코사인 유사도 측정을 통한 행위 기반 인증 연구

길선웅
인천대학교 정보통신공학과
swgil009@inu.ac.kr

Behavior-based Authentication Study By Measuring Cosine Similarity

Seon-Woong Gil
Dept. of Information and Communication Engineering,
Incheon National University

요 약

사용자 행위 기반 인증 기술은 다른 인증 기술들에 비해서 인증의 인식률을 높이는데 많은 데이터의 장기간 추출이 필요하다. 본 논문은 터치 센서와 자이로스코프를 이용하여 그동안의 행위 기반 인증 연구에서 사용 되었던 행위 특징 데이터들 중에서 핵심적인 최소한의 데이터들만을 사용하였다. 측정한 데이터들의 검증에는 그간 사용자 행위 기반 인증 연구에서 이용되지 않고 문서 검색의 유사도 측정에 사용되었던 코사인 유사도를 사용하였다. 이를 통해 최소한의 특징 데이터와 기준이 되는 데이터의 코사인 유사도 비교 검증만을 통해서도 인증 범위에 적용되는 임계값을 조절하는 방식을 통해서 최초 EER 37.637%에서 최종 EER 1.897%의 높은 검증 성능을 증명하는데 성공하였다.

1. 서론

행위 기반 인증 기술은 다른 인증 기술들에 비해서 인증에 필요한 사용자의 행동을 추출하는데 많은 데이터가 필요하고 인식률마저 부족하다는 문제점이 존재한다. 본 논문에서는 행위 기반 인증의 정확도 향상을 위해서 인증 기반으로 설정하는 사용자 행위 패턴 데이터를 다량으로 수집 하는 기존의 연구들을 분석하여 인증에 필요한 핵심적인 데이터를 선정해 최소한의 데이터 수집으로 기존에 행위 기반 인증에 사용되지 않던 코사인 유사도 측정 방법을 통해서 최대한의 사용자 인증 정확도를 얻을 수 있는 방법에 대해 연구하였다.

본 논문은 2장에서는 기존의 연구들을 분석하여 인증에 사용할 핵심적인 사용자 특징 데이터를 선정한다. 3장에서는 코사인 유사도를 이용한 사용자 행위 기반 인증 기술을 설계한다. 4장에서는 설계한 사용자 행위 기반 인증 기술을 실험하고 이를 분석한다. 마지막으로 5장에서 결론을 맺는다.

2. 기존 행위 기반 인증 기술들의 특징 데이터

행위 기반 인증 기술에 대한 다양한 연구가 진행되고 있는 가운데 최근 2013년도부터의 가시적인 결과를 보여준 연구결과를 정리해보면 아래 <표1>과 같다.

<표 1> 행위 기반 인증 기술 연구 동향

저자	특징 데이터	알고리즘	정확도(%)
Meng et al. (2013)	싱글 터치, 터치 이동, 멀티 터치 상황에서 터치 타입, 좌표, 시간, 방향, 행동 횟수	Neural Network	EER : 2.92
Xu et al. (2014)	키스트로크(크기, 시간, 압력), 슬라이드(위치, 크기, 압력, 속력, 길이), 서명(크기, 압력, 위치, 방향, 속력, 여백), 핀치(위치, 크기, 압력, 속력, 길이, 방향)	SVM	EER : 10
Meng et al. (2014)	터치 이동 수, 싱글 터치 수, 멀티 터치 수, 터치 이동 시간, 싱글 터치 시간, 멀티 터치 시간, 터치 이동속도, 터치 압력	Neural Network	EER : 2.46
Shen et al. (2015)	슬라이드(상하좌우 각 방향별 위치, 길이, 각도, 시간, 속도, 가속도, 압력)	One-class SVM	FAR : 0.03 FRR : 0.05
Sitova et al. (2016)	탭(시간, 크기, 속력), 키스트로크(누른 시간, 입력 간 시간), 쥐는 형태 (가속계, 중력계, 자력계)	one-class SVM	EER : 7.16(이동) EER : 10.05(정지)

각 연구의 특징 데이터와 사용한 알고리즘, 정확도를 보여준다. 지금까지 살펴본 행위 기반 인증 기술 연구들의 대부분이 터치와 각종 센서를 이용하여 사용자의 특징 데이터를 추출 하고 있다.[1][6]

본 논문에서는 기존 연구들에서 사용한 데이터들 중 핵심적인 특징 데이터로 터치 센서를 이용하여 터치 좌표와 시간, 모션 센서인 자이로스코프를 이용하여 스마트폰의 회전벡터를 사용한다.

3. 행위 기반 인증 기법 설계

본 논문에서는 안드로이드 스마트폰의 터치, 모션 센서로 추출한 사용자의 스크린 터치시의 좌표, 시간, 회전벡터를 검증에 사용하기 위해 코사인 유사도 값으로 가공한다.

코사인 유사도란 다차원의 양수 공간에서의 유사도 측정에 이용되며 벡터의 크기 값은 결과에 영향을 미치지 않고 차원의 개수가 많은 다차원의 벡터일수록 유사도를 뚜렷이 구분할 수 있는 장점이 있다. 그런 특징 때문에 벡터의 원소나 비교 벡터의 원소와의 크기 값이 차이가 많이 나는 경우에도 다른 가공 없이도 유사도를 비교할 수 있다. 또한 두 벡터의 내적과 외적을 통해서 아래 (그림 1)과 같이 간단한 공식만으로도 연산이 가능하다는 장점이 있다.[7]

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

(그림 1) 코사인 유사도 공식

본 논문에서 제안하는 인증 기법은 사용자가 다섯 번의 터치를 시도 할 때 본인만의 리듬으로 화면 좌표를 터치하고 스마트폰을 기울이며 측정을 마친다. 이렇게 한 차례의 측정을 마친 사용자에게 다섯 차례의 측정을 요구한다. 한 차례의 측정을 마친 사용자의 데이터는 아래 <표 2>과 같이 터치 X좌표, Y좌표, 측정 때의 스마트폰의 회전벡터 값 X, Y, Z, 터치시의 시간으로 한 차례의 측정 때마다 30개의 데이터가 측정된다.

<표 2> 한 차례의 측정 때 수집되는 데이터

터치 X좌표	터치 Y좌표	회전벡터 X	회전벡터 Y	회전벡터 Z	시간
touch X.1	touch Y.1	round X.1	round Y.1	round Z.1	T1
touch X.2	touch Y.2	round X.2	round Y.2	round Z.2	T2
touch X.3	touch Y.3	round X.3	round Y.3	round Z.3	T3
touch X.4	touch Y.4	round X.4	round Y.4	round Z.4	T4
touch X.5	touch Y.5	round X.5	round Y.5	round Z.5	T5

이 30개의 데이터에서 각 터치 사이를 구간으로 두면 총 4개의 구간이 생기게 된다. 각 구간 사이의 값은 터치 X좌표, Y좌표, 측정 때의 스마트폰의 회전벡터 X, Y, Z, 터치시의 시간 값으로 사용자가 측정을 할 때의 특징 데이터 값이다. 아래 <표 3>과 같이 다음 터치 횟수의 데이터 값에 현재 터치 횟수의 데이터 값을 차감하여 다음 터치로 이동 할 때의 각 데이터 항목의 변화량을 이용한다.

<표 3> 네 가지 구간별 사용자 행위 특징 데이터

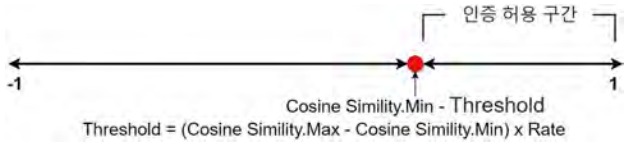
터치 X좌표	터치 Y좌표	회전벡터 X	회전벡터 Y	회전벡터 Z	시간
touch X.2 − touch X.1	touch Y.2 − touch Y.1	round X.2 − round X.1	round Y.2 − round Y.1	round Z.2 − round Z.1	T2 − T1
touch X.3 − touch X.2	touch Y.3 − touch Y.2	round X.3 − round X.2	round Y.3 − round Y.2	round Z.3 − round Z.2	T3 − T2
touch X.4 − touch X.3	touch Y.4 − touch Y.3	round X.4 − round X.3	round Y.4 − round Y.3	round Z.4 − round Z.3	T4 − T3
touch X.5 − touch X.4	touch Y.5 − touch Y.4	round X.5 − round X.4	round Y.5 − round Y.4	round Z.5 − round Z.4	T5 − T4

이렇게 구한 각 구간별 특징 데이터 값의 1회부터 5회까지 측정에서의 평균값을 구하고 이 평균값과 각 측정시도의 구간 값을 (그림 1)에서 보았던 코사인 유사도 공식에 사용되는 벡터로 사용하면 사용자의 검증을 위한 코사인 유사도 값을 구할 수 있다. 총 5차례의 측정이 있었으므로 각 구간 당 5개의 코사인 유사도 값 총 20개의 코사인 유사도 값이 생성된다. 아래 (그림 2)는 실제 실험자들의 각 구간별 최소 코사인 유사도 값을 보여준다.

	name	valid1min	valid2min	valid3min	valid4min
1	a	0.9979800015039642	0.994509343848845	0.992790538067006	0.989612523883865
2	b	0.988631548602581	0.986413805561361	0.99578486042711 8	0.993264592615508
3	c	0.985034300848266	0.981012263583957	0.990947984281981	0.979339312302176
4	d	0.996949306283853	0.994489861082515	0.998624731776395	0.996652134175104

(그림 2) 사용자의 각 구간별 최소 코사인 유사도 값

다섯 차례의 사용자 측정을 통해 구한 최소 코사인 유사도 값을 그대로 범위로 사용하면 FRR(False Rejection Rate)의 비율이 너무 높아져 버릴 것이기 때문에 최소 코사인 유사도 값에 특정한 임계값을 차감하는 형태로 아래 (그림 3)과 같이 인증 허용 범위를 늘리거나 줄여야 한다.



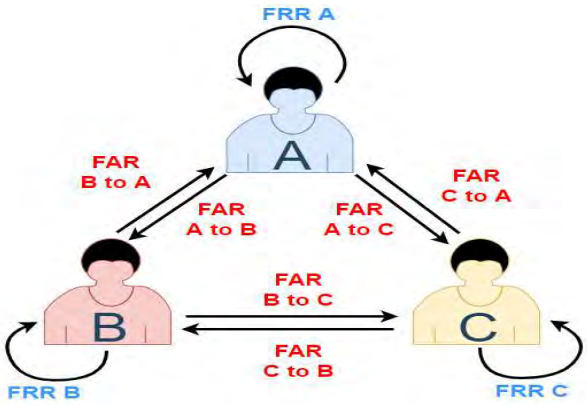
(그림 3) 인증 허용 범위

이때 차감할 임계값(Threshold)에 사용하는 값을 각 구간별로 구한 최대 코사인 유사도 값(Cosine Simility.Max)에서 최소 코사인 유사도 값(Cosine Simility.Min)을 빼준 값에 특정 임계 상수(Rate)를 곱한 값을 사용한다.

이러한 방법으로 최소 코사인 허용 범위를 설정하는 이유는 각 구간별로 최소, 최대 코사인 유사도 값이 다르기 때문에 FRR 비율이 높게 나올 수 있는 특정 구간에서는 최소, 최대 코사인 유사도 값의 차이가 크다. 이 때문에 사용자 마다 각 구간별로 유연한 허용 범위 설정이 측정과 동시에 가능하기 때문이다.

4. 실험 및 분석

본 실험에서 사용할 데이터 셋을 모으기 위해서 실험자 세 명이 실험에 참여하였다.



(그림 4) 실험자 A, B, C의 FAR, FRR 데이터 측정

실험자들은 위 (그림 4)처럼 스마트폰 화면이 서로에게 보이지 않을 수 있도록 삼각형 모양으로 마주보며 앉아 실험자 A부터 FRR 데이터 셋을 수집하기 위해 곧 바로 50회 가량의 본인에 대한 행위 기반 인증을 실시하였고, 이때 나머지 실험자 B와 C는 그 행위를 지켜보며 실험자 A의 인증 행동을 관찰하도록 하였다.

실험자 A의 FRR 측정이 끝난 후에는 실험자 A에 대한 FAR(False Acceptance Rate) 데이터 셋 수집을 위하여 실험자 A의 행동을 관찰하던 실험자 B와 C가 곧바로 실험자 A에 대한 침입자의 행위 기반 인증을 각각 25회 가량 진행하여 이전에 실험자 A의 인증 행동을 계속해서 따라 해보는 방식으로 실험하였다. 이렇게 실험자 A에 대한 FRR, FAR 데이터 셋을 수집한 후에는 실험자 B와 C 순서대로 같은 방식의 실험을 진행하였다.

이와 같이 실험한 결과 실험자 A와 B, C는 FRR에 대한 데이터 수집을 각각 53회, 57회, 72회 총 182회 진행하였고 이를 바탕으로 데이터 수집 1회당 1구간부터 4구간까지의 인증에 시도한 코사인 유사도 값 4개의 데이터가 수집되어 총 728개의 데이터 셋을 모을 수 있었다. 실험자 A와 B, C의 FAR에 대한 데이터 수집은 각 53회, 71회, 67회의 총 191회의 시도를 통해 총 764개의 데이터 셋이 모였다.

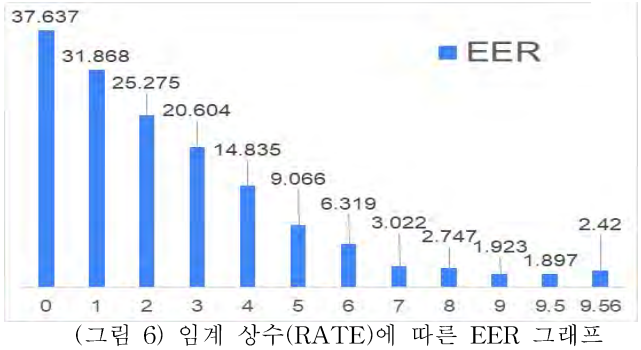
실험자 A와 B, C가 진행한 FRR, FAR 실험에서 측정한 인증 시도시의 각 네 가지 구간별 코사인 유사도 값 데이터 셋이 아래 (그림 5)와 같이 728개, 764개 총 1,492개의 데이터 셋이 모였다.

	_id	name	mnum	rate	tester
	필터	필터	필터	필터	필터
1487	1487	c	3	9.0	-0.00816233046606713
1488	1488	c	4	9.0	0.166972501599505
1489	1489	c	1	9.0	0.979234207482881
1490	1490	c	2	9.0	0.999786227953024
1491	1491	c	3	9.0	0.7069701770231 08
1492	1492	c	4	9.0	0.0736266626927253

(그림 5) 1,492개의 FAR, FRR 데이터 셋

이렇게 모인 데이터 셋을 가지고 사용자 행위 기반 인증을 허용하는 범위 조절에 사용하는 임계값(Threshold)의 임계 상수(Rate)값을 변화시키며 실험자 A와 B, C의 평균 EER을 구하여 임계 상수(Rate)에 따른 평균 EER을 서로 비교 방식하는 방법으로 실험을 분석하였다.

아래 (그림 6)은 임계 상수(Rate)의 변화에 따른 실험자 A와 B, C의 EER 값을 평균한 본 논문에서 제안하는 행위 인증 기법의 EER 수치이다. 본 실험 결과 최대, 최소 코사인 유사도 값의 차에 곱해지는 임계 상수(Rate) 값이 증가함에 따라 EER 수치가 점점 줄어드는 결과가 나타났다.



(그림 6) 임계 상수(RATE)에 따른 EER 그래프

최초 임계 상수(Rate)를 0으로 설정하여 임계값 없이 측정된 사용자 최소 코사인 유사도 값만으로 인증을 진행 할 시 평균 FRR은 75.275%, 평균 FAR은 0%로 평균 EER은 37.637%를 나타냈고, 최종적으로 임계 상수(Rate)가 9.5인 지점에서 평균 FRR은 2.747%, 평균 FAR은 1.047% 그리고 최종 EER은 1.897%로 가장 낮은 값을 나타내었다. 그 후에는 EER 수치가 증가하는 추세를 보였다.

본 실험에서 임계 상수(Rate) 값이 9.5 일 때 가장 성능이 좋은 인증이 되었던 이유는 실험자에 따라서 또 그 구간에 따라서 FAR, FRR 수치가 각기 다르며 코사인 유사도의 최대, 최소 값이 모두 다르기 때문에 이번 실험에 참여한 실험자 A와 B, C의 데이터 셋에서 각 구간 별 인증 범위를 가장 정확도 있게 설정할 수 있는 임계 상수(Rate)가 9.5였고 이 값은 유동적인 값이라고 분석한다. 즉 본 행위 기반 인증 기법의 EER수치를 변동시키는 임계 상수(Rate)값은 실험자의 수가 더 많아지거나 실험자가 인증 값을 생성할 때 5회 이상의 더 많은 측정 횟수를 시도하게 되거나 실험자의 제안 인증 기법의 숙련도에 따라 수시로 변동 할 수 있는 수치라고 분석한다.

5. 결론

본 논문을 통해서 안드로이드 스마트폰에 내재되어있는 터치 센서와 자이로스코프를 이용하여 최소한의 행위 특징 데이터들을 코사인 유사도 측정 비교 방식을 이용하여 행위 기반 인증을 설계 및 실험하였다. 사용자에게 다섯 차례의 측정을 요구하여 터치시의 X, Y좌표와 스마트폰 기울기에 해당하는 회전벡터 X, Y, Z와 터치 시간 값을 이용하여 총 6가지의 행위 특징 데이터를 수집하였고 다음 터치 측정으로 넘어가는 동안의 데이터들의 변화 값을 통해서 사용자를 특정 할 수 있는 데이터로 가공하였다. 측정 데이터의 유사도를 검증하기 위해서는 코사인 유사도를 이용한 비교 방식을 선택하였고 많은 수의 수집 데이터나 오랜 시간의 실험 없이도 수집한 데이터와 기준이 되는 데이터의 코사인 유사도 비교만을 통해서도 적은 수의 실험자와 특징 데이터의 환경에서도 인증 범위에 적용되는 임계값을 조절하는 방식을 통해서 최초 EER 37.637%에서 최종 EER 1.897%의 높은 성능을 증명하는데 성공하였다. 향후에는 사용자의 측정 횟수의 제한을 없애고 사용자의 수가 늘어날 때마다 EER 수치가 가장 낮은 임계값을 시스템 스스로 설정하여 변경하는 방식으로 보완하여 본 논문에서 제안하는 사용자 행위 기반 인증 기법의 신뢰도를 검증해 나갈 것이다.

참고문헌

[1] T. Feng, X. Zhao, B. Carbunar, and W. Shi, "Continuous mobile authentication using virtual key typing biometrics." 2th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2013.

[2] H. Xu, Y. Zhou, and M.R. Lyu, "Towards continuous and passive authentication via touch biometrics:An experimentalstudy on smartphones.," Symposium On Usable Privacy and Security (SOUPS 2014). 2014.

[3] Y. Meng, and D.S. Wong, "Design of touch dynamics based user authentication with an adaptive mechanism on mobile phones," Proceedings of the 29th Annual ACM Symposium on Applied Computing. ACM, 2014.

[4] C. Shen, Y. Zhang, Z. Cai, T. Yu, and X. Guan, "Touch-interaction behavior for continuous user authentication on smartphones," 2015 International Conference on Biometrics (ICB), IEEE, 2015.

[5] Z. Sitová, J. Šeďenka, Q. Yang, G. Peng, and G. Zhou, "HMOG: NewBehavioral Biometric Features for Continuous Authentication of Smartphone Users." IEEE Transactions on Information Forensics and Security, Vol. 11, No. 5, pp. 877-892, 2016.

[6] 김민우(2016), "안드로이드에서 앱 사용과 터치 정보를 이용한 행위 기반 사용자 인증 기술 연구", 석사학위 논문, pp.8 ~ 42

[7] https://www.ibric.org/myboard/cosinesimilarity