Internship proposal — 2014
Preimage search for the MD4 hash function

Lab.: Inria/IRISA, Rennes
Lab. head: Jean-Marc Jézéquel
Topic: Hash function cryptanalysis
Internship adviser: Pierre-Alain Fouque, Pierre-Alain.Fouque@irisa.fr
Internship co-adviser: Pierre Karpman, Pierre.Karpman@inria.fr

Context

Hash functions are one of the main primitives of symmetric cryptography. They have many applications, like *e.g.* verifying the integrity of digital documents, or being used to derive authentication codes, etc. The security of a hash function is mainly analysed thanks to two notions: its collision resistance, and its preimage resistance. The first notion is linked to the hardness of finding two messages m and m^* such that for the function H being considered, we have $H(m) = H(m^*)$. The second notion consists in the hardness of finding a message m such that $H(m) = t$, for a given, predefined target t. For a function with n-bit outputs, the complexity of generic attacks for finding collisions or preimages is respectively $\sim 2^{n/2}$ and $\sim 2^n$ calls to H.

Research in better-than-generic collision attacks for hash functions has strongly been influenced from 2005 onwards by the results of Wang *& al.* on the cryptanalysis of functions of the MD4 family (cf. *e.g.* [WLF⁺05, WY05, WYY05]). The first attacks from Wang have been well studied since, and they were improved in several occasions. Collisions for older members of this family can now be computed extremely fast (with for instance a cost equivalent to two calls to the function for MD4 [SWOK07]).

A major progress on preimage attacks for the same MD4 family dates back to 2008, with Leurent's attack on MD4 [Leu08]. Many improvements in this domain were later made by Aoki & Sasaki, who progressively developed advanced meet-in-the-middle techniques to analyse various members of the family (cf. *e.g.* [AS09, SA09]). More recently, Knellwolf & Khovratovich introduced a new differential formalisation of the techniques from Aoki & Sasaki, which allowed them to sensibly improve the best attacks on SHA-1 [KK12]. All of these attacks still have a much higher complexity than collision attacks, and many functions like SHA-1 are still attacked only for reduced versions. As an example, the best attack on MD4 has complexity $2^{78.4}$ [GLRW10] (but if no expensive precomputation is allowed, it is another attack in 2^{95} which is the best [ZL12]).

Purpose of the internship

This internship consists in studying the framework of Knellwolf & Khovratovich and to apply it to MD4.

The compression functions h of SHA-1 and MD4 are both built around a block cipher in *Davies-Mayer* mode, *i.e.* $h(iv, m) = f(m, iv) + iv$. It is therefore possible to view the problem of finding a preimage t as the one of finding a key m for the cipher f such that $f(m, p) = c$, for a fixed pair (p, c) (with typically $p = iv$ and $c = t - iv$).

The framework of Knellwolf & Khovratovich consists in a differential view of the meet-in-the-middle attacks. First a decomposition of f in two sub-ciphers f_1 & f_2 is found, such that $f = f_2 \circ f_1$, then one searches for mutually independent related-key differential paths for f_1 and ${f_2}^{-1}$. More precisely, if we make use of the notation $(\alpha, \beta) \xrightarrow[p]{f} \gamma$ to mean that $\Pr_{(x,y)}\left[f(x \oplus \alpha, y \oplus \beta) = f(x, y) \oplus \gamma\right] = p$ (where $\oplus$ is the bitwise exclusive OR), what we are searching for are paths $(\delta_1, 0) \xrightarrow[p_1]{f_1} \Delta_1$ and $(\delta_2, 0) \xrightarrow[p_2]{{f_2}^{-1}} \Delta_2$ with p_1 and p_2 not too small, such that the set of all δ_1 is disjoint from the set of all δ_2. Once we are equipped with such paths, it is possible to speed up the preimage search by testing all candidate messages of the form $m \oplus \delta_1 \oplus \delta_2$, *i.e.* 2^{2d} messages (where d is the number of differences δ_1 & δ_2) at the cost of only 2^d calls to f_1 and f_2 (this is actually only true if p_1 and p_2 are both 1. We refer to [KK12] for a more detailed description). The main interest of this approach is that it makes the formalisation of advanced techniques relatively easy. Such techniques include for instance the use of truncated differential paths, of a *splice & cut* decomposition of f, and of small bicliques.

The goal of this internship, then, consists in applying this framework to MD4. The relative simplicity of this function should lead to efficient attacks (basic attacks will be studied for a start, with the integration of advanced techniques being an interesting sub-goal), which should be useful in evaluating the relevance of the framework for functions other than SHA-1. A major motivation is also to find out if the current best attacks on MD4 can be improved thanks to these techniques.

Skills

Ideally, a candidate should have a good knowledge of symmetric cryptanalysis (including in particular some acquaintance with differential and meet-in-the-middle attacks), and good programming skills in C.

References

[AS09] Kazumaro Aoki and Yu Sasaki. Meet-in-the-Middle Preimage Attacks Against Reduced SHA-0 and SHA-1. In Shai Halevi, editor, *CRYPTO*, volume 5677 of *Lecture Notes in Computer Science*, pages 70–89. Springer, 2009.

[Cra05] Ronald Cramer, editor. *Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings*, volume 3494 of *Lecture Notes in Computer Science*. Springer, 2005.

[GLRW10] Jian Guo, San Ling, Christian Rechberger, and Huaxiong Wang. Advanced Meet-in-the-Middle Preimage Attacks: First Results on Full Tiger, and Improved Results on MD4 and SHA-2. In Masayuki Abe, editor, *ASIACRYPT*, volume 6477 of *Lecture Notes in Computer Science*, pages 56–75. Springer, 2010.

[KK12] Simon Knellwolf and Dmitry Khovratovich. New Preimage Attacks against Reduced SHA-1. In Reihaneh Safavi-Naini and Ran Canetti, editors, *CRYPTO*, volume 7417 of *Lecture Notes in Computer Science*, pages 367–383. Springer, 2012.

[Leu08] Gaëtan Leurent. MD4 is Not One-Way. In Kaisa Nyberg, editor, *FSE*, volume 5086 of *Lecture Notes in Computer Science*, pages 412–428. Springer, 2008.

[SA09] Yu Sasaki and Kazumaro Aoki. Finding Preimages in Full MD5 Faster Than Exhaustive Search. In Antoine Joux, editor, *EUROCRYPT*, volume 5479 of *Lecture Notes in Computer Science*, pages 134–152. Springer, 2009.

[SWOK07] Yu Sasaki, Lei Wang, Kazuo Ohta, and Noboru Kunihiro. New Message Difference for MD4. In Alex Biryukov, editor, *FSE*, volume 4593 of *Lecture Notes in Computer Science*, pages 329–348. Springer, 2007.

[WLF$^+$05] Xiaoyun Wang, Xuejia Lai, Dengguo Feng, Hui Chen, and Xiuyuan Yu. Cryptanalysis of the Hash Functions MD4 and RIPEMD. In Cramer [Cra05], pages 1–18.

[WY05] Xiaoyun Wang and Hongbo Yu. How to Break MD5 and Other Hash Functions. In Cramer [Cra05], pages 19–35.

[WYY05] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding Collisions in the Full SHA-1. In Victor Shoup, editor, *CRYPTO*, volume 3621 of *Lecture Notes in Computer Science*, pages 17–36. Springer, 2005.

[ZL12] Jinmin Zhong and Xuejia Lai. Improved preimage attack on one-block MD4. *Journal of Systems and Software*, 85(4):981–994, 2012.