Encryption of URL Address (Uniform Resource Locator) on PHP Web Base with MD5 Algorithm

1st Muhammad Basri *
Universitas Muhammadiyah Sumatera Utara
Medan 20238 Indonesia
mhdbasri@umsu.ac.id

2nd Oris Krianto Sulaiman
Dept. Informatic Engineering
Universitas Islam Sumatera Utara
Medan 20217 Indonesia
oris.ks@ft.uisu.ac.id

***Abstract*—The problem of an information (web base) is the security of the database displayed by the website manager. A common problem that is often encountered in information systems is how to make data displayed publicly have secure access without having to destroy the data in the web base system itself. A cryptographic system or often called a cipher is a system or collection of rules used to perform encryption and decryption. There are two kinds of cryptographic systems, namely secret key cryptographic systems or often referred to as symmetric cryptographic systems and public key cryptographic systems or often referred to as asymmetric cryptographic systems. MD5 is a one-way hash function created by Ron Rivest. MD5 is an improvement over MD4 after MD4 was successfully attacked by cryptanalysts. The MD5 algorithm takes as input a message of arbitrary size and produces a message digest that is 128 bits long.**

Keywords— Encryption, Web Base, PHP, URL, MD5

I. INTRODUCTION

Data security is very important for an information system (website). An information system is usually only intended for a certain group of individuals or communities, even an information system is intended for the general public. Therefore, it is very important to prevent important information from falling into the hands of unauthorized people. For this reason, many people have developed various ways to overcome data security problems in these information systems, so that unauthorized people cannot access or even damage the data intended for them.

One way to protect data is by using encryption techniques. Encryption is a process that changes a code from understandable to a code that cannot be manipulated by parties who are not entitled to access the data. The encryption technique is a system that is ready to be automated, so this technique is used to secure data in the system or that is being transmitted.

II. PROBLEM

With the wide scope and methods used in encryption, the discussion for encryption problems is given problem limitations. The problem limitations are

1. This encryption process (in the form of web pages) runs on all computer operating systems that have Apache Web Server installed.
2. This application uses the MD5 (*message digest*) encryption method.
3. This application is used only for the encryption process on the PHP web base url address to prevent Web Base injection (Sql-injection if there is a sql database).

III. LITERATURE REVIEW

Cryptography comes from the Greek language, consisting of two syllables: *crypto* and *graphia*. Crypto means to hide, while graphia means writing. Cryptography is the study of mathematical techniques related to aspects of information security, such as data confidentiality, data validity, data integrity, and data authentication (Menezes et al., 1996). But not all aspects of information security can be solved with cryptography.

Cryptography can also be defined as the science or art of keeping messages secure. When a message is sent from one place to another, the contents of the message may be intercepted by other parties who are not entitled to know the contents of the message. To protect the message, the message can be converted into a code that cannot be understood by other parties. Encryption is an encoding process that changes a code or message from an understandable one, called plaintext, into an incomprehensible code, called ciphertext. While the reverse process of converting ciphertext into plaintext is called decryption. The encryption and decryption process requires a certain mechanism and key.

A. *Cryptographic System*

A cryptographic system or often called a cipher is a system or set of rules used to perform encryption and decryption. There are two kinds of cryptographic systems, namely secret key cryptographic systems or often called symmetric cryptographic systems and public key cryptographic systems or often called asymmetric cryptographic systems.

B. *Secret Key Cryptography System*

A secret-key cryptographic system is a cryptographic system that uses the same encryption key as its decryption key. It requires the sender and receiver to agree on a specific key before they communicate with each other. The security of this system depends on the key, leaking the key means that others can encrypt and decrypt the message. For communication to remain secure, the existence of the key must remain secret. The nature of the key means that the sender must always ensure that the path to the message is secure.

used in key distribution is a secure path or ensuring that the person designated with the keys to be exchanged is a trustworthy person.

Problems will be complicated if communication is done together by *n* parties and for every two parties that exchange keys, there will be $C_2^n = \frac{1}{2}n(n-1)$as many secret keys that must be exchanged securely.

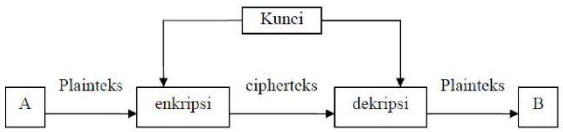

Figure 1. Secret Key Cryptography System

Examples of secret key cryptography systems are DES (*Data Encryption Standard*), Blowfish and AES (*Advanced Encryption Standard*).

C. *MD5 Algorithm*

MD5 is a one-way hash function created by Ron Rivest. MD5 is an improvement over MD4 after MD4 was successfully attacked by cryptanalysts. The MD5 algorithm accepts input in the form of a message of arbitrary size and produces a message digest that is 128 bits long. An overview of message digest generation with the MD5 algorithm is shown in Figure 2.

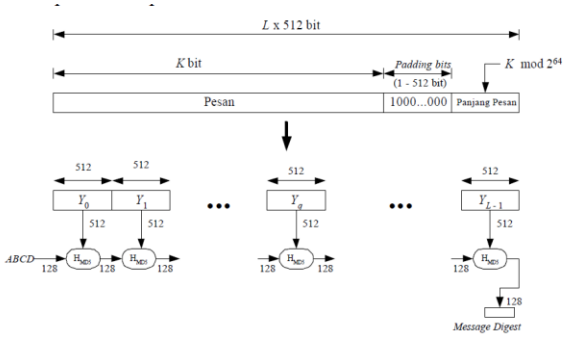

Figure 2. Generation of *message digest* with *MD5* algorithm

The steps for creating a message digest are as follows:

1. Addition of padding bits.
    - The message is appended with a number of buffer bits such that the message length (in bits) is congruent with 448 modulo 512. This means that the length of the message after adding the buffer bits is 64 bits less than a multiple of 512. The 512 number arises because MD5 processes messages in blocks of size 512.
    - Even messages with a length of 448 bits are still padded with bits. If the message is 448 bits long, then it is appended with 512 bits to become 960 bits. So, the length of the buffer bits is between 1 and 512.
    - The buffer bits consist of a 1 bit followed by the remaining 0 bits.

2. Addition of the original message length value.
   - The message that has been given the buffer bits is further augmented with 64 bits that represent the length of the original message.
   - If the message length is > 264 then what is taken is the length in modulo 264. In other words, if the original message length is K bits, then the 64 bits added express K modulo 264.
   - After adding 64 bits, the message length is now 512 bits.

3. MD buffer initialization.
   - MD5 requires 4 buffers, each of which is 32 bits long. The total buffer length is 4 x 32 = 128 bits. These four buffers hold the intermediate and final results.
   - These four buffers are named A, B, C, and D. Each buffer is initialized with values (in HEX notation) as follows:
     A = 01234567
     B = 89ABCDEF
     C = FEDCBA98
     D = 76543210

4. Processing of messages in blocks of 512 bits.
   - The message is divided into L blocks, each of which is 512 bits long (Y0 to YL - 1).
   - Each 512-bit block is processed along with the MD buffer into a 128-bit output, and this is called the HMD5 process. An overview of the HMD5 process is shown in Figure 3.

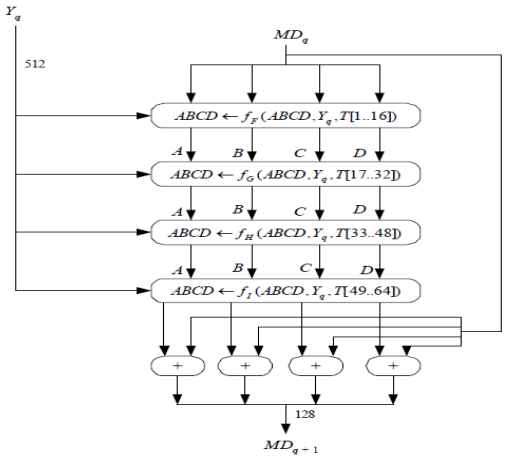

Figure 3. Processing of 512bit blocks (HMD5 process)

D. PHP Programming Language

PHP is the most widely used scripting language today. PHP is widely used to program dynamic websites, although it is possible to use it for other purposes. PHP has advantages over other programming languages, namely that PHP programming language is a script language that does not compile in its use and web servers that support PHP can be found everywhere from Apache, IIS, Lighttpd, nginx to Xitami with fairly easy configuration. PHP is very easy to develop, because there are many mailing lists and developers who are ready to help development. Besides having many advantages, PHP also has some disadvantages, namely in terms of language, PHP is not an ideal language for large-scale development. The main drawback is the absence of *namespaces*. *Namespace* is a way to group variable or function names in a hierarchical arrangement, in creating functions or classes in PHP can only be limited to one level. Because the programming is *embedded* (mixed with HTML), the development must be careful so that the application that will become large and medium to large, it is necessary to re-separate the HTML *template* and code / logic.

PHP stands for *Personal Home Page* or personal website (http://ilmukita.com/pengertian-php/) which is used as a server-side script language in web development that is inserted in HTML (*HyperText Markup Languages*) documents. The use of PHP allows the web to be made dynamic so that website maintenance becomes easier and more efficient.

PHP was first created by Ramus Lerdorf in 1994 (Peranginangin, Kasiman., 2006). Initially PHP was used to record the number and to find out who the visitors were on its homepage. Rasmus Lerdorf is a supporter of open source. Therefore, he released the *Personal Home Page Tools* version 1.0 for free, then added the capabilities of PHP 1.0 and launched PHP 2.0.
By 1996, PHP was widely used in websites around the world. A group of software developers consisting of Ramus, Zeew Suraski, Andi Gutman, Stig Bakken, Shane Caraveo, and Jim Winstead worked together to improve PHP 2.0. Finally, in 1998, PHP 3.0 was launched. Improvements continued to be made so that in 2000 4.0 was released. Not stopping there, PHP's capabilities continue to be added until the latest version released is PHP 5.0.x.

PHP has many advantages that are not owned by similar script languages. The advantages of PHP are:

1. PHP can be used on all operating systems, including Linux, Unix (including its variants HP-UX, solaris, and OpenBSD), Microsoft Window, Mac OS X, RISC OS.
2. PHP has the ability to process output images, PDF files, and Flash movies.
3. PHP can generate text such as XHTML and other XML files.

Program/Script syntax is written in PHP-specific tags. There are four kinds of PHP tag pairs that can be used to tag PHP script blocks:

1. <?php...>
2. <script language = "PHP> ...< /script>
3. <? ...?>
4. <% ...%>

Of the four methods above, methods 1 and 2 are the most commonly used.

E. PHP Identifier

Identifier is a name created by the programmer to give names to **variables, functions,** and **classes**. The identifier naming rules apply as follows:

1. Starts with a letter or underscore (_)
2. The next character can be a letter, number, or underscore (_)
3. It is *case sensitive*, except for functions that are already available in PHP that are *case sensitive*.
4. No punctuation is allowed.

IV. DISCUSSION

The result of this program implementation is to produce a php web base url that has been encoded so that not everyone can inject or inject which can damage data or even the web base itself.
The first experiment is the web base url without using encryption, shown in the following image:

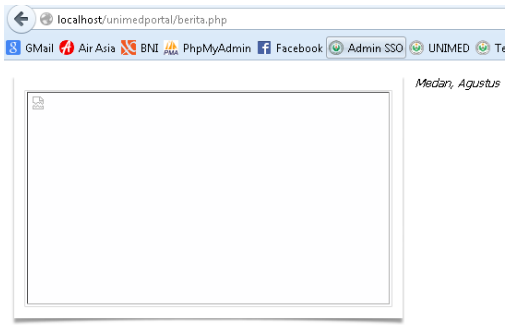

Figure 4. Web Base URL without The Use of Encryption.

In the above experiment when we type in the url without encryption it looks normal like a normal web base, but for professional people (crackers) this is a weak point of the web base so that several cracker tricks can be tried that aim to find out the data or information in the web base. For example, we enter a variable that is often used by crackers, namely the variable ?*id = 1*.

Figure 5. Web Base URL with id Variable

From the picture above, it can be concluded that the data on the web base uses one *id* column, so data breaches can be carried out on the web base (not explained in this discussion).

The second experiment is a web base url that uses encryption, shown in the following image:

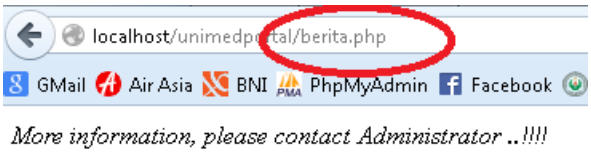

Figure 6. Web Base URL with Encryption in Use

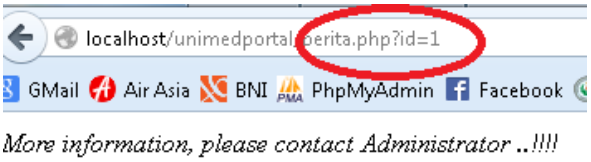

Figure 7. Web Base URL with id Variable

In both of the above images when we type the same web base url as in the first experiment, we will get a warning from the web base that we cannot access without being told the encrypted url. So if we have been given a url that has been encrypted then we can reopen the data on the web base.

Figure 8. Web Base URL with variable id and encrypted.

V. CONCLUSION

From the description that has been presented previously regarding the implementation of MD5 on this web base url, the following conclusions can be drawn: Web bases that do not use encryption or branding on their urls are very vulnerable to tampering, especially with the data in the web base. Encryption using the MD5 algorithm tends to make it difficult to access data freely. The MD5 algorithm on this web base is only a one-way process and matches the previous key.

REFERENCES

[1] Alfred J. Menezes, Paul C. van Oorschot dan Scott A. Vanstone, 1996, Handbook of Applied Cryptography, CRC Press, USA.
[2] Peranginagin, Kasiman., (2006), Aplikasi Web dengan PHP dan MySQL, Penerbit ANDI, Yogyakarta.
[3] Wikipedia, (2015), PHP, http://id.wikipedia.org/wiki/PHP Tanggal akses 24 Oktober 2015 Pukul : 16.00 WIB
[4] Skiena, Steven S.. The Algorithm Design Manual Second Edition. New York, USA : State University of New York. 2008
[5] Renaldi Munir,. Kriptografi,. Penerbit Informatika, Bandung . 2011