모바일 환경에서 키스트로크 다이나믹스 인증 성능 향상을 위한 사용자 맞춤형 특징 집합 연구

이성훈*, 노종혁**, 김수형**, 진승헌**
*과학기술연합대학원대학교 정보보호공학과
**한국전자통신연구원
e-mail:sunghoon1130@etri.re.kr

A Study of Adaptive Feature Subset for Improving Accuracy of Keystroke Dynamics Authentication on Mobile Environment

Sung-Hoon Lee*, Jong-Hyuk Roh**, Soohyung Kim**, Seung-Hun Jin**
*Information Security Engineering, University of Science and Technology
**Information Security Research Division, ETRI

요 약

키스트로크 다이나믹스 사용자 인증은 행위 기반 인증 방법 중의 하나로써, 사용자가 입력하는 비밀번호 혹은 PIN번호의 패턴을 분석하여 사용자를 인증한다. 비밀번호나 PIN번호가 다른 사용자에게 노출되어도 입력 패턴을 분석하여 사용자를 인증함으로써 지식기반(what you know) 인증의 단점을 보완할 수 있다. 하지만 사용자의 입력 패턴이 항상 일정하지 않고, 사용자별 터치하는 방법이 모두 다르기 때문에 모든 사용자에게서 동일한 특징을 추출하여 그 사용자의 패턴을 생성하고 인증 수단으로 사용하기에는 한계가 있다. 이에 본 논문에서는 사용자별 맞춤형 특징 집합과 전체 특징과의 사용자 인증 성능 변화를 실험을 통해 확인한다. 사용자별 맞춤형 특징이 전체 특징을 사용한 경우보다 평균적으로 EER 6% 이상의 성능 향상이 있었다.

1. 서론

키스트로크 다이나믹스 인증(Keystroke Dynamics Authentication, 이하 KDA)은 행위 기반 사용자 인증 기술 중의 하나이다. KDA는 1980년대 컴퓨터 환경에서 사용자가 키보드에 입력하는 비밀번호의 입력 패턴을 분석하여 사용자를 인증하는 연구 이후에 많은 연구가 진행되어 왔다[1]. 스마트폰의 폭넓은 보급으로 인해 최근에는 스마트폰에도 KDA를 적용하려는 노력이 있다. 스마트폰은 터치스크린과 가속도(accelerometer), 자이로스코프(gyroscope)와 같은 여러 센서를 탑재하여 사용자의 입력 패턴을 분석에 더 많은 정보를 이용할 수 있다.

컴퓨터 환경에서는 사용자가 키보드에서 입력하는 시간을 특징으로 추출하여 사용자의 입력 패턴을 생성하였다. 여러 센서가 탑재된 스마트폰에서는 기존의 시간 특징 이외에도 터치스크린, 가속도, 자이로스코프 센서에서 수집된 센서 데이터로부터 특징을 추출하여 입력 패턴을 생성하려는 연구가 진행되고 있다.

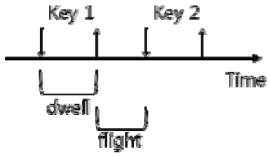
기존 연구에서는 모든 사용자에게 동일한 시간 특징과 센서 특징을 추출하여 사용자의 입력 패턴을 생성하였다. 본 논문에서는 추출된 특징으로부터 사용자 맞춤형 특징 집합을 선택하여 사용자의 키스트로크 패턴을 향상시킬 수 있는지 여부를 실험을 통하여 확인한다.

논문의 구성은 다음과 같다. 2장에서는 스마트폰에서의 KDA 연구에 대해서 살펴보고, 3장에서 실험 방법 및 실험 결과에 대해 설명하고, 4장에서 결론으로 마무리 한다.

2. 관련 연구

스마트폰에서의 키스트로크 다이나믹스 연구는 2013년 이후부터 활발히 진행되고 있다[2]. 키스트로크 다이나믹스에서 특징을 추출하는 방법은 전통적으로 <그림 1>과 같이 시간 기반 추출이 있다. 키를 입력하고 떼어지는 순간의 시간을 저장하여 dwell 혹은 flight 특징을 추출한다.

스마트폰에서의 키스트로크 다이나믹스 연구는 <표 1>과 같다. Zheng[3]은 4-PIN, 8-PIN을 80명의 사용자로부터 각각 수집하여 총 11,062개의 데이터를 수집하였다. 시



<그림 1> 시간 특징

$$\|a\| = \sqrt{a_x^2 + a_y^2 + a_z^2}, \quad (1)$$

<표 1> Keystroke Dynamics 연구 논문

paper	string	extracted features	normalization	feature selection	classifier	performance
Saevanee[6]	10 digits	flight, dwell, p	-	X	neural network	EER 1%
Zheng[3]	4, 8 digits	acc, p, s, flight, dwell	O	X	distance	EER 3.65
Giuffrida[4]	8, 9 alpha	acc, gyr, flight, dwell	-	X	OCSVM, kNN	EER 0.08%
Pin[5]	4,16 digits	dwell, flight, s, p	-	X	statistical-based	EER 5.49%
Paulo[7]	10 alpha	flight	O	X	immune-algorithm	EER 13%
Ho[8]	4 digits	acc, s, flight, dwell	-	X	distance	EER 15%

간 특징으로 dwell과 flight 특징을 사용하였다. 가속도 센서로부터 x,y,z축의 데이터를 유클리디안 놈(euclidean norm)으로 <수식 1>과 같이 계산하여 특징을 추출하였고, 터치 압력과 터치 사이즈도 사용하였다. 시간 특징과 센서 특징을 결합하여 EER 3.5% 결과를 얻었다.

Giuffrida[4]는 8자리(internet)와 9자리(satellite) 알파벳으로 비밀번호를 사용하였다. 가속도와 자이로스코프 센서에서 dwell 구간에서의 RMS, RMSE, Min, Max, AvgDeltas, NumMax, NumMin, TTP, TTC, RCR, SMA 등 11개 카테고리의 특징을 x, y, z 축별로 추출하였고, 시간 특징은 dwell과 flight를 추출하였다. 40명의 사용자로부터 데이터를 수집하여 실험한 결과, 센서 기반 특징으로 EER 0.5%의 성능을 얻었다.

Pin[5] 은 4-PIN과 16-PIN으로 150명으로부터 데이터를 수집하였고, 터치 사이즈와 터치 압력을 특징으로 사용하였다. 시간 특징은 앞선 연구와 마찬가지로 dwell과 flight 특징을 사용하였다. PIN 길이가 4-PIN에서 16-PIN으로 늘어남에 따라 EER이 8.55%에서 5.49%로 약 3.1% 향상되어 PIN 길이에 따른 약간의 성능 차이를 보여주었지만, 사용자가 더 많은 PIN을 입력해야하기 때문에 사용자 편의성은 다소 떨어진다.

3. 실험

PIN번호가 노출되는 경우는 여러 가지가 있다. 본 실험에서는 어깨 넘어 엿보기 공격으로 PIN번호가 노출되었을 경우에, 정상 사용자의 폰으로 정상 사용자와 비정상 사용자의 입력 패턴을 구분할 수 있는지 확인하고자 한다. 정상 사용자가 PIN번호를 입력할 때, 비정상 사용자들이 정상 사용자가 입력하는 모습을 보면서 그 사용자의 PIN 번호 및 입력 패턴을 확인한다.

데이터 수집은 6명의 사용자와 삼성 갤럭시S6 폰으로 진행되었다. 사용자의 키스트로크 다이나믹스 패턴을 생성하기 위한 학습 데이터는 한 번의 세션에서 10개씩 총 20개의 데이터를 수집하였다. PIN번호는 사용자별로 본인의 전화번호 6자리를 사용하였다. 사용자의 입력 패턴을 잘 표현할 수 있도록 본인에게 익숙한 번호를 입력함으로써 사용자의 자연스러운 입력 패턴을 생성하려고 하였다.

총 6명의 사용자들 중에서 각 사용자별로 정상 사용자인 경우를 가정하여 실험용 데이터를 수집하였다. 정상 사용자가 실험용 데이터 5개를 입력하는 동안에 비정상 사용자들은 정상 사용자의 입력 패턴을 확인한 후에, 실험용 데이터 5개를 비정상 사용자별로 수집하였다. 한 명의 정상 사용자의 25개 데이터(학습 데이터 20개, 실험 데이터 5개)와 그 사용자의 입력 패턴을 흉내 내는 5명의 비정상 사용자들의 실험용 데이터 25개를 수집하였다.

스마트폰에서 수집한 센서는 Accelerometer(Acc), Linear acceleration(Lacc), Gyroscope(Gyr), Uncalibrated gyroscope(Ugyr) 등 4개의 센서에서 데이터를 수집하였고, 각 센서로부터 x, y, z 축별로 dwell 시간 구간에서의 최소, 최대, 평균 값을 특징으로 추출하였다. 시간 특징 값은 센서 특징 값보다 범위가 크기 때문에 정규화를 적용하여 동일한 범위 내에서의 값으로 변경하였다. 정상 사용자와 비정상 사용자의 구분 성능을 확인하기 위해서 One-Class Classification(OCC) 분류기 중에 하나인 OCSVM(One-Class SVM)을 사용하였다. 해당 흐름은 <그림 2>와 같다.

추출된 특징을 정규화하기 위해서 Min-Max Scaling 기법을 사용하였다. Min-Max Scaling 기법은 <수식 2>와 같이 각 특징별로 최소, 최대 값을 기준으로 정규화한다. 정규화된 특징은 0에서 1사이의 값을 갖는다.

OCSVM으로 실험한 결과는 <표 2>와 같다. 각 사용

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (2)$$

<그림 2> 특징 추출 및 분류 흐름도

<표 2> OCSVM 분류 결과 (단위: EER %, 강조된 부분: 사용자별 낮은 EER의 특징 집합)

Feature / User	Time	Acc	Lacc	Gyr	Ugyr	All
User1	11.67	6.67	6.67	18.33	21.67	3.33
User2	0	31.67	41.67	33.33	15	30
User3	25	11.67	40	10	16.67	10
User4	21.67	25	33.33	25	11.67	21.67
User5	26.67	26.67	13.33	3.33	3.33	13.33
User6	13.45	0	41.03	13.79	0	0

자별로 시간 특징과 센서 특징을 결합하여 사용자 패턴을 생성하고 분류한 결과, 평균적으로 EER(Error Equal Rate) 13.05%의 결과를 보였다. 특징 타입 별로 사용자의 패턴을 생성한 경우에는, 최고 EER 0%에서 11.67%의 성능을 보였고 평균 EER은 5.27%를 얻었다.

사용자별로 하나 이상의 특징 타입에서 좋은 성능을 보여주고 있다. 정상 사용자의 입력 패턴을 확인하고 입력 패턴을 따라하더라도 정상 사용자만의 패턴을 완벽히 흉내내기는 어렵다는 것을 확인할 수 있다. User2와 User6의 경우에는 각각 Time과 Acc, Ugyr 특징 타입에서 EER 0%의 성능을 보였고, User5는 Gyr와 Ugyr 특집 타입에서 3.33%의 성능을 보였다. User2는 Time을 제외한 다른 특징 타입에서는 비정상 사용자를 정상 사용자로 분류한 FAR가 상대적으로 높아서 EER이 높았다.

User1을 제외한 모든 User들은 전체 특징을 사용한 경우보다 단일 특징 타입으로 학습한 경우에 더 좋거나 동일한 성능을 보였다. 이는 PIN번호나 사용자가 입력하는 방식에 따라, 다른 사용자와 구별되는 그 사용자만의 입력 패턴을 더 잘 표현하는 특징 집합이 있다고 할 수 있다.

4. 토론

머신 러닝 기법을 이용한 분류는 학습 데이터가 많을수록 정상 사용자의 더 많은 패턴을 분석할 수 있기 때문에 성능이 더 좋아질 수 있다. 이에 학습 데이터의 수를 줄였을 시에도 20개의 학습 데이터를 사용했을 경우와 차이가 있는지 실험을 하였다. 학습 데이터를 5개로 고정하고, <그림 3>과 같이 학습 데이터에 사용될 데이터를 변화하며 학습 데이터에 사용된 데이터의 변화에도 전체 데이터를 사용했을 경우와의 차이를 실험하였다.

정상 사용자의 실험 데이터 20개 중에서 5개의 데이터를 학습에 사용하고, 정상 사용자의 실험 데이터 5개와 비정상 사용자의 실험 데이터를 이용하여 성능을 확인하였다. <표 3>의 각 셀의 명도가 검은색에 가까울수록 분별력이 높은 특징 집합이고, 하얀색에 가까울수록 분별력이 낮은 특징 집합이다.

<표 3> 학습 데이터 변화에 따른 특징 집합

(검은색: 분별력 높은 특징 집합, 하얀색: 분별력 낮은 특징 집합)

Feature / User	Time	Acc	Lacc	Gyr	Ugyr
User1					
User2					
User3					
User4					
User5					
User6					

<그림 3 > 학습에 사용된 학습 데이터 범위

User2와 User5, User6은 전체 학습데이터를 사용한 경우와 차이가 없이 완벽히 동일한 분별력 높은 특징 집합이 있음을 보여주고 있다. User1은 전체 학습 데이터를 사용한 경우와 완벽히 일치하지는 않지만, Acc, Lacc 특집 집합이 제일 분별력 높은 특징임을 알 수 있다. User3의 경우에는 Gyr 특집 집합이 상대적으로 높게 나오긴 했지만, Time이나 Acc, Ugyr 특징 집합에서도 분별력이 어느 정도 있음을 확인할 수 있다. 이는 표 2에서 User3의 특징 집합별 EER이 Lacc 특집 집합을 제외하고는 비슷한 수준의 EER을 보여주었던 것과 마찬가지로 학습 데이터 변화에 따라서도 비슷한 특징 집합 분포를 보여주고 있다. User4는 전체 학습 데이터에서 특징 집합별로 비슷한 EER을 보여주었기 때문에 <표 3>에서도 두드러진 특징 집합이 나오지 않았다.

4. 결론

키스트로크 다이나믹스는 PIN이나 비밀번호 기반의 인증에서 해당 PIN 혹은 비밀번호가 노출되어도 입력 패

턴을 비교하여 정상 사용자인지 비정상 사용자인지 구분할 수 있다. 스마트폰에서 많이 사용되는 PIN기반 인증에도 키스트로크 다이나믹스를 적용하여 기존의 한계점을 보완할 수 있다.

과거 키스트로크 다이나믹스에서 사용자의 입력 패턴을 구분하기 위해서 시간과 센서 기반 특징을 모두 사용하였다. 하지만 사용자별로 PIN번호가 다르고 입력 방식 또한 다르기 때문에 실험을 통하여 사용자별 맞춤형 특징과 전체 특징과의 성능 차이를 확인하고, 사용자의 입력 패턴을 더 잘 표현하는 특징 집합을 사용한 경우에 더 좋은 성능을 얻을 수 있었다.

향후 연구에서는 더 많은 사용자로부터 실험 데이터를 수집하고, 사용자의 입력 패턴을 더 잘 표현하는 특징 집합 혹은 개개의 특징들을 선별할 수 있는 알고리즘에 대해 연구할 예정이다.

참고문헌

[1] GAINES, R. Stockton, et al. “Authentication by keystroke timing: Some preliminary results” RAND CORP SANTA MONICA CA, 1980.

[2] Teh PS, Teh PS, Zhang N, Teoh ABJ, Zhang N, Chen K, et al “A survey on touch dynamics authentication in mobile devices” Computers & Security. 2016;59:210 - 235.

[3] Zheng N, Bai K, Huang H, Wang H. “You Are How You Touch: User Verification on Smartphones via Tapping Behaviors” 2014 IEEE 22nd International Conference on Network Protocols. 2014;:221 - 232.

[4] Giuffrida C, Majdanik K, Conti M, Bos H. “I Sensed It Was You: Authenticating Mobile Users with Sensor-enhanced Keystroke Dynamics” In Detection of Intrusions and Malware, and Vulnerability Assessment. 2014;:92 - 111.

[5] Teh PS, Zhang N, Teoh ABJ, Chen K. “Recognizing Your Touch: Towards Strengthening Mobile Device Authentication via Touch Dynamics Integration” Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia - MoMM 2015. 2015;.

[6] Saevanee H, Clarke N, Furnell S, Biscione V. “Continuous user authentication using multi-modal biometrics” Computers & Security. 2015;53:234.

[7] Pisani PH, Lorena AC. “Emphasizing typing signature in keystroke dynamics using immune algorithms” Applied Soft Computing. 2015;34:178 - 193.

[8] HO, Grant. “Tapdynamics: strengthening user authentication on mobile phones with keystroke dynamics” Technical report, Stanford University, 2014.