https://doi.org/10.33778/kcsa.2020.20.5.053

가명정보 Life-Cycle에 대한 위험 분석을 통한 관리적/기술적 보호조치 방안에 대한 연구★

차 건 상*

요 약

개인정보보호법 등 데이터 3법 개정에 따라 통계작성, 과학적 연구, 공익적 기록보존 등을 위해서는 정보주체의 동의 없이 가명정보 처리가 가능하며 개인정보와 달리 개인정보 유출통지 및 개인정보 파기 등의 법적용 예외조항을 두고 있다. 가명정보는 국가별로 가명처리에 대한 기준이 상이하며 국내에서도 개인정보 비식별 조치 가이드라인에 비식별조치와 익명화를 동일시하고 있다는 점에서 개정이 필요하다 할 것이다. 본 논문에서는 4차 산업혁명에 따라 개인정보의 활용에 초점을 두고 새롭게 도입된 가명정보의 안전한 활용을 위해 가명정보의 개념을 살펴보고 국내외 비식별조치 기준과 가명정보의 생성/이용/제공/파기 단계에서 법 또는 시행령(안)의 주요내용 검토를 통해 향후 추진되는 관리적/기술적 보호조치 방안에 대한 제언을 하고자 한다.

Research on technical protection measures through risk analysis of pseudonym information for life-cycle

Gun-Sang Cha*

ABSTRACT

In accordance with the revision of the Data 3 Act, such as the Personal Information Protection Act, it is possible to process pseudonym information without the consent of the information subject for statistical creation, scientific research, and preservation of public records, and unlike personal information, it is legal for personal information leakage notification and personal information destruction There are exceptions. It is necessary to revise the pseudonym information in that the standard for the pseudonym processing differs by country and the identification guidelines and anonymization are identified in the guidelines for non-identification of personal information in Korea. In this paper, we focus on the use of personal information in accordance with the 4th Industrial Revolution, examine the concept of pseudonym information for safe use of newly introduced pseudonym information, and generate / use / provide / destroy domestic and foreign non-identification measures standards and pseudonym information. At this stage, through the review of the main contents of the law or the enforcement ordinance (draft), I would like to make suggestions on future management / technical protection measures.

Key words : Pseudonymisation, Anonymization, De-idenfication, Privacy, GDPR

접수일(2020년 11월 30일), 수정일(2020년 12월 10일), 게재확정일(2020년 12월 30일)

★ 본 논문은 2020학년도 건양대학교 학술연구비 지원에 의하여 이루어진 것임.

* 건양대학교 사이버보안공학과

1. 서 론

1.1 연구 배경

4차 산업혁명 시대에 인공지능, 빅데이터 등 미래 신산업 성장과 육성을 위해서는 데이터(Data)의 활용이 무엇보다 중요한 과제로 떠오르고 있다. 유럽연합은 데이터 활용을 위해 가명정보 개념 도입하였으며 이러한 글로벌 Trend는 국내 개인정보보호 관련 법령의 변화를 촉구하게 되었다. 이에 대통령 직속 4차 산업혁명위원회 주관으로 관계부처·시민단체·산업계·법조계 등 각계 전문가가 참여한 '해커톤' 회의 합의결과 및 국회 특별권고 등을 반영하여 "데이터 3법" 개정을 추진하였으며 2020년 1월 9일 관련 법안은 국회를 통과하였고 8월 시행을 앞두고 있다.

개정된 개인정보보호법에서 가명정보는 개인정보와 달리 법적 의무사항에 대한 예외조항을 두고 있다. 예를들면 가명정보의 활용시 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보 처리가 가능하며 제3자 제공도 가능하다. 또한 가명정보 처리시, 개인정보 유출통지, 개인정보의 열람, 정정, 삭제권 등 법적용 예외조항으로 인정되면서 개인정보처리자의 부담을 경감하고 있다.

그러나 가명정보의 특성상 추가정보 획득을 통해 언제든지 정보주체를 식별할 수 있다는 측면에서 가명정보의 안전한 관리는 무엇보다 중요하다고 할 수 있다. 이에 본 논문에서는 개인정보처리자가 가명정보 활용시 준수해야 할 최소한의 보호조치 방안에 대한 방향성을 제시하고자 한다.

1.2 관련 연구 및 연구 방법

가명정보와 관련한 선행 연구로써 이대희(2017)는 유럽, 미국, 영국의 비식별화 및 가명정보의 개념 및 동향을 연구하였으며 이루리(2016)는 개인정보 비식별화에 대한 쟁점 및 한계점을 연구하였으며 이현승/송지환(2016)은 가명화(휴라스틱 가명화, 암호화, 교환방법) 조치의 재식별 위험성을 지적하였으며 최광희(2019)는 위험도 기반의 가명정보 활용을 위한 프레임워크를 제시한 바 있다.

본 논문은 데이터 3법 개정에 따른 가명정보의 안전한 관리를 위한 방향성 제시를 위해 국내외 가명정보에 대한 관련 연구 및 문헌을 조사 분석하고 가명정보의 생성(수집), 이용·저장, 제공 파기 등 가명정보의 Life-Cycle에 대한 재식별, 오남용, 분실 등 공격기법 및 보호기술 Mapping 등의 위험분석을 통해 가명정보의 안전한 관리를 위한 방향성을 제시하고자 한다.

이를 위해 2장에서는 개인정보, 가명정보, 익명정보 및 비식별처리 등의 개념과 가명정보 관련 개정된 데이터 3법 개정내역, 가명화를 포함한 개인정보 비식별화 조치 기준을 살펴본다. 3장에서는 가명정보에 생성(수집), 이용, 제공, 파기 단계별 공격기법과 이에 따른 현행 법령상 대책 Mapping을 통한 위협을 분석하고 끝으로 4장에서는 가명정보의 안전한 활용을 위한 관리적/기술적 보호조치 방안을 기술하고자 한다.

2. 개인정보 비식별 조치

2.1 가명정보 개념 및 법률 규정

2.1.1 개인정보/가명정보/익명정보 개념

개인정보보호법 제정시 개인정보에 대한 개념은 "살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)"라고 규정되어 있었으나 개정된 개인정보보호법에서는 "살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보"는 동일하게 규정하고 있으나 "해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보"에 대해서는 "쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다."라는 문구 추가를 통해 개인정보의 범위를 명확히 규정하여 개인정보처리자의 혼선을 최소화하였다.

더불어 개인정보보호법 제2조제1호다목을 신설하여 가명정보의 개념을 도입하였다. 법에서 규정하고

있는 가명정보는 "원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보"로 규정하였으며 개인정보와 달리 "통계작성, 과학적 연구, 공익적 기록보존"등의 목적으로 정보주체의 동의없이 처리할 수 있도록 하는 한편, 가명정보의 처리에 관한 특례 조항을 신설하였다.

EU GDPR[1]의 경우 가명처리에 대해 제4조 정의 조항에서 "추가 정보를 사용하지 않으면 개인정보가 더는 특정 정보주체와 연결되지 않도록 개인정보를 처리하는 것을 의미한다. 단, 그러한 추가 정보는 별도로 관리되어야 하며, 개인정보가 식별되었거나 식별 가능한 자연인과 연결되지 않도록 하기 위한 기술적 조직적 조치가 이루어져야 한다."라고 규정하고 있다.

익명정보와 관련하여 개정된 개인정보보호법 제3조 제7항에서는 익명 또는 익명처리라는 용어가 사용되었으며 개인정보보호법 용어정의 별도로 규정되지는 않았지만 제58조의2(적용제외) 조항에서는 "이 법은 시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더이상 개인을 알아볼 수 없는 정보에는 적용하지 아니한다."라고 규정하고 있어 익명정보의 경우 개인정보보호법 적용 대상이 아님을 명시적으로 규정하고 있다.

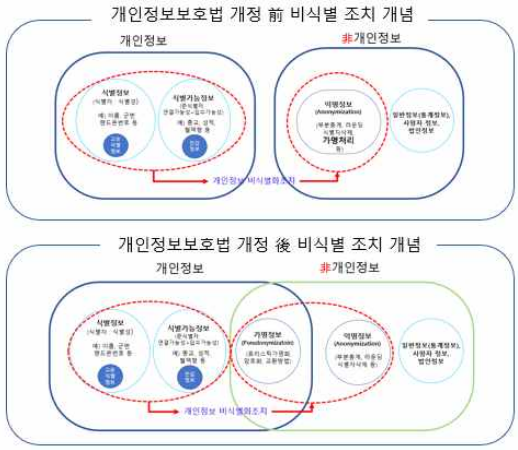
<표 1> 개인정보, 가명정보, 익명정보 비교

구분	개인정보	가명정보	익명정보
개념	식별가능한 개인정보 또는 다른정보와 결합하여 식별가능한 정보 포함 (단 입수가능성 고려)	개원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보	시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더이상 개인을 알아볼 수 없는 정보
식별가능성	가능	추가정보결합시 재식별가능	불가능
법규정	개인정보보호법 제2조제1호	개인정보보호법 제2조제1호 신용정보보호법 제2조제16호	신용정보보호법제2조제17호에서 익명처리 규정

2.1.2 가명정보 생성을 위한 비식별 조치 개념

비식별 조치에 대한 개념은 개인정보보호법과 신용정보보호법 모두 별도로 규정하고 있지는 않다. 국내의 경우 국무조정실 등 정부합동으로 2016년 발표한 "개인정보 비식별 조치 가이드라인"에서 비식별정보의 개념에 대해 개인정보를 비식별 조치한 정보로 규정하고 있다. 여기서 '비식별 조치'란 정보의 집합물에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체 등의 방법을 통해 개인을 알아볼 수 없도록 하는 조치라고 설명하면서 비식별조치를 EU 개인정보보호지침의 "anonymization, 익명화"와 같은 개념으로 규정하고 있다.

가이드라인에서는 비식별 조치 방법으로 가명처리 등을 포함하여 17개의 세부 기술을 언급하고 비식별 적정성 평가(K-익명성)에 따른 비식별정보는 "개인정보가 아닌 것으로 추정"하며 비식별 정보에 대한 산업적 활용을 허용하고 있다. 다만 가명처리 기술과 관련하여 가이드라인내에서는 가명처리(암호화 등)를 단독으로 사용하지 못하도록 규정하고 있다. 따라서 개정된 개인정보보호법에 따라 비식별 조치개념이 익명처리와 동일한 non-identification인지 가명처리를 포함한 de-identification인지 명확히 구분할 필요가 있는 것이다[2]. 이를 도식으로 정리하면 (그림 1)과 같다.



(그림 1) 개인정보보호법 개정 전후 비식별 조치 개념

또한 KISA에서는 개인정보 구성 요인을 다음과 같이 세가지로 구분한다. 첫째, 특정 데이터가 한 개인과 대응됨(single out), 둘째, 특정 데이터와 특정 개인이 연결됨(linkability), 셋째, 특정 데이터로부터 특정 개인을 추론할 수 있음(inference) 등이다. 여기서

세 가지 구성 요인을 모두 제거하는 것이 '익명화(anonymization)'라고 규정하고 있으며, 개인과 대응하는 경우는 허용하되 연결과 추론을 제거하는 경우 '가명화(pseudonymization)'라고 규정하고 있다[3].

국외의 경우, NISTIR 8053[4]에서는 비식별처리(de-identification)에 대한 개념을 '정보 집합에서 식별정보(identifying data)를 제거함으로써 개인정보를 특정한 인물과 연결할 수 없도록 한다.'라고 규정하고 있다. 즉 비식별처리는 익명처리와 가명처리를 포함한 개념으로 설명하고 있다.

소비자 프라이버시 권리장전법(CPBRA: Consumer Privacy Bill of Rights Act)에 비식별 데이터를 정의하고 특정인을 식별할 수 없도록 조치한 비식별 데이터는 개인정보가 아닌 것으로 취급하고 있다[5]. 의료개인정보보호법(HIPPA: Health Insurance Portability and Accountability Act)에 따라 제정된 개인의료정보 보호와 이용을 위한 규칙(HIPAA privacy rule)은 개인건강정보의 이용 및 공개와 관련하여 비식별 조치된 의료정보에 대해서는 규제를 면제하였다[6].

2.1.3 가명정보 관련 개인정보보호법 주요내용

가명정보와 관련한 개인정보보호법 및 동법시행령을 살펴보면 <표 2>와 같이 가명정보의 처리, 결합제한, 안전조치의무 등의 사항을 규정하고 있다. 다만 가명정보는 "정보주체 이외로부터 수집한 개인정보의 수집 출처 등 고지", "개인정보의 파기", "영업양도 등에 따른 개인정보의 이전 제한", "개인정보 유출 통지", "개인정보 열람, 정정, 처리정지권" 등을 예외사항으로 규정으로 있다.

<표 2> 가명정보 관련 개인정보보호법 주요내용

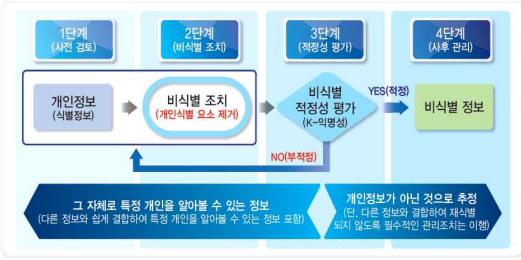
구분	조문	주요내용
개인정보보호법	제3조(개인정보 보호 원칙)	개인정보 처리시 익명 또는 가명정보로 처리
	제15조(개인정보의 수집·이용)	암호화 등의 가명처리시 정보주체 동의없이 이용가능
	제17조(개인정보의 제공)	암호화 등의 가명처리시 정보주체 동의없이 제공가능
	제28조의2(가명정보의 처리 등)	통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보 처리가능하며 제3자 제공도 가능함
	제28조의3(가명정보의 결합 제한)	통계작성, 과학적 연구, 공익적 기록보존 등을 위해 서로 다른 개인정보처리자 간의 가명정보의 결합시 전문기관 수행
	제28조의4(가명정보에 대한 안전조치의무 등)	가명정보의 안전성 확보조치 사항 규정 개인정보처리자는 가명정보 처리시 관련기록 작성 보관 의무
	제28조의5(가명정보 처리 시 금지의무 등)	누구든지 가명정보 재식별 처리 금지하고 가명정보처리시 특정인을 식별한 경우 처리중지 및 과기 의무
	제28조의6(가명정보 처리에 대한 과징금 부과 등)	가명정보의 불법적 재식별시 매출액의 3/100 과징금 부과
	제28조의7(적용범위)	가명정보는 제20조, 제21조, 제27조, 제34조제1항, 제35조부터 제37조까지, 제39조의3, 제39조의4, 제39조의6부터 제39조의8까지의 규정을 적용하지 아니한다.
	제71조(벌칙)	법제28조의3을 위반하여 처리하거나 특정 개인을 알아보기 위해 가명정보 처리시 5년 이하의 징역 또는 5천만원 이하의 벌금

개인정보보호법과 별도로 신용정보의 이용 및 보호에 관한 법률에서는 제2조에서 가명정보 및 가명처리에 대한 개념을 규정하고, 제20조의2(개인신용정보의 보유기간 등)에서는 가명정보의 보유기간을 규정하고, 제32조(개인신용정보의 제공·활용에 대한 동의)에서는 가명정보 제공시 동의 예외사항, 제40조의2(가명처리·익명처리에 관한 행위규칙), 제40조의3(가명정보에 대한 적용 제외)에서는 법적용 예외 사항을 규정하고 있다.

2.2 국내외 비식별 조치(가명화) 기준

2.2.1 국내 개인정보 비식별 조치 기준

정부에서 2016년에 발표한 '개인정보 비식별 조치 가이드라인'에서는 (그림 2)와 같이 총 4단계로 구분하여 단계별 조치사항을 제시하고 있다.



(그림 2) 개인정보 비식별화 조치 절차[7]

1단계(사전검토)에서는 개인정보 해당 여부를 검토하고 2단계(비식별 조치)에서는 (그림 3)과 같이 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 여러 가지 기법을 단독 또는 복합적으로 활용하여 비식별화 조치를 수행한다. 3단계(적정성 평가)에서는 비식별 조치가 충분하지 않은 경우 다른 정보와 결합, 다양한 추론 기법 등을 통해 개인식별이 될 우려가 있으므로 비식별 조치의 적정성 평가기준으로 k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness)을 제시하였다. 4단계(사후관리)에서는 비식별 정보의 안전한 관리를 위해 관리적 보호조치(비식별 정보파일에 대한 관리 담당자 지정, 비식별 조치 관련 정보공유 금지, 이용 목적 달성시 파기 등의 조치) 및 기술적 보호조치(비식별 정보파일에 대한 접근통제, 접속기록 관리, 보안 프로그램 설치·운영 등의 조치), 재식별 모니터링, 비식별정보 제공 및 위탁계약시 준수사항, 재식별시 조치요령 등을 제시하고 있다.

(그림 3) 개인정보 비식별화 조치 기법[7]

가이드라인에서는 '가명처리' 기법만 단독 활용된 경우는 충분한 비식별 조치로 보기 어렵다고 설명하고 있으며 가명처리와 관련한 3가지 기술을 세부적으로 기술하고 있다.

<표 3> 국내 가이드라인에서 가명처리 방법

구분	주요내용
Heuristic Pseudonymization	식별자에 해당하는 값들을 몇 가지 정해진 규칙으로 대체하거나 사람의 판단에 따라 가공하여 자세한 개인정보를 숨기는 방법
Encryption	정보 가공시 일정한 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 방법
Swapping	기존의 데이터베이스의 레코드를 사전에 정해진 외부의 변수(항목)값과 연계하여 교환하는 방법

2.2 국외 개인정보 비식별화 조치 기준

국외 개인정보 비식화 조치와 관련한 기준 크게 <표 4>와 같이 구분된다.

<표 4> 비식별화 조치 관련 국외 기준

구분	주요내용
국가별 주요 기준	- EU GDPR(Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques) - 유럽 ENISA Pseudonymisation Techniques and Best Practices - 일본 개인정보보호법 - 미국 NISTIR 8053 : De-idenfication of Personal Information - 미국 NIST 800-188 De-Identifying Government Datasets - 미국 HIPPA(Health Insurance Portability and Accountability Act) - 영국 UKAN The Anonymisation Decision-Making Framework - ENSIA Pseudonymisation Techniques and Best Practices
국제 표준화 그룹	- ISO/IEC 29100 Privacy Framework - ISO/IEC 20889:2018(Privacy enhancing data de-identification terminology and classification of techniques) - ITU-T X.1148 : Framework of de-identification processing service for telecommunication service providers

EU GDPR에서는 가명처리의 정의를 하면서 목적 외 처리를 광범위하게 허용하는 방향으로 법제화하고

있다. EU 개인정보보호작업반에 의해서 채택된 익명처리 기법에 대한 의견서(Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques)는 주요 익명처리 기법을 소개하고 익명처리 기법의 원칙, 강점과 약점 및 각 기법의 사용과 관련한 견해를 제시한다. 의견서는 주요 익명화처리 기법으로 무작위화(Randomization), 일반화처리(Generalizatoin)를 제시하고 있다.

과학기술 분야의 각종 표준을 담당하는 상무부 산하의 NIST(National Institute of Standards and Technology)는 2015년 10월 '개인정보 비식별화에 관한 보고서(NISTIR 8053 : De-idenfication of Personal Informaiton)'를 통해 준식별자에 대한 비식별처리를 위한 기술을 구분하여 제시하고 있다.

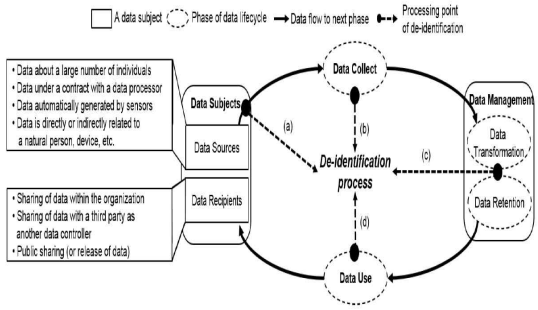
<표 5> De-identification technique of NIST IR 8053

구분	주요내용
교환 Swapping	규정된 일반화 수준 이내에서 개인 기록부 간에 준식별자 값을 교환할 수 있다. 교환은 통계적 특징을 유지할 필요가 있을 경우에는 주의하여 다루어야 한다.
범주화 Suppressing	준식별자를 범주화하거나 제거할 수 있다. 정보를 제거하면 프라이버시 보호가 극대화되지만 정보 집합의 효용이 감소할 수 있다.
일반화 Generalization	특정한 준식별자 값이 주어진 범위 안에 있다고 보고하거나 어떤 세트의 한 요소로서 보고할 수 있다. 예를 들면, 우편번호 12345를 12000과 12999 사이의 우편번호로 일반화할 수 있다.
인자변환 Perturbation	규정된 일반화 수준 이내에서, 각 개인에 대해 일관된 방식으로 특정한 값을 다른 값으로 교체할 수 있다.
하위샘플링 Sub-Sampling	비식별처리를 수행하는 기관은 전체 정보 집합을 공개하는 대신에 샘플을 공개할 수 있다. 만약 하위 샘플만을 공개한다면 재식별 가능성은 낮아진다.

국제표준화기구인 정보보호기술 연구반(ISO/IEC JTC 1/sc27)발간한 ISO/IEC 20889(PPrivacy enhancing data de-identification terminology and classification of techniques)에서는 De-Identification 방법에서 Pseudonymizayion techniques을 포함하여 8가지 방법을 구분하여 기술하고 있다.

ITU-T SG17에서 진행중인 신규 표준 아이템(Framework of de-identification processing service for telecommunication service providers: X.fdip)은 텔레콤 산업체들이 빅데이터 환경에서 비식별화에 대한 부분을 고려하고 있다.

X.fdip에서는 개인의 데이터 생명주기를 Data subjects{Data sources} → Data collect → Data management{Data analysis & Data storage} → Data usage → Data subjects{Data recipients}와 같이 정의하고 (그림 4)와 같이 데이터 생명주기 모델에서 비식별화 처리를 설명하고 있다.



(그림 4) X.fdip 데이터 생명주기별 비식별 조치 절차[8]

유럽 ENISA(European Union Agency for Cybersecurity)에서 2019년 발표한 "Pseudonymisation Techniques and Best Practices" 에서는 가명처리에 대한 기술을 5가지로 구분하여 설명하고 있다.

<표 6> PSEUDONYMISATION Techniques[9]

구분	주요내용
Counter	처음 숫자를 "0"으로 설정하고 하나씩 증가하는 함수를 사용하는 경우
Random number generator	임의의 숫자를 식별자로 할당하는 경우
Cryptographic hash function	단방향 암호화를 사용하는 경우
Message authentication code	가명처리를 위해 비밀키가 사용되며 HMAC이 대표적인 사례임
Encryption	가명처리를 위해 암호화를 수행하며 대표적으로 AES를 사용함

3. 가명정보 Life-Cycle별 위험분석을 통한 관리방안

가명정보에 대한 Life-Cycle별 위험분석은 크게 비식별정보에 대한 재식별 공격 유형 분석과 가명정보에 대한 Life-Cycle에 대한 위협 및 대책 분석을 통한 위험분석 방법을 적용하여 분석한다.

3.1 비식별정보에 대한 재식별 공격 유형

비식별정보(가명정보 포함)에 대한 재식별 공격에 대 이현승/송지환은 비식별화된 데이터는 이미 공개된 혹은 앞으로 추가로 공개될 데이터와 결합하여 재식별화될 가능성을 지적하였다.

<표 7> 비식별정보에 대한 재식별 공격 유형

구분	주요내용
동질성 공격 (homogeneity attack)	비식별화 이후 준식별자가 같은 경우 민감한 속성이 동일한 경우 발생
배경지식 공격 (background knowledge attack)	K 익명성의 경우 준식별자 내에서 민감한 속성에 대한 유출가능

비식별정보에 대해 동질성 공격과 배경지식공격을 통한 재식별 사례를 살펴보면 미국의 매사추세츠 주 사례(1997년), AOL 사례(2006년), Netflix 사례(2006년) 등이 있을 수 있다.

<표 8> 비식별정보의 재식별 사례

구분	주요내용
미국의 매사추세츠 주 사례 (1997년)	단체보험협회는 공무원 병원 기록을 식별자(이름, 주소, 사회보장번호)를 제외하고 준식별자(우편번호, 생년월일, 성별)을 공개하였으나 주지사 개인정보를 재식별함
AOL 사례 (2006년)	65만명의 3개월분 검색로그 2천만건을 공개(ID와 IP는 삭제)하였으나 질의문에 포함된 내용을 근거로 62세 미망인 Thelma Arnold 여사 재식별
Netflix 사례 (2006년)	영화추천 알고리즘의 정확성을 높이기 위해 50만명 이용자의 6년동안 평점기록 공개하였으나 온라인 영화전문사이트인 IMD에 공개된 사용자 리뷰와 매칭을 통해 재식별 성공

또한 비식별화된 정보에 대해 개별화 가능성, 연결가능성, 추론 가능성에 대해 “개인정보 비식별 조치 가이드라인”에서 제시된 17가지 기술에 대해 검토해 보면 <표 9>와 같으며 특히 가명화 기술과 관련한 “휴리스티 가명화”, “암호화”, “교환방법”은 개별화 가능성의 위함성이 있는 것으로 조사되었다.

<표 9> 비식별화기술 17종 위험도 비교[10]

분류	비식별화기술	개별화 가능성 위험	연결 가능성 위험	추론 가능성 위험
무작위화	휴리스틱 가명화	Yes	Yes	Yes
	암호화	Yes	Yes	May not
	교환 방법	Yes	Yes	May not
	재배열	Yes	May not	May not
	임의 잡음 추가	Yes	May not	May not
일반화	총계처리 기본 방식	일반적으로 No	일반적으로 Yes	일반적으로 Yes
	부분총계			
	라운딩			
	식별자 부분 삭제			
	감추기			
	랜덤 라운딩			
	범위 방법			
	제어 라운딩			
	공백과 대체			
기타	식별자 삭제	해당 없음	해당 없음	해당 없음
	레코드 삭제	해당 없음	해당 없음	해당 없음
	식별요소 전부삭제	해당 없음	해당 없음	해당 없음

NISTIR 8053에서도 비식별정보에 RISK를 크게 3가지로 구분하고 있다[4].

첫째, 신원 공개(identity disclosure)는 공격자가 특정한 정보 항목을 특정인에게 연결시킬 수 있을 때 일어나며 원인으로는 2006년 발생한 AOL사례처럼 비식별처리가 불충분한 경우, 식별자는 제거하였지만 IP 정보가 남아있어 연결을 통한 재식별이 가능한 경우, 가명을 역추적하는 경우 등이 있다.

둘째, 속성 공개(attribute disclosure)는 약간의 비밀 정보를 정보주체에 귀속시킬 수 있을 때 발생하는 경우로 예를 들면, 한 병원이 치료한 20세 여성 모두가 특정한 진단을 받았음을 나타내는 정보를 공개하고, Alice Smith가 20세 여성이며, 해당 병원에서 치료를 받은 사실이 알려져 있다면, 비록 그녀의 비식별처리된 치료 개인기록부를 다른 사람들의 것과 분간할 수 없더라도, Alice Smith의 진단 결과를 추론할 수 있다.

셋째, 추론적 공개는 공개된 정보의 통계적 특징으로부터 높은 신뢰도로 개인정보를 추론할 수 있는 경우에 발생한다. 예를 들면, 어떤 정보는 소득과 주택구매 가격 간에 높은 상관관계를 나타내고 있을 수 있

다. 주택의 구매 가격은 통상적으로 공개되는 정보이기 때문에 제3자는 이 정보를 이용하여 정보주체의 소득을 추론할 수 있다.

최광희(2019)는 “위험도 기반의 가명정보 활용을 위한 정책 프레임워크 연구”에서 재식별에 대한 공격 유형을 5가지로 설명하고 있다.

<표 10> 재식별 공격 유형[11]

재식별 공격	특징
Prosecutor attack	기존 지식을 사용하여 특정 데이터 주체에 속한 데이터를 재식별 ex) 유명인이나 친구, 친척 등을 찾아내는 것
Journalist attack	기존 지식을 사용하여 특정 데이터의 데이터 주체를 재식별 ex) 공개된 DB를 활용 (US voting registry 활용 등)
Marketer attack	기존 지식을 사용하여 가능한 많은 레코드에 상응하는 주체에 관하여 재식별 ex) 고객군 분류 등
Data membership attack	데이터 세트에서 특정 데이터 주체의 존재를 확인 ex) 특정 개인이 해당 데이터 집합에 있는지를 확인
Inference attack	다른 속성 그룹과 연관된 민감한 속성을 추론 ex) 서로 다른 데이터 집합 연계를 위한 특정 값 추론

3.2 가명정보 Life-Cycle별 위험분석

가명정보의 life-Cycle별 위험분석은 비식별 조치 기준 중 암호화와 해쉬 기법을 적용한다는 가정하에 위험분석을 실시하였다.

3.2.1 가명정보 수집/생성단계

개인정보의 Life-Cycle은 개인정보 수집, 이용,제공, 저장, 파기 등으로 구분되어 진다. 가명정보의 경우 개인정보와 유사하다고 할 수 있으나 개인정보의 경우 수집단계에서 시작하는 반면 가명정보는 타기관으로 부터 가명정보를 수집하거나 또는 생성, 기관간 결합을 통한 생성 등 여러 경로를 통해 가명정보가 생성·수집되며 이용, 제공, 저장 및 파기 등을 거치된다.

유럽 ENSIA는 가명처리와 관련한 생성 시나리오를 6가지로 구분하여 설명하고 있다. 여기서 국내 법과 비교하여 설명하면 Data Controller는 개인정보처리자에 해당하며 Data Processor는 개인정보 위탁자에 해당한다고 볼 수 있으며 특히 가명정보의 Life-Cycle 단계에서 가명정보의 생성단계에 촛점을 맞춰 구분하고 있다.

<표 11> PSEUDONYMISATION SCENARIOS[9]

구분	주요내용
시나리오 1	Data Controller가 개인정보 수집 및 가명처리 진행하는 경우
시나리오 2	Data Processor가 개인정보 수집하고 Data Controller가 가명처리 진행하는 경우
시나리오 3	Data Controller가 개인정보 수집 및 가명처리를 진행하고 가명정보를 Data Processor에게 전달하는 경우
시나리오 4	Data Processor가 개인정보 수집 및 가명처리를 진행하고 가명정보를 Data Controller에게 전달하는 경우
시나리오 5	Trusted Third Party 개인정보 수집 및 가명처리를 진행하고 Data Processor에게 전달하는 경우
시나리오 6	정보주체 스스로가 가명처리 과정에 참여하여 Data Processor에게 가명정보를 전달하는 경우

개인정보처리자내 가명정보의 생성 및 활용은 개인정보처리자가 이미 수집한 개인정보를 가명정보화하여 “통계작성, 과학적 연구, 공익적 기록보존” 목적으로 활용하는 경우로 재식별의 위험성이 발생하다고 하여도 기관내에서는 개인정보의 목적외 이용에 해당하는 사항이라고 할 수 있다. 다만 개인정보처리자간 가명정보의 결합 으로 인해 정보주체의 민감한 정보가 재식별된다거나 개인정보처리간 결합된 가명정보를 제공받는 경우 등이 앞서 언급된 동질성 공격과 배경지식 공격에 취약한 부분이라 할 수 있다.

가명정보 생성 또는 수집시 위험성은 비식별처리된 가명정보에 대해 승인되지 않은 재식별 이 이루어지거나 식별정보화 되는 것이 문제점이라 할 수 있다.

“NISTIR 8053”에서도 재식별 리스크(re-identification risk)에 대해 “재식별 능력은 원본 정보 집합, 비식별처리 기법, 공격자의 기술적 능력,

공격자의 가용한 자원, 비식별처리된 개인정보에 연결할 수 있는 추가 정보의 가용성에 달라 달라지기 때문에 이러한 리스크를 정량화하기는 어렵다."라고 설명하고 있다[12]. 따라서 기관내 가명처리의 경우 최소한의 기준을 제시하고 기관 스스로가 재식별성의 안전성을 담보하는 방향으로 추진하여야 한다. 다만 기관간 가명정보를 결합하여 제공되는 경우 재식별성에 대한 기준이 보다 엄격하게 관리되어야 한다. 암호화인 경우 결합전문기관이 Key 관리를 통해 재식별의 위험성이 관리된다고 하나 ENSIA가 제시한 Pseudonymisation Techniques 중 Cryptographic hash function의 경우 정보를 다량으로 보유한 기관의 경우 정보의 개인식별정보를 Input을 통해 개인정보가 재식별 될 위험성이 매우 높기 때문이다.

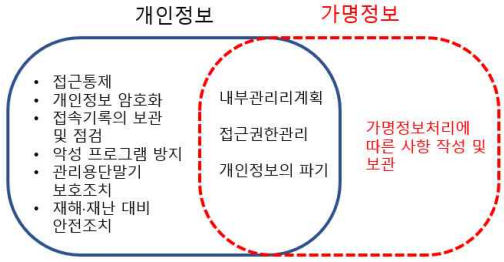
3.2.2 가명정보 저장/제공/파기단계

가명정보는 개정된 개인정보보호법을 살펴보면 암호화 등의 가명처리시 정보주체의 동의 없이 개인정보 수집·이용, 제공이 가능하며 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보 처리가 가능하며 제3자 제공도 가능하도록 규정하고 있다. 따라서 암호화 등의 기법을 통해 가명처리를 한 경우 암호화키에 대해서 엄격히 관리가 필요하다. 현재 행정안전부 홈페이지에 게시된 개인정보보호법 시행령 입법예고(안)[13]을 살펴보면 가명정보 및 추가정보에 대한 안전한 관리를 위해 다음의 사항을 의무조항으로 규정하고 있다.

<표 12> 개인정보보호법 시행령(안) 가명정보 보호조치 항목

구분	안전성 확보조치 사항
1	내부 관리 계획의 수립·시행
2	추가 정보의 별도 분리 보관 및 추가 정보에 대한 접근 권한 분리
3	가명정보 또는 추가 정보에 대한 접근 권한 관리 및 물리적·기술적 안전조치에 관한 내부 관리 계획 수립·시행
4	가명정보처리시 개인정보보호위원회가 정한 고시에 따라 가명정보의 처리 목적, 처리 및 보유기간, 추가 정보의 이용 및 파기 사항 작성 및 보관
5	개인정보처리자는 가명정보의 처리 목적이 달성되거나 가명정보 보유 기간이 경과한 때에는 그 가명정보의 지체 없는 파기

가명정보와 개인정보의 안전성 확보조치 사항을 비교하면 (그림 5)와 같다.



(그림 5) 개인정보와 가명정보의 안전성 조치기준 비교

개인정보와 가명정보의 안전성 확보조치 기준을 살펴보면 개인정보는 개인정보보호법 제29조 , 동법시행령제30조 및 고시에 따라 세부 내용을 정할 수 있나 가명정보는 개인정보보호법 제28조의4(가명정보에 대한 안전조치의무 등), 동법시행령(안) 제29조의5(가명정보 등의 안전성 확보조치 등)와 가명정보처리시 처리사항을 기재하는 내역에 대한 고시 등으로 구분된다. 따라서 암호화에 따른 암호화 키관리 등에 대한 사항등이 필요하다 할 것이다. 다만 개인정보보호법에 대해 모든 기준을 세부적으로 나열하는 것은 기술변화에 따라 제도변화가 늦다는 점과 기술의 발전을 제도가 앞서지 못한다는 점에서 바람직하다고 볼 수는 없다. 다만 가명처리 방법(또는 기술)에 따른 보호조치 사항을 개인정보처리자가 스스로 준수할 수 있는 제도적 장치는 필요하다 할 것이다. 예를 들면 암호화 강도, 키관리 등의 사항 등을 포함된 기관 자체 영향평가 등이 한 예일 것이다.

가명정보의 파기 측면에서는 제28조의7(적용범위)에서 법제21조 개인정보의 파기 조항을 적용 제외로 하였으나 시행령(안)의 안전성 확보조치 기준에서 "가명정보의 처리 목적이 달성되거나 가명정보 보유 기간이 경과한 때에는 그 가명정보의 지체 없는 파기"하도록 하였다는 점에서 개인정보처리자의 혼선이 야기될 수 있는 가능성이 있다. 가명정보는 재식별의 가능성이 있는 부분을 집중적으로 관리하되 가명정보의 활용은 개인정보에 비해 규제가 완화되어야 한다는 것이 입법취지라 판단된다.

4. 결 론

지금까지 가명정보의 국내외 개념과 가명처리를 위한 비식별 조치 기준과 가명정보 처리단계별 보호조치 사항에 대해 살펴보았다. 해쉬함수를 이용한 가명처리와 암호화 등을 이용한 가명처리시 안전한 관리를 전제로 한 이용활성화를 꾀하기 위해 앞으로 보호대책 수립시 고려사항을 정리하면 첫째, 개인정보보호법 개정에 따라 개인정보, 가명정보, 익명정보 및 비식별 조치 등에 대한 정의(definition)를 무엇보다 명확하게 하는 것이 중요하다. 둘째, 가명처리에 대한 기술적 조치방법(비식별조치)은 국내외 환경과 기준을 고려하여 제시하여야 한다. 셋째, 기존의 개인정보 비식별조치 가이드라인의 가명처리에 대한 적정성 평가와 같이 개인정보처리자가 스스로 평가할 수 있는 제도적 장치를 마련할 필요가 있다.

끝으로 신용정보보호법 시행령(안) 제34조의5(가명처리・익명처리에 관한 행위규칙) 제1항의 경우 금융위원회가 별도로 관리적·기술적·물리적 보호조치 기준을 별도 고시하도록 규정하고 있는데 이는 데이터 3법 개정 이전과 같이 규제 대상에 혼란이 야기될 수 있으므로 중장기적으로 안전성 확보조치의 경우 기본적으로 개인정보보호법의 규정을 준수하고 추가적인 사항만 신용정보보호법에 포함되도록 개정할 필요가 있다.

참고문헌

[1] https://gdpr.kisa.or.kr

[2] 오길영 “데이터 비식별화 논의의 쟁점과 맹점”, 한국공법학회 공법연구 제45집 제2호, pp. 289-321, 2016.

[3] 한국인터넷진흥원 개인정보보호 핫이슈 심층 분석 보고서, “개인정보 비식별화 관련 해외 현황 및 사례”, 2016.

[4] ‘NISTIR 8053’, http://dx.doi.org/10.6028/NIST.IR.8053.

[5] 김배현, 권영일, “개인정보 비식별 제도 해외 동향 및 사례”, 한국지능정보시스템학회 학술대회논문집, pp. 88-89, 2017.

[6] 이인호, “「개인정보 보호법」상의 '개인정보'개념에 대한 해석론”, 정보법학 제19권 제1호, pp. 59-87, 2015.

[7] 국무조정실 외 5개 관계부처, “개인정보 비식별 조치 가이드라인”, 2016.

[8] TTA ICT Standard Weekly, “ITU-T SG17 빅데이터 비식별화 표준화 동향”, 2017.

[9] ENSIA, “Pseudonymisation techniques and best practices”, 2019.

[10] EU (2014), “Opinion 05/2014 on Anonymization Techniques”, EU Article 29 Data Protection Working Party.

[11] ISO/IEC 20889:2018(en) Privacy enhancing data de-identification terminology and classification of techniques.

[12] 개인정보보호법학회, “개인정보보호법 개정안(인재근 의원 대표발의)에 대한 의견서”, 2019.

[13] 개인정보보호법시행령 입법예고(안), https://www.mois.go.kr/frt/sub/a05/publicHearing/screen.do#epeopleFrameFocus

[14] 이현승/송지환, ‘개인정보 비식별화기술의 쟁점 연구’, 소프트웨어정책연구소, 2016.

[15] 서인덕, ‘개인정보의 비식별화 방안에 대한 연구”, 서울대학교 융합과학기술대학원, 이학석사논문, 2020.

[16] 최광희, “위험도 기반의 가명정보 활용을 위한 정책 프레임워크 연구”, 전남대학교대학원 정보

보호 협동과정, 박사학위논문, 2019.
[17] 이대희, “개인정보 보호 및 활용 방안으로서의 가명·비식별정보 개념의 연구”, 정보법학회 논문지, 제21권, 제3호, pp 217-250, 2017.
[18] 이루리, ‘개인정보의 비식별화 제동의 쟁점 및 개선방안에 관한 연구”, 순천향대학교 석사논문, 2016.
[19] 고학수/최경진, “개인정보 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구”, 개인정보보호위원회, 2015
[20] 김나루, “빅데이터 환경에서 개인정보의 익명화 또는 비식별화에 관한 비교법적 연구”, 세계헌법연구 제25권 제2호, pp. 131-163, 2019.
[21] Mark Elliot, Elaine Mackey, “The Anonymisation Decision-Making Framework”, UKAN, 2016.

〔저자소개〕

차 건 상 (Gun-Sang Cha)

2012년 숭실대학교 컴퓨터학 박사
2006년 행정안전부 정보보안 전문위원
2015년 건양대학교 사이버보안공학과 교수
2019년 건양대학교 정보통신원장

email : chagunsang@konyang.ac.kr