[정보보호] ITU-T SG17 빅데이터 비식별화 표준화 동향

개요

텔레콤 산업체들은 클라우드환경에서(ITU-T Y.3600) 빅데이터 처리가 가능하도록 대량의 그리고 다양한 데이터로부터 대량의 값을 획득하도록 데이터 교환이 필요 하게 되었다. 이러한 환경에서 결과적으로 산업체는 서로 다른 데이터 흐름 시나리오들(예, 빅 데이터 교환)에서 데이터 비식별화 기술을 구현하는 것이 장려되고 있다. 빅 데이터 교환을 위한 데이터 흐름에 관여하는 관계자들 사이의 핵심은 데이터 비식별화가 수집 전이나, 후에 수행될 필요가 있는가 또는 빅데이터 교환의 다음 관계자와 공유하기 이전에 수행될 필요가 있느냐 하는 것이다. 다양한 빅데이터 교환 시나리오에 따라, 서비스 제공자는 적절히, 효과적으로 그리고 안전하게 빅데이터 서비스 고객에게 다음과 같은 고려사항을 통하여 비식별화가 제공될 필요가 있다.

a) 누가 텔레콤 서비스 제공자를 위하여 빅데이터 서비스 교환에 참여 하는가? 그리고 비식별화 서비스에서 그들의 역할이 무엇인가?
b) 무엇이 각 빅 데이터 교환 모델에서 텔레콤 서비스 제공자의 적절한 역할이 무엇인가?
c) 효율적인 비식별화 처리 서비스를 위하여 어떤 종류의 빅데이터 교환 시나리오가 선택될 수 있는가?
d) 빅데이터 교환 동안 각 시나리오나 모델에서 언제 비식별화 처리들이 수행되는 것이 좋은가 그리고 그러한 것에 어떤 리스크가 연계되어 있는가?
e) 텔레콤 서비스 제공자는 어떻게 비식별화 수준(level)을 선택하는가 그리고 그 수준에 어떤 리스크가 연계되어 있는가?
f) 텔레콤 서비스 제공자는 비식별화 처리가 적절히 수행되는지를 어떻게 확인하는가? 그렇지 않으면 텔레콤 서비스 제공자는 성능저하를 어떻게 조치 할 수 있는가?
g) 비식별화 처리로부터 나오는 데이터와 비식별화 서비스에 제공되는 데이터를 안전하게 전송 또는 저장을 위한 요구사항이 무엇인가?
h) 각 빅데이터 교환 모델에서 비식별화 처리의 요구사항이 무엇인가?
i) 비식별화 처리를 위하여 무엇이 텔레콤 서비스 제공자가 수행해야 하는 일반적인 절차와 방법인가?
j) 어떤 종류의 절차와 방법이 텔레콤 서비스 제공자가 각 빅데이터 교환 모델에서 비식별화 서비스를 제공하기 위해 선택할 수 있는가 그리고 그 것에 어떤 리스트가 연계되어 있는가?

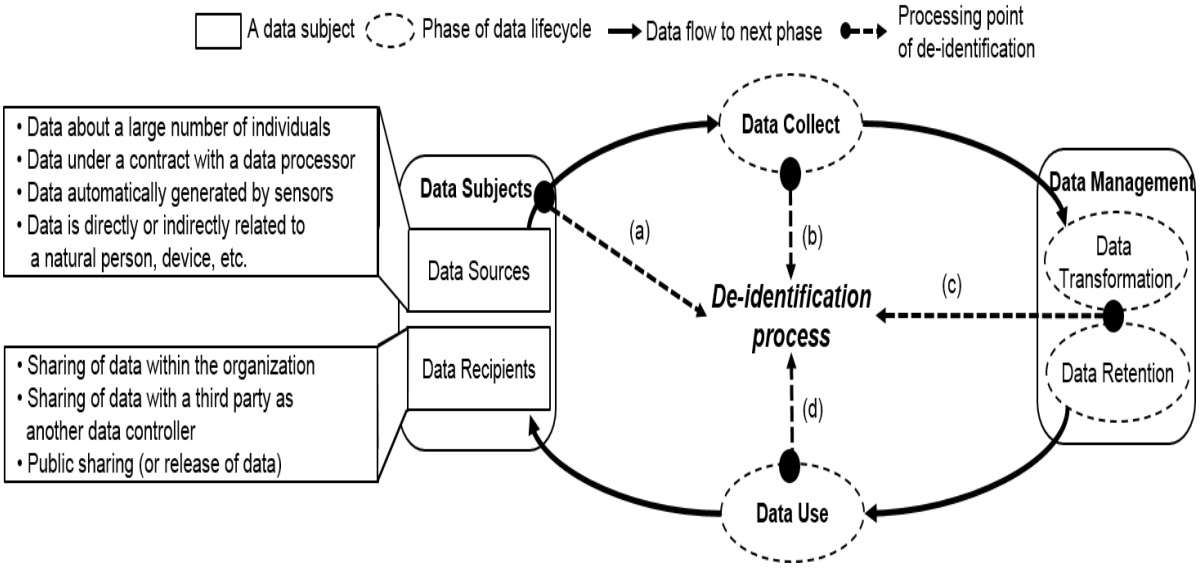
최근에는 표준화 단체의 빅데이터 표준 작업이 주로 클라우드 컴퓨팅과 IoT에 관련한 일반사항, 정의, 공통 요구사항, 유스케이스와 구조에 집중되어 있다. ITU-T SG17에서 진행중인 신규

아이템(X.fdip)은 텔레콤 산업의 특정적인 요구사항을 취급하지 않고 공통 요구사항과 솔루션을 취급하고 있다. 예로서, 모바일 텔레콤 환경에서는 이용자가 서로 다른 여러 개의 단말을 이용하고, 그리고 서로 다른 위치에서 서비스에 접속을 한다. 이러한 경우에, 오퍼레이터가 이용자로부터 수집하고 분석하는 빅데이터는 유선 오퍼레이터의 것과는 완전히 다른 것일 것이다. 전화 번호는 텔레콤 산업에서는 보통 이용자 식별번호로 사용된다. 이러한 것이 빅데이터 응용과 서로 다른 정보보호 이슈에 의하여 서로 다른 성질을 나타내고 있다.

ITU-T Y.Suppl.BigData-RoadMap은 텔레콤 서비스 제공자에 의한 정보보호, 데이터 보호, 익명화 그리고 식별 데이터의 비식별화를 포함하는 빅데이터 정보보호의 기술적 분야를 제안하였다. 현재 ITU-T SG17에서 개발하고 있는 빅데이터 정보보호 관련 표준 아이템은 고객의 데이터 정보보호와 클라우드 서비스를 위한 전반적인 정보보호 측면을 집중하고 있다. 이러한 표준 아이템은 빅데이터 교환 서비스를 위해 익명과 비식별화에 국한하기 보다는 전반적인 정보보호 측면을 취급하고 있으며, 신규 표준은(X.fdip, Framework of de-identification processing service for telecommunication service providers)은 빅데이터 교환 환경에서 비식별화를 위해 요구되는 지침을 개발한다.

비식별화 처리 개요

일반적으로 기관이 데이터 비식별화를 포함하여 영업의 니즈나 법과 규제와 일치하는 프라이버시나 정보보호 조치를 위한 목표를 설정한다. 기관이 비식별화를 활용한다면, 그것은 개인의 사적인 영역을 침해하지 않고 데이터 분석에 개인의 데이터를 이용할 수 있는 것이다. 신규 표준 아이템 X.fdip은 개인의 데이터 생명주기를 Data subjects{Data sources} → Data collect → Data management{Data analysis & Data storage} → Data usage → Data subjects{Data recipients}와 같이 정의 한다. 그림 1은 데이터 생명주기 모델에서 비식별화 처리의 개요를 보이고 있다. 이러한 데이터 생명주기는 데이터 비식별화로 예측되는 이득이나, 계획된(목표로 하는) 이용자, 프라이버시 위협 그리고 취약점들을 분석하는 지침으로 활용될 수 있다. 또한 데이터 생명주기 개념은 재식별화의 가능성에 대한 분석에서 적절한 제어를 선택하는 것에 활용 가능하다.



[그림] 데이터 흐름 환경에서 데이터 생명주기

데이터 생명주기 모델의 비식별화 프로세스

데이터는 데이터가 참조하는 데이터 주체(개인)로부터 수집된다. 수집된 개인 데이터는 개인 정보를 포함하는 데이터셋에 합치게 된다. 그리고 비식별화는 데이터를 식별할 수 없도록 새로운 데이터셋을 만든다. 이 데이터셋은 프라이버시 리스크를 줄이기 위하여 본래의 데이터셋을 이용하는 대신에 기관에 의하여 내부적으로 사용될 수 있다.

이 모델을 기반으로, 데이터 수집단계 (흐름 (b)) 또는 식별된 데이터가 수집되었지만 식별정보가 실제적으로 필요하지 않은 경우 (흐름 (a))에 비식별화가 실행될 수 있다. 즉, 데이터 관리 {Data Transformation & Data Retention} 에 위하여 필요하지 않는 식별자들은 수집할 필요가 없다. 그 대신에 비식별화는 식별정보 획득을 피하기 위하여 데이터 변환 후에 그리고 데이터 저장 전에 적용할 수 있다 (흐름 (c)).

완전히 식별된 데이터가 기관에서 필요하다면, 식별정보는 데이터가 데이터 사용(흐름 (d))을 위한 데이터셋으로 공개(배포)되기 전에 삭제되어야 한다. 그러면 이 데이터셋은 신뢰하는 데이터 수신자에게, 즉 데이터 사용 합의와 같은 추가적인 행정적인 제어와 연관된 수신자에게 제공될 수 있다. 또는 대신에 그 데이터는 예로 인터넷에 비식별화된 데이터를 공개하는 것과 같이 수많은 알려지지 않은 데이터 수신자들에게 이용 가능하게 공개 할 수 있다.

데이터 생명주기 모델을 기반으로 비식별화 프로세스를 적용하면, 프라이버시 리스크를 줄일 수 있고, 공공적 공개 프로세스를 쉽게 할 수 가 있다. 그러나 데이터 흐름에 관여하고 있는

참여자들간의 상호관계성으로 인하여 비식별화 프로세스 수행은 언제 할 것이냐에 영향을 줄 수 있다. 즉, 데이터 수집 전(흐름 (a)), 또는 데이터 수집 후(흐름 (b)), 또는 데이터 저장 전(흐름 (c)), 또는 데이터를 다음 참여자와 공유를 하기 이전(흐름 (d))에 수행될 수 있다. 이러한 결정은 순차적으로 정보보호의 타당성이나 타 기관의 비식별화 기술의 효율성을 향상하기 위한 조치에 영향을 주고 있다.

향후 전망

빅데이터 산업이 급속도로 상승하고 있는 상황에서 개인식별정보의 보호는 빅데이터 산업 육성에 저해요소로 작용하는 것으로 보는 것이 업계의 보편적 인식이다. 비식별화 기술은 이러한 애로점을 해소하기 위하여 데이터에 연관되어 있는 개인식별정보를 제거하여 빅데이터 서비스를 원활하게 운영하도록 환경을 변환시켜 주는 기술이다. 그러나 비식별화된 데이터로부터 다시 원 주체(사용자)를 재식별화하는 공격이 있어서 이에 대한 방어 방법도 같이 고려 하여야 한다. 현 아이템에 대하여 미국과 영국에서는 많은 관심을 갖고 있으며, 특히 비식별화 과정을 통하여 일어날 수 있는 프라이버시 이슈를 우려하여 면밀이 검토할 것을 요청하고 있는 실정이다. 또한 유사 표준으로 ISO/IEC JTC 1/SC27 WG5 에서는 20889 (Privacy enhancing data de-identification techniques) 문서가 2번째 CD로 진행하기로 4월회의에서 결의되었다. 이 표준은 미국의 MS(Microsoft)에서 지대한 노력을 기울여 개발을 하고 있는 표준이며, 향후 두 표준 간의 표준 범위와 적용에 관심이 집중되고 있다.

나재훈 (전자통신연구원 정보보호연구본부 전문위원, ITU-T SG17 WP4 부의장/Q.7 라포처, hnah@etri.re.kr)
임형진 (금융보안원, ITU-T SG17 에디터, hjlim@fsec.or.kr)