생체인증을 위한 프라이버시 중심 설계와 보안사용성에 대한 연구

문정현*, 김예은*, 최수빈**, 이일구*
*성신여자대학교 융합보안공학과
**성신여자대학교 수학과
moon_aver@naver.com, 20190893@sungshin.ac.kr, ssbb0322@gmail.com,
iglee@sungshin.ac.kr

A Study on Privacy by Design and Usable Security for Biometric Authentication

Jung-Hyun Moon*, Ye-Eun Kim*, Su-Bin Choi**, Il-Gu Lee*
*Dept. of Convergence Security Engineering, Sungshin Women's University
**Dept. of Mathematics, Sungshin Women's University

요 약

최근 편리하고 안전한 생체인증 기술이 핀테크 서비스의 필수 구성요소가 되고 있다. 그러나 생체인증의 보안성과 정확도 향상을 위해 많은 양의 생체 정보를 활용한다면 연산과정이 복잡해져서 속도 저하와 개인정보 유출 가능성이 높아지고, 실시간성을 제공하기 위해 생체 정보의 일부만 샘플링하여 활용하는 경우에는 보안성과 정확도가 열화되는 Trade-off 문제가 있다. 종래의 생체인증 기술은 메타데이터를 이용한 필터링의 경우 데이터 간의 상관성이 고려하지 않아서 개인정보보호와 사용자 편의성에 한계가 있었다. 본 연구에서는 상관성 높은 데이터를 활용하여 프라이버시 중심의 생체인증 설계와 보안사용성을 향상시키는 방안을 제안한다.

1. 서론

2020년 5월 국내 전자서명법의 개정으로 특정 기관에서만 발급 가능했던 인증서가 민간 기관에서도 발급 가능해져 편리한 민간 전자서명에 관한 연구가 활발히 이루어지고 있다[1]. ID/패스워드, OTP(One Time Password), 보안카드와 같이 지식 및 소유에 기반한 기존의 인증 방식은 분실되거나, 도난될 수 있다[2]. 이에 반해 생체인증 기술은 개인의 고유한 특징을 이용한 인증 방식으로 위조나 변조의 위험이 낮고 분실되기 어려워서 높은 신뢰성과 편의성을 제공하는 차세대 보안 기술로 주목받고 있다[3]. 그러나 생체인증은 보안성을 개선하기 위해 많은 생체정보를 이용할 경우 연산과정이 복잡해져 속도 저하와 개인정보 유출 가능성이 커지고, 속도를 개선하기 위해 인증에 활용하는 생체정보를 줄이면 보안성이 약화되는 Trade-off 문제가 존재한다. 실시간성이 요구되는 핀테크 애플리케이션에서 빠른 인증 속도와 사용자 편의성을 위해 보안성을 희생하는 방식으로 서비스가 구현되고 있다. 그러나 생체정보는 고유성과 불변성 때문에 한 번 유출되면 변경할 수 없으므로 다른 인증 수단보다 보안이 중요하다. 속도를 개선하기 위해 다른 개인정보를 활용할 경우 인증에 필요한 개인정보 양이 늘어날 뿐만 아니라, 데이터 3법으로 인해 개인정보 비식별 조치를 통해 개인 식별 요소를 제거하는 등의 솔루션이 필요하다. 본 논문에서는 최소한의 개인정보를 이용하여, 상관성 높은 메타데이터 필터링을 수행함으로써 빠르고 프라이버시 유출 위험을 줄이는 생체인증 시스템을 제안한다.

논문의 구성은 다음과 같다. 2장에서는 메타데이터를 이용한 필터링 기법에 관한 연구를 소개한다. 3장에서는 기존 연구의 한계점 및 해결법을 제시하고, 최근 인식 기법에 관한 연구 사례를 통해 문제의 가능성을 분석한다. 4장에서는 결론과 향후 과제를 제시하며 마무리한다.

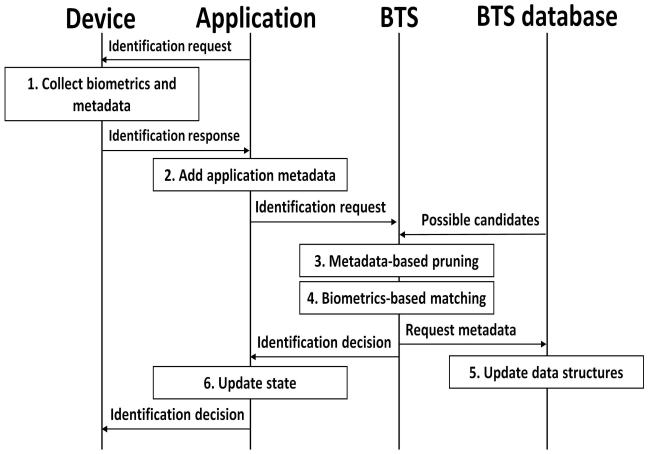
2. 관련 연구

이혜인[4]은 사용자의 정보보호를 위해 암호화된 청취자 선호도 데이터와 음악 메타데이터만을 사용해 추천 음악 목록을 제공하는 방법을 제시한다. 음악 메타데이터 추출 시 청취자가 식별될 수 있는 ID는 암호화하고 이름, 성별, 거주지 등 식별 가능한 정보는 필터링하여 개인정보를 보호한다. 이후 정규

화된 청취 횟수 데이터를 음악 선호도 데이터로 변환시키고, 음악 메타데이터와 선호도 데이터를 이용해 음악 간 유사도를 계산한다. 이를 통해 강화된 프라이버시와 함께 청취하지 못했던 음악에 대한 선호도를 예측하고, 사용자들은 추천 목록을 생성할 수 있다. 김재영과 이석원[5]은 영화와 영화 메타데이터의 유사성을 도출해 온톨로지를 구축하여 사용자가 예상하지 못한 영화를 추천할 수 있는 기법을 제안하였다. 정종진 외 2명[6]은 데이터의 특징에 맞춰 메타데이터를 자동 추출하여 데이터 검색 정확도를 개선하는 연구를 수행했다. Christian Gehrmann 외 2명[7]은 메타데이터 필터링을 사용하는 생체인증과 함께 대규모 사용자를 효율적으로 인증하고, FAR(False Acceptance Rate) 을 개선하는 방안을 제시하였다. 이를 위해 사용자는 생체정보 제출 시 위치, 이름, 나이 등의 정보를 포함한 메타데이터를 함께 제공하였다. 이를 통해 한 정보만 필터링에 이용하는 것보다 여러 정보를 결합하는 경우가 등록 유저가 증가하더라도 속도와 재현율 향상에 유리하다는 것을 보였다.

3. 생체인증을 위한 프라이버시 중심 설계와 보안 사용성

특히 연구[7]은 생체인증 서비스를 제공할 때, 사용자의 개인정보가 포함된 메타데이터를 필터링에 활용하여 인증 속도와 재현율을 개선했다. 그림 1은 선행연구에서 제안하는 시스템의 인증 과정을 도식화한 것으로, 단말기, 애플리케이션, 신뢰할 수 있는 중앙 집중식 식별 서비스(BTS, Biometrics Trusted Service), BTS 데이터베이스의 구성요소로 이루어져 있다. 인증 과정은 다음과 같다. 애플리케이션은 단말기로 식별 요청을 발송하고, 단말기는 생체정보와 위치, 이름, 나이 등의 메타데이터를 수집하여 애플리케이션에 응답한다. 이후 생체정보와 메타데이터는 BTS로 전달되어, 메타데이터를 기반으로 데이터를 축소한 후 생체인증을 진행한다. 마지막으로 인증 여부가 애플리케이션과 장치에 전달된다. 하지만 필터링에 사용된 위치, 이름, 나이 등의 경우, 구현에 사용된 지문과의 상관성을 확인하기 어렵다. 해당 연구에서는 필터링에 활용된 데이터와 지문 간의 상관성이 낮아 데이터 분포에 따라 성능 차이가 발생할 수 있다. 또한, 데이터간 상관성이 낮아서 생체인증을 위한 필수 정보가 아닌 정보를 수집·활용하고 있어 개인정보 유출 시 피해 규모가 커진다.



(그림 1) 메타데이터를 이용한 생체인증 절차[7]

따라서 본 논문에서는 생체 정보와 메타데이터의 상관성을 고려함으로써 보안사용성을 더욱 개선할 수 있는 메타데이터 기반 필터링 기법을 제안하고자 한다.

본 논문에서 제안하는 시스템은 단말기, 분석 서버, 인증 서버, 데이터베이스로 구성된다. 사용자 단말기에서는 지문 특징점에 관한 정보를 추출하고 암호화하여 분석 서버로 전달한다. 분석 서버는 수신한 정보를 바탕으로 생체정보의 특징점을 추출하여 메타데이터를 생성한다. 예를 들어 지문의 경우, 지문은 융선이 만드는 무늬의 배열에 따라 궁상문(Arch), 솟은 궁상문(Tented Arch), 좌제상문(Left loop), 우제상문(Right loop), 와상문(Whorl) 5종류로 나뉘게 되는데[8], 이를 기반으로 정규화를 시키는 것이다. 이때, 데이터 간의 상관성은 메타데이터에 포함되는 데이터로 충족되며, 이는 방향벡터, 특징점 인접 좌표와 같은 생체정보의 특징점을 추출한 정보로 구성되어 있다. 이를 바탕으로 사용자 인증을 수행할 경우, 상관성 높은 메타데이터 필터링이 수행된다. 이로 인해 사용자는 생체인증 시 속도와 정확도가 높은 편리한 인증을 진행하고, 인증에 필요한 개인정보의 양을 줄여 프라이버시 유출 가능성을 개선한 생체인증을 할 수 있다.

최근 인공지능 분야에서는 SVM(Support Vector Machine), CNN(Convolutional Neural Networks)과 같은 기계학습 알고리즘 기반의 생체 인식 연구가 활발히 이루어지고 있다. 이때, 데이터 간 상관성이 높은 경우 성능이 향상되는 결과를 얻을 수 있다.

Huizhong Chen 외 2명[9]은 수동적인 데이터 라벨링 없이 소셜 미디어에서 이름과 대응되는 얼굴 이미지를 사용하여 SVM 기반의 이름 추측 모델을

제시하였다. 이 모델은 나이 범주의 오류를 허용할 때 나이별 이름 분류에서 88%의 정확도를 보인다. 이름은 실제 정보를 바탕으로 부여되어 있어 주어진 이름과 다양한 얼굴 특징, 나이 등의 속성 사이에 상관성이 높게 나타난다. 따라서 무작위적으로 주어진 데이터를 기반으로 한 추측보다 이름 기반 추측 모델이 더 향상된 성능을 보인다. 이를 통해 데이터 간 상관관계가 성능 향상에 도움을 준다는 가능성을 확인할 수 있다.

Sunil Kumar와 Ilyoung Chong[10]은 상관관계 및 기계학습 기반 접근 방식을 통해 데이터 분류에 효과적인 영향을 미치는 데이터를 식별하고, 추출한 특징 간에 상관관계를 이용하여 우울장애와 감정 상태를 예측하였다. 상관관계와 분류 결과, 온도, 대기압, 오존이 우울증에 강한 영향을 미치며 9개의 데이터를 결합하여 사용할 때는 최고의 성능을 보이는 것을 입증하였다. 그리고 선형 상관 모델을 이용해 관계 식별함으로써, 상관성 높은 데이터를 적절히 사용할 경우 데이터 분류하는데 높은 효과를 얻을 수 있음을 보였다.

Grigory Antipov 외 3명[11]은 분석 대상, CNN 깊이, 사전 훈련, 단일/다중 작업 전략과 같은 CNN 훈련에 있어 중요한 요소를 분석함으로써 최첨단 GR(Gender Recognition) 및 AE(Age Estimation) 모델을 설계하였다. CNN 모델로 데이터를 학습할 때 대상 나이를 나타낼 수 있는 최적의 방법은 라벨 분포 나이 인코딩(LDAE)이며, AE의 경우 GR보다 깊은 CNN 아키텍처가 필요하고, 얼굴인식(FR) 사전 훈련은 성별과 나이에 대한 CNN의 효과적인 훈련을 가능하게 한다는 결과를 도출했다. 이 연구는 ChaLearn Apparent Estimation Challenge에서 0.2411 값으로 다른 경쟁업체(2위=0.3214) 대비 가장 낮은 MAE(Mean Absolute Error) 보이며 GR과 AE CNN의 효과적인 훈련을 가능하게 하였다. 이를 통해 나이와 같은 관련 데이터를 학습 모델에 활용할 경우 성능이 더 좋아진다는 것을 입증하였다.

위 연구들을 통해 기존에는 고려되지 않은 요소 간의 상관성을 고려한다면, 데이터 분류의 성능을 높일 수 있음을 알 수 있다. 따라서 생체인증에서 상관성 높은 데이터를 추출하여 메타데이터 필터링을 진행할 경우 최소한의 개인정보만으로 수집, 처리됨에 따라 속도 및 정확도와 같은 인증 성능을 개선할 수 있을 것이라 예상한다.

4. 결론

본 논문에서는 메타데이터 및 생체인증 개선 관련 연구 동향을 확인하였다. 메타데이터는 데이터에 대한 정보를 포함하고 있어 이를 활용할 경우 데이터를 효과적으로 검색할 수 있다. 그러나 메타데이터 필터링에서 데이터 간 상관성이 고려되고 있지 않아 프라이버시 중심의 생체인증 연구가 진행되어야 한다. 따라서 관련 연구 사례 분석을 통해 상관성 있는 데이터를 메타데이터에 적용할 경우 성능 향상이 가능함을 확인하였다. 후속 연구로는 생체정보와 상관성이 높은 특징점 연구 동향을 확인하고, 메타데이터를 필터링을 활용한 생체인증 연구[7]를 바탕으로 상관성 높은 데이터와 결합한 연구를 제안하고자 한다. 또한, 본 논문에서 제안한 내용을 구현하여 속도와 재현율을 측정하고, 선행연구와의 비교를 통해 상관성이 높은 데이터를 사용할 경우 기존 대비 편의성과 프라이버시 유출 위험이 개선하는 연구를 진행할 계획이다.

Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1F1A1061107)과 2021년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0008703, 2021년 산업혁신인재성장지원사업)을 받아 수행된 연구임.

참고문헌

[1] 강효관, “국내 인증 기술 및 서비스 현황″, 정보보호학회, vol.30, no.3, pp.31-36, 2020.

[2] 박영수, 이병엽, “일회용 세션을 활용한 인증정보 기반의 사용자 인증 방안″, 한국콘텐츠학회, vol.19, no.7, pp.421-426, 2019.

[3] 박희진, 이윤호, “생체 인증에서의 프라이버시 보호 기술″, 한국정보기술학회, vol.16, no.4, pp.109-122, 2018.

[4] 이혜인, “메타데이터를 이용한 음악 추천 기법″, 부경대학교 석사학위논문, 2019.

[5] 김재영, 이석원, “온톨로지 기반 영화 메타데이터간 연관성을 활용한 영화 추천 기법″, 지능정보연구, vol.19, no.3, pp.25-44, 2013.

[6] 정종진, 김경원, 김구환. “데이터셋 검색 지원을 위한 메타데이터 자동 추출에 관한 연구”, 한국통신

학회, pp. 867-868, 2020.
[7] Gehrmann, C. Rodan, M. Jönsson, N. “Metadata filtering for user-friendly centralized biometric authentication", EURASIP J. on Info. Security 2019, 2019.
[8] Kalle Karu, Anil K. Jain, “Fingerprint classification”, Pattern Recognition, Volume 29, Issue 3, pp.389-404, 1996.
[9] Huizhong Chen, Andrew C Gallagher, Bernd Girod, “The Hidden Sides of Names－Face Modeling with First Name Attributes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, no.9, pp.1860-1873, 2014.
[10] Kumar, Sunil, and Ilyoung Chong, “Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States”, International journal of environmental research and public health, vol.15, no.12, 2018.
[11] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, Jean-Luc Dugelay, “Effective training of convolutional neural networks for face-based gender and age prediction”, Pattern Recognition, Vol.72, pp.15-26, 2017.