Research Article

Enhanced Evaluation Model of Security Strength for Passwords Using Integrated Korean and English Password Dictionaries

Ki Hyeon Hong ¹, Un Gu Kang ², and Byung Mun Lee ²

¹Department of IT Convergence Engineering, Gachon University, Seongnam 13120, Republic of Korea

²Department of Computer Engineering, Gachon University, Seongnam 13120, Republic of Korea

Correspondence should be addressed to Byung Mun Lee; bmllee@gachon.ac.kr

Received 25 May 2021; Revised 24 August 2021; Accepted 6 September 2021; Published 22 September 2021

Academic Editor: Youwen Zhu

Copyright © 2021 Ki Hyeon Hong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the field of information security, passwords are a means of authenticating users. Passwords with weak security cannot perform the role of user authentication and personal information protection because confidentiality is easily violated. To ensure confidentiality, it is important to evaluate the strength of the password and choose a very secure password. Due to this fact, security evaluation models for various passwords have been presented. However, existing evaluation models evaluate security based on the English alphabet. Passwords depend on the memory of the user and are closely related to the language or environment used by the user. In this regard, there are limitations in applying the existing security evaluation models to passwords chosen by non-English speakers. We compose a non-English, Korean language-based password dictionary and propose a password security evaluation model based on this for Korean users. In addition, to verify the effectiveness of the proposed model, we conducted experiments to evaluate the security of Korean language-based passwords using a database of passwords that have been actually leaked. As a result, the proposed model showed 99.38% accuracy for Korean language-based leaked passwords. This is superior to the 80.06% accuracy shown by the existing model. In conclusion, the use of the Korean language-based password security evaluation model proposed in this paper will contribute to choosing more secure passwords for Korean language-based sites or users.

1. Introduction

Password-based user authentication methods are a powerful means of protecting the private information of users in the field of information security [1, 2]. When choosing a password, security and convenience conflict [3, 4]. For example, although complex passwords that combine numbers and special symbols like “toeq39bf@la” are inconvenient to use because they are difficult to remember, they are safe because they are not easily inferred by others. On the contrary, patterns that combine words in an English dictionary with consecutive numbers, such as “password123,” are easy to remember but can be inferred from a dictionary, making them unsuitable for authentication functions due to low security. Therefore, user convenience and security must be considered when choosing passwords. Existing sites or systems evaluate the security of easy-to-remember

passwords that the user selected and provide assistance in enhancing security through these analysis results [5, 6].

The simplest way to evaluate the security of user-selected passwords is to ensure that they do not choose already leaked passwords [7]. By collecting leaked passwords and building a database, it is possible to verify how many times the chosen password has been leaked. However, this approach to security evaluation is difficult to apply in low-spec environments, such as web browsers and small IoT devices, since it requires a very large set of leaked passwords, and lot of effort is needed to collect leaked password sets. The security of the password chosen by the user can be predicted through password security evaluation indicators. This can be done by implementing a password security evaluation model [8].

However, examples of existing evaluation models show that a lot of weight has been placed on predicting the security of English alphabet-based passwords for English-

speaking users [9]. For example, the National Institute of Standards and Technology (NIST) security guideline, which is the most commonly used evaluation method, predicts security by finding password complexity with combinations of uppercase and lowercase letters, numbers, and special symbols [10]. In addition, the zxcvbn evaluation model, presented as an example in a password security evaluation study by the University of Concordia, uses a variety of elements, including keyboard arrangement and English dictionaries, for security prediction [7]. Users tend to choose passwords by adding or modifying letters, numbers, and special symbols in character strings such as words, sentences, and keyboard sequences they are familiar with, since they have to remember the password [11]. At this point, since the words or sentences on which the password is based are deeply related to the language of the country, existing English-based evaluation models are less reliable when evaluating the security of passwords chosen by non-English-speaking users. To address this, we need a new password security evaluation model supplemented to fit the local language.

Therefore, in this paper, we propose a non-English evaluation model for password security strength based on the Korean language. This model can be divided into the formation of a dictionary for Korean language-based passwords and the password security prediction process. If the linguistic characteristics of the Korean language are examined, characters are composed by combining *chosung*, *joongsung*, and *jongsung* (initial consonants, vowels, and final consonants). Since they can be combined in various forms, the number of combinations for making a character is very large, making security predictions difficult. However, on the keyboard array, initial consonants, vowels, and final consonants corresponding to *chosung*, *joongsung*, and *jongsung* can be converted into the 26 characters of the alphabet [12]. If this method is used, Korean-based passwords, for which security evaluation is difficult, can be formed into an alphabet type dictionary. Password dictionaries converted to alphabets are used to predict the security of passwords. Security is predicted by whether a word in the Korean language-based password dictionary was used, like the traditional zxcvbn, and the expected number of times it takes to infer a password by combining numbers and special characters based on the words in the dictionary. We confirm this through experiments comparing the security prediction reliability of the existing password evaluation model with the Korean language-based password security evaluation model proposed in this paper.

In Section 2 of this paper, we consider the importance of password security evaluation and the existing password evaluation models that are used and examine the tendencies of non-English speakers when choosing passwords. Section 3 presents a Korean language-based password security evaluation model, defines a password dictionary that will be used in this model, and presents criteria for the evaluation results. In Section 4, the data collection process and an experiment are carried out to verify the effectiveness of the evaluation model and the results are analyzed. Section 5 presents the conclusions.

2. Related Research

2.1. Password Cracking Process and Method. In general, users are classified by a process in which the user chooses and registers a safe password and presents the same password later to be authenticated, as shown in Figure 1. In Figure 1(a), after the user chooses a password that is determined to be safe, it is converted into a hash value by a hashing algorithm and stored [13]. In such cases, since an inverse function for a hashing algorithm does not exist, there is the advantage that the original data cannot be inferred from the transformed hash value [14]. Therefore, even if a hacker acquires the hash values, the password chosen by the user cannot be inferred. The user performs user authentication through Figure 1(b). If the hash values match, authentication is accepted because they are identical passwords, and if they do not match, authentication is rejected.

Since hackers do not know the password chosen by the user, they try every possible combination of cases for authentication, as shown in Figure 1(c). At this time, if the hash values do not match, the hacker fails to authenticate because an incorrect password was entered, but if the hash values match, successful authentication results. Hackers can crack all passwords through a brute-force attack, which is basically a method that involves trying all possible combinations of strings [15]. However, since the number of possible combinations increases with respect to the length of the password, hackers can try different methods to crack the password.

There is a method in which users take a word from the dictionary and add, delete, or modify a few characters to choose a password instead of using the word directly [16]. Since slightly modified strings made from words in dictionaries are more likely to be chosen as passwords with this method than completely randomly composed strings, it can lead to reduced security effectiveness [17, 18]. Hackers can exploit this by composing a word dictionary and attempting to crack passwords by slightly modifying the words [19]. These methods are called dictionary attack and modified dictionary attack, respectively, and this means that dictionary-based words are more likely to be cracked than random strings [20]. In addition, hackers try to crack more easily by exploiting these vulnerabilities [20].

Hackers can use various threat models to crack passwords [21, 22]. For example, trawling guessing is a model that notes that universal users are more likely to choose commonly used words as passwords, and most password security evaluations present an evaluation model for trawling guessing. Targeted guessing is a method in which data that take into account individual characteristics, such as personally identifiable information (PII), are used for cracking after the target subject is determined when guessing a password. Online guessing is a model that attempts to crack through the interface provided when using online services. Offline guessing is a method in which the hacker attempts to crack the target system directly [23]. In addition, there is a threat model that obtains passwords directly from the user, such as shoulder surfing. In this study, the threat

FIGURE 1: Registration, authentication, and password cracking process.

model will be limited to trawling guessing because it is a study of passwords commonly chosen by Korean language-based speakers.

Since hackers can ultimately attempt cracking in a variety of ways, choosing a password with high security after fully understanding the password security guidelines is not an easy task [24, 25]. Therefore, a password security model that can proactively evaluate the security of a password that will be used is needed.

2.2. Password Security Evaluation Models and Their Limits.

A password security evaluation model must take into account the tendencies of users to choose passwords to obtain good predictive performance. Currently, the most popular method is the one that reflects the requirements for the number of uses of lowercase and uppercase letters, digits, and symbols (LUDS), which assesses the complexity of passwords by measuring the length, combination, and continuity of uppercase characters, lowercase characters, numbers, and special symbols [26]. The complexity of passwords can take other conditions into account. One of them is the effective character length. This is a password security assessment method that considers the use of uppercase letters, lowercase letters, numbers, and special symbols to compose passwords [26]. For example, passwords consisting of only eight numbers have fewer possibilities to consider compared with those with the same eight characters that consist of uppercase characters, lowercase characters, numbers, and special symbols. Therefore, effective character length is an indicator of the expected number of attempts a hacker must make through a brute-force attack before the password can be cracked.

However, neither LUDS requirements nor effective character length take into account the tendencies of users when they choose passwords. For example, the dictionary

word “password” and the randomly composed string “kkpauwds” have the same security prediction score according to LUDS requirements or effective character length because each password consists of only eight lowercase characters that contain consecutive alphabet letters “ss,” “kk.” However, given the fact that hackers are more likely to attempt “password” first in a prior attack, it is difficult to predict reliable security scores with the LUDS requirements or effective character length.

Considering such a limitation, the zxcvbn password security evaluation model that utilizes word dictionaries can be applied for evaluation. A word dictionary consists of a list of words that users are expected to use as passwords, as shown in Table 1.

The zxcvbn model organizes a list of words in seven categories. The dictionary consists of the 30,000 most commonly used English words in Wikipedia, 30,000 American movies and TV programs, 30,000 commonly used password statistics, and 35,494 American family names and other names [27]. Furthermore, by adding a dictionary for keyboard input patterns, the possibility of users choosing a keyboard input-based password rather than a word-based password, such as “q1w2e3r4,” is considered. Security is predicted in five steps, 0 to 4, by combining word lists with numbers and special symbols. According to existing studies, when security was predicted for a list of frequently leaked and low-security passwords with five different evaluation models, only zxcvbn evaluated that the security was low for all passwords on the list, displaying the best performance [28].

However, since zxcvbn predicts security based on words in word dictionaries, it is limited in evaluating the security of words that are not in the dictionary. For example, although American family names and other names, and American movies and TV programs are predicted to have low security because they are in the dictionary, names and words

TABLE 1: List of English words used in the zxcvbn evaluation model.

List	Number
English words in Wikipedia	30,000
American movies and TV programs	30,000
Frequently used passwords	30,000
Family names frequently used by Americans	30,000
Male names frequently used by Americans	1,219
Female names frequently used by Americans	4,275
Keyboard patterns	—

frequently used by Koreans are predicted to be very secure. Therefore, the existing zxcvbn lacks reliability to provide password security for non-English speakers.

2.3. Password Selection Trends for Non-English-Speaking Users. Since users must remember passwords, they are influenced by their language environment. Existing studies have shown that passwords leaked from English-speaking and Czech-speaking users are different. When the security of leaked Czech-language passwords was predicted using existing evaluation models, there were cases in which they were incorrectly evaluated to have excellent security [27]. This can be seen as an error caused by the existing model not considering the Czech language environment. To address this, a password security evaluation method that composed a Czech-based password dictionary and applied it to the zxcvbn evaluation model was proposed and resulted in improving security [27]. The fact that the Czech words used in this process are based on the Czech alphabet makes it easy to merge with the existing password dictionary composed of the English alphabet.

In comparison, the Korean language has a distinctive language system that is different from the English alphabet [29]. Figure 2(a) shows an example of combining *chosung*, *joongsung*, and *jongsung* (initial consonant, vowel, and final consonant) to form one Korean character. *Chosung* and *jongsung* consist of 14 consonants and *joongsung* consists of 10 vowels. In addition, one character can be composed with only *chosung* and *jongsung* as shown in Figure 2(b).

In the Korean language, words are a combination of characters, which are composed of *chosung*, *joongsung*, and *jongsung*. For example, “비밀번호”, the Korean word for “password,” consists of the following 10 consonants and vowels {ㅂ, ㅌ}, {ㅍ, ㅍ, ㅍ}, {ㅂ, ㅌ, ㄴ}, {ㅎ, ㅏ}. Combining these consonants and vowels results in four characters that combine to form the single word “비밀번호”. In contrast, each of the 26 uppercase characters and 26 lowercase characters in the alphabet is one character. From the fact that words are composed of 52 characters, the Korean language can be considered to be different from the language system based on the alphabet. This means that existing password dictionary composition methods cannot be applied to Korean password dictionaries.

In this regard, a novel dictionary composition method for password security evaluation must be applied in an independent language system rather than an alphabet-based one. Therefore, in this study, a method of composing a

Korean language-based password dictionary and a password security evaluation model based on it is proposed.

3. Password Security Evaluation Model Based on the Korean Language

3.1. Design of a Password Security Evaluation Model Based on the Korean Language. The proposed Korean language-based password security evaluation model predicts security using a Korean password dictionary. As shown in Figure 3, the model can be divided into two processes: the process to separate the collected Korean word data and convert them into alphabets to compose a password dictionary, and the process to predict the security of the password chosen by the user and provide the analysis results.

Figure 3(a) shows the process of collecting Korean word data and building it into a password dictionary. In order to convert them to alphabet-based words, each character is separated into *chosung*, *joongsung*, and *jongsung* and each of these are converted into an alphabet character. Words converted to alphabet characters are then merged again and used for composing the dictionary. In predicting the security of passwords not only Korean language-based passwords but also existing English alphabet-based passwords must be considered. For example, since English passwords such as “password,” “q1w2e3r4,” and “iloveyou” are frequently used by non-English speakers as well as English speakers, existing password dictionaries and Korean-based password dictionaries must be merged. This is why Korean language-based passwords should be converted to alphabet-based passwords.

Figure 3(b) shows the process of evaluating and predicting the security of a password chosen by a user using the password dictionary that was composed by following the process in Figure 3(a). For security predictions, the algorithm for the zxcvbn evaluation model is used. However, if numbers or special symbols are inserted between words, like “pas123word,” the security is predicted to be high in conventional zxcvbn because it is determined that there are no words that coincide with the dictionary. Considering this, a process of extracting strings from user-selected passwords and then comparing them with the dictionary is added. Therefore, the password security prediction process must evaluate the security of passwords using a merged dictionary and verify whether they match the words in the dictionary. The security level and security analysis are produced like this and serve as a standard for determining the stability of passwords. However, the performance of the evaluation model can be influenced by how the password word dictionary is built and by the criteria used during the evaluation process.

3.2. Composition of Korean Language-Based Password Word Dictionary. Password security evaluation models should include passwords that are frequently used by Korean users. This means adding names or words that are frequently used by Koreans, like items 3, 4, and 5 in Table 2. Since English words can also be used, the word dictionary used by the

FIGURE 2: Korean language character composition method.

FIGURE 3: Korean language-based password security evaluation model.

existing evaluation model such as items 0, 1, and 2 in Table 2 are also included. Some research studies have found that 77.38% of passwords used are reused by users [30]. In addition, even if it is not reused as is, there is a tendency to avoid blacklists with slightly modifications [30, 31]. These passwords are targeted guessing threat model because individual characteristics must be taken into account. However, in this paper, it is more suitable for the trawling guessing model than the targeted guessing model targeted to a specific individual because a large number of unspecified Korean language speakers are targeted. Therefore, we configure the items in Table 2 according to the trawling guessing threat model.

3.3. Word Selection Criteria. The Korean words chosen in this paper are based on data disclosed by the government of the Republic of Korea, as shown in Table 3, and they were taken from the statistical data of the system [32] registered in the Korean court. Most are data of the most commonly used names in Korea and names that were registered before and after renaming, including ones that were registered at the time of birth or adoption after 2008. The dictionary was established by choosing the top 1,000 from each.

In addition, the most frequently used common noun words in Korean are included in the dictionary, based on classification by the National Institute of Korean Language

TABLE 2: Password dictionary list based on the Korean language.

Number	List
0	English words in Wikipedia
1	Frequently used passwords
2	Keyboard patterns
3	Male names frequently used by Koreans
4	Female names frequently used by Koreans
5	Noun words with high frequency in the Korean language

[33]. Of the 82,502, the top 13,785 with use frequency from 20 to 13,594 were selected and included in the dictionary. Words with a frequency of less than 20, including slang, that are not frequently used in everyday life were excluded from the dictionary. Although words were selected through this process, it is difficult to integrate and organize them in the dictionary because English and Korean characters have different structures. Therefore, a word conversion process is needed.

3.4. Word Conversion Process. Unlike English words where alphabet letters are written in a series, Korean words have a structure that combines *chosung*, *joongsung*, and *jongsung* (initial consonants, vowels, and final consonants) to make one character. Therefore, Korean characters must be

TABLE 3: Word list based on the Korean language.

List	Source	Number
Male names frequently used by Koreans	Ministry of Culture, Sport and Tourism	1,000
Female names frequently used by Koreans	Ministry of Culture, Sport and Tourism	1,000
Noun words with high frequency in the Korean language	National Institute of Korean Language	13,785

converted to the same structure as English so that they can be managed consistently in the password word dictionary. Figure 4 gives examples of the conversion process. The words before conversion have a combined structure like “한국” (“Korea”) and “비밀번호” (“password”). For “한국,” this means separating the *chosung*, *joongsung*, and *jongsung* of each character and converting them into consonant or vowel data like “ㅎ,” “ㅏ,” “ㄴ,” “ㄱ,” “ㅓ,” and “ㄷ.” These consonants and vowels can be made to correspond to alphabetical characters on keyboards. In other words, “ㅎ” is converted to “g,” “ㅏ” to “k,” and “ㄴ” to “s.”

This is the same keyboard array as the qwerty keyboard, which is one of the standard keyboards in the United States and consists of 26 keys that allow all alphabetic characters to be entered. For the case of the Korean language, all consonants and vowels can be entered with the same 26-character keyboards, and most of them have an array that is common in Korea. The reason the keyboard is used as the standard is because most passwords are entered through the keyboard.

Korean word data words converted to English letters have no apparent meaning. For example, “한국” in Figure 4 means “Korea,” but it becomes “gksrnr” after going through the conversion process and loses meaning. However, since there is no “gksrnr” in the English dictionary, this is likely to be predicted to be very secure when evaluated with existing password security evaluation models. Therefore, considering these limitations, conversion to alphabet-based words, and composing by merging with the existing dictionary, may result in better performance for security evaluation, and an improved password evaluation method that reflects situations like this is also needed.

3.5. Improved Password Security Evaluation. The password security evaluation based on the Korean language will be improved by going through a two-step process. In the first step, the zxcvbn evaluation method is used to predict security, and in the second step, the results of the analysis are evaluated by verifying whether the word dictionary contains the same word. The password security evaluation score is calculated through this, and the evaluation index is shown in Table 4.

Table 4 shows the security evaluation index and evaluation criteria proposed in this paper. A 10-step security evaluation score ($2p + t$) is obtained by segmenting the zxcvbn evaluation score (p) with five steps into whether or not there is a password dictionary match (t). The zxcvbn evaluation score is a predictive value that diagnoses the number of guess attempts up to a consistency evaluation with a user-selected password by combining words, numbers, and special symbols in the dictionary. By determining

FIGURE 4: Conversion process from Korean word to English word.

whether it matches the dictionary, the password character string that was not considered in the zxcvbn evaluation score is supplemented.

For example, since “password” uses the word in the dictionary as it is, it does not require a combination of numbers and special symbols, resulting in a score of zero in the zxcvbn evaluation. However, for “pas123sword@” in Figure 5, it is predicted that more than 10^{10} attempts are needed to infer it, resulting in a score of 4 in the zxcvbn evaluation. However, the zxcvbn evaluation does not take into account the case in which a number or special symbol is included. If the example “pas123sword@” is considered, it can be separated into two characters, “pas” and “sword,” with respect to the numbers “123,” each of which is a meaningless string. In such a case, while a high score can be obtained in the zxcvbn evaluation, it is less reliable because it has the same evaluation score as a password composed of a randomly chosen string.

That is why we need a second step to remove numbers and special symbols from passwords to verify if a word that is identical to the word dictionary is included. For example, since “pas123sword@” contains the “password” string, it is calculated to be 0 when the password dictionary matching criteria are applied. However, in the second step, “pas123swort@” is calculated as 1 because there are no words that match the dictionary. This means that although the zxcvbn score is the same for both cases, there is a difference in the process that verifies whether there is a match in the password dictionary, resulting in a different security evaluation score. However, the matching score does not consider “leetspeak.” This is a method of producing visually similar character strings by replacing alphabet characters with

TABLE 4: Password security evaluation index.

Security evaluation score ($2p + t$)	zxcvbn evaluation score (p)	Matching score (t)	Password security evaluation criteria
0	0	True (0)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^3 times) (ii) Matches the password dictionary
1		False (1)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^3 times) (ii) Does not match the password dictionary
2	1	True (0)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^6 times) (ii) Matches the password dictionary
3		False (1)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^6 times) (ii) Does not match the password dictionary
4	2	True (0)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^8 times) (ii) Matches the password dictionary
5		False (1)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^8 times) (ii) Does not match the password dictionary
6	3	True (0)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^{10} times) (ii) Matches the password dictionary
7		False (1)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted less than 10^{10} times) (ii) Does not match the password dictionary
8	4	True (0)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted 10^{10} times or more) (ii) Matches the password dictionary
9		False (1)	(i) Evaluate consistency with a password that was chosen by combining a word from the merged dictionary with numbers and special symbols (attempted 10^{10} times or more) (ii) Does not match the password dictionary

special symbols or numbers [7]. For example, the alphabet character “o” can be replaced with the number “0” and the alphabet character “l” with the number “1.” Passwords replaced with leetspeak get a high security evaluation since they are composed of a combination of alphabet characters, numbers, and special symbols [7]. However, since the matching score removes numbers and special symbols to assess whether or not there is a match with words in the dictionary, the alphabet characters replaced with leetspeak may be removed and it may not be possible to extract words. To solve this problem, we introduced the zxcvbn evaluation score and proposed that it as an integrated model. The zxcvbn method is a method that evaluates by combining words in a dictionary with leetspeak [7].

The zxcvbn evaluation score and matching score are obtained by using a Korean language-based password dictionary. For example, if “!gksrnr123” is evaluated with the

existing model, security is evaluated to be high because it uses words that are not included in the dictionary. However, if a Korean language-based password dictionary is used, it is judged as a word that is included in the dictionary and is considered less secure. In other words, the zxcvbn evaluation score and matching score are evaluated strictly for Korean language-based passwords. In addition, since the matching score only determines simply whether the password and words in a dictionary match, the score drawn with the zxcvbn evaluation score, which takes into consideration capital letters, lowercase letters, numbers, special symbols, and leetspeak altogether, can be weighted. Since the weight is a value that is relative to the zxcvbn evaluation score and matching score, only the weight for the zxcvbn evaluation score is adjusted after setting the matching score as the norm and fixing the weight for it to 1. For example, if the zxcvbn evaluation score is given a weight that is less than 1 and less

FIGURE 5: Password security prediction model based on Korean language.

than the weight for the matching score, whether the Korean language-based word matches will be evaluated to be more important than numbers, special symbols, and leetspeak. In addition, if the weight is set to 2, the Korean language-based words and numbers, special symbols, and leetspeak combinations are considered to be more important than whether the word matches. Therefore, in this paper, the weight of the zxcvbn evaluation score is set to a value that is double the value of the matching score.

Figure 5 shows the process of predicting the security of a password chosen by a user with the password security evaluation index proposed in Table 4. The password dictionary used as a criterion in the evaluation index is a Korean language-based password dictionary merged with the existing English word dictionary, and Table 2 is established in advance by preprocessing the word list of Table 3 with the process of Figure 4 for evaluation. The security evaluation score is calculated by determining whether the returned zxcvbn evaluation score matches the password dictionary, and it is returned to the user to help determine the password.

To verify the effectiveness of the evaluation model that is presented like this, the security of passwords must actually be analyzed, and checking is needed to see if they were leaked. Performance experiments that compare and evaluate the improvement in accuracy of the prediction for passwords with low security compared to the existing model will be conducted.

4. Experiment and Evaluation

4.1. Experiment Data. In this experiment, vulnerable passwords that were used for cracking will be used to determine how low the resulting security outcome is when evaluated with the proposed security evaluation index. In this

experiment, two types of data were used: a password dictionary in which English and Korean were merged, and a list of leaked passwords. Before using the data, a dictionary was established after Korean words in Table 3 were converted into alphabet-based words, and it was merged with the existing English word dictionary. Figure 6 shows the password word dictionary used in this experiment. (a) Shows the existing English vocabulary dictionary, which corresponds to item 0 in Table 2. One word is separated by a comma (,), and they were used in the experiment as a set of meaningful words.

Figure 6(b) shows the word dictionary for Korean words that were converted to alphabet-based words. Each word is separated by a comma as in Figure 6(a). Unlike Figure 6(a), however, each separated word has no meaning. For example, “tkfka,” the first word in Figure 6(b), is the result of the alphabetical conversion of “사람,” which means a human being. By merging the dictionaries that are the criteria for password security evaluation, both existing English passwords and Korean passwords were included.

For the collection of low-security passwords, the established Korean word dictionary and database [34] of leaked passwords were used. It was provided by “Have I Been Pwned,” and it is data that is frequently used in other existing research examples.

As suggested in Table 5, the most leaked lists used in this experiment are the most common male and female names used by Koreans. For male names, data for which 943 out of 1,000 were leaked was used in the experiment. For female names, data for which 899 out of 1,000 were leaked was used. In addition, for the list of frequently used noun words in Korean, even words with a leakage rate of 53.21% out of 13,785 were used. By comparison, for strings composed of four to 10 random characters, data with only 91 out of 10,000 leaked history were also used in the experiment.

FIGURE 6: Merged password word dictionary for English (a) and converted Korean (b).

TABLE 5: Leakage status and number of data for passwords based on Korean language that were used in the experiment.

List	Number leaked	Number not leaked	Leakage rate (%)	Total
Male names frequently used by Koreans	943	57	94.3	1,000
Female names frequently used by Koreans	899	101	89.9	1,000
Noun words with high frequency in the Korean language	7,335	6,450	53.21	13,785
Strings composed of four to 10 random characters	91	9,909	0.91	10,000
Sum of Korean-based passwords	9,177	6,608	58.13	15,785

In this experiment, a total of 9,177 leaked Korean-based passwords were classified as low-security passwords and used in performance evaluation experiments like this.

4.2. Experiment Environment. The latest (2020.11) version of “Have I Been Pwned” was used for experimental data, the leaked Korean language-based passwords were collected in a *Python* 3.6.5 environment, and the experiment was carried out with Selenium 3.141.0, BeautifulSoup 44.6.0, and Chrome web browser version 89.0.4389.114.

In the experiment, the evaluation score must be calculated based on the evaluation index. This was carried out with the evaluation program code that was implemented and executed using the Node.js 14.15.3 framework. Note that the implemented program code, the merged password dictionary, and the experimental data can be found in Github [35]. The establishment of the merged dictionary for the evaluation model used the inko 1.1.1 module to convert Korean words to an alphabetical basis and zxcvbn version 4.4.2 to produce security evaluation scores.

4.3. Experiment Results and Performance Evaluation. The performance experiment for the password security evaluation model confirmed that the security evaluation score predicted by the model proposed in this study is reliable. The evaluation criteria are the 9,177 low-security Korean language-based passwords obtained in Section 4.1, and whether the model yields security evaluation scores based on the number of leaks was determined.

Table 6 shows the results of evaluation scores for passwords with low security that were calculated by using the proposed Korean language-based password security evaluation model and the existing zxcvbn evaluation model.

For comparison, the security evaluation scores produced by the proposed model were expressed in five steps. For the leaked passwords, the proposed model evaluated most with results that were classified as being weak or very weak for security, and 53 were classified as being average. In addition, there were three cases in which security was classified as strong, and none were evaluated as being very secure.

However, for the existing zxcvbn evaluation model, 279 cases were evaluated to have strong security for the leaked passwords, and 24 cases were evaluated to be very strong. This poses a risk of allowing users to choose low-security passwords by evaluating them as very secure. In comparison, the proposed model rated 23 of the 24 leaked passwords, which had been evaluated to be very strong for security by the existing model, as passwords with weak security, and one as average. With results like this, the proposed model evaluated the security to be weak for the leaked Korean language-based passwords with an accuracy of 99.38%. Compared with the existing model, which showed an accuracy of 80.06%, it can be seen that the performance has been improved. Consequently, existing evaluation models lack reliability in evaluating the security of Korean language-based passwords, and it was possible to verify the reliability of the Korean-based password security evaluation model proposed in this paper.

The security evaluation score of the evaluation model should be related to the actual number of password leaks since it should show the security of the password chosen by the user most intuitively. To confirm this, the number of password leaks according to the security evaluation score calculated in Table 6 was verified. The results are shown in Figure 7.

Figure 7(a) shows the number of passwords leaked that were predicted by the evaluation model proposed in this paper. For the group that was evaluated to have very low

TABLE 6: Security evaluation results of Korean language-based leaked passwords.

Evaluation model	Security evaluation score	Number
Korean language-based password security evaluation model	0~1 (very weak)	700
	2~3 (weak)	8,421
	4~5 (average)	53
	6~7 (strong)	3
	8~9 (very strong)	0
Existing zxcvbn evaluation model	0 (very weak)	695
	1 (weak)	6,653
	2 (average)	1,526
	3 (strong)	279
	4 (very strong)	24

FIGURE 7: Number of password leaks based on security evaluation score. (a) Evaluation model proposed in this paper. (b) Existing evaluation model.

security, there were passwords that were leaked more than 7,000 times, and most of the passwords that were leaked more than 1,000 times were found here. In addition, for the group that was evaluated to have low security, there were a number of passwords that were leaked more than 1,000 times, although it was less than for those that were evaluated to be very low in security. In comparison, for the group that was evaluated to have normal or strong security, passwords with very few leaks were included, and security was never evaluated as very strong.

Figure 7(b) shows the number of passwords leaked that was predicted by the existing evaluation model. For the group that was evaluated to have very low or low security, the evaluation results were shown to be similar to the evaluation model in (a). However, there was a case in which Korean language-based passwords that were leaked more than 1,000 times were evaluated to be moderate in security. We also confirmed that there were more than 100 leaked passwords in the group that was rated to have strong security, and there was a case in

which a leaked password was evaluated to have very strong security.

The results of identifying passwords leaked less than 500 times to compare groups that rated security as normal, strong, and very strong are shown in Figure 8. Figure 8(a) shows the security evaluation score evaluated by the proposed model, and it has a very small number of leaks in the group that was evaluated to have normal security. For example, the most leaked password was “Tkdenddl,” which was converted from “쌍둥이” meaning “twin,” and it was leaked 29 times in total. Three passwords were classified in the very secure group, and one, two, and seven leaks were identified, respectively, which are very small numbers.

Figure 8(b) shows the security evaluation score evaluated by the existing evaluation model. In the password group, which was evaluated to be very secure, there was a password with 94 leaks. In addition, 24 leaked passwords were identified in groups that were evaluated to be very secure.

In this regard, it was confirmed that the password security evaluation model proposed in this paper evaluated

FIGURE 8: Passwords with less than 500 leaks based on security evaluation scores. (a) Evaluation model proposed in this paper. (b) Existing evaluation model.

Korean language-based passwords to have lower security evaluation scores compared with the existing evaluation model. In addition, when the security evaluation score evaluated by the proposed evaluation model and the actual number of passwords that had been leaked were verified, there were passwords evaluated to be very weak or weak that had been leaked 1,000 times or more. For the case in which the security was evaluated to be normal, the number of times a password had been leaked was less than 100, and it was less than 10 for the case in which the security was evaluated to be strong. These results showed that the proposed Korean language-based password security evaluation model is more reliable than the existing evaluation model and can provide users with accurate security evaluation scores.

5. Conclusion

Protecting personal information through user authentication is an important part of the field of information security. Most of all, for passwords, it is necessary to ensure proper security while considering convenience, since passwords are chosen by users. However, because existing password security evaluation models do not take into account the tendencies of non-English language users when they choose passwords, these models have limitations in security evaluation and security cannot be ensured when non-English language users choose passwords.

Therefore, in this paper, a dictionary of Korean language-based words, with a language system that is independent even among non-English languages, was composed by collecting Korean words and converting them into an alphabet-based form. After merging this dictionary with the password dictionary used for traditional English-speaking user-based password security evaluation, an

evaluation model that predicts security was proposed. After collecting low-security passwords using a database of passwords that had been actually leaked, performance evaluation experiments were conducted along with the existing password security evaluation model. The results confirmed that the proposed model is superior. Therefore, if the Korean-based password security evaluation model is put into actual service, it is expected that it can contribute to having more security when users choose passwords. For example, it can be used on a website like Google. When a user changes their membership or password, if security evaluation based on a password dictionary that matches the language environment of the corresponding user is provided, it will be possible for the user to select a password with stronger security.

In this study, however, it was assumed that hackers use trawling guessing for a universal Korean language-based password security evaluation model. Passwords or PII-based passwords that are frequently reused by individual users were not considered. For future research, if targeted guessing attacks by hackers using PII-tag data such as name, birthday, and mobile phone number are considered, it will be possible to improve the performance of the password security evaluation model.

Data Availability

The data used to support the findings of this study are available at https://github.com/KiHyeon-Hong/Korean-based_password_security_model_paper.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Technology Development Program funded by the Ministry of SMEs and Startups (MSS, Korea) (Grant Nos. S2957039 and S2798371).

References

- [1] M. Jerry, Y. Weining, L. Min, and L. Ninghui, "A study of probabilistic password models," in *Proceedings of the 2014 IEEE Symposium on Security and Privacy*, Berkeley, CA, USA, May 2014.
- [2] A. Naiakshina, A. Danilova, C. Tiefenau, M. Herzog, S. Dechand, and M. Smith, "Why do developers get password storage wrong? A qualitative usability study," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, Dallas, TX, USA, October 2017.
- [3] D. Chang, A. Jati, S. Mishra, and S. K. Sanadhya, "Cryptanalytic time-memory trade-off for password hashing schemes," *International Journal of Information Security*, vol. 18, no. 2, pp. 163–180, 2019.
- [4] N. Woods and M. Siponen, "Improving password memorability, while not inconveniencing the user," *International Journal of Human-Computer Studies*, vol. 128, pp. 61–71, 2019.
- [5] S. Furnell, "—over a decade of progress?" *Computer Fraud & Security*, vol. 2018, no. 7, pp. 6–13, 2018.
- [6] S. Furnell, "An assessment of website password practices," *Computers & Security*, vol. 26, no. 7–8, pp. 445–451, 2007.
- [7] X. D. C. D. Carnavalet and M. Mannan, "A large-scale evaluation of high-impact password strength meters," *ACM Transactions on Information and System Security*, vol. 18, no. 1, pp. 1–32, 2015.
- [8] D. Pasquini, G. Ateniese, and M. Bernaschi, "Interpretable probabilistic password strength meters via deep learning," in *Proceedings of the European Symposium on Research in Computer Security*, Guildford, UK, September 2020.
- [9] V. Taneski, M. Heričko, and B. Brumen, "Password security—No change in 35 years?" in *Proceedings of the 2014 37th International Convention on Information and Communication Technology*, Opatija, Croatia, 26–30 May 2014.
- [10] C. Castelluccia, M. Dürmuth, and D. Perito, "Adaptive password-strength meters from Markov models," in *Symposium on Network and Distributed System Security*, San Diego, CA, USA, February 2012.
- [11] S. T. Haque, M. Wright, and S. Scielzo, "A study of user password strategy for multiple accounts," in *Proceedings of the third ACM conference on Data and application security and privacy*, San Antonio, TX, USA, February 2013.
- [12] S. Y. Lee, J. Y. Lee, and G. H. Lee, "Diagnosing and coping with mode errors in Korean-English dual-language keyboard," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, UK, April 2019.
- [13] G. Hu, "On password strength: a survey and analysis," in *Proceedings of the International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Springer, Cham, Buenos Aires, Argentina, June 2017.
- [14] J. Blocki and A. Datta, "CASH: a cost asymmetric secure hash algorithm for optimal password protection," in *Proceedings of the 2016 IEEE 29th Computer Security Foundations Symposium*, Lisbon, Portugal, June 2016.
- [15] E. Tirado, B. Turpin, C. Beltz, P. Roshon, R. Judge, and K. Gagneja, "A new distributed brute-force password cracking technique," in *Proceedings of the International Conference on Future Network Systems and Security*, Springer, Cham, Paris, France, July 2018.
- [16] X. Yu and Q. Liao, "User password repetitive patterns analysis and visualization," *Information & Computer Security*, vol. 24, no. 1, pp. 93–115, 2016.
- [17] C. Shen, T. Yu, H. Xu, G. Yang, and X. Guan, "User practice in password security: an empirical study of real-life passwords in the wild," *Computers & Security*, vol. 61, pp. 130–141, 2016.
- [18] M. Yildirim and I. Mackie, "Encouraging users to improve password security and memorability," *International Journal of Information Security*, vol. 18, no. 6, pp. 741–759, 2019.
- [19] A. K. Kyaw, F. Sioquim, and J. Joseph, "Dictionary attack on Wordpress: security and forensic analysis," in *Proceedings of the 2015 Second International Conference on Information Security and Cyber Forensics*, Cape Town, South Africa, November 2015.
- [20] Y. He, E. E. Alem, and W. Wang, "Hybritus: a password strength checker by ensemble learning from the query feedbacks of websites," *Frontiers of Computer Science*, vol. 14, no. 3, pp. 1–14, 2020.
- [21] D. Wang, P. Wang, D. He, and Y. Tian, "Birthday, name and bifacial-security: understanding passwords of Chinese web users," in *Proceedings of the 28th USENIX Security Symposium*, USENIX Security 19), Santa Clara, CA, USA, August 2019.
- [22] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted online password guessing: an underestimated threat," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, Vienna, Austria, October 2016.
- [23] D. Florêncio, C. Herley, and P. C. Van Oorschot, "An administrator's guide to internet password research," in *Proceedings of the 28th Large Installation System Administration Conference*, Seattle, WA, USA, November 2014.
- [24] F. Yu and Y. Huang, "An overview of study of password cracking," in *Proceedings of the 2015 International Conference on Computer Science and Mechanical Automation*, NW Washington, DC, USA, October 2015.
- [25] D. Pasquini, M. Cianfriglia, G. Ateniese, and M. Bernaschi, "Reducing bias in modeling real-world password strength via deep learning and dynamic dictionaries," in *Proceedings of the 30th USENIX Security Symposium*, Virtual event, August 2021.
- [26] M. M. Taha, T. A. Alhaj, A. E. Moktar, A. H. Salim, and S. M. Abdullah, "On password strength measurements: password entropy and password quality," in *Proceedings of the 2013 International conference on computing*, Khartoum, Sudan, August 2013.
- [27] P. Doucek, L. Pavlíček, J. Sedláček, and L. Nedomova, "Adaptation of password strength estimators to a non-English environment—the Czech experience," *Computers & Security*, vol. 95, 2020.
- [28] Naked Security, "Why you still can't trust password strength meters," 2016, <https://nakedsecurity.sophos.com/2016/08/17/why-you-still-cant-trust-password-strength-meters>.
- [29] S. J. Park, J. M. Byun, S. O. Baek, Y. S. Cho, and A. Oh, "Subword-level word vector representations for Korean," in *Proceedings of the 56th Annual Meeting of the Association For Computational Linguistics*, vol. 1, pp. 2429–2438, 2018.
- [30] D. Wang, D. He, H. Cheng, and P. Wang, "fuzzyPSM: a new password strength meter using fuzzy probabilistic context-free grammars," in *Proceedings of the 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Toulouse, France, June 2016.

- [31] M. Golla and M. Dürmuth, “On the accuracy of password strength meters,” in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, Toronto, Canada, October 2018.
- [32] Ministry of Culture, “Sport and tourism,” 2014, <https://efamily.scourt.go.kr/index.jsp>.
- [33] National Institute of Korean Language, 2005, https://www.korean.go.kr/front/reportData/reportDataView.do?mn_id=45&report_seq=1.
- [34] Have I Been Pwned, “Pwned passwords,” 2020, <https://haveibeenpwned.com/Passwords>.
- [35] GitHub, “Korean based password security model paper,” 2021, https://github.com/KiHyeon-Hong/Korean-based_password_security_model_paper.