Enhancing the resistance of password hashing using binary randomization through logical gates

Muhamad Zaki Anbari, Bambang Sugiantoro
Master of Informatics Department, Faculty of Science and Technology, Sunan Kalijaga Islamic State University, Yogyakarta, Indonesia

ABSTRACT

Digitalization in various sectors makes information security issues very crucial. Information security follows the authentication, authorization, and accounting (AAA) principle, where one of the most important parts is authentication. The most widely used authentication method is username-password. The best method to secure a user-pass is to convert the plaintext using a hash so that the converted plaintext cannot be recovered. However, with higher technology, hackers can crack the ciphertext using brute force.

This research proposes a username-password scrambling algorithm before it is fed into the hash function to improve resilience from attacks. This algorithm is named logical gates (LG). It works by converting the user pass into binary form, adding salt, and scrambling it with certain logical gates before inserting it into the hash function. Testing is divided into two: time of execution and attack resistance. Time of execution results show that LG takes 0.0443432033 s, while without LG takes 0.01403197646 s. The resistance of attack results show that the plaintext of the hash amplified by LG cannot be cracked at all and increases the attack time by 321.3% at prefix and 161.3% at postfix, while without LG, the plain text can be found for a certain duration of time.

1. INTRODUCTION

Today most people have utilized technology to support their various activities such as communication, business, study, entertainment, public services, and so on. Each application of the technology is closely related to data or information so that information security becomes very important for everyone. The rapid digitization on the one hand does facilitate various activities but also creates risks such as data leakage. Confidentiality, integrity, and availability are three parameters that must be referenced in building information security [1]. Confidentiality means ensuring that the use and storage of data can only be done by those who have the right. Integrity means ensuring that the data is intact and not deformed or contaminated by malware or viruses. Availability means ensuring that users who have data access rights can access the data when they need it [2].

Every application must have an authentication system. Authentication is very important because it serves to ensure that the data or information in the application can only be accessed by those entitled to access it [3]. The most commonly used authentication is a combination of username and password, so the issue of securing password data is very important [4]. Some methods of securing passwords are by encrypting passwords, entering passwords into a hash function before storing them in the database [5], and

Journal homepage: http://ijece.iaescore.com
creating strong passwords [6]. Strong password criteria include a minimum length of 6 characters, not using the same password for all accounts, using a combination of letters, numbers, symbols, and avoiding dictionary words as much as possible [7].

The method of securing passwords by storing hashes is the most secure method. Hackers will have difficulty finding the plaintext password due to the one-way nature of the hash function [8]. The use of hashes to secure data or passwords can be improved by adding salt [9]. Salt is random text that is combined with the plaintext password. Salt in addition to increasing the number of password characters will also make it more difficult to read and guess, making it more immune to various attacks such as dictionary attack, brute force attack, or rainbow table attack [10], [11]. One of the modules used to generate salt is Secrets. The Secrets module exists in the python programming language and is used to generate cryptographically strong random strings. Secrets has incorporated various parameters in cryptography theory to be able to generate random strings that are more immune to cryptographic attacks [12]. Apart from using salt, hash security is also affected by the combination scheme between password and salt. Research conducted by [13] clearly shows the effect of salt placement when used as prefix, postfix, and alternating. This means that the security of the hash function can be improved by increasing the complexity of the password before it is entered into the hash function.

Another method introduced to improve hash security is the SXR algorithm. SXR stands for split, XOR, and replace. This algorithm consists of 4 steps: the password is hashed, then the ratio and number of iterations of the secret key are calculated. Next, the hashed password is split into 2 based on the result of the ratio calculation. Finally, the obtained data is combined and stored in the database [14]. Another method introduced is by swapping the elements in the array proposed by Karrar et al. [15]. This algorithm works by randomizing the password and salt elements, then separating and hashing the salt when stored in the database, making it difficult for hackers to guess the stored password. Research conducted by De Guzman et al. [16] tried to strengthen the hash function by adding a hill chipher. Hill chipher is used to increase complexity when generating salt. This research focuses on overcoming hash-collision that may occur in large databases. The method used is that when the system detects hash-collision, the algorithm will take the first and last characters of the username, then input them into the matrix encryption key (MEK) which then generates a salt. The salt is combined with the original password to produce a new hash value so that hash-collision can be avoided.

Based on previous research, there are several methods to improve hash security. Starting from changing the hash function itself, adding salt, using complex passwords, or adding other algorithms before the hash function. In summary, this research aims to develop alternative algorithms in randomizing usernames, passwords, and salt before being entered into the hash function. The resulting algorithm is expected to have better complexity and robustness than other existing algorithms or methods. The proposed algorithm is named the logical gates (LG) algorithm. The way it works is that the username and password that have been combined with salt into binary form, then scramble them with a series of certain logic gates, the result is a binary circuit that is completely different from the initial binary circuit. Furthermore, the binary series is converted back to string form and entered into a hash function. In the final stage, testing is done using time of execution test and resistance of attack hash to compare the performance of LG algorithm with other algorithms.

2. PROPOSED ALGORITHM

At this stage the proposed algorithm is designed and developed. The proposed algorithm uses a combination of logic gates to scramble the username, password, and salt before being entered into the hash function. The flowchart of the proposed algorithm can be seen in Figure 1. Logic gates have many types with different functions. According to Garg and Kaur [17] there are 7 types of logic gates, namely AND, OR, NOT, NAND, NOR, XOR, and XNOR.

The LG algorithm uses some of these logic gates to randomize the binary username, password, and salt before entering the hash function. The first step is for the user to enter a username and password like most authentication methods. Second, the system will automatically generate a salt using the secrets module. The third salt and password will be combined using two different schemes in turn, namely the prefix and postfix schemes. The goal is to find out which scheme is better for strengthening the hash function.

The fourth step of the combined password and salt will be converted in binary form which only consists of the numbers 1 and 0. Then the binary will be entered sequentially through a series of certain logic gates. Determining the order of the logic gate circuit is very considering the function of the function of each logic gate. Errors in making sequences can cause functions between logic gates to be ineffective and eliminate each other. For example, when the NAND gate is followed by the NOT gate, it is the same as directly using the AND gate. The logic gate circuit used in the LG algorithm is as follows.
\[ A \text{ AND } B = AB \]  \hspace{1cm} (1)

\[ A \text{ OR } B = A + B \]  \hspace{1cm} (2)

\[ AB \text{ NAND} (A + B) = \overline{AB}(A + B) \]  \hspace{1cm} (3)

\[ AB \text{ NOR} (A + B) = \overline{AB} + (A + B) \]  \hspace{1cm} (4)

\[ \overline{AB}(A + B) \text{ XOR} \overline{\overline{AB}} + (A + B) = AB(A + B) \oplus \overline{AB} + (A + B) \]  \hspace{1cm} (5)

\[ \overline{AB}(A + B) \oplus \overline{\overline{AB}} + (A + B) \text{ NOT} = AB(A + B) \oplus AB + (A + B) \]  \hspace{1cm} (6)

After passing through a series of logic gates, the resulting binary will be converted back into string form. But sometimes the resulting binary cannot be converted immediately into string form because the arrangement of the resulting binary does not meet ASCII standards. So, the step before converting back into a string is to force the binary to meet the ASCII standard. Finally, the resulting string is entered into the hash function. The flowchart of LG algorithm is shown in Figure 1.

---

Figure 1. Flowchart of LG algorithm
3. METHOD

3.1. Research Setup

The research begins with preparing research equipment in the form of a computer with Intel Core i3-10100T @3.00 GHz processor specifications, 8 GB RAM, and Windows 10 operating system. Computer specifications will greatly affect the testing process. Especially in testing time of attack using the hashcat application, where the application is highly dependent on computer computing capabilities [18].

3.2. Experimental design

This research uses 8 different passwords and salts with the same username. Each password and salt will pass through the LG algorithm with variations in salt placement postfix and prefix. The following is the pseudo-code of the LG algorithm when using the prefix variation, while the postfix variation only needs to change the pseudo-code on line number 4 to $ps \leftarrow salt + p$. Algorithm 1 are showing the prefix variations of the LG algorithm:

```plaintext
Algorithm 1. LG algorithm
Input: a username U
a password P
Output: a hash value with modified password MP
Start
01: U \leftarrow char(00000000)
02: P \leftarrow char(000000)
03: salt \leftarrow generate_random_char (00)
04: ps \leftarrow p + salt
05: bin_U \leftarrow conv2bin(U)
06: bin_PS \leftarrow conv2bin(Ps)
07: bin_AND \leftarrow empty array
08: bin_OR \leftarrow empty array
09: bin_NAND \leftarrow empty array
10: bin_NOR \leftarrow empty array
11: bin_XOR \leftarrow empty array
12: bin_NOT \leftarrow empty array
13: ascii_char \leftarrow empty array
14: for i to len(bin_U) do :
15: bin_AND[i] \leftarrow bin_U[i] AND bin_PS[i]
16: bin_OR[i] \leftarrow bin_U[i] OR bin_PS [i]
17: end loop
18: for I to len(bin_U) do :
19: bin_NAND[i] \leftarrow bin_AND[i] NAND bin_OR[i]
20: bin_NOR[i] \leftarrow bin_AND[i] NOR bin_OR[i]
21: end loop
22: for I to len(bin_U) do :
23: bin_XOR \leftarrow bin_NAND[i] XOR bin_NOR[i]
24: end loop
25: for I to len(bin_U) do :
26: bin_NOT \leftarrow bin_XOR[i] NOT
27: end loop
28: ascii_char \leftarrow "01"+bin_NOT[3:len(bin_U)]
29: hash_value \leftarrow md5(ascii_char)
```

Testing is done with the aim of knowing how the hash resistance that has been strengthened with the LG algorithm when compared to hash without the LG algorithm. Testing is divided into 2, namely time of execution and resistance of attack. Time of execution test is a test of algorithm execution time, which is the time required for a computer to run an algorithm, generally the lower the execution time, the better the algorithm [19]. While the resistance of attack test is a test of hash resistance when attacked or hacked. In this test, the brute force attack method will be used, which is the easiest attack and is widely used by hackers because the method is simple [18]. The resistance of attack test uses hashcat v6.2.6 software to intentionally penetrate or brute force attacks. The more complex or secure a hash used, the longer it takes to hack it [20].

4. RESULTS AND DISCUSSION

The LG algorithm is tested through two types of testing, namely time of execution and resistance of attack. Some examples of passwords and salts used for testing purposes are taken based on [13] which are then limited to 6 characters, while the salt is generated using the secrets python module and is limited to 2 characters. Limiting the number of characters in passwords and salts aims to speed up testing due to limited computer specifications and computing power [21].

Enhancing the resistance of password hashing using binary randomization ... (Muhamad Zaki Anbari)
4.1. Test dataset

The username used in the research is made uniform, namely "uinjogja", while the list of passwords, salt, and hash values used can be seen in Table 1. The LG algorithm is implemented using the Python language by utilizing Google Colab. Here's an example of how the LG algorithm works. For example, the password inputted until the hash value is generated. This test uses the MD5 hash function. Finally, the resulting hash value is stored in the database.

<table>
<thead>
<tr>
<th>No.</th>
<th>Password</th>
<th>Salt</th>
<th>Hash value prefix (Non-LG)</th>
<th>Hash value prefix (LG)</th>
<th>Hash value postfix (Non-LG)</th>
<th>Hash value postfix (LG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>qwerty</td>
<td>b</td>
<td>a57cfbf9481b37302279</td>
<td>3c9d9eaca8b8b26363c</td>
<td>5400</td>
<td>5404</td>
</tr>
<tr>
<td>2</td>
<td>123456</td>
<td>d</td>
<td>2b10b38ccf1a1f33a</td>
<td>6ae36c9655a04b37fd</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>3</td>
<td>monyet</td>
<td>ya</td>
<td>bc06027d5f0b5c5a7</td>
<td>c401b18dd4</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>4</td>
<td>qazwsx</td>
<td>60</td>
<td>77fc2e5b3c4d549e0</td>
<td>51fafa</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>5</td>
<td>qwaszxe</td>
<td>e2</td>
<td>aac9ea9a263147296b51a</td>
<td>c401b18dd4</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>6</td>
<td>harkey</td>
<td>e7</td>
<td>c0c3cb45de71d6e00</td>
<td>74dcd454783b1c7747</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>7</td>
<td>654321</td>
<td>ef</td>
<td>77cbe5a08142a0f65330</td>
<td>41ebc4084e9a33512</td>
<td>5404</td>
<td>5407</td>
</tr>
<tr>
<td>8</td>
<td>abc123</td>
<td>67</td>
<td>3a13782a8981863ea0</td>
<td>b897728d5b6d4a8f</td>
<td>5404</td>
<td>5407</td>
</tr>
</tbody>
</table>

Thus, the algorithm without LG has better performance than the LG algorithm in the execution proposed algorithm. The execution without proposed algorithm takes 0.0469002723 seconds while the LG algorithm takes 0.0469002723 seconds.

4.2. Time of execution test

Time of execution test aims to determine the execution time of the LG algorithm and the algorithm without LG. The execution time is calculated from the time the username and password are inputted until the hash value is generated. This test uses software. The test results show that the algorithm without LG takes 0.01403197646 seconds while the LG algorithm takes 0.0443432033 seconds. The LG algorithm takes 4 times longer than the algorithm without LG. This data is obtained by testing each password from Table 1 using prefix and postfix variations for 10 repetitions and then taking the average. Thus, the algorithm without LG has better performance than the LG algorithm in the execution time section. However, with an execution time of less than 0.1 seconds, humans will not really notice it when the algorithm is implemented in the real world [22]. The time of execution test data is shown in Table 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Time execution proposed algorithm (s)</th>
<th>Time execution without proposed algorithm (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prefix</td>
<td>Postfix</td>
</tr>
<tr>
<td>1</td>
<td>0.0114199250</td>
<td>0.01403197646</td>
</tr>
<tr>
<td>2</td>
<td>0.0118825466</td>
<td>0.01502171516</td>
</tr>
<tr>
<td>3</td>
<td>0.0103435567</td>
<td>0.01403197646</td>
</tr>
<tr>
<td>4</td>
<td>0.010716843299</td>
<td>0.01403197646</td>
</tr>
<tr>
<td>5</td>
<td>0.010199652576</td>
<td>0.01403197646</td>
</tr>
<tr>
<td>6</td>
<td>0.01075904083</td>
<td>0.01403197646</td>
</tr>
<tr>
<td>7</td>
<td>0.01037818023</td>
<td>0.01403197646</td>
</tr>
</tbody>
</table>

4.3. Resistance of attack test

The resistance of attack test aims to compare the resistance of the hash value generated without the LG algorithm with the hash that has been reinforced with the LG algorithm. The test data is obtained by testing each password from Table 1 using prefix and postfix variations for 10 repetitions and then the average is taken. The resistance of attack data is shown in Table 3.

Brute force works by trying all possible combinations to guess the password to be attacked. The duration of brute force is highly dependent on how long the character set is and the computational capabilities of the device used by the hacker [21]. The following is the command used to perform brute force penetration using hashcat:

\[
\text{hashcat} -a 3 -m 0 \text{ hash\_value.txt} -f\text{orce}
\]

The above command consists of several arguments, namely:
- `hashcat`: to call hashcat application
- `-a`: to determine attack mode
- `3`: to determine brute force as attack mode
- `-m`: to determine hash function
- `0`: to determine md5 as hash function
- `hash\_value.txt`: to call file contains hash value
- `-f`: to ignore any warnings

In hashcat, brute force attacks are similar to mask attacks. The difference between mask attack and brute force attack is in determining the keyspace. Brute force attacks use the default keyspace, while mask attacks use a keyspace that has been adjusted to the hacker's target. Keyspace is a limitation of possibilities that are determined before carrying out an attack [23]. For example, when hackers use brute force, by default hashcat will use a combination of character sets that include uppercase letters, lowercase letters, and all numbers or known as mixalpha-numeric. The use of keyspace allows hackers to speed up the attack process as it minimizes the number of combinations that need to be executed by the computer. However, the use of keyspace may actually prolong the hacking process or even thwart it due to the hacker's mistake in determining the keyspace. Keyspace determination is closely related to social engineering techniques. Social engineering plays a role in the process of finding out how someone designs a password. By knowing how someone designs a password, hackers can determine a narrower keyspace so that the attack time becomes shorter [24].

This research uses a brute force attack with a default keyspace because brute force is more general and objective. In short, when a hash is immune to brute force, it is almost certainly immune to other types of attacks [25]. Here is the keyspace brute force attack on hashcat:

```
"?!?2?!?2?!?2?!?2?!?3?!?3?!?d?!?d?!?d"
```

where,

- `? 3`: : ? ? ? d * ! $ @ _ (lowercase letters, numbers, and five selected special characters)

The results of the resistance of attack test can be seen in Table 3.

<table>
<thead>
<tr>
<th>No</th>
<th>Prefix Non LG (s)</th>
<th>Cracked</th>
<th>Prefix LG (s)</th>
<th>Cracked</th>
<th>Postfix Non LG (s)</th>
<th>Cracked</th>
<th>Postfix LG (s)</th>
<th>Cracked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5639.3</td>
<td>Yes</td>
<td>28798.4</td>
<td>No</td>
<td>17774.7</td>
<td>Yes</td>
<td>28631.7</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>1450.5</td>
<td>Yes</td>
<td>28793.3</td>
<td>No</td>
<td>11637.6</td>
<td>Yes</td>
<td>28488.8</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>6316.8</td>
<td>Yes</td>
<td>28831.6</td>
<td>No</td>
<td>2469.8</td>
<td>Yes</td>
<td>28543.1</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>22421.9</td>
<td>Yes</td>
<td>28669.8</td>
<td>No</td>
<td>3277.1</td>
<td>Yes</td>
<td>28705.9</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>7506.7</td>
<td>Yes</td>
<td>28965.2</td>
<td>No</td>
<td>7759.2</td>
<td>Yes</td>
<td>28474.1</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>6345.7</td>
<td>Yes</td>
<td>20715</td>
<td>No</td>
<td>19217.9</td>
<td>Yes</td>
<td>28568.7</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>2161.7</td>
<td>Yes</td>
<td>28430.8</td>
<td>No</td>
<td>22854.2</td>
<td>Yes</td>
<td>28601.5</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>775.3</td>
<td>Yes</td>
<td>28456.5</td>
<td>No</td>
<td>2415.6</td>
<td>Yes</td>
<td>28660.3</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 3 shows that conventional or non-LG hashes can be cracked within a certain period of time. The difference in time is due to the varying level of complexity of the hashed password. Based on the experiments conducted, it is known that passwords that use numbers are more vulnerable than passwords that use letters. This is because numbers only have 10 different variations, namely numbers 0-9. While letters have 26 different variations, namely A-Z. For example, password number 1 (qwerty) has a longer hack time than password number 2 (123456). In addition, the order in which passwords are created has a significant effect on the duration of the hack [13]. Putting letters at the beginning of the password shows better strength than putting numbers at the beginning of the password. Password No. 2 when combined with salt (d8123456) has a longer duration than password No. 8 (67abc123) even though the number of letters in password No. 8 is more than No. 2. Meanwhile, the LG-strengthened hash has a much longer hacking duration than the conventional hash. Table 2 shows that the duration of the LG algorithm has a small difference between one password variation and another. This is very different from the non-LG algorithm whose duration varies greatly according to the level of password complexity. The similarity in the duration of the LG hash is because hashcat has run out of password candidates with a length of 8 characters. So, the attack duration on the LG hash is the total amount of time it takes hashcat to test password candidates from 1 character to 8 characters. In the LG hash, all password variations cannot be found in plaintext. This method is much more effective for improving password security because there is no need to force someone to create a complex password for security, but the computer will automatically strengthen the entered password no matter whether it is vulnerable or not [26].

5. CONCLUSION

The development of digitalization makes information security issues very important. Fraud, hacking, data theft that is increasingly happening shows that research in the field of information security is needed. The main focus of this research is on methods to improve password security using hash functions by utilizing the logical gates (LG) algorithm. The LG algorithm is an algorithm that functions to randomize the username and password at the binary level before being entered into the hash function. Tests were conducted to measure the execution time of the LG algorithm and measure the resistance of the hash generated using a brute force attack.

The test results show that the LG algorithm requires four times longer execution time than the use of hash without LG. However, the hash combined with the LG algorithm has a much higher resistance to brute force attacks. Experimental results show that none of the 8 password variations used can be cracked by brute force. Even though in the worst-case hackers can hack the hash of LG, hackers still need to try to find the original plaintext password from the modified password generated by LG. This research can be a solution to build a system that is user friendly, meaning that users do not need to create passwords that are too complicated, but still take into account in terms of information security. Future research is expected to develop the LG algorithm so that it can be embedded in a hardware so that its application in strengthening information security becomes wider.

REFERENCES


BIographies of Authors

Muhamad Zaki Anbari received his B.Sc degree in physics from Diponegoro University, Indonesia, in 2020. Currently, he is a master of computer science degree candidate at the Department of Informatics, Faculty of Science and Technology, Sunan Kalijaga State Islamic University. His research interests include automation, internet of things, computer networks, and cybersecurity. He can be reached at email: mzkianbari@gmail.com.

Bambang Sugiantoro earned his B.Sc and Dr degrees in computer science from Gadjah Mada University, Indonesia. MT degree in electrical engineering at Bandung Institute of Technology. He is currently the head of the informatics master study program at the Informatics Department of Sunan Kalijaga State Islamic University. His research interests include computer networks, cybersecurity, and machine learning. He can be contacted via email: bambang.sugiantoro@uin-suka.ac.id.