사물인터넷(Internet-of-Things) 플랫폼에서의 익명성 확보 기술에 관한 연구

김수빈, 송재승*

세종대학교, 정보보호학과

18011630@sju.ac.kr, jssong@sejong.ac.kr

A Study on the Solution for Securing Anonymity in the IoT Platform

Kim Subin, Song JaeSeung*

Sejong Univ.

요 약

최근 사물인터넷의 발달과 함께 공유되는 데이터의 양 또한 증가하고 있으며, 그 가운데 개인을 특정할 수 있는 데이터들도 상당수 포함이 되고있다. 사물인터넷 플랫폼에 저장된 개인정보가 노출이될 경우 보이스피싱 등과 같은 다양한 사회적 문제들을 야기할 수 있다. 기존 국제 표준에서 정의된 사물인터넷 플랫폼에서는 개인정보에 대한 식별방지 기술에 대한 반영이 미흡한 상황이다. 본 논문에서는 사용자에게 공개되어있는 비식별화된 개인정보를 활용하여 개인을 재식별할 수 있는 위험성에 대하여 인지하고, 개인의 익명성 확보를 위한 사물인터넷 플랫폼 설계 방식을 제안한다. 특히, 개인정보 식별 방지를 위하여 필요한 기능을 사물인터넷 플랫폼의 공통기능 중 하나로 제공하는 방안에 대해 분석하고, 이를 제공하기 위하여 고려해야 할 점을 연구하였다.

Ⅰ. 서 론

최근 사물인터넷(Internet of Things, IoT)을 구성하는 기계, 센서, 카메라 등 인터넷에 연결된 기기의 수가 꾸준히 증가하며, 사물인터넷 플랫폼에서 다뤄지는 데이터의 양 또한 증가하고 있다. 하지만 공개되어있는 방대한 양의 데이터는 개인의 프라이버시를 침해할 수 있다. 이에 각국에서는 개인정보보호법을 제정하여 개인의 프라이버시를 보호하고자 노력하고 있다. 특히, 2018년 5월 EU에서는 데이터 프라이버시를 보호하고 규제하기 위하여 강화된 개인정보보호법령 General Data Protection Regulation (GDPR)을 제정하였다. IoT 플랫폼은 스마트시티 등에 활용되는 등 수많은 데이터를 수집할 때 활용되기 때문에 이러한 법령을 준수하기 위한 제도 및 기술적 기능의 추가가 필요한 상황이다[1]. 특정 개인을 식별할 수 없게 하는 개념인 익명화는 프라이버시 침해 위협을 줄일 수 있으므로, GDPR을 준수하기 위하여 IoT 플랫폼에서 반드시 적용이 필요로 되어지는 기능이다.

개인을 식별할 수 없도록 이름, 전화번호, 주소 등 식별자를 암호화하거나 제거하여 개인정보 비식별화를 진행한다고 하더라도, 공개된 정보를 연결하여 개인을 식별하는 연결 공격(Linkage Attack)으로 비식별화된 개인정보를 재식별할 가능성이 존재한다. 특히, 오늘날에는 데이터의 양이 증가하여 공유되고 활용되며 쉽게 접근할 수 있으므로, 연결 공격은 심각한 개인정보 유출 문제가 될 수 있다. 이러한 재식별 공격을 방지하기 위하여 사물인터넷 플랫폼에서는 하나의 기본 기능으로 익명성을 확보할 필요가 있다[2].

익명성을 확보하는 기술로 K-익명성(K-anonymity), L-다양성(L-diversity), T-근접성(T-closeness) 등의 프라이버시 보호 모델이 개발되었다. 해당 프라이버시 보호 모델은 데이터 재식별화를 최소화하는 것을 목표로 한다. 그러나 이러한 기술들을 공통기능의 하나로 사물인터넷 플랫폼에서 제공하기 위한 노력과 연구는 아직 부족한 상황이다.

본 논문에서는 IoT 플랫폼에서 개인정보를 안전하게 비식별화하기 위하여 프라이버시 보호 모델을 적용할 수 있는 방식을 연구하였다. 이러한 기술은 앞으로 IoT 플랫폼에서 수집되는 개인정보 데이터가 보다 더 안전하게 저장되고 공유되는 기대효과를 가져올 것으로 보인다.

Ⅱ. 본 론

식별자(Identifier)는 어떤 대상을 유일하게 식별 및 구별할 수 있는 요소이다. 식별자의 일부 또는 전부를 삭제/대체 하거나 다른 정보와 쉽게 결합하지 못하도록 하여 대상을 식별할 수 없도록 하는 것을 비식별화(De-identification)라고 한다. 개인정보를 비식별화하게 되면 빅데이터 활용이 증가하는 환경에서 개인정보를 안전하게 이용할 수 있게 된다. 그러나 비식별화된 정보를 조합, 분석 또는 처리하는 과정에서 대상을 식별할 수 있게 되는 재식별화가 이뤄질 가능성이 존재한다.

이러한 연결 공격은 익명화 기법을 적용함으로써 데이터 주체를 더이상 식별할 수 없도록 요소를 제거함으로써 해결할 수 있다. 익명화를 위하여 재식별 가능성을 검토하는 프라이버시 보호 모델에는 K-익명성 기법, L-다양성 기법, T-근접성 기법이 있다.

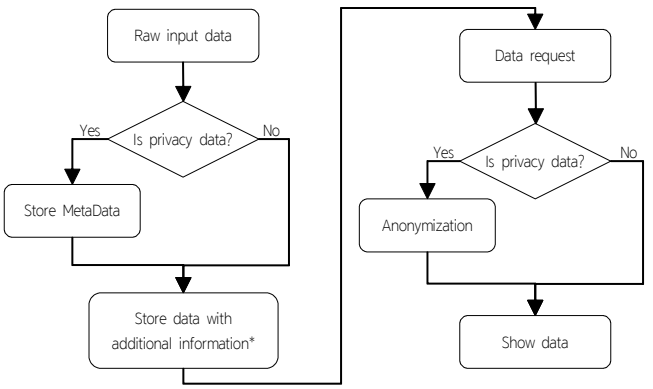
- K-익명성: 공개된 정보에 대한 연결 공격 등의 취약점을 방지하기 위하여 제시된 프라이버시 보호 모델로, 식별자 속성값의 동일한 데이터의 개수가 적어도 K개 존재하도록 한다. 그러나, K-익명화된 데이터는 동질성 공격이나 공격자가 배경지식을 가지고 있는 경우에 프라이버시를 보장할 수 없다는 단점이 존재한다.
- L-다양성: 공개된 정보에서 비식별 되는 레코드들은 해당 집합 내에 적어도 L개의 다른 정보를 포함해야 한다[3]. 그러나, L-다양성에도 취약점이 존재한다. 비식별화된 데이터의 정보가 유사한 경우이거나 정보가 특정 값에 치우쳐 있는 경우에는 L-다양성 기법을 사용하여도 프라이버시를 보장할 수 없다.
- T-근접성은 L-다양성의 취약점을 보완하기 위하여 개발된 기법이다. 정보가 치우쳐 있거나 유사한 경우에도 프라이버시를 보장하기 위하여, 동일한 집합 내에서 특정 정보의 분포와 전체 데이터에서 특정 정보의

분포가 유사해야 한다. 그 두 값은 T 이하의 차이를 가져야 한다[4].

이처럼 취약점을 보완한 프라이버시 보호 모델이 개발되고 있음에도 불구하고, 현재 획일화된 개인정보 관리체계가 수립되어 있지 않기 때문에 정보 관리 및 평가가 정상적으로 이루어지지 않고 있으며, IoT 서비스 개발자들이 상황에 맞게 익명성 확보를 위한 기능을 손쉽게 사용할 수 있도록 하는 API들이 개발되어있지 않기 때문에. 개인정보를 보호하기 위해서는 플랫폼 설계단계에서부터의 적용이 필요하다[5].

위에서 언급한 재식별화 방지 기법들을 실제 사물인터넷 플랫폼에 적용하기 위해서는, 플랫폼 설계단계에서 데이터를 수집할 때에 해당 데이터에 개인정보가 포함되어 있는지에 대한 식별이 필요하며, 개인정보가 포함된 경우에 해당 정보를 정해진 절차 및 다양한 비식별화 알고리즘을 통해서 처리한다면 플랫폼에서 프라이버시 보호 모델을 적용하고, 이를 준수하는 사물인터넷 서비스들을 제공할 수 있을 것이다.

IoT 플랫폼으로 수집되는 데이터에 개인정보가 있는지에 대하여 자동으로 식별하는 부분이 있는데, 데이터들이 비정형화되어있고, 또한 방대한 데이터를 실시간으로 분석해서 개인정보인지 여부를 판단할 때에 로드가 발생하기 때문에 기능 구현이 쉽지 않다. 본 논문에서는 수집되는 데이터 내에 개인정보 포함 여부를 판별하는 정보가 메시지에 포함되어 있다고 가정한다. 즉, 센서에서 플랫폼으로 오는 메시지에 데이터가 포함되어 있는데, 해당 데이터가 개인정보이기 때문에 비식별화가 적용되어야 한다는 정보가 함께 전달되는 것이다. 그렇게 되면 플랫폼에서는 해당 메시지를 받은 후 파싱과 같은 분석을 하면서 추가 정보들을 확인하고, 이 데이터는 비식별화를 진행해야 한다는 것을 판단하여 이러한 메타데이터를 실제 데이터와 함께 저장한다. 이후 누군가 해당 데이터를 요청할 때, 플랫폼에서는 해당 데이터가 개인정보라는 메타데이터를 가지고 있기 때문에 비식별화를 먼저 적용한 뒤에 비식별화된 정보를 그 누군가에게 보여주게 되는 것이다. 그림 1에서는 이와 같은 절차에 대해 설명하고 있다.



* 데이터에서 어떤 부분이 익명화가 필요한지, 어떤 알고리즘을 적용해서 익명화를 할 것인지, 원본을 볼 수 있는 사용자와 익명화된 정보에 대해서만 접근이 가능한 사용자가 누구인지에 대한 정보 등이 포함된다.

그림 1. IoT 플랫폼에서 데이터 저장 및 전달 절차

IoT 플랫폼은 개인정보가 포함된 데이터가 저장되고 사용될 때, 익명성을 제공하기 위해서 추가적으로 필요한 정보들을 플랫폼에 원본 데이터와 함께 저장하고, 상황에 맞추어 비식별화를 적용하게 된다. 이를 위해 IoT 플랫폼은 데이터를 수집한 후 데이터를 저장하기 전에 데이터를 처리하는 과정이 필요하다. 사물인터넷 플랫폼에서는 일반적으로 데이터를 접근이 가능한 식별자 (Uniform Resource Identifier-URI)를 가지는 리소스를 만들고 그 안에 각종 데이터 및 데이터의 속성을 저장하는 방식으로 데이터를 저장한다. 익명화를 수행하기 위한 각종 추가 데이터도 이와 같은 방식으로 사물인터넷 플랫폼에 저장될 수 있다. 익명화를 수행하는 데 필요한 데이터로는, 익명화 수행 데이터 표시, 익명화 대상 데이터, 익명화 알고리즘 등이 포함된다. 데이터가 저장된 뒤에 다른 IoT 애플리케이션에서 익명화 대상 데이터에 대해 접근을 요청할 경우, 추가적으로 저장된 익명화 관련 정보들을 토대로 사전에 정의된 익명화 알고리즘 등을 적용해서 익명화된 데이터를 제공하게 된다. 이러한 익명화 관련 메타데이터는 데이터를 포함하는 리소스의 추가적인 속성을 정의함으로써 구성할 수 있다. oneM2M 에서는 *contentInstance* 리소스에 실제 데이터가 저장되고, 해당 리소스는 실제 데이터를 저장하는 *content,* 데이터의 사이즈를 나타내는 *contentSize* 등의 속성을 가진다. 익명화 관련 추가 메타데이터들은 이러한 속성과 같이 *contentInstance* 리소스의 추가적인 attribute로 포함될 수 있다.

제안한 방식대로 플랫폼이 동작하기 위해서는 개인정보 레코드 중에서도 어떠한 정보가 범주화 또는 마스킹 되어야 할지 등의 추가 정보가 필요하다. 다양한 IoT 플랫폼이 존재하기 때문에 서로 다른 레코드를 비식별화 한다면 연결 공격에 취약점을 가지게 되기 때문이다. 위 사항을 고려한다면 더욱 효과적인 설계를 할 수 있을 것이다.

Ⅲ. 결 론

본 논문에서는 사용자의 익명성을 확보하기 위하여 어떻게 사물인터넷 플랫폼을 설계하고, 고려되어야 할 점이 무엇이 있을지를 확인하였다. 플랫폼에서 데이터를 수집하는 단계에서 개인정보가 포함되어 있는지 판단하고, 개인정보가 포함된 데이터에 대해서는 사용자에게 제공할 때 비식별화를 진행한 후 제공하는 방식이다. 하지만 현실적으로 적용하기에는 이미 설계가 완료된 다양한 IoT 플랫폼이 사용되고 있고 설계 시 고려해야 할 사항도 존재하기 때문에 구현에 제약이 있을 수 있다.

따라서 이미 구현된 플랫폼에서도 보다 효율적으로 익명성을 확보할 수 있는 솔루션을 개발해야 할 것이며, 더하여 특정 데이터에 개인정보가 포함되어 있는지 여부를 로딩 없이 효율적으로 판단할 수 있는 솔루션을 고안해야 할 것이다. 또한, 다양한 IoT 플랫폼에서 동일한 기준으로 익명화가 제공될 수 있도록 제도적 개정이 필요하다.

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 20190004260022002, IoT 기반 이식-침습형 고위험 의료장치를 위한 능동형 킬 스위치 및 바이오 마커 활용 방어 시스템 개발)

참 고 문 헌

[1] Seongeun H. "A study on Personal Information Control System for Information Protection in IoT Environment," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, February 2020.

[2] Simi M. S. "An Extensive Study on Data Anonymization Algorithms Based on K-Anonymity," IOP Conference Series: Materials Science and Engineering, August 2017.

[3] Machanavajjhala A. "L-Diversity: Privacy Beyond K-Anonymity," ACM Transactions on Knowledge Discovery from Data(TKDD), March 2007

[4] Ninghui L. "T-Closeness: Privacy Beyond K-Anonymity and L-Diversity," 2007 IEEE 23rd International Conference on Data Engineering, pp. 106-115.

[5] Aeri L. and Soomin S. "Improving Personal Data Protection in IoT Environments," Journal of The Korea Institute of Information Security & Cryptology, pp. 1000-1002.