technische universität
dortmund
Faculty for Computer Scinece

Bachelor's Thesis

Collsisions in Hash-Functions

Marco Bellmann

First Reviewer:
Prof. Dr. Coja-Oghlan

Second Reviewer:
Msc. Arnab Chatterjee

Dortmund, August 2023

Contents

Chapter 1

Introduction

1.1 May Contains

1. about hashfunctions
2. Definition of md5
3. my impelmentation of md5
   (a) padding (missing in Stevens code)
   (b) md5compress
   (c) potential for improvement
4. collisions for md5
   (a) Blocks 1 and 2 finding
   (b) 00, 01, 10, 11
   (c) MMM
5. conclusion
6. Definition of SHA1
7. why not in SHA1
8. notation helper

Notation

Stevens	Wang	Definition
$h(m) = H$		the hash H of some message m
$RL\,(X, Y)$	$ROTL^Y\,(X)$	cyclic left shift X by Y (mod 31)
$RR\,(X, Y)$	-	cyclic right shift X by Y (mod 31)
$RC\,(t)$	$S\,(t)$	rotation Constant of t
Block 1, Block 2	Block N, Block M	pair of first blocks for collisions finding
Block 0, Block 1	Block N, Block M	same pair but im Code

Motivation

This thesis is about a deeper look on MD5. We take a closer look at the Master thesis of M. Stevens: a fast collisions finding algorithm [1]. The goal is to work out a more clear and understandable code, which is not necessarily faster, to reevaluate the code on modern systems and the difference to SHA1.

Notes

Example for Stevens bit manipulation for Q_t mit $t = 3$:
Bit condition for $Q_t|t = 3$:

1. the val to set the zeros (and with 0xfe87bc3f)
2. the val to set the ones (or with $0x017841c0$)
3. the new bit Conds
4. the old bit Conds

1.&	11111110	10000111	10111100	00111111	$0xfe87bc3f$
2.$\|\|$	00000001	01111000	01000001	11000000	$0x017841c0$
3.	....	.1111...	.1....01	11......	
4.	....	....0...	....0...	.0......	

the & flips the 0 correct, the $||$ flips the 1 correct

1.2 Notation Helper and Ideas

$RR\,(X, Y) :\equiv X \circledgtr Y$
$RL\,(X, Y) :\equiv X \circledlt Y$
$X \oplus Y :\equiv X"XOR"Y$

$$10000000000000000000000000000001 \circledlt 7 \equiv 00000000000000000000000011000000$$

Chapter 2

About MD5

2.1 Definition

MD5 stands for *Message-Digest Algorithm 5* since it generates a digest for any given text. MD5's output length is 128 Bit, represented in hexadecimal. Since the amount of possible text it close to and the amount of MD5s is limited by 32^{16}, some text may have the MD5. If two different inputs create have the same hash value, that is what we call a collision. MD5 is usually seen as 4 steps as seen in fig md5

1. padding
2. procssesing
3. md5 sum
4. output

Figure 2.1: MD5 algo

Chapter 3

Collision Finding

3.1 About collisions

Hash functions have resistance:

1. First Preimage Resistance: for a hash function $h(m) = H$ the massage m is hard to find.
2. Second Preimage Resistance: for a given messages m_1 it is hard to find an m_2 with $m_1 \neq m_2$ and $h(m_1) = h(m_2)$.
3. Collision Resistance: two arbitrary messages m_1 and m_2 with $\neq m_2$ and $h(m_1) = h(m_2)$ are hard to find.

3.2 Differential Path

Stevens starts with Wangs attack, which trys to finde to pairs of blocks: (B_0, B_0') and (B_1, B_1') that $IHV = IHV'$, with the goal to create two massages M and M', with the same hash value:

$$\begin{array}{ccccccccc}
IHV_0 & \xrightarrow[M_{(1)}]{}\cdots & \xrightarrow[M_k]{} IHV_k & \xrightarrow[B_0]{} & IHV_{k+1} & \xrightarrow[B_1]{} & IHV_{k+2} \xrightarrow[M_{k+1}]{} & \cdots \xrightarrow[M_N]{} & IHV_N \\
= & & = & & \neq & & = & & = \\
IHV_0 & \xrightarrow[M_{(1)}]{}\cdots & \xrightarrow[M_k]{} IHV_k & \xrightarrow[B_0]{} & IHV'_{k+1} & \xrightarrow[B_1]{} & IHV'_{k+2} \xrightarrow[M_{k+1}]{} & \cdots \xrightarrow[M_N]{} & IHV_N
\end{array}$$

The idea to manipulate a block B such that $Q_1 \ldots Q_{16}$ maintain their conditions and that Q_17 to some Q_k do not change at all. We try to make k as large as possible.

3.3 Bit Conditions

Bit conditions describe the differential path on bits. We need the bit conditions to avoid a carry, so a manipulation in step t stays in step t and does not propagate beyond the 31st bit. We look at conditions and restrictions. The restrictions leads to conditions, which we calculate in the following. A restriction e.g. $\Delta T_2\,[\ 31] = +1$ leads to conditions $Q_1\,[16]\,Q_2\,[16] = Q_3\,[15] = 0$ and $Q_2[15] = 1$. Notice, conditions are on $\Delta T_t[i]$ a state in md5-algorithm before the rotation and restrictions are on $Q_t[i]$ states of the md5-algorithm after the rotation.
We calculate the bit conditions by using the Add-Difference for two massage blocks containing tow blocks $N|M$ and $N'|M'$. The XOR-Difference is useful, too.

$$\begin{aligned} \delta X &= X' - X\ (mod 32)\ \text{ Add-Difference} \\ \Delta X &= X' \oplus X \text{ XOR-Difference} \\ \lambda\,[i] &= \text{ our guess for the ith bit: } X\,[i] \end{aligned}$$

$$\text{if } \Delta X = \lambda \Rightarrow \delta x \text{ can be determined}$$

For $\lambda\,[i]$ we only need to consider $i < 31$, since $X\,[31]$ as msb always creates a add difference of 2^{31}.

We calculate a δ for each f_t, Q_t, T_t and R_t for our add difference, to calculate Q_{t+1} . Additional we need the rotation constant RC for each t. In general we begin with the f_t since we want f_t to be in a particular state. Since we want to avoid a carries in our calculation

t	$RC(t)$
0	7
1	12
2	17
3	22
4	7
5	12
6	17
7	22
8	7
9	12
10	17
11	22
12	7
13	12
14	17
15	22

$$\begin{aligned} \delta F_t &= f\left(Q'_t, Q'_{t-1}, Q'_{t-2}\right) \\ \delta T_t &= \delta F_t + \delta Q'_{t-3} + \delta W_t \\ \delta R_t &= RL\left(T'_t, RC_t\right) - RL\left(T_t, RC_t\right) \\ \delta Q_{t+1} &= \delta Q_t + \delta R_t \end{aligned}$$

1. $t \in \{0, 1, 2, 3\}$:
   $Q_t = 0$ since here is no influence by an message and no calculation of f, there is nothing to change:

2. $t = 4$
   $\Delta T_4 = -2^{31}$, because we must not have a carry, we *lock*

the last bit. Since $RL(T_4, RC_4) = RL(-2^{31}, 7) = -2^6$ and
$\delta Q_4 = 0 \Rightarrow \delta Q_5 = -2^6$

$$m_t = RR\left(Q_{t+1} - Q_t, RC_t\right) - f_t\left(Q_t, Q_{t-1}, Q_{t-2}\right) - Q_{t-3} - AC$$

Appendix A

Tables

+—+
|abc|
+—+

Appendix B

Code

a +=a;

List of Figures

Algorithmenverzeichnis

Bibliography

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie Zitate kenntlich gemacht habe.

Dortmund, den May 31, 2023

Muster Mustermann