정 보 보 호 학 회 지
제 33 권 제6호, 2023. 12

개인정보 비식별화 현황 및 비식별 조치 가이드라인 보완 연구

손 지 민*, 신 민 호*

요 약

최근 AI와 로봇기술 등으로 개인정보를 포함한 데이터의 처리가 일상화됨에 따라 한국정부는 개인정보 비식별 조치 가이드라인 및 데이터 3법을 발표함으로써 개인정보 비식별화를 돕고자 하였다. 하지만 복잡한 비식별화 절차와 이의 효과에 대한 불명확함으로 기업들이 개인정보를 포함한 빅데이터의 활용에 어려움을 겪고, 동시에 시민단체나 소비자단체에서는 현 가이드라인에 따른 비식별화 절차가 개인정보를 보호하기에 충분하지 않다고 지적하고 있다. 본고에서는 비식별화 현황과 기술을 검토하고 현 가이드라인의 한계점을 보완 함으로써 데이터 활용 업체와 기관들의 정확한 비식별화를 돕고 빅데이터 활용의 활성화에 기여하고자 한다.

Ⅰ. 서 론

AI와 로봇기술 및 사물인터넷 등으로 인터넷과 정보통신기술의 사용이 일상화되고 IT기술과 타 산업간 융합이 발생하게 되면서 빅데이터 활용이 확산되었다. 특히 개인정보가 포함된 데이터의 활용가치가 증가하고 대규모의 데이터를 활용하는 과정에서 개인 식별 가능성이 높아지면서 개인정보 침해의 우려가 높아지고 있다. 이러한 우려로 데이터에 포함된 개인정보를 안전하게 보호하는 동시에 데이터를 효과적으로 활용할 수 있는 방안이 모색되어 왔다. 이를 위한 방법으로써 미국, 캐나다, 호주 등 많은 국가들이 공개되는 데이터에 대한 비식별화를 의무화하고 있다. 데이터의 비식별화에 대한 법률을 체계적으로 제정하여 비식별화된 정보는 더 이상 개인정보가 아님을 명시하였으며, 비식별화를 왜 수행해야 하는지에 대한 여러 문서들을 발간하여 국민을 상대로 비식별화의 필요성과 개인정보 보호수준에 대한 교육을 꾸준히 제공하고 있다.

2016년 대한민국 정부는 ‘비식별 조치 가이드라인’을 발표함으로써 개인정보가 포함된 자료를 활용하기 위해 취해야 하는 비식별화 절차와 방법들을 소개하여 비식별화를 활성화하고자 하였고, 2018년 데이터 3법 개정안이 발의되면서 데이터 산업이 활성화될 것으로 기대되고 있다. 데이터 3법에서 주목해야 할 점은 개인의 명시적 동의 없이 연구 목적으로 사용하기 위해서는 비식별 조치가 이뤄져야 한다는 것인데, 국내에서 데이터 활용을 위한 움직임으로 인해 활용되는 데이터의 품질도 크게 향상될 전망이다.

하지만 비식별화라는 기술의 복잡한 절차 및 규제로 인하여 많은 기업들이 데이터 활용에 어려움을 겪으며, 시민단체 및 소비자단체에서 현 비식별 가이드라인은 개인정보를 충분히 보호하지 못하고 있다는 지적으로 기업들의 움직임은 소극적으로 변할 수밖에 없었다. 법제화된 것이 아닌 ‘가이드라인’이라는 한계로 본격적인 데이터 활용은 시행령 등 구체적인 제도가 마련되어야 할 것이다.

현 비식별 조치 가이드라인에서는 정형데이터에 대하여 k-익명성 및 관련모델인 ℓ-다양성과 t-근접성을 기반으로 한 비식별 처리기법이 소개되었다. 하지만 데이터의 종류나 공개형태 및 공개 빈도에 따라서 k-익명성 모델이 적용될 수 없거나 k-익명성 모델만으로는 부족한 경우가 빈번히 존재한다.

본고에서는 국내외의 다양한 비식별 정보 활용사례를 분류하고 이에 따른 다양한 비식별화 모델 및 기법들을 정리하였다. 또한, 프라이버시 위협의 종류인 신

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다. (NRF-2021R1F1A1055324)
* 명지대학교 컴퓨터공학과 (대학원생, gml9785@naver.com, 교수, mhshin@mju.ac.kr)

분노출, 속성노출, 추론 노출, 소속 노출, 자취 노출 등을 분류하여 활용방안을 제시하였으며, 현 가이드라인의 적정성 평가를 새로운 기준으로 활용하여 적정성 평가 개선 방법을 제시하고 있다. 또한, 정보집합물 결합 절차로 임시대체키에 대한 보완점 및 정보집합물 결합 시 비식별화 조치 등에 관한 개선안을 제시하고 있다.

Ⅱ. 배 경

많은 나라들의 비식별화 정책과 국내 비식별화 정책 활용 사례가 무엇인지 등 비식별화 정책 및 기술에 대해 살펴보고자 한다.

2.1. 비식별화 정책

NIST 보고서에 따르면 비식별화는 “데이터셋과 정보주체 간 연계성을 제거하는 과정을 지칭하는 일반적인 용어”라고 정의하고 있으며 일반적으로 원본 데이터를 주어진 모델에 따라서 일방적으로 처리하여 비식별화된 데이터를 생성하는 기법을 말한다.

익명화는 비식별화와는 다른 용어로써 “식별 가능한 데이터 셋과 정보주체 간의 연계성을 제거하는 과정”이라고 정의하였다. 비식별화와 익명화는 같은 의미로 사용되기도 하지만 정확히 분류를 할 경우 비식별화는 비식별을 하는 과정을 말하고 익명화는 비식별을 수행하는 기법 중 하나로 특히 비식별화된 데이터로부터 원본 데이터를 복구시킬 수 없는 비식별화를 말한다.

가명처리는 “데이터 주체와의 연관성을 제거함과 동시에 하나 이상의 가명과 그 데이터 주체와 관련된 특성들을 연결시키는 비식별화 종류”로 정의하고 있으며, 가명화된 데이터를 잠재적으로 재식별 가능한 데이터로 간주하고 있다. 신용정보법 개정안에서는 “가명처리”를 추가정보를 사용하지 않고서는 특정 신용정보주체를 알아볼 수 없도록 개인신용정보를 처리하는 것을 의미한다고 명시되어 있다.

2.1.1. 국제 비식별화 정책 동향

미국 정부에서는 방대하고 다양한 데이터셋을 활용하기 위하여 다양한 시도를 해왔으며, 2016년에는 NIST에서 “De-Identifying Government Datasets“라는 제목으로 정부 데이터셋의 비식별화에 대한 내용의 특별 출판본을 발행하였다. 미국은 여러 분야에서 시민들이 반드시 정부에 각 개인의 특정 정보를 제공해야 하는 경우가 있으며, 이 때의 개인정보 보호는 필수적이다. 정부에 제공된 시민들의 정보들은 기밀이 유지되어야 하고, 공식 통계자료를 위해서만 사용되어야 한다. 각 정부 기관은 각자 보유하고 있는 데이터셋을 국민들이 사용할 수 있도록 공개할 경우, 반드시 비식별화를 수행하여 공개할 것을 규제하고 있다.[1]

캐나다는 그동안 여러 해에 걸쳐 비식별화에 대한 여러 문서들을 발간하였는데, 2016년 6월에 공개된 가이드라인은 정형데이터에 대한 비식별화에 대해 초점을 맞추고 있다. 여기서 정형데이터란 각 레코드가 정해진 속성에 해당되는 정보의 집합으로 이루어진 데이터를 말한다. 캐나다 가이드라인에서는 비식별화된 데이터셋은 더 이상 개인정보를 포함하지 않으므로, 비식별 정보를 활용 및 공개해도 개인 프라이버시가 침해되지 않는다고 명시적으로 기술하고 있다.

호주는 2013년 이후부터 비식별화와 관련된 문서들을 발간하였으며[2], 2017년 9월에는 ”DDF(The De-identification Decision-Masking Framework)“ 라는 제목의 신규 라이드라인을 발표하였다. 호주의 프라이버시 법에서는 비식별화된 정보는 개인을 더 이상 식별할 수 없거나 합리적인 수준으로 식별이 되지 않는 정보를 의미한다. 하지만 어떤 종류의 데이터가 어떤 상황에서는 항상 개인정보이고, 그렇지 않을 때는 비식별화된 정보인지 판단할 수 있는 명확한 기준을 제공하지는 않는다. 호주의 가이드라인에서는 합리적으로 식별이 가능한 지 판단하는 것은 데이터셋의 상황에 대한 고려가 필요하다고 지적하고 있다. 정보의 본질과 양, 누가 그 정보를 보유하고 있고 접근할 수 있는지, 접근이 가능한 다른 정보, 그리고 그 다른 정보를 이용하여 개인을 식별할 수 있는 실제 가능성 등의 항목을 고려하여 합리적으로 식별이 가능한 데이터셋이라고 판단될 경우에 비식별화된 정보가 아닌 개인정보라고 판단한다.

앞서 조사한 세 국가가 공식적으로 발표한 기술 문서 또는 가이드라인들을 분석한 결과, 이에 대한 공통점은 데이터셋의 비식별화에 대한 법률이 체계적으로 제정되어 있다는 것과 데이터셋에 대한 비식별화를 국가적으로 규제하기 위한 자국민들에게의 지식전달이다. 법적으로 비식별화를 공개하는 데이터셋에 대해 비식별화를 수행할 것을 규정하고 있고, 비식별화된

완료된 데이터는 더 이상 개인정보가 아님을 법적으로 명시하고 있으며, 왜 비식별화를 수행해야 하는지 등의 비식별화에 대한 여러 안내 문서들을 발간함으로써, 이를 통해 자국민들이 비식별화의 필요성을 점차 깨달을 수 있도록 해왔다.

2.1.2. 국내외 산업별 비식별 정보 활용 사례

앞서 해외 동향 분석을 통해 해외 주요국들이 이미 비식별화에 대한 법률은 물론이고, 비식별화가 여러 기업에서 정확한 목적으로 사용될 수 있도록 여러 가이드라인을 공표하고 체계를 갖추려고 하는 모습을 확인할 수 있었다.

2014년 Google은 “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response”라는 논문을 발표하였다[3]. RAPPOR은 사용자 소프트웨어의 사용 통계 내역 등을 수집하되 강력한 프라이버시를 보장하는 기술이다. 사용자의 소프트웨어로부터 사용 내역 통계 데이터 등을 수집할 때 수집 결과에 Random response를 통한 Noise를 추가시켜 각 사용자의 프라이버시는 유지하는 동시에 전체적인 수집 통계는 정확하게 유추할 수 있도록 했다. 2017년 4월에는 Google이 이어서 “Secured Aggregation”라는 논문을 발표하였는데, 이는 Google App인 Gboard에 사용자의 쿼리 검색 추천 기능을 학습시키고 이러한 학습 결과 및 사용 내역 등의 집계 데이터를 Google의 서버로 보내는 것이다. 이 때, 데이터들이 그대로 전송되는 게 아니라 사용 내역들에 대한 평균값들이 암호화되어 전송되기 때문에 상세 내역을 전혀 알 수 없는 평균값들에 대해서만 접근이 가능하게 된 것이다. 다음 사례는 2016년 6월 Apple이 Apple WWDC Keynote에서 사용자의 데이터를 보호하는 데에 Differential Privacy가 적용되었음을 발표하였는데, iOS 10의 새로운 메신저 앱의 이모티콘 교체 추천 기능에 전 세계의 아이폰 사용자들의 집계된 데이터가 사용된다. 많은 사람들이 ‘엉덩이’라는 단어 대신 복숭아 모양 이모티콘을 사용한다고 가정해보자. 다른 사용자들이 ‘엉덩이’라고 타이핑을 하면, 추천 이모티콘에 복숭아가 나타나게 된다. 이 때 사용되는 집계 데이터는 Differential Privacy[4]가 적용되어 ‘엉덩이’ 라는 단어에 대해 복숭아 이모티콘을 많은 사람들이 사용한다는 것은 알 수 있지만, 실제 누가 해당 이모티콘을 사용했었는지는 알 수 없다. 또한 예측 기능에도 사용자에게 예측 단어 추천 시, 로컬의 사용자 데이터로만 추천했지만 새로운 버전의 텍스트 예측 기능에는 Differential Privacy를 적용하여 전 세계 사용자들의 텍스트 데이터를 활용하여 단어 추천을 한다.

2019년 페이스북 AI 연구원들이 동영상에서 딥페이크를 찾아낼 수 있는 비식별화 기술을 개발하였다. “Live face de-identification in video”라는 논문을 발표하였는데, 이는 높은 프레임 속도에서 완전 자동 비디오 수정을 가능하게 하는 얼굴 비식별화 방법을 제안한다. 목표는 인식(포즈, 조명, 표정)을 고정시키면서 identity를 최대한 비상관시키는 것이다. 페이스북의 비식별화 기술 작동 방식은 출력 비디오의 페이스 디스크립터를 대상 이미지의 페이스 디스크립터와 거리를 두는데, 대상 이미지는 입력 비디오의 프레임을 기반으로 할 필요가 없으며, 이는 라이브 비디오에 적용할 수 있다. 연구진은 사람들의 높은 수준의 얼굴 이미지에 대해 훈련된 새로운 Feed-forward encoder-decoder 네트워크 아키텍쳐를 사용하였다. 또한 네트워크는 주어진 비디오나 주어진 ID에 대해 재교육될 필요가 없다는 점에서 글로벌하며, 시간이 지남에 따라 왜곡이 거의 없는 자연스러운 이미지 시퀀스를 생성한다고 설명했다.

Ⅲ. 비식별화 조치 가이드라인

3.1. 개인정보 비식별 조치 가이드라인 개요

2016년 대한민국 정부는 ‘비식별 조치 가이드라인’을 발표하며 비식별 조치 기준, 지원 및 관리체계에 관한 안내 기준을 정하였다[5]. 빅데이터, IoT 등 IT 융합기술 발전으로 데이터 이용 수요가 급증함에 따라 주요 선진국들은 데이터 산업 활성화를 위한 정책을 추진 중에 있다. 이에 빅데이터 활용에 필요한 비식별 조치 기준 및 절차, 방법 등을 구체적으로 안내하여 안전한 빅데이터 활용 기반 마련과 개인정보 보호 강화를 도모해야 한다.

2018년 11월 데이터산업 활성화를 위한 ‘데이터 3법’ 개정안이 발의됐다. 데이터 3법이란 개인정보보호법, 정보통신망법, 신용정보법을 말한다. 2019년 12월 4일 개정안이 국회를 통과했으며 이후 국회 본회의를 최종 통과하여 2020년 8월 5일부터 시행되고 있다. 데이터 3법에서 개인의 명시적 동의 없이 연구 목적으로

사용하기 위해서는 비식별 조치가 이뤄져야 한다는 것이 주목해야 할 점이며 데이터 3법 개정안에서는 데이터 활용을 위한 가명정보 개념 도입, 비식별 조치 후 산업적 통계 등 연구 목적 활용, 개인정보 보호위원회 격상 등이 법제화됐다. 기존에는 법제화된 것이 아닌 '가이드라인'이었기 때문에 한계가 있었고, 데이터 3법이 개정되기 이전에는 데이터 활용이 사실상 불가능했다. 이런 관점에서 가이드라인의 한계를 해결함으로써 데이터 활용의 기반이 마련될 것으로 보인다.

3.1.1. 비식별 조치 기준

본 가이드라인은 개인정보를 비식별 조치하여 이용 또는 제공하려는 사업자 등이 준수하여야 할 조치 기준을 제시한 것이며, 단계별 조치사항으로 사전 검토, 비식별 조치, 적정성 평가, 사후관리 4단계로 나뉜다.

사전 검토 단계에서는 개인정보에 해당하는지 여부를 검토 후, 개인정보가 아닌 것이 명백한 경우에 법적 규제 없이 자유롭게 활용한다.

개인정보에 해당한다고 판단되는 경우에는 데이터 셋에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체하는 등의 방법을 활용하며 개인을 알아볼 수 없도록 하는 비식별 조치기법을 적용한다. 개인 또는 개인과 관련된 사물에 고유하게 부여된 값 또는 이름인 식별자는 원칙적으로 삭제 조치하고, 개인과 관련된 정보로서 다른 정보와 쉽게 결합하는 경우 특정 개인을 알아볼 수도 있는 정보인 속성자 또한 데이터 이용 목적과 관련이 없는 경우에는 원칙적으로 삭제한다. 데이터 이용 목적상 반드시 필요한 식별자는 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 여러 가지 기법으로 비식별 조치 후 활용한다.

비식별 조치가 충분하지 않은 경우 공개 정보 등 다른 정보와의 결합 및 다양한 추론 기법 등을 통해 개인이 식별될 우려가 있기 때문에 개인정보 보호책임자 책임 하에 외부전문가가 참여하는 비식별 조치 적정성 평가단을 구성하여 개인식별 가능성에 대한 엄격한 평가가 필요하다. 적정성 평가 시 프라이버시 보호 모델 중 k-익명성을 활용하는데, k-익명성은 최소한의 평가 수단이며 필요시 추가적인 평가모델인 l-다양성, t-근접성을 활용한다. 적정성 평가는 기초자료 작성, 평가단 구성, 평가 수행, 추가 비식별 조치, 데이터 활용으로 나뉜다. 개인정보처리자는 적정성 평가에 필요한 데이터 명세, 비식별 조치 현황, 이용기관의 관리 수준 등의 기초자료를 작성하고, 3명 이상인 개인정보 보호책임자로 평가단을 구성한다. 평가단은 개인정보 처리자가 작성한 기초자료와 k-익명성 모델을 활용하여 비식별 조치 수준의 적정성을 평가하고, 개인정보처리자는 평가결과가 '부적정'인 경우 평가단의 의견을 반영하여 추가적인 비식별 조치를 수행한다. 비식별 조치가 적정하다고 평가 받은 경우에는 빅데이터 분석 등에 이용 또는 제공이 허용된다.

비식별 조치된 정보가 유출되는 경우에 다른 정보와 결합하여 식별될 우려가 있으므로 관리적 및 기술적 보호조치를 필수적으로 이행한다. 비식별 정보파일에 대한 관리 담당자 지정 및 비식별 조치 관련 정보 공유 금지 등의 관리적 보호조치, 비식별 정보파일에 대한 접근통제 및 보안 프로그램 운영 등의 조치가 필요한 기술적 보호조치가 있다. 또한 재식별 가능성을 열어두고 정기적으로 모니터링을 해야 한다.

3.1.2. 지원 및 관리체계

본 가이드라인은 비식별 조치를 통해 개인정보를 안전하게 활용할 수 있는 지원 및 관리체계를 제시하였다. 분야별 전문기관을 정하였으며, KISA에 개인정보 비식별 지원센터를 설치하여 운영을 개시하였다.

빅데이터 분석에 활용하기 위해 서로 다른 사업자가 보유하고 있는 정보집합물을 결합하는 경우 개인별로 부여된 식별자가 매칭키로 사용되는데, 정보집합물 간 결합·분석을 위해서는 결합 과정에서만 임시로 매칭키 역할을 하는 임시 대체키의 활용이 필요하다. 임시 대체키를 활용한 결합을 허용하는 경우에도 무분별한 결합을 통한 개인정보 침해 소지를 방지하기 위해 전문기관에서만 결합을 하도록 하는 등 지원 및 관리체계가 필요하다.

정보집합물을 결합 절차는 먼저 A사와 B사가 같은 알고리즘을 적용하여 식별자를 임시 대체키로 전환하고, 결합대상 정보집합물도 비식별 조치 및 적정성 평가를 수행한다. 그리고 비식별 조치된 정보를 전문기관에 제공하여 결합을 요청한다. 임시 대체키를 활용하여 전문기관에서 결합을 수행한 후 임시대체키를 삭제하고, 결합 데이터베이스를 필요한 기업에게 제공한다. 이 때 A사와 B사는 분야별 전문기관과 임시 대체

키 생성 알고리즘에 대한 정보공유를 금지해야 하며, 임시 대체키 생성을 위해 주민등록번호를 활용하는 것 또한 금지되어야 한다.

3.2. 비식별화 기술

3.2.1. 프라이버시 보호 모델 개요 및 속성 분류

마이크로데이터는 각 레코드가 한 개인의 정보를 담고 있는 정보집합물을 말하는데, 프라이버시 보호 모델은 개인정보를 제거하기 위해 비식별화 처리된 마이크로데이터에 대해 그 비식별화 조치가 소정의 목적을 달성했는지를 객관적으로 평가하는 기준을 제시한다. 그 객관적인 기준을 제시함으로써 공격자가 해당 데이터로부터 도출해 낼 수 있는 개인정보의 수준과 그 가능성을 제한하는 것이 프라이버시 보호모델의 목적이다.

마이크로데이터의 각 레코드는 다수의 속성으로 이루어져 있다. 그 중, 식별자는 그 값 자체로 한 개인을 유일하게 식별할 수 있는 속성을 말하는데, 예를 들어 주민등록번호, 운전면허번호, 휴대폰번호와 같이 한 개인에게만 부여되는 값이나 이름, 주소와 같이 한 개인을 거의 유일하게 식별할 수 있는 것으로 여겨지는 정보를 말한다. 미국의 HIPAA에서는 이러한 식별자는 반드시 제거되어야 하는 정보로 규정하고 있다. 준식별자는 그 자체로는 개인을 유일하게 식별할 수는 없으나 두 개 이상이 모이면 해당되는 개인의 수가 급격히 줄어들어 대상 데이터에서 유일하게 식별될 수도 있는 속성을 말한다. 공격자는 그 대상의 준식별자 정보를 활용해서 개인을 식별하려는 시도를 할 수 있다. 민감속성은 개인의 민감한 정보로써 타인에게 알려지는 경우 프라이버시를 심각하게 침해한 것으로 간주되는 정보를 말한다. 예를 들면 의학적 병명, 연 수입, 온라인 쇼핑내역 등을 공격자는 재식별 공격을 통해서 대상자의 민감속성을 알아내고자 하는 경우가 많다. 일반속성은 데이터에 포함되어 있는 속성 중 식별자, 준식별자, 민감속성에 포함되지 않는 나머지 속성을 의미한다. 예를 들면 데이터 관리자가 관리 목적으로 임의로 부여한 레코드번호나 데이터 수정날짜 등 일반속성은 개인과 전혀 무관한 정보이므로 개인 프라이버시와 무관하지만 데이터 분석에 필요한 정보가 아닌 경우가 대부분이어서 비식별화 과정에서는 삭제 대상이다.

3.2.2. 현행 프라이버시 보호 모델 소개 및 분석

현 가이드라인에서는 비식별 조치기법 중 프라이버시 보호모델 기법으로 k-익명성, 기본형 ℓ-다양성, t-근접성 기술을 소개하였다.

k-익명성은 1998년 Sweeney가 처음 학계에 소개한 후로 25년간 데이터 비식별화의 가장 정통적이고 효과적인 방법으로 알려져 왔다[6]. k-익명성은 신분노출 공격을 방어하고자 나온 개념이다. 즉, 마이크로데이터의 특정 레코드가 실제로 누구의 것인지를 100% 확률로 지적할 수 없도록 데이터를 수정하는 것이 목적이다. 하지만 공격자가 특정 공격대상자의 준식별자 값들을 알고 있다면, 이 값들을 데이터의 준식별자 값과 비교했을 때 공격대상자의 레코드일 가능성이 있는 후보들이 매우 작은 집합으로 좁혀진다.

준식별자가 같은 값으로 이루어져 있는 레코드들끼리 묶어 이를 동질집합(Equivalence Class)이라고 부르고, 모든 동질 집합이 적어도 k개 이상의 레코드를 포함하도록 함으로써 데이터에 포함된 모든 사람이 준식별자를 기준으로 다른(k-1) 사람들과 구분될 수 없도록 익명성을 보장하는 모델이다. 예를 들어, 어떤 데이터가 나이, 결혼여부, 성별의 준식별자 기준으로 k=3인 k-익명성을 보장한다면 이 데이터에 포함된 모든 사람은 자신과 같은 나이, 결혼여부, 성별 값을 가진 사람이 적어도 두 명 더 존재하기 때문에, 공격자가 설사 자신의 레코드를 지목하더라도 항상 그럴듯한 부인(plausible deniability)을 가능하게 해준다.

주어진 k값에 대한 k-익명성을 만족하지 않는 데이터는 해당 준식별자의 값을 수정하여 동질집합들의 크기를 증가시킬 수가 있다. 이에 사용되는 방법이 일반화(Generalization) 기법이다. 즉, 준식별자의 본래값의 의미를 유지하면서 그 값을 보다 일반적인 값으로 수정하면 서로 준식별자 값이 같은 레코드가 여럿이 생길 수가 있다. 그림 1은 각 동질집합의 크기가 2 이상인, 즉 k=2인 k-익명성을 만족하게 된다.

앞에서 소개한 바와 같이, k-익명성은 준식별자의 값을 일반화함으로써 달성할 수 있다. 숫자 값을 갖는 준식별자(예를 들어 혈압수치 “105”)의 일반화는 그 값이 포함되는 범위 값으로 일반화가 가능하며(예를 들어 “100이상 110미만”), 분류 값을 갖는 준식별자(예를 들어 병명 “위염”)의 일반화는 다수의 값을 통

Race	DOB	Gender	ZIP	Problem
black	1965/*/*	male	0214*	diabetic
black	1965/*/*	male	0214*	chest pain
black	1965/*/*	female	0213*	painful eye
black	1965/*/*	female	0213*	wheezing
black	1964/*/*	female	0213*	obesity
black	1964/*/*	female	0213*	chest pain
white	1964/*/*	male	0213*	short of breath
black	1965/*/*	female	0213*	hypertension
white	1964/*/*	male	0213*	obesity
white	1964/*/*	male	0213*	fever
white	1967/*/*	male	0213*	vomiting
white	1967/*/*	male	0213*	back pain

〔그림 1〕 k-익명성 실현 알고리즘 예시(k=2)

틀어 지칭하는 보다 일반화된 용어로 대체함으로써 일반화가 가능하다(예를 들어 “소화기관 장애”). 이 때, 각 준식별자의 일반화 수준을 미리 정해놓고 전략적으로 적용하게 되는데, 이 때 사용되는 기법이 VGH (Value Generalization Hierarchy) 기법이다. 이 기법은 미리 준식별자의 모든 값들을 조사한 후 각 일반화 단계에서 어떻게 일반화 할 지를 트리로서 표현해 놓는 방법이다. k-익명성은 신분노출을 방어하는 모델로써, 속성노출 공격에 대한 방어로는 한계가 있다. 예를 들어 그림 2의 데이터는 동질집합의 수가 2 이상인 k-익명성을 만족하지만, 첫 번째와 두 번째 동질집합은 동질집합 내 민감속성의 값이 같기 때문에 공격자는 동질집합 내 어떤 레코드가 공격대상자의 레코드인지는 모르지만(신분노출 실패) 공격대상자의 민감속성에 대해서는 100%의 확률로 알아낼 수 있다(속성노출 성공).

ℓ-다양성은 k-익명성이 동질집합의 민감속성 값의 다양성을 고려하지 않기 때문에 발생하는 한계를 극복하여 속성노출 공격에 대한 방어를 하기 위해서 제안되었다[7]. 현 가이드라인에서 소개되고 있는 기본형 ℓ-다양성은 각 동질집합의 민감속성 값들이 적어도 l개

Race	DOB	Gender	ZIP	Problem
black	1965	male	0214*	diabetic
black	1965	male	0214*	diabetic
black	1965	female	0213*	painful eye
black	1965	female	0213*	painful eye
black	1964	female	0213*	obesity
black	1964	female	0213*	obesity
black	1964	female	0213*	hypertension

〔그림 2〕 속성노출에 대한 k-익명성 모델의 한계 예시

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	1305*	≤ 40	*	Heart Disease
4	1305*	≤ 40	*	Viral Infection
9	1305*	≤ 40	*	Cancer
10	1305*	≤ 40	*	Cancer
5	1485*	> 40	*	Cancer
6	1485*	> 40	*	Heart Disease
7	1485*	> 40	*	Viral Infection
8	1485*	> 40	*	Viral Infection
2	1306*	≤ 40	*	Heart Disease
3	1306*	≤ 40	*	Viral Infection
11	1306*	≤ 40	*	Cancer
12	1306*	≤ 40	*	Cancer

〔그림 3〕 속성노출에 대한 ℓ-다양성의 방어(l=3)

의 다른 값들로 이루어져 있어야 한다. 그림 3의 데이터에서는 각 동질집합이 모두 세 가지 이상의 다른 민감속성 값을 갖고 있고 속성노출을 방어하고 있다. 하지만 기본형 ℓ-다양성은 민감속성의 다른 값들의 개수만 고려하고 그 값들의 균형은 고려하지 않기 때문에 민감속성의 분포를 고려한 공격에 대해서는 한계를 갖고 있다. 예를 들어, 만약 동질집합의 크기가 8이고 그 민감속성 값들이 {Heart Disease, Viral Infection, Cancer, Cancer, Cancer, Cancer, Cancer, Cancer}라고 하면, 동질집합 내 세 가지 다른 민감속성을 갖기 때문에 l=3이지만 8개의 레코드 중 6개의 레코드가 Cancer이기 때문에 공격자는 6/8=75%의 확률로 공격대상자의 민감속성을 알 수 있다. 즉, 민감속성 값들의 분포를 고려하지 않기 때문에 속성노출의 성공률이 높아졌다.

t-근접성은 공격자가 비식별화된 데이터 전체를 분석함으로써 얻는 공격대상의 민감속성에 대한 정보와 공격대상이 속한 동질집합을 분석함으로써 얻는 공격대상의 민감속성에 대한 정보가 크게 다르지 않게 함으로써 공격자에게 노출되는 민감속성에 대한 정보를 제한하는 것을 목표로 하고 있다[8]. 예를 들어 데이터 전체를 분석해 보면 평균 혈압이 150이지만 공격대상자가 속한 동질집합만 보았을 때 평균 혈압이 120 정도라면 공격자는 공격대상자가 데이터 내 다른 사람들보다 혈압이 낮을 거라는 추론을 할 수 있다. 이 노출의 성공 여부는 공격자가 공격대상자에 대한 배경지식이 없을 때 얻을 수 있는 정보인 전체 데이터의 민감속성 분포와 배경지식이 있을 때 얻을 수 있는 정보인 동질집합 내의 민감속성 분포의 상대적 차이에 의해

결정된다는 점이 핵심이다. 만약 공격대상자의 동질집합을 분석했을 때 평균 혈압이 120이므로 공격대상자의 혈압이 120 근처일 확률이 높다고 판단하면 이는 속성노출 공격에 해당된다. 하지만 이 정보의 가치는 공격자가 공격대상에 대한 배경지식이 없어도 어차피 알 수 있는 정보와 다를 때만 가치가 있다. 즉, 전체 데이터의 평균 혈압도 120이고 동질집합의 평균 혈압도 120이라면 공격자는 공격대상자에 대한 배경지식으로 인해 추가적으로 얻은 정보가 없기 때문에 추론노출 공격은 실패한 것으로 판단된다.

주어진 비식별 데이터의 추론노출에 대한 방어 여부는 민감속성의 전체 데이터에서의 분포(전체분포)와 동질집합에서의 분포(동질집합분포)의 차이에 의해서 결정된다. 두 분포가 유사할수록 추론노출에 대해서 안전하며, 두 분포 간의 유사성은 여러 가지 방식으로 계산될 수 있는데, t-근접성 모델은 EMD(Earth Moving Distance)라는 방법을 채택하고 있다. 두 분포 간의 유사성을 측정하는 기법인 EMD는 한 확률분포를 그 모양대로 쌓인 모래더미로 비유하여 두 분포 간의 거리를 한 모래더미에서 다른 모래더미로 바꾸기 위해서 필요한 모래의 이동량으로 측정한다. 모래의 이동량 계산은, 가장 최소의 움직임으로 모래를 이동할 수 있는 최적의 움직임을 Linear Programming 기법으로 계산하고, 이 때의 최적의 비용을 정규화(Normalize)해서 계산한다. 비식별화 데이터가 t-근접성을 만족시키려면 모든 동질집합의 동질집합분포와 전체분포의 EMD 거리가 미리 설정한 t값보다 작거나 같아야 한다.

t-근접성 모델은 비식별화의 이상적인 목표를 제시하고 있으나, 실질적으로 t-근접성을 만족시키는 비식별화를 수행하는 것이 기술적으로 어려울 때가 많고, 적절한 t값에 대한 이론적인 뒷받침이 부족한 상태이다. 특정 t값이 공격자가 민감속성에 대해서 추가적으로 얻는 정보의 양과 어떤 관련이 있는지에 대한 기준이 부재한 점이 t-근접성의 이론적 한계로 여겨진다.

3.2.3. 신규 프라이버시 보호 모델 소개 및 분석

ℓ-다양성은 기본형, 엔트로피형, 재귀형 이 세 가지 실현방법이 있는데, 기본형 ℓ-다양성은 현 가이드라인에서 소개하였다. 이 절에서는 엔트로피와 재귀형, δ-노출, δ-소속 등 현 가이드라인에서는 다루지 않았으나 기술적으로 중요한 모델을 검토하고자 한다.

기본형 ℓ-다양성은 동질집합의 “다양성 수준을 그 동질집합에 포함된 서로 다른 민감속성 값들의 개수로 측정하지만, 단순히 다른 값들의 개수로 다양성을 판단하는 데는 한계가 있다. 예를 들어 열 개의 레코드 중 5개가 "교육직”이라는 민감속성을 갖고 있고 나머지 5개가 "공무원"이라는 값을 가질 때와 열 개의 레코드 중 9개가 "교육직"이고 1개가 "공무원"인 경우를 생각해 보자. 두 경우 모두 다른 민감속성의 개수는 2개로서 ℓ=2인 ℓ-다양성을 만족하고 있다. 하지만 후자의 경우는 동질집합의 레코드들 중 90%가 "교육직"이므로 공격자는 매우 높은 성공률로 공격대상자가 "교육직"이라고 추측할 수가 있다. 반면에 전자의 경우에 공격자는 공격대상이 "교육직"인지 "공무원"인지 어느 하나도 더 유리한 확률로 추측할 수가 없기 때문에 더 높은 비식별 수준을 유지한다고 해석된다. 따라서 "정보의 다양성"을 측정할 수 있는 고도화된 측정방법이 필요하게 된다.

엔트로피는 확률변수의 불확실성을 표현하는 수학적 지표이다. 불확실성이 전혀 없는 경우, 예를 들어 열 개의 레코드가 모두 "교육직”인 경우는 엔트로피 값은 0이다. 반대로 불확실성이 가장 높은 경우는 열 개의 레코드 중 5개가 "교육직"이고 나머지 5개가 "공무원”으로 균등한 분포를 가질 때, 엔트로피는 가장 커지고 log(2)=1로 계산된다. 그 외의 경우는 모두 엔트로피 값이 0과 1사이의 값으로 결정된다. 엔트로피 ℓ-다양성은 모든 동질집합의 엔트로피가 log(ℓ)보다 커야한다는 조건을 제시하고 있다. 예를 들어 ℓ=4인 경우는 각 동질집합이 4개 이상의 다른 값을 포함하고 있어야 할 뿐만 아니라, 엔트로피 값이 log(4)=2 이상이어야 한다. 다른 값의 개수가 4 미만이면 무조건 조건을 만족할 수 없다. 만약 다른 값의 개수가 4이면 반드시 네 가지 값이 균등하게 나타나야 한다. 다른 값의 개수가 5 이상이어도 적당히 균등하게 값들이 나타나지 않으면 엔트로피값이 2가 안 될 수 있다. 이 경우는 ℓ=4인 엔트로피 ℓ-다양성을 만족할 수 없다.

엔트로피형 ℓ-다양성은 각 동질집합의 엔트로피가 log(ℓ) 이상이어야 하는데, 이를 위해서는 전체 테이블의 해당 민감속성 값이 log(ℓ)이상의 엔트로피를 가져야 한다. 그 이유는 한 집합의 부분집합의 엔트로피는 항상 그 모집합의 부분집합보다 클 수가 없기 때문이다. 하지만 전체 데이터의 민감속성이 균등하지 않다

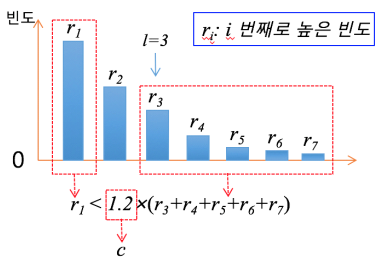
면 엔트로피가 log(ℓ)이 안 될 수도 있고, 이 경우 이 기법을 적용하기에는 무리가 있을 수 있다. 따라서 이 기법을 적용하기 전에 전체 테이블의 엔트로피를 먼저 계산해 보고 적용 여부를 판단하여야 한다.

재귀형 ℓ-다양성은 엔트로피형 ℓ-다양성의 엄격한 조건 때문에 이를 만족하는 데이터를 생성하기가 어려운 점을 해결하기 위해서 제안된 모델로써 민감속성의 다양성 조건을 민감속성 값 간의 비율에 대한 조건으로 표현하고 있다. 그리고 각 동질집합이 기본형 조건(ℓ개 이상의 다른 값)에 추가적으로 가장 빈도수가 높은 값의 빈도(r_1)가 ℓ번째로 빈도수가 높은 값(r_ℓ)과 그 이하로 빈도수가 높은 모든 값들의 빈도를 합친 값보다 c배 미만이면 (c, ℓ)-다양성을 만족시키는 것으로 정의한다.

그림 4에서 ℓ=3인 경우 한 동질집합에 포함된 민감속성을 빈번한 순으로 나열하면 가장 빈번한 민감속성의 포함 횟수가 ℓ번째보다 낮거나 같은 빈도로 포함된 횟수들을 모두 더한 값에 c를 곱한 값보다 작아야 한다. 재귀형 ℓ-다양성도 엔트로피형과 비슷한 이유로 쏠림현상이 발생할 수 있고, 따라서 추론노출에 대해서 취약하다.

δ-노출은 t-근접성과 같은 맥락 하에서 프라이버시 모델을 제시하고 있다. 즉, 동질집합 내 민감속성의 분포가 전체데이터의 분포와 동일할수록 프라이버시 수준이 높은 것으로 간주된다. 하지만 t-근접성이 전체 분포와 동질집합의 분포를 분포 대 분포로서 비교하는 반면 δ-노출은 각 민감속성 값의 동질집합 내 비율과 전체 데이터에서의 비율의 유사성을 비교함으로서 간접적으로 분포간의 비교를 하고 있다. 모든 동질집합이 각 민감속성 값 s에 대해서는 그림 5을 만족하면 된다.

δ-노출의 이러한 선택은 몇 가지 이점을 염두에 두고 있다. 첫 번째 이점은 계산이 t-근접성보다 훨씬 간



〔그림 4〕 재귀형 ℓ-다양성의 정의

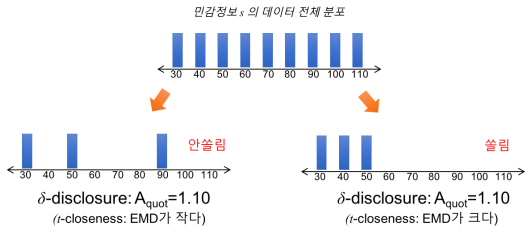
$$\left|\log \frac{p(\langle t \rangle, s)}{p(T, s)}\right| < \delta$$

〔그림 5〕 δ-노출 정의

단하다. 두 번째 이점은 제안된 계산식이 정보이론 관점에서의 공격자의 정보획득과 관련이 있기 때문에 경험적 근거로만 접근해야 하는 t-근접성의 t값 설정에 비해서 δ-노출에서의 δ값 설정은 의미론적으로 유의한 접근방식이다.

하지만 δ-노출은 t-근접성과는 달리 각 민감속성 값 간의 거리에 대한 고려 없이 단순 비율검색에만 의존하기 때문에 쏠림현상을 해결할 수가 없다. 그림 6에서 민감속성의 분포가 전체 데이터에서는 30에서 110까지 골고루 균등하다고 하자. 이 때 왼쪽 동질집합의 경우는 EMD가 작고 골고루 퍼져 있으나 오른쪽 동질집합은 값들이 모두 40 주변에 몰려있어서 EMD값이 크고 공격자가 공격대상자의 민감속성이 40 근처일 거라고 추론하는 공격에 취약하다. 하지만 δ-노출의 δ값은 1.10으로 두 경우가 같게 나타나고 있다.

δ-소속은 소속노출 공격을 막기 위한 모델이다. 개인이 특정 데이터에의 포함여부 자체가 그 개인의 프라이버시에 심각한 영향을 미칠 수 있는 경우에 그 포함여부 사실을 공격자가 알아내지 못하도록 하는 데 그 목적이 있다. 예를 들어 치매 환자 데이터가 비식별화 되어 공개가 되었을 때 공격자가 자신의 공격 대상이 그 데이터의 어떤 레코드에 해당이 되든 상관없이 그 데이터 어딘가에 포함되어 있는지 여부만 알면 공격이 성립한다.



〔그림 6〕 δ-노출의 쏠림현상

3.3. 유용도와 보호수준

데이터 비식별화는 데이터를 타인에게 공유하기 전에 데이터에 포함된 개인정보를 보호하기 위해서 수행

되는 작업이다. 하지만 데이터 공유의 목적은 데이터가 담고 있는 정보를 제공하는 것이기 때문에 프라이버시의 보호와 동시에 비식별화를 통한 정보의 손실을 줄여야 하는 상반된 목표를 달성하여야 한다. 데이터 사용자가 데이터를 얼마나 유용하게 사용할 수 있는지를 측정하는 척도를 유용도(Utility)라고 한다. 현 가이드라인에는 유용도에 대한 언급은 없다.

비식별 데이터의 유용도는 상대적인 유용도만 의미가 있다. 원본데이터 고유의 유틸리티(절대적 유용도)보다는 원본데이터의 유용도 대비 비식별 후의 유용도를 비교하는 데에 그 의의가 있다. 비식별 처리에 의해서 감소된 유용도를 측정하려면 비식별 처리 알고리즘을 기반으로 유용도 측정방법을 결정하여야 한다. 예를 들어 k-익명화를 통한 유용도 감소를 측정하기 위해서는 k-익명화가 수행하는 그룹화 수준, 또는 그룹화의 결과물인 동질집합을 기반으로 유용도를 측정해야 한다. 비식별화 기법 후의 유용도 측정방법을 논하고 프라이버시와 유용도의 관계에 대해서 살펴보기로 하자.

3.3.1. 유용도 측정

비식별화 기법 후의 감소된 유용도를 측정하는 방법에는 일반화에 의한 정보손실, 분별력, 평균동질집합크기가 있다.

정보손실 척도는 준식별자의 일반화에 의해 야기된 정보의 손실을 각 속성의 테이블 전체에서의 값의 범위 대비(예: 20살~60살) 각 레코드의 준식별자 값의 범위(예: 20대)의 평균비율로 측정하고 있다. 레코드의 값의 범위는 항상 전체 테이블에서의 범위보다 작거나 같기 때문에 이 척도는 항상 0부터 1까지의 실수 값을 갖는데, 0에 가까울수록 일반화가 많이 안 된 상태로 유용성이 증가하고, 1에 가까울수록 일반화가 많이 되어 유용성이 떨어짐을 표시한다.

비식별화된 데이터 T*의 일반화 정보손실 GenILoss(T*)는 모든 레코드(j=1, …, |T|)의 모든 준식별속성(i=1, …, n)에 대해서 그 속성의 전체범위(Ui-Li) 대비 일반화된 범위(Uij-Lij)의 비율을 모두 더한 후 정규화시킨다(0과 1 사이의 값을 갖도록 한다). 계산방식은 그림 7과 같다[9].

$$GenILoss(T^*) = \frac{1}{|T| \cdot n} \times \sum_{i=1}^{n} \sum_{j=1}^{|T|} \frac{U_{ij} - L_{ij}}{U_i - L_i}$$

〔그림 7〕 일반화 정보 손실

분별력 척도는 각 레코드가 다른 레코드들과 구분되는 정도를 나타내기 위해서 그 레코드가 포함된 동질집합의 크기가 클수록 비유용성 점수를 높이 부과하는 척도이다[10]. 분별력 척도 값이 낮을수록 유용도가 높은 것으로 해석이 되며 분별력의 정의는 그림 8과 같다. 원본데이터 T에 대한 비식별 조치를 취한 데이터 T*의 분별력 DM(T*)는 각 동질집합 크기의 제곱을 더한다. 만약 동질집합의 크기가 k보다 작아서 제외(suppression)된 레코드가 있다면 각 동질집합의 크기에 원본데이터 크기 |T|를 곱한 값을 추가한다.

$$DM(T^*) = \sum_{\forall EQ s.t. |EQ| \geq k} |EQ|^2 + \sum_{\forall EQ s.t. |EQ| < k} |T| \cdot |EQ|$$

〔그림 8〕 분별력 척도

평균동질집합크기의 척도는 모든 동질집합 크기가 k인 경우에 가깝게 비식별이 되었는지를 판단하는 척도이며 그림 9와 같이 정의된다. 원본데이터 T를 비식별화한 데이터 T*의 평균동질집합크기 척도 CAVG(T*)는 평균 동질집합 크기 |T|/|EQs|, 즉 전체 레코드 수를 동질집합 개수로 나눈 값을 k로 나눈 값으로 정의한다. CAVG(T*)가 1이면 가장 이상적인 비식별화에 가까우며 1보다 크면 클수록 유용도가 떨어진다. k-익명성이 적용된 경우 CAVG값은 항상 1보다 크거나 같다. 동질집합 크기가 작아서 제외(supression)된 데이터가 있는 경우는 평균 동질집합 크기가 커지는 영향을 준다.

$$C_{AVG}(T^*) = \frac{|T|}{|EQs| \cdot k}$$

〔그림 9〕 평균동질집합 크기 공식

3.3.2. 보호수준 측정

데이터 공유를 위해 정보의 손실을 줄이는 것도 중요하지만 반대로 프라이버시를 보호하는 것 또한 비식

별화의 중요한 목적이다. 프라이버시를 보호하기 위해서는 공개된 데이터가 재식별 될 위험성이 낮아야 한다. 재식별 위험도를 낮추기 위해서는 앞서 소개된 유용도를 낮추어 프라이버시를 강화할 수 있다.

주어진 데이터의 프라이버시 수준은 방어하고자 하는 공격에 따라 다르게 측정이 되어야 한다. 표 1은 공개된 데이터에 대한 공격자의 프라이버시 침해 위협을 분류한다.

신분노출은 공격자가 데이터 내 레코드 중 특정 레코드가 공격대상자의 것임을 밝혀내는 위협을 말한다. 이 때 공격자는 공격대상자에 관한 배경지식을 바탕으로 각 레코드의 준식별자가 자신이 알고 있는 배경지식과 일치하는지를 판단하여 이에 부합하는 레코드들을 모은 후 이 중 하나를 무작위로 추출하여 공격대상자의 것으로 추측하는 공격이 최선의 공격이다. k-익명성을 비롯한 모든 그룹핑을 통한 비식별화 기술은 준식별자가 동일한 그룹으로 레코드들을 묶어 동질집합이라고 칭하는데 결국 한 동질집합에 속해있는 개인들에 대한 신분노출 확률은 그 동질집합의 크기 분의 1로 계산할 수 있다. 하지만 각 동질집합의 크기가 다를 수 있어 데이터에 속한 개인들의 신분노출 확률은 각기 다를 수 있다. 따라서 한 정보집합물의 프라이버시 위협 수준은 그 정보집합물에 속한 모든 개인이 갖는 프라이버시 위협 수준 중 가장 큰 값으로 정의하는 것이 합리적이다. 이러한 정의는 정보집합물을 공개 혹은 공유하는 업체는 그 정보집합물에 포함된 모든 개인이 안심할 수 있는 수준의 프라이버시 수준을 보장할 의무가 있으며 어느 한 개인의 프라이버시도 포기할 수 없음을 의미한다. 이는 그림 10과 같은 정의가 성립된다. 이 정의는 모든 종류의 프라이버시 위협 수준에 대해서 적용하며 3.4절에서 논의되는 가이드라인 개선안에서도 일관성 있게 활용되고 있다. 이 정의에 입각하여 한 데이터의 신분노출 수준을 그림 11와 같이 정의한다.

〔표 1〕 프라이버시 침해 분류

종류	내용
신분 노출 (Identity Disclosure)	특정 공격대상이 주어진 마이크로데이터의 레코드들 중 어떤 데이터에 해당되는 지를 알아내거나 마이크로데이터의 특정 레코드가 누구의 데이터인지 알아내는 위험
속성 노출 (Attribute Disclosure)	공격대상의 민감한 정보를 주어진 마이크로데이터에서 직접적으로 알아내는 위험
추론 노출 (Inferential Disclosure)	공격대상의 민감한 정보를 주어진 마이크로데이터에 있는 정보를 바탕으로 간접적으로 추론하는 위험
소속 노출	공격대상이 마이크로데이터에 포함되어 있는지 여부를 알아내는 위험으로, 특정 마이크로데이터의 포함되는 사실 자체가 그 대상자의 민감한 정보를 노출하는 경우 성립
자취 노출 (Trace Disclosure)	각 개인에 대한 정보가 시기별로 다수 공개되는 경우 한 개인의 정보를 서로 연결하여 그 자취를 추적하는 위험으로, 예를 들어 주기적인 위치정보, 시간에 따른 환자상태 정보 등

(정의) 한 정보집합물의 프라이버시 위협 수준은 그 정보집합물에 속한 모든 개인이 갖는 프라이버시 위협 수준 중 가장 큰 값으로 정의한다.

〔그림 10〕 정보집합물의 최소 프라이버시 수준 원칙

$$\text{신분노출 수준} = \max_{EQ_i \in T^*} \frac{1}{|EQ_i|}$$

- T^*: 비식별정보집합물
- EQ_i : i번째 동질집합

〔그림 11〕 신분노출 수준 정의

속성노출은 공격자가 자신의 공격대상자가 속한 동질집합 내 민감속성 값들 중 어떤 것이 공격대상자의 것인지를 알아내는 위협을 말한다. 이를 위해서 공격자의 유일한 전략은 동질집합 내 가장 많이 존재하는 민감속성 값이 공격대상자의 것이라고 추측하는 것이다. 따라서 한 동질집합의 속성노출 수준은 그 동질집합에 속한 민감속성 값들 중 가장 빈번히 나타나는 값의 비율로 계산할 수 있고 정보집합물의 속성노출 수준은 이 값들의 최대값으로 측정한다. 속성노출 수준 정의는 그림 12과 같이 정의한다.

$$\text{속성노출 수준} = \max_{EQ_i \in T^*} \max_{s \in EQ_i} \left(\frac{freq_{EQ_i}(s)}{|EQ_i|} \right)$$

- T^*: 비식별 테이블
- EQ_i : i번째 동질집합
- $freq_{EQ_i}(s)$: EQ_i에 속한 s값의 갯수

〔그림 12〕 속성노출 수준 정의

추론노출은 공격자가 공격대상자에 대한 배경지식(준식별자 값)을 가정하지 않았을 때 비식별 데이터로부터 얻을 수 있는 공격대상자에 대한 정보를 기준으로, 공격대상자에 대한 배경지식을 알고 있을 때 공격자가 추가적으로 얻을 수 있는 정보로 인한 위협을 말한다. 예를 들어 공격자는 공격대상자의 혈당 수준에 대해서 알고 싶고 비식별 데이터의 전체 통계에 의하면 환자의 평균 혈당 수준이 75mg/dL이라고 가정하자. 이후에 공격자가 공격대상자의 준식별자를 알게 되었고 이를 바탕으로 공격대상자가 속한 동질집합을 알아낸 후 그 동질집합에 속한 환자들의 혈당수준을 평균을 내보니 90mg/dL이라고 가정하자. 이 경우 공격자는 자신의 공격대상자의 혈당수준이 다른 환자들에 비해서 높은 혈당을 갖고 있다는 정보를 배경지식으로부터 추가적으로 얻을 수 있다. 하지만 만약 데이터 전체 환자의 평균 혈당도 90mg/dL이었다면 공격자의 배경지식은 추가적인 정보를 제공하지 않게 된다. 이에 착안한 프라이버시 모델이 t-근접성이고 k-익명성 연관 모델 중 정보이론 측면에서 프라이버시 수준에 대한 가장 근본적인 접근방식을 보여주고 있다. 추론노출에서 추론되는 민감속성의 값은 반드시 동질집합 내에서 보여지는 민감속성의 값이 아닐 수도 있다는 점이 속성노출과의 차이점 중 하나이다. 따라서 이 절에서는 주어진 데이터의 추론노출 수준을 t-근접성에서 취하고 있는 계산방식을 따라 그림 13과 같이 정의한다.

$$\text{추론노출 수준} = \max_{EQ_i \in T^*} EMD(EQ_i, T^*)$$

- T^*: 비식별 데이터
- EQ_i : i번째 동질집합
- $EMD(EQ_i, T^*)$: EQ_i내 민감속성과 T^*내 민감속성 간 EMD거리

〔그림 13〕 추론노출 수준 정의

2016년 가이드라인 발표 후 산업계 및 관련 단체에서 현행 비식별화 정책 및 가이드라인 내용에 대한 다양한 의견을 제시하였다. 본고에서는 가이드라인 개선사항을 표 2과 같이 정리하여 명시하였다.

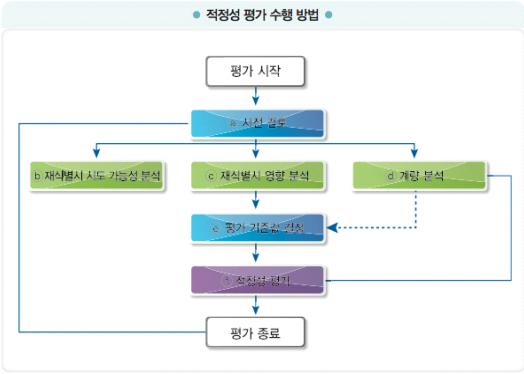
〔표 2〕 가이드라인 개선사항 개요

주제	현행	개선안
속성의 분류 세분화	식별자와 속성자	- 식별자, 준식별자, 민감속성, 일반속성 - 각 조치기준 제시
프라이버시 보호모델 보완	k-익명성, ℓ-다양성 (기본형), t-근접성	- 기존모델 정의 보완 - 엔트로피형, 재귀형 ℓ-다양성 추가 - δ-노출, δ-소속 모델 추가
적정성 평가 재정비		- 적용기법의 다양화 - 평가절차의 완화 - 평가기준 산정 재정비 - 단계적 적용방안 제시
적정성 평가시 평가기준의 다양화	k-익명성 ℓ-다양성, t-근접성	- 데이터 분류에 따라서 다양한 모델 적용 가능
적정성 평가기준 도출방식 정량화	정성적 평가방식, 주관적 기준설정	- 요소분석방식을 이용한 정량적 평가 - 객관적 기준설정 - 재식별 위험도는 재식별 시도 가능성, 재식별 시 영향, 소속노출 가능성에 의해 측정 - 재식별 위험도에 따라서 소속노출 방어기준, 신분노출 방어기준, 속성노출 방어기준, 추론노출 방어기준을 선정
데이터 보호수준 도출방식 정의	모델에 입각한 계량	- 프라이버시 보호목표에 따른 계량 - 소속노출: 모 데이터 기준으로 계량 - 신분노출:소속노출+1/k - 속성노출:소속노출+민감속성 분포 - 추론노출:소속노출+t-근접성

3.4. 비식별화 적정성 평가

3.4.1. 현 가이드라인의 적정성 평가 개요

현 가이드라인은 그림 14와 같은 절차로 적정성 평가를 수행한다. 현 가이드라인에 의하면, 평가대상자는 기초자료를 평가단에 제공하고, 평가단은 이 기초자료와 평가대상 비식별 데이터를 바탕으로 재식별 시도



〔그림 14〕 현 가이드라인 적정성 평가 절차

가능성 분석, 재식별시 영향 분석, 계량분석을 수행한 후, 평가 기준값을 결정하여 적정성 평가를 수행한다.

재식별 시도 가능성은 “재식별 의도 및 능력 분석 평가 지표”와 “개인정보 보호 수준 평가 지표”를 작성하여 결정하게 되는데, “재식별 의도 및 능력 분석 평가 지표”의 평가단 평균 결과 값(0~9점)을 기준으로 “낮음”(0~2점), “중간”(3~4점), “높음”(5~9점)으로 분류하고, “개인정보 보호 수준 평가 지표”의 평가단 평균결과 값(0~9점)을 기준으로 “없음”(일반공개), “낮음”(0~3점), “중간”(4~5점), “높음”(6~9점)으로 분류한 후 그림 15를 기준으로 재식별 시도 가능성을 분류하게 된다.

또한 “재식별시 영향분석 평가지표”(0~4점)를 통해서 “낮음”(0점), “중간”(1점), “높음”(2~4점)으로 분류한다.

계량분석에서는 비식별 데이터를 분석하여 데이터가 만족하고 있는 k값, ℓ값, t값을 계산한다.

평가 기준값은 위에서 도출한 “재식별 시도 가능성”(거의 없는, 가끔, 가능한, 빈번한)과 “재식별시 영향”(침해위험 낮음, 중간, 높음)을 바탕으로 평가단이 논의 후 결정하게 되는데 그림 16을 참고 기준값으로

● 재식별 시도 가능성 분석표 ●

2)개인정보 보호 수준				
없음	빈번한	빈번한	빈번한	
낮음	가능한	가능한	빈번한	
중간	가끔	가끔	가능한	
높음	거의 없는	거의 없는	가끔	
	낮음	중간	높음	1) 재식별 의도 및 능력

〔그림 15〕 재식별 시도 가능성 분석표

● 평가 기준 값 사례 *●

재식별시 영향					
침해위험 높음	k = 5 ℓ = 2	k = 10 ℓ = 3	k = 10 ℓ = 4	k = 20 ℓ = 5 t < 0.3	
침해위험 중간	k = 3 ℓ = 2	k = 5 ℓ = 2	k = 10 ℓ = 3	k = 10 ℓ = 4	
침해위험 낮음	k = 3 ℓ = 2	k = 5 ℓ = 2	k = 5 ℓ = 2	k = 10 ℓ = 3	
	거의 없는	가끔	가능한	빈번한	재식별 시도 가능성

〔그림 16〕 평가 기준값 사례

제시하고 있다.

평가단은 이와 같이 기준이 되는 k값과 ℓ값을 선정한 후 계량분석의 결과를 비교하여 비식별화 적정성 판단을 한다.

3.4.2. 현 가이드라인의 적정성 평가 한계점

현 가이드라인은 적정성 평가 기준을 도출하기 위해서 ‘재식별 시도 가능성’과 ‘재식별 시 영향도’를 측정하고 있다. 하지만 이를 바탕으로 도출하는 목표 프라이버시 수준이 평가단의 논의에 의해 자의적으로 결정되기 때문에 평가의 일관성과 객관성을 달성하기가 어려울 수 있다. 그리고 이러한 주관적 판단시점이 전체 평가절차에 여러 번 있어 최종 평가결과에 대한 신뢰성이 떨어질 수 있다.

또한 “재식별 시도 가능성”과 “재식별 시 영향도”를 3, 4가지로 분류를 한 후 이를 가지고 기준을 도출하는 방법은 평가의 결과가 왜곡될 수 있는 여지가 있다. 이는 점수 1점의 차이로 “낮음”과 “중간”으로 다르게 분류가 될 수 있어, 평가방법이 연속적이지 않아 평가 결과에 대한 불만이 예상된다. 또한 각 분류의 기준점, 예를 들어 “개인정보 보호수준 분석”의 “낮음”과 “중간”의 기준점이 왜 4점인지에 대한 근거를 제시하기가 어려운 실정이다.

그리고 k, ℓ, t 모델에 대한 평가기준에 대한 한계가 존재한다. k-익명성의 경우는 그 모델의 목표가 신분노출 공격이고, k값이(즉, 가장 작은 동질집합 크기) 신분노출 공격의 성공확률(1/k 이하)에 그대로 적용이 된다. 따라서 주어진 데이터의 신분노출 보호수준은 k값으로 정확하게 표현될 수 있다. t-근접성은 민감속성의 동질집합 내 분포와 테이블 전체의 분포의 유사도를 측정하여 공격자가 동질집합으로부터 추가적으로 얻어낼 수 있는 정보를 제한하는 데 목적이 있고 이는 두 분포의 차이로 직접 표현될 수 있기 때문에 기준이

되는 t값은 추론노출 공격의 성공확률을 직접 제어하는 변수로 볼 수 있다. 추론노출 공격의 성공확률과 t의 직접적인 관계는 도출하기 힘드나 추론노출은 공격자의 추가적인 정보획득량으로 판단을 해야 하며 t값은 이를 직접적으로 제어를 하고 있다.

하지만 ℓ-다양성(기본형)의 ℓ값은 모델이 목표로 하고 있는 속성노출의 보호수준을 표현하는 데 어려움이 있다. 이는 ℓ값이 아무리 커도 민감 정보의 쏠림현상이 존재하면 공격자가 높은 확률로 공격대상의 민감 정보를 추측해 낼 수 있기 때문이다. 이 문제는 차후에 논의할 엔트로피형과 재귀형 ℓ-다양성에도 그대로 적용된다. 따라서 개선안에서는 모델의 파라미터인 ℓ에 의존하지 않고 속성노출의 방어수준을 나타낼 수 있는 새로운 기준을 활용해야 할 것으로 보여진다.

3.4.3. 적정성 평가 개선안 개요

적정성 평가 개선안은 다음과 같이 요약될 수 있다. 평가점수의 연속성을 확보하기 위해서 각 부분평가단계의 결과 값들을 정해진 방법으로 조합하면 최종 평가수치가 나오고 이 평가수치에 따른 적절한 비식별 수준에 대한 가이드라인을 제공하고자 한다. 즉, 각 부분평가의 평가수치를 0에서 1까지의 실수로 표현하고, 전체 평가수치는 각 부분평가의 곱으로 표현하는 방식을 제안한다. 이는 리스크 분석에 많이 사용되는 기법이다.

데이터의 재식별 위험도를 측정함에 있어서 공격자가 자신의 공격대상이 비식별 데이터에 포함되어 있음을 100% 확신하고 있음을 가정하고 측정하는 경향이 있다. 예를 들어, 동질집합의 크기가 k이면, 공격대상자의 준식별자 값을 알고 있는 공격자가 대상자 레코드를 식별할 가능성은 1/k 이라고 평가하는데, 이는 그 k 중에 한 명이 반드시 공격대상자일 거라는 믿음을 가정하고 있을 때만 의미가 있다. 하지만, 실제로는 그렇지 않은 경우도 많다. 예를 들어, 한 통신사에서 자신의 고객정보 중 일부를 비식별화 하여 공개하는 경우, 공개되는 고객 중에 공격대상자가 포함되어 있는지 여부가 공격자에게 확실하지 않을 경우가 있다. 이러한 경우 대상자가 포함되어 있는 총 데이터(모집단 데이터)보다 비식별 데이터가 작으면 그만큼 공격자의 대상자 포함여부에 대한 확신이 감소하고, 이는 비식별 데이터에 대한 모든 공격의 성공률에 영향을 미친다. 따라서 데이터의 프라이버시 수준을 평가할 때 소속노출 가능성을 함께 고려해야 한다.

비식별 데이터에 대한 재식별 위험은 환경적인 요소, 공격자의 능력과 의도, 속성노출 수준, 재식별 시 파급효과 등 여러 가지를 복합적으로 고려해야 하며, 재식별 위험 수준에 따라서 데이터 비식별화 수준을 결정해야 한다. 즉 침해 위험요소를 파악하고 각 요소별 위험 수준을 정량화하여 그 요소별 수준들의 곱으로 침해 위험도를 계산하고, 이와 비례하여 프라이버시 수준을 강화하는 방식을 취하는 것이 합리적이다. 재식별 위험도에 대한 적절한 비식별화 수준은 모델에 따라 다르고 정책에 의존하는 값이기 때문에 각 모델의 기준값 선정은 모델별로 정의하고 프라이버시에 대한 현 인식수준과 경험적인 판단기준을 활용하여 결정하여야 한다. 구체적으로, 침해 위험도는 그림 17과 같이 계산할 수 있다.

"개인정보 침해 수준"은 현 가이드라인에서 평가대상 기관에서 제출하는 개인정보 보호 수준의 반대 개념이고 현재 개인정보 보호 수준의 평가 기준표를 활용하면 측정이 가능하다. 재식별의도 및 능력, 개인정보 침해수준, 재식별 시 영향수준 측정값이 정해지면 다음과 같은 절차로 프라이버시 모델 보호수준을 결정한다.

침해 위험도 = (재식별의도 및 능력)
X(개인정보 침해수준)
X(재식별 시 영향수준)

〔그림 17〕 재식별 위험도 정량화 공식

3.4.4. 소속노출에 대한 보호

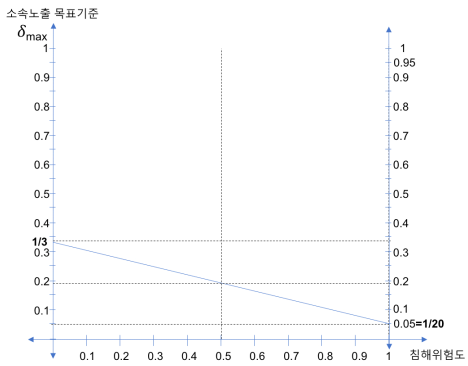
만약 데이터의 특성상 속성노출에 대한 보호가 우선인 경우는 침해 위험도에 따라서 δ-소속 모델의 기준값인 δ_{max}를 정해진 기준에 따라 정한다. δ_{max}는 공격자가 비식별 데이터 내에 공격대상자가 존재한다고 추측해서 맞출 확률을 의미하므로 결국 속성노출의 성공 확률을 의미한다. 이에 대한 목표값을 선정하는 것은 결국 개인이 심리적으로 받아들일 수 있는 익명 수준을 고려해서 결정해야 한다. 이는 현재 널리 사용되고 있는 k-익명성의 수준에 입각해서 결정하는 것이 합리적이다.

즉, k-익명성을 적용했을 때 가장 완화된 k값으로 사용되는 값은 3이고 이는 프라이버시 보호의 중요도

가 가장 낮은 상황에서 개인이 심리적으로 받아들일 수 있는 프라이버시 침해확률 중 가장 높은 값이 1/3=0.333 임을 의미한다. 또한 가장 강화된 k값으로는 20이 사용되는데, 이는 프라이버시 보호의 중요도가 가장 높은 상황에서 개인이 받아들일 수 있는 프라이버시 침해확률 중 가장 높은 값이 1/20=0.05 임을 의미한다. 이 기준은 사회적, 문화적, 정책적 배경에 따라 바뀔 수 있으며, 본 가이드라인 개선안에서는 상기한 침해확률 범위 최소 0.05에서 최대 0.333 사이의 값들을 프라이버시 침해 목표기준으로 사용하고, 차후에 환경변화를 고려하여 조정하는 것으로 한다. 따라서, 소속노출 목표기준은 그림 18과 같이 선정하는 것으로 한다. 혹은 주어진 침해위험도에 대해서 소속노출 목표기준은 그림 19와 같은 공식으로 계산한다.

만약 공격대상자가 해당 데이터에 포함되지 않은 사실이 공격자에게 노출될 경우도 프라이버시 침해가 발생하는 경우에는 δ_{min}에 대한 기준도 설정할 수 있으나 이에 대한 기준은 평가단의 토의에 의해 결정하는 것으로 한다.

소속노출 목표기준이 정해지면 평가대상 데이터의 소속노출 수준을 계산한다. 주어진 데이터의 소속노출 수준은 배경지식 기반 소속노출 수준과 데이터 기반 소속노출 수준이 있다. 배경지식 기반 소속노출 수준은 공격자가 주어진 데이터의 생성조건에 대해서 배경지식을 갖는 경우를 고려한다. 만약 공격자가 알고 있는 공격대상자의 준식별자 정보를 바탕으로 비식별화



〔그림 18〕 소속노출 목표기준 선정 기준

$$\text{소속노출 목표기준} = -\frac{17}{60} \times \text{침해위험도} + \frac{1}{3}$$

〔그림 19〕 소속노출 목표기준 선정공식

데이터에 공격대상자가 포함됨을 판단할 수 있는 경우는 배경지식 기반 소속노출은 1이 된다. 예를 들어서 한 병원의 모든 환자들 중에서 2017년에 입원한 40대 환자에 대한 데이터를 비식별화 해서 공개하는 경우에 만약 공격자가 자신의 공격 대상자가 2017년에 해당 병원에 입원했었고 40대라는 배경지식을 갖고 있고 해당 비식별 데이터의 조건을 알고 있다면 공격자는 공격대상자가 해당 비식별 데이터에 포함되어 있음을 100% 확신할 수 있고 따라서 소속노출 가능성은 비식별 데이터의 내용과 상관 없이 1로 계산할 수 있다. 경우에 따라서는 배경지식에 의한 소속노출 가능성이 0과 1 사이의 확률로 측정될 수도 있다. 예를 들어 모 데이터로부터 비식별화 데이터의 추출 기준이 민감속성을 기준으로 했다면 공격자는 그 기준을 안다고 하더라도 공격대상자의 민감속성에 대한 확신이 없으면 소속여부에 대한 확신을 갖기가 힘들다. 만약 비식별화 데이터가 고혈압 환자만 모 데이터에서 추출했다면 배경지식 기반 소속노출 수준은 공격자가 자신의 공격대상자가 고혈압 환자라고 추측할 가능성과 같아진다. 이 때, 비식별화의 목적이 환자의 민감속성인 혈압 수준을 감추는 것이 목적이라고 하면 순환논리에 빠질 수 있다. 따라서 이 경우 배경지식 기반 소속노출의 수준은 공격자가 알고 있는 일반적 상식, 우리나라 국민 중 공격대상자와 비슷한 조건을 가진 사람이 고혈압일 확률로 계산하여야 한다. 만약 공격자가 비식별화 데이터의 조건을 전혀 모르고 있는 경우에는 배경지식 기반 소속노출은 0으로 계산한다.

데이터 기반 소속노출 수준을 측정하기 위해서는 우선 비식별화 데이터의 원천데이터인 모 데이터에 대한 파악이 필요하다. 모 데이터는 비식별화 데이터에 포함된 개인이 모두 포함되어 있으며 공격자가 접근할 수 있는 데이터 모두를 말한다. 여기서 공격자의 모 데이터에 대한 접근성은 모 데이터의 레코드 내용을 읽을 수 있는 경우(레코드 접근성)와 둘째, 모 데이터의 레코드 수만 알 수 있는 경우(크기 접근성)를 모두 고려한다. 여기서 모 데이터가 비식별화 데이터의 개인을 모두 포함한다는 사실은 공격자도 알고 있음을 가정한다. 그리고 모 데이터 또한 비식별화된 데이터일 수도, 아닐 수도 있다. 모 데이터가 포함한 준식별자의 집합은 대상 비식별화 데이터의 준식별자의 집합과 다를 수 있다. 모 데이터의 준식별자 집합과 비식별화 데이터의 준식별자 집합이 중복이 없는 경우에도 이 모

데이터는 소속노출 수준 계산에 사용할 수 있다.

어떤 비식별화 데이터에 대한 모 데이터는 다수가 존재할 수 있다. 예를 들어 한 병원에서 2013년 해당 병원 신경정신과에 입원한 환자에 대한 비식별화 데이터를 공유한다고 가정하자. 이 경우에 모 데이터는 2013년 해당 병원에 입원한 모든 환자의 데이터 (원본 데이터 혹은 비식별 데이터 둘 다 가능)도 되고, 2010~2015년도에 서울시에 있는 병원에 입원한 환자 데이터(비식별화 혹은 원본)도 되며, 또한 통계청에서 발행한 2013년 인구통계도 포함된다. 이 경우, 해당 비식별화 데이터의 소속노출 수준을 계산하기 위해서는 각각의 모 데이터를 기준으로 소속노출 수준을 계산하여야 한다.

모 데이터는 그 레코드 수만 접근할 수 있어도 되기 때문에 항상 존재한다. 예를 들어 한 병원의 2017년 40대 여성의 입원데이터가 비식별 데이터라면 그 병원의 2017년 전체 입원 환자수를 알면 모 데이터가 존재하는 것이고, 만약 그것도 존재하지 않는다면 2017년 우리나라 전체 인구수를 모 데이터의 레코드 수로 사용하면 된다. 각 모 데이터에 대한 데이터 기반 소속노출 수준은 그림 20과 같이 계산한다.

$$\text{데이터기반 소속노출 수준} = \frac{\text{비식별데이터 수}}{\text{모 데이터 수}}$$

〔그림 20〕 모 데이터 크기접근성 소속노출 수준 공식

모 데이터에 대한 접근 가능한 정보가 레코드 수인 경우는 모 데이터의 준식별자 값에 대한 정보가 없으므로 모 데이터로부터 무작위 추출을 가정하여야 하고, 따라서 데이터 기반 소속노출 수준은 그 레코드 수의 비율로 다음과 같이 판단한다. 모 데이터의 레코드가 접근이 가능한 경우에는 우선 모 데이터에 포함된 준식별자 집합과 비식별화 데이터의 준식별자 집합을 비교하여 그 교집합을 계산한다. 만약 교집합이 없으면(즉, 중복되는 준식별자 부재) 레코드 수만 접근 가능한 것으로 보고 크기 접근성 소속노출 수준 공식을 사용하여 계산한다. 만약 교집합이 존재하면 데이터 기반 소속노출 가능성은 그림 21과 같이 δ-소속 모델의 계산방법을 따른다. 비식별화 데이터의 각 동질집합에 대해서 그 준식별자 값에 해당되는 모 데이터의 레코드 수 분의 동질집합 크기 비율을 구하고, 모든 동질집합에 대한 최댓값을 구한다. 주어진 모 데이터의 집합에 대해서 비식별화 데이터의 소속노출 수준은 그림 22와 같이 계산할 수 있다. 소속노출 목표수준과 데이터의 소속노출 수준을 비교하여 적정성 여부를 판단한다.

$$\text{데이터기반 소속노출 수준(레코드접근성)} = \max_{EQ_i \in T^*} \frac{|EQ_i|}{EQ_i\text{와 준식별자가 매칭되는 모데이터 레코드수}}$$

- T^*: "비식별" "데이터"
- EQ_i: i"번째" "동질집합"

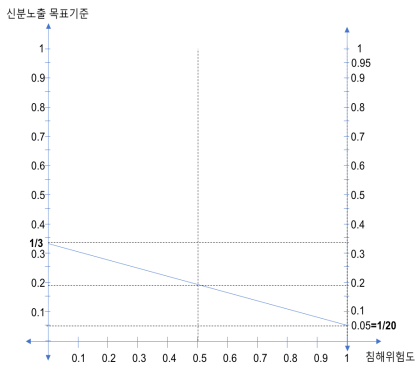
〔그림 21〕 소속노출 가능성 공식

$$\text{소속노출 수준} = \max(\text{배경지식기반 소속노출수준}, \max_{\text{모든모데이터집합}} \text{데이터기반 소속노출수준})$$

〔그림 22〕 소속노출 수준 계산공식

3.4.5. 신분노출에 대한 보호

신분노출 공격은 공격대상자가 비식별화 테이블의 특정 레코드와 연결시키는 공격이다. 신분노출 목표기준은 침해 위험도에 따라서 결정하되, 그림 23과 같이 결정한다.



〔그림 23〕 신분노출 목표기준 선정기준

혹은 주어진 침해위험도에 대해서 신분노출 목표기준은 그림 24와 같은 공식으로 계산한다.

$$\text{신분노출 목표기준} = -\frac{17}{60} \times \text{침해위험도} + \frac{1}{3}$$

〔그림 24〕 신분노출 목표기준 선정공식

주어진 데이터의 신분노출 수준은 동질집합의 크기와 함께 소속노출 수준도 함께 고려해야 한다. 즉, 데이터의 가장 작은 동질집합의 크기가 5라고 해서 달성한 신분노출 수준이 1/5가 아니라 이 확률에 소속노출 수준을 곱해줘야 한다. 이는 공격자가 이 데이터에 공격대상자가 포함되어 있다는 100% 확신이 없이는 신분노출 확률이 1/5라고 단정 지을 수 없음을 의미한다.

데이터의 소속노출 수준은 상기한 방법대로 계산하는 것으로 한다. 데이터의 최소 동질집합 크기와 소속노출 수준이 주어지면 그림 25와 같은 식으로 신분노출 수준을 계산하고 이를 신분노출 목표기준과 비교하여 비식별화 적정성을 판단한다.

$$\text{신분노출수준} = \text{소속노출수준} \times max_{EQ_i \in T^*} \frac{1}{|EQ_i|}$$

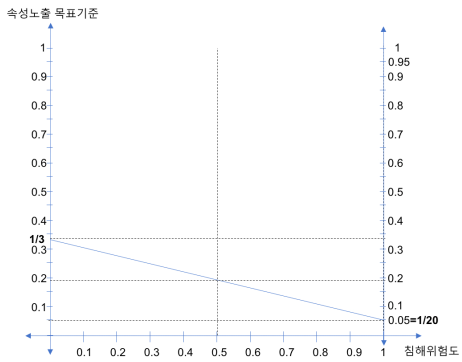
- T^*: 비식별 데이터
- EQ_i: i번째 동질집합

〔그림 25〕 신분노출 가능성 공식

3.4.6. 속성노출에 대한 보호

적정성 평가단이 속성노출에 대한 방어가 필요한 것으로 판단한 경우는 침해 위험도에 입각하여 속성노출 목표기준을 정하고 주어진 데이터의 속성노출 수준을 계산하여 이를 바탕으로 속성노출에 대한 적정성 평가를 한다.

속성노출에 대한 목표기준은 침해 위험도에 입각하여 그림 26과 같이 결정한다.



〔그림 26〕 속성노출 목표기준 선정기준

혹은 주어진 침해위험도에 대해서 속성노출 목표기준은 그림 27과 같은 공식으로 계산한다.

$$\text{속성노출 목표기준} = -\frac{17}{60} \times \text{침해위험도} + \frac{1}{3}$$

〔그림 27〕 속성노출 목표기준 선정공식

기본형/엔트로피형/재귀형 l-다양성은 속성노출에 대한 방어를 목표로 하고 있다. 하지만 속성노출 위험은 동질집합 내 민감속성의 분포 따라 결정되므로 데이터의 속성노출 수준과 l값 사이의 관계는 규정적이지 않다. 이러한 해당 프라이버시 모델의 한계 때문에 속성노출에 대한보호수준의 평가는 모델과는 다른 기준이 필요한 실정이다.

본 개선안에서는 한 동질집합의 속성노출 수준(그림 28)은 그 동질집합에 속한 각 민감속성의 값 중 가장 많이 존재하는 값의 비율의 최댓값으로 계산하고, 이에 대한 전체 테이블에서의 최댓값에 소속노출 수준을 곱한 값을 그 데이터의 속성노출 수준으로 계산한다.

$$\text{속성노출수준} = \text{소속노출수준} \times max_{\substack{EQ_i \in T^* \\ S_j \in S}} \times \frac{\max_{s \in S_j} freq_{EQ_i}(s)}{|EQ_i|}$$

- T^*: 비식별 테이블
- EQ_i: i번째 동질집합
- S: T^*에 존재하는 민감속성집합
- S_j: T^*에 존재하는 j번째 민감속성
- $s \in S_j$: S_j "민감속성" "값"
- $freq_{EQ_i,S_j}(s)$: EQ_i내 S_j민감속성값 s의 갯수

〔그림 28〕 속성노출 수준 공식

3.4.7. 추론노출에 대한 보호

데이터가 추론노출 수준까지의 보호수준이 필요한 경우는 추론노출에 대한 방어수준을 측정하여야 한다. 추론노출을 목적으로 제안된 모델로는 δ-노출과 t-근접성이 있다. 하지만 δ-노출의 경우는 실제로 테이블 내 민감속성 분포와 동질집합 내의 민감속성 분포의 유사성을 보장하지 못하기 때문에 주어진 테이블의 추론노출 보호수준 측정은 t-근접성이 제시하고 있는 방법으로 측정하는 것이 바람직하다.

추론노출에 대한 프라이버시 달성수준을 나타내는 t값의 선정은(작을수록 프라이버시 고수준 보장) 주어

진 “침해 위험도”를 기준으로 선정하되, 침해 위험도가 높을수록 작은 t값을 선정하는 방식으로 한다. δ-노출과 t-근접성 모델의 경우 공격자에 노출되는 정보의 양을 기준으로 설계된 모델이기 때문에 그 기준값인 δ와 t를 침해 위험도에 직접적으로 연관시키는 것은 어렵고 여러 시행착오 후에 자연스럽게 정립되는 시간이 필요하다. 주어진 데이터의 추론노출 수준은 각 동질집합의 민감속성의 분포와 테이블 전체의 민감속성의 분포의 차이를 EMD 알고리즘으로 계산하고 그 최대값을 소속노출 가능성에 곱한 값에 그 테이블의 추론노출 수준으로 계산한다(그림 29).

추론노출 수준 t^*
$=$ 소속노출 수준 $\times \max_{\substack{EQ_i \in T^* \\ S_j \in S}} EMD_{S_j}(EQ_i, T^*)$

- T^*: ”비식별”“데이터”
- EQ_i: i”번째”“동질집합”
- S: T^*”에”“존재하는”“민감속성”“집합”
- S_j: T^*”에”“존재하는”“한”“민감속성”
- $EMDvert_{S_j}(EQ_i, T^*)$: 민감속성 S_j의 EQ_i내 분포와 T^*내분포의 EMD”거리”

〔그림 29〕 추론노출 수준 정의

이를 앞에서 선정한 추론노출 위험도 기준 t값과 비교하여 적정성을 판단한다.

3.5. 정보 집합물 결합

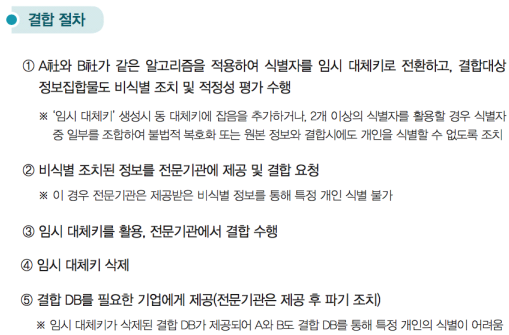
정보 집합물의 결합이란 다수의 정보집합물에 동시에 속해있는 각 개인에 대한 데이터를 한 곳에 모음으로써 정보의 질을 향상시키는 방법이다. 두 개 이상의 기관이 소유하고 있는 정보집합물을 결합하면 하나의 데이터로부터는 얻을 수 없는 유용한 정보를 도출할 수 있다. 예를 들어 전문분야가 다른 두 병원의 데이터를 모으면 환자의 치료에 도움이 되는 중요한 정보를 도출해 낼 수도 있다. 이러한 이유로 산업계에서는 정보집합물의 결합에 대한 요구가 점점 높아지고 있는 실정이다.

하지만 정보집합물에 포함된 개인의 프라이버시를 보호하기 위해서 정보집합물의 결합은 신중히 이루어져야 한다. 특히 결합으로 인해서 개인정보의 보호수준이 결합 이전보다 낮아지는 경우를 대비해야 하며 이를 예방하기 위해서는 기술적, 행정적, 법률적 조치가 필요한 실정이다.

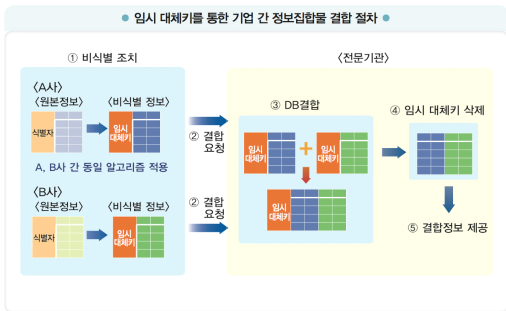
이 절에서는 정보집합물의 결합을 지원하기 위한 기본적인 방안을 현 가이드라인을 통해서 알아보고, 보다 안전한 정보집합물 결합을 위한 방안을 모색하고자 한다.

3.5.1. 현 가이드라인 정보집합물

현재 가이드라인은 정보집합물의 절차에 대해서 그림 30, 31과 같이 기술하고 있다.



〔그림 30〕 현 가이드라인의 정보집합물 결합 절차

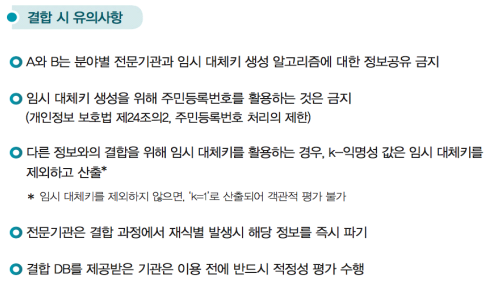


〔그림 31〕 현 가이드라인의 정보집합물 결합 절차 흐름도

또한 현 가이드라인은 정보집합물 결합에 있어서의 주의사항을 그림 32와 같이 기술하고 있다.

현 가이드라인은 정보집합물의 결합절차에 대해서 다음 중요한 세가지 제안을 하고 있다.

- ○ 정보집합물 결합은 정보 제공사가 아닌, 신뢰할 수 있는 제3의 기관이 수행하여야 한다.
- ○ 적절한 결합과 동시에 개인정보 보호를 위해서 임시대체키를 활용하여야 한다.
- ○ 정보집합물의 결합 이전과 이후 모두 반드시 비식



〔그림 32〕 현 가이드라인의 정보집합물 결합 시 유의사항

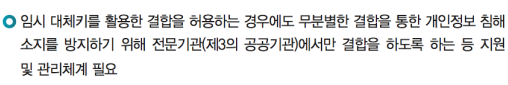
별화 절차를 거쳐야 한다.

위 세 가지 원칙은 정보집합물 결합에 있어서 개인정보보호를 위한 중요한 원칙을 적절히 제시하고 있다. 다음 절에서는 각각에 대해 면밀히 살펴보고 현 가이드라인의 보완점을 분석하고자 한다.

3.6. 정보집합물 결합 가이드라인 보완 사항

3.6.1. 전문기관에 의한 정보집합물의 결합

가이드라인은 정부에서 지정한 신뢰할 수 있는 전문기관만 정보집합물을 결합해야 하는 이유를 그림 33과 같이 설명하고 있다.



〔그림 33〕 전문기관에 의한 정보결합물 집합

가이드라인이 지적한 대로 임시대체키를 이용하여 결합을 하는 경우에도 개인정보 침해의 가능성을 완전히 배재할 수가 없다. 이는 임시대체키의 잘못된 생성 및 관리의 가능성이 있기 때문이다. 또한 임시대체키가 제대로 활용되었고 결합 이전에 비식별화가 적절히 수행되었다 하더라도 결합의 특성상 비식별화 수준이 현저히 저하될 가능성이 여전히 존재한다.

따라서, 정보집합물 결합에 포함된 개인의 프라이버시 보호를 위해서는 신뢰할 수 있는 전문기관의 도움이 절실한 상황이다. 만일에 하나 결합과정에서 비식별화 수준의 저하가 발생하더라도 전문기관이 이를 안전하게 관리하고, 결합데이터의 사용기관에게 전달하기 이전에 철저한 비식별화 평가 및 조치를 수행한 후 전달할 수 있게 된다. 따라서 전문기관이 정보집합물 결합을 대행하는 방안은 적절한 것으로 보여진다.

하지만 아직 해결해야 할 법률적으로 보완되어야 할 사항이 남아있다. 전문기관이 결합을 수행하는 과정에서 발생할 수 있는 개인정보 침해를 대비해서 법률적으로 전문기관의 역할과 권한을 강화할 필요가 있다. 만약 전문기관이 정보집합물 결합으로 인해서 법률적 제재를 당한다면 정보집합물의 결합은 실질적으로 어려워지며 이로 인한 잠재적인 사회적 경제적 이익은 포기해야 할 것으로 보여진다.

또 다른 측면으로는, 전문기관의 법적 행정적 부담을 덜기 위해서는 전문기관이 관여하지 않는 정보집합물 결합방안을 고려해 보아야 한다. 이는 비식별화 관련 사업을 활성화하는 데 기여할 수 있다.

3.6.2. 임시대체키의 활용

임시대체키는 비식별화된 두 정보집합물을 결합할 때 대응되는 개인을 인식하는데 필수적인 요소이다. 개인 식별자의 유일성과 일관성을 유지한다는 측면에서 임시대체키는 가명(pseudonym)과 같은 개념이다. 개인마다 다른 임시대체키를 생성하되, 임시대체키로부터 본래 개인을 판별할 수 없어야 하며 두 정보집합물에 모두 속한 개인의 임시대체키는 양쪽이 일치해야 한다.

현 가이드라인은 임시대체키의 안전을 위해서 임시대체키 생성 시 유의사항을 기술하고 있다. 임시대체키는 주민번호를 사용해서 생성하면 안 되며, 다수의 식별자를 조합해서 사용하는 경우 일부만 조합을 함으로써 만일의 경우에 대비를 하도록 권고하고 있다. 또한 임시대체키는 결합이 된 후 즉시 파기하도록 명시하고 있다.

하지만 아래와 같은 안전하지 않은 임시대체키 생성을 사전에 예방하기에는 그 한계가 존재한다.

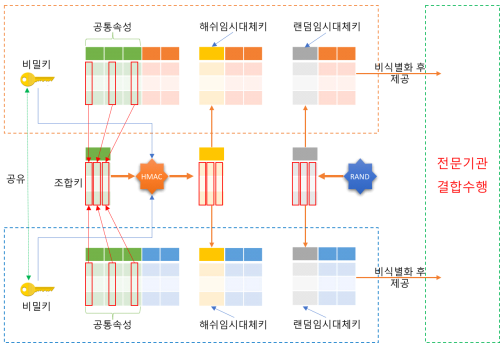
- ○ 주민번호를 제외한 식별자에 해쉬함수를 적용하여 임시대체키를 생성하는 경우는 사용된 해쉬함수가 안전하다 하더라도 공격자는 전수조사를 통해서 원래 주민번호를 복구할 수 있다. 이는 가능한 주민번호의 수가 적기 때문이다.
- ○ 주민번호를 제외한 식별자에 정보제공사끼리 공유하는 랜덤 숫자(혹은 비밀키)를 앞 혹은 뒤에 덧붙

여서 해쉬함수를 적용하는 경우는 앞의 경우보다 훨씬 안전하지만 고도의 분석기술을 통한 공격을 방어하기 위해서는 가장 안전한 방법으로 알려져 있는 HMAC기술을 통해서 식별자와 암호키의 조합에 대한 해쉬를 계산하는 것이 안전하다.

따라서 본고는 임시대체키의 안전성을 강화하기 위해서 다음과 같은 보완사항을 제안한다.

○ 조합키는 두 기관의 정보집합물 내 중복되는 속성들의 부분들의 조합으로 유일성을 유지하는 값을 정보집합물 제공사들끼리 합의를 한다.
○ 또한 정보제공사끼리 256bit이상의 비밀키를 정하고 SHA-256 이상의 해쉬함수를 정하며, 이를 조합키와 함께 HMAC 방법을 통해서 해쉬임시대체키를 생성한다.
○ 각 해쉬임시대체키에 대해서 랜덤임시대체키를 생성하여 그 매핑테이블을 공유한다.
○ 각 업체는 비밀키를 삭제하고 해쉬임시대체와 랜덤임시대체키 간의 맵핑테이블을 삭제한다.

위에서 제안한 임시대체키 생성방법은 그림 34로 표시할 수 있다.



〔그림 34〕 임시대체키 생성방법 개선안

3.6.3. 정보집합물 결합 시 비식별화 조치

현 가이드라인은 아래와 같이 기술하고 있다.

> ① A社와 B社가 같은 알고리즘을 적용하여 식별자를 임시 대체키로 전환하고, 결합대상 정보집합물도 비식별 조치 및 적정성 평가 수행

〔그림 35〕 현 가이드라인 임시대체키

자료 제공사는 임시대체키를 생성 후 결합전문기관에게 자료를 제공하기 이전에 비식별조치를 취하도록 되어 있다. 이는 전문기관도 제3기관이며 따라서 비식별화된 정보집합물을 제공하는 것이 바람직하다.

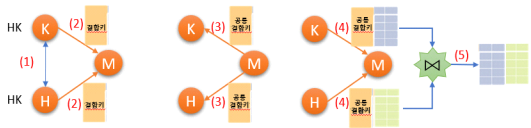
하지만 경우에 따라서 다수의 비식별화 된 데이터가 결합이 되면서 프라이버시 보호수준은 저하될 수가 있다. 예를 들어, A사가 k=5로 비식별화를 하였고 B사가 k=5으로 비식별화한 경우, 만약 두 회사가 비식별화에 사용한 준식별자 집합이 다르다면 결합 후 k-익명성 수준은 k<5로 저하될 가능성이 높다.

임시대체키로 대체 수행할 때 임시대체키의 안전한 생성에 대한 가이드라인이 부재하며, 전문기관에게 임시대체키 및 결합데이터가 노출될 가능성도 배제할 수 없다. 또한 전문기관 결합대행 시 비용이 발생한다. 이러한 이유로 현재 가이드라인은 결합된 데이터를 활용하기 이전에 이를 보완할 필요가 있다.

따라서 본고는 다음과 같은 보완사항을 제안한다.

○ 정보 제공자들간 키교환 프로토콜을 수행하여 해쉬키(HK)를 도출한다.
○ 각 정보제공자가 보유한 키집합과 해쉬키를 이용하여 결합키를 생성하고, 결합키로 인증코드를 생성해 정보활용자에게 제공한다.
○ 정보활용자는 M개의 제공자들로부터 수신한 결합키들 중 공통 결합키를 도출하여 각 정보제공자에게 데이터를 요청한다.
○ 각 정보제공자는 정보활용자가 전달한 공통결합키 리스트에 해당하는 데이터를 붙여서 제공하고, 정보활용자는 결합을 수행한다.
○ 결합 후 공통결합키는 제거한다.

위에서 제안한 분산 정보집합물 분산결합 방법은 그림 36으로 표시할 수 있다.



〔그림 36〕 분산 정보집합물 분산결합 방법

이 방법을 통해서 정보제공자는 자신의 데이터 중 어떤 데이터가 다른 제공자들도 공통으로 보유하고 있

는지 외 다른 제공자가 보유한 데이터에 대한 어떠한 정보도 알 수 없고, 다른 제공자의 데이터에 대한 조작이 불가능하다. 활용자는 공통결합키를 거짓으로 요청할 수 없기에 결합에 불필요한 데이터가 활용자에게 전달되지 않으며, 결합키로부터 원래의 식별자를 추측할 수 없다. 게다가 해쉬키(HK)를 모르기 때문에 전수조사를 통한 특정 고객의 정보를 알아낼 수가 없다. 제3자인 관찰자는 제공자 간의 통신을 통해서 해쉬키(HK)를 알아낼 수 없으며, 제공자와 활용자 간의 통신을 통해서 고객데이터를 접근할 수 없다.

Ⅳ. 결 론

본고에서는 국내외의 비식별화 정책과 정책 동향 및 정보 활용 사례를 알아보고, 현 정부의 비식별화 조치 가이드라인의 보호 모델과 비식별화 기술 및 정보 집합물 결합절차에 관한 개선안을 살펴보았다.

AI, 사물인터넷, 빅데이터 등 대규모의 데이터를 처리하는 과정에서 개인정보의 활용가치가 증가함에 따라 데이터를 활용할 수 있는 방안이 모색되어 왔다. 일찌감치 미국은 개인정보의 활용가치에 눈을 뜨고 1996년 HIPAA 법을 만들어 개인정보의 활용에 기틀을 만들었고, 캐나다, 호주, 일본 등의 여러 나라에서 잇따른 가이드라인, 법률, 정책의 형태로 이 흐름에 동참하고 있다. 2018년 5월 25일부터 효력이 발생하는 EU의 GDPR(General Data Protection Regulation)은 모든 개인에 관한 데이터의 주된 제어권을 각 개인에게 돌려주고 이를 적극적으로 보호함으로써 유럽 전역에 걸친 개인정보 활용에 대한 규약을 통일시키고 비즈니스 환경을 정비하는데 목적이 있다[11].

2016년 국내에서 6월 발표한 “개인정보 비식별 조치 가이드라인”은 비식별 조치의 까다로움과 그 복잡한 절차 및 비용, 그리고 비식별화 조치에 따라 손상된 데이터의 활용가치에 대해 우려의 목소리를 들을 수 있었다. 민간 소비자 단체 및 시민단체는 개인의 정보를 본래의 목적과 달리 활용하고 제3의 단체에게 공유하는 행위에 대한 우려를 표해왔고 가이드라인에 기술된 비식별 조치 기준의 효과에 대한 불신을 표현해 왔다. 발표된 지 1년 만에 정부는 본 가이드라인의 개선방안을 모색하기 시작하였고, 2018년 데이터 3법 개정안이 발의되었다. 하지만 법제화된 것이 아닌 ‘가이드라인’이라는 한계로 본격적인 데이터 활용은 시행령 등 구체적인 제도가 마련되어야 할 것으로 보인다.

산업계에서는 유용한 데이터 공유방식으로 판단하고 있는 두 개 이상의 정보집합물 결합에 대해서는 많은 논란의 여지가 있어왔다. 특히 임시대체키의 안전성과 결합시 프라이버시 수준의 저하, 그리고 전문기관의 결합대행에 대한 우려의 목소리가 컸다. 이에 본고에서는 기존에 사용되던 해쉬를 이용한 임시대체키 생성방법을 더욱 보완하였고, 나아가 랜덤 임시키를 사용하여 본 식별자와의 연관성을 완전히 차단하는 기법을 소개하였다. 또한 비식별화 된 두 데이터가 결합되었을 때 프라이버시 수준이 현격히 떨어질 수 있음을 설명하였고 이러한 데이터를 대행하여 결합하여야 하는 전문기관의 역할과 책임에 대한 문제는 입법적으로 해결할 수밖에 없음을 지적하였다.

본고는 개선안을 통해 개인정보의 남용을 막고 비식별화 조치를 활성화함으로써 데이터 활용과 개인정보 보호에 대한 긍정적인 기여를 할 수 있을 것으로 기대된다.

참 고 문 헌

[1] Health Insurance Portability and Accountability Act of 1996, Public Law No. 104-191, 110 Statutes. 1936 (1996)

[2] Nergiz, Mehmet Ercan and Atzori, Maurizio and Clifton, Chris, “Hiding the Presence of Individuals from Shared Databases”, Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, 2007

[3] Erlingsson, Pihur, Korolova, “A. Rappor: Randomized aggregatable privacypreserving ordinal response.” In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (2014), ACM, pp. 1054 – 1067, 2014

[4] Cynthia Dwork, “Differential Privacy”, Proceedings of International Conference on Automata, Languages, and Programming (ICALP 2006), 2006

[5] Office for Government Policy Coordination, Ministry of the Interior and Safety, Korea Communications Commission, Financial Services Commission, Ministry of Science, ICT and Future

Planning, Ministry of Health and Welfare, “Guideline on De-identification of Personal Information”, 2016

[6] Latanya Sweeney and Pierangela Samarati, “Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression”, Proceedings of the IEEE Symposium on Research in Security and Privacy, 1998.

[7] Ashwin Machanavajjhala, Johannes Gehrke, amd Daniel Kifer, “l-Diversity: Privacy Beyond k-Anonymity”, ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), 2007

[8] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian, “t-Closeness: Privacy Beyond k-Anonymity and l-Diversity”, IEEE 23rd International Conference on Data Engineering, 2007

[9] R. J. Bayardo and R. Agrawal. Data Privacy Through Optimal k-Anonymization. In Proceedings of the 21st International Conference on Data Engineering, ICDE ’05, pages 217–228, 2005.

[10] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Mondrian Multidimensional K-Anonymity. In Proceedings of the 22nd International Conference on Data Engineering, ICDE ’06, page 25, 2006.

[11] European Union. General Data Protection Regulation. Official Journal of Eurpean Union, 49:L119, 2016, Retrieved from https://gdpr-info.eu

〈저자소개〉

손 지 민 (Jimin Son)

학생회원

2021년 3월~현재 : 명지대학교 컴퓨터공학과 석사과정

<관심분야> 컴퓨터보안, 데이터 프라이버시

신 민 호 (Minho Shin)

종신회원

서울대학교 계산통계학과 학사 졸업

메릴랜드주립대 전산과 석사 졸업

메릴랜드주립대 전산과 박사 졸업

2010년 3월~2011년 2월 : 삼성전자 종합기술원

2011년 3월~2016년 2월 : 명지대학교 컴퓨터공학과 조교수

2016년 3월~2020년 2월 : 명지대학교 컴퓨터공학과 부교수

2020년 3월~현재 : 명지대학교 컴퓨터공학과 교수

<관심분야> 컴퓨터보안, 모바일보안, 데이터 프라이버시, 전기차 충전통신, 블록체인 보안