Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.5, Oct. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.5.1059

얼굴과 사용자 입력정보를 이용하여 안전한 키를 생성하는 방법*

김 혜 진,[1†] 최 진 춘,[1] 정 창 훈,[1] 양 대 헌,[1] 이 경 희[2‡]
[1]인하대학교, [2]수원대학교

A Method for Generating Robust Key from Face Image and User Intervention*

Hyejin Kim,[1†] JinChun Choi,[1] Chang-hun Jung,[1] DaeHun Nyang,[1] KyungHee Lee[2‡]
[1]Inha University, [2]The University of Suwon

요 약

바이오해싱 기법은 생체 인식 템플릿으로부터 효과적으로 비트스트링 키를 생성할 수 있지만, 토큰 같은 사용자 입력 요소에 대한 의존도가 높아 토큰 도난 시 성능이 하락한다. 이러한 한계점을 개선하기 위하여, 본 논문에서는 얼굴 사진과 사용자 입력정보로부터 안전한 키를 생성하는 기법을 제시한다. 바이오해싱 기법과 GPT 기법을 사용하여, 인증 시 사용자 입력정보에 대한 의존도를 조정하고, 충분한 길이의 안전한 키를 생성하도록 구성하였다. 제시한 기법을 입증하기 위하여 다양한 실험을 진행하고 결과를 보였다.

ABSTRACT

Even though BioHashing scheme can effectively extract binary string key from analog biometrics templates, it shows lower performance in stolen-token scenario due to dependency of the token. In this paper, to overcome this limitation, we suggest a new method of generating security key from face image and user intervention. Using BioHashing and GPT schemes, our scheme can adjust dependency of PIN for user authentication and generate robust key with sufficient length. We perform various experiments to show performance of the proposed scheme.

Keywords: biometrics, key generation, cryptography, BioHashing, GPT

I. 서 론

공개키 기반 구조(Public Key Infrastructure)는 부인 방지(non-repudiation)가 가능한 특성 덕분에 전자 서명 시스템 구축에 널리 이용되고 있다. 특히 온라인 금융 서비스를 이용하기 위해, 사용자들은 패스워드와 공인 인증서를 결합한 전자 서명 시스템을 사용하고 있다. 그러나 2015년, 공인 인증서 의무 사용 폐지 법안이 시행되고, 불편한 기존 공인 인증서 시스템의 대한 사용자들의 불만의 목소리가 높아지며 이를 대체할 수 있는 새로운 기술의 필요성이 대두되고 있다. 그리고 이러한 대체 기술 중 하나로써 각광받고 있는 것이 생체 인식(biometrics)을 이용한 본인 인증 기술이다[1].

생체 인식은 개개인의 고유한 얼굴, 지문, 홍채 등의 생체 인식 정보를 이용하여 개별 사용자를 인식하는 방식이다. 생체 인식 정보는 기존의 패스워드와 달리 사용자의 기억에 의존할 필요가 없고, 토큰 장치처럼 분실할 위험도 없어 차세대 본인 인증 요소로

Received(05. 31. 2017), Modified(09. 26. 2017), Accepted(10. 11. 2017)
* 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2016-0-00097, 비대면 본인확인을 위한 바이오 공개키 기반 구조 기술 개발)
† 주저자, sunnyq@isrl.kr
‡ 교신저자, khlee@suwon.ac.kr(Corresponding author)

써 이점을 가지고 있다. 더욱이 최근 생체 인식 센서가 소형화되어 모바일 기기 등에 탑재되는 등 보편화되었고, 이를 이용한 본인 인증 서비스도 확대되고 있다[1].

그러나 생체 인식이 전자 서명에 사용되는 패스워드와 공인인증서를 완전히 대체하기는 쉽지 않다. 생체 인식 정보는 연속적이고 불확실성이 높은 아날로그 데이터이기에, 이를 패스워드나 공인인증서와 같은 디지털 데이터로 가공하기 쉽지 않기 때문이다. 센서가 읽어 들인 생체 인식 데이터는 많은 노이즈 데이터를 포함하기 때문에, 항상 동일한 값을 가지는 템플릿을 도출해내기 어렵다. 일부에서는, Fast IDentity Online(FIDO) 연합과 같은 국제 생체 인증 기술 표준을 이용해 기존의 시스템을 대체하려는 노력을 기울이고 있지만, 부인 방지가 가능한 전자 서명 시스템에 적용하기 쉽지 않다[2]. 이와 같은 이유로 생체 인식은 아직까지 낮은 단계의 본인 인증 수단으로만 활용되고 있는 실정이다.

이러한 한계점을 극복하기 위해, 본 논문은 얼굴과 사용자 입력정보를 이용하여 암호학적 시스템에 적용 가능한 키를 생성하는 방법을 제안한다. 얼굴 전체 영역에서 추출한 얼굴의 특징 정보와 사용자 입력정보(PIN, 패스워드, 토큰 등)를 이용하여 항상 동일하며 충분한 길이의 비트스트링 형태의 키를 추출하고, 기존 기법들이 가지고 있던 사용자 입력정보에 대한 보안의 편중성을 낮추어 토큰 도난(stolen-token) 시나리오의 취약점을 해결하였다.

본 논문의 구성은 다음과 같다. 2장에서는 대표적인 얼굴 인식 알고리즘과 생체 인증 기법들에 대해 설명한다. 3장에서는 본 논문이 제안하는, 얼굴과 사용자 입력정보를 사용한 이중(two-factor) 키 생성 시스템을 설명하고, 4장에서 얼굴 이미지 데이터베이스를 사용한 키 생성 실험 결과와, 정상 사용자와 위조자들 사이에서 키 생성 실험 결과를 보인다. 5장에서는 실험결과에 대한 평가와 활용, 더 보완해야 할 점에 대해 논의하고 결론으로 마무리 짓는다.

II. 관련 연구

2.1 얼굴 인식 알고리즘

2.1.1 PCA 기반 분류기: PCA, 2DPCA, DiaPCA

주성분 분석(Principal Component Analysis, PCA)은 고차원 데이터에서 가장 분산이 큰 주성분 축을 찾아, 데이터 간의 차이를 잘 구분할 수 있는 저차원 데이터로 변환하는 분석 기법이다. M.A. Turk 등[3]은 얼굴 이미지에 PCA를 적용하여 얼굴 특징 벡터를 생성하는 기법을 제안하였고, 이를 eigenface 기법이라고 한다. 기존 eigenface 기법에서는 2차원 행렬로 표현되는 이미지를 1차원 벡터 형태로 변환한 뒤, 특징 벡터를 생성했다. 그러나 크기가 작은 이미지일지라도 1차원으로 변환할 경우, 차원의 크기가 기하급수적으로 증가하기 때문에, 특징 벡터를 구하기 위한 연산량도 같이 증가하게 된다. J. Yang 등[4]은 2차원 이미지를 변환 없이 PCA 기법에 적용할 수 있는 2DPCA 기법을 제안하였는데, 연산량이 적어 고해상도의 이미지 사용이 가능하며, 인식률도 기존 성능보다 향상됨을 보였다. D. Zhang 등[5]은 2DPCA를 기반으로 하여, 얼굴 이미지의 대각선 변환으로 이미지에 비대칭성을 부여해 2DPCA보다 개선된 성능을 보이는 DiaPCA 기법을 제안했다.

2.1.2 LDA 기반 분류기 : LDA, R-LDA

선형 판별 분석(Linear Discriminant Analysis, LDA)은 클래스로 분류된 데이터들을 클래스 내 분산(within-class scatter)는 최소화하고 클래스 간의 분산(between-class scatter)는 최대화 하는 축을 찾아 데이터를 투영시켜 분류하는 선형 분류 기법이다. 기본 LDA는 PCA처럼 2차원의 이미지를 1차원 벡터로 변환하여 입력해야 하고, 결과물로 차원이 감소한 특징 벡터를 추출한다. 그러나 입력된 샘플의 차원이 샘플의 수보다 클 경우, 클래스 내 분산 행렬이 보이는 특이성으로 인해 인식률이 감소한다(Small Sample Size problem, SSS)[6]. J. Lu 등[7]은 이를 해결하기 위하여 클래스 간 분산 행렬을 정규화 파라미터로 결합된 클래스 간 분산, 클래스 내 분산 조합 행렬로 대체하여 SSS 문제를 개선한 Regularized LDA(R-LDA) 기법을 제안하였다.

2.2 생체 키 추출 방법 및 인증 기법 시스템

2.2.1 Random Projection

랜덤 프로젝션(Random Projection)은 가공하기 어려운 고차원의 데이터를 저차원의 부분 공간 상으로

투영하여 차원을 축소하는 기법이다[8]. 랜덤 프로젝션 행렬의 원소는 가우시안 분포를 따르며, 데이터의 차원을 축소할 때도 데이터 간의 거리 등과 같은 데이터들의 특성이 유지된다[9].

2.2.2 BioHashing

A. Goh 등[10]은 PCA를 적용한 얼굴 특징 벡터를 양자화시키고, 임계값(threshold)을 이용한 이진화의 결과로 비트스트링을 생성하는 바이오해싱(BioHashing) 방법론을 제안했다. A. Teoh 등[11]은 바이오해싱 방법론을 발전시켜 사용자 토큰과 지문을 이용하여 비트스트링을 생성하는 이중 인증 기법을 제안하였다. Teoh의 바이오해싱 기법은 지문으로부터 추출한 특징 벡터에 사용자의 토큰으로 생성한 랜덤한 숫자들을 내적하고 이진화를 거쳐 비트스트링을 생성한다. 이 방법을 골격으로, 2006년 D. Ngo 등[12]은 얼굴 이미지를 이용한 바이오해싱 기법을 제시하였는데, 지문 데이터와는 달리 고차원인 얼굴 이미지의 차원 축소를 위하여 사용자의 비밀 키로 생성된 랜덤 프로젝션 행렬을 적용하고, 항상 동일한 비트스트링 생성을 위하여 오류를 보정하는 방법인 Error-Correcting Code(ECC)로 보정하는 과정을 추가하였다. 그러나 A. Kong 등[13]이 바이오해시가 토큰으로 만들어진 랜덤한 숫자들에 대한 의존도가 너무 높아, 토큰 도난(stolen-token) 시나리오에서는 외려 생체 인식 정보만 사용할 때보다 인식 성능이 하락하여, 동일한 토큰을 가진 다른 사용자에 대한 FAR이 높다는 것이 확인되었고, 이를 보완하기 위해 수정된 바이오해싱 기법들이 제기되었다[14][15][16].

2.2.3 Helper Data Scheme(HDS) 시스템

2005년 Kevenaar 등[17]은 보조 데이터 기법(Helper Data Scheme, HDS)을 이용하여 얼굴로부터 이진(binary) 특징 벡터를 생성하는 방법을 제안하였다. 등록 시 사용자의 얼굴 이미지로부터 특징 추출을 거쳐 비트스트링을 생성하고, 이를 위한 ECC를 생성해 저장해 두었다가 인증 시 이를 활용한다. 인증 시 생성된 비트스트링에 ECC 코드를 붙여 디코딩(decoding)하면, 등록한 비트스트링과 차이를 보이는 에러들을 보정하여 동일한 비트스트링을 만들 수 있다.

2.2.4 General Permutation Transformation(GPT) 기법

2005년 강전일 등[18]은 사용자의 입력정보로부터 치환 행렬을 생성하여 프라이버시 보호 및 템플릿 취소가 가능한 이중(two-factor) 얼굴 인증 기법을 제안하였다. 단순 치환 행렬 변환 기법(Simple Permutation Transformation Scheme, SPT Scheme)이라고 명명된 이 기법은, 프로젝션 행렬을 통하여 얼굴 정보를 복원할 수 있다는 점을 보완하기 위하여 사용자의 비밀번호로부터 생성한 치환 행렬 $\mathbf{P}$를 이용하여 프로젝션 행렬 $\mathbf{U}$를 보호하는 방법으로 사용자의 생체 인식 정보의 유출을 막는다. 또한, 연산량이 적은 행렬 계산을 이용하여 생체 템플릿에 취소 가능 특성을 부여하였다. 2014년에 강전일 등[19]은 SPT 기법이 가지고 있던 사용자 입력정보의 의존도를 낮추고, 프로젝션 행렬 $\mathbf{U}$의 유출 보호를 강화하는 General Permutation Transformation (GPT) 기법을 제안하였다.

III. 제안하는 기법

본 기법은 얼굴 이미지와 사용자 입력정보를 이용하여 동일한 사용자일 경우 동일한 비트스트링을 생성한다.

제안하는 기법에서는, 등록 시에 등록하고자 하는 사용자의 얼굴 이미지 여러 장을 전처리 과정을 거쳐 이미지 내 노이즈를 제거한다. 전처리를 거친 얼굴 이미지에 얼굴 인식 알고리즘을 이용해 특징 벡터를 추출하고, 특징 벡터에 랜덤 프로젝션 행렬을 내적하여 템플릿을 생성한다. 생성된 템플릿 벡터를 임계값을 기준으로 0과 1로 이진화하여 키를 생성하고, 키에 대한 ECC를 생성하여 저장한다.

이 기법의 인증 시에는, 등록과 동일한 전처리 과정을 거친 후, GPT와 랜덤 프로젝션 행렬이 적용된 기저(basis) 행렬을 통해 템플릿을 만들고 키를 생성한다. 이후, 등록 시 생성했던 ECC를 이용해 만들어진 키의 오류를 보정한다. 오류가 수정된 키를 등록할 때 생성한 키와 비교하여 사용자 인증을 진행한다.

3.1 전처리(Preprocessing)

전처리는 사용자의 얼굴 이미지 $\mathbf{x}$의 노이즈를 제거하고, 광원과 같은 다양한 환경적 요인들의 영향을 최소화하기 위한 이미지 보정 작업을 의미한다. 얼굴 이미지를 이목구비를 중심으로 가로 N, 세로 N 픽셀 크기로 재단하여 배경을 제거하고, 이미지 촬영 시 광원으로 인하여 고르지 못한 색을 그레이스케일 필터와 히스토그램 이퀄라이징 등의 작업으로 이미지를 보정한다. 적용되는 특징 추출 알고리즘에 따라 이미지 $\mathbf{x}$의 형태 가공이 달라진다. R-LDA를 사용할 경우 $N \times N$ 크기의 이미지를 $N^2 \times 1$ 크기의 벡터 형태로 변환하고, DiaPCA를 사용할 경우는 $N \times N$인 2차원 행렬 형태를 유지하고 대각선 방향으로 순환 시프트를 시켜 비대칭성을 부여한다.

3.2 템플릿 생성(Generating Templates)

3.2.1 Feature Extraction

얼굴 이미지로부터 특징을 추출하기 위해서는 얼굴 인식 알고리즘을 통해, 다수의 샘플 이미지들을 사용하여 프로젝션 행렬 $\mathbf{U}$를 만들어야 한다. 프로젝션 행렬 $\mathbf{U}$에 $\mathbf{x}$를 내적하여 특징 벡터 $\mathbf{U} \cdot \mathbf{x} = \mathbf{y}$를 만들 수 있다. 샘플 이미지의 경우 사용하는 선형 분류 기법에 따라 사용자 별로 1장 혹은 여러 장이 필요하다.

얼굴 인식 알고리즘으로 R-LDA를 사용할 경우, 프로젝션 행렬 $\mathbf{U}$를 생성하기 위해서 클래스(사용자) n개의 해당 클래스 샘플 이미지가 필요하다. n은 임력 $\mathbf{x}$의 차원 크기에 따라 적정한 값이 달라지는데, $\mathbf{x}$의 차원이 높을수록 많은 샘플 이미지가 필요하다. R-LDA 프로젝션 행렬 $\mathbf{U}$의 크기는 $\mathbf{U} \in \mathbb{R}^{d \times N^2}$ 이고, $d \ll N^2$이다. $\mathbf{U}$ 행렬 크기의 d는 R-LDA 알고리즘 안에서 클래스 수에 따라 결정된다.

DiaPCA를 사용할 경우, 클래스 구분 없이 얼굴의 대표성을 부여할 수 있는 샘플 이미지들을 통해 프로젝션 행렬 $\mathbf{U}$를 생성한다. DiaPCA는 $N \times N$크기의 분산이 큰 순서대로 정렬된 기저 행렬을 생성하게 되는데, 그 중 상위 d개의 기저 벡터들을 선택하여 사용한다. 따라서, 프로젝션 행렬 $\mathbf{U}$의 크기는 $\mathbf{U} \in \mathbb{R}^{d \times N}$ 이고, $d \ll N^2$인 d개의 상위 기저 벡터의 개수를 의미하며, 출력하고자 하는 키의 길이에 따라 조정이 가능하다.

3.2.2 General Permutation Transformation

GPT 기법으로 생성되는 행렬 $\mathbf{P}_i$는 사용자 i의 입력정보로부터 생성된 치환 행렬로써 이미지 $\mathbf{x}$를 변환시켜 본래의 정보를 보호하기 위하여 사용된다. 전처리를 거친 $\mathbf{x}_i$에 $\mathbf{P}_i$를 내적하면 열과 행이 랜덤하게 섞이면서 본래의 얼굴 정보를 복원하기 어렵게 된다. $\mathbf{P}_i\mathbf{x}_i$로부터 특징 벡터 $\mathbf{y}_i$를 추출하기 위해서는 $\mathbf{U}$ 행렬에 $\mathbf{P}_i^{-1}$ 행렬을 내적하면 $\mathbf{P}_i^{-1}$와 $\mathbf{P}_i$가 단위행렬 $\mathbf{I}$로 바뀌고, $\mathbf{U}$에 $\mathbf{x}$를 내적할 수 있다.

$$\mathbf{U}\mathbf{P}_i^{-1}\mathbf{P}_i\mathbf{x}_i = \mathbf{U}\,\mathbf{I}\,\mathbf{x}_i = \mathbf{U}\mathbf{x}_i = \mathbf{y}_i \quad (1)$$

사용한 얼굴 인식 알고리즘에 따라서 서버에 저장되는 $\mathbf{W}_i$ 행렬을 만들 때, 행렬 $\mathbf{U}$와 $\mathbf{P}_i^{-1}$를 곱하는 순서가 달라진다. R-LDA의 경우는 isolation 행렬 $\mathbf{S}_i$, GPT의 역행렬 $\mathbf{P}_i^{-1}$, 프로젝션 행렬 $\mathbf{U}_{\mathrm{RLDA}}$가 $\mathbf{W}_i = \mathbf{S}_i\mathbf{U}_{\mathbf{RLDA}}\mathbf{P}_i^{-1}$ 순서대로 곱해진다. DiaPCA를 사용하는 경우, $\mathbf{W}_i = \mathbf{P}_i^{-1}\mathbf{U}_{\mathbf{DiaPCA}}\mathbf{S}_i$ 순서대로 곱하고, 특징 벡터를 생성할 때도 $\mathbf{x}_i\mathbf{P}_i\mathbf{W}_i = \mathbf{y}_i\mathbf{S}_i$ 과 같이 생성된다. 이러한 방법으로 만들어진 $\mathbf{W}_i$를 인증용 키를 생성할 때 불러와 사용한다.

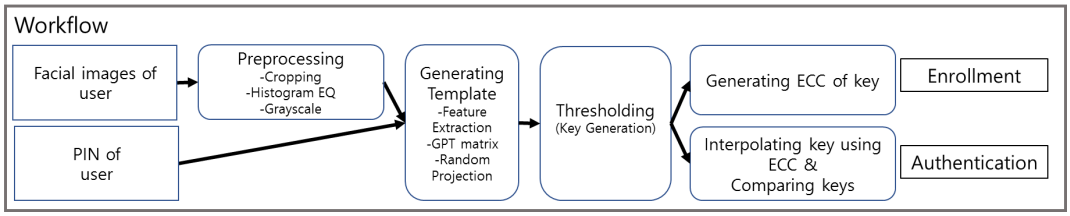


Fig. 1. Entire workflow of the suggested scheme

3.2.3 Random Projection

바이오해싱에서는 이 랜덤 프로젝션 행렬을 각 사용자의 보안 토큰을 이용해 생성하여 사용자마다 각기 다른 부분공간으로 데이터를 사상한다[14]. 반면 C. Soutar 등[20]과 M. Savvides 등[21]은 모든 사용자에게 공통 랜덤 숫자를 곱하여 랜덤 프로젝션과 유사한 취지의 효과를 적용하였다.

본 기법에서는 기존 바이오해싱의 기법처럼 데이터를 각기 다른 부분공간으로 투영시켜 거리를 비교할 경우, 특징 벡터의 차이가 왜곡될 수 있다고 판단하여 모든 사용자에 대하여 공통의 랜덤 프로젝션 행렬을 사용한다. R-LDA를 사용할 경우, $\mathbf{R}\in\mathbb{R}^{d\times d}$인 랜덤 프로젝션 행렬 $\mathbf{R}$을 $\mathbf{RW}_i=\mathbf{RS}_i\mathbf{U}_{\mathbf{RLDA}}\mathbf{P}_i^{-1}$ 순으로 하나의 행렬로 만들어 활용할 수 있다. 그러나 DiaPCA의 경우 $\mathbf{P}_i\mathbf{x}_i$가 행렬 $\mathbf{U}$의 앞쪽에 곱해지기 때문에 $\mathbf{W}_i$와 $\mathbf{R}$을 하나의 행렬로 미리 곱해놓을 수가 없다. 따라서 $\mathbf{R}$ ($\mathbf{R}\in\mathbb{R}^{N\times N}$) 을 별도의 행렬로 보관하고, 다음과 같이 연산할 수 있다.

$$\begin{aligned}\mathbf{Rx}_i\mathbf{P}_i\mathbf{W}_i&=\mathbf{Rx}_i\mathbf{P}_i\mathbf{P}_i^{-1}\mathbf{U}_{\mathbf{DiaPCA}}\mathbf{S}_i\\&=\mathbf{Ry}_i\mathbf{S}_i\end{aligned}\quad(2)$$

3.3 임계값을 통한 이진화(thresholding)

템플릿 $\mathbf{t}_i$는 $\mathbf{t}_i=\mathbf{RS}_i\mathbf{U}_{\mathbf{RLDA}}\mathbf{x}_i=\mathbf{RS}_i\mathbf{y}_i$ 혹은 $\mathbf{t}_i=\mathbf{Rx}_i\mathbf{U}_{\mathbf{DiaPCA}}\mathbf{S}_i=\mathbf{Ry}_i\mathbf{S}_i$ 연산을 통해 생성된 얼굴 특징을 담은 크기 $m\times1$의 템플릿이다. R-LDA의 경우 m은 클래스 수에 비례하여 R-LDA 알고리즘에서 계산되는 특징 벡터의 크기 값이고, DiaPCA의 경우 $m=d\times N$이다. 이진화는 $\mathbf{t}_i$의 각 실수 요소들을 임계값 τ 값을 기준으로 0과 1로 변환하는 작업이다.

$\mathbf{t}_i=\{t_{<i,1>},t_{<i,2>},\cdots,t_{<i,m>}\}$ 일 때,

$$b_{<i,j>}=\begin{cases}1 & \text{if } t_{<i,j>}\geq\tau\\0 & \text{if } t_{<i,j>}<\tau\end{cases}\quad(3)$$

와 같은 비트스트링 키 $\mathbf{b}_i=\{b_{<i,1>},b_{<i,2>},\cdots,b_{<i,m>}\}$를 생성한다. 이 실험에서는 기존의 바이오해싱 기법과 같이 $\tau=0$으로 임계값을 설정한다.

등록 시에는 샘플 이미지 n장으로부터 가장 사용자를 잘 나타낼 수 있는 비트스트링을 생성한다. 각 이미지로부터 생성된 n개의 비트스트링 $\mathbf{B}_i=\{\mathbf{b}_{<i,1>},\mathbf{b}_{<i,2>},\cdots,\mathbf{b}_{<i,n>}\}$를 생성하고, mode 연산을 사용하여 n개의 비트스트링의 각 1번째 요소에서 0과 1중 가장 빈도수가 큰 값을 취하여 등록용 비트스트링 $\mathbf{b}_i$의 $b_{<i,1>}$ 요소로 설정한다. 이와 같은 방법으로 1번째 요소부터 m번째 요소까지 mode 연산을 이용하여 등록용 $\mathbf{b}_i$를 생성한다.

3.4 ECC: BCH Code

얼굴 이미지에 대한 전처리와 성능이 좋은 얼굴 인식 알고리즘을 사용한다 하더라도, 포즈나 광원과 같은 노이즈를 모두 제거할 수 없기 때문에 항상 동일한 키를 뽑아내기는 어렵다. 이 기법에서는 바이오해싱 기법과 HDS 기법과 같이 동일한 사용자가 항상 동일한 키를 생성하기 위하여 ECC 코드 중에서 BCH Code를 이용해 전체 비트의 일부 에러 비트를 보정한다. 사용자 등록 시 생성된 키 $\mathbf{b}_i$의 ECC 코드 ecc_i를 만들어 저장하고, 인증용 키 $\mathbf{b}_i^{test}$에 ecc_i를 디코딩하여 보정한다. 비트 형식의 경우 ECC로 BCH code를 사용할 수 있고, $\mathbf{b}_i$의 길이에 따라 ecc_i의 길이, 복구 가능한 최대 오류 비트의 개수(bit capacity)를 적절하게 설정해야 한다. bit capacity가 작으면, 공격자가 인증되기 어려워지지만 정상 사용자의 거부율 역시 같이 높아진다. 반면, bit capacity가 크면, 정상 사용자의 키와 많은 차이를 보이는 공격자의 가짜 키도 보정을 통해 정상 키로 변환될 수 있다.

3.5 취소 가능한 생체 인식 키

생체 인식 정보는 변경이 불가능하다는 특성 상, 유출 사고에 매우 취약하다. 따라서 생체 인식을 이용한 시스템을 만들 때 반드시 따라와야 할 특성 중 하나가 취소가 가능해야 한다는 점이다. 본 기법은 GPT와 랜덤 프로젝션의 취소 가능한 생체 인식이라는 특성을 적용하여 생체 인식 정보의 유출 사고 시에도 생체 인식 정의 노출 없이 사용자 입력정보를 갱신하거나 새로운 키를 생성할 수 있다.

우선, 사용자의 사용자 입력정보가 노출되어 갱신해야 하는 경우가 있다. 공격자가 사용자의 입력정보

를 탈취했을 경우, 사용자 입력정보를 모를 때 보다는 인증에 성공할 확률이 높아지기 때문에 새로운 사용자 입력정보로 교체해 주어야 한다. GPT 기법에서는 다음과 같이 기존 사용자 입력정보를 새로운 입력정보로 갱신한다. 서버에 저장되어 있는 행렬 $\mathbf{W}_i$에 새로운 isolation 행렬 $\mathbf{S}_{new}$, 기존 사용자 입력정보로부터 만든 치환 행렬 $\mathbf{P}_{old}$, 새 사용자 입력정보로부터 만든 $\mathbf{P}_{new}^{-1}$를 곱해주면, 행렬 $\mathbf{U}$를 드러내지 않고도 새로운 사용자 입력정보로 인증을 진행할 수 있게 된다.

$$\begin{aligned}\mathbf{S}_{new}\ \mathbf{W}_i\mathbf{P}_{old}\mathbf{P}_{new}^{-1} &= \mathbf{S}_{new}\,\mathbf{S}_{old}\mathbf{U}\mathbf{P}_{old}^{-1}\mathbf{P}_{old}\mathbf{P}_{new}^{-1} \\ &= \mathbf{S}_{new}^{'}\ \mathbf{U}\mathbf{P}_{new}^{-1} = \mathbf{W}_i^{'}\end{aligned} \tag{4}$$

만약 보안을 위하여 키 자체를 교체해야 하거나, 다수의 사용자의 키가 드러났을 경우에는 사용자 입력정보를 교체하는 것은 물론 새로운 키를 생성할 수 있도록 해야 한다. 이를 위해 모든 사용자에 대하여 공통으로 사용하는 랜덤 프로젝션 $\mathbf{R}$을 교체하여, 모든 사용자들이 새로운 키를 생성할 수 있도록 한다.

IV. 실 험

본 장에서는 제안한 기법을 실험용으로 공개된 얼굴 이미지 데이터를 이용하여 등록된 비트스트링과 일치율을 통해 제안 기법의 성능을 측정한다.

실험 환경은 다음과 같다. CPU i5-4570, RAM 12GB, 운영체제 Windows 10, 코드 구현 프로그램은 MATLAB 2016a을 사용하였다. 사용된 얼굴 이미지 데이터베이스는 Essex FACE94이고, 151 클래스(사용자)에 각 클래스별로 19장의 이미지, 총 2,869개의 이미지를 사용했다.

각 사용자당 19장중에 임의로 1장을 선택하여 인증 테스트용으로 사용하고, 18장은 등록 시 샘플 이미지로 사용하였다. 선택되는 1장의 테스트용은 매 단일 실험을 실행할 때마다 바뀌게 되고, 이미지는 이목구비를 포함한 40×40 픽셀 크기의 그레이스케일과 히스토그램 이퀄라이징 전처리 작업을 거친 이미지를 사용하였다. 사용자 입력정보는 6자리 숫자인 PIN을 생성하여 이용하였고, GPT 기법에서 설정하는 사용자 입력정보 의존도(kd) kd＝0.6으로 설정하였다. 사용자 입력정보 의존도는 GPT 행렬의 영향을 나타내는 수치로써, 사용자 입력정보로부터 생성되는 GPT 행렬의 무작위성을 나타낸다. GPT 행렬의 무작위성이 높을수록, 사용자 입력정보가 가지는 보안성이 높아져 정확한 사용자 입력정보 없이는 동일한 템플릿을 생성할 수 없다. 해당 논문[19]에서는 반드시 kd의 값이 0.6이상 최대 1.0으로 설정할 것을 권고하였다. 본 논문에서는 얼굴과 사용자 입력정보가 주는 영향을 동일한 비중으로 비교하기 위하여 kd 값을 권고 최소 사항인 0.6으로 설정하였고, 실제 응용에서는 보안을 위하여 1.0으로 설정하는 것이 좋다.

키 생성의 결과물로, R-LDA를 이용했을 때는 $m_{\mathrm{RLDA}}=120$ 길이의 비트스트링 키를 생성하였고, DiaPCA를 이용했을 때는 $d=6$으로 설정하여 $m_{\mathrm{DiaPCA}}=240$ 길이의 비트스트링 키를 생성하였다. 실험의 표본 종류는 정상 사용자(Genuine User, 얼굴과 사용자 입력정보 모두 일치), 얼굴 위조자(Face Imposter, 얼굴 불일치 사용자 입력정보 일치), 사용자 입력정보 위조자(PIN Imposter, 얼굴 일치 사용자 입력정보 불일치), 일반 위조자(General Imposter, 얼굴, 사용자 입력정보 모두 불일치)로 4가지 경우로 나누고, R-LDA, DiaPCA 알고리즘으로 각각 100번씩 실행하였다.

결과는 Fig. 2, Fig. 3, Table 1, Table 2과

Table 1. FAR, FRR results using R-LDA

prob. / bit	FRR	FAR-Ⅰ*	FAR-Ⅱ**	FAR-Ⅲ***
5	0.4172	0	0	0
10	0.0927	≈0	0	0
15	0.0132	0.0015	0	0
20	0.0066	0.0146	0	0
25	0	0.0621	0	0
30	0	0.1950	0	0

*Face imposter **Password imposter ***General imposter

Table 2. FAR, FRR results using DiaPCA

prob. / bit	FRR	FAR-Ⅰ*	FAR-Ⅱ**	FAR-Ⅲ***
10	0.5003	0	0	0
20	0.0809	0	0	0
30	0.0102	≈0	0	0
40	0.0009	0.0002	0	≈0
50	0.0004	0.0008	0.0003	≈0
60	0	0.0030	0.0032	0.0012

*Face imposter **Password imposter ***General imposter

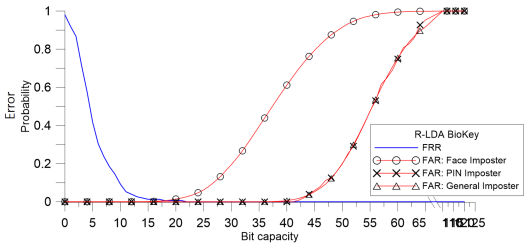


Fig. 2. FAR, FRR graph using R-LDA

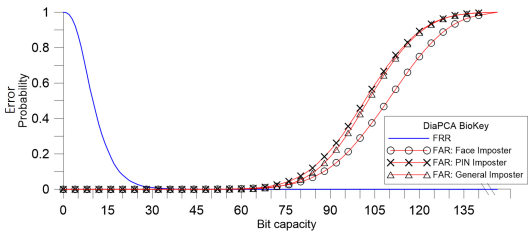


Fig. 3. FAR, FRR graph using DiaPCA

같다.

4.1 FRR

정상 사용자에 대한 인식률 결과는 FRR(False Reject Rate, 오거부율) 표와 그래프를 통해 파악할 수 있다. *bit_capacity*는 ECC를 이용해 복구 가능한 최대 에러 비트 개수를 의미하고, *error_probability*는 정상 사용자가 키 생성에 실패하여 인증에 실패할 확률을 의미한다. *bit_capacity*를 높게 설정할수록 정상 사용자가 거부당하지 않고, 인증될 확률이 높아져 용이하다. 그러나 타인 역시 정상 사용자로 인증될 확률이 높아져 위조자가 접근하기 쉬워질 수 있다. R-LDA 기법을 사용한 실험에서는 *bit_capacity* = 15일 때, *error_probability* = 0.0132으로 약 1.32%의 사용자가 본인의 키를 정상적으로 생성하지 못했다는 것을 의미한다. *bit_capacity* = 22 이상일 때, *error_probability* = 0으로 모든 사용자가 자신의 키를 정상적으로 생성했으며, 이 때 *bit_capacity*는 전체 길이의 18.3%에 해당한다. DiaPCA의 경우 *bit_capacity* = 20일 때, *error_probability* = 0.0809으로 약 8.09%의 사용자가 본인의 키를 정상적으로 생성하지 못했다. *bit_capacity* = 55 이상일 때 모든 사용자가 자신의 키를 정상적으로 생성했으며 이 때 *bit_capacity*는 전체 길이의 22.91%에 해당한다.

4.2 FAR

공격자가 정상 사용자의 정보를 탈취하고 인증을 시도하려 할 때의 보안성은 FAR(False Acceptance Rate, 오인식률) 표와 그래프를 통해 분석할 수 있다. R-LDA 실험에서는 PW(PIN)을 탈취할 경우, 얼굴을 탈취하거나 아무 정보 없이 공격을 하는 경우에 비해 인증에 성공할 확률이 높다. 이는 얼굴에 비하여 PIN에 대한 의존도가 약간 더 높다는 것을 의미한다. 반면, DiaPCA는 아무 정보가 없거나 얼굴을 탈취한 공격자가 PIN을 탈취한 공격자보다 인증에 성공할 확률이 약간 높다.

FAR 표와 그래프에서 눈여겨봐야 할 부분은 EER(Equal Error Rate, 동일 오류율) 값이다. EER은 FRR과 FAR의 교차점으로 생체 인식 성능의 임계점을 파악하기 용이하다. R-LDA의 EER 지점은 *bit_capacity* = 18에 *error_probability* = 0.0066이다. DiaPCA의 EER 지점은 *bit_capacity* = 44에 *error_probability* = 0.0004이다. 또한, 모든 공격자가 인증에 실패하도록 FAR 값이 0일 때의 *bit_capacity*는 R-LDA와 DiaPCA 각각 8, 27이고, FRR의 *error_probability*는 각 0.1854, 0.0169이다.

V. 결 론

본 연구에서는 바이오해시와 GPT 기법을 응용하여 사용자의 얼굴 이미지로부터 비트스트링 형식의 안전한 키를 생성하는 방법을 제안하였다. 기존 바이오해싱 연구는 토큰에 대한 높은 의존성으로 인하여 토큰 도난 시나리오에서 높은 FAR 값을 보이는 취약점을 가지고 있었으며, 키의 길이도 충분히 길지 않다는 한계점을 가지고 있었다. 본 연구에서는 공통 랜덤 프로젝션, 높은 인식률을 보이는 얼굴 인식 알고리즘과 GPT 기법을 통해 얼굴로부터 충분한 길이의 키를 생성하고 실험을 통해 낮은 EER 수치를 보였다.

키의 보안성을 평가할 때 비트 길이에 의거하여 엔트로피를 측정한다. 본 실험에서 R-LDA는 120 비트의 키를 생성했고 DiaPCA는 240 비트의 키를 생성했다. R-LDA의 경우 비트 길이가 클래스 수에 비례하여 조절하는 데 제약사항이 있으나, DiaPCA는 d값을 조절해서 비트스트링의 길이를 조정할 수 있다. ECC를 이용하게 되면 본래의 길이보다 엔트

로피가 감소하게 되는데, ECC로 보정 가능한 최대 비트의 개수의 두 배를 제한 나머지 길이를 실제 엔트로피로 계산한다. 모든 사용자가 인증에 성공 가능한 0-FRR 지점을 기준으로 삼았을 때 R-LDA는 $120-22\times2=76$ 비트 길이의 엔트로피를, DiaPCA는 $240-55\times2=130$ 비트 길이의 엔트로피를 갖는다.

이러한 방법으로 생성된 비트스트링을 인증 시스템과 공개키 시스템의 개인키(Private key)를 생성하는 등 다양한 활용이 가능할 것으로 예상된다. 그리고 기법에 따라 연산 비용(computational cost)이 작아, 모바일 같은 제한적인 환경에서도 도입이 가능하다. 또한, 얼굴뿐만 아니라 지문, 홍채 등 다른 생체 인식 정보로부터 키를 생성할 수 있도록 응용이 가능하다.

그러나 실제 환경에서는 광원 등에 의한 노이즈의 영향이 크고, 얼굴 각도에 대한 영향이 존재하기 때문에 실제 성능이 실험 환경에서 나온 결과와 차이를 보일 것으로 예상된다. 따라서 다양한 환경에서도 안정적으로 키를 추출할 수 있도록 추후 연구가 진행되어야 한다고 여겨진다.

References

[1] Soo-yeon Lim, "National Innovation Trend: FinTech" *Science & Technology Policy Periodicals*, No. 210, pp. 14-21, Jan. 2016

[2] Seon-Jong Kim, "A Method to support possession and biometric authentication using public certificate in smartphone environment," Review of Korea Institute of Information Security and Cryptography, vol. 25, no. 6, pp. 13-17, Dec. 2015.

[3] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," Guide to Cryptography, CVPR'91, pp. 72-86, 1991

[4] Jian Yang, D. Zhang, A.F. Frangi and Jing-yu Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 26(1). pp. 131-137, 2004

[5] D. Zhang, Z. Zhou and S. Chen, "Diagonal principal component analysis for face recognition," Pattern Recognition, vol. 39(1), pp. 140-142, 2006

[6] Jae-Hyun Oh and Nojun Kwak, "A Resampling Method for Small Sample Size Problems in Face Recognition using LDA," Signal processing-Journal of the institute of electronics engineers of Korea, vol. 46(2), pp. 78-88, 2009

[7] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition,", Pattern Recognition Letters, vol. 26(2), pp. 181-191, 2005

[8] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," Proceddings of seventh ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '01, pp. 245-250, Aug. 2001

[9] Hosik Sohn and YongMan Ro, "Privacy analysis of random projection based biometrics template," Conference of The Institute of Electonic Engineers of Korea, pp. 213-214, Nov. 2009

[10] A. Goh and D. Ngo, "Computation of Cryptographic Keys from Face Biometrics," IFIP International Conference on Communications and Multimedia Security, vol. 2828, pp. 1-13, 2003

[11] Andrew Teoh Beng Jin, David Ngo Chek Ling and Alwyn Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," Pattern Recognition, vol. 37(11), pp. 2245-2255, 2004

[12] David Ngo, Andrew Beng Jin Teoh and Alwyn Goh, "Biometric hash: High-con-

fidence face recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16(6), pp. 771-775, June 2006

[13] Adams Kong, King-Hong Cheung, David Zhang, Mohammed Kamel and Jane You, "An analysis of BioHashing and its variants," Pattern Recognition, vol. 39(7), pp. 1359-1368, July 2006

[14] Andrew Beng Jin Teoh and Chong Tze Yuang, "Cancelable Biometrics Realization With Multispace Random Projections," IEEE Transactions on System, Man, and Cybernetics, Part B: Cybernetics, vol. 37(5), pp. 1096-1106, 2007

[15] Andrew Beng Jin Teoh, Yip Wai Kuan and Sangyoun Lee, "Cancellable biometrics and annotations on BioHash," Pattern Recognition, vol. 41(6), pp. 2034-2044, 2008

[16] Meng-Hui Lim, Min-Yi Jeong and Andrew Beng Jin Teoh, "A Novel Two-Stage Approach in Rectifying BioHash's Problem under Stolen Token Scenario," Journal of information and communication convergence engineering, vol. 8(2), pp. 173-179, 2010

[17] T.A.M. Kevenaar, G.J. Schrijen, M. Van Der Veen and A.H.M. Akkermans, "Face recognition with renewable and privacy preserving binary templates," Fourth IEEE Workshop on Autmatic Identification Advanced Technologies, AutoID'05, pp. 21-26, Oct. 2005

[18] Jeonil Kang, DaeHun Nyang and KyungHee Lee, "Two Factor Face Authentication Scheme with Cancelable Feature," Journal of the Korea Institute of Information Security and Cryptography, vol. 16(1), pp. 13-21, Feb. 2006

[19] Jeonil Kang, DaeHun Nyang, and KyungHee Lee, "Two-factor face authentication using matrix permutation transformation and a user password," Information Sciences, vol. 269, pp. 1-20, June 2014

[20] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy and B.V.K. Vijaya Kumar, "Biometric Encryption using image processing," ICSA Guide to Cryptography, pp. 649-675, 1999

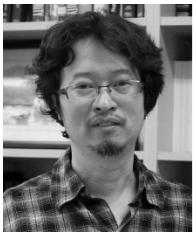
[21] M. Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla "Cancelable biometrics filters for face recognition,", Proc. Int. Conf. Pattern Recognition, vol. 3, pp. 922-925, 2004

〈저자소개〉

김 혜 진 (Hyejin Kim) 학생회원
2016년 2월: 인하대학교 컴퓨터정보공학과 학사
2016년 3월~현재: 인하대학교 컴퓨터정보공학과 석사과정
〈관심분야〉 암호이론, 생체인증, 네트워크 보안

최 진 춘 (JinChun Choi) 학생회원
2011년 2월: 인하대학교 컴퓨터 정보공학과 졸업
2014년 2월: 인하대학교 컴퓨터 정보공학과 석사
2014년 3월~현재: 인하대학교컴퓨터 정보공학과 박사 과정
〈관심분야〉 네트워크 보안, 무선 센서

정 창 훈 (Chang-hun Jung) 학생회원
2014년 9월~현재: 인하대학교 컴퓨터정보공학 석사과정
〈관심분야〉 정보보호, 인증 프로토콜, 금융 보안, IoT



양 대 헌 (DaeHun Nyang) 종신회원
1994년 2월: 한국과학기술원 과학기술 대학 전기 및 전자공학과 졸업
1996년 2월: 연세대학교 컴퓨터과학과 석사
2000년 8월: 연세대학교 컴퓨터과학과 박사
2000년 9월~2003년 2월: 한국전자통신연구원 정보보호연구본부 선임연구원
2003년 2월~현재: 인하대학교 컴퓨터정보공학과 교수
〈관심분야〉 암호이론, 암호프로토콜, 인증프로토콜, 무선 인터넷 보안, 네트워크 보안

이 경 희 (Kyung-Hee Lee) 정회원
1993년 2월: 연세대학교 컴퓨터 과학과 학사
1998년 8월: 연세대학교 컴퓨터 과학과 석사
2004년 2월: 연세대학교 컴퓨터 과학과 박사
1993년 1월~1996년 5월: LG 소프트(주) 연구원
2000년 12월~2005년 2월: 한국전자통신연구원 선임연구원
2005년 3월~현재: 수원대학교 전기공학과 부교수
〈관심분야〉 바이오인식, 정보보호, 컴퓨터비전, 인공지능, 패턴인식