SECURING PASSWORD AUTHENTICATION USING MD5 ALGORITHM

SITI NABILAH BINTI MOHAMMAD ZAIN

THESIS SUBMITTED IN FULFILMENT OF THE DEGREE OF COMPUTER SCIENCE (COMPUTER SYSTEM AND NETWORKING)

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING
2014

ABSTRACT

Process of authentication can be defined as developing a unique mapping process from given password to some other unique information in a defined domain. This project is about to make sure that system integrity is secure by using cryptographic based encryption method.Multi –application usability of password today forcing users to have proper memory aids. Apart from protection, a step toward perfect security has taken by using MD5 algorithm

TABLE OF CONTENTS

SECTION	CONTENT	PAGE
	SUPERVISOR'S DECLARATION	I.
	STUDENT'S DECLARATION	II.
	ACKNOWLEDGEMENT	III.
	ABSTRACT	IV.
	CONTENTS	V-VI
	LIST OF ABBREVIATION	VII
	INTRODUCTION	
1.0	Overview	1
1.1	Problem Statement	2
1.2	Objective	3
1.3	Scope	3
1.4	Methodology	4
1.5	Thesis Organization	5
1.6	Discussion	6
	LITERATURE REVIEW	
2.0	Overview	7
2.1	Hashing	8
2.2	Cryptographic Hashing	10
2.2.1	Purpose of Cryptographic Hashing Function	10
2.2.2	How Cryptographic Hashing Function Work	11
2.3	Comparison Between Cryptographic Hash Function	12
2.4	MD5 Algorithm	13
2.4.1	Overall Structure of MD5 Algorithm	16
2.5	Development Tools	17
	METHODOLOGY	
3.0	Overview	18

3.1	Methodology	19
3.1.1	Planning Phase	21
3.1.2	Analysis Phase	21
3.1.3	Design Phase	21
3.1.4	Coding Phase	22
3.1.5	Testing Phase	22
3.1.6	Troubleshooting and Support	22
3.1.7	Evaluation	23
3.1.8	Documentation Phase	23
3.2	Discussion	23
	DESIGN IMPLEMENTATION AND RESULT	
4.0	Overview	24
4.1	Proposed Design	25
4.2	Design Model	25
4.2.1	Flow Chart	26
4.3	Experimental Architecture and Expected Result	27
4.4	The Design and Implementation of The System	27
4.5	The characteristic of the System	34
	DISCUSSION AND CONCLUSION	
5.0	Overview	35
5.1	Advantage and Disadvantage	36
5.2	Recommendation for Future Work	36
	REFERENCES	37

LIST OF ABBREVIATION

CRC	Cyclic Redundancy Check
MAC	Message Authentication Codes
MDC	Message Detection Codes

CHAPTER 1

INTRODUCTION

1.0 Overview

Security is broad topic and covers many issues. Malicious people trying to gain some benefit, attention or to harm some intentionally cause most security problems [1]. Hashing is a procedure that is used in sorting and retrieving the information about the database [2]. The information is associated with key properties and makes use of individual character, numbers in key itself. In hashing transformation of a string of characters into frequently shorter fixed-length value or key that represent the original string is done [3]. It's really tough to do the work in a faster manner like to discover the item using the shorter hashed key than to find it using the original value so for this reason hashing is very capable. Moreover, it is also used in many encryption algorithms. Hashing is a procedure that is used in sorting and retrieving the information about the database. The information is associated with key properties and makes use

of individual character, numbers in key itself. It's really tough to do the work in a faster manner like to discover the item using the shorter hashed key than to find it using the original value so for this reason hashing is very capable. Moreover, it is also used in many encryption algorithms [4]

1.1 Problem statement

1. According to Micheal in his article A logic of Aunthentication , in distributed system and similar networks of computers it is necessary to have procedures by which various pairs of principle(people, computers, services) satisfy themselves mutually about each other's identity [5].
2. A common way to approach this is by means of secret, usuallly encryption keys. In Barrest outline, an authentication protocol guarantees that if the principals really are who they say they are then will end up in possesion of one or more shared secrets in example in Needham & Schroeder [3].
3. Most of common authentication system, though not safest. Both of them are saving in clear text. It should be encrypting to security.

1.2 Objective

1. To identify common technology in securing password
2. To implement authentication of securing password using hashing technique.
3. To test the accuracy of the technology of securing data integrity.

1.3 Project Scope

The scope of this project is to provide secure authentication password of web based application. Password hashing is mainly used in password authentication.

The scope of this project also important to make sure the flow of the system will run smoothly.

1. The system only can be access by system's administrator but must meet the type of user that has different permission.
2. System can only be managed on platform that has operating system installed with the web server that provides *Hypertext Preprocessor (PHP)* services and MySql database architecture.
3. System can be view by any web browser that have connection with internet

1.4 Methodology

There are a several steps should be performed in this study to build the project successfully. Figure 1.1 shows the steps to complete this project:

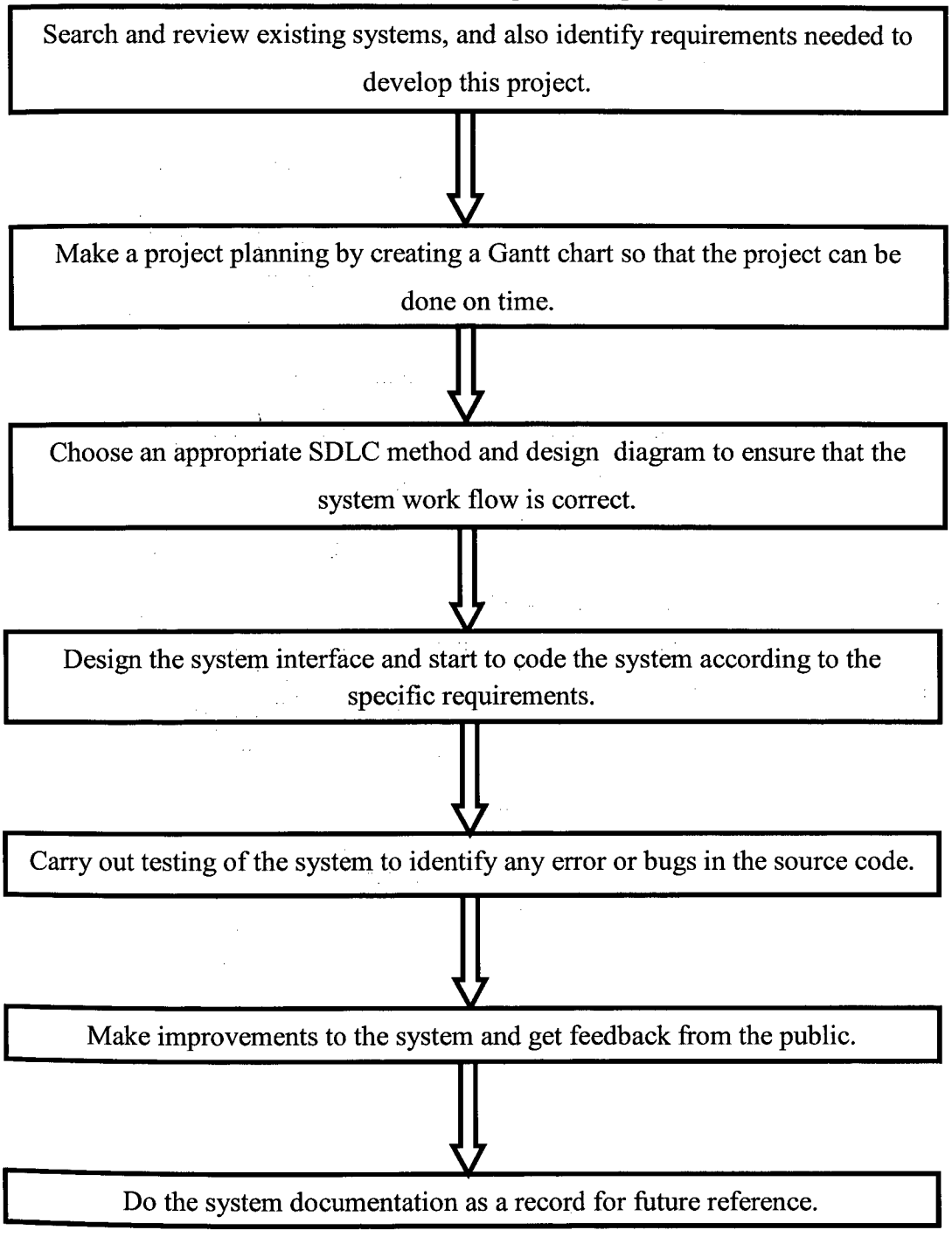

Figure 1.1 Project development flow chart

1.5 Thesis organization

This thesis consists of five chapters. Chapter one will discuss on introduction to research of cryptology hash function for securing password.

Chapter two will discuss on literature review of system well. The discussion on this chapter consist three sections: The first section is describes the hashing. The second section will explain about the cryptographic hashing function and comparison between the cryptographic hashing algorithms. The third section introduces the chosen algorithm that will be use to develop this system. The last section will briefly explain about the development of the system. Cryptographic hashing is basically used in computer and network security. Like for checksum and authentication purpose.

In chapter three, it will describe on how project was done, explanation about activity of research development. This project will develop by using waterfall SDLC.

Chapter four will discuss about design of the project. The development of framework and model through flow work should be done. Continuously designing the research which include any planning of data analysis. This chapter will focus on analysis and design of the project based on the requirement

Last chapter which is chapter five, conclusion is made for research that has been done. The part of conclusion consist of conclusion of the research, all data retrieve and observe how far

it been fit into the research and its objectives, methodology and research implementation conclusion and lastly future suggestion and enhancement of research topic or technique.

1.6 Discussion

In fact, the system administrator is always responsible for making decisions about how to control and solve authentication issues. Therefore, they have to make good decisions to select great quality algorithm in efforts to make sure the authentication process is secure enough for any existing issues associated to the server can be resolved. However, before developing new system for the authentication process, depth study and research must be performed regarding the server and software. The system must also be tested for actual network environment whether it is ideal to use. That is what going to be done in the next chapter of literature review where a very detail of study and analysis regarding the system.

CHAPTER 2

LITERATURE REVIEW

2.0 Overview

On this chapter, the project will focus on literature reviews based on the architecture of the project to be developed later. According to Dena , Director of Health Science Writing Center from University of Toronto (2007); "In writing the literature review, your purpose is to convey to your reader what knowledge and ideas have been established on a topic, and what their strengths and weaknesses are." From this statement, the literature review is like way to gain the ideas on how the developed system is going to be.

The discussion on this chapter consist three sections: The first section is describes the hashing. The second section will explain about the cryptographic hashing function and comparison between the cryptographic hashing algorithms. The third section introduces the chosen algorithm that will be use to develop this system. The last section will briefly explain about the development of the system

2.1 Hashing

Figure 2.1 shows the functioning maps of input data with hash function.

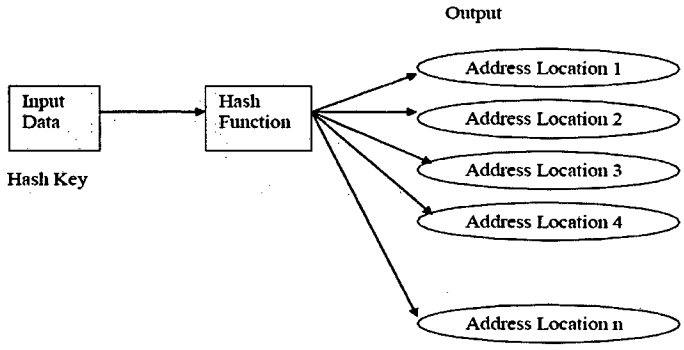

Figure 2.1 Hash Function

Based on figure 2.1, Hash functions are functioning that map input of arbitrary length to a string fixed length, the hash code. It is a procedure that is used in sorting and retrieving the information about the database. Input data will locate to address location,. There are many address location will be locate. The information is associated with key properties and makes use of individual character, numbers in key itself. It's really tough to do the work in a faster manner like to discover the item using the shorter hashed key than to find it using the original value so for this reason hashing is very capable [1]. Moreover, it is also used in many encryption algorithms.

There are four types of hashing which is cyclic redundancy checks, checksums, non-cryptographic hash functions and cryptographic hash functions .

Checksum or hash sum is a small size datum compute from an arbitrary block of digital data for the purpose of detecting errors that may have been introduced during its transmission or storage.

A cyclic redundancy check (CRC) is an error-detecting code commonly used in digital network and storage device to detect accidental changes to raw data. Blocks of data entering these systems get a short check value attached, beside on the remainder of a polynomial division of their contents: on retrieval the calculation is repeated, and corrective action can be taken against presumed data corruption if the check value does not match.

Cryptographic hashing is basically used in computer and network security. Like for checksum and authentication purpose. Cryptographic hashing is basically used in computer and network security. Like for checksum and authentication purpose.

2.2 Cryptographic Hashing Function

Cryptographic hashing is basically used in computer and network security. Like for checksum and authentication purpose [3].

The use of cryptographic hash function like MD5 or SHA1 for message authentication has become standard approach in many Internet applications and protocols. Though very easy to implement, these mechanisms are usually based on ad hoc technique that lack a sound security analysis [5].

Cryptographic hash function map strings of different lengths to short, fixed-size, outputs. In addition to the basic collision property, cryptographic hash function are usually designed to have some randomness-like properties, independence of input/output, unpredictability of the output when the part of predictability are unknown. Not only do these properties help in making it harder to find collision, but also help to randomize the input presented to the signature algorithm.

2.2.1 Purpose of cryptographic hashing function.

Cryptographic hash function is mainly used for security reason like information, notably in digital signatures, message authentication codes (MACs) and other forms of authentication. As it major applications are password verification related applications and to authenticate a user, the password presented by the user is hashed and compared with the stored hash. This is sometimes referred to one way encryption [3]. For integrity purpose of message verification, cryptography can detect any changes applied to message or not, by comparing message content calculated

before, after transmission. Security and performance reason like digital signature algorithms specify that only the digest to identify a file in a several source code management system with reliability. Focused on password hashing instead of storing password it also can store the password, the system can store a hash of password.

According to SANS Institute Reading Room Site , another application of cryptographic hash functions is data authentication. Data authentication is the process is the process of being able to verify the source of data. With data authentication, one can distinguish messages originating from the intended sender and an attacker [2].

2.2.2 How Cryptographic Hashing Function Work

When a password is supplied, it computes the password's hash and compares it with the stored value. If they match, the password is deemed correct. If the hashed password file is obtained by an attacker, it is not immediately useful because the password can't be derived from the hashed.

Refer to the article by John the SHA1 and MD5 algorithms are considered secure because there are no known techniques to find collision except by brute force clearly the properties of cryptographic hash functions have many application in the realm of computer security [5].

2.3 Comparison between cryptographic hash function.

The use of cryptographic hash functions like MD5 or SHA-1 for Message authentication has become a standard approach in many applications, Particularly Internet security protocols. Though very easy to implement, these mechanisms are usually based on ad hoc techniques that lack a sound security analysis. The constructions and analysis presented here are free from any dependency on the peculiarities of the underlying hash function. We only exploit the general structure of functions like MD5 and SHA-1, as being built on top of a basic compression function which works on fixed length messages, and is then iterated multiple times in order to process variable length inputs [1].

Comparison between MD5 algorithm and SHA1 algorithm.

- Variation of SHA1 algorithm it will produces longer digest than MD5 algorithm.
- Note that SHA1 more costly to compute a message than MD5 algorithm.
- MD5 have shorter time to compute than MD5 to produces larger output.

2.4 MD5 algorithm

The MD5 message digest algorithm takes as input a message of arbitrary length and produces as output a 128-bit message digest. MD5 is intended by its authors for use in digital signature applications. These applications require that it be computationally infeasible to produce two message having the same digest, for it is the digest which is signed, not the original message [2].

Figure 2.2 briefly describe about the process of MD5 algorithm step by step [8].

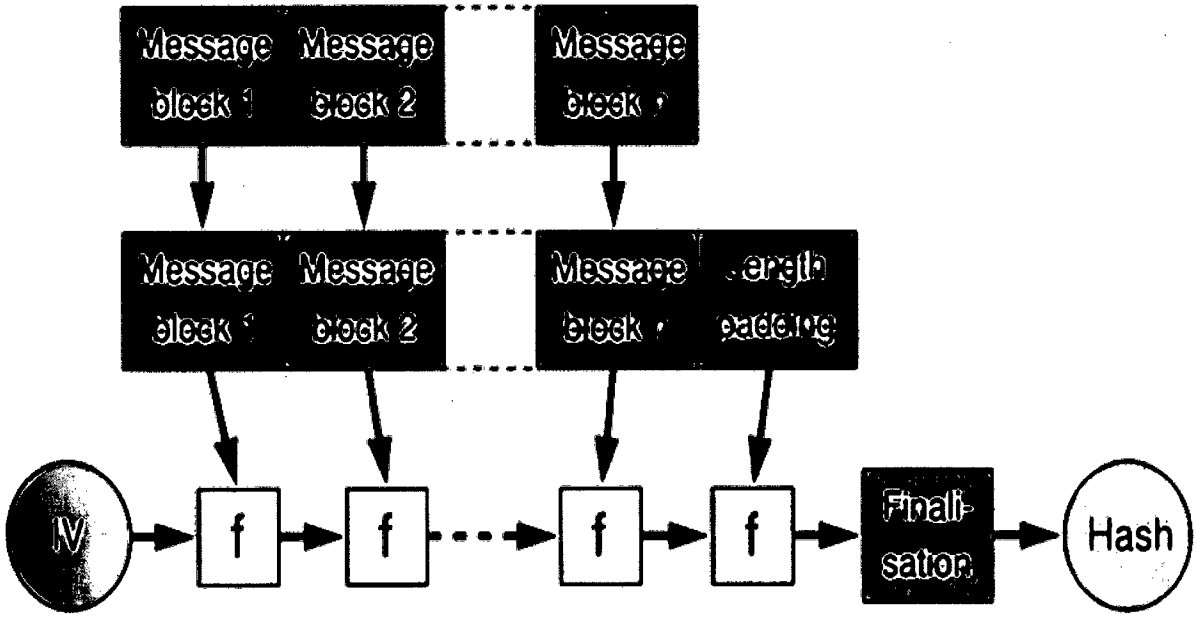

Figure2.2 : MD5 Algorithm

There are four step involves in MD5 algorithm which is:

- Step 1 – Append padded bits

  For step 1, the message is padded so that its length is

  congruent to 448, modulo 512. Its Means extended to just 64 bits shy of being of

  512 bits long. A single "1" bit is appended to the message,

  and then "0" bits are appended so that the length in bits equals 448 modulo 512.

- Step 2 – Append length

  For stetp 2, A 64 bit representation of b is appended to

  the result of the previous step. The resulting message has a length that is an exact multiple of 512 bits

- Step 3 – Initialize MD Buffer

  A four-word buffer (A,B,C,D) is used to compute the message digest. Here each of A,B,C,D, is a 32 bit register. These registers are initialized to the following values in hexadecimal:

  word A: 01 23 45 67

  word B: 89 ab cd ef

  word C: fe dc ba 98

  word D: 76 54 32 10

- Step 4 – Process message in 16-word blocks.

Four auxiliary functions that take as input three 32-bit words and produce as output one 32-bit word.

$$F(X,Y,Z) = XY \text{ v not}(X) Z$$

$$G(X,Y,Z) = XZ \text{ v } Y \text{ not}(Z)$$

$$H(X,Y,Z) = X \text{ xor } Y \text{ xor } Z$$

$$I(X,Y,Z) = Y \text{ xor } (X \text{ v not}(Z))$$

If the bits of X, Y, and Z are independent and unbiased, the each bit of F(X,Y,Z), G(X,Y,Z), H(X,Y,Z), and I(X,Y,Z) will be independent and unbiased. The message digest produced as output is A, B, C, D. That is, output begins with the low-order byte of A, and end with the high-order byte of D.

2.4.1 Overall Structure of MD5 algorithm.

MD5 uses the following four functions (one for each round) to process the input. They all take a 3-word input and produce a single word of output [8].

	Round i			
	1	2	3	4
si1	7	5	4	6
si2	12	9	11	10
si3	17	14	16	15
si4	22	20	23	21

Figure 2.3 : The 16 different shift constant of MD5

Step 1 of round 1 through 48 (step 16 of round 4). For a complete specification of MD5 the reader is referred to the original description. Note that in this original description of MD5 the designation f.g,h and I is used for respectively the round function f1, f2, f3, and f4.

Derivation of the round function input condition of MD5 algorithm .

The basis of the md5 algorithm is a compression function G that takes as input a 4-word buffer (A, B, C, D) and 16 word message block (X[0], X[1], ...i X[15]), and produces a 4-word output (AA, BB, CC, DD)

(AA, BB, CC, DD) =G((A, B, C, D) (X[0], X[1],....x[15]))

2.5 Development Tools

This system will be developing using selected tools to support the development of the system. Above table is the description of the tools:

1.1.1 Software and Hardware Tools

The hardware and software for development of the project are identified.

No	Name	Quantity	Purpose
1	Laptop / Computer	1	Research,development, documentation
2	Printer	1	To print thesis.
3	Pendrive 8 GB	1	Data backup
4	Compact disk (CD)	1	Burn the project

Table 1: List of hardware

No	Name	Purpose
1.	Adobe Dreamweaver CS5.5	To develop MD5 Algorithm system
2.	Xampp	For database
3.	Microsoft Office Visio 2007	Draw diagram
4.	Microsoft Office Word	Documentation
5.	Microsoft Power Point	Presentation
6.	Google Chrome	Research

Table 2: List of Software

CHAPTER 3

METHODOLOGY

3.0 Overview

In order for computer system to perform specific acts on behalf of specific individual, an identification and authentication step is needed. In this chapter the step will briefly explained. The specific selection of methodology, project methods, tools and techniques can ensure the accurate and appropriate project schedule is used. A methodology is usually a guideline system for solving a problem, with specific components such as phases, tasks, methods, techniques and tools [9]. In developing the project, the selected methods, techniques and tools that are accurate and appropriate play a very important role to ensure the smooth and orderly scheduled project. The use of a systematic methodology has been able to produce software that is easy to maintain, low dependency rate, effective in use, and easy to use. This chapter consists of six sections mainly. The first section will be discussing about the project development phases. Not all phases will be included in this study. The next section will be discussing about the system requirements in developing this study. The system requirement can be categorize into two parts which are hardware requirement and software requirement.

3.1 Methodology

This is according to Wasson (2005), where the waterfall model is a model for the development of an activity For the methodology, it is needed to explain the steps required to build this system. Each phase should be explained and state the activity to be carried out. In the process of this methodology, it is reasonable to choose the waterfall SDLC where it is commonly used to develop a project.

Figure 3.1 shows the waterfall SDLC for this project which consists of eight phases. To all phases are conducted sequentially, in which progress can be seen as flowing steadily downwards through the phases of planning, analysis, design, coding, testing, troubleshooting and support, evaluation and documentation.