

Key-Stroke 기반 Two-Factor 인증 기술★

안 준 연*, 고 광 필*, 이 태 진**

요 약

ID/Password 기반 인증기술은 간편하면서 일정한 보안수준을 제공하기에 대부분의 시스템에서 사용되고 있으나, 이미 무수히 많은 개인정보 노출사고가 있었으며, 무엇보다 한번 노출된 패스워드를 회수하기 어렵다. 이에, 다양한 two-factor 인증기법들이 도입되었으나, 이들은 많은 비용이 필요하며, 무엇보다 사용자의 불편함을 초래하게 된다. 본 논문에서는 기존과 같이 단 한번의 ID/Password 인증과정에서 사용자마다 고유한 Key-Stroke를 동시에 인증하여 비용대비 효과적이면서, 사용자의 불편함을 초래하지 않고, ID/Password 노출 시에도 Key Stroke Dynamics 패턴이 달라 실패하여 높은 보안성을 보장할 수 있는 기술을 제안한다. 본 논문의 제안 모델은 시스템으로 구축하여 효과성을 확인하였다.

Two-factor Authentication technology based on Key-Stroke

Jun-Yeon An*, Gwang-Feel Ko*, Tae-jin Lee**

ABSTRACT

Password based authentication technology is yet certain and id to provide a level of security being used in most systems, but already a myriad of personal information exposure to the accident. Above all, and once exposed, it is difficult to recover the password. Thus, the various authentication techniques - factor two was introduced, but they are expensive and discomfort to users, to lead. In this paper, the existing unique to users in such a single accreditation process / password id key - stroke, user authentication and cost effectively and at the same time. And not cause discomfort, suggested technologies that can also ensure high security exposure, password id. This paper's proposals and determine the effectiveness of the system to build model.

Keywords: Key-Stroke Dynamics, Two-Factor Authentication, User Authentication

접수일(2020년 02월 19일), 수정일(2020년 06월 02일)
게재확정일(2020년 07월 06일)

* 호서대학교
** 호서대학교(교신저자)

★ 이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2018-0-0276, 딥러닝 기반 악성코드 패턴물셋 생성 자동화 원천기술 개발).

1. 서 론

2019년 Cloud Storage 서비스인 MEGA에 약 11억 개의 고유 이메일 주소와 비밀번호 조합, 7억 7,300만 개의 이메일 주소, 2000만 개의 비밀번호가 포함된 총 27억 개의 데이터가 ‘컬렉션 #1(Collection #1)’이란 이름으로 유출되었다고 발표했다. 또한 현재 ID 및 Password 기반 인증 방식이 널리 쓰이고 있지만 위의 결과처럼 ID 및 Password 등 개인정보 유출 사고는 빈번하게 발생하고 있고 지속적으로 증가하는 추세이다.

이를 보완하기 위해 홍채 인식, 지문 인식 등 다양한 인증 방식이 도입/운영되고 있다. 전문가들은 사용자 인증에 사용자가 알고 있는 것, 사용자가 가지고 있는 것, 사용자의 존재를 나타내는 것을 확인하는 3가지 방법 중 적어도 2가지는 사용되어야 한다고 말하지만, 대부분은 그렇지 못하다.[7] 문제점은 높은 도입 비용, 기존 Legacy 시스템과의 연계 문제, 이용자들의 불편함 등으로 현실적으로 도입이 쉽지 않다.

본 논문에서는 사용자 인증 시, 사용자의 ID 및 Password가 노출되더라도 사용자 계정이 보호받을 수 있는 Key-Stroke Dynamics 기반 인증 기술을 제안한다. Keystroke Dynamics를 활용할 때 첫 번째로 컴퓨터의 모든 사용자가 키보드로 입력하기 때문에 실용적이라는 점. 추가 부품의 필요가 없어 저렴하다는 점. 인증 과정이 지난 후에도 여전히 타이핑 리듬을 사용할 수 있다는 장점이 있다[8][13].

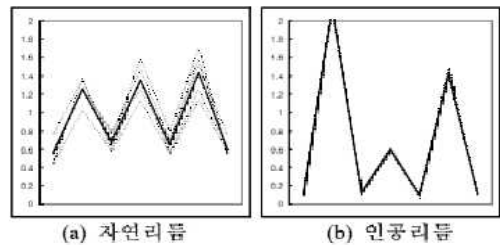
본 논문의 구성은 다음과 같다. 2장에서 Key-Stroke Dynamics 데이터를 사용한 관련 연구를 설명하고, 3장에서 본 논문이 제안하는 Two-Factor 인증 예시 모델을 설명한다. 4장에서는 제안 모델의 시험 결과를 보여주고, 5장에서는 4장의 결과로 본 논문이 제안하는 기술의 의미와 결론, 추후의 연구 방향을 제시한다.

2. 관련연구

2.1 Keystroke Dynamics 분석을 이용한 모바일 인증

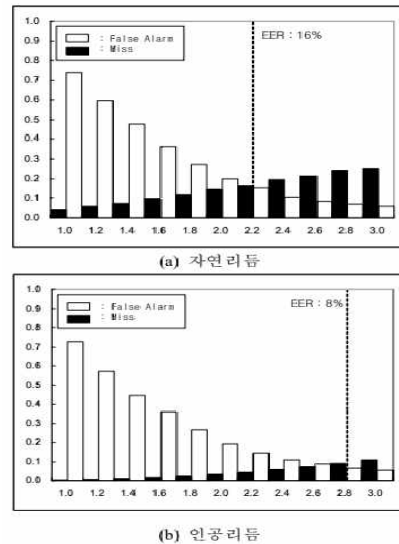
최근 핸드폰 같은 휴대용 단말기의 용도는 통화 이외에도 예금, 증권, 결제, 신원확인 등과 같은 다양한 어플리케이션으로 발전하고 있다. 이에 따라 핸드폰에도 높은 보안이 필요해지고 있다.

숫자만으로 이루어진 네 자리의 비밀번호를 자연 리듬과 인공리듬, 두 종류의 Keystroke Dynamics로 나누어 30회 학습 후 인증을 진행하였다.



(그림1) 자연리듬과 인공리듬

자연리듬의 경우 사용자 패턴 인증 결과의 ERR이 16%였으며, 인공리듬을 이용할 때는 8%로 ERR이 절반 가량 하락한 것을 알 수 있다. 또한 인공 리듬의 경우 포즈 삽입의 횟수나 길이의 증가가 타이핑 패턴의 품질 향상으로 이어져 높은 성능의 사용자 인증을 할 수 있음을 보였다.[1]

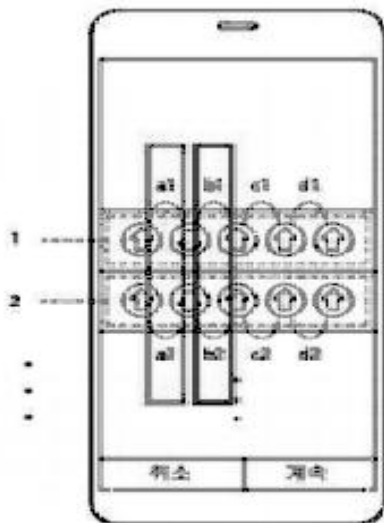


[그림2] 사용자 인증 결과

2.2 Keystroke Dynamics가 적용된 볼륨버튼 기반의 패스워드 인증법

기존에 스마트폰에서 사용되는 PIN, 패턴인식 등의 기법은 어깨너머 공격 및 레코딩 공격에 취약하고 비교적 낮은 보안성을 가지므로 현재 생체 인식을 이용한 사용자 인증 기법이 보편적으로 사용된다. 그러나 생체 인식은 앞서 서론에서 서술한 단점들이 있어 이러한 문제를 해결하기 위한 연구가 진행중이었다.

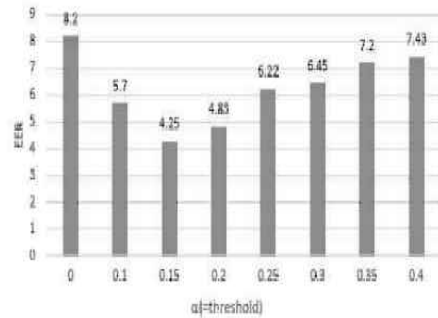
기존의 모바일기기에 장착된 볼륨버튼을 음량 조절, 화면 캡처 등의 원기능이 아닌 패스워드로 사용한 연구는 이런 단점을 극복하였고, 복잡한 비밀번호가 아닌 단순한 볼륨 +, - 두 버튼을 사용하여 모바일기기의 사용이 서툰 사람들의 접근성을 높였다.



(그림3 볼륨 버튼 누름 시간 간격 계산)

인증시 미리 학습한 5회의 사용자 볼륨 패턴과 비교해 발생한 이벤트 버튼 시간을 측정하여 평균값으로 DB의 기준 데이터와 비교해 사용자를 판단하였다.

15명의 피실험자와 5명의 공격자, 총 20명이 5회의 연습을 거쳐 10회 인증 시도를 하였고 오타를 제외한 985개의 샘플을 수집하였다. 4.25%라는 낮은 ERR을 보였으며, 기존의 Keystroke Dynamics 연구들과 비교하여 높은 정확도를 보여 주었다.[2]



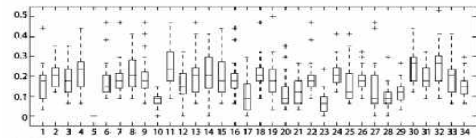
(그림4 임계값에 따른 ERR 결과값)

2.3 자유로운 문자열의 Keystroke Dynamics를 활용한 사용자 인증 연구

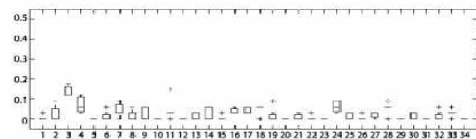
사용자 인증 보안에서는 비밀번호의 복잡성 역시 중요한 항목 중 하나이다. 복잡성이 높은 자유로운 문자열에 대한 Keystroke Dynamics 연구도 진행되었다.

키 입력시간(latency)를 기본적인 측정 단위로 삼았으며, 입력키의 전후 관계를 고려하지 않은 K-S 스코어와 고려한 R-A 스코어로 키 입력시간을 나누어 연구하였다.

3500~4000자에 해당하는 국문 및 영문 두 가지 언어를 각각 34명, 35명이 인증하였다. 각각 짧은 2~3문장과 1~2문단에 해당하는 문단 입력으로 나누어 실험하였고, 결과에 대해 Keystroke Dynamics는 입력 언어에 크게 영향을 받지 않는 개인의 고유한 행동 속성임을 다시 보여주었으며, 비밀번호의 복잡성을 향상시킨 자유로운 문자열에 대해서도 Keystroke Dynamics를 이용해 사용자 인증이 가능함을 보였다[3].

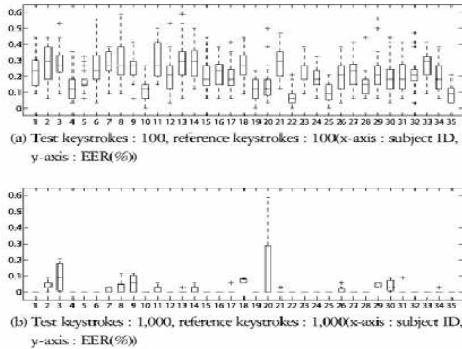


(a) Test keystrokes : 100, reference keystrokes : 1000(x-axis : subject ID, y-axis : EER(%))



(b) Test keystrokes : 1,000, reference keystrokes : 1,000(x-axis : subject ID, y-axis : EER(%))

(그림5) 국문 입력시 테스트 Keystroke Dynamics 100회, 1000회 결과



(그림6) 영문 입력시 테스트 Keystroke Dynamcis 100회, 1000회 결과

2.4 기존 Two-Factor의 문제점

기존의 Two-Factor 기술이 가지고 있는 문제점은 많은 비용과 불편함이다. 별도의 추가적인 기기가 필요하기 때문에 이러한 과정은 사용자에게 불편함을 느끼게 한다.

또한 오탐지율이 개선되지 않는다. 생체 인식의 경우 모든 인증 때와 생체 정보 등록시의 생체정보가 100% 동일하지 않다. 인증된 사용자임에도 불구하고 오탐지로 인하여 여러번 시도해야 하는 경우가 발생한다.

따라서 기존의 Two-Factor 기술이 가지고 있는 많은 비용, 사용자의 불편함, 동결된 오탐지율을 개선 및 보완할 수 있는 방안이 필요하다.

2.5 개발 기술의 요구 사항

본 개발 기술의 요구사항으로는 간편한 Legacy 연동, 오탐율의 개선, 사용자의 편리성 개선이 있다.

Keystroke Dynamics를 활용하기 때문에 기존의 사용자 인증을 요구하는 시스템들에 간편하게 연동이 가능해야 한다. 이 경우 타 생체 인증 방법에 비해 비용 절감이 가능하다.

허가된 사용자인지 판단할 때 오탐지를 할 경우 사용자는 재입력을 통해 불편함을 겪지만, 인증의 성공시마다 해당 데이터가 사용자의 Keystroke Dynamics 표본에 추가되어 더욱 정확한 학습이 가

능해지도록 하여, 인증을 시도할 때마다 오탐율을 개선할 수 있어야 한다.

사용자가 ID/Password 입력 행위 외에 추가적인 별도의 행위가 필요하지 않은 점과 위의 사항들을 포함하여 사용자의 편리성 역시 개선할 수 있어야 한다.

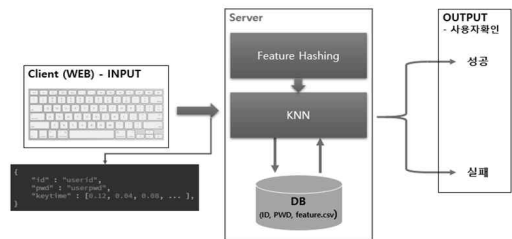
핵심 요구사항으로는 ID/Password 공격시에 공격자와의 Keystroke Dynamics 데이터 패턴이 다른 것을 이용해 ID/Password 유출 시에도 높은 보안을 유지할 수 있어야 한다.

3. 제안 모델

3.1 시스템 모델

본 논문에서 제안하는 모델의 Two-Factor 인증 기법은 1차적으로는 타 시스템과 동일하게 ID/Password 일치 유무를 확인하고, 2차적으로는 ID/Password 입력시의 Key-Stroke Dynamics 데이터 패턴을 통해 실제 사용자가 맞는지 판단하여 인증한다.

제안하는 모델은 등록 단계와 인증 단계로 구성된다. 등록 단계는 사용자의 Key-Stroke Dynamics 데이터 패턴을 추출 획득하는 단계이며, 인증 단계는 인증 시도 시 입력된 데이터 패턴을 샘플과 비교하는 방식이다.



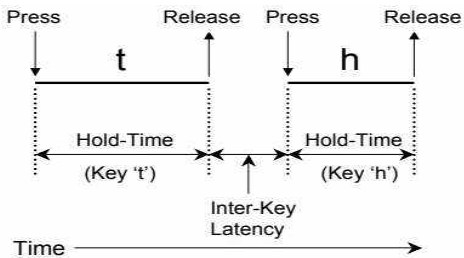
(그림 7) 제안 모델의 예시

KeyStroke Dynamics

Keystroke Dynamics는 이름 그대로 사용자가 키보드의 입력을 수천 번 모니터링하여 단말기에서 타 이평하는 방법을 분석한다는 것을 의미한다. 초당, 그리고 특정한 습관적인 타이핑 리듬 패턴에 근거하여

사용자를 식별하는 것을 목표로 한다.[4]

Keystroke Dynamics는 별도의 하드웨어 장비를 필요로 하지 않고 구현이 쉽다는 장점이 있다. Keystroke Dynamics 기반의 인증 방법은 1975년에 사용자의 키보드 타이핑 리듬에서 시간 패턴, 키에 가해지는 압력을 Spillane가 이용하여 처음으로 제안되었다[12]. 일반적으로 Keystroke Dynamics는 한 키를 누른 시점부터 다음 키를 누를 때까지의 시간인 키 입력시간(latency)과 한 키를 눌렀다가 뗄 때까지의 시간인 키 지속시간(hold-time) 등의 시간을 체크한다.[2]



(그림 8) 이중글자 “th” 입력 시 키간의 지속시간과 입력시간[2]

Keystroke Dynamics 데이터의 등록 과정에서 자연스러운 Keystroke Dynamics 방법과 인공적인 Keystroke Dynamics 방법이 있다. 때로는 인공적인 리듬을 추가함으로써 효과적인 인증 및 인증의 성능을 향상시킬 수 있다. 그러나 인공적인 인증을 할 때 자신이 기억할 수 있는 인공적인 리듬을 기억해야만 한다. 인공적인 리듬을 사용할 때, 특정할 문자를 입력할때 3초간 누르는 것도 리듬이 가미된 Keystroke Dynamics 방법이라 할 수 있다[15]

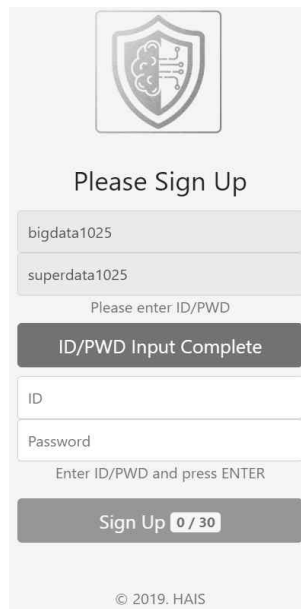
생체 정보는 개인의 고유한 특징 값으로 한번 노출되면 다시 사용할 수 없다는 한계점을 가지고 있다. Keystroke Dynamics는 개인 식별 번호나 패턴 모양에 따라 같은 사용자더라도 값이 다르므로 생체 인증이 벗어날 수 없는 한계점을 벗어나는 장점을 가지고 있다.[9]

1985년에 데스크톱 컴퓨터에서 키 입력 시간 특징을 이용한 Keystroke Dynamics 기반의 인증 실험을 통해 12%의 오거부율 6%의 오인식률을[10], 1990년

에는 13.3%의 오거부율, 0.17%의 오인식률을 결과로 보여주며 Keystroke Dynamics 인증 방법의 가능성을 제시했다[11].

3.2 등록 단계

등록 단계는 회원가입 시 사용하려는 ID/Password에 대해 30번 입력 요구를 해서 사용자의 Key-Stroke Dynamics 데이터 샘플을 추출 획득한다.



(그림 9) 등록 단계의 예시

0.00	1.23	0.19	0.08	1.48	0.06	0.1	0.9	0.06	0.07	0.12	0.17	0.46	0.05	1.37	0.02	0.27	1.05	0.04	1.06		
0.14	1.01	0.06	0.07	0.15	0.03	0.12	0.19	1.01	0.06	0.19	1.16	0.02	0.14	1.16	0.01	0.03	0.19	1.02	0.06	0.06	
0.04	1.10	0.06	0.09	1.16	0.05	0.06	0.16	0.03	0.04	0.16	0.16	0.06	0.14	1.16	0.02	0.02	0.19	1.02	0.06	0.06	
0.08	1.19	0.05	0.09	1.01	0.02	0.03	0.13	0.02	0.05	0.08	1.47	0.06	0.05	1.44	0.46	0.03	0.17	1.04	0.03	0.17	
0.05	1.08	0.03	0.05	1.47	0.07	1.12	0.14	0.04	1.06	0.05	0.12	0.19	0.02	0.01	0.14	0.03	0.07	0.92	0.07	0.92	
0.06	1.24	0.08	0.04	0.19	0.07	0.03	0.16	1.03	0.16	0.02	1.07	0.02	0.06	0.12	0.04	0.07	0.19	0.02	0.03	0.99	
0.07	0.14	0.14	0.04	1.03	0.02	0.09	0.09	0	0.14	0.09	1.02	0.05	0.13	1.47	0.04	0.06	0.15	1.04	0.01	0.92	
0.07	1.08	0.01	0.02	0.12	0.06	0.02	0.15	1.03	0.05	0.12	0.17	0.05	0.12	1.07	0.05	1.03	0.19	1.01	0.01	0.15	
0.10	1.14	0.06	0.03	0.19	0.08	0.19	0.15	0.17	1.42	0.01	0.02	1.48	0.05	0.05	0.17	0.02	0.04	0.19	1.07	0.09	0.92
0.11	1.16	0.09	0.07	1.14	0.09	0.07	0.19	1.04	0.07	0.19	1.04	0.07	0.19	1.16	0.04	0.19	1.12	0.01	0.14	1.07	0.14
0.12	1.27	0.07	0.05	1.04	0.02	0.06	0.05	1.02	0.05	0.16	1.19	0.47	0.01	0.14	0.03	0.03	0.19	1.03	0.03	0.14	
0.07	1.20	0.09	0.07	0.16	0.09	0.08	0.16	1.07	0.10	0.18	0.15	0.46	0.14	1.12	0.03	1.03	1.03	0.12	0.06	0.01	0.91
0.03	1.16	0.02	0.07	0.14	0.07	0.07	0.14	1.47	0.19	0.16	1.08	0.03	1.11	0.16	0.02	0.07	0.12	1.03	0.04	0.91	
0.15	1.19	0.19	0.03	1.14	0.06	0.12	0.16	1.05	1.07	0.09	1.42	0.01	0.16	1.04	0.02	0.04	1.04	0.16	1.07	0.12	0.42
0.08	0.14	0.07	0.07	0.16	0.06	0.03	0.05	1.02	0.01	0.1	1.17	0.07	0.14	1.12	0.04	0.06	0.12	1.04	0.06	0.14	
0.06	0.21	0.19	0.03	0.16	0.06	0.03	0.19	1.05	0.05	0.19	1.46	0.03	0.14	1.19	0.09	0.09	0.19	1.04	0.06	0.49	
0.06	0.14	0.06	0.01	0.14	0.06	0.05	0.17	1.12	0.09	0.14	1.14	0.06	0.12	1.11	0.03	0.06	0.17	1.07	0.06	0.46	
0.07	1.07	0.09	0.07	0.16	0.03	0.17	0.19	1.02	0.01	0.16	1.04	0.03	0.19	1.19	0.03	0.05	0.19	1.04	0.03	0.53	
0.07	1.16	0.11	0.02	0.16	0.04	0.12	0.19	0.05	0.17	0.06	1.03	0.04	1.11	0.19	0.03	0.02	0.12	0.07	0.07	0.38	
0.04	0.14	0.19	0.07	0.14	0.04	0.17	0.19	1.02	0.14	0.17	1.01	0.04	0.13	1.12	0.04	0.15	0.12	1.02	0.05	0.42	
0.06	0.14	0.19	0.07	0.14	0.04	0.17	0.14	1.05	0.12	0.06	1.06	0.07	0.11	0.09	0.04	0.19	1.03	0.04	0.19	0.36	
0.04	1.02	0.10	0.06	0.19	0.07	0.12	0.19	1.04	0.01	0.06	1.03	0.04	0.16	1.16	0.19	1.17	0.17	1.07	0.07	0.91	
0.01	1.16	0.14	0.09	0.14	0.04	0.11	0.14	1.05	0.03	0.16	1.09	0.01	0.13	1.04	0.01	0.09	0.18	1.46	0.06	0.49	
0.11	1.14	0.14	0.14	0.12	0.02	0.09	0.11	1.02	0.14	0.11	1.07	0.07	0.12	1.16	0.01	0.07	0.19	1.03	0.04	0.19	
0.07	1.19	1.09	0.09	1.14	0.07	0.05	0.09	1.01	0.11	0.15	0.14	0.16	0.17	1.16	0.14	0.06	0.19	1.09	0.01	0.36	
0.08	1.14	0.12	0.09	0.14	0.01	0.11	0.17	1.01	0.06	0.16	1.06	0.04	0.15	1.42	0.07	0.04	0.14	1.05	0.04	0.57	
0.02	1.18	0.19	0.02	0.16	0.05	1.03	0.09	0.05	0.17	1.05	0.05	0.18	0.17	1.06	0.15	0.19	1.03	0.01	0.44	0.44	
0.18	1.12	0.06	0.11	0.16	0.07	0.05	0.09	1.01	0.11	0.15	0.14	0.16	0.17	1.16	0.14	0.06	0.19	1.09	0.01	0.36	
0.12	1.01	0.11	0.07	0.19	0.04	0.16	0.17	1.01	0.19	1.01	0.05	1.00	0.04	1.14	0.05	0.19	1.09	0.09	0.57	0.57	
0.06	1.19	0.13	0.09	1.18	0.02	0.16	0.05	1.12	0.05	0.16	1.02	0.05	0.14	1.17	0.02	0.05	0.12	1.05	0.04	0.19	

(그림 10) 등록된 Feature 예시

이 때 사용자마다 ID/Password가 달라서 데이터의 길어도 다르기 때문에 Key-Stroke Dynamics 데이터들은 MultiLabelBinarizer를 이용해 Feature Hashing을 수행한 후 사용한다.

(그림 11) MultiLabelBinarizer를 이용한 Feature Hashing 후 데이터 예시

3.3 인증 단계

인증 단계는 인증 시도 시 입력한 ID/Password에 대한 Key-Stroke Dynamics 데이터가 사용자의 데이터 샘플과 어느정도 유사한지를 KNN 알고리즘으로 비교하여 인증한다. 인증에 성공할 시 해당 데이터는 샘플에 추가되어 인증에 성공할 때 마다 정확도가 올라간다. 입력한 Key-Stroke Dynamics 데이터가 허용 범위를 넘어서는 경우 이상징후로 하여 모니터링을 통한 별도의 조치를 한다.

KNN(K-Nearest_Neighbors)

1968년 Cover, Hart에 의해 KNN(K-Nearest neighbor)은 제안된 알고리즘이다. KNN은 k- 최근접 이웃 알고리즘이라고 부른다. k개의 가장 가까운 이웃을 이용하기 때문에 불리는 명칭인데, 훈련 데이터 집합에 있는 표본 간의 유사도에 따라 라벨이 붙지 않은 표본들을 매우 직관적인 방법으로 분류하는 방법이다. 즉 라벨이 없는 표본이 주어질 경우, 훈련 데이터 집합에서 가장 가까운 k개의 라벨 표본을 찾고, k개씩 묶었을 때 나타나는 집합 안에 빈도수가 많은 그룹에 주어진 표본은 할당하는 방법이다.[5]

- (a) Euclidean Distance $D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- (b) Manhattan Distance $D = \sum_{i=1}^n |x_i - y_i|$
- (c) Minkowski Distance $D = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$

(그림 12) KNN의 유사도 측정 방법[5]

KNN의 유사도 측정 방법에는 여러 가지 방법이 있다. 데이터의 흐름과 분포에 따라 어떤 측정법을 사용하는지에 따라 결과가 달라진다. 대부분의 경우, 유클리드 거리 측정을 사용하여 유사도를 측정한다. KNN은 비 모수적 방법이기 때문에 어떤 분포이든 상관없이 사용할 수 있고, 알고리즘의 특성상 쉽기 이해하기 직관적이다.

K의 값을 크게 줄 경우, 대세의 흐름을 알 수 있지만, 세분하게 분류하지 못한다는 단점이 있다. 반대로 K의 값이 작을 경우 너무 미세하게 구분되어 오류가 생길 확률이 커진다. 데이터의 개수 와 클래스의 개수에 따라 K를 적절히 선택해야 올바른 분류가 가능해진다.[5]

4. 시험 결과

본 장은 본 논문에서 제안한 Key-Stroke Dynamics 기반의 Two-Factor 인증 시스템 모델이 어느 정도의 성능이 있는지 5-Cross Validation 기반으로 검증한 결과를 다룬다.

4.1 제안 모델의 결과

본 논문의 제안 모델은 ID/Password의 일치 여부를 확인하고, 입력된 Key-Stroke Dynamics 데이터 패턴과 사용자의 Key-Stroke Dynamics 데이터 패턴이 어느정도 유사한지를 통해 인증 성공 여부를 결정한다.

인증에 성공하게 되면 입력한 ID/Password에 맞는 사용자의 Key-Stroke Dynamics 데이터 패턴과

입력된 Key-Stroke Dynamics 데이터 패턴의 유사도가 어느정도로 일치해서 인증에 성공했는지 화면에 출력한다.

인증에 실패한다면 입력한 ID/Password에 맞는 사용자가 아니라 회원가입된 사용자들의 Key-Stroke Dynamics 데이터 패턴과 가장 높은 유사도를 보이는 사용자로 의심된다는 화면을 출력한다.

4.2 성능 검증 과정

본 논문에서 제안한 모델의 성능 검증은 5-Cross-Validation을 이용한다.

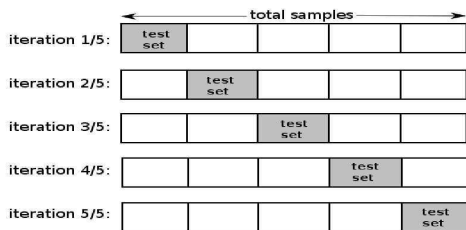
본 제안 모델이 등록 단계에서 입력받은 사용자의 Key-Stroke Dynamics 기본 샘플 30개를 5-Cross-Validation을 사용하여 성능 검증을 하였다.

교차검증(Cross Validation)

교차검증은 특정 Dataset에만 잘 맞는 과도적합(Overfitting)을 방지하고 다양하게 학습데이터와 검증 데이터를 지정하여 검증하는 방법으로 명시적인 검증을 할 수 없는 경우 간접적으로 모델을 검증할 수 있는 효과적인 방법이다.[6]

K-Fold

단순히 데이터를 분리하면 신뢰도가 떨어진다. 그렇다고 검증 데이터 수를 증가시키면 학습용 데이터 수가 적어지므로 정상적인 학습이 되지 않는다. 이러한 문제를 해결하기 위한 검증방법인 K-Fold 교차검증 방법을 사용한다.



(그림 13) 5-Fold 교차검증(Cross Validation)[6]

원본 데이터를 5개로 분할 하여 각각 한번씩 검증

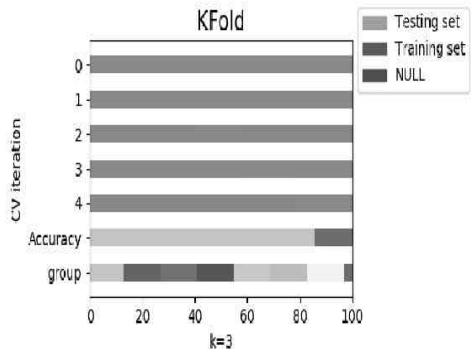
데이터로 사용하고 나머지는 학습데이터로 사용하여 5번의 검증을 통하여 평균을 측정한다.[6]

성능 검증

7명의 사용자에게서 등록 단계에 입력받은 Keystroke Dynamics 데이터를 5개로 분할하여 분할된 데이터를 각각 한번씩 테스트 데이터로 사용하고 나머지 4개의 데이터를 학습데이터로 사용하였다.

K=3일때의 KNN 알고리즘을 사용하였고 cross_val_score을 사용해 정확도를 구하고, 평균을 측정하였다.

4.3 성능 검증 결과



(그림 14) 5-Cross Validation의 결과

평균 Accuracy 87%의 결과가 산출되었다. Key-Stroke Dynamics 기반의 본 모델이 사용자를 높은 정확도로 판단하는 것을 알 수 있다.

사용자 판단에 대한 오탐지율은 인증 성공 시의 Key-Stroke Dynamics 데이터를 사용자의 데이터 샘플에 추가하는 방식으로 샘플을 늘려나가 낮출 수 있다.

본 기술을 도입하지 않았을 경우 일반적인 사용자 인증 방식은 사용자에게 ID/Password 일치 여부만 확인하기 때문에 일치할 경우 인증에 성공한다. 실제 사용자가 아닌 공격자가 인증을 시도해도 인증에 성공하므로 오탐지율이라는 말 자체가 무의미하다.

5. 결론

5.1 제안 모델의 결론

본 논문은 Key-Stroke Dynamics를 기반으로 한 별도의 행위가 필요없는 Two-Factor 인증 방식을 제안하였다. Key-Stroke Dynamics를 이용한 결과 ID/Password를 올바르게 입력하더라도 사용자의 Key-Stroke Dynamics 데이터 패턴과 높은 일치율을 보여야만 인증이 성공하기 때문에 ID/Password를 출력시에도 높은 보안을 유지할 수 있음을 보였다.

또한 Key-Stroke Dynamics 데이터가 사용자 인증에 적합함을 보였다. 123123과 같은 단순한 비밀번호의 경우 Key-Stroke Dynamics 데이터 패턴도 유사하게 나타날 수 있으나 대,소문자와 숫자, 특수문자 등이 다양하게 쓰일 경우 Key-Stroke Dynamics 데이터 패턴은 사용자만의 고유한 리듬 패턴이 될 수 있다.

Keystroke Dynamics를 활용하는 본 기술은 앞서 서술한 Keystroke Dynamics의 장점을 그대로 가져올 수 있다. 해당 기술은 별도의 행위나 장치가 필요하지 않으므로 비용적인 부분에서의 지출이 없고, ID/Password를 입력하는 기존의 Legacy에 적용이 쉽다.

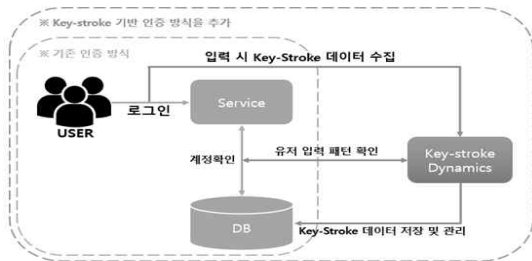
평균 정확도는 87%로 높게 나타났으며 인증 성공시마다 사용자 샘플의 추가하기 때문에 인증을 성공하는 것만으로 오탐지율을 낮추고 보안율을 더욱 높일 수 있다.

5.2 향후 과제

향후의 연구 과제로는 최초 로그인 및 회원가입시 학습을 위한 번거로운 수 차례의 입력 과정은 샘플에서만 사용한 임시 방안이므로 보다 나은 방법을 연구한다.

본 논문이 제안하는 시스템의 실제 적용시에는 개선된 방법과 함께 1차 인증으로 ID/Password 일치 여부를 판단한 후 보조 인증으로서 Key-Stroke Dynamics를 사용한다.

또한 Key-Stroke Dynamics 자체가 100% 정확한 인증을 담보할 수 없으므로 사용자별 2차 인증의 누적된 결과에 대한 해석을 통해서 계정 사용자가 맞는지 정책적으로도 판단할 수 있다. 상시 운용할 필요는 없고 해당 사용자의 결과가 의심될 경우에 유연하게 정책 운용이 가능하다.



(그림 15) 기존 인증 방식에 2차인증을 추가한 방식의 전체 프레임

	Key-Stroke 기반 Two-Factor 인증 기술	기존 ID/Password 인증 기술	기존 two-factor 인증 기술
인증의 편리성	○	○	△
계정정보 노출 시 대응 가능성	○	△	○
사용자들의 불편함 해소	○	○	△
Legacy와의 연계 운용	○	○	○
구축 비용	○	○	△

(그림 16) 타 기술과의 비교표

참고문헌

- [1] 황성섭, et al. (2006). "키스트로크 다이나믹스 분석을 이용한 모바일 사용자 인증." 한국경영과학회 학술대회논문집, 652-655.
- [2] 선아영, and 정일용. "키스트로크 다이나믹스가 적용된 볼륨버튼 기반의 패스워드 인증 기법." 예술인문사회 융합 멀티미디어 논문지 9.5 (2019): 855-863.
- [3] 강필성, and 조성준. (2012). "자유로운 문자열의 키스트로크 다이나믹스를 활용한 사용자 인증 연구." 산업공학, 25(3), 290-299.
- [4] Monrose, Fabian, and Aviel Rubin. "Authentication via keystroke dynamics." Proceedings of the 4th ACM conference on Computer and communications security. ACM,

- 1997.
- [5] 임상미, et al. "KNN(K-nearest neighbors) 기반 음악 매칭 애플리케이션을 이용한 사운드 아트" VOL.- NO.- (2018)
- [6] 황찬웅, and 이태진. "KNN기반 악성코드 분석 최적화 방안 연구." 2019 한국정보보호학회 하계학술대회 논문집 Vol. 29, No. 1
- [7] Foresi, Andrew, and Reza Samavi. "User Authentication Using Keystroke Dynamics via Crowdsourcing." 2019 17th International Conference on Privacy, Security and Trust (PST). IEEE, 2019.
- [8] Darabseh, Alaa, and Akbar Siami Namin. "On Accuracy of Classification-Based Keystroke Dynamics for Continuous User Authentication." 2015 International Conference on Cyberworlds (CW). IEEE, 2015
- [9] 이현구. "성별 일치 여부에 따른 모션 센서를 추가한 키스트로크 다이내믹스" VOL.- NO.- (2018)
- [10] Umphress, David, and Glen Williams. "Identity verification through keyboard characteristics." International journal of man-machine studies 23.3 (1985): 263-273
- [11] Joyce, Rick, and Gopal Gupta. "Identity authentication based on keystroke latencies." Communications of the ACM 33.2 (1990): 168-176.
- [12] Spillane, R. "Keyboard apparatus for personal identification." IBM Technical Disclosure Bulletin 17 (1975): 3346.
- [13] Darabseh, Alaa, and Akbar Siami Namin. "Effective user authentications using keystroke dynamics based on feature selections." 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 2015.
- [14] Ahmad, Abd Manan, and Nik Nailah Abdullah. "User authentication via neural network." International Conference on

Artificial Intelligence: Methodology, Systems, and Applications. Springer, Berlin, Heidelberg, 2000.

- [15] 김천식, et al.(2008).이러닝 시스템에서 사용자 인증을 위한 키스트로크의 응용 기술.전자공학 회논문지-CI,45(5),25-31.

【 저 자 소개 】



안 준 연 (Jun-Yeon An)
2015년 3월 ~ 현재 : 호서대학교 정보보호학과 재학
<관심분야> 정보보호, 기계학습



고 광 필 (Ko-Gwang Feel)
2015년 3월 ~ 현재 : 호서대학교 정보보호학과 재학
<관심분야> 정보보호, 기계학습



이 태 진 (Tae-jin Lee)
2003년 1월~2017년 2월 : 한국인터넷진흥원 팀장
2017년 3월~현재: 호서대학교 컴퓨터정보공학부 교수
<관심분야> 시스템 보안, 악성코드 분석, 기계학습
kinjecs0@gmail.com