국회가 이른바 '데이터3법' 입법을 통해 가명화된 개인정보를 통계·연구 및 금융 등의 목적으로 이용규제를 풀면서 가명정보의 기술안전성에 대한 궁금증이 부각되고 있다. '가명화'는 데이터를 통해 제3자가 특정 개인을 누구인지 알 수 없도록 하는 비식별화 방식의 한 종류다. 현재 가명화 기술은 현존하는 기술로는 뚫기가 매우 까다로워 안전성이 높다고 전문가들은 분석하고 있다.가명정보의 원본이 되는 개인정보를 어떻게 가공, 저장, 관리하느냐에 따라 가명화 기술의 보안성이 좌우되므로 이에 대한 상시적인 모니터링과 관리·감독이 필요할 것으로 보인다.

데이터3법 중 개정 개인정보보호법은 가명처리의 정의에 대해 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체함으로써 추가 데이터를 사용하지 않는 이상 특정 개인을 알아 볼 수 없도록 하는 것으로 규정하고 있다. 다만 구체적인 가명처리 방식은 규정되지 않아 후속조치가 필요하다고 한국과학기술단체총연합회는 제언하고 있다. 이에 대해 행정안전부 관계자는 "국민들이 헛갈리지 않도록 가명처리 등에 대해 구체적인 사례 등을 담아서 해설서나 안내서를 내려고 한다"고 설명했다.

일반적으로 가명화 기법은 비밀키 암호화, 토큰화(Tokenization), 해시(Hash) 함수의 3가지 방식으로 분류해볼 수 있다. 이중 가장 간단한 방식은 비밀키 암호화 기법이다. 개인정보를 암호화 형태로 가공해 해당 암호키를 가진 사람이 아니면 해당 데이터의 주체가 되는 개인이 누구인지 모르도록 하는 방식이다. 그 간편성 때문에 보편화된 방식이지만 암호키를 가진 사람이 악용할 경우 개인을 식별할 수 있고, 암호 자체를 해커 등이 복호화(암호를 푸는 것)할 가능성을 배제할 순 없다. 이에 대비해 데이터3법은 형사처벌과 더불어 매출의 3%를 과징금을 매기는 강력한 제재를 담았다.

토큰화 방식은 특히 금융권에서 애용됐다. 신용카드나 체크카드 등의 아이디(ID·서비스가입자 신원) 정보를 해커 등 데이터베이스 공격자가 활용할 수 없는 값(토큰)으로 대체함으로써 가입자 개인의 식별을 불가능하게 하는 기술이다. 즉 개인정보의 원본 자료를 제3자가 풀어볼 수 없도록 난수화하는 방식 등을 의미한다. 은행 및 신용카드회사 등 전통적인 금융사는 물론이고 삼성페이, 안드로이드페이, 애플페이와 같이 스마트폰을 활용한 새 모바일결제서비스에도 토큰화 기술이 활용된 것으로 알려져 있다. 국내의 한 대형금융사 정보보안 담당자는 "토큰화는 답(개인정보)을 풀 수 있는 방정식에 난수를 적용해 아무리 풀어도 답(개인정보)이 안 나오는 방정식을 만드는 것과 같다"며 "우연하게 라도 해당 난수가 답으로 연결될 수 없도록 작업해야 하기 때문에 다소 작업능률은 떨어지지만 그만큼 안전하다고 볼 수 있다"고 설명했다.

토큰이 어떻게 활용되는 지는 평소 신용카드 결제 과정을 보면 이해하기 쉽다. 신용카드 회원 A씨가 상점에서 물건을 산 뒤 카드로 결제하면 해당 상점의 카드결제 단말기가 해당 카드의 개인정보(신용카드번호)를 토큰서버로 전송한다. 토큰서버는 해당 개인정보를 저장하는 동시에 이를 난수화한 토큰 데이터를 생성해 상점 단말기로 전송한다. 상점 단말기는 이렇게 받은 토큰을 다시 응용 및 거래서버로 보내 결제를 처리한다. 이 과정에서 만약 해당 토큰의 원본 개인정보인 신용카드번호가 필요하게 되면 거래서버는 해당 토큰을 토큰 저장소로 보낸다. 토큰 서버는 이렇게 받은 토큰을 원래 저장됐던 개인정보 데이터와 대조해 해당 거래의 원본이 되는 신용카드 번호를 조회해 준다.

사실 토큰화 방식에도 약점은 있다. 난수화 처리 되기 전의 원본인 개인정보 자체가 토큰서버에 집중돼 저장돼 있으므로 해당 서버가 공격 당할 경우 대규모 개인정보 유출 사고가 빚어질 가능성이 상존한다. 토큰서버의 사이버 공격 대응 능력을 키우는 것이 핵심 이슈로 떠오를 수 밖에 없다. 아울러 개인데이터를 난수로 치환할 때 해당 작업의 효율성을 높이려고 암호화되지 않은 난수로 치환하는 경우(일명 의사난수생성기법)가 있는데 이는 암호화된 난수를 사용할 때보다 개인정보 보호 수준이 낮으므로 가급적이면 암호화 난수방식을 적용하도록 유도하는 게 바람직하다.

해시함수는 원래의 데이터값이 어느 정도의 길이가 됐든 무조건 고정된 크기의 값으로 변환시켜 원본 정보를 알아 볼 수 없도록 하는 방식이다. 예를 들어 4글자가 됐든, 100글자가 됐든 무조건 8자 길이의 값으로 변환해 원본 값을 알아볼 수 없도록 뭉게 버리는 것이다. 여타 비식별화 방식에 비해 알고리즘이 비교적 간소하다. 그만큼 고도의 컴퓨팅 인프라 없이도 비식별화 처리를 할 수 있어서 경제성과 작업 효율성면에서 선호된다. 특히 전자서명, 전자상거래, 서비스 회원의 가입 비밀번호 등에 많이 활용된다.

해시함수는 원본 데이터값의 크기를 알 수 없도록 뭉개버리기 때문에 해커가 깨는 것이 사실상 어렵다. 다만 이론상으로 원본 데이터 값의 길이를 안다면 아주 무식한 방법으로 깰 수 있다는 주장도 있다. 이른바 '무차별 대입법'이다. 만약 8글자라고 한다면 8글자에 넣을 수 있는 모든 문자값을 마구잡이로 넣어서 우연히 하나 걸리면 풀 수 있다는 것이다. 다만 이는 천문학적인 연산능력으로 공격해야 극히 희박한 확률로 우연히 얻을 수 있는 결과이기 때문에 현재의 보편적인 컴퓨팅 능력으로는 실효성이 떨어진다는 게 정보기술(IT)업계의 대체적인 평가다.

여기에서 더해 보안성을 높인 해시함수 방식들이 있다. '저장키를 사용하는 키 해시함수 방식'이다. 이 방식은 비밀키를 입력해야 해시함수 값을 재전송할 수 있어 정보 유출 위험을 줄여준다. '키 삭제를 통한 결정적 암호화 또는 키 해시 함수방식'도 있다. 이는 개인정보에 대한 가명을 난수로 선택한 뒤 그에 상응하는 데이터값의 테이블을 지워버리는 것이다. 이렇게 하면 가명 정보와 개인정보간 연결성이 감소돼 그만큼 프라이버시 침해 위험이 낮아지게 된다.