EPH - International Journal of Science And Engineering
ISSN (Online): 2454 - 2016
Volume 01 Issue 01 January 2015

DOI: https://doi.org/10.53555/eijse.v1i1.9

VULNERABILITY OF DATA SECURITY USING MD5 FUNCTION IN PHP DATABASE DESIGN

Neyole Misiko Jacob[1]*
**[1]*Lecturer, Collage of Human Resource Development, Jomo Kenyatta University of Agriculture and Technology. Kitale Campus*

Corresponding author:

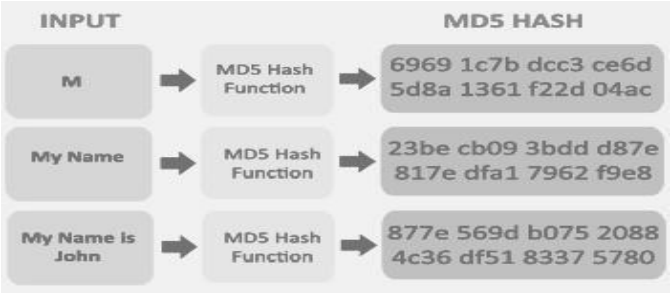
Abstract: -

Database integrity and security are vital measures in database development. Database designers in this respect have a number of approaches to secure data during database design and development. Some of the approaches include the use of the HASH functions such as the MD5 to encrypt database passwords. But with this comes technological challenges of attacks and vulnerability. Current new technologies such as the use of the has killers and descriptor's. This paper brings to light the challenges faced in using MD5 by database designers a give alternatives to secure data during database design.

Keywords: - *Md5 algorithms, Database design, DBMs,*

INTRODUCTION

MD5 stands for Message Digest algorithm 5, it is a widely used cryptographic hash function that was invented by Ronald Rivest in 1991. The idea behind this algorithm is to take up a random data inform of text or binary as an input and generate a fixed size hash value as the output. The input data can be of any size or length, but the output hash value size is always fixed. All the hash values share the following properties: Hash length which is the length of the hash value is determined by the type of the used algorithm, and its length does not depend on the size of the file. The most common hash value lengths are either 128 or 160 bits. The non-discoverability property where every pair of non- identical files will translate into a completely different hash value, even if the two files differ only by a single bit. Repeatability where each time a particular file is hashed using the same algorithm, the exact same hash value will be produced. Lastly Irreversibility where all hashing algorithms are one-way. Example as shown below.



Source: - Secondary data http://www.gohacking.com/what-is-md5-hash/

Each input size that one inputs during data entry in database the algorithm generates a fixed size 32 digit hex MD5 hash. The MD5 message-digest algorithm is a widely used cryptographic hash function producing a 128-bit hash values, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity in database designs. Apart from hiding password in databases, hash MD5 is Cryptographic hashing has a number of other uses and there are a vast number of algorithms such as MD4 designed to do a similar job. One of the main uses for cryptographic hashing is for verifying the contents of a message or file after transfer.

Background

MD5 is an algorithm that is used to verify data integrity through the creation of a 128-bit message digest from data input which may be a message of any length that is claimed to be as unique to that specific data as a fingerprint is to the specific individual [4]. The MD5 algorithm is an extension of the MD4 message digest algorithm. MD5 is slightly slower than MD4, but is more "conservative" in design.

According to Automated data classification drives security, storage convergence an Information Security magazine. Data classification products from a variety of start-ups are leading efforts to increase convergence of security, compliance and storage management. Many organizations have turned to use database s to store their volumetric data in databases whose designers and managers used the primitive Md5 algorithm to encrypt their passwords especially in the password field of the database tables.

But with sophistication of technology much of this security measures are vulnerable the hacking and cracking to reveal the hashed or encrypted information. Hash Killer website allows one to input an MD5 hash and search for its decrypted state in their database, basically, it's a MD5 cracker / decryption tool. According to the data they have on the decryptions in their database they had about 43.745 billion unique decrypted MD5 hashes since August 2007.

With advances in technology this security threats and problem measures need to be looked at to protect organization data that would otherwise cost an organization growth and development. Technologies such as cloud computing underpins an important part of economic activity today, and has the potential to make a major contribution to future growth. For many DBMs data integrity control such as control on the possible values a field can assume can be built into the physical structure of the field and controls enforced by the DBMs on the fields.

Problem definition

The security of the MD5 hash function is severely compromised. Currently using MD5 for file integrity may bring about practical problem. The attacks against MD5 are collision attacks, not pre- image attacks. This means an attacker can produce two files with the same hash, if he has control over both of them. Apart from this current technologies such as hash killers and crackers deem this security check a compromised function. The MD5 password hash algorithm is no longer considered safe by the original software developer, a day after the leak of more than 6.4 million hashed LinkedIn passwords.

Objectives of the study

The general objective to the study was to bring to the fore the compromises of data security in using md5 algorithm for password hash by database designers and database administrators. The specific objective of the study include:-

i) To create awareness on the weaknesses of data integrity using MD5.
ii) To sensitize database designers and DBA on the challenges experienced when dealing MD5.
iii) To propose alternatives to MD5 password hashing.

ISO requirement

The ISO 9000 family of quality management standards define quality as the features of a product or service which are required by the customer. Quality management is what an organization does to ensure that its products or services satisfy the customers' quality requirements and comply with applicable regulations. ISO/IEC 7498 Open Systems Interconnect (OSI) security model This multi-partite standard defines the OSI reference model, describing an architecture to secure network communications through security services such as access control, authentication, data integrity, data confidentiality and nonrepudiation and security mechanisms such as decipherment, digital signature, access control, data integrity, authentication exchange, traffic padding, routing control and notarization

HASH function and integrity

The principal objective of a hash function is data integrity. A "good" hash function has the property that the results of applying the function to a large set of inputs will produce outputs that are evenly distributed, and apparently random. The kind of hash function needed for security applications is referred to as a cryptographic hash function. A cryptographic hash function is an algorithm for which it is computationally infeasible because no attack is significantly more efficient than brute force to find either:-

(i) A data object that maps to a pre-specified hash result this is the one-way property.
(ii) Two data objects that map to the same hash result the collision-free property.
(iii) Hash functions determine whether or not data has changed.

Desired qualities of HASH functions

A good cryptographic hash function should be:

(i) Computationally infeasible to find data mapping to specific hash (one-way property)- robust enough to withstand brute force attacks
(ii) Computationally infeasible to find two data to same hash (collision-free property)-robust enough

The Specifications for good hash functions may include; essentially it must be extremely difficult to find two messages with the same hash. As well as the hash should not be related to the message in any obvious way such that it should be a complex non-linear function of the message.

Alternative methods to secure data

Encrypt passwords with SHA

MySQL offers a function called SHA () that applies an encryption algorithm to a string of text. The SHA function stands for Secure Hash Algorithm. The result is an encrypted string that is exactly 40 hexadecimal characters long, regardless of the original password length. So the function actually generates a 40-character code that uniquely represents the password. Since SHA () is a MySQL function, not a PHP function, you call it as part of the query that inserts a password into a table.

SHA functions

In cryptography, SHA-1 is a cryptographic hash, it produces a 160-bit (20-byte) hash value. A SHA-1 hash value is typically rendered as a hexadecimal number, 40 digits long. SHA stands for "secure hash algorithm". The four SHA algorithms are structured differently and are named SHA-0, SHA-1, SHA-2, and SHA-3. SHA-0 is the original version of the 160-bit hash function. SHA-1 is one of the most secure hash algorithms. It is used in SSL (Secure Sockets Level), PGP (Pretty Good Privacy), XML Signatures, and in Microsoft's Xbox etc. It is defined in the NIST (National Institute of Standards and Technology) standard 'FIPS 180-2'.

Salting

In cryptography, a salt is random data that is used as an additional input to a one-way function that hashes a password or passphrase. The primary function of salts is to defend against dictionary attacks versus a list of password hashes and against pre-computed rainbow table attacks. A new salt is randomly generated for each password. In a typical setting, the salt and the password are concatenated and processed with a cryptographic hash function, and the resulting output (but not the original password) is stored with the salt in a database. Hashing allows for later authentication while defending against compromise of the plaintext password in the event that the database is somehow compromised.

Conclusion

Database designers, security and risk professionals should keenly be aware that protecting data is vital to an organization's continued success development and growth. Data security is a high or critical concern for most organizations. As the threat situations with data and information continue to evolve, chief to the awareness and management should be up to data. Information security officers must adjust their risk management strategies accordingly to counter the next frontline.

References

[1]. Cuschieri Daniel and Po Wah Yau (2013). *Encryption in the cloud.* Royal Holloway Information Security Thesis Serie Encryption in the cloud. University of London. London. Accessed at http://searchsecurity.bitpipe.com/fulfilment/1397142251_919?ff=1

[2]. Fast Sum integrity control accessed at http://www.fastsum.com/support/md5-checksum-utility-faq/md5hash.php

[3]. Hashkiller.co.uk accessed at http://www.hashkiller.co.uk/md5-decrypter.aspx

[4]. Information Security magazine of December 2006. Accessed at http://searchsecurity.techtarget.com/magazineContent/Automated-data-classificationdrives security-sto age-convergence.

[5]. Rouse Margaret (2005) *MD5*. Accessed at http://searchsecurity.techtarget.com/definitinMD5

[6]. Rivest Ronald L. (April, 1992). *The MD5 Message-Digest Algorithm*. Massachusetts Institute of Technology Laboratory for Computer Science NE43-324 545 Technology Square Cambridge, MA 021391986. Accessed at http://www.ietf.org/rfc/rfc1321.txt

[7]. Srikanth Ramesh (January, 2010) what is MD5 Hash and How to Use it. Accessed at http://www.gohacking.com/what-is-md5-hash/

[8]. Wikipedia the free encyclopaedia accessed at http://en.wikipedia.org/wiki/MD5