https://doi.org/10.5392/JKCA.2021.21.09.042

동형 암호의 데이터 수집 프로토콜 적용 방안 연구

A Study on Data Collection Protocol with Homomorphic Encryption Algorithm

이종덕*, 정명인, 유진철***

육군사관학교 컴퓨터과학과*, 육군사관학교 수학과**

Jongdeog Lee(jdlee6461@kma.ac.kr)*, Myoungin Jeong(mangjj@kma.ac.kr), Jincheol Yoo(jyoo@kma.ac.kr)***

요약

인터넷 사용 환경이 발전함에 따라 스마트폰과 각종 센서로부터 발생하는 대량의 데이터를 수집 및 분석하여 활용하는 데이터 기반 애플리케이션의 사용은 지난 10여 년간 폭발적으로 증가하였다. 그러나 이러한 사용자 데이터 기반의 애플리케이션을 사용하는 것은 언제든지 개인정보가 승인되지 않은 제3자에게 유출될 수 있다는 문제점을 내재하고 있다. 이러한 문제를 해결하기 위해 학자들은 데이터 교란과 암호화를 포함한 여러 기법을 사용해 왔다. 동형 암호는 암호화된 데이터를 복호화과정 없이 그대로 연산하더라도 결괏값이 보존되므로 원하는 연산을 수행할 때 개별 데이터를 복호화할 필요가 없어 기존의 방식보다 더 나은 개인정보보호를 제공한다. 본 연구에서는 개인정보를 보호하기 위해 사용되는 두 가지 알고리즘인 데이터 교란 방식과 전통 암호 방식 알고리즘을 구분하여 살펴보고, 두 가지 알고리즘의 단점을 보완할 수 있는 동형 암호를 이용한 데이터 수집 방법을 제안한다.

■ 중심어 : | 데이터 수집 프로토콜 | 센서데이터 | 동형 암호 | 개인정보보호 |

Abstract

As the Internet environment develops, data-analysis-based applications have been widely and extensively used in the past decade. However, these applications potentially have a privacy problem in that users' personal information may be leaked to unauthorized parties. To tackle such a problem, researchers have suggested several techniques including data perturbation and cryptography. The homomorphic encryption algorithm is a relatively new cryptography technology that allows arithmetic operations for encrypted values as it is without decryption. Since original values are not required, we believe that this method provides better privacy protection than other existing solutions. In this work, we propose to apply a homomorphic encryption algorithm that protects personal information while enabling data analysis.

■ keyword : | Data Collection Protocol | Sensor Data | Homomorphic Encryption Algorithm | Privacy Protection |

I. 서 론

최근 4차 산업혁명 기술의 발전과 함께 빅데이터 분석에 기반한 애플리케이션들이 많이 사용되고 있다. 대표적인 예로 MIT에서 개발한 CarTel[1]과 같은 교통정보 분석 앱을 들 수 있다. CarTel은 사용자 차량의

* 본 연구는 2020년 육군사관학교 사이버전 연구센터 연구과제로 수행되었습니다.

접수일자 : 2021년 07월 23일 심사완료일 : 2021년 08월 18일

수정일자 : 2021년 08월 18일 교신저자 : 정명인, e-mail : mangjj@kma.ac.kr

속도 정보를 이용하여 특정 도로구간의 혼잡 여부를 파악하여 이를 내비게이션 서비스에 활용하는 앱이다. 또한 해외에서는 사용자의 몸무게 정보를 공유하며 다이어트 진행 상황을 비교하는 다이어트 트래커(diet tracker)[2]와 같은 애플리케이션도 큰 인기를 얻었다. 이 외에도 사용자의 데이터를 수집, 분석하여 활용하는 애플리케이션으로 BikeNet, MMM2, ImageScape 등이 있다[3-5].

개인정보를 수집하여 분석하는 애플리케이션들의 가장 큰 문제점은 개인정보(privacy) 보호이다. 사용자들은 사용 데이터를 제공할 때 개인정보가 노출되는 것을 원하지 않는다. 예를 들면, 제공한 차량 속도 정보에서 개인정보 식별이 가능한 경우 속도 데이터를 통해 속도 위반인 것이 확인되면 경찰로부터 과속티켓을 받을 수도 있다. 또한 다이어트 트래커 애플리케이션에서 자신이 속한 그룹의 다이어트 진행 상황을 확인하고는 싶지만, 본인의 몸무게가 공개되는 것을 원하지는 않을 것이다. 이러한 개인정보 노출 문제는 사용자들이 정보제공에 대해 부정적인 인상을 받게 만들어 애플리케이션의 실패로 이어질 수 있다.

데이터를 수집하는 단계에서 개인정보가 노출되는 문제를 해결하기 위해서 여러 연구가 선행되어왔다. 그 중 대표적인 기법 중의 하나가 데이터 교란(data perturbation) 방식이다. 원본 데이터에 적정한 노이즈(noise) 신호를 추가하여 다른 사람이 원본 데이터를 알아볼 수 없게 하는 방법으로, 원본 데이터를 복원하기 위해서는 노이즈 신호를 제거하면 된다. 이때, 추가한 노이즈와 제거한 노이즈는 일반적으로 무작위 함수를 이용하므로 완전히 동일하지 않다. 따라서 복원한 값은 원본값에 근사하지만 일치하지는 않는다. 이처럼 데이터 교란 방법은 단순하고 효율적이지만 정확한 값을 복원할 수 없다는 단점이 있다. 높은 정확성이 요구되는 시스템의 경우, 데이터 교란 방법은 허위경보(false alarm)와 같은 오류를 야기할 수 있다. 예를 들면, 어떤 시설 특정 구역의 평균 온도가 기준치를 초과하는 경우 비상벨이 울리는 시스템이 있다고 가정하자. 데이터 교란 방식의 경우 이러한 오류로 인해 기준치가 초과하여도 비상벨이 울리지 않거나 반대로 초과하지 않았는데도 벨이 울리는 문제가 발생할 가능성이 있다.

또 다른 대표적인 방법은 전통적인 암호 알고리즘(encryption algorithm)을 이용하는 것이다. 암호 알고리즘은 키를 가지고 있는 사용자 간에 안전하게 데이터를 공유할 수 있는 검증된 방법이다. 센서들이 측정한 데이터를 암호화한 후 전송하면 수신 측에서는 수신한 데이터를 모두 복호화하여 분석을 시행한다. 이 방식은 암호키가 노출되지 않는 한 원본 데이터의 기밀성(confidentiality)이 유지된다는 장점이 있다. 그러나 데이터 분석을 위해서는 개별 데이터를 모두 복호화하여야 하므로 수신 측에 높은 컴퓨팅 파워가 요구된다는 단점이 있다. 특히, 데이터의 수가 많을 경우 수신 노드에 과부하가 발생할 수 있다. 또한 데이터 분석을 위해서는 모든 개별 데이터를 분석 전에 복호화하여야 하므로 복호화된 데이터를 처리하고 저장하는 과정에서 개인정보 노출의 위험성이 매우 높다.

이러한 대표적인 기법들의 문제점을 해결하기 위해 본 연구에서는 동형 암호를 이용한 데이터 수집 프로토콜을 제안한다. 동형 암호는 암호화된 데이터를 복호화하지 않고도 연산을 수행할 수 있는 암호기법 중 하나이다. 즉, 동형 암호를 이용하면 데이터가 암호화된 상태로 원하는 연산을 수행한 뒤 복호화를 하여 얻은 값과 원본 데이터에 같은 연산을 수행하여 얻은 값이 동일하다. 데이터 수집 프로토콜에 이를 적용할 경우, 송신 측이 데이터를 암호화해서 전송한다는 점에서는 전통적인 암호 알고리즘을 이용한 방식과 차이가 없다. 그러나 수신 측에서는 이를 복호화할 필요 없이 암호화된 데이터 자체에 대해 연산을 수행한 후 최종 결과만 복호화하면 된다. 따라서 전통 암호기법과는 달리 복호화된 데이터가 노출되지 않으며, 그로 인해 더욱 높은 수준의 개인정보보호가 가능하다.

지금까지 의료, 금융정보와 같은 민감한 데이터에 대한 동형 암호기법의 적용 필요성은 많이 논의되었으나 실제 시스템 레벨에서 여러 데이터 생산자들이 동형 암호를 통해 데이터 소비자와 어떻게 자료를 공유하는지에 대한 구체적인 프로토콜에 관한 연구는 많이 이루어지지 않았다. 본 연구에서는 2장에서 기존 연구들을 살펴보고, 3장에서 동형 암호 알고리즘을 소개한다. 그리고 4장에서는 동형 암호를 데이터 수집 프로토콜에 적용하는 방안을 제안하고, 5장에서 개인정보보호 측면에

서 다른 알고리즘들과 비교분석을 한 뒤, 마지막으로 6장에서 결론과 향후 연구에 관해 서술하며 마무리한다.

II. 관련 연구

2000년대 후반부터 센서 데이터에 포함된 개인정보를 보호하기 위한 연구들이 많이 제안되었다. 이를 크게 두 가지로 데이터 교란 방식과 암호 알고리즘 방식으로 구분할 수 있다.

데이터 교란 방식에 관한 대표적인 연구는 2008년에 Raghu Ganti 등이 제안한 PoolView이다[6]. 이 연구는 데이터에 단순히 노이즈 정보를 넣는 것이 아니라 애플리케이션에서 제공하는 모델을 이용하여 수학적으로 오차의 바운드를 계산할 수 있음을 증명하였다. 무작위 노이즈 신호를 입력하는 경우 데이터를 복원하는 과정에서 오차율의 예측이 어렵다는 문제점이 있는데 PoolView는 이러한 문제점을 수학적 방식으로 해결하였다. PoolView에서는 2가지 케이스 스터디를 통해 실제로 제안된 방법을 실험적으로 검증하였다. 이 방법은 애플리케이션에서 제공하는 모델 정보가 필요하므로 이 정보가 부정확할 시 오차율이 높아지는 단점이 있다.

2010년에 J. Shi 등은 PriSense를 제안하였다[7]. 이 연구에서 도심지역에 일어나는 이벤트들에 대해 사람이 제공하는(people-centric) 데이터를 이용하여 분석하는 도심 센싱(urban sensing) 환경을 가정한다. 이때, 사람들이 제공하는 데이터를 슬라이싱(slicing)과 믹싱(mixing) 기법을 통해 개인정보가 드러나지 않도록 보호하고, 분석 단계에서 정확한 합계, 평균, 최댓값, 그리고 최솟값 등의 통계적 수치를 계산할 수 있음을 시뮬레이션을 통해 확인하였다.

Qinghua Li는 2012년에 모바일 센싱 환경에서 additive homomorphic encryption algorithm을 이용하여 효율적으로 최솟값을 종합(Min aggregation)하는 프로토콜을 정의하였다[8]. 본 연구와 같이 동형암호를 사용하긴 하였지만, 기본적으로 애드혹 네트워크(ad-hoc network)에서 동작하는 프로토콜을 정의하였으며, fully homomorphic encryption algorithm을 사용한 본 연구와는 차이점이 있다.

2013년에 R. Zhang 등은 사용자 데이터의 homomorphic authentication code를 계산하여 종합을 담당하는 서버(aggregation server)로 보낸 뒤, 계산된 값이 원본과 일치하는지 확인하는 방법을 제안하였다[9]. 이 방식은 기본적으로 데이터의 무결성, 즉 공격자에 의해 원본 데이터에 어떤 악의적인 데이터가 삽입되거나 변조되지 않았는지를 확인하는 방법으로, 동형 암호 알고리즘을 이용하여 기밀성을 유지한 상태에서 데이터 분석을 하고자 하는 본 연구와는 방향성과 다르다.

C. Borcea는 2017년 PICADOR라는 Pub/Sub 프로토콜에 암호화 알고리즘을 이용하여 종단 간(end-to-end)의 기밀성을 달성한 프로토콜을 제안하였다[10]. 이 연구는 publisher와 subscriber 간에 사전에 키를 공유할 필요가 없다는 데 의의를 두고 있다. 이를 위해 중간 브로커가 PRE(proxy re-encryption) 방식을 통해 publisher와 subscriber 간 메시지 전달의 역할을 하여 Publisher가 암호화를 하고 브로커가 암호화된 데이터를 재 암호화하며, 마지막으로 subscriber가 복호화를 한다. 이 과정을 성공적으로 수행하기 위해 PALISADE이라는 동형 암호 알고리즘을 사용하였는데, 이 연구는 Pub/Sub 프로토콜에 동형 암호를 적용한다는 점에서는 본 연구와 유사하지만, 그 목적이 다르다. PICADOR는 publisher와 subscriber의 키 공유 없이 암호화된 메시지를 전송하기 위하여 동형 암호를 사용하였으나 본 연구는 암호화된 상태에서 연산할 수 있게 하여 개인정보가 유출되지 않도록 하는 데 그 목적이 있다.

가장 최근에 S. Li 등은 산업용 사물 인터넷 환경에서 동형 암호를 이용한 개인정보 보호가 가능한 경량 스킴을 제안하였다[11]. 이 연구에서는 공기 질 데이터를 동형 암호를 이용하여 암호화된 상태에서 원하는 연산을 수행한 결과를 도출하였다. 그러나 이 논문에서는 외부에서의 위협 상황은 고려하지 않았기 때문에 본 연구와는 차이가 있다.

Ⅲ. 동형 암호 알고리즘

일반적으로 암호는 패스워드(password)와 암호화(encryption)를 포함하는 기술을 뜻한다. 암호는 크게 단순 인증(authentication) 기능만을 가지는 패스워드와 같은 1세대 암호, 데이터 암호화가 가능한 대칭키 암호(symmetric encryption)를 통칭하는 2세대 암호, 복잡한 키 공유 알고리즘이 필요 없어서 비대칭키 암호(asymmetric encryption)라 부르는 3세대 암호로 구분할 수 있다. RSA 알고리즘으로 대표되는 3세대 암호 알고리즘이 개발되기 전까지는 데이터 송신자와 수신자가 사전에 동일한 키를 공유하여야 비밀 메시지를 교환할 수 있었다. 그러나 RSA 알고리즘은 데이터 송신자가 수신자의 공개키(public key)를 이용하여 암호화를 하고 수신자는 본인의 개인키(private key)로 이를 복호화하는 방식으로 별도의 키 공유 과정 없이 비밀 데이터를 교환할 수 있게 하였다. 동형 암호 알고리즘은 4세대 암호로 분류되며 암호화된 데이터를 복호화하지 않고도 덧셈과 곱셈 같은 연산을 수행할 수 있는 알고리즘이다. 동형 암호 중 XOR, AND, 덧셈, 곱셈 등의 연산 중 일부가 보존되는 성질을 가진 것을 '제한 동형 암호(somewhat homomorphic encryption)'라 하고, 덧셈과 곱셈을 포함하여 XOR, AND와 같은 모든 논리 연산을 보존하는 알고리즘을 '완전 동형 암호(fully homomorphic encryption)'라 한다.

동형 암호 알고리즘은 1978년 Rivest, Adleman, Dertouzous가 제안한 privacy homomorphism에서 처음 소개되었다[12]. 이 연구에서는 RSA 암호시스템의 변형을 포함한 5가지 방법이 제안되었으나 안정성을 검증하지 못하였다. 이후 1999년에 Paillier는 일부 연산이 보존되는 제한 동형 암호인 Paillier Cryptosystem을 발표하였다[13]. 이후 Paillier Cryptosystem은 2001년 Damgård와 Jurik에 의해 일반화되었고[14], 2002년 S. Galbraith에 의해 타원곡선 상에서 적용한 방법[15]으로 개선되는 등 동형 암호 알고리즘에 관한 연구가 활발하게 이루어진다. 2005년에는 Boneh, Goh, Nissim에 의해 타원곡선 상에서 정의된 겹 선형 사상을 이용하여 덧셈과 한 번의 곱셈이 가능한 동형 암호 알고리즘이 소개되었다[16].

2000년 후반에 접어들면서 제한된 연산만이 가능했던 제한 동형 암호에서 원하는 횟수만큼의 연산을 보존할 수 있는 완전 동형 암호 알고리즘으로 연구가 발전되었다. 1978년 처음 동형 암호가 제안된 이후 30년 동안 안전성이 증명된 완전 동형 암호를 개발하지 못하다가 2009년 Gentry에 의해 최초로 안전성이 증명된 완전 동형 암호 알고리즘이 제안되었다[17]. 2010년에는 Dijk, Gentry, Halevi, Vaikuntanathan가 격자 기반이 아닌 정수 집합에 Gentry의 방법을 적용한 제한 동형 암호를 발표하였다[18]. 하지만 두 가지 방법 모두 연산을 거듭할수록 잡음(noise)이 증폭되어 일정 횟수 이상의 연산을 반복하게 되면 에러가 커져서 정확한 복호화가 어렵다는 문제점을 가지고 있다.

2013년에는 중국인의 나머지 정리(CRT, Chinese Remainder Theorem)에 기반을 둔 완전 동형 암호 방식이 국내 연구진에 의하여 제안되었다[19]. CRT 기반의 알고리즘은 원래 1978년에 최초로 동형 암호가 제안되었을 때 함께 소개되었으나 알려진 평문 공격(known-plaintext attack)에 취약하여 활용이 제한되었다. 이후에도 동형 암호 알고리즘을 실용적으로 적용할 수 있도록 계산의 효율성과 연산의 정확성 등의 측면에서 성능을 향상시키기 위한 다양한 연구가 진행되고 있다.

Ⅳ. 프로토콜 디자인

본 절에서는 개인정보보호를 위한 데이터 수집 프로토콜 설계에 대해 논의한다. 이를 위해 먼저 위협 모델을 설정하고, 기존에 제안되었던 대표적 해결방안을 분석한다. 마지막으로 동형 암호를 이용하여 설계된 새로운 프로토콜을 제안한다.

1. 위협 모델

본 프로토콜의 위협 모델은 크게 두 가지이다. 첫 번째는 공격자가 통신 채널을 도청하는 것이다. 본 연구에서 가정하고 있는 환경에 해당하는 무선 노드의 경우, 신호 반경 내에 공격 노드를 위치시켜 전송되는 메

시지를 쉽게 엿들을 수 있다. 두 번째는 일부 노드를 공격하여 공격자에게 협조적인 노드(compromised node)로 만드는 것이다. 이 경우 노드에 저장된 자료 중 보호되지 않은 영역에 있는 데이터는 공격자에게 노출될 수 있다. 예를 들어, 보호가 필요한 암호키의 경우 메모리 또는 디스크의 특수한 영역에 저장하고 노드 손상이 의심되는 경우 사용이 불가하도록 하드웨어적으로 구성할 수 있다. 하지만 이 경우, 일반 영역에 있는 데이터는 여전히 공격자들의 접근이 가능하여 그곳에 저장된 개인정보 등이 유출될 수 있다. 두 가지 모델 중 특히 두 번째 위협 모델은 단순한 암호 알고리즘을 이용하여 개인정보를 보호하기 어렵게 만든다. 데이터를 분석하기 위해서는 개별 데이터를 반드시 복호화해야 하는데, 복호화된 데이터가 비보호 영역에 저장되면 개인정보의 유출이 발생할 수 있기 때문이다.

본 연구에서는 위의 두 가지의 위협 모델을 가정하고 위협에 대응하여 데이터 제공자의 개인정보를 효과적으로 보호하는 방안을 강구한다.

2. 기존 프로토콜 분석

관련 연구에서 분석한 바와 같이 대표적인 개인정보 보호 방법으로 데이터 교란 방식과 암호 알고리즘을 이용한 방식을 들 수 있다. 본 연구에서는 동형 암호와 기존 암호를 구분하기 위하여 대칭키와 공개키 암호를 통칭하여 전통 암호 방식으로 지칭한다.

2.1 데이터 교란 방식

데이터 교란 알고리즘은 앞에서 언급했듯이 원본 데이터에 노이즈를 추가하여 공격자가 원본 데이터를 유추하기 어렵게 만드는 방법이다. 송신자는 특정 분포에서 난수값을 생성하여 원본 데이터에 이를 더하여 송신하고, 수신자는 동일한 분포에서 난수값을 생성하여 수신한 값에서 이를 빼는 방식이다. 예를 들어, 어떤 사용자가 자신의 몸무게인 70kg에 노이즈를 추가한다고 가정하자. 노이즈는 -30에서 +30 사이의 균등분포(uniform distribution)에서 발생한 난수값 20이라 하자. 이 경우 교란된 자룟값은 90kg이 되어 공격자는 원래 몸무게 70kg을 유추하기 어렵다. 수신자는 수신한 값 90kg에 다시 생성한 노이즈를 이용하여 이를 제거한다. 이번에는 난수값 0이 생성되었다고 가정하자. 노이즈가 0이므로 복원된 값은 그대로 90kg이 된다. 복원된 값이 원본값과 20kg의 차이가 나므로 오류라고 생각할 수 있지만, 프로토콜의 목적은 수집한 개별 데이터를 복원하는 것이 아니라 전체 데이터를 분석하는 것이므로 데이터 샘플의 양이 많아짐에 따라 정확도가 자연스럽게 높아지게 된다. 다시 말해, 동일한 방식으로 대량의 데이터에 노이즈를 추가 및 제거하게 되면 교란된 자료의 분석값(합산, 평균 등)은 원본 자료의 분석값에 매우 가까워진다.

이 방식은 노이즈를 생성하여 추가하고 제거하기만 하면 되기 때문에 상대적으로 연산속도가 빠르며, 송수신되는 데이터가 도청 등을 통해 노출되더라도 원본 데이터를 유추하기 어렵다는 장점이 있다. 그러나 난수값의 범위에 따라서 정확도의 오차가 커질 수 있으므로 프로토콜 설계 전에 데이터의 특성에 대한 이해가 선행되어야 한다는 한계가 존재한다. 예를 들어, 몸무게 분석에 추가되는 난수값을 -1000에서 1000까지의 값으로 지정하면 정확도가 낮아지게 되므로 몸무게라는 변수의 특성에 맞는 적절한 난수 범위를 선정해야 한다. 더불어 이러한 가정을 모두 만족하더라도 100%의 정확성은 달성할 수 없다는 문제점이 있다.

2.2 전통 암호 방식

전통 암호 알고리즘 중 대칭키 암호가 공개키 암호보다 효율적이기 때문에 본 연구에서는 대칭키 암호를 이용한다고 가정한다. 단, 대칭키 암호를 사용하기 위해서는 키 교환이 선행되어야 하므로 대표적인 키 교환 알고리즘인 디피 헬만(Diffie-Hellman) 키 교환 알고리즘을 이용하거나 공개키 알고리즘을 이용하여 비교적 길이가 짧은 키를 사전에 교환해야 한다.

전통 암호 방식을 이용하여 송신 노드가 정보를 전달하기 위해서는 단순히 공유한 키를 이용하여 데이터를 암호화하여 전송하면 된다. 수신 측에서는 마찬가지로 공유된 키를 이용하여 데이터를 복호화한 뒤 이를 분석에 사용하면 된다. 예를 들어, 몸무게가 70kg인 사용자가 70이라는 데이터를 암호화해서 송신한다면 수신자는 복호화를 통해 70이라는 원본값을 얻을 수 있다.

이 방식의 장점은 정확한 데이터를 복원할 수 있다는

점이다. 암호화된 데이터를 복호화하므로 원본 데이터가 변조되지 않으며, 따라서 원본값과 동일한 분석값을 얻을 수 있다. 또한, 키가 유출되지 않는 한 공격자가 암호화된 데이터를 획득한다고 하더라도 원본값을 복원할 수 없다. 그러나 앞의 위협 모델에서 설명한 바와 같이 수신 측에서는 개별 데이터를 복원할 수 있으므로 잠재적인 개인정보의 유출 위험이 존재한다. 또한 암호화 및 복호화 연산에는 높은 계산 비용이 요구되므로 대량의 데이터에 이를 적용하기는 부담이 따른다. 송신자/수신자 모두 모든 데이터를 암호화/복호화해야 하는데 연산에 많은 시간과 컴퓨팅 파워가 소모되므로 데이터가 많아질수록 감당해야 할 비용이 데이터의 양만큼 증가한다.

3. 제안하는 방식

본 연구에서는 개인정보보호를 위해 동형 암호를 활용한 데이터 수집 프로토콜을 제안한다. 본 프로토콜에서 송신 측은 전통 암호 방식과 동일하게 수신 측과 공유한 키를 이용하여 데이터를 암호화해서 송신한다. 그러나 수신 측에서는 전통 암호 방식과는 다르게 동형 암호 알고리즘을 이용하여 개별 데이터의 복호화과정 없이 암호화된 상태의 데이터에 원하는 연산을 적용하여 값을 계산한다. 마지막으로 계산된 값을 복호화하면 전통 암호 방식으로 계산한 데이터의 분석값과 동일한 값을 구할 수 있다. 세부적인 프로토콜 동작 방식은 아래와 같다.

표 1. Pub/Sub protocol API(Application Program Interface)

구분	API	기능
publisher	pub(topic, content)	주제에 콘텐츠를 출판하는 함수
subscriber	sub(topic, [callback])	주제를 구독하는 함수 (데이터가 수신될 때 필요 시 callback 함수를 호출)
	sum(topic)	현재까지 수신된 값들의 합계를 구하는 함수
	average(topic)	현재까지 수신된 값들의 평균을 구하는 함수
	reset(topic)	현재까지 수신된 값을 삭제하는 함수

네트워크에는 다수의 데이터 생산자와 소비자가 존재하며 이들은 [표 1]에 제시된 API를 통해 데이터를 생산하고 소비할 수 있다. 본 논문에서는 pub/sub 프로토콜을 기준으로 데이터를 생산하거나 소비할 때 어떤 방식으로 개인정보를 보호하는지를 설명한다.

프로토콜의 제어 흐름(control flow)은 [그림 1]과 같다. 주제(topic) 기반 pub/sub 프로토콜은 해당 주제에 대한 publish를 실행하면 동일 주제를 구독(subscribe) 하는 노드에 데이터가 전달되는 방식이다.

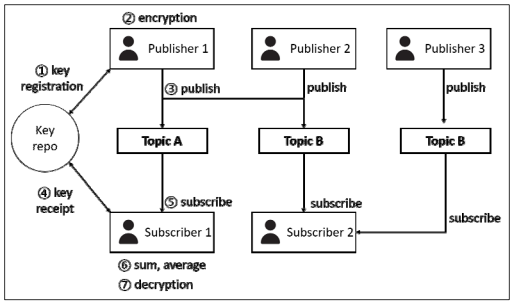

그림 1. 프로토콜 제어 흐름

이를 위해 먼저 ①단계에서 데이터 생산자가 해당 주제에 대한 키를 등록한다. 이때, 먼저 키 저장소(key repository)를 확인하여 해당 주제가 이미 등록이 되어 있는지를 확인하여 만일 이미 등록된 주제라면 단순히 그에 해당하는 키를 받는다. 그렇지 않을 경우 최초로 등록된 주제이므로 새로운 키를 생성하여 이를 키 저장소에 등록한다. 키 저장소는 시스템의 단일 장애점(single source of failure)이 될 수 있으므로 시스템 안정성을 높이기 위하여 여러 레플리카(replica)에 분산 저장하는 방식을 사용할 수 있다.

②단계에서는 ①단계의 키를 이용하여 데이터를 암호화한다. 암호화된 데이터는 ③단계에서 네트워크에 전송이 되며 해당 주제를 구독하고 있는 데이터 소비자들에게 전달된다. 데이터 생산자 입장에서는 [표 1]의 pub 함수를 호출하면 ②단계와 ③단계가 자동으로 수행이 된다. 즉, ②단계는 데이터 생산자에게 명시적이지 않기 때문에 사용자 API([표 1])에 포함되지 않는다.

데이터 소비자는 ④단계에서 sub 함수를 호출하여 원하는 주제에 해당하는 콘텐츠를 수신할 수 있다. 이

때, 프로토콜은 키 저장소에 접근하여 해당 주제에 해당하는 키를 받는다. 곧바로 ⑤단계에서 생산자가 데이터를 게시하면 이를 수신하여 수신자의 애플리케이션에 데이터를 전달한다. 만일 수신될 때마다 특정 이벤트를 발생시키고 싶다면 [표 1]에 표현된 것처럼 사용자 정의 콜백(user-defined callback) 함수를 정의할 수 있다.

사용자는 현재까지 수신된 값들에 대한 통계적 데이터를 요구할 수 있으며 이는 ⑥단계에서 일어난다. 본 프로토콜에서 제공하는 함수는 sum과 average이지만, 필요에 따라 이는 더욱 확장될 수 있다. 만약 데이터 소비자가 특정 시점까지의 값을 삭제하기를 원한다면 reset 함수를 사용할 수 있다. 물론 통계 데이터가 소비자에게 전달되려면 ⑦단계와 같은 복호화 작업이 이루어져야 한다. 이는 앞의 ②단계와 마찬가지로 소비자가 직접 사용하지 않고 사용자가 값을 요구할 때 자동으로 실행되므로 [표 1]의 프로토콜 API에 제공되지 않는다.

V. 분 석

본 장에서는 제안된 동형 암호 알고리즘을 이용한 데이터 수집 프로토콜이 다른 들보다 개인정보보호에 탁월한 이유를 정성적으로 분석하고자 한다. 먼저 전통 암호 알고리즘과 비교해보자.

동형 암호 알고리즘을 이용하면 수집한 데이터를 통계적으로 처리하기 위해 개별 데이터들을 복호화할 필요가 없다. 암호화된 상태에서도 연산이 가능하므로 개별 데이터를 복원하지 않고도 통계 수치를 계산할 수 있다. 반면에 기존의 암호 알고리즘을 이용하는 경우 데이터를 분석하기 위해 반드시 개별 데이터를 복호화하는 과정이 필요했다. 따라서 두 번째 위협 모델에서 살펴본 바와 같이 개별 데이터가 복호화된 상태에서 메모리나 디스크에 저장되는 경우, 공격자들에 의해 정보가 유출될 수 있는 위험이 존재한다. 데이터를 처리하는 노드가 여러 개인 경우라면 문제는 더욱 심각해진다. 복호화의 경우 일반적으로 높은 연산 비용이 발생한다. 특히, 처리해야 할 데이터가 수백, 수천만 개의 빅데이터라면 싱크 노드와 같이 데이터를 최종 수신하는 노드에서 이를 모두 처리하기 어려울 것이다. 효율성을 위해 데이터를 분산해서 처리하는 방법이 있지만 대칭키 알고리즘의 경우에는 분산 처리기관들이 암호키를 공유하여야 하므로 키 관리에 문제가 생길 수 있다. 만약 암호키를 안전하게 공유한다고 하더라도 복호화된 데이터가 저장되어 있는 장소가 늘어나고 복호화된 데이터를 안전하게 송수신하는 문제 때문에 개인정보를 보호하기 더욱 어려워진다.

데이터 교란 알고리즘은 노이즈를 추가하고 제거하는 상대적으로 단순한 연산을 이용하여 효율적인 개인정보보호가 가능하다. 데이터를 최종 수신하는 노드에서도 개별 데이터가 정확히 무엇이었는지 확인할 수 없으므로 개인정보를 효과적으로 보호할 수 있다. 그러나 개별 데이터를 완전히 복원할 수 없다는 태생적인 문제점 때문에 높은 정확성을 요구하는 애플리케이션에는 사용이 제한된다. 반면에 동형 암호 알고리즘은 이론적으로 원본 데이터와 동일한 계산 값을 보장한다. 따라서 높은 정확성을 요구하는 애플리케이션에서도 큰 문제 없이 사용할 수 있다.

이와 같이 전통적인 암호 알고리즘은 개별 데이터를 반드시 복원해야만 데이터를 분석할 수 있다는 문제점이 있고, 데이터 교란 알고리즘은 태생적인 정확도 문제로 인해 정확한 데이터 분석이 제한된다는 문제점이 있다. 따라서 동형 암호 알고리즘을 데이터 수집 프로토콜에 적용하여 효과적으로 개인정보를 보호하면서도 다양한 수치를 정확히 분석할 수 있게 해주는 효율적인 방안이 요구된다.

VI. 결론 및 향후 연구

본 연구에서는 사물 인터넷 환경하에서 센서 또는 데이터 생산자로부터 데이터를 수집할 때 개인정보를 효율적으로 보호하기 위한 방안에 대하여 논의하였다. 기존 연구에서는 대표적으로 원본 자료에 변형을 가하여 공격자가 중간에 데이터를 가로채더라도 원본값을 유추하기 어렵게 만들거나 암호기법을 이용하여 암호키가 유출되지 않는 한 기밀성을 보장받는 방식을 사용하였다. 본 연구에서 제안하는 방법은 동형 암호 알고리

즘을 데이터 수집 프로토콜에 적용하는 것이다. 본 프로토콜에서 송신 측은 수신 측과 공유한 키를 이용하여 데이터를 암호화하여 송신하지만 송신 측은 암호화된 데이터를 그 자체로 연산하여 원하는 분석값을 도출하므로 원본 데이터에 접근할 필요가 없어서 개인정보보호에 탁월하다.

동형 암호 알고리즘이 데이터 수집 프로토콜에 적용될 때 염려되는 부분은 비용이다. 데이터 암호화 및 복호화에 대한 부분을 차치하더라도 암호화된 데이터를 그대로 연산하면 원본 데이터에 대한 연산보다 높은 비용이 발생한다. 또한 데이터가 암호화되면 원본 데이터보다 메시지 길이가 늘어나게 되고 따라서 네트워크 대역폭(bandwidth)이 낭비될 수도 있다. 제안된 방안을 실제 프로토콜로 구현하여 사용하기 위해서는 먼저 비용 분석을 통해 효율성을 검증할 필요가 있다. 또한 주제별 단일키 사용과 같은 단점들을 보완하는 방안을 강구하여야 한다. 차후 연구에서는 본 연구에서 수행한 정성적 분석을 바탕으로 실제 알고리즘 구현을 통해 제안한 프로토콜을 정량적으로 분석할 예정이다. 더불어 앞에서 언급한 주제별 단일키 사용 등의 단점을 보완하는 방안에 관한 연구를 진행 중이다. 이러한 연구는 실제로 데이터 수집 프로토콜에 동형 암호 알고리즘을 적용하기 위한 중요한 연구 주제가 될 것이라 확신한다.

참 고 문 헌

[1] Bret Hull, “CarTel: A distributed mobile sensor computing system,” Proc. ACM SenSys, pp.125-138. 2006.

[2] https://www.myfitnesspal.com/

[3] Shane Eisenman, “BikeNet: A Mobile Sensing System for Cyclist Experience Mapping,” TOSN, 2006.

[4] Marc Davis, “MMM2: mobile media metadata for media sharing,” CHI EA, 2005.

[5] Sasank Reddy, “Image browsing, processing, and clustering for participatory sensing: Lessons from a DietSense prototype,” Proceedings of the 4th Workshop on Embedded Networked Sensors, EmNets, 2007.

[6] Raghu Ganti, “PoolView: Stream privacy for grassroots participatory sensing,” SenSys - Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems, pp.281-294, 2008.

[7] Jing Shi, “PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems,” 2010 Proceedings IEEE INFOCOM, San Diego, CA, 2010.

[8] Qinghua Li, “Efficient and privacy-preserving data aggregation in mobile sensing,” 2012 20th IEEE International Conference on Network Protocols (ICNP), Austin, TX, 2012.

[9] Rui Zhang, “Verifiable Privacy-Preserving Aggregation in People-Centric Urban Sensing Systems,” in IEEE Journal on Selected Areas in Communications, Vol.31, No.9, pp.268-278, 2013.

[10] Cristian Borcea, “PICADOR: End-to-end encrypted Publish-Subscribe information distribution with proxy re-encryption,” Future Generation Computer Systems, Vol.71, Pages pp.177-191, 2017.

[11] S Li, S Zhao, G Min, L Qi, and G. Liu, “Lightweight privacy-preserving scheme using homomorphic encryption in industrial Internet of Things,” IEEE Internet of Things Journal, 2021.

[12] R. Rivest, L. Addleman, and M. Dertouzos, “On data banks and privacy homomorphism,” In Foundations of Secure Computation, pp.169-177, 1978.

[13] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” EUROCRYPT, pp.223-238, 1999.

[14] I. Damgård and M. Jurik, “A Generalization, a Simplification and Some Applications of Paillier's Probabilistic Public-Key System,” Public Key Cryptography - PKC, pp.119-136, 2001.

[15] S. Galbraith, “Elliptic curve Paillier schemes,” Journal of Cryptology - JOC, Vol.15, No.2, pp.129-138, 2002.

[16] D. Boneh, E. Goh, and K. Nissm, “Evaluating 2-DNF formulars on ciphertexts,” Theory of Cryptography, pp.325-341, 2005

[17] C. Gentry, “Fully homomorphic encryption using ideal lattices,” ACM Symposium on Theory of Computing - STOC, pp.169-178, 2009.

[18] M. V. Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, “fully homomorphic encryption over the integers,” Advances in Cryptology - EUROCRYPT, Lecture Notes in Computer Science, Vol.6110, pp.24-43, 2010.

[19] J. H. Cheon, J. S. Coron, J. Kim, M. S. Lee, T. Lepoint, M. Tibouchi, and A. Yun, “Batch fully hoomorphic encryption over the integers,” Advances in Cryptology - EUROCRYPT, Lecture Notes in Computer Science, Vol.7881, pp.315-335, 2013.

저 자 소 개

이 종 덕(Jongdeog Lee) 정회원

- 2005년 3월 : 육군사관학교 전산학과(B.S.)
- 2019년 12월 : 일리노이대학교 컴퓨터과학과(Ph.D.)
- 2019년 8월 ~ 현재 : 육군사관학교 컴퓨터과학과 조교수

〈관심분야〉 : 사물인터넷, 정보보안

정 명 인(Myoungin Jeong) 종신회원

- 2004년 3월 : 육군사관학교 운영분석(B.S.)
- 2018년 8월 : 뉴욕주립대학교 수학과(Ph.D.)
- 2018년 9월 ~ 현재 : 육군사관학교 수학과 조교수

〈관심분야〉 : 암호학, 정보보안

유 진 철(Jincheol Yoo) 정회원

- 1989년 3월 : 육군사관학교 전산학과(B.S.)
- 2003년 5월 : 펜실베니아 주립대학교 컴퓨터공학과(Ph.D.)
- 2012년 11월 ~ 현재 : 육군사관학교 컴퓨터과학과 정교수

〈관심분야〉 : 컴퓨터시스템, 정보보안