Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.5, Oct. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.5.1235

패스워드 표기 방식이 패스워드 생성에 미치는 영향*

김 승 연,† 권 태 경‡
연세대학교 정보대학원

A Study of Interpretation Effect of Passwords to Password Generation*

Seung-Yeon Kim,† Taekyoung Kwon‡
Graduate school of Information, Yonsei University

요 약

본 논문은 웹페이지의 로그인 또는 패스워드 변경 인터페이스에서 제공하는 패스워드 표기 방법이 국내 사용자의 패스워드 구성(composition)에 영향을 주는지 설문을 통해 살펴보고, 보안 향상을 위한 표기법을 제안한다. 특히 현재 혼용되고 있는 외국어 '패스워드' 표기와 이를 우리말로 번역한 '비밀번호' 표기는 의미적인 차이가 있다. 국내 S대학교 재학생 200명을 대상으로 설문조사를 통해 '비밀번호' 표기를 사용할 때 더 많은 학생들이 숫자 위주의 패스워드를 만드는 것을 확인하였다. 숫자 위주의 패스워드는 그렇지 않은 경우에 비해 가능한 조합의 수가 크게 감소하므로 이는 보안에 좋지 않은 영향을 줄 우려가 있다. 따라서 본 논문이 국내 사용자들의 패스워드 보안을 향상시킬 방법을 찾는 연구의 참고 자료로 활용될 수 있을 것이라 기대한다.

ABSTRACT

The purpose of this study was to find if the password composition of domestic users is affected by the different form of the word 'Password' in the interface of login or password change. In particular, 'Password', foreign notation, and 'Secret Number', notation translated by Korean, have a semantic difference. According to the survey of 200 students in S university, passwords made under the word 'Secret Number' are heavy on numbers than alphabet. Because these passwords make much smaller composition space than another case, they have bad security impact. We expect to make use of this paper as a base line data for study to find how improve domestic user's password security.

Keywords: Password, Secret Number, Number-oriented password

I. 서 론

1.1 배경

패스워드는 사용자 인증을 위해 사용되는 비밀 문자열을 의미하며 일반적으로 영문자와 숫자, 그리고 경우에 따라 특수문자를 조합하여 구성된다. 영어권 웹사이트에서는 이러한 패스워드를 일률적으로 'Password'로 표기하지만, 국내에서는 '패스워드' 보다는 '암호' 또는 '비밀번호' 표기가 널리 쓰이고 있다(Fig.1.). 그 중 '암호' 표기는 통신에서 제 3자가 통신 내용을 알아볼 수 없도록 하는 기술과 관련이 있는 암호(cryptograph)와 혼동의 우려가 있어서 최근 국내 주요 웹사이트에서 거의 찾아볼 수 없는

접수일(2015년 5월 29일), 수정일(2015년 9월 15일), 게재확정일(2015년 10월 8일)
* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학 ICT연구센터육성 지원사업의 연구결과(IITP-2015- H8501-15-1008)로 수행되었음. 또한, 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2015R1A2A2A01004792)
† 주저자, tribunus000@yonsei.ac.kr
‡ 교신저자, taekyoung@yonsei.ac.kr(Corresponding author)

Fig. 1. Login interface(N portal)

표기이며 '비밀번호' 표기가 훨씬 널리 쓰이고 있다.

1.2 문제제기

국내에서 패스워드의 의미로 널리 사용되는 '비밀번호' 표기는 사실 전자식 현관 자물쇠, 통장 비밀번호처럼 숫자만 입력 가능한 인증 수단(PIN)에 더 어울리는 표기이다. 그럼에도 '패스워드'를 '비밀번호'로 표기하는 이유는 그 역할이 서로 유사하나 후자가 직관적으로 더 이해하기 쉽기 때문인 것으로 짐작된다. 그러나 본 연구에서는 이러한 번역 및 표기가 더 많은 사용자들이 '비밀번호'의 본래 의미인 PIN에 가까운, 숫자 위주의 패스워드를 만들도록 하는 취약점의 원인이 될 가능성이 있다고 보았다. 주진국[5]은 이와 같은 번역된 단어의 의미적 차이에 대해 "*… 흔히 동일한 사물이나 개념을 지시하는 것으로 간주되는 서로 다른 언어의 어휘들은 그 원형적 의미 영역만을 공유할 뿐 그 주변 영역에 있어서는 상당한 격차가 있을 수밖에 없다.*"라고 하였다.

더 많은 사용자가 숫자 위주의 패스워드를 만드는 것이 취약점이 되는 이유는 다음과 같다. 숫자가 k자리 포함된 길이 n자리의 패스워드의 가능한 조합의 수 $P(n,k)$는 특수문자를 고려하지 않을 때 n과 k에 대해 다음과 같은 관계를 가진다.

$$P(n,k)=\binom{n}{k}|M|^k|D|^{(n-k)} \quad (1)$$

식 (1)에서 $|M|$은 집합 $M=\{0,1,\cdots,9\}$의 크기이고, $|D|$는 집합 $D=\{a,b,\cdots,z,A,B,\cdots,Z\}$의 크기이다. 식 (1)을 변형하면 다음과 같다.

$$P(n,k)=\binom{n}{k}\left(\frac{10}{52}\right)^k 52^n \quad (2)$$

n이 고정되어 있을 때, ${}_nC_k$는 $k=0$ 또는 $k=n$이면 최솟값을 갖는다. $(10/52)^k$는 물론 k값에 반비례하므로 식 (2)는 $k=n$, 다시 말해 패스워드가 숫자만으로 구성되었을 때 가능한 조합의 수가 가장 적음을 의미한다. Table 1.은 $P(8,8)=1$이라 할 때 가능한 모든 k에 대해 $P(8,k)$의 상대적인 크기를 보여주고 있다.

Table 1. Relative values of $P(8,k)$

k	$P(8,k)/10^8$
0	534597
1	822457
2	553577
3	212914
4	51181
5	7874
6	757
7	42
8	1

전수 공격 위협에 대해 숫자 위주의 패스워드가 문자 위주의 패스워드에 비해 취약함은 명백하다. 또한 사전 공격 등 보다 진보된 공격에 대해서도 숫자 위주의 패스워드가 특별히 더 안전하다고 보기 어렵다. 예를 들어 특정 사용자의 패스워드를 공격하기 위해 가능한 숫자 후보(예: 생일, 전화번호 등)를 준비하는 것이 단어 후보(예: 이름, 관심사 등)를 준비하는 것에 비해 크게 어려워야 할 이유가 없다.

본 연구는 설문조사에 의한 실험을 통해 '비밀번호' 표기를 사용할 때 숫자 위주의 패스워드가 더 많이 발생한다는 것을 보임으로써 국내 웹사이트에서 널리 사용되고 있는 '비밀번호' 표기의, 좀 더 정확히는 패스워드를 '비밀번호'로 표기하는 것에서 발생할 수 있는 보안 취약성을 보인다. 또한 이를 근거로 국내 웹사이트들이 외래어 표기인 '패스워드'를 사용하거나 더 직관적으로 이해하기 쉬우면서도 안전한 표기 방법을 적극적으로 도입할 것을 권고한다.

1.3 논문의 구성

본 논문은 5장으로 구성되어 있다. 2장에서 본 연구의 질문과 가설, 가설검정을 위한 데이터 수집 절차를 설명한다. 3장에서는 수집된 데이터를 처리하고 분석하는 절차와 분석 결과를 기술하고, 4장에서 관련 연구를 소개한 뒤, 마지막으로 5장에서 본 논문의 결론을 제시한다.

II. 연구 설계 및 실험

2.1 연구 질문

본 연구의 질문은 *"패스워드의 표기 방식에 따라 사용자들이 생성하는 패스워드 구성에 차이가 있는가?"* 이다. 패스워드의 표기 방식은 'Password', '패스워드', '비밀번호', '암호' 등 다양한 방식이 있으나, 'Password'와 '패스워드'는 의미적으로 차이가 없는 표기 방식이므로 '패스워드'로 통합하고, '암호'는 국내 주요 웹사이트에서 최근 찾아보기 어려운 표현이므로 이는 본 연구에서 다루지 않는다. 즉, 본 연구는 '패스워드' 표기와 '비밀번호' 표기에서 사용자들이 생성하는 패스워드 구성에 차이가 있는지를 알아볼 것이다.

2.2 연구 가설

'패스워드' 표기보다 '비밀번호' 표기에서 더 많은 사용자들이 숫자 위주의 패스워드를 만든다는 것이 본 연구의 가설이다. 이 때 숫자 위주의 패스워드란 무엇인지 정의해야만 이러한 가설을 검증할 수 있다. 단순히 숫자가 문자보다 많은 경우(H1)를 숫자 위주의 패스워드로 생각할 수 있으나, 이는 보편타당한 관점이 아닐 수 있다. 따라서 숫자가 문자보다 2배 이상 많은, 다시 말해 숫자 위주의 패스워드가 될 기준이 훨씬 엄격한 경우(H2)에 대해서도 가설검정을 수행하였다. 아래 가설에서 $|N|$은 패스워드에서 숫자의 개수를, $|C|$는 패스워드에서 문자의 개수를, $|S|$는 특수문자의 개수를 의미한다.

- H1: '비밀번호' 표기에서 $|N| > |C|$인 패스워드가 더 많이 발생한다.
- H2: '비밀번호' 표기에서 $|N| > 2|C|$인 패스워드가 더 많이 발생한다.

그러나 가설 1과 가설 2에서는 숫자 위주의 패스워드를 판정함에 있어서 특수문자를 고려하지 않았다. 특수문자는 분류하기가 애매하나, 적어도 이것을 숫자로 생각하기는 어렵기 때문에 숫자보다는 문자에 가깝다고 간주하고 다음의 가설을 세웠다.

- H3: '비밀번호' 표기에서 $|N| > |C|+|S|$인 패스워드가 더 많이 발생한다.

가설 1과 가설 2는 반대로 말하면 '패스워드' 표기에서 문자 위주의 패스워드가 더 많이 발생한다는 해석이 가능하나, 가설 3은 그와 같은 해석이 어렵다. 따라서 가설 3에 대응되는 '패스워드' 표기 측면 가설인 가설 4를 수립하였다.

- H4: '패스워드' 표기에서 $|C| \geq |N|+|S|$인 패스워드가 더 많이 발생한다.

단, 특수문자가 숫자보다는 문자에 가깝다고 간주하였으므로 패스워드의 주축이 될 조건을 가설 3보다 약간 약하게 설정하였다.

2.3 연구 방법

본 연구는 종이 설문지를 사용한 대면 설문조사로 수집된 데이터를 기반으로 진행된다. 대면 설문은 온라인 설문과 달리 설문에서 악성 소프트웨어로 참가자의 ID와 같은 계정 정보를 수집할 수 있는 가능성이 전혀 없고, 참가자 또한 그 사실을 직관적으로 이해할 수 있으므로 패스워드에 관한 민감한 질문에도 비교적 진실한 답변을 할 것으로 기대하여[9] 종이 설문지를 통한 대면 설문을 수행하였다.

2.3.1 설문지 구성

설문 내용은 동일하면서 패스워드 표기 방식만 '비밀번호'와 '패스워드'로 각각 다른 두 종류의 설문지를 만들었다. 설문 문항은 논문에 설문 문항이 포함된 기존 연구들[3][9]을 참고하여 개발되었다. 설문 내용은 다음과 같다. 먼저 응답자가 학사정보시스템 계정의 패스워드를 변경하는 상황을 가정하여 새로운 패스워드를 가상으로 설정하게 한다. 그리고 성별을 포함한 인구통계 정보와, 패스워드의 전체 길이, 사용된 문자(a~z, A~Z)의 길이와 의미, 숫자의 길이와 의미 정보를 수집한다. 그리고 패스워드의 재사용 여부와 사용된 구체적인 수정 규칙, 그리고 일반적인 계정관리에 관한 몇 개의 문항을 추가하였다.

각 설문지에는 현실적인 패스워드의 데이터를 얻기 위해[10] 실제 패스워드 변경화면(Fig.2.)의 패스워드 표기를 '비밀번호' 또는 '패스워드'로 완전히 통일시킨 스크린 샷을 첨부하였다. Fig.3.은 Fig.2.와 같은 실제 패스워드 변경화면에서 패스워드 표기를 '비밀번호'로 통일한 결과이다.

설문 응답 시 절대로 현재 패스워드를 기준으로 답하지 말 것을 강조했는데, 이는 설문 응답자의 정보 유출을 방지함과 동시에 '비밀번호' 표기와 '패스워드' 표기의 차이에 따른 패스워드 생성 패턴의 차이를 관찰하기를 원했기 때문이었다. 구체적인 수정 규칙을 묻는 것과 같은 다소 민감한 문항에는 '대답하고 싶지 않습니다' 선택을 포함시켰다[9].

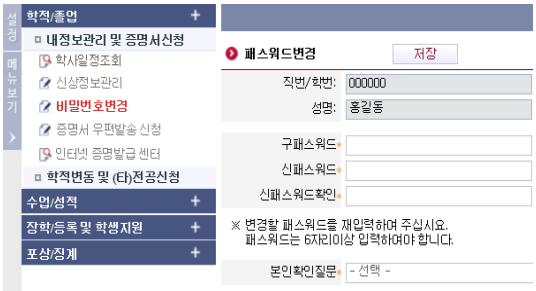


Fig. 2. *S* university's password change interface

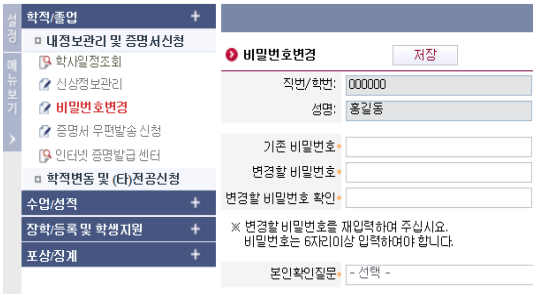


Fig.3. Modified password change interface ('Secret Number' version)

2.3.2 설문 대상

국내 S대학교 재학생 200명[6]을 대상으로 설문조사를 계획했다. S대학교의 학사정보시스템은 로그인 화면에서 '비밀번호' 표기를 사용하고 있고, 패스워드 변경화면에서는 Fig. 2.와 같이 '비밀번호' 표기와 '패스워드' 표기를 혼용하고 있다.

따라서 기존 패스워드가 '비밀번호' 표기나 '패스워드' 표기 어느 한쪽에만 크게 영향을 받았다고 보기 어렵다. 영향을 받았다 하더라도 패스워드를 변경하는 패널에서 '패스워드' 표기를 사용하고 있으므로 '패스워드' 표기의 영향을 더 많이 받았을 것이다. 그러므로 많은 사용자들이 기존 패스워드를 재사용한다는 연구결과[3][9][10]들을 고려하면 S대학교 재학생들은 패스워드를 변경하는 가상 시나리오를 주더라도 여전히 ('패스워드' 표기에 영향을 받았을 가능성이 높은) 기존 패스워드와 유사한 패스워드를 만들 가능성이 높다. 그럼에도 '비밀번호' 표기를 사용하는 설문지의 응답에서 숫자 위주의 패스워드가 유의하게 많이 발견된다면, 이는 본 연구 가설을 뒷받침하는 근거가 될 것이다.

2.3.3 설문조사

'비밀번호' 표기 설문지와 '패스워드' 표기 설문지를 각각 100부씩 준비하여 무작위로 섞은 다음, S대학교 캠퍼스에서 학생들에게 설문에 참여해 줄 것을 요청하였다. 수락하면 어떤 설문인지를 설명한 후 설문 응답을 받고, 소정의 답례를 제공하였다.

III. 실 험

3.1 인구통계

Table 2.과 Table 3.에 도시된 것처럼, '비밀번호' 또는 '패스워드' 표기 설문지를 받은 학생들은 성별과 단과대학 면에서 유사한 분포를 보였다.

설문 응답자 중에는 IT 관련학과(컴퓨터공학과, 정보보호학과 등)이 가장 많았는데, 이는 비교적 보안의식이 있으리라고 예상되는 IT분야 전공 학생[9]들을 다수 포함시킴으로써 본 연구의 신뢰성을 높이기 위해 해당 단과대학 학생들이 많은 건물 근처에서 집중조사를 잠시 진행하였기 때문이다.

Table 2. The distribution of Gender

Gender	'Secret Number'	'Password'
Male	61	64
Female	39	36
Total	100	100

3.2 데이터 처리

가설검정에 앞서, 수집된 응답 중 S대학교 패스워드 규칙에 맞지 않거나 관련 문항의 응답과 모순되는 등 신뢰성이 떨어지는 응답을 무시하거나 사용 가능한 응답을 적절하게 포함하였다. 데이터 처리는 다음의 순서로 진행되었다.

• $|C|+|M|>|L|$ 응답 무시: 문자의 개수($|C|$)와 숫자의 개수($|M|$)의 합이 전체 길이($|L|$)를 넘는 응답을 제외하였다('비밀번호' 표기 응답 7개, '패스워드' 표기 응답 7개).

• $|L|<6$인 응답 무시: S대학교 학사정보시스템의 패스워드는 최소 6글자 이상이어야 한다. 그럼에도 4글자 패스워드를 만든 응답자의 데이터를 제외하였다('비밀번호' 표기 응답 1개).

• 재사용 여부 불일치 응답 무시: 설문조사에서는 가상으로 패스워드를 변경하면서 기존 패스워드를 재사용했는지 여부를 조사하고, 재사용한 경우 구체적인 변환 규칙을 선택 또는 서술하도록 하였다. 앞의 질문에서 완전히 새로운 패스워드를 만들었다고 응답했으나, 다음 질문에서는 변환 없이 그대로 재사용했다고 답한 응답을 제외하였다('비밀번호' 표기 응답 3개, '패스워드' 표기 응답 3개).

• '?' 응답 포함: $|L|=12$, $|C|=7$, $|M|=$'?'로 응답한 응답은 '?'에 0부터 5까지 모든 값을 대입해도 가설 1~4에 모두 적용 가능하므로 포함시켰다.

• '~' 응답 포함: $|L|=12$, $|C|=3\sim4$, $|M|=8\sim9$로 답한 응답은 $|C|=4$이고 $|M|=9$인 경우를 제외($\because |C|+|M|>|L|$)한 모든 조합이 가설 1~4에 모두 적용 가능하므로 포함시켰다.

데이터 처리 결과, '비밀번호' 표기에서 89개, '패스워드' 표기에서 90개 응답이 유효했다.

3.3 가설검정

Table 4.는 2.2절에서 제시한 가설을 검정하기 위해 '비밀번호' 표기와 '패스워드' 표기의 응답 중 각 가설의 조건을 만족하는 응답의 수(2, 3열)와 가설 검정 결과를 보여준다[7]. 6행은 3.2절에서 얻은 유효 데이터를 의미한다. 단측 t-test 결과 가설 1과 가설 2가 유의한 것으로 나타났다($p \leq 0.05$). 즉 '비밀번호' 표기에서 단순히 문자보다 숫자가 많은 패스워드는 물론이고 문자 개수의 2배보다도 숫자의 개수가 많은, 다시 말해 명백히 숫자 위주인 패스워드도 유의하게 더 많이 발견할 수 있었다. 특수문자까지 고려한 가설 3($p=0.067$)과 가설 4($p=0.056$)는 유의하지 않다고 나타났다.

실험 결과에 대한 해석은 다음과 같다. 특수문자 개수를 정확하게 확인 할 수 있는 데이터(유효 데이터 중 '?' 응답과 '~' 응답을 제외한 177개 데이터)를 기준으로 패스워드에 특수문자를 포함시킨 경우는 '비밀번호' 표기 설문지와 '패스워드' 표기 설문지 각각 23%, 26%였다. 이는 90%가 패스워드에 문자와 숫자를 함께 포함시킨 것에 비해 매우 낮은 비율이다. 특수문자를 포함시킨 경우 중에서도 각각 75% 이상이 1~2자였으며 두 종류의 설문지 모두 1자가 가장 많았다. 즉, 패스워드에 특수문자를 포함시키는 경우는 적었고 특수문자를 3자 이상 쓴 경우는 그 중에서도 적었다.

이러한 점들로 미루어 볼 때 특수문자가 패스워드의 주축이 되는 경우가 매우 희박함을 알 수 있다. 또한 패스워드에 숫자 또는 특수문자를 덧붙이는 것은 사용자들이 가장 선호하는 패스워드 수정 방법이며, 특수문자가 패스워드의 뒤쪽 2자리에서 가장 많이 발견된다는 기존 연구결과[3][9]를 참고한다면 특수문자는 (숫자 위주 또는 문자 위주 중 하나로) 완성된 패스워드의 보안을 증가시키기 위한 추가적인 문자의 개념으로 보아야 한다. 그렇다 하더라도 패스워드의 표기 방법에 따라 이러한 추가적인 문자(특수문자)의 사용 형태에 차이가 생길 가능성이 있으나, 양측 t-test 결과 177명의 특수문자의 평균 길이와 사용 여부 비율 각각에 대해 표기법에 따른 유의한 차이가 없다고 나타났다[7]. 따라서 가설 1과 가설 2의 결과를 통해 '비밀번호' 표기에서 숫자 위주의 패스워드가 더 많이 발견된다고 할 수 있다.

Table 3. The distribution of College

College	'SN'	'PW'
Liberal Arts	2	0
Social Science	3	4
Business Administration	9	14
Hospitality&Tourism Management	10	10
Natural Sciences	20	21
Life Science	4	4
Engineering	22	16
Electronics & Information	29	25
Arts & Physical Education	1	5
Other	0	1
Total	100	100

Table 4. Results of hypothesis test

Hypo-thesis	'SN'	'PW'	t	p-value	Result
H1	37	26	1.78	0.038	accept
H2	28	17	1.94	0.026	accept
H3	33	24	1.50	0.067	reject
H4	48	59	1.59	0.056	reject
Valid data	89	90			

3.4 계정 관리 실태

설문조사에는 가설을 검증하기 위한 문항 외에도, 국내 사용자의 계정 관리에 관한 기초자료를 수집하기 위한 문항이 포함되어 있었다. 다음은 그러한 문항에서 드러난 몇몇 특징에 관한 설명이다. 3.2절에서 얻은 유효한 응답뿐 아니라 200개의 모든 응답을 대상으로 조사하였다. 계정 관리에 관한 문항은 복잡한 실험 문항에 비해 참가자의 일반적인 경향을 조사하는 비교적 간단한 문항이므로 복잡한 실험 문항에서는 부적절한 응답을 했더라도 간단한 경향 조사 문항에서는 진실에 가까운 응답을 했을 가능성이 높다고 판단하였기 때문이다.

- 패스워드의 문자 의미: 가상 패스워드에 문자를 포함시켰다고 응답한 198명 중, 69%(137명)가 영어단어(96명) 또는 한글단어(40명), 또는 둘 다(1명) 포함시켰고, 단어의 의미를 조사하는 복수응답이 가능한 문항에서 137명 중 41%(56명)가 본인의 이름(56명)을, 39%(54명)가 본인과 직접적인 관련이 없는 의미를 포함시켰다고 답했다.

- 패스워드의 숫자 의미: 가상 패스워드에 숫자를 포함시켰다고 응답한 197명 중, 숫자의 의미를 조사하는 복수응답이 가능한 문항에서 64%(127명)이 본인과 직접적인 관련이 없는 의미를 포함시켰다고 답했다. 16%(32명)는 생일 등 기념일을 패스워드에 포함시켰다고 답했고, 10%(20명)가 전화번호 관련 숫자를 패스워드에 포함시켰다고 답했다.

- 패스워드 재사용: 응답한 183명 중 78%(143명)가 가상으로 패스워드를 만들면서 기존에 사용하고 있는 패스워드를 재사용했다고 답했다. 또한 중복 응답을 허용한 재사용 수정 방법 문항에서 이 143명 중 50%(72명)가 수정 없이 그대로 재사용했다고 답했고(단, 이 답변만 복수응답 불허), 앞 또는 뒤에 숫자, 특수문자를 덧붙였다고 응답한 참가자는 각각 15%(22명), 17%(25명)였다.

- 패스워드 관리방법: 중복 응답을 허용한 패스워드 관리 방법 문항에서, 응답자 198명 중 15%(30명)만 패스워드를 기록해둔다고 응답하였다. 자동로그인 기능을 사용한다고 응답한 참가자는 14%(27명)였다.

- 패스워드 매니저: 패스워드 매니저에 관한 문항에 응답한 참가자 198명 중 87%(172명)가 패스워드 매니저가 무엇인지 모른다고 답했다. 패스워드 매니저를 사용하고 있는 참가자는 6%(12명)에 불과했다.

- 오프라인 패스워드: 문서, 압축파일, 또는 PC계정 등에 설정하는 암호는 Table 5.에 도시된 것처럼 웹사이트에서 자주 쓰는 패스워드와 비교하여 대체로 비슷하거나 더 간단한 경향이 있었다.

Table 5. Off-line password complexity

Complexity	'SN'	'PW'
Much more complexity	5	4
More complexity	8	10
Similar	45	43
Simpler	17	10
Much simpler	8	13
Not use off-line password	16	19
Total	99	99

IV. 관련 연구

최근(2015년 7월) SOUPS에 발표된 논문에서 Ur 등은 49명의 사용자를 대상으로 패스워드에 관한 45분~1시간 규모의 심층 면접을 진행하여 사용자들이 안전한 패스워드에 대해 잘못된 관념(misconception)을 가지고 있음을 보였다[12]. 이 논문에서는 단어 및 어구 선택에 관한 내용을 한 단락으로, 그리고 숫자와 특수문자를 묶어서 한 단락으로 기술하였고 또한 사용자들이 숫자와 특수문자를 추가함으로써 패스워드가 더 안전해졌다고 믿는 경향을 관찰했다고 하였다. 이로부터 해당 연구의 연구자들과 참가자들은 대체로 패스워드를 (직역한 의미에 걸

맞게) 단어 중심으로 구성함과 동시에 숫자를 특수문자와 함께 보안을 위한 추가 문자로 여김을 짐작할 수 있다. 이는 사전적으로 단어를 의미하는 '패스워드' 표기에 영향을 받은 결과로 짐작된다.

사용자들의 일반적인 패스워드 관리 방법에 관한 다양한 연구들이 있다. Shay 등은 약 8천명을 대상으로 실시한 온라인 설문조사를 통해 복잡한 조건의 패스워드 정책보다 조금 더 간단한 조건으로 길이가 긴 패스워드를 만드는 정책이 더 나은 보안 강도와 사용성(usability)을 가질 수 있음을 보였다[8]. Just와 Aspinall은 사용자들이 패스워드를 잊었을 때 보조 인증 수단으로 사용하는 패스워드 질문(challenge question)을 사용자가 직접 작성하도록 한 경우에도 낮은 엔트로피의 정답을 가지는 질문을 생성하는 경향이 있고, 동시에 공격자가 그 정답을 추측하기 어려울 것으로 믿는 것을 보였다[4].

실제 사용자들의 패스워드 리스트를 통해 패스워드 추측 효율을 향상시키는 연구도 활발히 진행되고 있다. Das 등은 사용자들이 비슷하거나 같은 패스워드를 서로 다른 사이트에 재사용하는 것을 고려하여 한 사이트의 유출된 패스워드 리스트를 바탕으로 다른 사이트의 패스워드를 추측하는 실험을 하였고 단순한 알고리즘으로도 추측이 훨씬 더 쉬워짐을 보였다[3]. Veras 등은 RockYou 유출 패스워드 리스트를 기반으로 사용자들이 패스워드에 자주 사용하는 영어 단어와 함께 일반 영문법이 아닌 패스워드 특유의 문법을 가지고 패스워드의 의미까지 고려한 패스워드 추측 알고리즘을 개발하였다. 알고리즘을 바탕으로 LinkedIn과 MySpace 유출 패스워드를 추측하는 실험을 수행하였고, 그 결과 최신 기법보다 추측 성공률이 크게 높아지는 것을 보였다[11].

패스워드 강도 지표에 관한 연구는 실제 패스워드 목록을 바탕으로 좀 더 현실적인 패스워드 강도 판정 방법을 제안하고 있다. Carnavalet과 Mannan은 유출된 패스워드 목록을 기반으로 11개 사이트의 패스워드 미터를 분석한 결과, 사이트마다 패스워드 강도 판정이 크게 달라서 사용자가 패스워드 선택 시 혼란을 느낄 수 있다고 주장하였다[2]. Bonneau는 Yahoo!의 협조를 얻어 7천만 사용자들의 실제 패스워드와 개인정보를 수집하였고, 이러한 실제 패스워드 분포를 가지고 새로운 패스워드 강도 지표를 만들어냈다[1].

V. 결 론

본 연구는 웹페이지의 로그인 또는 패스워드 변경 인터페이스에서 패스워드를 '비밀번호'로 표기할 경우 '패스워드'로 표기한 경우에 비해 숫자 위주의 패스워드가 더 빈번히 발생함을 보였다. 숫자 위주의 패스워드는 그렇지 않은 경우에 비해 명백하게 전수 공격에 취약하므로 국내 웹사이트들은 '비밀번호' 표기를 사용하는 것을 지양하고, '패스워드' 표기나, 더 직관적으로 이해하기 쉬우면서도 보안에 악영향을 줄 우려가 적은 표기 방법을 도입해야 할 것이다.

본 연구는 S대학교 재학생 200명을 대상으로 한 설문조사 결과를 바탕으로 진행하였으므로 설문 대상이 비교적 컴퓨터와 친숙한 세대이며 어느 정도 교육을 받았다는 점, 그리고 의도적으로 IT 전공자를 다수 포함시켰다는 점에서 주목할 만하다. 또한 본 연구는 국내 사용자의 일반적인 계정 관리 조사에 참고가 될 수 있는 몇몇 특징을 제공한다. 예를 들어, 사용자들은 본인과 관련된 여러 가지 단어 중에서도 특히 이름을 선호하였다. 같은 이름이라도 다양한 방법으로 표현 가능하므로 어떤 방식의 표현 방법이 많이 사용되는지에 관한 후속 연구가 가능할 것이다.

본 연구의 한계는 종이 설문지를 이용한 직접 설문의 특성상 매우 큰 규모의 설문이 어려웠다는 점이다. 본 연구보다 복잡한 연구 모델을 통해 사용자의 패스워드 선택에 따른 보안효과에 관한 연구를 진행한 기존 연구[6]에서도 본 연구와 비슷한 규모 및 형태의 표본을 사용하였으므로 연구 모델이 비교적 단순한 본 연구는 충분히 유의미하다고 생각되나, 온라인 설문을 이용하여 매우 큰 규모의 설문을 통한 후속 연구도 가능할 것이다.

본 연구의 결과는 해외, 특히 영어권 국가의 연구에서는 찾아보기 어려운 사항이다. 이러한 점에서 본 연구가 국내 사용자의 특성에 맞춘 보안 정책 및 기술 개발의 참고 자료로 활용될 수 있을 것이다.

References

[1] J. Bonneau, "The science of guessing: analyzing an anonymized corpus of 70 million passwords," Proceedings of the 33th IEEE Symposium on Security and Privacy, pp. 538-552, May. 2012.

[2] X.de.C.de Carnavalet, and M. Mannan,

"From very weak to very strong: Analyzing password-strength meters," Proceedings of the Network and Distributed System Security Symposium, Feb. 2014.

〔3〕 A. Das and J. Bonneau, "The tangled web of password reuse," Proceedings of the Network and Distributed System Security Symposium, Feb. 2014.

〔4〕 M. Just and D. Aspinall, "Personal choice and challenge questions: a security and usability assessment," Proceedings of the 5th Symposium on Usable Privacy and Security, pp. 8, July. 2009.

〔5〕 Jinkook Joo, "A Contrastive Semantic Analysis of English and Korean News Terminology," The Journal of translation studies, 12(3), pp. 263-279, Sep. 2011.

〔6〕 Jongki Kim and Dayeon Kang, "A Study on the Factors Affecting the Information Systems Security Effectiveness of Password," Asia Pacific Journal of Information Systems, 18(4), pp. 1-26, Dec. 2008.

〔7〕 Wonwoo Lee, Statistics clearly written, Pakyoungsa, pp. 157-172, Sep. 2009.

〔8〕 R. Shay, S. Komanduri, A.L. Durity, P. Huh, M.L. Mazurek, S.M. Segreti, B. Ur, L. Bauer, N. Christin, and F. Cranor "Can long passwords be secure and usable?," Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pp. 2927-2936, Apr. 2014.

〔9〕 R. Shay, S. Komanduri, P.G. Kelley, P.G. Leon, M.L. Mazurek, L. Bauer, N. Christin, and L.F. Cranor "Encountering stronger password requirements: user attitudes and behaviors," Proceedings of the Sixth Symposium on Usable Privacy and Security, pp. 2:1-2:20, July. 2010.

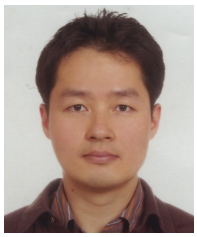
〔10〕 E. Stobert and R. Biddle, "The password life cycle: user behaviour in managing passwords," Proceedings of the Symposium on Usable Privacy and Security, pp. 243-255, July. 2014.

〔11〕 R. Veras, C. Collins, and J. Thorpe, "On the semantic patterns of passwords and their security impact," Proceedings of Network and Distributed System Security Symposium, Feb. 2014.

〔12〕 B. Ur, F. Noma, J. Bees, S.M. Seqreti, R. Shay, L. Bauer, N. Christin, and L.F. Cranor, ""I Added '!' at the End to Make It Secure":Observing Password Creation in the Lab," Proceeding of the 11th Annual Symposium on Usable Privacy and Security, pp. 123-140, July. 2015.

〈저자소개〉

김 승 연 (Seung-Yeon Kim) 학생회원
2015년 2월: 세종대학교 응용통계학, 컴퓨터공학 학사 (자연과학대학 수석졸업)
2015년 3월~현재: 연세대학교 정보대학원 석박통합과정
〈관심분야〉 Usable Security, Social Engineering



권 태 경 (Taekyoung Kwon) 종신회원
1992년 2월: 연세대학교 컴퓨터과학과 학사
1995년 2월: 연세대학교 컴퓨터과학과 석사
1999년 8월: 연세대학교 컴퓨터과학과 박사
1999년~2000년: U.C. Berkely Post-Doc.
2001년~2013년 8월: 세종대학교 컴퓨터공학과 교수
2007년~2008년: Univ. Maryland at College Park 교환교수
2013년 9월~현재: 연세대학교 정보대학원 부교수
〈관심분야〉 암호 프로토콜, 네트워크 프로토콜, IoT 보안, Usable Security, HCI 등