Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.26, NO.3, Jun. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.3.757

국내 웹 사이트 패스워드 미터 분석*

김 경 훈,† 권 태 경‡
연세대학교 정보보호연구실

An Analysis of Password Meters for Domestic Web Sites*

KyoungHoon Kim,† Taekyoung Kwon‡
Information Security Lab., Graduate School of Information, Yonsei University

요 약

패스워드 인증은 가장 대표적인 사용자 인증 기법이며 그 중에서 특히 문자 기반 패스워드가 가장 많이 사용되고 있다. 그러나 사용자들이 선택하는 패스워드가 갖는 취약성으로 인하여 사용자로 하여금 강한 패스워드 생성을 유도하기 위해 많은 웹사이트에서는 패스워드 강도를 측정하는 패스워드 미터를 제공하고 있다. 하지만 여기에는 패스워드 미터 결과가 일관되지 않고, 정확하지 않은 강도로 피드백하는 문제가 있다. 이에 본 논문에서는 패스워드 미터의 문제점에 대해 알아보며 패스워드 미터의 개선 방향을 제시하고자 한다.

ABSTRACT

Password authentication is the representative user authentication method and particularly text-based passwords are most widely used. Unfortunately, most users select weak passwords and so many web sites provide a password meter that measures password strength to derive the users to select strong passwords. However, some metering results are not consistent and incorrect strength feedbacks are made. In this paper, we tackle these problems regarding password meters and present an improvement direction.
Keywords: Password, Password Meter, Meter Accuracy.

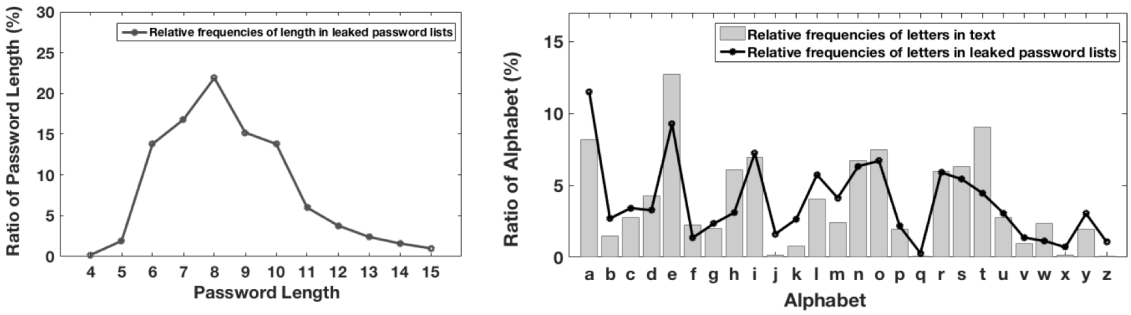
I. 서 론

사용자 인증 방법 가운데 가장 많이 사용되는 방법으로 텍스트 패스워드 입력을 통한 인증방법은 현재 웹 환경뿐만 아니라, 모바일 환경에서도 많이 사용되고 있다. 하지만 사용자들은 텍스트 패스워드를 많이 사용함에도 불구하고 여전히 약한 강도의 패스워드를 생성하고 있다.

사용자들이 생성한 패스워드는 Fig. 1.(a)와 같이 특정 길이에 집중되고 있으며, 자신이 알기 쉬운 단어 및 단순한 키보드 배열 패턴을 사용하고 있다 [3, 14]. 알기 쉬운 단어 사용 및 단순한 키보드 배열 패턴의 패스워드는 약한 강도의 패스워드로 사전 공격 및 무작위 공격 등에 취약하다. 약한 강도의 패스워드는 해커의 공격 목표가 될 수 있으며, 패스워드 노출로 인해 패스워드로 보호되는 암호키 또는 주요 개인 정보 유출사고로 이어져 개인에 피해를 가져올 수 있다. 개인의 피해를 방어하고자 강한 강도의 패스워드 생성을 유도하기 위해 웹 사이트에서는 패스워드 미터를 이용하고 있다.

패스워드 미터는 사용자가 생성하고 있는 패스워드 강도를 피드백 하고 있다. 패스워드 미터의 경우 동일 패스워드에 대하여 서로 다른 미터에서 서로 다른 강도로 판정됨에 따라 패스워드 미터 본연의 목적을 훼손시킬 수 있는 문제점을 밝혔다 [1]. 국외에서는 패스워드 인증 체계와 취약성 관련 연구가 활발

Received (02. 16. 2016), Modified (05. 27. 2016), Accepted (05. 27. 2016)
* 논문은 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다.(No. NRF-2015R1A2A2A01004792)
† 주저자, rickyboss@yonsei.ac.kr
‡ 교신저자, taekyoung@yonsei.ac..kr(Corresponding author)



(a) (b)

Fig. 1. Analysis on leaked passwords (a) Length composition of leaked passwords. (b) Comparison of appearance frequency of alphabets. (This analysis is based on leaked passwords from Hotmail, Myspace, Phpbb, PWlist, and Rockyou and it is a prelimiary result of password vulnerabilities.)

Table 1. Composition ratio per password

	Hotmail	Myspcae	Phpbb	Pwlist	Rockyou
Letter	69.4%	74.9%	74.5%	56.6%	65.8%
Number	29.1%	23.2%	25.2%	42.7%	32.9%
Symbol	1.5%	1.9%	0.4%	0.7%	1.3%
평균 Chunk 수	1.6개	2.2개	1.7개	1.8개	1.7개

히 진행되고 있지만, 국내에서는 활발히 진행되고 있지 않다. 이에 본 논문은 다음과 같은 공헌을 한다.

- 패스워드 생성 정책에 따른 패스워드 미터의 일관성 분석
- 패스워드 미터 정확성 분석
- 패스워드 미터 개선점 제시

II. 연구배경

2.1 패스워드 체계 취약성 사전 분석

텍스트 패스워드 입력을 통한 사용자 인증 방법은 웹 환경뿐만 아니라 모바일 환경에서도 많이 사용되고 있다. 하지만 많이 사용됨에도 불구하고 여전히 약한 강도의 패스워드가 많이 생성되고 있다. 본 연구를 진행하기에 앞서 유출 패스워드 Hotmail, Myspace, Phpbb, PWlist, Rockyou를 수집하여 패스워드 체계 취약성을 분석하였다 [24].

유출 패스워드 분석 결과, 패스워드 구조가 특정 형태에 집중되고 있음을 확인하였다. 생성된 패스워드는 Fig. 1.(a)와 같이 약 78%가 길이 6자리에서 10자리에 집중되고 있다. 또한 패스워드에 사용된 알파벳 출현 빈도와 일반 단어에서 알파벳 출현 빈도가 유사함을 Fig. 1.(b)에서 확인 할 수 있다. 또한 패스워드를 구성함에 있어서 Table 1과 같이 문자 class가 높은 비율을 차지하고 있음을 알 수 있다.

유출 패스워드 청크 분석을 통해 패스워드 구성 시 사용자들이 많이 사용하는 구조가 있음을 확인하였다. 청크는 덩어리의 의미로, 문자(대소문자 포함), 숫자, 특수문자 청크가 있다. Password1234!는 3개(Password / 1234 / !)의 청크로 구성되어 있으며, 문자-숫자-특수문자로 구성된 패스워드이다.

청크 분석 결과 사용자들은 평균적으로 1.75개의 청크를 이용하여 패스워드를 구성하고 있음을 알 수 있었다. 패스워드 구성 시, 문자-숫자, 문자-숫자-문자, 문자-숫자-문자-숫자와 같은 구조로 패스워드를 많이 생성하고 있다.

2.2 패스워드 미터에 대한 기존 연구 결과

약한 패스워드 생성을 막기 위하여 웹 사이트에서는 패스워드 미터를 이용하여 패스워드 강도에 대해 피드백 하면서 강한 패스워드를 사용하도록 유도하고 있다. 패스워드 미터는 지시자로 다양한 표현 방법이 있고,[13], 내부 점수 함수를 통해 사용자가 입력한 패스워드에 대하여 강도를 측정한다. 다양한 표현 방식뿐만 아니라 패스워드 강도를 측정함에 있어서 규칙 기반 패스워드 미터, *n*-gram을 이용한 Adaptive 패스워드 강도 미터 (APSM), *PCFG*를 이용한 Analyzer and Modifier 패스워드 미

Table 2. Examples of passwords in accordance with password rules.

Rule1	18.4 (9.46)	password	iloveyou	veronica	princesca	tequiero	alejandra	diciembre	Chilinflas
Rule2	28.9 (9.73)	mama1939	kli89rty	tqdbr9gv	girl&love	ra7bi4mi7	212carlos	*bentikus*	sebastian.
Rule3	34.7 (8.22)	Beto1984	atv_1978	*zalena6	ladoce-12	062692@dr	info7_tec	dabren/02	Elotroyo21

* Standard deviation and average entropy of password groups in complying with rules.

터 (AMP)가 있다[2]. 대부분의 웹 사이트에서는 규칙 기반 패스워드 미터를 반영하고 있지만, 사용자들이 사용하고 있는 패스워드의 다양성을 반영하고 있지 못한 한계점이 있다. 웹 사이트에서는 표현 방법 및 패스워드 등급 측정에 있어 서로 다른 기준을 가지고 있는 문제점 또한 존재한다[6].

강한 패스워드 생성을 유도해야 하는 패스워드 미터에 대하여 2014년 Carnavalet 등의 연구에서는 유출 패스워드(Top500, Rockyou, Phpbb)와 패스워드 사전을 이용하여 패스워드 미터의 일관성 측정을 통해 비일관 된 강도를 측정하고 있는 문제점을 밝혔으며 [1], 비일관성 문제는 패스워드 생성 시 혼동을 유발 할 수 있음을 주장한다.

2.3 기존 연구 결과와 차별성

본 논문은 사전연구의 방법을 이용하여 유출 패스워드(Hotmail)를 사용하여 국내 접속 상위 웹 사이트에 대하여 패스워드 미터 일관성 측정을 진행한다. 일관성을 측정함에 있어서 기존 연구에서는 대량의 패스워드를 사용하였지만, 본 연구에서는 패스워드 생성 규칙에 따라 그룹을 나누어 일관성 검사를 진행하였다. 또한 기존 연구에서는 일관성 검사까지만 진행하였지만, 기존의 연구에서 더 나아가 유출 패스워드의 Hashcat [23], Zxcvbn [25]을 이용하여 패스워드 안전성 검증 체계를 추가하여 패스워드 미터의 정확성을 측정하였다. 이에 본 연구는 기존 연구와 차별성을 가진다.

III. 연구방법

3.1 연구목적

유출 패스워드 분석 및 기존 연구들에서 밝힌 바와 같이 사용자들은 여전히 약한 패스워드를 사용하고 있다. 이에 본 연구에서는 “국내 상위 접속 웹 사이트의 패스워드 미터는 본연의 역할을 정확히 하고 있는가?”의 연구 질문을 가졌다. 세부적인 질문으로 “국내 상위 접속 웹 사이트의 패스워드 미터는 일관된 강도를 측정하고 있는가?”와 “국내 상위 접속 웹 사이트의 패스워드 미터는 정확한 강도를 측정하고 있는가?”를 가진다.

연구 질문에 따라 국내 상위 웹 사이트의 문제점을 밝히며, 패스워드 미터의 발전 방향을 제시하고자 한다. 패스워드 미터의 일관성, 안전성을 측정하기 위하여 다음과 같이 국내 접속 상위 웹 사이트와 패스워드를 선정한다.

3.2 웹 사이트 선정

본 연구의 실험 대상 웹 사이트는 Alexa [21]에서 제시한 사이트 중 상위 접속 웹 사이트를 선택하였다. Alexa 인터넷 주식회사는 미국 아마존의 자회사로, 인터넷 사이트 별 트래픽 조사기관이다. 웹 사이트 순위는 한 달간 하루 평균 방문자 수와 페이지 뷰, 이용 평균수를 바탕으로 측정된다. 이에 따라 본 논문에서 선정한 웹 사이트는 Naver, Daum, Gmarket, 11st, Clien, Auction, Saramin, Nate이다. 패스워드 미터가 존재하는 8개 웹 사이트에 대하여 패스워드 미터의 일관성, 정확성을 측정한다.

3.3 패스워드 선정

실험에 사용한 패스워드는 2009년 피싱 공격으로 유출된 Hotmail 패스워드 리스트를 이용한다. 패스워드를 선정하기에 앞서 다음과 같은 규칙에 따라 3개의 그룹으로 나누어 그룹별 80개, 총 240개의 패스워드를 선정하였다. 각 그룹에 대한 설명은 다음과 같다.

- Class: 대문자, 소문자, 숫자, 특수문자의 집합
- 규칙 1: 패스워드를 구성함에 있어서 하나의 class

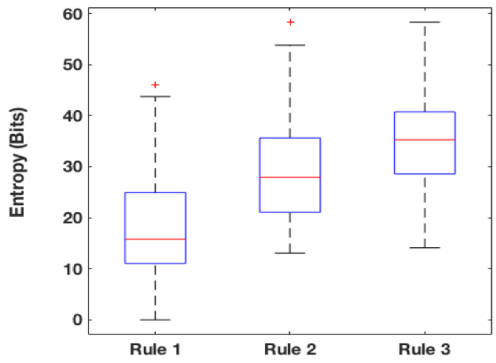


Fig. 2. Entropy bits of password rules. (Minimum entropy bits of Zxcvbn)

사용

- 규칙 2: 패스워드를 구성함에 있어서 두개의 class 사용
- 규칙 3: 패스워드를 구성함에 있어서 세개의 class 사용

규칙 1 보다 규칙 2가 조금 더 까다로운 패스워드 생성 정책에 의해 생성된 패스워드, 규칙 2 보다 규칙 3이 좀 더 까다로운 패스워드 생성 정책에 의해 생성된 패스워드로 각 규칙에 대하여 의미를 부여 할 수 있다.

Fig. 2.는 본 논문에서 선정한 규칙에 대한 그룹의 엔트로피 측정 결과이다. 규칙 3으로 갈수록 엔트로피의 Median, Lower Extreme, Upper Extreme 값이 증가하고 있음을 확인할 수 있다. 엔트로피는 Shannon에 의해 만들어진 개념으로 불확실성(Uncertainty), 메시지 당 평균 정보량을 의미하며, 다음과 같은 수학적 특성이 있다.

- 불확실성은 0보다 크거나 같은 수로 대응되어야 한다.
- 불확실성이 클수록 그 수는 커져야 한다.
- 독립적 확률 변수의 불확실성을 수치화하면 각각의 불확실성을 수치화 한 값들을 더해서 표현 가능하다. Shannon 엔트로피는 일반화되어 다양한 엔트로피의 기본이 되었고, Empirical 엔트로피, NIST 엔트로피 등으로 변형되어 사용된다. 따라서 규칙이 어려워짐에 엔트로피가 증가하는 것은 해당 규칙 패스워드의 불확실성, 메시지 당 평균 정보량은 증가하고 있음을 의미한다.

Table 3. Cracking results via Hashcat and Zxcvbn.

	#	Hashcat	Zxcvbn
Rule 1	80	63 (78.75%)	57 (71.25%)
Rule 2	80	33 (41.25%)	26 (32.5%)
Rule 3	80	12 (15%)	13 (16.25%)

(Experimental groups are the passwords that would take less than a day to be compromised based on Zxcvbn estimation)

3.4 패스워드 안전성 검증

Hashcat, Zxcvbn을 이용하여 패스워드 안전성 검증을 진행한다. Hashcat은 빠른 속도를 장점으로 하여 해쉬 암호화 된 패스워드를 크래킹 하는 툴로 많은 사람들이 사용하고 있는 툴이다 [23]. Zxcvbn은 패스워드 강도를 측정함에 있어서 키보드 배열, 영어 사전 등 다양한 요소를 고려하여 사용자가 생성하는 패스워드에 대한 강도를 측정하는 것으로 Dropbox에서 개발하였다 [25]. Zxcvbn은 크래킹 툴이 아니기에 자체에서 제공하는 예상 크래킹 시간을 바탕으로 크래킹을 측정한다. 예상 크래킹 시간은 즉시(instant), 분, 시간, 일, 년 순으로 나타남에 따라 1일 이하 예상 크래킹 시간을 가지는 패스워드를 크래킹 된 패스워드라 선정하였다. Table 3은 Hashcat, Zxcvbn을 이용한 크래킹 결과이다. 규칙 1의 경우 Hashcat은 78.75%, Zxcvbn은 71.25%에 해당할 만큼 크래킹이 됨에 따라 약한 규칙임을 알 수 있다. 규칙이 어려워짐에 따라 크래킹 되는 패스워드는 점점 줄어드는 것을 확인 할 수 있다.

IV. 웹 사이트 패스워드 미터 분석 결과

4.1 패스워드 미터 표현방법

패스워드 미터는 Fig. 3.과 같이 다양한 방법으로 표현된다. 하지만 웹 사이트 간 서로 다른 표현 및 통일 되지 않은 단어를 선택하여 사용자들에게 피드백 할 수 있고, 통일 되지 않은 스케일 등 문제점이 존재한다. 스케일은 패스워드 강도의 단계를 구분하는 것으로, 강도가 높을수록 높은 스케일에 속한다.

Daum의 패스워드 미터의 경우, 안전 혹은 완벽한 패스워드의 경우 동그라미 표시와 함께 표현을 하고

Web Site	Scale	Standard Scale	Password Meter Figure
Clien	6	-	아주안좋습니다 / 안간안좋습니다 / 조금괜찮습니다 / 양호수준입니다 / 좋은암호입니다 / 아주괜찮습니다
Naver	4	1	사용불가 적정 / 낮음 높음
Daum	4	2	조금 더! 비밀번호는 8자 이상이에요! / 유추하기 쉬운 비밀번호는 사용할 수 없어요! / 안전한 비밀번호입니다. 바로 사용하세요! / 완벽한 비밀번호입니다. 바로 사용하세요!
Nate	4	1	사용불가 / 비밀번호는 6자리 이상으로 입력해 주세요. / 보안상태낮음 / 좀 더 안전한 비밀번호로 설정해 주세요. / 보안상태안전 / 사용가능한 안전한 비밀번호 입니다. / 보안상태높음 / 예측이 어려운 강력한 비밀번호 입니다.
Gmarket	4	2	비밀번호 조합기준에 적합하지 않습니다. / 보안에 매우 취약하여 사용할 수 없습니다. / 적정 수준의 안전한 비밀번호입니다. / 매우 안전한 비밀번호 입니다.
11st	2	1	사용 불가 / 사용 가능
Auction	4	2	사용불가 취약 적정 안전
Sarmain	4	2	사용불가 (안전성 강도 매우 약함) / 사용불가 (안전성 강도 약함) / 사용가능한 비밀번호입니다. (안전성 강도 보통) / 사용가능한 비밀번호입니다. (안전성 강도 매우 강함)

Fig. 3. Scales and expression styles of password meters from different web sites.

있지만 그 외의 강도에서는 문장으로만 표현하고 있는 통일 되지 않은 표현하고 있다. 또한 단어 선택에 있어서 안전과 완벽이라는 단어의 애매함이 존재한다.

Clien의 경우 아주, 조금과 같은 수식어를 사용하여 표현 하였지만, 사용자 개인마다 이해하는 기준이 달라 애매한 의미로 해석될 수 있다.

Naver의 경우 서로 다른 스케일을 가지고 있는 경우에도 동일한 색상으로 표현을 하고 있고, 적정이라는 모호한 강도를 의미하는 단어를 사용하고 있는 문제를 가지고 있다.

4.2 패스워드 미터 일관성 측정 결과

패스워드 미터 일관성 측정에 있어서 패스워드 생성 정책을 만족하는 패스워드 규칙에 대하여 일관성 있게 측정하는 지를 알아본다. 또한 동일한 규칙을 적용하였을 경우, 각 웹 사이트의 패스워드 미터가 비교적 일관되게 측정하고 있는 지 알아본다.

Fig. 4.는 조사 대상 웹 사이트 7개에 대하여 규칙별 패스워드 미터 강도 측정 결과이다. 패스워드 미터 일관성 측정 결과, 동일 규칙 패스워드 그룹 내에서도 웹 사이트에서는 서로 다른 강도로 판정하고 있음을 확인 할 수 있었으며, 동일한 패스워드 생성 정책을 가진 웹 사이트 간에도 차이가 남을 알 수 있다.

Fig. 4.(a).은 규칙 1 그룹의 결과이다. 규칙 1에 해당하는 패스워드에 대하여 Naver에서는 Lower Extreme, Upper Extreme의 값이 많이 차이가 남을 알 수 있다. 웹 사이트 Daum의 경우, 규칙 1에서 중앙값이 가장 높게 측정됨을 알 수 있다. Nate의 경우 규칙 1에 해당하는 패스워드는 전혀 사용할 수 없는 결과를 보인다. 이에 규칙 1 그룹에 대하여 Naver는 일관되지 않은 결과로 표현하고 있다.

Fig. 4.(b).은 규칙 2 그룹의 결과이다. 규칙 2에서는 웹 사이트 11st은 중앙값이 100%에 해당하여, 모든 패스워드가 강한 패스워드로 측정하고 있다. 하지만 11st, Gmarket, Auction의 경우 동일한 패스워드 생성을 가진 웹 사이트이지만, 서로 다른 결과를 보여주고 있는 것을 확인하였다. 또한 규칙 2에서는 Nate가 상대적으로 일관되지 않은 결과가 나옴을 확인하였다.

Fig. 4.(c).은 규칙 3 그룹의 결과이다. 중앙값은 다른 그룹에 비하여 대부분 70%이상에 위치하고 있음을 확인 할 수 있지만, 이상 값이 상대적으로 많이 발견되었다. 규칙 3의 경우 Clien, Naver가 상대적으로 일관되지 않은 결과가 나옴을 확인하였다.

Fig. 4.를 종합한 결과 패스워드 생성 정책에 따른 패스워드 그룹 내에서도 서로 다른 결과를 측정하고 있음을 확인하였다. 또한 동일한 패스워드 생성 정책을 가진 경우에도 각 웹 사이트에서는 서로 다른 결과를 측정하기도 하며, 동일 그룹 내에서도 각 웹 사이트에서는 일관되지 않게 패스워드 강도를 측정하고 있다.

따라서 패스워드 미터 일관성 측정 결과 동일한 패스워드 생성에 따른 패스워드들이 서로 다른 강도로 판정되고 있다. 일관되지 않은 결과에 따라 패스워드 미터는 사용자들에게 혼동을 야기할 수 있다. 그 이유는 사용자들은 패스워드를 재사용하고 있기 때문이다. 2010년 Shay 등의 연구에 따르면, 사용자들은 약 70%정도 재사용하고 있음을 밝혔고[16], 2014년 Das의 연구에서는 설문 참여자 77%가 재사용하고 있음을 밝혔으며 [3], 2014년 Stobert의 연구에서는 심층 인터뷰 참여자 96%가 패스워드를 재사용하고 있음을 밝혔다[17]. 많은 사용자들이 패스워드를 재사용함에 따라 자신이 사용하고 있는 패스워드의 강도를 정확히 파악하지 못할 가능성이 크다.

혼동을 불러일으키는 근본적인 문제는 웹 사이트 패스워드 미터의 스케일과 내부 점수 함수가 상이하다는 것이다. 조사대상 웹 사이트의 경우 패스워드 미터 스케일은 최소 2부터 최대 6까지 존재한다. 이와

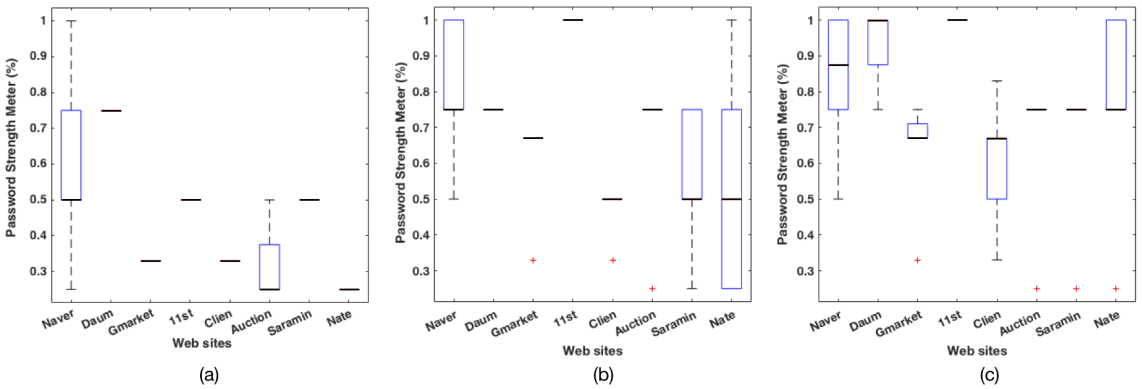


Fig. 4. Results produced from different password meters with the same password rules (consistency-test).

더불어 패스워드 강도를 측정함에 있어서 Fig. 3.에서 제시한 기준점이 되는 스케일이 다르기도 하며, 심지어 기준점이 되는 스케일이 존재하지 않는 웹 사이트도 존재한다. 본 논문에서 이야기하는 기준점 스케일은 패스워드 사용 가능·불가의 기준이 되는 것을 의미한다. 또한 패스워드 강도를 측정하기 위한 내부 함수가 웹 사이트마다 상이하기 때문에 패스워드 강도를 일관성 있게 측정하지 못하고 있다. 따라서 일관된 패스워드 강도 측정하기 위해서는 패스워드 미터의 표준 규격의 필요함을 주장한다.

4.3 패스워드 미터 정확성 측정 결과

패스워드 미터 안전성 측정은 Hashcat, Zxcvbn을 통한 패스워드 검증 결과를 바탕으로 하여, 해당 패스워드 사용 가능·불가능 기준 스케일에 따라 선정 웹 사이트의 오판에 대해 알 수 있다. 사용 가능·불가능의 기준 스케일은 웹 사이트 간 서로 다르다. 기준이 되는 스케일 이하의 패스워드는 사용이 불가능, 기준 스케일보다 큰 패스워드는 사용 가능한 패스워드다. 웹 사이트의 패스워드 미터 정확성 측정은 크래킹 저항 강도가 전혀 없는 패스워드에 대하여 안전하다고 판정하는 경우, 크래킹 저항 강도가 있음에도 불구하고 약한 패스워드라 판정하는 경우가 있다. 전자의 경우 사용자가 패스워드를 생성함에 있어서 크래킹에 약한 패스워드의 경우에도 안전한 패스워드로 판정하기 때문에 더욱 큰 문제가 된다.

Fig. 5.와 Fig. 6.은 패스워드 미터 정확성 측정 결과이다. Fig. 5.를 통해 규칙에 따라 웹 사이트에서 오판하는 경우 각각에 대해 알 수 있다. Fig. 6.을 통해서 규칙이 강화됨에 따라 웹 사이트에서는 오판하는 경우가 현저하게 줄어들고 있음을 알 수 있다. 하지만 규칙이 강화됨에도 불구하고 여전히 오판하는 경우가 발생하고 있다. 규칙 1(Fig. 5.(a))의 경우 평균적으로 77.3%로, 패스워드 그룹에 대하여 오판하고 있다. 최대로 91.3%의 패스워드 그룹에 대하여 오판하는 웹 사이트도 존재한다. Auction의 경우, 패스워드에 대해 상대적으로 다른 웹 사이트에 비해 정확하게 판단하고 있다.

규칙 2(Fig. 5.(b))의 경우 규칙 1에 비해 오판하는 경우가 적음을 알 수 있다. 규칙 2에서는 평균적으로 54.2%로, 패스워드 그룹에 대하여 오판하고 있다. 규칙 2에서 최대로 56.3%로, 패스워드 그룹에 대하여 오판하는 웹 사이트가 존재한다. 규칙 2에서는 11st 웹 사이트에서 다른 웹 사이트에 비해 오판하는 경우가 상대적으로 적지만, 50% 이상의 패스워드에 대하여 오판하고 있음에 정확하게 판단하고 있다고 보이지 않는다.

규칙 3(Fig. 5.(c))의 경우 규칙 2에 비해 오판하는 경우가 적음을 알 수 있다. 규칙 3에서는 평균적으로 24.5%로, 패스워드 그룹에 대하여 오판하고 있다. 규칙 3에서 최대 30%로, 패스워드 그룹에 대하여 오판하는 웹 사이트가 존재한다. 규칙 3에서는 Naver, Daum이 타 웹 사이트에 비해 오판하는 경우가 상대적으로 적음을 알 수 있다.

패스워드 미터 본연의 역할은 강한 패스워드를 만들게 유도하는 것 뿐 아니라 사용자가 생성하고자 하는 패스워드에 대하여 정확한 판정을 내리는 역할도 있다. 오판에 따라 사용자들은 자신이 사용하고 있는 패스워드에 대하여 강한 판정을 받겠지만, 실질적으

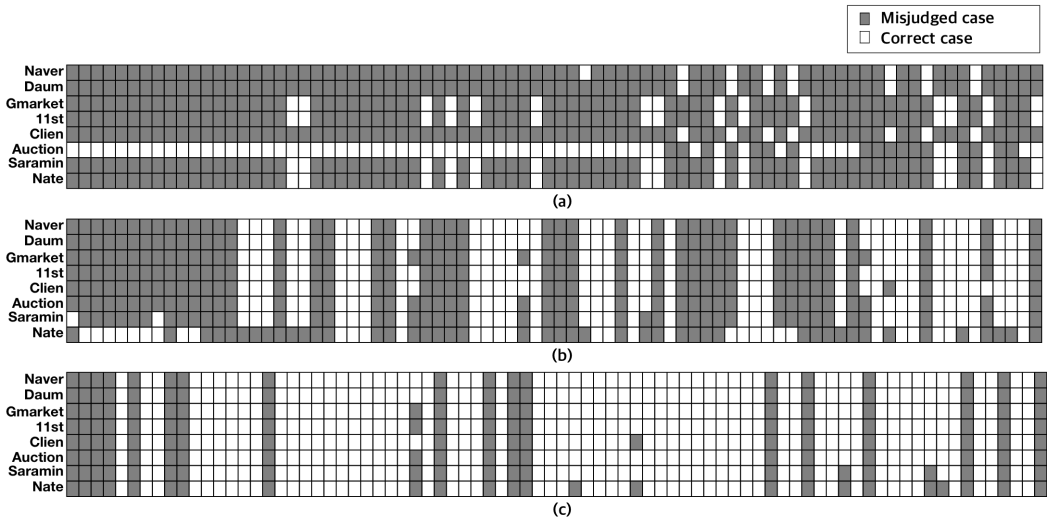


Fig 5. Accuracy test results of password meters in accordance with the password rules. (a) Rule1, (b) Rule2, (c) Rule3.

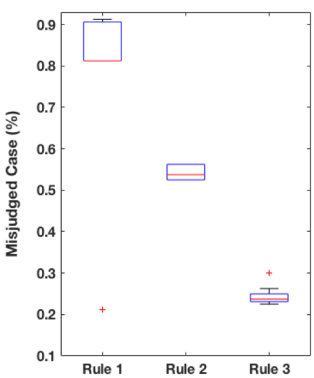


Fig. 6. Accuracy test of results (graphs) of password meters in accordance with the password rules.

로 강하지 않은 패스워드인 경우가 발생할 수 있다. 오판을 방지하기 위하여 패스워드 강도 측정 함수인 점수 함수(Scoring Function)에 크래킹 측정 관련 함수가 추가되어야 하며, 뿐만 아니라 leet 변환, 엄격한 사전 검사 등 다양한 검사 등이 반영되어야 한다. 점수 함수 내에 크래킹 측정 함수를 통해 더욱 안전한 패스워드를 사용자들이 생성 할 수 있을 것이며, 웹 사이트에서는 사용자들이 생성하고 있는 패스워드에 대하여 한층 더 정확한 패스워드 강도 측정을 할 수 있을 것이다.

V. 패스워드 미터 토의 및 제언

본 연구 결과, 강한 패스워드를 생성하도록 유도해야 하는 패스워드 미터에서 일관되지 않은 표현, 일관되지 않은 강도 판정, 정확하지 않은 강도 판정의 문제가 존재함을 알 수 있다. 이에 패스워드 미터에 대하여 다음과 같이 제안한다.

- 패스워드 미터도 패스워드 생성 정책의 가이드 라인 NIST[10], KISA[18]에서 발표한 내용과 같이 표준 표현 방법 가이드 라인을 만들어야 함을 주장한다. 표준화된 표현 방법을 이용하면 사용자들이 사용함에 있어서 통일된 시각적 표현방법과 애매하지 않은 단어를 선택함에 따라 사용자들은 더욱 더 강한 패스워드를 생성 할 것이다.
- 패스워드 강도를 측정하는 내부함수에 본 연구의 사전연구에서 밝힌 패스워드가 특정 구조에 집중되는 현상을 막기 위하여, 패스워드 구조를 검사하는 함수가 추가되어야 한다. 이와 관련한 함수는 생성하고자 하는 패스워드에 대한 청크 기반 구조 검사이다. 생성하고자 하는 패스워드에 대하여 청크 기반의 구조 구분하여, 사용자들이 많이 사용하는 패스워드 구조에 대해서는 약한 강도로 피드백하는 기능이 포함되어야 한다. 사용자들 개개인 마다 서로 다른 방법으로 패스워드를 생성하지만, 결과적으로는 유사한 구조를 가지고 있기

때문에 이를 방지하기 위하여 다양한 구조로 생성할 수 있도록 해야 한다.

- 패스워드 강도를 측정하는 내부함수에 크래킹 저항성 측정 관련 함수가 포함되어야 한다. 사용하고자 하는 패스워드 전체 길이에 대한 크래킹 검사를 하기 에는 오랜 시간이 걸리기 때문에 8자리에서 10자리 사이의 패스워드에 대해서, NIST에서 제시하고 있는 5만개 단어가 포함되어 있는 패스워드 사전을 이용하여 검사하는 함수로 구성되면 될 것이다. 제한된 패스워드 길이에 대해서 검사를 할 경우에 미처 발견하지 못하는 패스워드가 있을 수 있지만, 패스워드 사전으로 이는 보완 할 수 있다. 패스워드 크래킹과 관련한 함수를 포함할 경우, 패스워드가 유출되었을 경우에도 안심할 수 있는 기반을 마련할 수 있다.
- 규칙기반의 패스워드 미터 환경의 한계점을 극복하기 위하여 사용자가 생성하고 있는 패스워드에 대한 다양한 측정 방법이 마련되어야 한다. 사용자가 패스워드를 생성함에 있어서 기억에 의존하기 때문에 기존에 사용하고 있던 패스워드를 재사용 하는 행동은 많이 알려진 사실이다. 이에 패스워드 미터 내부함수에서는 기계학습을 통해 기존에 사용자가 사용했던 패스워드 생성 패턴 혹은 기존에 사용했던 리스트를 이용하여 자주 사용하는 패턴 또는 단어에 대해서 약한 강도의 패스워드로 판단하거나 좀 더 다른 패턴의 패스워드로 생성할 수 있도록 유도하는 함수가 추가되어야 한다.

이와 같은 기능을 패스워드 미터 내 점수 함수에 포함을 하면, 사용자는 더욱 더 강한 패스워드를 생성 할 것으로 예상한다. 내부 함수에서 측정한 각 항목의 결과를 피드백하면 사용자들은 충분히 수긍할 것이며, 패스워드 미터 자체는 일관되고, 정확한 결과를 측정할 수 있는 기능이 될 수 있을 것이다. 하지만 사용자가 패스워드를 생성함에 있어서 너무 강한 패스워드 미터를 제시 할 경우, 사용성이 많이 떨어질 수 있을 것이다. 이에 대해서는 패스워드 미터의 역할에 추가적으로 패스워드 추천 시스템을 도입하게 되면, 사용자가 원하는 단어를 이용하여 크래킹에도 저항가능 하며, 기존 청크구조에서 약간의 변형을 통해 기억하기에 무리 없는 환경을 제공할 수 있을 것이다. 이에 대해서는 후속 연구로 진행되어야 한다.

VI. 관련연구

패스워드는 시스템 보안 중 하나의 구성 요소이자 필수요소다 [9]. 사용자들은 텍스트 패스워드를 이메일, 온라인 커뮤니티 등에서 많이 사용하고 있으며 [4], 다양한 패스워드 입력 방법 중 텍스트 패스워드 인증 방법은 과거부터 현재까지 많이 사용하고 있는 방법이다. 하지만 여전히 사용자들은 약한 패스워드를 사용하고 있는 문제점이 있다. 사용자들은 단순한 패턴으로 자신만의 방법으로 키보드 배열 패턴, 자신과 관련한 단어, 숫자를 사용하여 유추하기 쉬운 패스워드를 사용하고 있는 문제점이 있다 [14]. 사용자들은 패스워드 내 사랑과 관련한 단어, 성(Sexual)과 관련한 단어, 음식, 동물, 돈과 관련한 알기 쉬운 단어를 사용하고 있는 문제점을 밝혔다 [11]. 또한 사용자들은 사용하고 있는 패스워드를 다중 계정에서 재사용하고 있는 문제점이 있다 [3]. 사용자들은 하나의 패스워드로 평균적으로 6개의 사이트에 재사용하고 있고 [5], 기억하기 쉽기 때문에 재사용하고 있음을 밝혔다 [7]. 사용자들은 여전히 무작위 공격과 사전 공격 등에 노출이 될 수 있는 취약점을 가진 약한 패스워드를 생성하여 사용하고 있으며, 하나의 패스워드를 재사용하고 있는 문제점을 가지고 있다.

패스워드 생성 시, Morris 등의 연구에서는 사용자는 패스워드 생성 시 길이가 긴 혹은 다양한 종류의 문자사용 해야 함을 주장하고 있다 [9]. 강한 패스워드 생성을 유도하기 위하여 서비스 제공자들은 최소 길이 제한 및 대소문자, 숫자, 특수문자를 이용하여 패스워드를 생성하도록 패스워드 생성 정책을 적용하고 있다. 하지만 웹 사이트에서는 서로 다른 요구 사항을 제시하고 있으며, 보안성을 고려하여 패스워드 생성 정책을 어렵게 만들면 사용자들의 편리성 및 사용성에 문제점을 유발한다. 2015년 Ur 등의 연구에 따르면 이러한 문제점에 대하여 패스워드 생성 정책은 보안성과 사용성 모두 고려한 정책을 만들고, 정책 기반한 패스워드 생성 시 공격자로부터 방어함과 동시에 사용자가 기억하기 쉽게 해줘야 한다고 주장한다 [14]. 또한 사용자들은 강력한 패스워드 생성 정책을 제시하였을 경우, 패스워드 생성에 있어서 어려움을 겪고 있고, 어려운 생성 정책을 제시하면 사용자들은 좌절감을 느끼는 문제점이 있다 [19]. 사용자들은 어려운 생성 정책이 제시되면 불편함을 느끼지만, 복잡한 패스워드를 이용하면 자신의 계정과 패스워드에 대하여 안전해 졌다고 믿고 있음을 밝혔다 [16].

NIST(National Institute of Standards and Technology)에서는 패스워드 생성 및 관리에 관하여 문서화하여 가이드를 제시하고 있다 [10]. 서비스 제공자뿐만 아니라 2008년 한국인터넷진흥원(KISA)에서는 안전한 패스워드 설정 방법, 패스워드 보안 지침 등을 소개하는 "패스워드 선택 및 이용 가이드" 배포를 통해 안전한 패스워드 생성 및 보안 지침을 제시하고 있다 [18].

강한 패스워드를 만들게 하기 위하여 패스워드 생성 정책과 더불어 패스워드 미터를 이용하여 사용자들에게 패스워드 강도를 피드백 한다. 패스워드 미터는 지시자로 다양한 종류의 패스워드 미터 지시자가 존재하며 [13], 사용자가 입력하는 패스워드에 대하여 내부 점수 함수를 이용하여 입력한 패스워드의 강도를 보여준다. 패스워드 강도를 측정함에 있어서 규칙 기반 패스워드 미터, *n*-gram을 이용한 Adaptive 패스워드 강도 미터 (APSM), *PCFG*를 이용한 Analyzer and Modifier 패스워드 미터 (AMP)가 있다 [2]. 2012년 Ur 등의 연구에서 엄격하게 패스워드를 선택하도록 하는 미터들은 확실히 사용자들이 더 많은 숫자, 특수문자, 대문자들을 포함하여 더 긴 패스워드를 생성하게 한다는 것을 보였다 [13]. 또한 시각적 표현과 텍스트 표현을 같이 사용하여 사용자에게 피드백을 할 경우, 사용자는 더 안전한 패스워드를 생성함을 밝혔으며 [13], 패스워드 미터의 표현 방식에 따라 사용자가 느끼는 차이점 또한 밝혔다 [15]. 패스워드 미터는 사용자들에게 더 긴 패스워드를 사용하게 유도하고 있다 [12]. 하지만 중요하지 않은 계정의 패스워드에 대해서는 패스워드 미터는 영향이 미미한 결과를 보이기도 했다 [15]. 패스워드 미터의 본연의 역할은 사용자에게 강한 패스워드를 만드는 역할이다. 하지만 2014년 Carnavalet과 Mannan의 연구에서는 동일 패스워드에 대하여 서로 다른 미터에서 서로 다른 강도로 나오는 문제가 있다 [1]. 서로 다른 패스워드 미터 결과는 패스워드 미터의 강한 패스워드를 선택함에 있어서 혼란을 줄 수 있으며, 패스워드 미터 본연의 목적을 훼손시킬 수 있는 문제점을 밝혔다.

VII. 결 론

본 논문은 국내 접속 상위 웹 사이트의 패스워드 미터에 대한 연구를 진행하였다. 기존연구와 차별하여 규칙 기반 패스워드 그룹을 이용한 패스워드 미터 일관성 측정을 하였고 추가적으로 Hashcat, Zxcvbn을 통한 안전성 검증을 통한 패스워드 미터의 정확성 측정을 하였다. 연구 결과, 웹 사이트에서는 일관되지 않고, 정확하지 않은 결과를 측정하고 있음을 알 수 있었다.

이에 본 연구에서는 패스워드 미터에 대한 표준화된 가이드라인을 제시할 것을 주장하며, 패스워드 미터 내부 함수 개선에 대해 논의 하였다.

References

[1] X. de C. de Carnavalet and M. Mannan, *"From Very Weak to Very Strong: Analyzing Password-Strength Meters,"* In Proc. of NDSS, Interent Society, 2014.

[2] D.J. Gusaas ,*"Password Strength Meters: Implementations and Effectiveness,"* In Proc. of Csci, Dec. 2015.

[3] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. F. Wang, *"The Tangled Web of Password Reuse,"* In Proc. of NDSS, Vol. 14, pp. 23-26, Feb. 2014.

[4] H. Eiji, and H. Jason I, *"A Diary Study of Password Usage in Daily Life,"* In Proc. of SIGSCHI, ACM, pp. 2627-2630, May. 2011.

[5] D. Florencio and C. Herley, *"A Large-Scale Study of Web Password Habits,"* In Proc. of WWW, pp. 657-666, May. 2007.

[6] S. Furnell, *"Assessing password guidance and enforcement on leading websites,"* In Proc. of Computer Fraud&Security, 2011(12), pp. 10-18, Dec. 2011.

[7] S. Gaw and E. W. Felten, *"Password Management Strategies for Online Accounts,"* In Proc. of SOUPS, pp. 44-55, July. 2006.

[8] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. V. L. Bauer, N. Christin, L. F. Cranor, and J. Lopez , *"Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms,"* In Security and Privacy on IEEE, pp.523-537, May, 2012.

[9] R. Morris and K. Thompson, *"Password*

Security: A Case History," In Proc. of ACM, 22(11), Nov. 1979.

〔10〕 Scarfone, Karen, and M. Souppaya, *"Guide to Enterprise Password Management."* NIST Special Publication 800-118, 2009.

〔11〕 R. Veras, C. Collins, and J. Thorpe, *"On the Semantic Patterns of Passwords and their Security Impact,"* In Proc. of NDSS, 2014.

〔12〕 B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, *"Helping Users Create Better Passwords,"* In Proc. of USENIX, 2012.

〔13〕 B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Aranor, *"How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation,"* In Proc. of USENIX Security, 2012.

〔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,"* In Proc. of SOUPS, pp. 123-140, July. 2015.

〔15〕 E. Serge, S. Andreas, M. Ildar, B. Konstantin, and H. Cormac, *"Does my password go up to eleven?: the impact of password meters on password selection."* In Proc. of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 2379-2388, 2013.

〔16〕 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 Behaviros,"* In Proc. of SOUPS, p.2, July. 2010.

〔17〕 Stobert, Elizabeth, and Robert Biddle. *"The password life cycle: user behaviour in managing passwords."* In Proc. SOUPS. pp. 243-255, July. 2014.

〔18〕 방송통신위원회, KISA, *"패스워드 선택 및 이용 안내서," KISA 안내·해설 제2010-22호.*

〔19〕 S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, *"Of Passwords and People: Measuring the Effect of Password-Composition Policies,"* In Proc. of CHI, pp. 2595-2604, May, 2011.

〔20〕 M. Weir, S. Aggarwal, M. Collins and H. Stern, *"Testing metrics for password creation policies by attacking large sets of revealed passwords,"* In Proc. of CCS, pp. 162-175, Oct. 2010.

〔21〕 Alexa website, *http://www.alexa.com/topsites*

〔22〕 Relative frequencies of letters in text, *Wikipedia. https://en.wikipedia.org/wiki/Letter_frequency*

〔23〕 Hashcat, http://hashcat.net/hashcat/

〔24〕 Leaked Password Lists, Skullsecurity. https://wiki.skullsecurity.org/index.php?title=Passwords,

〔25〕 Dropbox TechBlog, zxcvbn: realistic password strength estimation. https://blogs.dropbox.com/tech/2012/04/zxcvbn-realistic-password-strength-estimation

〈저자소개〉

김 경 훈 (KyoungHoon Kim) 학생회원
2015년 2월: 성공회대학교 컴퓨터공학 학사
2015년 3월~현재: 연세대학교 정보대학원 석사과정
〈관심분야〉 Authentication, HCI 보안 등

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〉 암호프로토콜, 네트워크 프로토콜, 사물인터넷 보안, HCI 보안 등