University of Technology, Graz

Master Thesis

Differential cryptanalysis with SAT solvers

Author:
Lukas Prokop,
BSc BSc

Supervisor:
Dipl.-Ing. Dr.techn.
Florian Mendel

A thesis submitted in fulfillment of the requirements for the master's degree in Computer Science

at the

Institute of Applied Information Processing and Communications

March 29, 2016

Abstract

Hash functions are ubiquitous in the modern information age. They provide preimage, second preimage and collision resistance ensuring data and origin integrity. They are used as cryptographic primitives in protocols and applications.

In August 2006, Wang et al. showed efficient attacks against several hash function designs including MD4, MD5, HAVAL-128 and RIPEMD. With these results differential cryptanalysis has been proven useful to break collision resistance in hash functions. Over the years advanced attacks based on those differential approaches have been developed.

In this thesis we encode differential attack settings as SAT problems. SAT solvers are utilized to solve the problem revealing actual message collisions for a defined message difference.

Keywords: hash function, differential cryptanalysis, MD4, collision resistance, satisfiability, SAT solver

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

(digital signature)

Acknowledgements

First of all I would like to thank my academic advisor for his continuous support during this project. Many hours of debugging were involved in writing this master thesis project, but thanks to Florian Mendel, this project came to a release with nice results.

I would also like to thank Maria Eichlseder for her great support. Her unique way to ask questions brought me back on track several times. Mate Soos supported me during my bachelor thesis with SAT related issues and his support continued with this master thesis in private conversations.

Also thanks to Roderick Bloem and Armin Biere who organized a meeting one year before submitting this work defining the main approaches involved in this thesis.

And finally I am grateful for the support by Martina, who also supported me during good and bad days with this thesis, and my parents which provided a prosperous environment to me to be able to be where I am today.

Thank you!

All source code is available at lukas-prokop.at/proj/megosat and published under terms and conditions of Free/Libre Open Source Software. This document was printed with LuaLaTeXand Linux Libertine Font.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Hash functions are used as cryptographic primitives in many applications and protocols. They take an arbitrary input message, considered as byte string, and compute a hash value. The hash value is of fixed length and satisfies several properties which make them useful to enable data and origin integrity.

1.1 Hash Function Properties

Definition 1 (Hash function)**.** A *hash function* is a mapping $h : X \to Y$ with $X = \{0,1\}^*$ and $Y = \{0,1\}^n$ for $n \in \mathbb{N}_{\geq 1}$.

- Let $x \in X$. $h(x)$ is called *hash value of* x.
- Let $h(x) = y \in Y$. y is called *preimage of* y.

One example showing the use of hash functions as primitives is PKCS #5 specified in RFC 2898 [2]. Section 5.2 specifies PBKDF1 and PBKDF2 using hash algorithms as pseudorandom functions to derive password-based keys. Given a minimum iteration count of 1000 as defined in section 4.2, it suggested to allow fast computations of hash value for given $x \in X$.

Definition 2 (Preimage resistance)**.** Given $y \in Y$. A hash function h is preimage resistant iff it is computationally infeasible to find $x \in X$ such that $h(x) = y$.

Definition 3 (Second-preimage resistance)**.** Given $x \in X$. A hash function h is second-preimage resistant iff it is computationally infeasible to find $x_2 \in X$ with $x \neq x_2$ such that $h(x) = h(x_2)$. x_2 is called *second preimage*.

Definition 4 (Collision resistance)**.** A hash function h is collision resistant iff it is computationally infeasible to find any two $x \in X$ and $x_2 \in X$ with $x \neq x_2$ such that $h(x) = h(x_2)$.

As far as hash functions accept input strings of arbitrary length, but return a fixed size output string, existence of preimages and collisions is unavoidable [8]. However, good hash functions make it very difficult to determine collisions or preimages.

1.2 Cryptanalysis of Hash Functions

In August 2004, Wang et al. published results at Crypto'04 [10] which revealed that MD4, MD5, HAVAL-128 and RIPEMD can be broken practically using differential cryptanalysis. On an IBM P690 machine, an MD5 collision can be computed in about one hour using this approach. Collisions for HAVAL-128, MD4 and RIPEMD were found as well.

TODO: discuss their approach in more detail, introduce diff cryptanalysis

1.3 Boolean functions

TODO: and, or, not

Definition 5. A *boolean function* is a mapping $h : X \to Y$ with $X = \{0, 1\}^n$ for $n \in \mathbb{N}_{\geq 1}$ and $Y = \{0, 1\}$.

Definition 6. An *algorithm* is a step-wise set of instructions to solve a problem.

1.4 Satisfiability

TODO

Definition 7. A boolean function is *satisfiable* iff there exists at least one input $x \in X$ such that $h(x) = 1$.

Theorem 1. Every algorithm can be represented as boolean function.

1.5 Thesis Outline

This thesis is organized as follows.

In Chapter 1 we discussed the basic properties and fundamentals of the tools in discussion including hash functions and SAT solvers.

In Chapter 2 we introduce the MD4 hash function and discuss possible approaches in differential cryptanalysis.

In Chapter 3 we discuss SAT solving and potential approaches to speed up SAT solvers for cryptographic problems.

In Chapter 4 we show results of our work and discuss its implications.

Chapter 2

Differential cryptanalysis

2.1 MD4

MD4 is a cryptographic hash function originally described in RFC 1186 [5], updated in RFC 1320 [6] and obsoleted by RFC 6150 [9]. It was invented by Ronald Rivest in 1990 with properties given in Table 2.1. Since 1995 [1] successful attacks have been found to break collisions, preimage and second-preimage resistance in MD4; including but not limited to [7] and [3]. A Python 3 implementation derived from a previous Python version is available at github [4].

block size	512 bits	namely variable block in RFC 1320 [6]
digest size	128 bits	as per section 3.5 in RFC 1320 [6]
internal state size	128 bits	namely variables A, B, C and D
word size	32 bits	as per section 2 in RFC 1320 [6]

Table 2.1: MD4 hash algorithm properties

A short summary of MD4's design is given: Once padding and length extension has been applied, the message is split into 512-bit blocks (i.e. 16 32-bit words). Four state variables A, B, C and D are initialized with hexadecimal values:

$$[\mathbf{A}]\ 01234567 \quad [\mathbf{B}]\ 89\mathrm{abcdef} \quad [\mathbf{C}]\ \mathrm{fedcba}98 \quad [\mathbf{D}]\ 76543210$$

To process one block, three auxiliary functions are defined:

$$F(X,Y,Z) = (X \wedge Y) \vee (\neg X \wedge Z) \tag{2.1}$$

$$G(X,Y,Z) = (X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z) \tag{2.2}$$

$$H(X,Y,Z) = X \oplus Y \oplus Z \tag{2.3}$$

2.2 Differential notation

2.3 Addition example

2.4 Differential path

Chapter 3

Satisfiability

“What idiot called them logic errors rather than bool shit?”

3.1 Basic SAT solving techniques

3.2 SAT solvers in use

3.3 Encodings

Chapter 4

Results

4.1 Benchmark results

Chapter 5

Conclusion