Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.26, NO.4, Aug. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.4.961

국내 사용자의 패스워드 사용 현황 분석*

김 승 연,† 권 태 경‡
연세대학교 정보대학원

A Case Study of Password Usage for Domestic Users*

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

요 약

패스워드 기반 인증 기법의 안전성 강화를 위해서는 충분한 길이와 무작위 문자열 성질을 갖는 강한 패스워드의 선택과 관리가 필요하다. 하지만 이미 알려진 바와 같이 많은 사용자들은 기억하기 쉬운 약한 패스워드를 선택하는 경향이 있으며 이에 대한 관리 또한 취약하다. 본 논문에서는 국내 사용자들이 어떠한 경향으로 패스워드를 선택하며 관리하는지에 대해 연구하였다. 직접 327명 규모의 설문조사를 진행하여, 패스워드 생성과 갱신에 관한 경향을 분석하였으며 또한 패스워드의 구조와 계정관리에 관한 경향을 분석하였다. 마지막으로 서버의 패스워드 생성 규칙이 사용자 패스워드 구성에 주는 영향에 대해 분석하였다. 분석 결과 사용자들이 유의하게 선호하는 패스워드 구조, 특수문자가 있는 반면 서버의 패스워드 생성 규칙이 미치는 영향은 유의하지 않음을 발견하였다.

ABSTRACT

For securing password-based authentication, a user must select and manage a strong password that has sufficient length and randomness. Unfortunately, however, it is known that many users are likely to choose easy-to-remember weak passwords and very poorly manage them. In this paper, we study a domestic user case of password selection and management. We conducted a survey on 327 domestic users and analyzed their tendency on password creation and update strategies, and also on the password structure and account management. We then analyzed an effect of a server's password creation rule on a structure of a user-chosen password. Our findings include that there are password structures and special characters that users significantly prefer while the effect of server's password creation rule is insignificant.

Keywords: Password, authentication

I. 서 론

패스워드 인증은 컴퓨팅 환경에서 가장 널리 이용되어온 사용자 인증 기법이며 안전성 강화를 위해 안전한 패스워드의 선택과 관리를 요구한다. 안전한 패스워드란 충분한 길이와 무작위 문자열로 구성된 강한 패스워드를 의미하지만 실제로 사용자가 선택하는 패스워드는 대부분 기억하기 쉬운, 즉 안전성이 약한 패스워드이다[4][8][12][13]. 이와 같이 약한 패스워드는 그 길이가 짧거나 사용자들이 선호하는 문자열을 포함하는 경향이 있으며, 따라서 패스워드에 대한 추측 공격에 노출되는 매우 취약한 문제가 있다. 국내 사용자들 또한 이러한 문제에서 자유로울 수 없으며 본 논문에서는 국내 사용자들의 패스워드 보안 향상을 위해 패스워드 사용 현황 분석을 수행하였다.

Received(03. 18. 2016), Modified(1st: 07. 27. 2016, 2nd:08. 05. 2016), Accepted(08. 05. 2016)
* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학 ICT연구센터육성 지원사업의 연구결과(IITP- 2016-R2718-16-0003)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)

중국어 사용자와 영어 사용자의 패스워드를 비교한 연구[9]와 스페인어 사용자와 영어 사용자의 패스워드를 비교한 연구[1]의 결과에서 관찰할 수 있는 사용 언어별 패스워드 경향의 차이를 고려한다면, 본 연구와 같이 국내 사용자의 패스워드 경향을 분석하는 연구는 국내 사용자의 패스워드 보안을 위한 정책 수립, 기술 개발에서 최적의 결과를 이끌어내기 위해 반드시 선행되어야만 할 것이다.

본 논문은 국내 사용자들의 패스워드 선택과 관리, 그리고 서버의 패스워드 생성정책이 주는 영향에 대한 설문 조사 및 분석 연구를 진행한다. 관련 선행연구를 간략히 살펴보면, 패스워드 규칙 변경의 영향을 조사하는 설문을 진행하면서 패스워드에 포함된 숫자의 의미를 함께 조사한 연구와[12] 대규모 유출 패스워드를 단어 사전 기반으로 분석함으로써 추측 공격의 효율을 크게 높인 연구[15] 등의 선행 연구들이 있다. 그러나 전자는 숫자의 의미를 조사하는 문항이 설문 전체의 목적과 큰 관련이 없는 비교적 중요도가 낮은 문항이었고, 후자는 사용자가 패스워드를 만들 때 어떠한 의미를 부여했는지 확인할 방법이 없었다는 한계가 있다. 이러한 한계로 인해 사용자에게 개인적으로 의미 있는 문자를 무작위 문자로 취급함으로써 패스워드의 추측 난이도를 과대평가할 우려가 있다. 따라서 본 연구는 국내 사용자를 대상으로 하는 설문조사 과정에서 패스워드의 구성적, 의미적 구조를 조사하는 문항을 가장 중요하게 취급함과 동시에 사용자가 그의 패스워드의 구성적, 의미적 구조를 직접 그리게 함으로써 앞서 언급한 문제들을 보완하였다. 또한 패스워드 생성 전략 및 관리 방법에 대해 조사하여 국내 패스워드 사용 현황을 다각적으로 분석하였다.

II. 기존문헌 연구

패스워드에 관한 연구는 1979년 Morris와 Thompson의 연구[10]에서부터 시작되었다. 그들은 3천개 이상의 패스워드를 수집했고 그중 86%가 취약한 패스워드임을 보였다. 여기서 취약하다는 것은 너무 짧거나, 단일 구성 청크로 구성되어 있거나, 또는 사전 등에서 발견할 수 있는 것을 포함한다. 이 연구가 수행되고 30년 이상 경과했으나 문자 기반 패스워드의 취약성에 관한 연구는 여전히 진행되고 있다. 다음은 최근에 발표된 관련 연구들이다.

2.1 패스워드 관련 설문 연구

2010년 Shay 등은 CMU (Carnegie Mellon University)의 구성원 470명을 대상으로 패스워드 관련 설문을 진행하였다[12]. CMU는 설문조사 시점을 기준으로 최근에 패스워드 정책을 엄격하게 변경하였다. 연구자들은 이 시기에 설문을 진행하여 같은 사용자가 같은 웹 사이트의 다른 패스워드 정책에 갖는 느낌에 관한 비교적 신뢰성 높은 데이터를 수집할 수 있었다. 설문 참가자들은 대체로 강화된 정책이 불편하다고 느꼈지만 동시에 자신의 계정이 더 안전해졌다고 생각하는 모습을 보였다. Das 등은 패스워드 생성 과정에서의 사용자의 생각을 이해하기 위해 224명 규모의 설문을 진행하였다[4]. 그 결과 새 전자 메일 계정의 패스워드를 설정할 때 기존 패스워드를 재사용한다는 비율이 77%에 달했으며 패스워드를 수정할 때 패스워드 앞뒤에 숫자나 특수문자를 덧붙이거나 일부 소문자를 대문자로 바꾸는 것이 가장 일반적인 수정 방법이라는 것 등을 보였다.

2.2 패스워드 관련 질적 연구

Stobert 등은 27명 규모의 인터뷰 연구를 통해 컴퓨터 관련 전공자가 아닌 사람들의 패스워드 경향에 관한 데이터를 수집했고 이를 근거 이론으로 분석하여 패스워드의 생성부터 소멸까지의 주기를 단계별로 정리하였다[13]. Ur 등은 49명의 참가자를 대상으로 금융, 이메일, 뉴스 사이트에 대한 패스워드를 가상으로 설정하는 실험을 진행하였다[14]. 그 후 인터뷰를 진행한 결과 참가자들 중 일부는 패스워드에 철자가 어려운 단어를 사용하는 것이 철자가 쉬운 단어를 사용하는 것에 비해 더 안전하다고 믿는 등 패스워드에 관한 잘못된 관념을 가지고 있었다. 연구자들은 이러한 잘못된 관념이 일부 사용자가 약한 패스워드를 만들도록 하는 것을 발견하였다. Rinn 등은 20명의 저학력자를 대상으로 인터뷰 연구를 진행하였다[11]. 그 결과 고학력자들이 패스워드 보안 수칙을 잘 실천하지 않을 뿐 올바르게 인식하고 있는 반면, 실험에 참가한 저학력자들은 패스워드에 관해 잘못된 관념을 가지고 있음을 보였다.

2.3 패스워드 관련 양적 연구

2.1절에서 언급한 연구와 본 절에서 언급하는 연

구의 차이점은 본 절의 연구가 보다 대규모의 사용자를 상대로 하였으며 연구 방법으로 설문조사를 사용하지 않았다는 점이다. 2007년 Floêncio와 Herley는 Windows Live Toolbar 클라이언트에 사용자의 패스워드 사용에 관한 데이터를 수집하는 컴포넌트를 삽입하여 약 50만명 규모의 데이터를 수집하였다[5]. 데이터 수집은 사용자의 동의하에 수행되었으며 사용자의 식별 데이터나 실제 패스워드, IP 주소 등 민감한 데이터는 수집하지 않거나 분석의 용도로만 사용 후 폐기되었다. 이 연구는 (저자들이 밝힌 바에 의하면)최초의 대규모 사용자 집단을 대상으로 한 패스워드 사용 경향 분석 연구이며 그 결과를 통해 기존의 소규모 사용자 집단을 대상으로 한 연구에 의해서만 확인할 수 있었던 패스워드 사용 경향을 재확인할 수 있었다. Bonneau는 2015년에 Yahoo!의 협조를 얻어 사용자 7천만명의 개인정보 및 패스워드를 수집하였고 이를 통해 새로운 패스워드 강도 지표를 제안하였다[2]. Chatterjee 등은 MTurk를 통해 약 4천명의 패스워드 입력 오류에 관한 데이터를 수집하여 분석하였고 이를 통해 이론적으로 패스워드 보안을 유의하게 낮추지 않는 입력 오류 허용 프레임워크를 제안하였다[3].

2.4 유출 패스워드 분석 연구

2.1절에서 언급한 Das 등은 설문 및 유출 패스워드 분석을 통해 확인된 사용자의 패스워드 재사용 경향을 패스워드 추측 공격에 적용하였다[4]. 공격자가 유출 패스워드를 가지고 있다면 그렇지 않은 경우에 비해 추측 효율이 크게 증가함을 보였다. Veras 등은 유출 패스워드 리스트, 사전, 그리고 추가 단어 리스트를 사용하여 패스워드에 사용된 단어를 의미에 따라 분류하였다[15]. 이를 통해 일반적인 영문법이 아닌 패스워드 특유의 문법 구조를 구성하는 알고리즘을 개발했고, LinkedIn과 MySpace에서 유출된 패스워드에 대해 추측 공격을 수행하는 실험을 진행하였다. 그 결과 Veras 등이 최신 기법으로 지목한 Weir의 기법보다 추측 효율이 훨씬 높음을 보였다.

최다 인구를 보유하고 있는 중국의 인터넷 사용자에 관한 연구도 활발히 이루어지고 있다. Wang 등은 2011년 말에 발생한 1억 명 규모의 패스워드 유출 사고 데이터를 활용하여 중국인 인터넷 사용자의 패스워드 습관이 취약함을 보였다[16]. Li 등은 중국인의 패스워드를 영어 사용자의 패스워드와 비교한 결과 중국인의 패스워드에서는 날짜가 상대적으로 더 많이 관찰되었음을 보였고, 사전에 중국어의 로마자 표기법인 핀인(Pinyin)을 적절히 추가하는 방법을 사용하면 중국인의 패스워드 추측 효율이 34% 증가함을 보였다[9].

2.5 비-텍스트 기반 패스워드 연구

Karapanos 등은 일반적인 이중 인증 과정에서 사용자가 그의 휴대전화를 조작해야만 한다는 불편함을 해결하기 위해 사용자가 로그인을 시도한 장치와 사용자의 휴대전화간 거리를 두 번째 인증수단으로 하는 이중 인증 기법을 개발하였다[7]. 해당 연구에서는 주변 소음을 비교하는 방식으로 장치와 휴대전화의 거리를 측정한다. Hang 등은 패스워드 분실 시 위치 기반 질문을 통해 이를 복구하는 방법을 제안하였다[6]. 이는 대부분의 패스워드 복구 수단이 이메일 또는 보안 질문을 사용하는데, 이메일은 모든 상황에 적합하지는 않고 보안 질문은 사용성과 안전성에 문제가 있다는 단점을 보완하기 위한 것이다.

III. 연구 설계

데이터 수집은 설문지를 활용한 설문 조사를 통해 이루어졌으며 설문지의 구조가 Table 1.에 정리되어 있다.

Table 1. Survey structure

Question category	Objects of investigation
Demographic	Age, Gender, Job
Password creation	Password creation methods
Password structure	Length, Password pattern, Used meanings
Password management	The number of account, ID, Password, Password management methods
Password rule	Difference of password patterns between users under two rules

3.1 설문 항목

3.1.1 패스워드 생성 설문 항목

패스워드 생성 방법은 신규 생성과 재사용으로 나

눌 수 있다. 패스워드를 재사용한다는 것은 다른 사이트를 포함하여 과거에 사용했던 패스워드를 일부 변경하거나 완전히 동일한 패스워드를 이용하여 생성하는 것을 의미한다. 패스워드 재사용 방법은 매우 다양하다. 본 연구에서는 이 중 사용자들이 선호하는 방법이 무엇인지 조사하였다.

3.1.2 패스워드 구조 설문 항목

패스워드를 대문자, 소문자, 숫자, 특수문자의 4가지 구성 청크 관점에서 분석한다면 특수문자는 패스워드의 뒤쪽, 대문자는 패스워드의 앞쪽에 있을 확률이 높다는 점 등은 기존 연구를 통해 알려져 있다[4][12]. 또한 사전과 자연어 처리 알고리즘을 통해 유출된 패스워드에서 발견된 단어를 사전적 의미에 따라 분류한 후 이를 통해 패스워드 문법을 구축하여 패스워드 추측 공격 효율을 높인 연구도 있다[15]. 그러나 이는 같은 단어라도 사용자에게 어떤 의미가 있는지는 알 수 없고 머리글자 등 단어가 아닌 것의 의미는 더 알기 어렵다는 문제가 있다. 예를 들어 사용자가 패스워드의 가장 앞에 'A'를 넣었다고 할 때, 해당 문자가 사용자의 이름인지, 단순히 패스워드의 첫 번째 자리에 알파벳 첫 글자를 넣은 것인지는 확인할 수 없다. 이러한 문제를 해결하고자 본 연구에서는 사용자가 패스워드에 포함시킨 문자열이 본인에게 어떤 의미가 있는지 조사하였다.

3.1.3 패스워드 관리 설문 항목

각 계정마다 서로 다른 패스워드를 사용하는 것이 안전하지만 사용자가 관리해야하는 계정의 수가 증가함에 따라 이는 지켜지기 어려운 보안수칙이 되었다. 본 연구에서는 사용자들이 몇 개의 계정을 보유하고 있으며 이 계정들에 몇 개의 ID와 패스워드를 사용하는지, 그리고 패스워드들을 어떤 방법으로 관리하며 얼마나 자주 변경하는지를 조사하였다.

3.1.4 패스워드 규칙 설문 항목

대부분의 국내 웹 사이트에서는 패스워드 생성 시 패스워드의 최소 길이 또는 구성 청크 조합을 권고 또는 강제하고 있다. 이때, 많은 국내 웹 사이트들이 구성 청크 조합을 권고하면서 구성 청크를 특정한 순서로 제시하는 것을 관찰하였다(Fig.1.). 본 연구에서는 사용자들이 패스워드를 만들 때 이러한 순서를 따르는지, 또한 제시된 순서가 바뀌면 패스워드의 구조도 바뀌는지 실험을 통해 확인하였다.

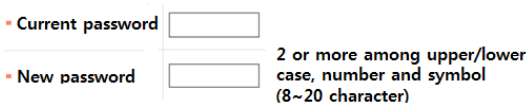


Fig. 1. A password policy example

3.2 설문 문항

3.2.1 패스워드 생성 문항

패스워드 생성에 관한 경향을 조사하기 위해 패스워드 생성 시 대체로 기존 패스워드를 재사용하는지 확인한 후, 이 문항에 긍정적으로 답변한 사람들을 대상으로 선호하는 패스워드 수정 방법을 복수응답이 가능한 문항으로 조사하였다. 또한 일반적으로 패스워드 생성 시 선호하는 특수문자를 조사하였다. 이때 사용자들이 떠올리기 쉽도록 키보드의 이미지를 첨부하여 제공하였다.

3.2.2 패스워드 구조 문항

패스워드 구조 경향 조사를 위해 응답자가 다음 절차에 따라 응답을 작성하도록 한다(Table 2.).

1) 보유한 계정 중 가장 자주 접속하는 등 비교적 가치가 높은 계정의 패스워드 1개를 떠올린다.
2) 패스워드를 지정된 기호로 바꾸어 적는다.
3) 기호로 변환한 패스워드의 일부 또는 전체를 패스워드 생성 시 본인이 부여한 의미별로, 묶이지 않은 문자 없이 묶는다.
4) 각 묶음이 가진 의미에 가장 가까운 의미 기호

Table 2. Expected responses to a password structure question

Step	Result
1	20hee0615@
2	XX▽▽▽XXXX☆
3	(XX)(▽▽▽)(XXXX)(☆)
4	(XX)(▽▽▽)(XXXX)(☆) ⓑ ⓐ ⓥ
5 (Expected response)	(XX)(▽▽▽)(XXXX)(☆) ⓑ ⓐ ⓥ ⑧

Table 3. Password semantic components

Type	Mark	Keyword	Description
Chunk relevant to you	ⓐ	Date	Your personal anniversary (e.g. birthday, wedding anniversary)
	ⓑ	Name	Your name (e.g. full, first, last, nickname, initials)
	ⓒ	Phone number	Your phone number (e.g. home, cellular phone number)
	ⓓ	Belongings	Relevant to your possession (e.g. car number, serial number, brand name)
	ⓔ	Hobby	Your pastime (e.g. sports, movie, novel, game, comic book)
	ⓕ	Interest	Your interest (e.g. future goal, desire, concern, political issue, social issue)
	ⓖ	Pet	Information relevant to your pet (e.g. pet name, pet birthday)
	ⓗ	Residence	Information relevant to your residence (e.g. address, building name, bus number)
	ⓘ	Occupation	Related to your current or past job (e.g. student ID number, employee ID number or technical term)
	ⓙ	Favorites	Items of personal preference (e.g. specific food, cigarette, alcohol, tea)
	ⓚ	Religion	The thing related to your religions (e.g. date, great men)
	①	Acquaintance	A information of people yon know directly (e.g. family, friends)
Containing meanings	ⓜ	Website information	The name, color or feature of website.
	ⓝ	ID	A part or all of the string used on ID
	ⓞ	'Password'	A string means password (e.g. P@$$w0rd, passcode, PIN)
	ⓟ	Preferred number/word	A number, word or phrase that you just prefer
	ⓠ	Other number/word	A number, word or phrase that are irrelevant to you (e.g. repetitively heard commercial number, TV ad numbers)
Meaning-less	ⓡ	Spatial characters	Sequential characters on a keyboard for your convenience (e.g. 123456, qwerty, !@#$, 1q2w3e4r)
	ⓢ	Alphabetical orders	Alphabetical order keys without special meanings (e.g. abcd, efgh)
	ⓣ	Repetition	Using the same character more than 3 times for your convenience without special meanings
	ⓤ	Punctuation marks	Using symbols (e.g. . , ! " (@) as if they are used in daily words, sentences, URL, or email address
	ⓥ	Random key	The key(s) for making password longer and complex without special meaning
Others	ⓩ	Others	Cannot find relevant keywords above

Table 4. Word expressions

Mark	Expression	Example
①	English	goodbye, Iloveyou, SonOGong, bbangzib
②	Hangul	rneqkdl, dkdlfjqmdb, thsdhrhd, Qkdwlq
③	Initial, Acronyms	GB, ILY, SOG, BbZ
④	Other	Anagram(goodbye → obeygod), borrowed notation(dhrhd → 50, rlfehd → roadeast), Leet transformation(Pass → P@$$)

를 설문지에 제시된 키워드 및 기호 표(Table 3.)에서 찾아 키워드 기호를 묶음 밑에 표기한다.

5) 패스워드에 한글 및 영어 단어 또는 이를 변형한 묶음이 있다면 Table 2. 의 5단계 결과와 같이 설문지에 제시된 표현 방법 기호를 찾아 키워드 기호 밑에 표기한다(Table 4.). 패스워드의 모든 묶음이 단어 기반이 아니거나, 한글 또는 영어가 아닌 단어에 기반을 둔 묶음만 있다면 이 단계는 무시한다.

Table 2.는 설문 참가자들에게 본 문항을 설명하는 과정에서 참가자의 이해를 돕기 위해 사용했던 응답 예시이다. 예시는 이름이 '이영희'인 가상의 사용자가 본 문항에 응답하는 시나리오를 따른다. 가상의 사용자 이영희는 본인이 가장 자주 접속하는 계정에 '20hee0615@'라는 패스워드를 사용하고 있다. 이 패스워드의 문자열 중 '20hee'는 사용자 이름인 '이영희'에서 따온 것이며 '0615'는 이영희의 생일이다. 마지막 '@'는 패스워드를 길고 복잡하게 만들기 위해 넣은 아무 의미 없는 키이다.

3.2.3 패스워드 관리 문항

먼저 사용자들이 보유한 계정 중에서도 비교적 가치 있는 계정만을 선별하도록 하였다. 이는 설문 응답자가 본인이 보유한 모든 계정에 관한 사항을 정확히 떠올리는 것은 매우 어려운 일이므로 조사 대상을 한정할 필요가 있고, 그 대상에는 공격 대상이 될 가능성이 높은, 다시 말해 비교적 가치 있는 계정이 포함되는 것이 합리적이기 때문이었다. 이러한 계정을 이후 '주요 계정'으로 표기한다. 설문 시 주요 계정의 구체적인 기준은 주 1회 이상 로그인으로 설정했으나 모두 기억하기 어렵다면 즉시 기억나는 계정만 적도록 했고 로그인 횟수에 상관없이 본인이 가치 있게 여기는 계정도 포함시킬 수 있었다.

사용자들의 주요 계정 수를 조사하기 위해 웹 사이트를 13종류(금융, 인터넷 쇼핑, 포탈, 정보 검색, 정보 교환, SNS, 온라인 게임, 동영상 감상, 음악 감상, 온라인 강의, 주요 뉴스, 전산시스템, 기타)로 나누고 종류별 웹 사이트 예시를 제공하였다. 이 문항으로 각 종류에 해당하는 본인의 주요 계정 수를 조사하고 그 합계를 각 사용자의 주요 계정 수로 사용하였다. 13종류로 나눈 것은 본인이 보유한 주요 계정 수를 한 번에 정확히 계산하는 것이 어려운 일이므로 사용자가 나누어 생각함으로써 쉽게 떠올릴 수 있도록 돕기 위한 설계였다.

앞서 조사한 주요 계정만을 대상으로, 주요 계정에 사용하는 아이디와 패스워드의 수를 조사하였다. 그 후 주요 계정들의 암기, 기록, 자동 로그인 3가지 패스워드 관리 방법에 대한 각각의 의존도를 조사하였다. 패스워드 관리 방법 중 '기록'은 패스워드를 종이에 적어두거나 전자기기에 저장해두고 로그인 시 이를 참고하는 방법이며 자동 로그인은 브라우저에 내장된 패스워드 저장 기능을 사용하는 등 패스워드를 사용자가 직접 입력하지 않는 방식이다. 의존도는 사용자 개인이 보유한 모든 주요 계정 중 해당 관리 방법을 사용하여 패스워드를 관리하고 있는 계정의 비율을 의미한다. 예를 들어 주요 계정을 4개 보유한 어떤 사용자가 그 중 2개 계정의 패스워드를 외워서 사용하고 모든 계정의 패스워드를 적어두었으며 항상 패스워드를 직접 입력해서 사용한다고 가정한다면 이 경우 암기, 기록, 자동 로그인에 대한 사용자의 의존도는 각각 50%, 100%, 0%이다. 추가적으로 최근 1년간 주요 계정의 패스워드에 대한 패스워드 변경 횟수와 웹 사이트의 패스워드 변경 권고에 대한 반응을 조사하였다.

3.2.4 패스워드 규칙 문항

먼저 설문 응답자들은 본인이 보유한 계정 중 가장 자주 로그인하거나 가장 중요하게 여기는 계정 하나를 선택한다. 그리고 해당 계정의 패스워드를 설문지에서 제시한 규칙에 따라 변경하는 가상 시나리오를 따른다. 설문지에는 현실성을 높이기 위해 스크린샷이 포함되었으며[13] 규칙을 제시하는 방법에 따라 두 종류를 준비하였다[8]. 두 종류의 설문지에 인쇄된 규칙의 내용은 각각 다음과 같다.

- 비밀번호는 영어 대문자/소문자, 숫자, 특수문자 중 2종류 조합 시 10자리 이상, 3종류 이상 조합 시 8자리 이상으로 설정하세요.
- 비밀번호는 숫자, 영어 소문자/대문자, 특수문자 중 2종류 조합 시 10자리 이상, 3종류 이상 조합 시 8자리 이상으로 설정하세요.

첫 번째 패스워드 규칙은 한국인터넷진흥원(KISA)의 패스워드 선택 및 이용 안내서에 제시된 규칙[17]의 내용을 참고하여 결정하였다. 이를 이후 '대소숫특(ULNS) 규칙'이라 한다. 두 번째 규칙의 나열 순서는 금융결제원 전자인증센터(yessign)의

공인인증서 비밀번호 규칙[18]에 나열된 순서를 참고하여 첫 번째 규칙을 변형한 결과이다. 이를 이후 '숫소대특(NLUS) 규칙'이라 한다.

'대소숫특 규칙'이 인쇄된 설문지를 이후 '대소숫특 설문지'라 하고, '숫소대특 규칙'이 인쇄된 설문지를 이후 '숫소대특 설문지'라 한다. 각 설문지에서 패스워드 규칙 문항을 제외한 나머지 문항의 내용은 모두 같다. 설문 응답자가 규칙에 맞추어 만든 패스워드에서 구성 요소의 순서를 관찰하여 패스워드 규칙 표현 방식이 패스워드 구성에 영향을 주는지 검증한다.

각 응답자의 답변을 관찰하여 다음과 같은 규칙에 따라 점수를 부여한다.

• 100점 : 패스워드 구성 청크의 사용 순서가 설문지에 주어진 규칙에 나열된 순서의 전체 또는 일부와 일치하는 경우이다. 예를 들어 대소숫특 설문지에서 패스워드를 대-소-숫 또는 소-특 또는 대-소-숫-특 순서로 만든 응답자 등이 여기에 해당한다.

• 50점 : 패스워드 구성 청크의 최초 사용 순서가 설문지에 주어진 규칙에 나열된 순서의 전체 또는 일부와 일치하는 경우이다. 예를 들어 대소숫특 설문지에서 패스워드를 대-소-숫-소-숫 순서로 만든 응답자가 있다고 하면, 이 응답자가 만든 패스워드 구성 청크의 최초 등장 순서는 대-소-숫이므로 50점을 부여한다.

• 0점 : 100점과 50점 중 어느 경우에도 해당하지 않는 경우로, 예를 들어 특-소-숫 순서로 만든 경우와 같이 설문지에서 뒤쪽에 나열된 구성 청크가 앞쪽에 나열된 구성 청크보다 먼저 사용된 경우에 0점을 부여한다.

IV. 설문 결과

설문은 서울 시내 여러 장소에서 진행되었다. 조사원들은 행인을 대상으로 설문 참여를 요청한 후 이를 수락하면 설문 내용을 설명하고 응답을 수집하였다. 그 후 설문을 마친 응답자에게 소정의 답례품을 제공하였다. 총 327개의 응답이 수집되었고, 3개의 불완전한 응답은 분석에서 제외되었다.

4.1 응답자 인구 통계

유효 응답자 324명의 성별 및 직업 분포가 각각 Table 5. 와 Table 6. 에 나타나 있다. 연령별로는 주로 20대(165명)와 30대(99명) 연령층이 많았으며, 직업별로는 주로 학생(90명)과 사무직/관리직(113명) 및 전문직/자유직(52명) 종사자들이 많이 참여하였다.

Table 5. Gender & Age distribution of participants

Age	Male	Female	Sum
10s	2	1	3
20s	56	109	165
30s	66	33	99
40s	27	17	44
50s	8	3	11
over 60	0	2	2
Sum	159	165	324

Table 6. Job distribution of participants

Job	Sample
Student	90
Blue collar job	3
Private business	8
Office/managerial job	113
Sales/service job	21
Profession/free lancer	52
Agricultural/forestry/fishery/livestock industry	1
Full-time homemaker	9
Unemployed	8
Teaching/researching job	12
Other	7
Sum	324

4.2 응답 분석

4.2.1 패스워드 생성 응답 분석

전체 응답자 324명 중 298명(91%)이 패스워드를 만들 때 대체로 기존 패스워드와 유사하게 만든다고 응답하였다. 이들 298명을 기준으로 한 특별히 선호하는 패스워드 변경 방법의 분포가 Fig.2. 에 나타나 있다. 타 사이트와 완전히 같은 패스워드를 사용하거나(38%) 기존의 패스워드에 숫자 또는 특수문자를 추가/삭제하는 방법(29%, 39%)을 가장 선호하였다. 이 문항은 선택 가능 응답 수 제한 없이 복수응답이 가능했다.

전체 응답자 324명 중 218명(67%)이 패스워드를 만들 때 선호하는 특수문자가 있다고 응답하였다. 이들 218명을 기준으로 선호하는 특수문자를 최대 5개까지 선택하도록 한 결과 중 상위 10개 특수문자

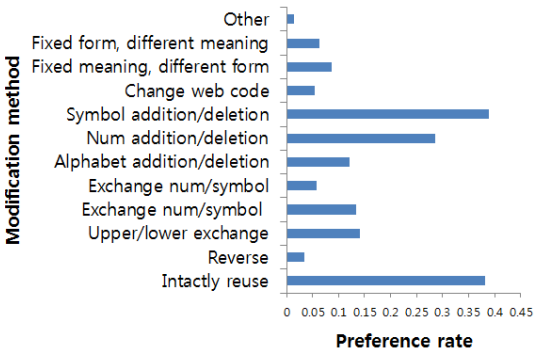


Fig. 2. Preference rate of password modification methods

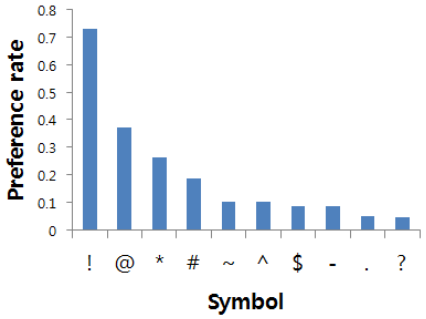


Fig. 3. Preference level of symbols

의 선호도 분포가 Fig.3.에 나타나 있다. '!', '@' 등 사용자들이 유의하게 선호하는 특수문자가 존재함을 확인하였다($\chi^2-test, p<0.001$).

4.1.1 패스워드 구조 응답 분석

패스워드 구조 관련 문항에 대한 설문 응답자 324명의 응답을 분석한 결과 패스워드 길이는 평균 9.63자였고 평균 2.59개의 의미가 포함되어 있었다 (Table 7.). Fig.4. 는 대문자(U), 소문자(L), 숫자(N), 특수문자(S)의 조합 방법 중 많이 사용된 몇몇 방법의 사용 빈도를 보여주고 있다. 소-숫 또는 소-숫-특의 순서로 구성된 패스워드를 유의하게 많이

Table 7. Password length and semantic chunks

	Length	Number of semantic chunks
Average	9.63	2.59
Std. deviation	1.98	0.86
Minimum	4	1
Maximum	17	6

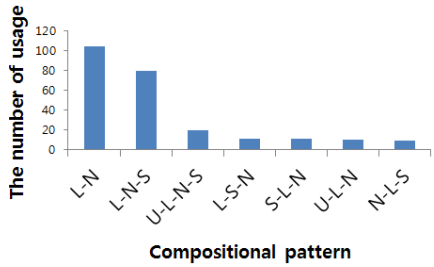


Fig. 4. Compositional patterns of passwords

사용하고 있었다($\chi^2-test, p<0.001$).

패스워드의 의미 구조는 본 연구에서 사용한 기준을 적용할 경우 사용자마다 천차만별임이 드러났다. 설문 응답에서 총 207개의 의미 구조가 관찰되었다. 그 중 170개는 각각 한 사람이 사용하는 구조였고 20개는 2명이 사용하고 있었다. 이 210명을 제외한 나머지 114명이 사용하는 의미 구조 중 사용자 수 기준 상위 12개(상위 10순위 구조)의 의미 구조가 Fig.5.에 나타나 있다. 한 가지 흥미로운 발견은 상위 12개의 의미 구조 중에서 패스워드에 '이름(ⓑ)'이 포함되었다면 '무작위 키(ⓥ)'를 제외한 다른 의미 청크보다 먼저 사용되었다는 점이다.

각 의미 청크의 사용률이 Fig.6.에 나타나 있다. 예를 들어 '이름'의 경우 약 46%의 사용률을 보이고 있는데 이는 응답자 324명 중 149명(약 46%)의 응답에 본인의 이름이나 별명 등을 어떤 형태로든 1회 이상 포함시켰음을 의미한다. 전체 응답자 324명 중 262명(약 80%)이 본인과 관련된 의미(ⓐ~①) 중 하나 이상을 패스워드에 포함시켰다. 본인과 관련된 여러 의미 중에서도 '이름(46%)'과 '날짜(25%)'가 유의하게 빈번히 포함되었다($\chi^2-test, p<0.001$). 일부 의미 청크는 다양한 방법으로 표현이 가능하다. 예를 들어 '날짜'는 거의 숫자로 표현하는 반면, '이

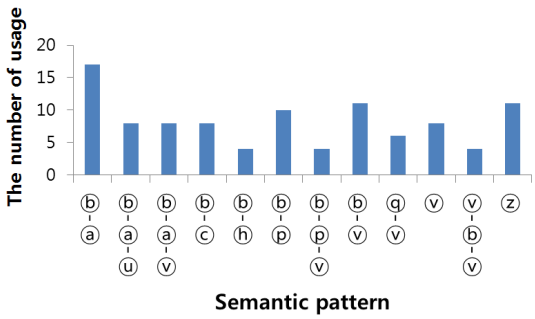


Fig. 5. Semantic patterns of passwords

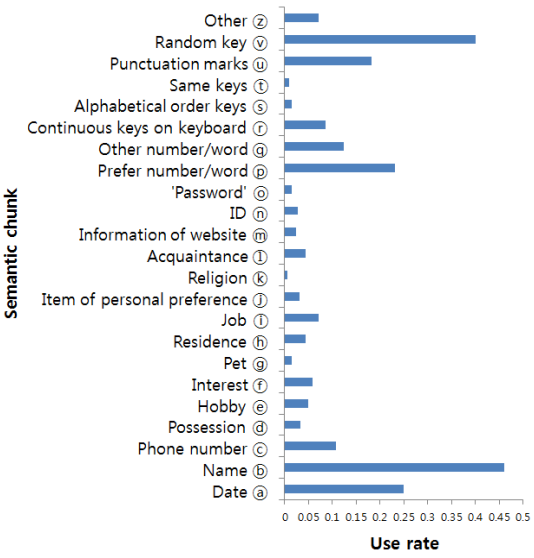


Fig. 6. Semantic chunk usage rate

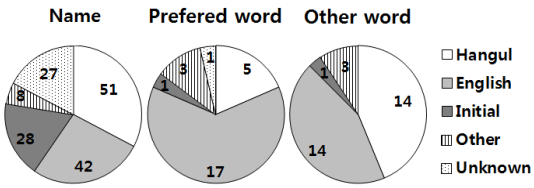


Fig. 7. Word expressions in user-chosen passwords

름'은 영문 표기법, 한글, 머리글자, 기타 외국어, 또는 더 복잡한 변환을 거친 방법으로 표현할 수 있다. 본 연구에서는 이러한 다양한 방법으로 표현이 가능한 의미 청크들 중 빈번히 관찰된 '이름', '선호 단어', '기타 단어'에 대해 어떠한 표현 방법을 사용하는지 분석하였고, 그 결과가 Fig.7.에 나타나 있다. '선호 단어'와 '기타 단어'는 각각 '선호 숫자/단어'와 '기타 숫자/단어'에서 숫자로 사용된 경우를 제외한 나머지이다. Fig.7.에서 'Unknown'은 사용자가 표현 방법 표기를 생략하여 표현 방법을 알 수 없는 경우이다.

4.2.2 패스워드 관리 응답 분석

상관분석 결과 주요 계정과 아이디 개수는 0.16의 상관계수를 가지고 있었으며 주요 계정과 패스워드 개수는 0.15의 상관계수를 가지고 있었다. 즉, 주요 계정의 개수와 아이디, 패스워드 개수의 사이에는 거의 상관관계가 없었다.

사용자들이 보유한 주요 계정 수와 한 개의 아이디 또는 패스워드의 평균 중복 사용수가 Fig.8.에 나타나있다. 막대그래프를 통해 절반 이상의 사용자들이 15개 이하의 주요 계정을 관리하고 있음을 알 수 있다. 꺾은선그래프는 다음과 같은 의미를 가지고 있다. 예를 들어 주요 계정을 31~35개 가지고 있는 사용자들은 평균적으로 11.7개의 계정에 같은 아이디를 중복 사용하며 9.7개의 계정에 같은 패스워드를 중복 사용하는 것으로 해석할 수 있다. 이를 통해 관리하는 주요 계정 수가 많아질수록 같은 아이디와 패스워드를 더 많이 재사용함을 확인할 수 있으며, 대체로 패스워드의 재사용이 ID의 재사용보다 적음을 확인할 수 있다.

패스워드 관리 방법별 의존도가 Fig.9.에 정리되어 있다. 이상 값(outlier) 점 옆에는 해당 답변을 선택한 사용자의 수가 표기되어 있다. 주로 암기에 의존하는 경향이 관찰되었으며, 동시에 기록에는 거의 의존하지 않고 자동 로그인에도 비교적 의존하지 않는 경향을 보였다. 추가로 조사한 패스워드 변경 주기 및 패스워드 변경 권고에 대한 반응에서 응답자 324명중 245명이 최근 1년간 주요 계정의 패스워드를 4회 이하 변경하였으며 동시에 웹 사이트가 패스워드 변경을 권고하면 대체로 '나중에 변경하기'를 누른다고 응답한 결과는 이처럼 패스워드를 주로 암기에 의존하여 관리하는 경향을 뒷받침한다.

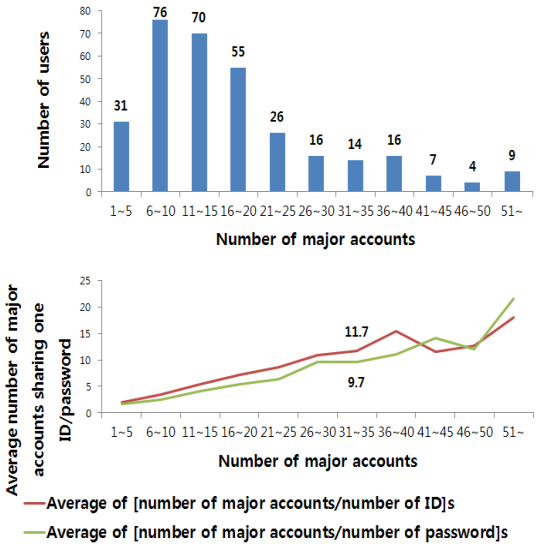


Fig. 8. Number of major accounts and duplicated usage rates

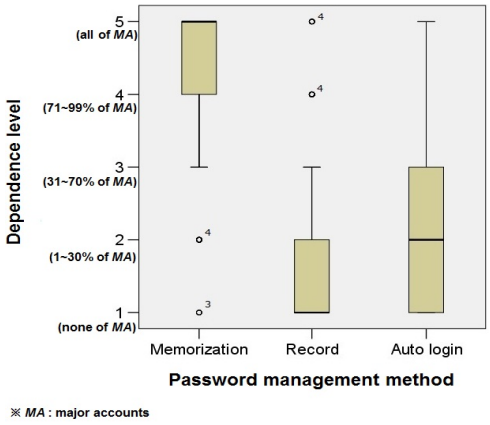


Fig. 9. Users' dependency on password management methods

4.2.3 패스워드 규칙 응답 분석

유효한 응답자 324명 중 168명은 대소숫특 설문지를, 156명은 숫소대특 설문지를 받았다. 각 응답에 대해 2.2.4절에서 설명한 것과 같이 점수를 매긴 결과가 Table 8. 에 정리되어 있다. 대소숫특 설문지의 응답들을 대소숫특 규칙의 관점에서 점수를 부여한 후 평균을 구한 결과는 60.7점이었다. 숫소대특 설문지의 응답들을 숫소대특 규칙의 관점에서 점수를 부여한 후 평균을 구한 결과는 7.3점이었다. 평균 점수에 유의한 차이가 있다고 나타났으며 $(t-test, p \leq 0.05)$ 이는 사용자들이 패스워드를 설문지에서 제시한 규칙과 비슷하게 만든다고 볼 수 없음을 의미한다. 또한 숫소대특 설문지의 응답들을 대소숫특 규칙의 관점에서 점수를 부여하고 평균을 구한 결과는 58.6점으로 앞서 구한 60.7점과 유의한 차이가 없었다$(t-test, p=0.69)$. 이는 사용자들이 설문지에서 제시한 규칙과 상관없이 대소숫특 규칙과 비슷하게 만든다는 것을 의미한다.

Table 8. The correlation between passwords and policies

	ULNS	NLUS
Participants number	168	156
Average score on ULNS policy	60.7	58.6
Average score on NLUS policy	-	7.3

V. 결론 및 시사점

본 논문에서는 패스워드의 생성, 구조, 관리, 규칙에 관한 사용자 설문 연구를 진행하였다. 총 327개의 설문 응답을 수집하였고 응답자는 주로 서울 지역의 20~30대 학생 또는 사무직, 전문직 종사자였다.

사용자들은 매번 완전히 새로운 패스워드를 만들기보다는 기존에 사용하던 패스워드와 같거나 비슷한 것을 선호하였으며 패스워드에 사용하는 특수문자는 상당히 편중되어 있었다. 같은 패스워드 정책이라도 이를 설명하는 방법에 따라 사용자들이 생성하는 패스워드의 구조가 달라지는지 실험을 진행하였으나 유의한 차이가 나타나지 않았으며 이는 앞서 언급한 결과와도 합치된다. 따라서 기존에 사용하던 패스워드를 재사용하지 않도록 유도하는 정책이 필요하다.

이러한 정책의 일환으로 복잡한 문자열과 웹사이트 정보를 활용하여 웹사이트마다 패스워드를 다르게 설정하는 방법이 권장되고 있다[19]. 이는 매우 복잡한 문자열을 중심으로 앞 또는 뒤에 웹사이트 이름 등을 일부 추가하여 사실상 단 하나의 패스워드만 암기하면서 동시에 웹사이트마다 다른 패스워드를 설정할 수 있는 방법이다. 자동화된 공격에 대해서는 효과적인 방법으로 생각되나, 이러한 설정 방법을 숙지한 공격자가 보안이 취약한 웹사이트를 공격하여 패스워드를 확보한다면 이 방식의 사용자가 오히려 더 위험할 것이라 사료되며 그 이유는 다음과 같다.

첫째로 본 연구의 결과에 의하면 패스워드에 웹사이트 정보를 포함시키는 사람이 적었고(Fig.6.), 둘째로 해당 패스워드의 주인이 비교적 보안에 관심 있다고 생각 될 수 있으며, 셋째로 사용자들이 대체로 패스워드에서 웹사이트 정보를 공격자가 찾아내기 어렵게 만들 것이라 기대하기 어렵다. 실제로 본 연구에서 웹사이트 정보를 Table 4. 의 '기타'의 방법으로 포함시킨 사용자는 없었다. 즉, 공격자는 가치 있는 계정을 보유할 가능성이 비교적 높은, 적은 수의 사용자를 찾아낼 수 있으며 다른 웹사이트의 패스워드 또한 쉽게 추측할 수 있는 것이다. 따라서 상기한 설정 방법은 충분한 검증이 필요하다 사료된다.

사용자들은 패스워드를 기록해 놓거나 패스워드 매니저 등을 사용하기보다는 주로 암기에 의존하여 관리하고 있었다. 이는 사용자가 보유한 계정 수와 사용하는 패스워드 수가 낮은 상관관계를 보이는 결과를 잘 설명한다. 다시 말해 한정된 기억력에만 의존하므로 다양한 패스워드를 사용하지 못하고 같거나

비슷한 패스워드를 반복 사용한다고 해석할 수 있다. Ur 등의 연구[14]에 따르면 사용자들이 항상 웹사이트의 중요도에 따라서 패스워드 강도를 다르게 한다고 생각하기 어려우며 이는 앞서 언급한 결과와 연관 지어 생각해 볼 때 많은 사용자들이 같거나 비슷한 패스워드를 반복 사용하는 것의 위험성을 올바르게 인식하지 못하고 있다고 볼 수 있다. 그러므로 사용자에게 보안 정책 준수를 요구하는 것 외에도 이러한 위험성을 충분히 잘 전달할 필요가 있다.

본 연구는 패스워드의 구조를 가장 중점적으로 조사하였다. 대소문자, 숫자, 특수문자 조합 관점에서의 구조는 소문자-숫자 또는 소문자-숫자-특수문자 순서로 매우 정형화되어 있었다. 의미 관점에서의 구조를 분석한 결과 많은 사용자들이 패스워드에 본인의 신상정보, 특히 본인의 이름과 관련 있는 문자열을 포함시키는 것을 확인하였다. 또한 패스워드의 의미 구조 중 사용 빈도가 가장 높은 구조의 대부분이 이름 뒤에 다른 단어나 키를 추가하는 방식으로 구성되어 있었다. 따라서 사용자들이 패스워드에 포함시켜서는 안 될 정보를 제시할 때, 그중 첫 번째는 본인의 이름이 되어야 할 것이다.

본 연구는 연구 방법으로 설문 조사를 사용하였다. 이는 대규모 유출 패스워드를 분석한 기존 연구[9][15]에 비해 패스워드에 사용된 문자열과 사용자의 관계를 용이하게 파악할 수 있다는 장점이 있으나 표본의 수가 적다는 한계 또한 존재한다. 따라서 본 연구를 보다 대규모의 국내 사용자를 대상으로 수행하려 한다면 실제 인터넷 서비스 사용자의 데이터를 분석한 연구[2][5]와 같이 데이터를 수집하는 방법이 있을 수 있다. 그러나 기존 연구[2][5]와 달리 패스워드의 문자열과 사용자의 관계를 파악하는 것에 있어서는 개인정보 또한 매우 중요하므로 실제 연구 수행에 앞서 관련법과 연구 윤리에 대한 충분한 조사와 철저한 숙지가 필요할 것이다.

References

[1] J. Abbott and V.M. Garcia, "Password differences based on language and testing of memory recall," NNGT Int J. on Information Security, vol. 2, pp. 1-6, Feb. 2015.

[2] 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.

[3] R. Chatterjee, A. Athalye, D. Akhawe, A. Juels and T. Ristenpart, "pASSWORD tYPOS and How to Correct Them Securely," Proceedings of the 37th IEEE Symposium on Security and Privacy, pp. 799-818, May, 2016.

[4] A. Das, J. Bonneau, M. Caesar, N. Borisov and X.F. Wang, "The tangled web of password reuse," Proceedings of the Network and Distributed System Security Symposium, Feb. 2014.

[5] D. Florêncio and C. Herley, "A Large-Scale Study of Web Password Habits," Proceedings of the 16th international conference on World Wide Web. ACM, pp. 657-666, May, 2007.

[6] A. Hang, A. De Luca, M. Smith, M. Richter and H. Hussmann, "Where Have You Been? Using Location-Based Security Questions for Fallback Authentication," Proceedings of the Symposium on Usable Privacy and Security, pp 169-183, July. 2015.

[7] N. Karapanos, C. Marforio, C. Soriente, and S. Čapjun, "Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound," Proceedings of the 24th USENIX Security Symposium, pp. 483-498, Aug. 2015.

[8] S.Y. Kim and T.K. Kwon, "A Study of Interpretation Effect of Passwords to Password Generation," Journal of the Korea Institute of Information Security and Cryptology, 25(2), pp. 1235-1243, Oct. 2015

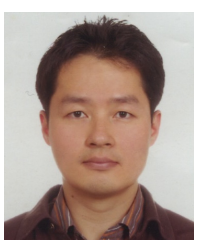
[9] Z. Li and W. Han, "A Large-Scale Empirical Analysis of Chinese Web Passwords," Proceedings of the 23rd USENIX Security Symposium, pp. 559-574, Aug. 2014.

[10] R. Morris and K. Thompson, "Password

security: A case history," Communications of the ACM, vol. 22, no. 11, pp. 594-597, 1979.
[11] C. Rinn, K. Summers, E. Rhodes, J. Virothaisakun and D. Chisnell, "Password Creation Strategies Across High- and Low- Literacy Web Users," Proceedings of the 78th Association for Information Science and Technology vol. 52, no. 1, 2015.
[12] R. Shay, S. Komanduri, P.G. Kelly, 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 Symposium on Usable Privacy and Security, pp. 243-255, July. 2014.
[13] 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.
[14] B. Ur, F. Noma, J. Bees, S. M. segreti, 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," Proceedings of the 11th Annual Symposium on Usable Privacy and Security, pp. 123-140, July. 2015.
[15] R. Veras, C. Collins, and J. Thorpe, "On the semantic patterns of passwords and their security impact," Proceedings of the Network and Distributed System Security Symposium, Feb. 2014.
[16] W. Wang, H. Wang and Y. Meng, "A Large-scale Survey on Password Habits of Internet Users in China," Jounal of Convergence Information Technology, vol. 8, no. 4, pp. 71-80, 2013.
[17] Korea Internet & Security Agency, "Password choice and using guidance," 2008.
[18] The reinforcement of authorized certificate rule : http://www.yessign. or.kr/common/popup/home/28.do
[19] How to Create a Secure & Memorable Password : http://www.ahnlab.com/kr / site/securityinfo/secunews/secuNewsView.do?seq=24062

〈저자소개〉

김 승 연 (Seung-Yeon Kim) 학생회원
2015년 2월: 세종대학교 응용통계학 및 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 교수
〈관심분야〉 암호 프로토콜, 네트워크 프로토콜, IoT 보안, Usable Security, HCI 등