Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.6, Dec. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.6.1419

온라인 간편 결제 환경에서 기계학습을 이용한 무자각 인증 기술 연구*

류 권 상,† 서 창 호, 최 대 선‡
공주대학교

A Study on Unconsciousness Authentication Technique Using Machine Learning in Online Easy Payment Service*

Gwonsang Ryu,† Changho Seo, Daeseon Choi‡
Kongju National University

요 약

최근 환경기반 인증 기술로 사용자의 로그인 히스토리를 계정도용 또는 정상 로그인으로 분류한 후 사용자별로 통계모델을 만들어 사용자를 인증하는 Reinforced authentication이 제안되었다. 하지만 Reinforced authentication은 사용자가 과거에 계정도용을 당한 적이 없으면 공격을 당할 가능성이 높다. 본 논문은 이러한 문제점을 해결하기 위해 기계학습 알고리즘을 이용하여 사용자 환경정보와 타인의 환경정보를 함께 학습시켜 2-Class 사용자 모델을 만드는 무자각 인증 기술을 제안한다. 제안한 기술의 성능을 평가하기 위해 목표 사용자에 대해 아무 정보도 없는 무 지식 공격자와 목표 사용자에 대해 한 가지의 정보만 알고 있는 정교한 공격자에 대한 Evasion Attack을 실험하였다. 무 지식 공격자에 대한 실험 결과 Class 0의 Precision과 Recall 각각 1.0과 0.998로 측정되었으며, 정교한 공격자에 대한 실험 결과 Class 0의 Precision과 Recall 각각 0.948과 0.998로 측정되었다.

ABSTRACT

Recently, environment based authentication technique had proposed reinforced authentication, which generating statistical model per user after user login history classifies into account takeover or legitimate login. But reinforced authentication is likely to be attacked if user was not attacked in past. To improve this problem in this paper, we propose unconsciousness authentication technique that generates 2-Class user model, which trains user's environmental information and others' one using machine learning algorithms. To evaluate performance of proposed technique, we performed evasion attacks: non-knowledge attacker that does not know any information about user, and sophisticated attacker that only knows one information about user. Experimental results against non-knowledge attacker show that precision and recall of Class 0 were measured as 1.0 and 0.998 respectively, and experimental results against sophisticated attacker show that precision and recall of Class 0 were measured as 0.948 and 0.998 respectively.

Keywords: Authentication, Machine Learning, Account Takeover, Fraud Detection

Received(10. 20. 2017), Modified(1st: 11. 23. 2017, 2nd: 12. 11. 2017), Accepted(12. 11. 2017)

* 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(No.2016-0-00173, 핀테크 서비스 금융사기 방지를 위한 비대면 본인확인) 및 한국연구재단의 지원을 받아 수행된 연구임 (No. 2016R1A4A1011761, 핀테크 서비스를 위한 금융 보안 핵심 기술 개발)

† 주저자, gsryu1026@smail.kongju.ac.kr
‡ 교신저자, sunchoi@kongju.ac.kr(Corresponding author)

I. 서 론

오늘날 대부분의 온라인 서비스에서 비밀번호를 사용하여 사용자를 인증한다. 하지만 비밀번호는 간단하거나 추측하기 쉬운 비밀번호 사용[1], 여러 서비스에서 동일한 비밀번호 사용[2], Shoulder surfing 공격[3] 등 많은 보안 결함을 가지고 있다. 이러한 비밀번호의 문제점을 다루기 위해 행위기반 인증기술 및 환경기반 인증기술이 연구되고 있다.

행위기반 인증기술은 사용자의 행동 정보를 인증 요소로 활용하여 사용자를 인증하는 기술로 Key stroke dynamic[4][5][6], 동적서명[7], 음성 인증[8] 등이 있다. 하지만 행위기반 인증기술은 사용자 디바이스에서 사용자의 행위 데이터를 수집해야하기 때문에 별도의 어플리케이션이나 플러그인이 필요하다. 환경기반 인증기술은 IP주소, 사용한 브라우저 및 운영체제, 위치 등 컴퓨터나 스마트폰 이용환경에 대한 정보를 기반으로 하여 사용자를 인증하는 기술을 말한다. 예를 들어 구글, 페이스북 등은 사용자가 평소에 접속하는 위치 정보, 브라우저 정보 등의 환경정보를 이용하여 계정 도용시도를 탐지하여 사용자에게 알린다. 환경기반 간편 인증기술의 최근 연구로 Reinforced authentication[9]이 있다. Reinforced authentication은 의심스러운 로그인을 탐지하기 위해 LinkedIn의 사용자 로그인 히스토리를 이용하여 사용자별로 계정 도용된 환경정보의 패턴에 따라 분류 경계수준을 조절하는 방식을 제안하였다. 이 방식은 과거에 계정 도용을 당했던 사용자에게는 효과적이지만, 과거에 계정 도용을 당했던 적이 없는 사용자의 경우 분류 경계수준이 낮아져 공격 성공할 가능성이 높다는 단점이 있다. 따라서 계정 도용여부와 상관없이 효과적으로 사용자를 인증할 수 있는 인증 기술이 필요하다.

본 논문에서는 계정 도용여부와 상관없이 사용자를 인증하기 위해 온라인 간편 결제 환경정보를 활용한 무자각 인증 기술을 제안한다. 실제 간편 결제환경에서는 과거에 계정 도용을 당했던 사용자에 비해 계정 도용을 당했던 적이 없는 사용자가 비교적 더 많기 때문에, 기계학습 알고리즘으로 사용자 환경정보와 타인의 환경정보 사이의 경계를 설정하는 2-Class 사용자 모델을 만들어 사용자를 인증한다.

본 논문의 구성은 다음과 같다. 2장에서는 행위기반 인증기술과 환경기반 인증기술에 대한 관련 연구에 대해 설명하고, 3장에서는 제안한 무자각 인증 기술과 공격 모델을 설명한다. 4장은 제안한 기술의 성능을 평가하기 위한 실험과 안정성 및 유용성에 대해서 고찰한다. 마지막으로, 5장에서 결론을 맺는다.

II. 관련 연구

2.1 행위기반 인증기술

Key stroke dynamic은 키보드를 이용하여 사용자가 패스워드의 각 문자를 입력할 때 시간차의 특징을 추출하여 사용자를 인증하는 방법이다. 최근에는 모바일 디바이스의 가상 키보드 상에서 사용자의 Key stroke 패턴을 이용하여 사용자를 인증하는 기술이 많이 연구되고 있다[4][5][6].

Signature dynamic은 인터넷이나 모바일에서 본인 확인을 위해 사용자가 입력한 서명에서 전체 획수, 좌표 점의 수, 교차점의 개수, 획 상의 Pen up이나 Down시간, 획 사이의 시작점과 끝점간의 길이 등 다양한 특징을 추출하여 사용자 인증에 사용한다. 최근에는 스마트 폰에서 손가락으로 서명할 때 스마트폰에 내장된 가속도센서 값을 추가로 사용하여 사용자를 인증하는 기술이 연구되었다[7].

음성 인증 시스템은 화자의 음성 신호에서 특징을 추출하여 사용자를 인증하는 시스템이다. 최근에는 스마트폰에 내장된 2개의 마이크를 이용하여 사용자가 말하는 단어의 음소(Phoneme) 위치를 기반으로 사용자를 인증하는 기술이 연구되었다[8].

위와 같은 기술 외에 스마트폰에서 사용자의 터치 정보 또는 행위 정보를 이용하여 사용자를 인증하는 기술도 연구되었다[10][11]. 또한 사용자의 눈 움직임 특징을 기반으로 사용자를 인증하는 기술도 연구되었다[12][13][14].

2.2 환경기반 인증기술

웹 브라우저는 소프트웨어 및 하드웨어 정보, WebGL 정보, 시스템 시간, 배터리 정보와 같은 다양한 환경정보를 제공한다. 사용자의 IP주소, 접속 시간대, 화면 해상도와 같은 인증된 세션에서 다양한 환경정보를 저장하고, 저장되어 있는 환경정보와 추후 수집된 환경정보 간의 유사도를 측정하여 검증하는 인증기술인 SmartAuth[15]가 연구되었으며, IP주소를 기반으로 사용자의 지리적 위치를 파악하는 사용자 인증기술이 연구되었다[16].

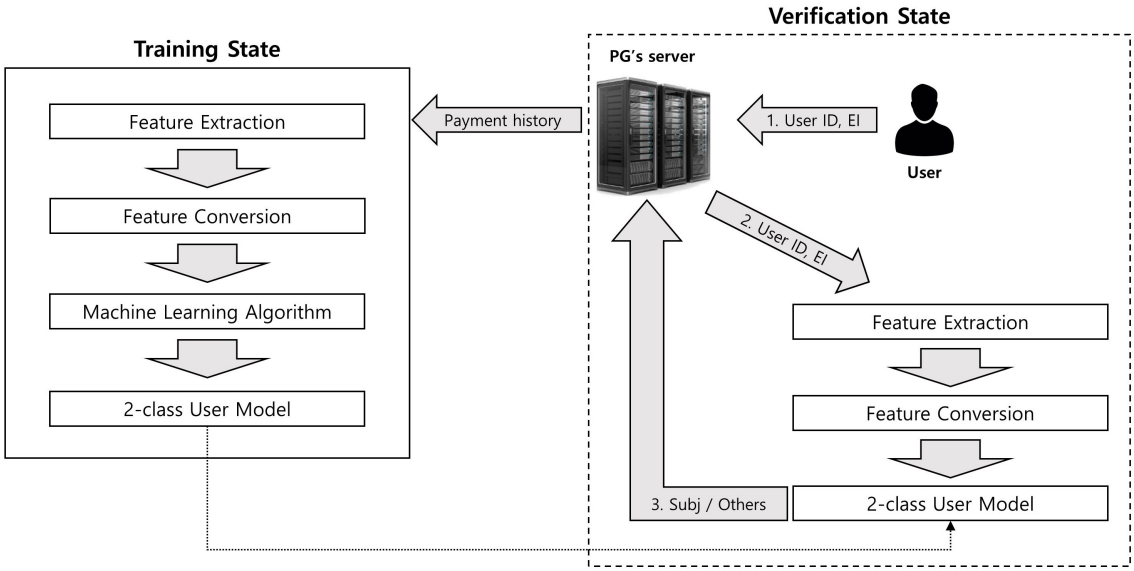

Fig. 1. Architecture of proposed unconsciousness authentication technique

최근에는 의심스러운 로그인을 탐지하기 위해 IP 주소, User-Agent 등과 같은 사용자의 로그인 환경정보를 이용한 Reinforced authentication[9]이 연구되었다.

III. 결제 환경정보 기반 무자각 인증 기술

본 논문에서 제안한 간편 결제 환경정보기반 무자각 인증 기술의 구조는 Fig.1.과 같다. 먼저 PG(Payment Gateway)사 서버는 사용자의 결제 히스토리를 가지고 특징을 추출한 다음 기계학습 알고리즘으로 사용자 데이터와 타인의 데이터를 학습시켜 사용자별로 2-Class 모델을 생성한다. 사용자는 온라인 쇼핑몰에서 물건을 구매하기 위해 간편 결제 서비스에 로그인하여 사용자의 ID와 환경정보(EI, Environmental Information)를 PG사 서버로 전송한다. PG사 서버는 전송받은 환경정보로 특징을 추출하여 2-Class 사용자 모델에 입력한다. 2-Class 사용자 모델이 입력받은 특징 벡터를 사용자로 분류하면 인증 승인, 타인으로 분류하면 인증 거부를 할 것이다.

3.1 데이터 소개 및 특징 선택

본 논문에서 사용한 데이터는 국내 모 PG사의 간편 결제 서비스로 2017년 6월 1일부터 2017년 6월 10일까지 10일 동안 7번 이상 결제한 사용자 3880명의 데이터이며, PG사에서 계정 도용이 아니라고 판단한 사용자의 데이터이다.

사용자 데이터는 IP주소, User-Agent, 결제 금액 등 총 161개의 필드로 이루어져 있다. 161개의 필드 중 의미 있는 필드를 선별하기 위해 Python 라이브러리인 Scikit-Learn에서 제공하는 Tree 기반 특징 선택 알고리즘을 사용하였다. Tree 기반 특징 선택 알고리즘은 해당 특징을 사용하는 노드가 평균적으로 불순(Impurity)을 줄일 수 있는지 살펴보고 특징의 중요도를 측정하는 방식이다. Tree 기반 특징 선택 알고리즘에서 필드가 사용자별로 가장 높은 중요도를 가진 횟수를 측정하였다. Fig.2.는 필드가 사용자별로 가장 높은 중요도를 가진 횟수를 히스토그램으로 보여준다. 특징 선택 알고리즘은 기계적인 계산을 통해 보편적인 선택을 도와주는 방법이기 때문에 의도와 다른 특징이 선택될 수 있다. 따라서 전문가의 경험에 의한 주관적인 의견을 반영하는 휴리스틱 기반 데이터 분석을 통해 필드를 추가하였다. 사용된 필드는 Table 1.에 정리되어 있으며, 괄호 안의 값은 각 필드의 데이터에 대한 예시이다.

PC, 모바일 모두 고려했을 때 국내에서 사용가능한 IP주소는 상당히 많기 때문에 IP대역을 사용하였다. User-Agent는 브라우저 및 운영 체제를 포함하고 있는 정보이며 한 줄의 문자열로 이루어져 있다. Shipping address는 배송 주소로써, 도로 번

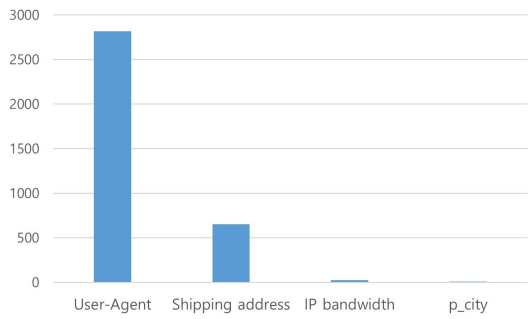

Fig. 2. Histogram of fields that importance was measured high

Table 1. Used fields

IP bandwidth (127.0, ...)	User-Agent (Mozilla/5.0 (Windows NT 6.1; WOW64; ... , ...)
Shipping address (Seoul Gangnamgu Nonhyeon-ro 507, ...)	P_city (Seoul, ...)

호까지 사용하였다. P_city는 사용자가 접속한 도시 정보를 나타낸다.

3.2 특징 변환

사용한 필드는 모두 명목형 데이터이기 때문에 특징 값을 변환해서 사용해야 한다. 특징 변환의 첫 번째 방법(FC1, Feature Conversion)은 각각의 필드마다 고유한 모든 특징 값을 0부터 번호를 부여하는 방식인 Label Encoder를 사용하였다.

두 번째 방법(FC2)은 일반적인 One Hot Encoding 방식으로 FC1처럼 각 필드의 고유한 모든 특징 값들을 0부터 번호를 부여한다. 그 다음 각 필드의 특징 값의 개수만큼 0을 만들어 해당 특징 값이 부여된 번호 위치에 있는 0을 1로 바꾼다. 따라서 해당 필드의 고유한 특징 값이 많을수록 차원수가 늘어나 학습시간이 길어지는 단점이 있다.

세 번째 방법(FC3)은 일반적인 One Hot Encoder를 변형한 방법이다. FC2와 달리 변형 One Hot Encoder는 사용자 본인의 데이터만 인코딩하기 때문에 특징의 차원수가 사용자가 필드별로 사용한 특징 값의 개수만큼 늘어나기 때문에 학습시간에 큰 영향을 주지 않는다. Table 2.는 특징 변환 방법의 예를 보여준다.

3.3 공격 모델

본 논문에서는 현실적으로 공격자가 학습 데이터에 접근하기 어렵다는 가정 하에 제안한 무자각 인증 기술에서 현실적으로 발생할 가능성이 있는 Evasion Attack[17]을 시뮬레이션 하였다. Evasion Attack은 학습 데이터에 영향을 주지 않고, 입력 데이터를 최소한으로 변조하여 학습이 완료된 모델이 공격 데이터를 정상으로 오 분류하도록 하는 것이다. 공격자가 사용자의 주변인이라면 공격자는 사용자가 평소에 사용하는 IP대역, User-Agent와 같은 추가 환경정보를 알고 있을 가능성이 높기 때문에 Evasion Attack을 충분히 시도할 수 있다.

Evasion Attack에서 공격자의 목표는 계정 도용 시도를 정상으로 오 분류하도록 일부 입력 값을 조작하는 것이다. 공격자의 능력은 시스템이 작동중일 때만 입력 데이터를 조작하여 공격할 수 있으며, 학습 데이터는 조작할 수 없다. 로그인 시도에서 각각의 입력 값은 공격자의 지식에 따라 조작된다.

공격자의 지식은 다음과 같이 두 가지로 정의하고 있다. 첫 번째는 ID, 비밀번호를 제외하고 목표 사용자에 대한 임의의 정보가 없는 무 지식 공격자이다. 따라서 공격자는 모든 필드가 가질 수 있는 값에 대해 전수조사 공격(Brute-Force Attack)을 수행해야한다. 본 데이터에서 사용된 IP대역의 수는

Table 2. Example of feature conversion

Field 1	Field 1 Label Encoding	Field 2	Field 2 Label Encoding	Features that user used	Modified One Hot Encoding
'A1'	0	'B1'	0	Field 1: 0, Field 2: 3	10000001
'A2'	1	'B2'	1	Field 1: 2, Field 2: 1	00100100
'A3'	2	'B3'	2	Field 1: 1, Field 2: 0	01001000
'A4'	3	'B4'	3	Field 1: 5, Field 2: 2	00000010

1063개이고 User-Agent의 수는 5733개이므로 전수조사 공격을 수행하기에는 경우의 수가 굉장히 많아 비효율적이다. 따라서 전수조사 공격 보다 더 적은 개수의 데이터로 공격 성공하기 위해 IP대역과 User-Agent를 랜덤하게 10,000건을 조합하여 공격시도를 하였다. 두 번째는 목표 사용자에 대해 한 가지의 필드 값을 알고 있는 정교한 공격자이다. 만약 공격자가 사용자가 사용하는 IP대역만 알고 있다면, 여러 개의 User-Agent를 조합하여 공격시도를 할 수 있다. 또한 공격자가 사용자가 사용하는 User-Agent만 알고 있다면, 여러 개의 IP대역을 조합하여 공격시도를 할 수 있다. 이 공격 또한 알고 있는 입력 값에 대해서 랜덤하게 10,000건을 조합하여 공격시도를 하였다.

IV. 실험 및 고찰

이번 장에서는 2-Class 사용자 모델의 성능을 평가하기 위한 실험을 수행하였으며 안전성 및 유용성에 대한 고찰을 다룬다. 사용한 데이터에서 거래 후, 취소 및 실패한 거래를 제외하고 6번 이상 결제한 3,507명의 사용자를 대상으로 실험하였으며, 사용자 데이터를 Class 0, 타인의 데이터를 Class 1로 설정하였다. 학습 비율은 7:3(학습 데이터:테스트 데이터)으로 설정하였다.

본 실험에서 Python 라이브러리 Scikit-Learn에 구현된 기계학습 알고리즘인 Decision Tree, SVM(Support Vector Machine), Logistic Regression 알고리즘을 사용하여 2-Class 사용자 모델을 만들어 비교 평가하였다. Decision Tree는 기준을 엔트로피, Tree의 최대 깊이를 7로 설정하였다. SVM은 비선형 문제를 고려하기 위해 RBF(Radial Basis Function) 커널을 사용하였으며 오버피팅을 조절하기 위한 변수인 γ는 0.1로 설정하였다. Logistic Regression에서 바이어스-분산 간 균형을 조절하기 위한 변수 C를 500으로 설정하였다.

성능 검증의 평가 척도는 Precision, Recall F1-Score, Confusion Matrix를 사용하였으며 이전 연구와의 성능을 비교하기 위한 평가 척도로 TPR(True Positive Rate), FPR(False Positive Rate)를 사용하였다. 각각의 의미는 다음과 같다.

- Precision: 각 Class로 분류된 데이터 중에서 해당 Class로 올바르게 분류된 데이터의 비율
- Recall: 각 Class에 해당하는 데이터가 각 Class로 올바르게 분류된 데이터의 비율
- F1-Score: Precision과 Recall을 조합한 통합 성능 지표
- TPR: 타인 또는 공격 데이터를 사용자로 오분류한 비율
- FPR: 사용자 데이터를 타인 또는 공격으로 오분류한 비율
- Confusion Matrix: 만들어진 모델의 성능을 나타낸 행렬

4.1 Baseline 및 사용자 모델 비교

본 논문에서 Baseline은 Table 3.과 같이 사용자의 과거 로그인 데이터를 한 줄의 문자열로 만들어 저장한 뒤, 현재 로그인 데이터를 한 줄의 문자열로 만들어 과거에 이용했던 것 중에 완전히 일치하는 것이 있는지 실험하였다. 로그인 데이터를 한 줄의 문자열로 만들어서 비교하므로 한글자라도 다르면 타인으로 분류하기 때문에 Class 0의 Recall이 0.764로 낮게 측정되었다. 또한 4명의 사용자에 대해서 정확히 일치한 26건의 타인 데이터가 있었다. 본 실험에서는 Class 0의 Precision과 Recall 둘 다 최대로 하는 것이 목표이다.

각각의 특징 변환 방법을 이용하여 기계학습 알고리즘으로 만들어진 사용자 모델 중 FC1을 이용하여 SVM으로 사용자 모델을 만들었을 때 Class 0의 recall이 0.997로 가장 높았으며 1명의 사용자에 대해서 사용자로 분류된 타인 데이터는 1건으로 Baseline보다 적었다. 기계학습 알고리즘으로 만들어진 모든 사용자 모델이 Baseline보다 사용자와 타인 데이터를 잘 분류하였다.

세 가지의 특징 변환방법 중 FC1을 이용하여 기계학습 알고리즘으로 사용자 모델을 만들었을 때 FC2 및 FC3보다 사용자와 타인 데이터를 잘 분류하였다. 또한 FC2를 이용했을 때 특징 벡터의 차원 수가 상당히 늘어나 학습시간이 가장 오래 걸렸으며,

Table 3. Example of baseline

Field1	Field2	Baseline
'A1B1'	'C1D1'	'A1B1C1D1'
'A2B2'	'C2D2'	'A2B2C2D2'

Table 4. Performance comparison of baseline and user model

	Feature Conversion	Average Training Time	Confusion Matrix		Precision	Recall	F1-Score
Baseline	-	-	[9252 2855(1768)] [26(4) 129905480]	Class 0	0.997	0.764	0.865
Decision Tree	1	0.015sec	[12023 84(71)] [48(36) 38971657]	Class 0	0.996	0.993	0.995
	2	6.291sec	[11612 495(397)] [95(77) 38971610]	Class 0	0.992	0.959	0.975
	3	0.004sec	[11564 543(430)] [75(53) 38971630]	Class 0	0.994	0.955	0.974
Support Vector Machine	1	2.581sec	[12067 40(36)] [1(1) 38971704]	Class 0	**0.999**	**0.997**	**0.998**
	2	NA	NA	NA	NA	NA	NA
	3	0.033sec	[11589 518(431)] [27(21) 38971678]	Class 0	0.998	0.957	0.977
Logistic Regression	1	0.230sec	[11645 462(392)] [22(19) 38971683]	Class 0	0.998	0.962	0.980
	2	1.931sec	[11213 894(723)] [5(5) 38971700]	Class 0	0.999	0.926	0.961
	3	0.122sec	[11231 876(701)] [2(2) 38971703]	Class 0	0.999	0.928	0.962

FC2와 FC3을 이용하여 기계학습으로 만들어진 사용자 모델의 성능은 큰 차이를 보이지 않았다. Table 4.는 Baseline과 각 특징 변환방법을 이용하여 기계학습 알고리즘으로 만들어진 2-Class 사용자 모델의 성능을 비교한다.

4.2 Evasion Attack에 대한 성능 평가

Evasion Attack에 대한 성능을 평가하기 위해 실험에 사용한 데이터에서 필드의 특징 값들을 샘플링하여 공격데이터를 조합하는데 사용하였다. 또한 배송 주소와 접속 도시는 사용자를 구분하는데 중요하지만 드러날 경우 공격자가 불리하기 때문에, 배송 주소가 필요 없는 전자상품권을 구매하여 공격 시도한다고 가정한다. 그러므로 IP대역 및 User-Agent를 제외한 다른 입력 값은 조작하지 않았다. 사용자의 로그인 데이터 전부와 IP대역 및 User-Agent를 랜덤하게 조합한 공격데이터를 실험에 사용하였다.

4.2.1 무 지식 공격자

무 지식 공격자는 목표 사용자에 대해 아이디와 비밀번호를 제외하고 아무 정보가 없는 공격자이다. 따라서 사용된 IP대역과 User-Agent를 랜덤하게 조합하여 공격하였다. FC1로 특징 변환한 후 SVM으로 사용자 모델을 만들었을 때, Class 0의 Recall이 0.999로 성능이 가장 좋게 측정되었으며, Precision은 0.999로 공격 시도를 대부분 탐지하는 모습을 보였다. Decision Tree로 사용자 모델을 만든 경우 대부분의 사용자 모델이 Root노드에서 사용자를 구분하는데 특정 필드의 특징 값만 보고 판단하기 때문에 Precision이 모두 0.9 미만으로 다른 기계학습 알고리즘보다 공격 시도를 탐지하는 성능이 떨어졌다. Logistic Regression으로 만들어진 사용자 모델은 다른 기계학습 알고리즘보다 사용자를 타인으로 분류한 데이터가 많았지만, Class0의 Precision이 모두 0.99 이상으로 측정되어 공격 시도를 탐지하는 성능은 상당히 좋게 측정되었다. 무 지식 공격자에 대한 실험 결과는 Table 5.에 정리되어 있다.

Table 5. Performance evaluation against non-knowledge attacker

	Feature Conversion	Confusion Matrix		Precision	Recall	F1-Score
Decision Tree	1	[35777 91(74) / 8202(2568) 35061798]	Class 0	0.814	0.997	0.896
	2	[35366 502(412) / 4385(1627) 35065615]	Class 0	0.890	0.986	0.935
	3	[35318 550(444) / 4508(1756) 35065492]	Class 0	0.887	0.985	0.933
Support Vector Machine	1	[35821 47(37) / 4(3) 35069996]	Class 0	**0.999**	**0.999**	**0.999**
	2	NA	NA	NA	NA	NA
	3	[35335 533(435) / 2084(924) 35067916]	Class 0	0.944	0.985	0.964
Logistic Regression	1	[35212 656(507) / 865(181) 35069135]	Class 0	0.976	0.982	0.979
	2	[34959 909(725) / 205(79) 35069795]	Class 0	0.994	0.975	0.984
	3	[34977 891(703) / 207(87) 35069793]	Class 0	0.994	0.975	0.985

4.2.2 정교한 공격자

정교한 공격자는 목표 사용자에 대해서 IP대역 또는 User-Agent의 특징 값만 추가로 알고 있는 공격자이다. 따라서 특정 필드의 공격자가 알고 있는 특징 값에 다른 필드의 특징 값을 랜덤하게 조합하여 공격하였다.

FC2을 이용하여 SVM으로 사용자 모델을 만들었을 때 Precision이 0.971로 측정되어 목표 사용자가 어떤 IP대역을 사용하는지 알고 있는 정교한 공격자의 공격 시도 탐지 성능이 가장 좋게 측정되었다. FC1을 이용하여 SVM으로 만들어진 사용자 모델과 FC2 및 FC3을 이용하여 Logistic Regression으로 만들어진 사용자 모델을 제외한 나머지 방법들은 Class 0의 Precision이 0.8 미만으로 IP대역을 알고 있는 정교한 공격자에게 상당히 취약하였다. 목표 사용자의 IP대역을 알고 있는 정교한 공격자에 대한 실험 결과는 Table 6.에 정리되어 있다.

목표 사용자가 어떤 User-Agent를 사용하는지 알고 있는 정교한 공격자에 대한 실험 결과 FC1을 이용하여 SVM으로 사용자 모델을 만들었을 때 Precision은 0.859로 공격 시도를 가장 잘 탐지하였다. FC1을 이용하여 SVM으로 만들어진 사용자 모델을 제외한 나머지 방법 모두 Precision이 0.8 미만으로 User-Agent를 알고 있는 정교한 공격자에게 상당히 취약한 모습을 보였다. 특히 Decision Tree의 경우 대부분의 사용자 모델이 root노드에서 User-Agent의 특징 값을 보고 사용자인지 아닌지 분류하기 때문에 성능이 가장 좋지 않았다. 목표 사용자의 User-Agent를 알고 있는 정교한 공격자에 대한 실험 결과는 Table 7.에 정리되어 있다.

정교한 공격자는 목표 사용자에 대해 추가 정보를 알고 있어 입력 값을 조작하기 때문에 2-Class 사용자 모델이 공격 데이터를 목표 사용자로 오 분류하여 안전성이 떨어진 것을 볼 수 있다.

4.3 이전 연구와의 성능 비교

본 논문에서 제안한 2-Class 사용자 모델은 FPR이 0.1%로 Reinforced authentication[9]보다 사용자와 타인 데이터를 잘 구분하였으며, 목표 사용자의 User-Agent를 알고 있는 정교한 공격자에 대한 실험에서는 TPR이 25% 향상되어 Reinforced authentication보다 공격 시도를 잘 탐지하였다. Table 8.은 이전 연구와 본 연구의 성

Table 6. Performance evaluation against sophisticated attacker that knows IP bandwidth

	Feature Conversion	Confusion Matrix		Precision	Recall	F1-Score
Decision Tree	1	[35777 91(74) / 172117(2598) 34897883]	Class 0	0.172	0.997	0.294
	2	[35366 502(412) / 506590(1699) 34563410]	Class 0	0.065	0.986	0.122
	3	[35318 550(444) / 404628(1854) 34665372]	Class 0	0.080	0.985	0.148
Support Vector Machine	1	[35821 47(37) / 1073(186) 35068927]	Class 0	**0.971**	**0.999**	**0.985**
	2	NA	NA	NA	NA	NA
	3	[35335 533(435) / 73377(1465) 34996623]	Class 0	0.325	0.985	0.489
Logistic Regression	1	[35212 656(507) / 22809(373) 35047191]	Class 0	0.607	0.982	0.750
	2	[34959 909(725) / 5891(479) 35064109]	Class 0	0.856	0.975	0.911
	3	[34977 891(703) / 1308(649) 35068692]	Class 0	0.964	0.975	0.970

Table 7. Performance evaluation against sophisticated attacker that knows user-agent

	Feature Conversion	Confusion Matrix		Precision	Recall	F1-Score
Decision Tree	1	[35777 91(74) / 27950381(2861) 7119619]	Class 0	0.001	0.997	0.003
	2	[35366 502(412) / 15879711(1970) 19190289]	Class 0	0.002	0.986	0.004
	3	[35318 550(444) / 17329156(2161) 17740844]	Class 0	0.002	0.985	0.004
Support Vector Machine	1	[35821 47(37) / 5860(189) 35064140]	Class 0	**0.859**	**0.999**	**0.924**
	2	NA	NA	NA	NA	NA
	3	[35335 533(435) / 8753591(1867) 26316409]	Class 0	0.004	0.985	0.008
Logistic Regression	1	[35212 656(507) / 721121(555) 34348879]	Class 0	0.047	0.982	0.089
	2	[34959 909(725) / 673966(722) 34396034]	Class 0	0.049	0.975	0.094
	3	[34977 891(703) / 777592(973) 34292408]	Class 0	0.043	0.975	0.082

Table 8. Performance comparison with previous research

Research	Non-knowledge Attacker	Sophisticated Attacker that knows user-agent
Proposed method(FC1, SVM)	FPR: 0.001, TPR: 0.99	FPR: 0.001, TPR: 0.99
Reinforced authentication[9]	FPR: 0.1, TPR: 0.99	FPR: 0.1, TPR: 0.74

능을 비교한다.

4.4 고 찰

제안한 인증 기술의 안전성을 평가하기 위해 Evasion Attack을 실험하였다. 무 지식 공격자 및 정교한 공격자에 대한 Evasion Attack 실험 결과 FC1을 이용하여 SVM으로 만들어진 사용자 모델을 만들었을 때, 이전 연구보다 성능이 향상된 것을 볼 수 있다. 이는 이전 연구보다 높은 정확도로 사용자를 인증할 수 있으며, 공격 시도 탐지를 잘할 수 있다는 것을 의미한다. 실제 간편 결제환경에서 공격 시도를 잘 탐지하는 것은 이상거래탐지를 잘할 수 있다는 것을 의미하기 때문에 사용자의 금전적인 손해를 줄일 수 있다.

유용성에 대해서, 제안한 인증 기술은 행위기반 인증 기술과는 달리 사용자에게 임의의 행동을 요구하지 않으며 플러그인 또는 어플리케이션 설치를 요구하지 않는다. 사용자 모델은 사용자의 과거 간편 결제서비스 히스토리를 기반으로 만들어지기 때문에 사용자는 단지 자신이 평소에 간편 결제서비스를 사용하던 환경에서 접속하면 된다. 사용자의 행동 특징을 지속적으로 모니터링 하는 지속인증과는 달리, 제안한 인증 기술은 사용자가 간편 결제 서비스를 이용하여 결제할 때 사용자가 자각하지 못하도록 한 번만 인증하는 무자각 인증이다.

V. 결 론

본 논문에서는 계정도용 여부와 상관없이 효과적으로 사용자를 인증하기 위해 기계학습 알고리즘으로 사용자 환경정보와 타인 환경정보의 경계를 설정하여 2-Class 사용자 모델을 만드는 간편 결제 환경정보를 활용한 무자각 인증 기술을 제안하였다. 제안한 인증 기술의 성능을 평가하기 위해 국내 PG사의 실제 데이터를 샘플링하여 Evasion Attack을 실험하였다.

목표 사용자에 대한 임의의 정보가 없는 무 지식 공격자에 대한 실험 결과 FC1을 이용하여 SVM으로 사용자 모델을 만들었을 때, Precision이 0.999으로 공격 시도를 대부분 탐지하는 모습을 보였다. 목표 사용자가 사용하는 IP대역을 알고 있는 정교한 공격자에 대한 실험 결과 FC2를 이용하여 SVM으로 사용자 모델을 만들었을 때, Precision이 0.991로 가장 높게 측정 되었으며, 목표 사용자가 사용하는 User-Agent를 알고 있는 정교한 공격자에 대한 실험 결과 FC1을 이용하여 SVM으로 사용자 모델을 만들었을 때 Precision이 0.859로 다른 알고리즘보다 공격 시도를 잘 탐지하였다. 또한 이전 연구보다 FPR과 TPR이 상당히 향상된 것을 볼 수 있다. 실제 간편 결제 데이터를 분석하여 실험한 결과 높은 성능을 보였기 때문에 실제 간편 결제 서비스에서 활용될 수 있을 것으로 기대된다.

그러나 가장 성능이 좋은 사용자 모델도 목표 사용자가 사용하는 User-Agent를 알고 있는 공격자에 대한 실험에서 Precision이 0.859로 측정되어 이를 더 높이기 위한 연구가 필요하다. 또한 사용한 데이터의 기간이 짧아 브라우저 버전, OS 버전 등의 업데이트 정보를 반영하지 못해 장기간의 데이터를 분석하여 환경정보의 업데이트를 고려하는 추가 연구가 필요하며, 명목형 데이터를 수치 데이터로 바꾸기 위한 방법에 대한 연구와 다양한 학습 알고리즘을 이용한 연구도 필요하다.

References

[1] R. Shay, S. Komanduri, P.G. Kelley, M.L. Mazurek, L. Bauer, and L.F. Cranor, “Encountering stronger password requirements: user attitudes and behaviors,” Proceedings of the Sixth Symposium on Usable Privacy and Security, ACM, Jul. 2010.

[2] A. Das, J. Bonneau, M. Caesar, N.

Borisov, and X. Wang, "The tangled web of password reuse," Proceedings of Network and Distributed System Security Symposium, Feb. 2014.

[3] F. Tari, A. Ozok, and S.H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," Proceedings of the Sixth Symposium on Usable Privacy and Security, ACM, pp. 56-66, Jul. 2006.

[4] E. Maiorana, P. Campisi, N. González-Carballo, and A. Neri, "Keystroke dynamics authentication for mobile phones," Proceedings of the 2011 ACM Symposium on Applied Computing, pp. 21-26, 2011.

[5] T. Feng, X. Zhao, B. Carbunar, and W. Shi, W. "Continuous mobile authentication using virtual key typing Biometrics," 12th International Conference on Trust, Security and Privacy in Computing and Communications, IEEE, pp. 1547-1552. Jul. 2013.

[6] H. Crawford and E. Ahmadzadeh, "Authentication on the go: assessing the effect of movement on mobile device keystroke dynamics," In Thirteenth Symposium on Usable Privacy and Security, USENIX, pp. 163-173. Jul. 2017.

[7] Seungsoo Nam, Changho Seo, and Daeseon Choi, "Mobile finger signature verification robust to skilled forgery," Journal of The Korea Institute of Information Security & Cryptology, 26(5), pp. 1161-1170, Oct. 2016.

[8] L. Zhang, S. Tan, J. Yang, and Y. Chen, "Voicelive: a phoneme localization based liveness detection for voice authentication on smartphones," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 1080-1091. Oct. 2016.

[9] D. Freeman, S. Jain, M. Dürmuth, B. Biggio, and G. Giacinto, "Who are you? a statistical approach to measuring user authenticity," Proceedings of Network and Distributed System Security Symposium, pp. 1-15, Feb. 2016.

[10] N.Z. Gong, M. Payer, R. Moazzezi, and M. Frank, "Forgery-resistant touch-based authentication on mobile devices," Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, pp. 499-510, May. 2016.

[11] Minwoo Kim, Seungyeon Kim, and Taekyoung Kwon, " A study of behavior based authentication using touch dynamics and application usage on android," Journal of The Korea Institute of Information Security & Cryptology, 27(2), Apr. 2017.

[12] D. Liu, B. Dong, X. Gao, and H. Wang, "Exploiting eye tracking for smartphone authentication," International Conference on Applied Cryptography and Network Security, Springer, pp. 457-477, Jun. 2015.

[13] I. Sluganovic, M. Roeschlin, K.B. Rasmussen, and I. Martinovic, "Using reflexive eye movements for fast challenge-response authentication," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 1056-1067, Oct. 2016.

[14] S. Eberz, K.B. Rasmussen, V. Lenders, and I. Martinovic, "Preventing lunchtime attacks: fighting insider threats with eye movement biometrics," Proceedings of Network and Distributed System Security Symposium, Feb. 2015.

[15] D. Preuveneers, and W. Joosen, "SmartAuth: dynamic context fingerprinting for continuous user authentication," Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp. 2185-2191, Apr. 2015.

〔16〕 J.A. Muir, and P.C.V. Oorschot, "Internet geolocation: evasion and counter-evasion," ACM Computing Surveys (CSUR), vol. 42, no. 1, Dec. 2009.

〔17〕 Sohee Park and Daeseon Choi, "Artificial intelligence security issues," Review of The Korea Institute of Information Security & Cryptology, 27(3), pp. 27-32, Jun. 2017.

〈저자소개〉

류 권 상 (Gwonsang Ryu) 학생회원
2016년 2월: 공주대학교 응용수학과 졸업
2016년 3월~현재: 공주대학교 융합과학과 석사과정
〈관심분야〉 인증, 정보보호, 이상거래탐지, 머신러닝

서 창 호 (Changho Seo) 종신회원
1990년 2월: 고려대학교 수학과 학사
1992년 2월: 고려대학교 수학과 석사
1996년 8월: 고려대학교 수학과 박사
1996년 8월~2000년 2월: 한국전자통신연구원 선임연구원, 팀장
2000년 3월~현재: 공주대학교 응용수학과 교수
〈관심분야〉 암호알고리즘, PKI, 무선 인터넷 보안 등

최 대 선 (Daeseon Choi) 종신회원
1995년 2월: 동국대학교 컴퓨터공학과 학사
1997년 2월: 포항공과대학교 컴퓨터공학과 석사
2009년 1월: 한국과학기술원 전산학과 박사
1997년 1월~1999년 6월: 현대정보기술 선임
1999년 7월~2015년 8월: 한국전자통신연구원 인증기술연구실 실장/책임연구원
2015년 9월~현재: 공주대학교 의료정보학과 부교수
2016년 현재: 정보보호학회 이사
〈관심분야〉 인증, 개인정보보호, 이상거래탐지, 의료정보보안, 머신러닝