호 성질을 이용하여 사용자의 암호문에 노이즈를 첨가하고, 두 번째 서버는 노이즈가 첨가된 값에 대하여 연산을 수행한다. 그 후 다시 첫 번째 서버에 전송하면, 첫 번째 서버는 동형암호의 성질을 이용하여 복호화 하지 않고 노이즈를 제거하고, 이렇게 수행한 연산결과는 다시 각자의 공개키로 암호화 되어 사용자에게 전송된다. 각각의 서버는 비밀 값을 나눠가짐으로서 사용자의 데이터를 복호화 할 수 없기 때문에 안전하게 연산 수행이 가능하다.

2.2 Osman 등의 연구[11]

Osman 등은 생체정보를 이용하여 환자의 상태를 분석 할 수 있는 모니터링 기법을 제안하였다. 환자의 상태를 측정할 수 있는 센서가 사용자의 몸에 부착되고 스마트폰에 수집된 사용자 데이터는 인터넷을 통해 실시간으로 의료진에게 전달된다. 의료진은 사용자 데이터를 이용하여 환자의 상태를 지속적으로 모니터링하고 이상치가 탐색되었을 경우 위급상황임을 알리는 서비스를 제공한다.환자는 주기를 두고 혈압, 심박수, 맥박수, 포화산소농도, 호흡수 총 5가지의 생체정보가 수집되며 이 정보는 서버로 전송된다. 안정적인 상태에서 수집된 사용자 데이터는 학습단계를 거치고 사용자 데이터가 전송되었을 때 거리를 측정하기 위한 대푯값으로 사용 된다. 전송된 데이터와 대푯값 사이의 거리를 통해 이상치 라고 판단했을 때에 응급상황을 알리게 된다.

Osman등은 두 단계의 이상치 탐색을 통하여 시스템의 정확도를 높였다. 1단계는 환자의 상태를 MD로 측정한다. 1단계에서 이상데이터라고 판단 시 2단계는 커널 밀도 함수(Kernel Density Estimation) 를 통하여 한 번 더 이상치 탐색을 시도한다. 각각의 속성들 마다 커널 밀도 함수를 구하고 k개 이상의 속성이 이상치로 판단되었을 때, 센서의 오작동이 아닌 환자의 응급상황이라고 판단하게 된다. 실제 한 사람에게서 수집되는 생체데이터를 이용하여 실험을 진행하였고 알람이 잘못 울릴 확률인 False Positive Rate(FPR)은 5.5%이하로 높은 정확도를 보였다.

III. 배경 지식

Peter 등은 덧셈에 대하여 동형암호의 성질을 만족하는 BCP 프로토콜을 이용하여 시스템을 설계하였으며, Osman등은 Mahalanobis Distance (MD)와 커널밀도함수를 이용하여 이상치를 탐색하였다. 본 논문에서 사용되어지는 동형암호화 방법과 MD에 관하여 설명한다.

3.1 동형암호(Homomorphic Encryption)

Homomorphic Encryption은 평문과 암호문 공간에 대하여 덧셈 혹은 곱셈 연산을 보존하는 암호체계이다. 쉽게 말해 평문에 연산을 하여 암호화 한것과 평문을 암호화하고 연산을 한 것이 같게 되는 암호체계이다. 동형암호를 식으로 표현하면 (1)과 같다.

$$Enc(m_1 * m_2) = Enc(m_1) \diamond Enc(m_2) \qquad (1)$$

Fig. 1은 비대칭키 기반의 동형암호 기법에 대하여 보여주고 있다. 암호화 된 두 값에 대하여 복호화를 하지 않고 메시지를 연산할 수 있으며, 키를 갖고 있는 사람만 그 값을 복호화 할 수 있다. 자신의 공개키 pk로 메시지 m_1, m_2를 각각 암호화 하여 서버에 전송하면, 서버는 암호문에 대하여 복호화 하지 않고 결과 값 $Enc_{pk}(f(m_1, m_2))$를 연산할 수 있다. 서버는 연산만 수행할 수 있으며, 데이터를 복호화

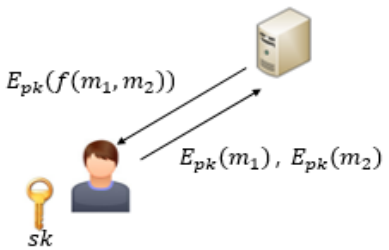


Fig. 1. Homomorphic Encryption using Asymmetric key.

하여 연산결과 $f(m_1, m_2)$를 얻을 수 있는 것은 비밀키를 갖고 있는 사용자만 가능하다[7].

동형암호는 연산 수행능력이 부족한 환경에서 비밀값을 보호하며 서버에 연산을 요청하는 아웃소싱 환경에서 일반적으로 사용된다. 동형암호 방법은 덧셈 혹은 곱셈과 같이 일부 연산만 수행이 가능한 Somewhat Homomorphic Encryption과 모든 연산이 가능한 Fully Homomorphic Encryption (FHE)으로 나뉜다. FHE는 모든 연산이 가능하지만, 많은 연산량을 필요로 하며 속도가 느리거나 특정 횟수 이상으로 연산이 불가능하기 때문에 실제 환경에 적용하는데 있어 제약사항이 존재 한다[8].

3.2 Mahalanobis Distance(MD)

MD는 독립적인 다변수의 데이터에 대하여 이상치를 탐색하고자 할 때에 적합한 거리 계산법이다. 연산 시 두 개의 데이터 사이의 절대적인 거리뿐만 아니라, 변수의 특성을 나타내는 표준편차와 상관계수가 함께 고려되기 때문에 이상치를 탐지하기 적합하다[15]. 두 개의 데이터 사이의 거리를 구하기 위하여 주로 Euclidian Distance(ED)를 사용하지만 데이터가 변동이 크다면 정상적인 데이터도 이상치로 판단할 수 있다. 반대로, MD는 다변량의 데이터에 대하여 분산을 고려하여 거리를 구하는 방법이다. 거리를 계산할 때에 분산을 함께 고려하기 때문에 데이터의 변동 폭이 큰 데이터에 대하여 클러스터링 하거나 이상치를 탐색할 때 유용하게 사용된다. Fig. 2는 ED와 MD를 비교하는 그림으로, 벡터변수 x_1, x_2에 대하여 중간값 μ와 같은 거리를 갖고 있는 점들에 관하여 나타내고 있다. ED는 같은 거리를 갖고 있는 점이 원으로 나타내어지는 반면에 MD는 변수의 분산을 고려하기 때문에 전체 데이터와 같이 타원 모양으로 표시 된다[9].

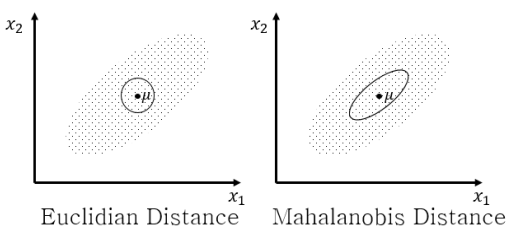


Fig. 2. Comparison of Euclidian Distance and Mahalnobis Distance.

IV. 시스템 모델

이번 장에서는 우리가 고려하고 있는 시스템 모델에 대하여 논의한다. 제안하는 기법은 1장에서 설명하고 있는 EMS 서비스를 대상으로 하고 있다. EMS를 서비스를 위한 시스템 모델은 크게 2가지 영역으로 구분된다. 사용자 신체에 부착되어 있는 여러 개의 바디센서가 이루고 있는 Body Area Network(BAN) 영역과 서비스 공급자(Service Provider)와 CSP를 포함한 서비스 제공자(SP)영역으로 구분된다. BAN는 사용자의 몸에 부착되거나 삽입되는 디바이스로 구성된 네트워크로, 센서를 통하여 사용자의 몸에 생체정보를 측정 할 수 있으며 기기 간 근거리 통신이 가능하다. SP영역은 사용자의 생체정보를 분석하고 결과 값을 도출하여 사용자에게 맞는 서비스를 공급하는 영역을 말한다. Fig. 3.은 IoT 환경에서 EMS 서비스를 위한 시스템 모델을 보여주고 있다. 본 논문이 제안하는 시스템은 4가지의 개체 Body Sensor, Gateway, Service Provider(서비스 공급자), CSP로 이루어진다. Table 1는 각각의 개체에 대하여 설명한다.

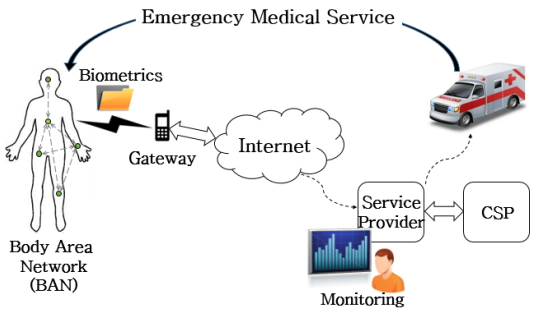


Fig. 3. Emergency Medical Service in IoT Environment.

4.1 요구 사항

이번 절에서는 앞서 설명한 시스템 모델에서 필요한 요구 사항을 설명 한다. EMS서비스는 서비스 공급자가 사용자의 생체정보를 분석하기 때문에 사용자의 민감한 정보가 노출될 수 있는 문제가 있다. 본 논문이 제안하는 기법의 요구사항은 크게 효율성(Efficiency) 측면과 보안(Security) 측면으로 나뉜다. 모니터링 환경에서는 생체 데이터를 분석하여 질병의 유무를 알 수 있다. 뿐만 아니라 실시간으로

Table 1. Entities of our system

Body Sensor	User's biometric informa- tion is extracted in real time by a body sensor. Body sensors are attached to the body while forming a Body Area Network (BAN). And it is connected to the Internet via gateway.
Gate Way	Gateway refers to a device such as a mobile phone to collect and send a user's biometric information over the Internet.
Service Provider (SP)	Service Provider extracts meaningful values by analy- zing the data collected from the user. Using the results, it is possible to provide a service to the user.
Crypto Service Provider (CSP)	Crypto Service Provider (CSP) works for security. When the service provider analyzes health related data, CSP restrict service provider not to decrypt bio-information. Both service provider and CSP compute distance using 4 operations (+, −, ×, ÷).

사용자 데이터가 전송되기 때문에 이 정보를 이용하여 수면 시간, 활동 정보 등 사용자의 생활 패턴을 알 수 있으며, 합계, 평균, 분산 등 중간 연산 결과는 사용자의 개인적인 특성을 반영하므로 노출 시 문제가 된다. 따라서 이러한 정보는 제 3자로부터 보호되어야 하며, 서비스 공급자는 프라이버시를 보장하면서 서비스를 제공하는 보안을 고려한 시스템이 요구된다. 또한 단말기 환경에서는 효율성이 고려되어져야 한다. 전력이 지속적으로 공급되기 힘든 단말기 환경에서 사용자와 서비스 공급자가 지속적으로 인터랙션(Interaction)을 하게 되면 배터리의 소모량이 많아 현실적으로 사용이 어렵다. 따라서 단말기 환경을 고려한 효율성이 고려된 환경이 필요하며 이는 즉, 사용자는 데이터를 전송하면 더 이상 연산에 관여하지 않는 Non-Interactiveness환경을 의미한다. 다음은 본 논문이 제시하는 EMS환경에서 필요한 요구사항에 대하여 설명한다.

- **Non-Interactiveness**: BAN 영역의 바디 센서는 사용자 데이터를 측정하여 이를 서비스 제공자에게 전달한다. 서비스 제공자는 사용자에게 서비스를 제공하기 위하여 특정 연산을 수행하는데, 연산 수행과정에서 바디센서가 참여하지 않는 것을 Non-Interactiveness라고 한다. IoT환경에서는 단말기의 배터리 문제와 사용자의 온/오프라인 상태를 고려하여 데이터를 한번 보내면 연산에 더 이상 참여할 필요가 없는 사용자와 서비스 공급자 간에 통신을 최소화해야할 필요가 있다.

- **기밀성**: 데이터가 전송되는 모든 부분에서 공격자 혹은 서비스 공급자가 사용자의 정보를 도청할 수 있는 위험이 존재한다. 생체정보를 전송하는 각 구간은 암호화되어야하며 서비스 공급자와 CSP 그리고 외부 공격자에게 기밀성이 지켜져야 한다. 서비스공급자는 센서로 부터 수집되는 데이터뿐만 아니라 개인의 생체 특성을 알 수 있는 중간 연산 값(평균, 분산)도 알지 못하며, 이상치 여부만을 판단 할 수 있다. CSP는 서비스공급자가 프라이버시를 보장하며 연산을 수행할 수 있도록 도와주는 개체로, 연산 수행 시 사용자의 생체데이터에 대한 정보를 알아낼 수 없어야한다. 또한, 외부 공격자는 전체 시스템을 도청하고 나아가 능동적 공격이 가능하므로, 데이터가 전송되는 전체 구간에서 기밀성이 지켜져야 한다.

4.2 공격자 모델

공격자는 서비스 공급자와 CSP 그리고 외부 공격자 세 가지 유형으로 나눌 수 있다. 모든 공격자의 목표는 사용자의 프라이버시 정보 즉, 생체정보를 알아내는데 있으며 각 공격자의 능력은 다음과 같다.

- **서비스 공급자**: 사용자는 데이터를 암호화 하여 서비스 공급자에게 전송하기 때문에 서비스 공급자는 사용자의 public key(pk)로 암호화된 모든 생체데이터를 수집할 수 있다. 또한 서비스 공급자는 특정 연산결과를 계산할 때, 암호화된 중간 연산결과를 알 수 있다.
- **CSP**: CSP는 Master Key(MK)를 이용하여 사용자의 데이터를 복호화 할 수 있는 능력을 갖고 있다. 서비스 공급자가 연산을 요청하면, CSP는 노이즈가 추가된 암호문을 전송 받아 MK를 이용하여 복호화하고 정해진 프로토콜대로 연산을 수행한다.

· **외부 공격자:** 사용자-CSP, 사용자-서비스공급자, 서비스 공급자-CSP사이에서 주고받는 데이터를 도청할 수 있다.

4.3 가정

서비스 공급자와 CSP는 수동적 공격자(Passive Attacker), 외부 공격자는 능동적 공격자(Active Attacker)를 가정한다. 서비스 공급자와 CSP는 Semi-Trusted한 객체로, 정해진 프로토콜대로 연산을 수행하지만 각각의 데이터에 대해서는 복호화하거나 의미 있는 정보를 알아내고자 하는 시도는 할 수 있다. 또한 둘은 수동적 공격자이기 때문에 두 서버간 공모공격(Collude Attack)을 하지 않는다고 가정한다. 외부 공격자는 능동적 공격자로 데이터가 전송되는 전 구간에 걸쳐 도청이 가능하며, 이를 이용하여 사용자의 생체정보를 알아내고자 한다.

V. 제안 기법

제안하는 기법에 대한 이해를 위하여, 데이터 흐름과, 사용되는 표기법에 대하여 설명하겠다.

Fig. 4는 전체적인 데이터 흐름을 도식화하여 나타내고 있다. CSP는 Public Parameter(PP)와 Master Key(MK)를 생성 후 사용자에게 전송한다. 사용자는 그것을 이용하여 자신의 public key(pk)와 secret key(sk)를 만들고 데이터를 암호화 하여 서비스 공급자에게 전송한다. 덧셈만 가능한 동형암호 프로토콜이기 때문에 덧셈이외의 연산에 대해서는 서비스 공급자와 CSP가 협력하여 연산을 수행한다. 최종 결과 값은 서비스 공급자에게 전송되고 이 결과 값을 통해 서비스를 제공한다. 제안하는

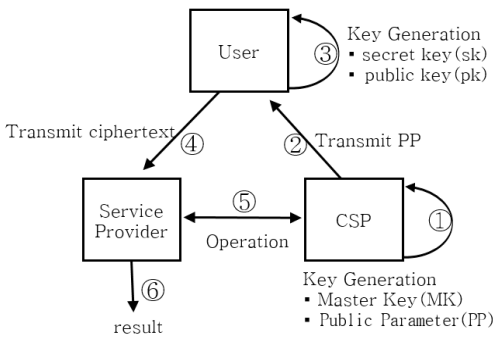


Fig. 4. Data flow

Table 2. Notation

PP	Public Parameter
MK	Master Key
Z_N	Plaintext space N
Z_{N^2}	Plaintext space N^2
p', q'	Distinct prime
p, q	Distinct prime p=2p'+1, q=2q'+1
N	N=p·q
k ($k \in [1, N-1]$)	Constant such that $g^{p'q'} \bmod N^2 = 1 + kN$
g ($g \in Z^*_{N^2}$)	Random element g of order pp'qq'
sk	User's secret key
pk	User's public key

기법에서 사용되는 주요 표기법은 Table 2와 같다.

제안하는 기법은 i) 키 생성, ii) 데이터 전송, iii)데이터 모니터링으로 구성되어 있다. 그리고 데이터 모니터링 과정에서는 서비스 공급자와 CSP 간 연산이 필요하다. 데이터 모니터링을 위해 두 서버간 연산을 수행하는 부분은 5.3에서 자세히 설명하겠다. 5.1에서는 CSP서버의 키 생성과 사용자의 키 생성 과정을 설명하고 있으며 , 5.2장에서는 사용자가 데이터를 암호화 하여 서비스 공급자에게 전송하는 과정을 설명한다. 5.4장은 서비스 공급자가 이상치를 탐색하기 위하여 데이터를 분석하는 과정을 설명하고 있다. 이때에 사칙연산은 서비스 공급자와 CSP가 협력하여 진행하며 이 프로토콜은 5.3장 서버간 연산에서 설명하고 있다.

5.1 키 생성

5.1.1 CSP서버 마스터키 생성

Algorithm 1은 CSP 서버가 Master Key (MK)와 Public Parameter (PP)을 만드는 과정을 나타내고 있다. PP는 사용자가 키를 만들기 위해 사용되는 공개된 값으로, 서비스를 이용하는 사용자에게 전송된다. MK는 CSP 서버의 트랩도어를 위

Algorithm 1. CSP key generation

Set prime p', q', p, q

$p = 2p' + 1$
$q = 2q' + 1$
$N = p \times q$
Set $PP \leftarrow (N, k, g)$
Set $MK \leftarrow (p', q')$

Output the PP, MK

한 마스터키로, CSP만 보유하게 되며 이를 이용하여 사용자의 데이터를 복호화 할 수 있다.

5.1.2 사용자 키 생성

Algorithm 2. User key generation

Generate User Key using PP

Pick random $a \in Z_{N^2}$
$h = g^a \mod N^2$
$sk \leftarrow a$
$pk \leftarrow h$

Output the sk, pk

Algorithm 2에서 사용자는 CSP로부터 받은 PP를 이용하여 공개키 pk와 개인키 sk를 생성한다. 사용자는 개인키 sk를 랜덤 값으로 선택하고 PP와 sk를 이용하여 공개키 pk를 생성한다.

5.2 데이터 전송

Algorithm 3은 사용자의 공개키 pk와 CSP로부터 받은 PP를 이용하여 데이터를 암호화 한다.

사용자는 랜덤 값 r을 선택하고, 메시지 m에 대하여 (A, B)의 형태로 암호화 한다.

센서에서 매 시간 마다 생성되는 데이터는 암호화 과정을 거쳐 사용자의 게이트웨이 노드에 전송되며, 게이트웨이 노드에서는 p개의 속성 값을 갖는 데이터 셋$(E_{pk}(A_{i1}), E_{pk}(A_{i2}), ..., E_{pk}(A_{ip}))$의 형식으로 묶어 서비스 공급자에게 전송한다.

Algorithm 3. Data Encryption

User Encrypt message m using PP, pk

Pick random $r \in Z_{N^2}$
$A = g^r$
$B = h^r(1 + mN) \mod N^2$

Output the ciphertext (A, B)

5.3 서버간 연산

본 장은 5.4장의 데이터 모니터링을 위하여 사용되는 기본 연산인 $(+, -, \times, \div)$을 수행하는 방법에 대하여 설명한다. 서비스공급자와 마스터키를 갖고 있는 CSP 두 서버는 상호 통신 (인터랙션)을 통하여 연산을 수행한다. Fig. 5는 서버 간 연산과정을 보여주고 있다. 서비스공급자는 동형암호 성질을 이용하여 암호문에 노이즈를 첨가하고 CSP에게 전송한다. CSP는 MK를 이용하여 노이즈가 첨가된 암호문을 복호화하고 뺄셈, 곱셈, 나눗셈 연산을 수행한다. 여기서, 덧셈 연산은 CSP 없이 서비스제공자 스스로 수행할 수 있다. 연산된 값은 다시 서비스 공급자에게 전송되고 서비스 공급자는 첨가했던 노이즈를 동형암호 성질을 이용하여 제거한다. 5.3.1에서는 서비스 공급자가 노이즈를 추가하는 과정을 설명하고 있으며, 5.3.2에서는 CSP가 서비스 공급자로부터 받은 데이터를 MK를 이용하여 복호화 하는 과정을 설명하고 있다.

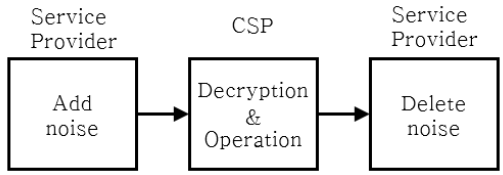


Fig. 5. Operation process between Service Provider and CSP.

5.3.1 노이즈 추가

서비스 공급자는 MK를 이용하여 데이터를 복호화 할 수 있는 능력을 지닌 CSP가 사용자의 암호화

된 데이터를 알 수 없도록 노이즈를 사용한다. 식(1)과 같이 서비스 공급자는 동형암호 성질을 이용하여 데이터를 복호화 하지 않고 암호문에 노이즈 $-\sigma$를 추가할 수 있다.

$$(C,D) \leftarrow Add((A,B), Enc_{pk}(-\sigma)) \quad (1)$$

5.3.2 복호화

Algorithm 4. Data Decryption of CSP

Master Server Decrypt (A, B)
using PP, pk, MK

$$a \bmod N = \frac{h^{p'q'} - 1 \bmod N^2}{N} \cdot k^{-1} \bmod N$$

$$r \bmod N = \frac{A^{p'q'} - 1 \bmod N^2}{N} \cdot k^{-1} \bmod N$$

$$m = \frac{(B/g^{ar})^{p'q'} - 1 \bmod N^2}{N} \cdot (p'q')^{-1} \bmod N$$

Output the message m

MK를 갖고 있는 CSP는 비밀정보 a값과 랜덤하게 선택한 r값을 복구할 수 있으며 이렇게 구한 a와 r을 이용하여 암호화된 메시지를 복호화 할 수 있다. Algorithm 4는 CSP가 암호문 (A,B)를 PP, pk 그리고 MK를 이용하여 메시지 m을 복호화 하는 과정을 설명하고 있다.

5.3.3 $+, -, \times, \div$ 연산

서비스 공급자가 노이즈를 추가하여 CSP에게 전송하면, CSP는 그것을 복호화하고 노이즈가 포함된 채 연산을 수행한다. 여기서 동형암호가 덧셈만 가능하기 때문에 연산을 편리하게 수행하기 위하여 노이즈를 음수로 더해준다.

본 논문에서 사용하는 암호기법은 덧셈연산이 가능한 동형암호 성질을 이용하므로 서비스 공급자 스스로는 암호문 간에 덧셈연산만 가능하다. 덧셈을 제외한 뺄셈, 곱셈, 나눗셈은 서비스공급자 스스로 수행할 수 없기 때문에 CSP와 함께 연산을 수행한다.

연산과정을 설명하기 위한 표기법은 Table 3에서

Table 3. Notations for Data Operation

Add Noise	
(C, D)	$Add((A,B), Enc_{pk}(-\sigma))$
(C', D')	$Add((A', B'), Enc_{pk}(-\sigma'))$
CSP Decryption	
z	$mDec(C, D)$
z'	$mDec(C', D')$

설명되어있으며, 덧셈연산 과정은 아래 식(2)에서 나타내고 있다.

Input Data (2)

$$(A,B) = (g^r \bmod N^2, h^r(1+mN) \bmod N^2)$$
$$(A',B') = (g^{r'} \bmod N^2, h^{r'}(1+m'N) \bmod N^2)$$

Addition

$$(\overline{A}, \overline{B}) = (A \times A', B \times B')$$
$$= (g^{r+r'} \bmod N^2, h^{r+r'}(1+mN)(1+m'N) \bmod N^2)$$

서비스 공급자와 CSP은 함께 프로토콜을 수행하여 뺄셈, 곱셈, 나눗셈(상수) 연산을 할 수 있다. 덧셈은 동형암호 성질을 이용하여 서비스 공급자 스스로 연산 수행이 가능하지만 뺄셈, 곱셈, 나눗셈은 CSP의 도움을 받아 수행이 가능하다. 서비스 공급자는 노이즈를 추가하여 CSP에게 전송하고, CSP는 전송받은 데이터를 복호화 하여 뺄셈, 곱셈, 나눗셈 연산을 수행한다. 나눗셈은 상수 나눗셈 연산을 다루고 있다. 평균, 분산 등 데이터를 분석할 때에는 총 데이터의 개수와 같이 고정된 상수 n에 대한 나눗셈이 필요한 경우가 있다. 따라서 본 논문에서 제시하는 나눗셈은 암호화된 두 개의 데이터 사이의 나눗셈이 아닌 암호화된 하나의 값과 상수 값에 대한

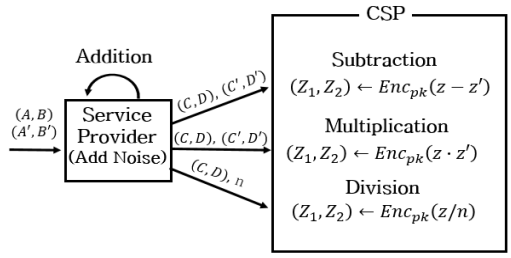


Fig. 6. CSP Operation

연산을 설명한다. Fig. 6은 CSP가 데이터를 복호화 하여 뺄셈, 곱셈, 나눗셈 연산을 수행하는 과정을 나타낸다.

5.3.4 노이즈 제거

CSP는 연산을 수행하고 사용자의 pk로 다시 암호화 하여 서비스 공급자에게 전송한다. 서비스 공급자는 CSP로부터 받은 값에서 동형암호 성질을 이용하여 노이즈를 제거한다. 뺄셈연산은 $-\sigma-(-\sigma')$이 포함되어있으며 이를 제거하기 위하여 서비스 공급자는 $\sigma+(-\sigma')$를 더한다. 곱셈연산은 두 데이터 m, m'에 대하여 노이즈가 포함된 채 연산 $(m-\sigma)(m'-\sigma')$이 수행되기 때문에 서비스 공급자는 $+m\sigma'+m'\sigma-\sigma\sigma'$의 값을 더해 노이즈를 제거하고 곱셈 결과 mm'을 얻을 수 있다. 나눗셈은 σ를 n으로 나눈 σ/n값 을 더해주어 노이즈를 제거한다. 위의 연산은 모두 복호화 하지 않고 동형암호 성질을 이용하여 암호화된 상태에서 수행된다.

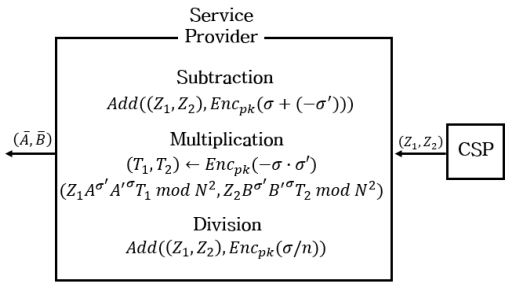


Fig. 7. Delete Noise of Service Provider

5.4 데이터 모니터링

5.4.1 학습 단계

학습 단계는 이상치를 탐지할 수 있는 대표 값을 구하기 위한 단계이다. 사용자가 안정적인 상태일 때에 측정된 생체 정보를 이용하여 학습을 수행한다. 5.4.2에서는 이를 이용하여 실시간으로 전송된 사용자의 데이터와 대푯값 사이의 차이를 통해 MD을 구한다. 일정 간격으로 데이터 $X=(A_1,A_2,..,A_p)$가 수집되며 이것을 1부터 n까지의 시계열 개념을 포함하여 나타내면 다음과 같이 행렬로 나타낼 수 있다. MD를 구하기 위한 대푯값으로 평균 벡터와 공분산 행렬이 필요하다. MD를 계산하기 위하여 필요한 기본 연산은 덧셈, 뺄셈, 곱셈, 나눗셈(상수)으로 충분하다. 학습단계에서 대푯값을 구하기 위하여 5.3의 연산과정을 수행한다. 식(3)은 사용자로부터 전송받은 생체 데이터를 행렬로 표현하고 있으며, 간략화하기 위해 암호화 기호를 생략하였다.

$$X=\begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array}\begin{array}{c} \begin{array}{cccc} A_1 & A_2 & \cdots & A_p \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \end{array} \tag{3}$$

1~n행은 시계열 단위를 나타내며, 1~p 열은 속성을 나타낸다. 이렇게 모아진 사용자 데이터는 서비스 공급자에게 전송되며 CSP와 인터랙션을 통해 다음과 같이 평균 벡터와 공분산을 계산한다.

a) 평균 벡터

식 (4)는 각 속성의 n개의 데이터의 평균값을 의미한다. μ_k는 k번째 속성의 평균값으로, 해당 속성을 대표하는 값이다. p개의 속성에 대하여 각 평균값은 $(1\times p)$의 행렬 $\mu=(\mu_1,\mu_2,...,\mu_p)$로 나타낼 수 있다.

$$\mu_k=\frac{1}{n}\sum_{i=1}^{n}x_{ik} \tag{4}$$

b) 공분산 행렬

다음으로 평균 벡터를 이용하여 공분산 행렬을 계산한다. 공분산은 여러 개의 속성 값 사이의 상관정도(correlation)를 의미하기 때문에, 변동폭이 큰 데이터인 생체 데이터에서 이상치를 탐지하는데 적합하다. 식(5)는 공분산 행렬에서 각각의 공분산 값을 구하는 방식을 설명하고 있으며 전체 공분산 행렬은 식(6)과 같다.

$$\Sigma_{ij}=\frac{1}{n-1}\sum_{k=1}^{n}(x_{ki}-\mu_i)(x_{kj}-\mu_j) \tag{5}$$

$$\Sigma = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1p} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \Sigma_{p1} & \Sigma_{p1} & \cdots & \Sigma_{pp} \end{vmatrix} \quad (6)$$

5.4.2 이상치 판단

$(1\times p)$사이즈의 평균벡터와 $(p\times p)$사이즈의 공분산 행렬의 연산을 통해 MD_i를 구할 수 있다. 이상치 판단 역시 5.3에서 설명한 서비스공급자와 CSP간 연산과정을 따른다. 식 (7)의 연산을 통해 학습 후 새롭게 전송 된 사용자데이터 $X_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$에 대하여 MD값을 구한다.

$$MD_i^2 = (X_i - \mu)^T \Sigma^{-1} (X_i - \mu) \quad (7)$$

$$MD_i^2 \geq \chi^2_{p,0.975} \quad (8)$$

$\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ 이며, 연산 수행 후 CSP에게 복호화를 요청하여, MD가 임계값을 넘으면 이상치로 탐지하게 된다. 예를 들어 MD가 97.5%이상의 차이를 보이는 데이터에 관하여 이상치로 탐지한다면 식(8)과 같이 나타낼 수 있다.

VI. 평 가

6.1 실험 환경

본 논문에서는 제안한 기법을 평가하기 위해 PhysioNet에서 제공하는 생체정보를 Data set으로 사용하였다[13]. PhysioNet은 연구를 목적으로 여러 사람의 생체 정보를 제공하고 있으며, 본 논문에서는 한 사람에게서 동 시간에 측정한 4가지 종류의 생체정보인 MIMICⅡ를 사용하였다. 4가지 종류의 생체정보는 HR, PULSE, RESP, SpO2이며, 이를 이용하여 프라이버시 보장 EMS 시스템을 구현하고 평가하였다. 평가 방법으로는 키 사이즈 |N|에 따른 수행시간(s)과 데이터를 학습 시 필요한 Window size에 따른 수행시간을 평가한다.

Table 4. Experiment Environment

CPU	Intel Core i5-4590 CPU running at 3.30GHz
RAM	16,384MB
Programmning Language	Java jdk1.8.0_51
Data Set	MIMIC Ⅱ: HR, PULSE, RESP, SpO2 (from www.Physionet.org)

6.2 윈도우 사이즈(Window size)

첫 번째 성능평가 방법으로 Window size의 변화에 따라서 전송받은 사용자의 데이터와 학습을 통해 얻은 대푯값 간의 MD를 계산하는 수행 시간을 측정하였다. Window size는 학습단계에서 대푯값을 계산할 대, 사용되는 데이터의 개수를 의미한다. 본 성능평가에서 사용된 생체정보는 1초에 1번씩 수집되는 정보이기 때문에, Window size = 20은 총 20초 동안 수집된 정보를 이용하여 MD를 계산하는 것을 말한다.

첫 번째 실험은 키 사이즈 |N|을 256, 512, 1024, 2048 로 고정시키고 Window size를 20, 40, 60, 80, 100으로 변화시키며 실험을 진행하였다. Window size가 증가함에 따라서 수행시간이 선형으로 증가하는 것을 볼 수 있었다. 키 사이즈 |N|을 1024로 고정하고 Window size가 60일 때에 MD를 구하기까지 약 22초의 시간이 소요되었고, 키 사이즈 |N|이 2048비트이고 Window size 20일 때에 이상치를 탐색하기 까지 약 7-8분이 소요되었다. 또한 키 사이즈 |N|이 클수록 Window size에 따라서 이상치 탐색 소요시간이 빠르게 증가하였다.

추가적으로 암호화를 하지 않은 MD는 Window size가 증가함에 따라서 성능에 큰 차이를 보이지 않았으며 약 2-3ms의 결과가 도출되었다.

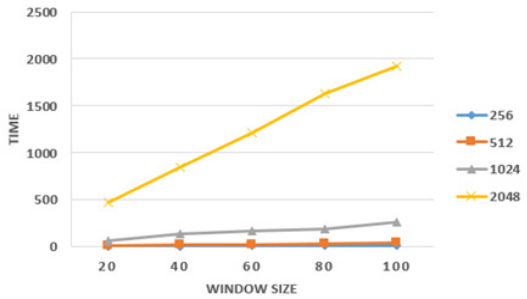


Fig. 8. Performance with respect to window size

6.3 키 사이즈

두 번째 성능평가 방법으로 윈도우사이즈를 고정시키고, 키 사이즈 |N|을 256, 512, 1024, 2048 비트로 증가시키며 수행시간을 측정하였다. 연산 수행시간은 키 사이즈에 비례하여 증가하였다. 1024 비트를 기준으로 Window size가 20일 때에 약 1분정도의 시간이 소요되었으며, 512비트는 10초의 시간이 소요되었다. 사용자가 EMS 서비스를 사용하고자 할 때에 사용자의 빠른 응답속도와 암호화의 안전성 사이에서 Trade-Off가 되어야 하는 부분이다. 속도를 결정하기 위하여, 사고직 후 사용자의 생명을 살릴 수 있는 골든타임시간(최소 5분~최대 10분)을 고려하여야한다[14].

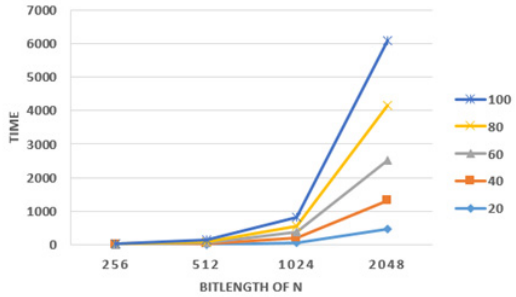


Fig. 9. Performance with respect to key size

VII. 분 석

이번 장에서는 제안한 기법이 4.1에서 정의한 요구사항을 만족함을 설명한다. 제안하는 기법이 Non-interactiveness와 기밀성을 만족하는 이유는 다음과 같다.

· **Non-Interactiveness**: 사용자 몸에 부착된 바디 센서로 부터 측정된 데이터는 게이트웨이 노드를 통해 인터넷에 연결 된다. 이 때, 일반적으로 게이트웨이 노드는 배터리 동작하기 때문에 지속적인 서버와의 인터랙션은 제한된다. 이러한 환경에서 Non-Interactiveness를 만족하는 시스템은 사용자가 지속적으로 서버와 통신할 필요가 없게 만들어 주고 배터리 소모를 줄여 서비스의 가용성을 확보할 수 있는 장점이 있다. 본 논문에서 제안한 기법에서 사용자는 CSP의 PP에 의해 만든 pk로 암호화된 암호문을 서비스공급자에게 전송하여 주면, 이후의 연산은 사용자를 제외한 서비스제공자와 CSP 간의 통신만이 필요하게 된다. 따라서 사용자의 바디센서와 게이트웨이노드는 연산 수행에 있어 다른 개체와 인터랙션이 없게 된다.

· **기밀성**: BAN, 사용자 단말기, 서비스 공급자, CSP 네 개의 개체 사이에 통신을 하는 전 구간에서 사용자 프라이버시 보호를 위하여 사용자 데이터에 대한 기밀성이 지켜져야 한다. 서비스 공급자는 노이즈를 추가하며 CSP는 MK를 이용하여 노이즈가 추가된 값을 복호화 하여 연산한다. 따라서 서비스 공급자와 CSP는 비밀값을 나눠 갖음으로서 사용자 데이터를 복호화 할 수 없으며 노이즈가 추가되어 실제 데이터를 알 수 없다. 사용자는 PP를 이용하여 자신의 키를 생성하고 CSP는 MK를 이용하여 복호화하기 때문에 외부공격자에 의해 전과정을 도청되더라도 기밀성이 지켜질 수 있다. 또한 본 논문에서 제안하는 기법은 사용자의 특성을 나타낼 수 있는 중간 값인 평균벡터와 공분산 행렬도 암호화로 보호하고 있다. CSP가 뺄셈, 곱셈, 나눗셈 등 연산을 수행 후 사용자의 pk로 암호화하여 서비스 공급자에게 전송하기 때문에 서비스 공급자는 중간 연산값에 대한 정보를 얻을 수 없다. 최종적으로 MD를 요청 했을 시에만 CSP는 결과 값을 반환한다.

VIII. 결 론

본 논문은 바이오 센서를 활용한 EMS 서비스 모델을 제안하였으며, 이 때, 동형암호 성질을 이용하여 사용자 프라이버시는 보호하며 EMS 서비스를 제공받을 수 있는 기법을 제안하였다. IoT환경에서는 실시간으로 생성되는 센서 정보를 활용하여 다양한 서비스가 제공될 것이다. 하지만 이러한 정보는 생활패턴, 질병의 유무 등 사용자 프라이버시와 관련된 정보를 담고 있으며 이것이 노출되었을 경우 사용자의 사생활 침해로 이어질 수 있다. 따라서, 제안한 기법을 통하여 바이오 센서를 사용하여 EMS 서비스를 제공받을 때 민감한 문제가 되는 사생활 침해 문제와 서비스공급자의 권한이 과도하게 주어져 사용자의 데이터를 복호화 할 수 있는 문제를 해결하였다.

References

[1] Sejin Park, Hogyu Lee, Jongseon Park, and Cheolu Kim, "Analog Front-End Design Techniques and Method for Saturation of Hemoglobin with Oxygen Sensor," Journal of Institute of Electrical and electronic engineers, 18(48), pp. 172-178, Mar. 2014.

[2] http://h21.hani.co.kr/arti/special/special_general/33985.html, hani.co.kr, Mar. 2013.

[3] Kyoung Sik Min, "Internet of Things," KISA NetTerm, Jun. 2013.

[4] Huijeong Hwang, "U-health system example based on standard," 2013 Health ICT Seminar, Oct. 2013.

[5] Seonhui Lee and Seonsil Yoo, "Status and View of Mobile Healthcare Application," 26(17), KISDI, Sep. 2014.

[6] Yeongju Jeon, "Utilization of electronic medical records (EMR) and protection of patient information," Korea Society of Health Service Management, 7(3), pp. 213-224, Sep. 2013.

[7] Jonghyeok Im, Jeongeun Song, and Mungyu Lee, "Speeding up encryption operation for a somewhat homomorphic encryption scheme over the integers," Korean Institute of Next Generation Computing, 10(2), pp. 6-18, Apr. 2014.

[8] Myeongin Jeong, "Technical Trend of Fully Homomorphic Encryption, Journal of Korea Contents Association," 13(8), pp 36-43, Jul. 2013.

[9] Monghyeon Lee, "Multivariate Spatial Cluster Analysis Using Mahalanobis Distance," The Korean Cartographic Association, 12(2), pp 37-46, Aug. 2012.

[10] Adreas Peter, Erik Tews, and Stefan Katzenbeisser. "Efficiently outsourcing multiparty computation under multiple keys," Information Forensics and Security IEEE Transactions, vol. 8, no. 12, pp. 2046-2058, Dec. 2013.

[11] Osman Salem, Yaning Liu, and Ahmed Mehaoua. "Anomaly detection in medical wireless sensor networks," The Korean Institute of Information Science and Engineering, 7(4), pp. 272-284. Oct. 2013.

[12] Jihye Kim, "Prosecution·Police controversial of 3000 people kakaotalk dialogue recording inspection," SISA Focus, Oct. 2014.

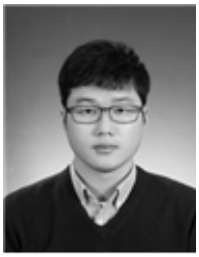
[13] www.Physionet.org,〔accessed Aug-2015〕

[14] https://en.wikipedia.org/wiki/Golden_hour_(medicine), 〔accessed Aug-2015〕

[15] Byeongman Cho and Dongyun Kim, "A Study of the Utility of Mahalanobis Distance for Decision of the Results of Health Examination," Journal of Korea industrial medicine, 6(2), pp 270-275, Sep. 1994.

〈저자소개〉

이 보 영 (Bo Young Lee) 학생회원
2013년 2월: 덕성여자대학교 컴퓨터시스템과 졸업
2016년 2월: 고려대학교 정보보호학과 석사
〈관심분야〉 정보보호, IoT, Privacy-preserving Analysis



최 원 석 (Wonsuk Choi) 학생회원
2008년 2월: 서울시립대학교 수학과 학사
2013년 2월: 고려대학교 정보보호대학원 석사
2013년 3월~현재: 고려대학교 정보보호대학원 박사과정
〈관심분야〉 정보보호, 의료기기 보안

이 동 훈 (Dong Hoon Lee) 종신회원
1983년 8월: 고려대학교 경제학사 졸업
1987년 12월: Oklahoma University 전산학과 석사 졸업
1992년 5월: Oklahoma University 전산학과 박사 졸업
1993년 3월~1997년 2월: 고려대학교 전산학과 조교수
1997년 3월~2001년 2월: 고려대학교 전산학과 부교수
2001년 3월~현재: 교려대학교 정보보호대학원 교수
〈관심분야〉 암호프로토콜, 암호이론, USN이론, 키 교환, 익명성 연구, PET기술