Research for the Application and Safety of MD5 Algorithm in Password Authentication

Xiaoling Zheng
Department of Computer Science and Technology
Capital University of Economics and Business
Beijing, China

Jidong Jin
Department of Computer Science and Technology
Capital University of Economics and Business
Beijing, China

***Abstract*—MD5 algorithm takes an important position in the application of password authentication. This essay analyses the security features of the passport authentication and the methods of application safety improvement of MD5 algorithm in password authentication, on the basis of which, it focuses on the methods of the application safety improvement of MD5 in the passport authentication by switching or interfering with the treatment process of MD5.**

Keywords- MD5; Collision Attack; Password Authentication

I. FOREWORD

As the continuous development of computer network technology and rapid popularization of Internet application, e-commerce and e-government businesses are used more and more widely. However, at the time that these applications bring great convenience to our work and lives, they also bring increasingly serious security problems. Identity authentication is an important method to ensure safety of various e-commerce activities, e-government business information and internet information. Certificate based digital signature authentication and password authentication are most common at present. But traditional password authentication with basic mode of user name and password authentication faces many security problems, which concluded like these below:

- Eavesdrop through user's host. There are various Trojan horse programs coming from different channels nowadays, which could be activated by user's carelessness and give the remote hacker a chance to record user's activities, including his password.
- Attain data through system flaws. The computer system storing and verifying password may have security flaws in different aspects, which would be used by hackers to hack into the system and obtain user's password files and decipher them by special tools.
- Monitor through network data. As authentication information is usually transmitted through network, hackers could monitor these data by special tools and abstract user's name, password and other authentication information through analyzing.

The biggest problem of the traditional password authentication method is that the user's name, password and other key information used in authentication are transmitted on the internet with plain codes and stored in plain codes. In order to prevent problems mentioned above threatening the computer security, the common resolution is to use MD5 algorithm to convert the content of user's password information and transmit or store the results converted.

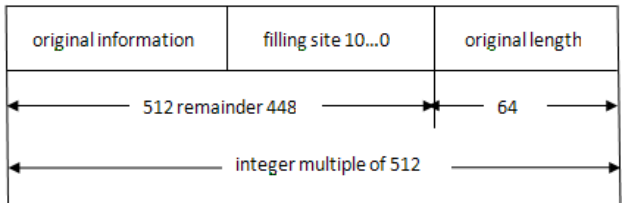
MD5 is the abbreviation of Message-Digest Algorithm 5, which was developed jointly by MIT Computer Science Laboratory and Ronald L. Rivest from RSA Data Security Inc. in early 90s in 20th century and evolved from MD2, MD3 and MD4. It compresses a piece of information with plain code and random length into 128 bits value by hash algorithm, which is called information distract. MD5 algorithm is irreversible and cannot recover the original plain code information from information abstraction, thus it is always believed safe. However, some researches indicate that MD5 algorithm could be deciphered by collision attack, and the security of its application has received the challenge. This essay analyzes the application of MD5 algorithm in password authentication and its security, and probes into the physical measures of the application security of MD5 algorithm in password authentication.

II. HASH FUNCTION AND MD5 ALGORITHM

Hash function compresses a piece of information with random length by hash algorithm into fixed length value, which is called information abstraction. An information abstract generated from two pieces of different plain code is the so called "collision". The requirements for safe Hash function include: first, two pieces of different plain code generate the same information abstract which should not be calculated and is called collision; second, a certain information abstract cannot be calculated through the other plain information generating the same abstract, which means the initial state cannot be deduced by the results. The so call "cannot be calculated" means that it is too expensive to get the results through the algorithm. Thus, Hush function usually contains two obvious features: first, no matter how long of the plain code information, the information abstract with certain length after calculating; second, if only the plain code information has any change, no matter how small the change is, the corresponding information abstract will be totally different. Thus, it is usually called "digital finger print", which could distinguish identities and ensure the unique and integrity of plain code information and the main effect is password authentication.

978-1-4673-0024-7/10/$26.00 ©2012 IEEE

MD5 algorithm is one of the most common Hash function. Its basic principle is to process the input information divided groups by 512 bits, and each group divided into 16 sub-groups with 32 bits. After a series of processing, the algorithm output composed by 4 groups with 32 bits, and cascade this 4 groups will generate a hash value with 128 bits. In the process of MD5 algorithm, fill information first to make its length 64 less than the multiple numbers of 512. The filling method is to attach a 1 and millions of 0, and add an information length before filling indicated by binary system with 64 bits. These two steps are to make the information length be the integer multiple of 512, and ensure the difference after different information filling. The data model after filling is shown as the diagram below:



After filling, each information group with 512 bits is divided into 16 bytes blocks with 32 bits each. Set variable a, b, c ,d in each cipher block chaining, a= 0x67452301， b= 0xefcdab89，c= 0x98badcfe，d= 0x10325476. The main loop of algorithm after initialization is 4 cycles, each cycle will process 16 bytes blocks with 64 steps total. The function formula could be shown as M (a, b, c, d, m[i] , k[i], s[i]): a= b+ ((a+ f(b, c, d) + m [i] + s[i]) <<< k[i]), in which, "+"is modulo 32 adder, s[i] and k[i] are the settled constants, m[i] is the divided information block with 32 bits, the calculation method for s[i] is the integer part in s[i] = 232*abs(sin(i)), in which the unit of i is radian, x<<< k means x loop move left for k bits, f(b,c,d) is a Boolean function. The functions of four cycles loop are:

$F(b,c,d) = (b \wedge c) \vee ((\text{not } b) \wedge d)$

$F(b,c,d) = (b \wedge d) \vee (c \wedge (\text{not } d))$

$H(b,c,d) = b \oplus c \oplus d$

$I(b,c,d) = c \oplus (b \vee (\text{not } d))$

In the 16 steps of each cycle of loop, 16 information blocks with 32 bits are used by different orders. The first cycle is according to the original order; the next three cycles are used after exchange order, the exchange formulas are: ρ2(i) = (1+ 5i)mod16， ρ3(i) = (5+ 3 i) mod16， ρ4(i) = 7imod16. Presume that the 16 bytes blocks are m[0], m[1], …, m[15], the first cycle process with the order of 0…15, and the sub-block with i block processing in second cycle isρ2(i), for example, the 0 block is (1+ 5* 0)mod16= 1, which is m[1]; empathy, No 1 block is m[6], No 2 is m[11], and so on, set the order of byte blocks processing according to the corresponding exchange formulas in each cycle, and update the chain variables according to the orders of M (a, b, c, d)， M (d, a, b, c)， M (c, d, a, b)， M(b, c, d, a)in each cycle, which means update for 4 times for each chain variable; the value of k[i] and s[i] calculated according to steps, and each step has a certain constant. The constants and Boolean formula in the algorithm are all gotten from the long term experiment and research by Rivest, the purpose of which is to make the algorithm more complicated and less collision. For example, the first four steps in the 16 steps of the second cycle are:

M (a, b, c, d, m [1], 5, 0x f61e2562)

M (d, a, b, c, m [6], 9, 0x c040b340)

M (c, d, a, b, m [11] , 14, 0x265e5a51)

M (b, c, d, a, m [0], 20, 0xe9b6c7aa)

Four loop variables are upgraded in order. Every variable in the second cycle conducts for the third time according to exchange formulas and the same order and different k[i] and s[i]. Add initial value and value of a, b, c, d after 64 steps finishing by mode 32, and cascade, at last the MD5 code with 128 bits is calculated.

III. Application of MD5 Algorithm in Password Authentication

The premise of security of various internet applications is effective identification and authentication to identities. There are several identity authentication methods realizing the users' identification. Current commonly used methods are user name/password, digital signature, figure print, hardware ID card, and so on. And the user name/password is the easiest and most common method, which is to confirm the user validity through password matching. The general method is to build a form storing user information in the database, which should at least contain 3 fields, user name, password and privilege level. As user visits the system, it will verify the match of user name and password with the relevant information stored in the database. In general, user information in the database is transmitted and stored in the form of initial plain code, the security of which is obviously relatively low.

In order to increase the security of password to prevent users' password leaking, MD5 algorithm could be used to exchange the content of users' password. Realization method is to use MD5 algorithm to process the users' password, and store the changed value in database. The length of the password content is value of 128 bits. When verify the users' identifications, same process the password input by MD5 algorithm, and compare with the MD5 value stored in database, if the two values are the same, the user is valid. As the MD5 algorithm is irreversible, the value of information abstract calculated by MD5 cannot get the initial information by inverse algorithm. This method cannot only avoid the system administrator obtaining the password, but also increase the difficulties to decipher password in some degree.

IV. Application Security Analysis and Countermeasures Research

MD5 algorithm is mainly applied in digital signature, password authentication and other fields. For the sake of the password stored in database in the form of MD5 information abstract value, its security is threatened by MD5 being deciphered, and also by dictionary attack from hackers.

Considering that the attack to MD5 is mainly from collision attack, which means to find two different kinds of initial

information, the values of calculated information abstract should be same. For example, presume a kind of initial information (M1) calculates the value of information abstract being H by MD5 algorithm, if the other kind of initial information (M2) is found, the value of information abstract by MD5 algorithm is still H, then M1 and M2 are a pair of collision. Deciphering MD5 is the process of looking for collision. Once MD5 is deciphered, the password stored in database in the form of MD5 information abstract value will be in danger.

Otherwise, there are 3 methods attacking the password stored in database in the form of MD5 information abstract value. First, look up MD5 information abstract value online. Some websites provide the look-up service for MD5 value online, after inputting MD5 value, if it is in the database, the password value could be obtained soon. Second, use MD5 deciphering tools. There are many special tools to decipher MD5 through dictionary setting. Third, obtain of reset the users' passwords through social engineering. Therefore, simple MD5 encryption cannot be absolute safe. To the attack mentioned above, MD5 algorithm processing could be exchanged, or use encryption algorithm, interference by add information to make the information abstract value processed by MD5 in database is not simple MD5 information abstract value anymore.

A. Exchange the MD5 Processing

- Increase the times of MD5 processing. The basic thought of this method is to conduct MD5 algorithm for N times to the user's password needing encryption, which means conduct MD5 algorithm again to the obtained information abstract value with 128 bits. The practical times should be decided by the administrator.
- Intercept the information abstract value. The basic thought of the method is to use the user's password to conduct the MD5 algorithm first to get the MD5 information abstract value; and then select part of it (the first 18 bits for instance) to conduct the algorithm again to get the final result.
- Divide information abstract value. The basic thought of the method is to use the user's password to conduct the MD5 algorithm first to get the MD5 information abstract value (presume the value is H); then divide H into two groups of 64 bits each on right and left sides, and conduct MD5 algorithm separately to get the corresponding information abstract value, shown as HL and HR; make HR and HL connect as an alphabetic string, and conduct MD5 algorithm again to get the final result.

B. Additional Encryption Algorithm Interference

Before MD5 calculating, add an encryption algorithm to interfere the MD5 processing. The basic thought is to encrypt the user's password through self-defined encryption algorithm to get cipher code; get MD5 information abstract value by MD5 algorithm. There are countless kinds of self-defined encryption. Take user password "QDTONEWR" as an example to explain one of the kinds self-defined encryption algorithm. Presume the key is "xsw".

Step1：Take the length of key as the index to make the plain code as a matrix (fill by spaces with insufficiency), the key and its length are self-defined. The results are like below:

Q D T

O N E

W R _

Step2：Conduct exclusive-or operation to the elements in each row and corresponding character;

x s w

↑ ↑ ↑

Q D T

O N E

W R _

The results are:

) 7 #

7 = 2

/ ! W

Step3：Form intermediate processing string by row priority,)7/7=!#2w;

Step4：Move left for n bits (n is the length of the key) to form cipher code, 7=!#2W)7/

C. Additional Information Interference

During MD5 processing, add an alphabetic string with certain content and interfere the processed data. Its basic thought is to make the user name input and system time as additional information, and conduct MD5 processing to "user name + password + system time". To data through additional information interference processing, even the hacker decipher the MD5 algorithm, he is also difficult to lead out a comparison table with abundant characters from the dictionary established to decipher abundant users' passwords and decrease the possibility of password deciphered.

Under the condition without revising the MD5 algorithm, the methods described above are the effective methods to increase the password security. If some flaws cause the user password data exposure, the MD5 exchange algorithm discussed above could increase the difficulty to decipher the password greatly and increase MD5 security in the password authentication application.

V. CONCLUSION

The MD5 algorithm has resolved the transmission and storage by form of plain code and other security problems in traditional password though; it is also threatened by the collision attack and dictionary attack. It needs more secure methods. This essay has analyzed the application of MD5

algorithm and security in the password authentication, and suggests the method of the process of exchanging or interfering MD5 pointed to collision attack and dictionary attack to improve the application ofMD5 and security.

[1] Zhang Shaolan, Xing Guobo, Yang Yixian, Improvement to MD5 and Security Analysis [J]. Computer Application, 2009,vol.29(4):947-949

[2] Mao Ming, Qin Yinguang, Chen Shaohui, [J].Journal of Computer Application, 2009, 29(12):3174-3177

[3] Zhang Qing, Iterative Hashing Algorithm Base on MD5 [J]. Computer Engineering,2011,vol.37(18) 124-126

[4] Chen Shaohui, Zhai Xiaoning, Yan Na, Su Guoxing.Cryptanalysis of deciphering Mf5 algorithm, Computer Engineering And Applications, 2010, vol.46(19):109-112