Baltic J. Modern Computing, Vol. 13 (2025), No. 4, pp. 919–932
https://doi.org/10.22364/bjmc.2025.13.4.09

Dictionary Attack with Transformed Russian Words using QWERTY Keyboard Layout

Lea MÜLLER[1], Aušrius JUOZAPAVIČIUS[2], Volodymyr OKHRIMCHUK[3], Stefan SÜTTERLIN[1,4]

[1] Albstadt-Sigmaringen University, Albstadt, Germany
[2] General Jonas Žemaitis Military Academy of Lithuania, Vilnius, Lithuania
[3] Korolov Zhytomyr Military Institute, Zhytomyr, Ukraine
[4] Østfold University College, Halden, Norway

muellerlea@hs-albsig.de, ausrius.juozapavicius@lka.lt, okhrimchuk84@ukr.net, suetterlin@hs-albsig.de

ORCID 0009-0009-6538-8046, ORCID 0000-0002-8852-8605, ORCID 0000-0001-7518-9993, ORCID 0000-0002-4337-1296

Abstract. Despite the known vulnerabilities of passwords, the username-password combination remains the most widely used authentication method. Many users still choose simple, memorable passwords, making them susceptible to dictionary attacks. These attacks are especially effective when using target-specific wordlists. This paper introduces a novel wordlist tailored to Russian-speaking users who may type passwords using the QWERTY layout while writing in Russian, leading to seemingly random character strings. Based on this assumption, a dictionary of transformed Russian words was compared with one million unique Russian passwords. The analysis revealed that around 1% of the passwords exactly matched transformed entries, and an additional 6% partially matched, supporting the effectiveness of this new wordlist approach.

Keywords: Dictionary Attack, Passwords, Password Security, Transformed Passwords, Keyboard Layout

1 Introduction

Notwithstanding the vulnerability of passwords to a variety of attacks (Bošnjak et al., 2018; Alkhwaja et al., 2023; Klein, 1990), username and password remain the most prevalent authentication method globally (Statista, 2024; Wash and Rader, 2021; Taneski et al., 2019). While the necessity for complex passwords that are difficult to guess is widely acknowledged (Wash and Rader, 2021), a considerable number of users opt for straightforward passwords that can be easily recalled (Bryant and Campbell, 2006;

Woods and Siponen, 2018; Taneski et al., 2019; Shen et al., 2016), such as simple keyboard patterns (Shen et al., 2016; Klein, 1990; Ur et al., 2015), dictionary words (Shen et al., 2016; Klein, 1990), or a combination of a word and a basic pattern of numbers (Wash and Rader, 2021; Ur et al., 2015).

The potential for exploitation of a given password varies depending on its length, complexity, and the specific attack method employed. Examples of the various types of attacks include exhaustive search or dictionary attacks. An exhaustive search, or brute force attack, involves systematically testing all possible combinations of characters for an unknown password (Bryant and Campbell, 2006). This type of attack is therefore not a suitable option for long passwords (Alkhwaja et al., 2023; Bryant and Campbell, 2006). Dictionary attacks employ wordlists for the purpose of guessing passwords (Alkhwaja et al., 2023; Yan et al., 2000). This method poses a risk in instances where users select passwords that can be found in dictionaries or other wordlists, such as those resulting from previous password breaches. Although this type of attack is less time-consuming than an exhaustive search (Bryant and Campbell, 2006), it is not feasible for passwords that are not included in the wordlist used in the attack (Alkhwaja et al., 2023; Bryant and Campbell, 2006). However, a combination of different approaches can be used to optimise password guessing attacks.

This paper will focus on a dictionary attack using a novel wordlist based on the transformation of Russian words through the substitution of Cyrillic letters with the corresponding characters as found on a QWERTY keyboard layout. The aim of this paper is to ascertain whether the proposed method of selecting transformed Russian words as passwords is a common practice among Russian-speaking users and, if so, whether it could be employed to enhance dictionary attacks by utilising target-specific wordlists. The efficacy of this approach is evaluated through the use of Russian words transformed through a simple substitution of Cyrillic letters with the characters they share a key with on the QWERTY keyboard layout. However, it is conceivable that the same approach can be applied to further keyboard layouts.

The paper is structured as follows: Section 2 will describe the concept of dictionary attacks and provide examples of potential enhancements to this attack type. Section 3 will put forth a novel approach for dictionary attacks against Russian-speaking users. The construction of the wordlist in this attack is based on the assumption that Russian-speaking users select a Russian word as their password and type it in accordance with the Russian keyboard layout, despite having their keyboard configured to the QWERTY layout, thereby creating seemingly random patterns of characters. This assumption is tested against a list of the one million most frequent passwords used by a Russian-speaking audience. Section 4 will present a summary of the findings and offer an outlook on potential future work.

2 Dictionary Attacks

Dictionary attacks represent a category of attack employed for the purpose of guessing unknown passwords or usernames. This is achieved through a systematic process of testing words contained in a dictionary or wordlist (Alkhwaja et al., 2023; Yan et al., 2000). Dictionary attacks are based on the premise that passwords chosen by users are susceptible to being easily guessed. This is particularly the case when words contained in a dictionary are used as a password. Additionally, as many users reuse the same or similar passwords for multiple purposes (Wash and Rader, 2021; Bryant and Campbell, 2006; Woods and Siponen, 2018; Taneski et al., 2019; Wash et al., 2016), dictionary attacks utilising password lists from previous security breaches can be employed to rapidly deduce a password (Bryant and Campbell, 2006; Taneski et al., 2019).

The selection of secure passwords is a fundamental aspect of information security, yet users continue to rely on easily memorable passwords (Bryant and Campbell, 2006; Woods and Siponen, 2018; Taneski et al., 2019; Shen et al., 2016). Although the use of dictionary words may safeguard a password from being brute-forced (assuming a sufficiently lengthy word is selected), this approach leaves the password vulnerable to dictionary attacks.

Consequently, users seek methods to circumvent the use of words that precisely match dictionary entries while simultaneously striving for a password that is memorable and straightforward to recall. For example, one frequently employed strategy is to append digits to a dictionary word (Wash and Rader, 2021; Ur et al., 2015), resulting in a password such as *password123*.

The approach presented in this paper is based on the premise that Russian-speaking users may select a Russian word as their password and type it in accordance with the Russian keyboard layout, despite having their keyboard configured to the QWERTY keyboard layout. This approach is discussed in detail in Section 3, along with a consideration of how it may be applied in the context of dictionary attacks.

2.1 Optimisation of Dictionary Attacks

Although dictionary attacks, which involve using all entries in a dictionary to guess passwords, are effective when a password matches an existing entry exactly (Alkhwaja et al., 2023), they are less successful when users make minor alterations to their passwords, such as adding an additional character (Wash and Rader, 2021; Ur et al., 2015). Consequently, dictionary attacks can be enhanced by combining them with other types of attacks. The following section outlines some of the approaches that can be employed to improve the efficacy of dictionary attacks.

The optimisation of a dictionary attack through the integration of an exhaustive search methodology entails the utilisation of words from a pre-defined wordlist, which are employed to guess passwords. In contrast to a traditional dictionary attack, this process involves the combination of these words with additional characters, thereby expanding the search space. Many users opt for passwords that comprise a word in conjunction with additional characters, such as letters, numbers, or a year (Wash and Rader, 2021; Rinn et al., 2015; Ur et al., 2015). To illustrate, a combination of a dictionary attack with an exhaustive search methodology would not only try the word *password* but also combinations of *password* with other characters, such as *password1* or *password12*.

Some users create passwords by repeating or concatenating words (Bošnjak et al., 2018; Bryant and Campbell, 2006; Shen et al., 2016; Klein, 1990). Consequently, dictionary attacks can be optimised to not only repeat a dictionary entry once, but to test repetitions of dictionary words, such as *passwordpassword*, or to combine two or more words from a dictionary, such as *hellopassword*. Consequently, the combination of dictionary attacks with an exhaustive search methodology renders passwords comprising a word and additional characters, repeated words, or concatenated words susceptible to relatively straightforward exploitation.

An additional method for enhancing the efficacy of dictionary attacks is the utilisation of target-specific wordlists. To illustrate, a German dictionary could be employed in a dictionary attack against German-speaking users. Additional customisation is possible should the attacker have access to supplementary personal information regarding the target (Bryant and Campbell, 2006; Klein, 1990; Taneski et al., 2019), such as the names or dates of birth of family members (Shen et al., 2016; Rinn et al., 2015). This can be achieved through the use of open-source intelligence (OSINT).

This paper puts forth a method for optimising dictionary attacks through the utilisation of a novel type of target-specific wordlist. This new approach is predicated on the substitution of the Cyrillic alphabet with Latin letters and a select set of special characters. Section 4 will address the potential applications of this approach in enhancing the efficacy of dictionary attacks.

3 Dictionary Attack with Transformed Russian Words using QWERTY Keyboard Layout

3.1 Premise

In addition to the aforementioned approaches for optimising dictionary attacks, a novel approach is presented and its efficacy is evaluated in the following section.

In the area of password usage and dictionary attacks, research is primarily focused on an English-speaking audience (Wang et al., 2019), while languages that use non-ASCII characters have been less extensively investigated. The approach presented in this article is based on the premise that Russian-speaking users may input their passwords according to the Russian keyboard layout, while having their keyboard configured to the QWERTY layout. Typing a word in this manner would result in a seemingly random sequence of characters. To test this assumption, the Russian language was selected for analysis. The rationale for this decision is threefold. Firstly, Russian is a widely used language (Zeidan, 2023b,a). Secondly, it does not use the QWERTY or a similar keyboard layout (Wilcock and Dempsey, 2024; Unicode, 2021). Thirdly, all Cyrillic letters are mapped to a single ASCII character on the keyboard (compare Figure 1).

As there are multiple keyboard layouts for Cyrillic characters, the most prevalent layout was selected for this study. According to Chumachenko and Burkov, the ЙЦУКЕН (JCUKEN) keyboard layout, a layout specifically adapted for the Russian language, is the most popular Cyrillic layout (Chumachenko and Burkov, 2023). This is also the default Russian keyboard layout available in different operating systems, such as Windows (Wilcock and Dempsey, 2024), ChromeOS (Unicode, 2021) or Ubuntu. The Russian keyboard layout, according to (Wilcock and Dempsey, 2024; Unicode, 2021), is illustrated in Figure 1. In this illustration, the QWERTY keyboard layout is presented in black letters, with the Russian layout superimposed in red. When typing

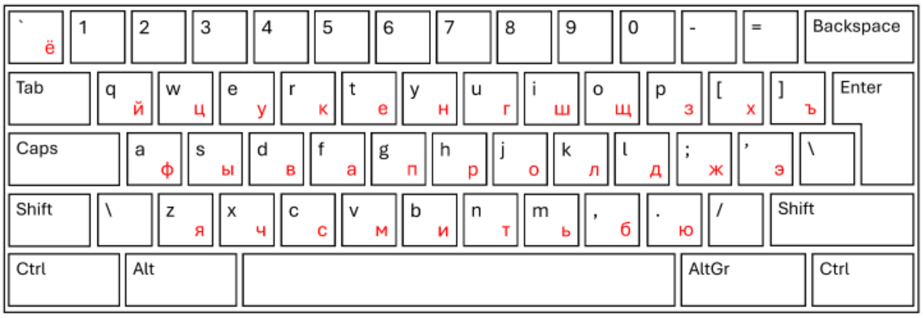

Fig. 1. QWERTY keyboard layout with Russian keyboard layout superimposed in red letters.

a Russian word according to the Russian layout with the keyboard configuration set to QWERTY, a simple substitution of Cyrillic letters for Latin letters, as well as some special characters, will occur. This implies that each Russian letter is precisely matched to a single letter or special character within the QWERTY keyboard layout. To illustrate, the Cyrillic letter А is mapped to the Latin letter *F*.

Table 1 illustrates the correspondence between the Cyrillic letters of the Russian keyboard layout and the characters of the QWERTY keyboard layout, namely the characters by which they are substituted. Accordingly, the Russian word for *password*, пароль, would be substituted with the seemingly random sequence of characters *gfhjkm*.

Table 1. Cyrillic letters and the characters by which they are substituted (QWERTY).

Cyrillic	ё	й	ц	у	к	е	н	г	ш	щ	з	х	ъ	ф	ы	в	а	п	р	о	л	д	ж	э	я	ч	с	м	и	т	ь	б	ю
QWERTY	‘	q	w	e	r	t	y	u	i	o	p	[	]	a	s	d	f	g	h	j	k	l	;	’	z	x	c	v	b	n	m	,	.

3.2 Preparation

In order to assess the efficacy of this novel approach, a list of the 50,000 most prevalent Russian words, obtained from (Hingston, 2018), was transformed in accordance with the aforementioned substitution. This entailed the replacement of the Cyrillic characters with the corresponding characters on the QWERTY keyboard layout. The result was a list of 50,000 seemingly random sequences of Latin letters and special characters.It should be noted that in this step all Cyrillic letters were transformed to lowercase Latin letters and no uppercase letters were used.

The first part of this study compares the list of transformed words with a list of the one million most frequently used passwords among Russian-speaking users (Sharsi1, 2019) to identify all exact matches between a transformed Russian word and an entry in the password list. The second part of the study compares the same two lists, this

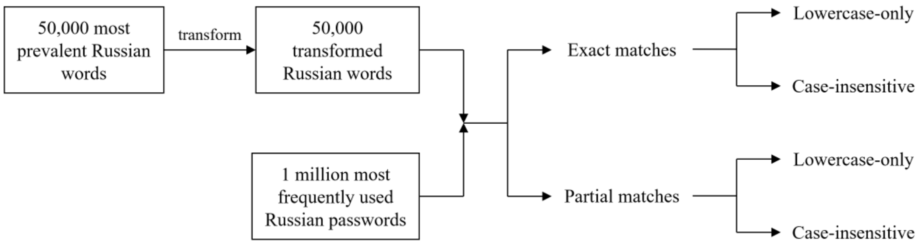
50,000 most prevalent Russian words
transform
50,000 transformed Russian words
1 million most frequently used Russian passwords
Exact matches
Partial matches
Lowercase-only
Case-insensitive
Lowercase-only
Case-insensitive

Fig. 2. The steps of the analysis.

time considering partial matches. In the following, the term ‘exact match’ is used to describe transformed Russian words that are present in the password list in the same form as stated, without the addition of any characters or numbers. To illustrate, the transformed word *gfhjkm* is only deemed an exact match if it occurs in the password list identically. *gfhjkm123*, however, would not be regarded as an exact match. In contrast, ‘partial matches’ may encompass additional characters. Consequently, both *gfhjkm* and *gfhjkm123* would be identified as a match in this instance.

In the search for both exact and partial matches, lowercase-only and case-insensitive matches were sought in two stages. As the list comprising the 50,000 transformed Russian words was created exclusively using lowercase letters, in the initial phase of each analysis, only lowercase matches are taken into consideration. In the second step of each analysis, all matches regardless of case are searched for. To illustrate, in the lowercase-only search, only *gfhjkm* would be considered a match, whereas in the case-insensitive search, versions such as *Gfhjkm* and *GFHJKM* would additionally be considered a match. The steps of the analysis are illustrated in Figure 2.

To reduce the number of false positives, Russian words comprising a length of less than five characters were excluded from the analysis. In this context, a false positive refers to a situation where a word is identified as a match, despite it being a random sequence or one selected for a different purpose, such as keyboard patterns. Given the prevalence of such instances with shorter words, words comprising four characters or less were excluded. This resulted in the exclusion of 1,565 exact matches with a length of less than five characters when considering lowercase-only matches, and the exclusion of 1,900 case-insensitive matches.

3.3 Analysis

To evaluate the aforementioned premise, the list of 50,000 transformed Russian words was compared with a list of the one million most frequently used passwords among Russian speakers, retrieved from (Sharsi1, 2019). As indicated in the GitHub repository, this is a list of common passwords used by Russian speakers. The data set was uploaded in 2019 and was compiled from a number of data breaches, with the passwords ordered according to the frequency of occurrence. In addition to the passwords, a count is provided for each password. The most frequently used password is *qwerty*, with a count of 8,226,502, and the least frequently used passwords have a count of 65.

In the first step of the analysis, only exact matches were considered. In a subsequent step, partial matches were also taken into consideration. To minimise the number of false positives, only those matches with a minimum length of five characters were considered.

3.3.1 Searching for Exact Matches

To perform a comparative analysis between the 50,000 transformed Russian words and the list of the one million most frequently used Russian passwords, the logic illustrated in Listing 1.1 was used. Given two sets, one containing the transformed Russian words and one containing the Russian passwords, computing the intersection of these two sets yields all exact matches. Matches comprising four or fewer characters were discarded.

This yielded 9,580 exact matches, representing approximately 0.96% of the one million most frequent passwords used by Russian speakers.

Listing 1.1. Pseudocode showing the logic for finding all exact matches of a transformed Russian word with a Russian password.

```
temp = set_intersection(transformed Russian words, Russian passwords)
common words = set()
for word in temp:
  if length(word)>=5:
    add word to common words
```

As the 9,580 matches comprise solely those passwords consisting of lowercase letters, the analysis was repeated with a case-insensitive comparison. To achieve this, all Russian passwords were converted to lowercase and subsequently compared with the 50,000 transformed Russian words. This implies that, in addition to lowercase passwords such as *gfhjkm*, uppercase passwords such as *Gfhjkm* or *GFHJKM* were also identified as matches. This search yielded 12,448 results, representing approximately 1.24% of the one million most frequent passwords used by Russian speakers. This number also includes the 9,580 lowercase-only matches.

3.3.2 Searching for Partial Matches

In addition to exact matches, the lists of transformed Russian words and one million passwords were compared to identify instances where a transformed Russian word forms a part of a password in the password list.

The logic used to identify all passwords in the list of the one million most frequently used Russian passwords that at least partially consist of a transformed Russian word is shown in Listing 1.2. For each of the transformed Russian words with five or more characters, it was tested whether it is part of any of the Russian passwords.

The comparison yielded 72,950 results, which also included all exact matches. Upon exclusion of the 9,580 previously identified exact matches, the remaining set comprises 63,370 partial matches, representing approximately 6.34% of the one million most frequently used Russian passwords.

Listing 1.2. Pseudocode showing the logic for finding all partial matches, i.e., matches where a transformed Russian word is part of a Russian password.

```
transformed words = transformed Russian words
```

```
passwords = Russian passwords

subwords = set()
for transformed word in transformed words:
  if length(transformed word)>=5:
    for password in passwords:
      if transformed word is part of password:
        add "transformed word | password" to subwords
```

As with the search for exact matches, analysis was repeated with a case-insensitive comparison. To achieve this, all Russian passwords were converted to lowercase and then compared with the 50,000 transformed Russian words once more. Again, the logic presented in Listing 1.2 was used. This yielded 81,031 results, which also included all exact matches. Upon exclusion of the 12,448 previously identified exact matches, the remaining set comprises 68,583 partial matches, representing approximately 6.86% of the one million most frequently used Russian passwords.

As a common approach among many users is to create a password by adding numbers at the beginning or end of a word, such as a birth year or a simple pattern like *123*, the list with partial matches was subsequently filtered for passwords that consist of a transformed Russian word with preceding or following numbers.

The logic used to identify these matches is presented in Listing 1.3. This assumes that all partial matches have been identified according to Listing 1.2, implying that the transformed word that matches a password is already known. Initially, all numerical characters are removed from the password. A match is deemed to be present if the transformed Russian word and the password devoid of numerals are found to be identical. In this instance, exact matches, defined as passwords that corresponded precisely to a transformed Russian word prior to the removal of numbers, were not deemed to be matches.

The comparison yielded 21,355 lowercase-only and 23,245 case-insensitive results, indicating that approximately 2.32% of the one million unique passwords were a combination of a transformed Russian word with preceding or following numbers, such as *gfhjkm1977*.

Other observed passwords were combinations of several transformed Russian words (for example, the transformed words *rfrjqnj* and *gfhjkm* were combined to form *rfrjqnjgfhjkm*) or repeated stringing together of the same transformed Russian word (for example, *gfhjkm* was repeated to form *gfhjkmgfhjkm*).

Listing 1.3. Pseudocode showing the logic for filtering all partial matches that are a combination of a transformed Russian word with added numerals. In this case, the matching transformed word is already known, as all partial matches have been identified previously.

```
def remove_numbers(string):
  return string with numbers removed

passwords = passwords that consist in part of a transformed word

matches with numerals = set()
for password in passwords:
```

```
if remove_numbers(password) equals transformed word and password does
    ↪ not equal transformed word:
  add "transformed word | password" to matches with numerals
```

3.4 Results

Table 2 presents the results of the comparative analysis of the 50,000 most frequently used Russian words, transformed in accordance with the proposed methodology, and the one million most frequently used passwords by Russian speakers.

Table 2. Results as absolute values and as percentages of the 1 million most frequenlty used Russian passwords. It should be noted that case-insensitive matches also include all lowercase-only matches.

	Lowercase-only	Case-insensitive
Exact match	9,580 (0.96%)	12,448 (1.24%)
Partial match (all)	63,370 (6.34%)	68,583 (6.86%)
Partial match (added numbers)	21,355 (2.14%)	23,245 (2.32%)
Total	72,950 (7,30%)	81,031 (8.10%)

A comparison between the 50,000 transformed Russian words and the one million passwords yielded 9,580 exact matches and 63,370 partial matches, when only passwords composed of lowercase letters were considered. When passwords comprising both lowercase and uppercase letters were considered, 12,448 exact matches and 68,583 partial matches were identified. An exact match was defined as a transformed Russian word with a minimum length of five characters that corresponded exactly to an entry in the password list, whereas a partial match was defined as a transformed Russian word that was part of an entry in the password list. Among the partial matches, 23,245 were instances where a transformed Russian word with digits added was used as a password.

Table 2 additionally illustrates the proportion of the one million most frequently used Russian passwords that align with the proposed scheme. While only a minor proportion of the passwords (approximately 1.24%) are an exact match for a transformed Russian word, a larger proportion (approximately 6.86%) consist of a transformed Russian word in part.

It is important to note that only unique passwords were considered in this analysis, and the frequency with which these passwords are used was not taken into account. The presented percentages do not necessarily reflect the actual prevalence of users employing this method. Rather, they represent the proportion of unique passwords within the list of the one million most frequently used Russian passwords that correspond to this method.

Figure 3 shows the distribution of the number of exact matches, both in lowercase-only and case-insensitive categories, across a range of password lengths. The lengths of exact matches ranged from five characters (due to the exclusion of matches with four or fewer characters from the analysis in order to reduce the number of false positives) to 21 characters.

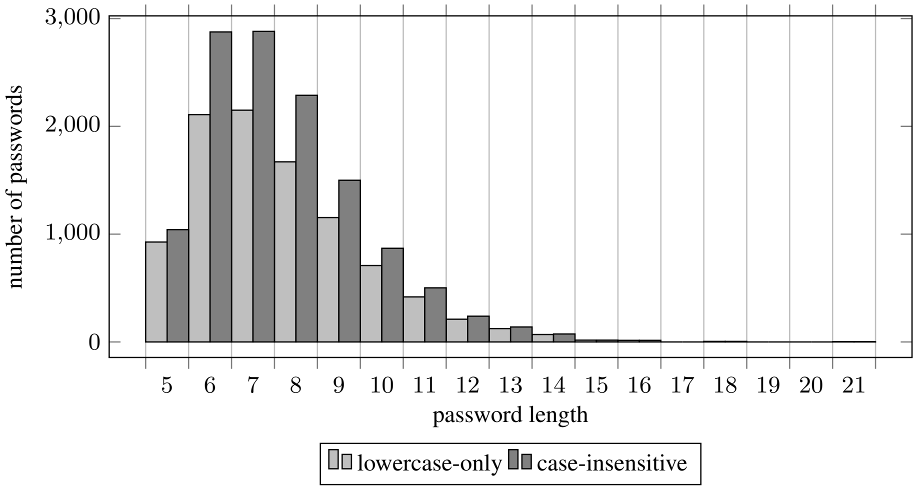

Fig. 3. Distribution of exact matches across different password lengths. Lowercase-only: passwords that only use lowercase letters. Case-insensitive: passwords that use lowercase and/or uppercase letters.

4 Discussion

A novel approach to the creation of target-specific wordlists for dictionary attacks against Russian-speaking users has been presented. This approach is predicated on the assumption that Russian speakers may opt for a Russian word as their password and input it in accordance with the Russian keyboard layout, despite having their keyboard configured to the QWERTY layout. This would result in a seemingly random sequence of Latin letters and special characters.

In order to determine whether this method is employed by Russian speakers, a list comprising the 50,000 most frequently used Russian words was transformed in accordance with the aforementioned scheme and subsequently compared with a list of the one million most frequently used passwords by Russian-speaking users. In order to reduce the number of false positives, only transformed Russian words comprising a minimum of five characters were considered.

The comparison of transformed Russian words with the one million most frequently used Russian passwords revealed that this approach is employed by a subset of Russian-speaking users. Approximately 1% of the passwords were an exact match with one of the transformed Russian words, while an additional approximately 6% of the passwords at least partially consisted of a transformed Russian word.

While the passwords that represented exact matches could be guessed using a traditional dictionary attack, which simply tests the words contained in a wordlist, the passwords that partially consist of transformed Russian words could be attacked by combining a dictionary attack with an exhaustive search approach. One such approach would be to add additional numbers at the beginning or end of a word. The findings suggest that utilising a dictionary of transformed Russian words could facilitate more effective dictionary attacks against Russian-speaking users. Furthermore, they illustrate how the deployment of target-specific wordlists can enhance the efficacy of dictionary attacks.

The presented technique introduces a wordlist that has not previously been considered in the context of dictionary attacks. The proposed approach represents a novel contribution to the field of dictionary attacks. Rather than merely considering existing wordlists for target-specific attacks, such as using a German dictionary for an attack on German-speaking targets, it suggests an entirely new set of words. This is achieved by substituting Cyrillic letters with Latin letters and some special characters.

This reinforces the necessity for users to select robust passwords and demonstrates that circumvention strategies, such as the one outlined in this paper, will inevitably result in the generation of weaker passwords that are susceptible to specific forms of attack.

4.1 Threats to Validity

This section provides a structured analysis of the threats to internal, external, and construct validity.

Internal validity refers to the extent to which the findings can be attributed to the premise of this work, namely that transformed Russian words appear in password lists because users type Russian words while having their keyboard configured to the QWERTY layout. One potential threat arises from alternative explanations for the observed matches. Some matched sequences may originate from simple keyboard patterns or random character strings, rather than from the hypothesised behaviour. Even after excluding strings of fewer than five characters to reduce accidental matches, longer coincidental matches may still occur. A second threat relates to the quality of the password list used in this study. The dataset comprising the one million most commonly used Russian passwords was compiled from publicly available breaches in 2019, and its accuracy cannot be verified independently. Noise, bias or artefacts in the dataset may influence the number and nature of detected matches, thereby affecting internal validity. A third threat stems from the incomplete coverage of the Russian lexicon. The present study relies on a list of the 50,000 most frequent Russian words, which does not include names, slang, domain-specific terms and less common words often used in passwords. This may lead to an underestimation of the prevalence of the hypothesised behaviour. Additionally, reported rates are sensitive to analytical choices, such as the minimum length of five characters and how case is handled. These analytical decisions introduce additional uncertainty into the interpretation.

External validity refers to the extent to which the results can be generalised beyond the specific data and conditions of this study. One threat that affects external validity is temporal generalisability. The password list is based on data that was leaked and collected prior to 2019, so it may not accurately reflect present-day behaviour, websites, or demographics. Password behaviour and security practices may have changed since then. Therefore, the prevalence of the studied phenomenon in contemporary settings may differ. Another threat concerns population validity. The dataset only includes Russian-speaking users from unspecified services that have been affected by historical breaches. It does not necessarily reflect the broader population of Russian-speaking users or subgroups such as mobile users, different age groups, or users of specific platforms. However, concerns about population validity affect the majority of studies analysing password behaviour based on leaks, as analyses are limited by the availability of lists of leaked passwords. In this case, it was essential to employ a password list for a specific user group, Russian-speakers, which complicated the search for password lists further. Generalising the findings to other keyboard layouts and Cyrillic languages poses another threat. While the present study focuses on the Russian JCUKEN keyboard layout, other layouts map Cyrillic characters differently. The same substitution patterns cannot be assumed to apply across languages or layouts. Although this study relies on publicly available datasets and provides pseudocode, its results may be contingent on the specific list of leaked passwords and the analytical choices employed.

Construct validity refers to whether the operationalisation used in the study accurately captures the behaviour of Russian-speaking users entering passwords using the QWERTY keyboard layout. A primary threat to this arises from the definition of a match. In the present study, matches were defined as exact or partial occurrences of transformed Russian words within leaked passwords. However, these matches may not always reflect the deliberate selection of a Russian word typed while having the keyboard configured to the QWERTY layout. Some matches may be the result of accidental substrings or other password selection habits unrelated to the behaviour under analysis. A second threat is the limitation to a list of 50,000 Russian words. Because this wordlist does not contain a comprehensive set of words, such as names, slang or domain-specific terms that users may select as their passwords, the prevalence of the hypothesised behaviour may be incorrectly represented. These threats may lead to both false negatives (missed legitimate instances) and false positives (matches that do not reflect the behaviour under analysis). A further threat arises from preprocessing decisions, such as case normalisation, the exclusion of short passwords and the criteria for partial matches. These decisions influence how the construct is instantiated in the analysis. As partial matches include passwords that merely contain the transformed sequence anywhere, the operationalisation may at times be broader than the underlying construct. Finally, as the study infers behaviour purely from leaked passwords, there is no direct behavioural validation available. Consequently, the study provides an inferred approximation rather than a directly measured behavioural phenomenon.

4.2 Additional Limitations and Future Work

In addition to the threats to the validity of this research, as discussed in Section 4.1, investigating one language and keyboard layout constitutes a primary limitation of the research. While analysing additional languages and keyboard layouts would significantly strengthen the claim of a novel type of target-specific wordlist presented in this article, no such analysis was conducted. This was due to the fact that analyses of password behaviour are dependent on the availability of password lists. In this case, password lists of a specific user group (i.e., of a specific language) are required, which further limits their availability. Most password lists do not contain passwords from a single user group characterised by their language. Furthermore, not all languages and keyboard layouts are suitable for the presented approach, as an exact mapping from the language-specific keyboard layout to the QWERTY layout is required. It can generally be assumed that the approach can be applied to all languages and keyboard layouts where the language-specific characters map exactly to one letter or special character on the QWERTY keyboard layout. However, this assumption is subject to the condition that users exhibit the hypothesised behaviour when selecting passwords. Due to the limited scope of this article, we encourage further research analysing password lists of specific languages.

The aim of this study was to ascertain whether the proposed methodology of selecting transformed Russian words is employed by Russian-speaking users. However, only a relatively limited set of Russian passwords, comprising one million unique entries, was subjected to analysis. Moreover, the number of matches identified merely reflects the number of distinct passwords that adhere to this scheme. It may not accurately reflect the proportion of users who employ this strategy when creating passwords.

Further tests could be conducted using a larger set of Russian passwords to enhance the reliability of the findings. An investigation into the frequency of the passwords could provide insight into the prevalence of this scheme among Russian-speaking users. Furthermore, the password list employed in this study is from 2019, thus the utilisation of a more recent dataset could also enhance the reliability of the findings.

There are numerous alternative keyboard layouts in use beyond the QWERTY layout, including Ukrainian, Kazakh, or Belarusian (Chumachenko and Burkov, 2023). To determine whether this same approach to selecting transformed words as passwords is also utilised by other audiences, further analysis could be conducted on additional password datasets to confirm or refute the generalisability of the approach.

References

Alkhwaja, I., Albugami, M., Alkhwaja, A., Alghamdi, M., Abahussain, H., Alfawaz, F., Almurayh, A., Min-Allah, N. (2023). Password cracking with brute force algorithm and dictionary attack using parallel programming, *Applied Sciences* **13**(10), 5979.

Bošnjak, L., Sreš, J., Brumen, B. (2018). Brute-force and dictionary attack on hashed real-world passwords, *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1161–1166.

Bryant, K., Campbell, J. (2006). User behaviours associated with password security and management, *Australasian Journal of Information Systems* **14**(1).

Chumachenko, O., Burkov, A. (2023). Development of an efficient ukranian keyboard layout using a genetic algorithm, *Electronics and Control Systems* **2**(76), 35–39.

Hingston (2018). 50000-russian-words-cyrillic-only, GitHub Repository.
https://github.com/hingston/russian/blob/master/50000-russian-words-cyrillic-only.txt

Klein, D. V. (1990). Foiling the cracker: A survey of, and improvements to, password security, *Proceedings of the 2nd USENIX Security Workshop*, pp. 5–14.

Rinn, C., Summers, K., Rhodes, E., Virothaisakun, J., Chisnell, D. (2015). Password creation strategies across high- and low-literacy web users, *Proceedings of the Association for Information Science and Technology* **52**(1), 1–9.

Sharsi1 (2019). stat_russkiwlst_top_1m, GitHub Repository. https://github.com/sharsi1/russkiwlst/blob/master/stat_russkiwlst_top_1M.txt

Shen, C., Yu, T., Xu, H., Yang, G., Guan, X. (2016). User practice in password security: An empirical study of real-life passwords in the wild, *Computers & Security* **61**, 130–141.

Statista (2024). Authentication methods deployment worldwide 2023. https://www.statista.com/statistics/1441144/companies-authentication-methods-deployment-status-worldwide/

Taneski, V., Heričko, M., Brumen, B. (2019). Systematic overview of password security problems, *Acta Polytechnica Hungarica* **16**(3), 143–165.

Unicode (2021). Layouts: Russian (ru). https://www.unicode.org/cldr/charts/40/keyboards/layouts/ru.html

Ur, B., Noma, F., Bees, J., Segreti, S. M., Shay, R., Bauer, L., Christin, N., Cranor, L. F. (2015). "i added'!'at the end to make it secure": Observing password creation in the lab, *Eleventh symposium on usable privacy and security (SOUPS 2015)*, pp. 123–140.

Wang, D., Wang, P., He, D., Tian, Y. (2019). Birthday, name and bifacial-security: Understanding passwords of chinese web users, *28th USENIX Security Symposium (USENIX Security 19)*, USENIX Association, pp. 1537–1555.

Wash, R., Rader, E. (2021). Prioritizing security over usability: Strategies for how people choose passwords, *Journal of Cybersecurity* **7**(1).

Wash, R., Rader, E., Berman, R., Wellmer, Z. (2016). Understanding password choices: How frequently entered passwords are re-used across websites, *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pp. 175–188.

Wilcock, J., Dempsey, P. (2024). Russian keyboard. https://learn.microsoft.com/en-us/globalization/keyboards/kbdru

Woods, N., Siponen, M. (2018). Too many passwords? how understanding our memory can increase password memorability, *International Journal of Human-Computer Studies* **111**, 36–48.

Yan, J., Blackwell, A., Anderson, R., Grant, A. (2000). The memorability and security of passwords – some empirical results, *Technical Report UCAM-CL-TR-500*, University of Cambridge, Computer Laboratory.

Zeidan, A. (2023a). languages by number of native speakers. https://www.britannica.com/topic/languages-by-number-of-native-speakers-2228882

Zeidan, A. (2023b). languages by total number of speakers. https://www.britannica.com/topic/languages-by-total-number-of-speakers-2228881

Received July 28, 2025 , revised November 26, 2025, accepted December 11, 2025