Generic collision attacks on narrow-pipe hash functions faster than birthday paradox, applicable to MDx, SHA-1, SHA-2, and SHA-3 narrow-pipe candidates

Vlastimil Klima[1] and Danilo Gligoroski[2]

[1] Independent Cryptologist - Consultant and KNZ, s.r.o., Prague, Czech Republic, e-mail: v.klima@volny.cz

[2] Faculty of Information Technology, Mathematics and Electrical Engineering, Institute of Telematics, Norwegian University of Science and Technology, Trondheim, Norway, e-mail: Danilo.Gligoroski@item.ntnu.no

Abstract. In this note we show a consequence of the recent observation that narrow-pipe hash designs manifest an abberation from ideal random functions for finding collisions for those functions with complexities much lower than the so called generic birthday paradox lower bound. The problem is generic for narrow-pipe designs including classic Merkle-Damgård designs but also recent narrow-pipe SHA-3 candidates. Our finding does not reduces the generic collision security of $n/2$ bits that narrow-pipe functions are declaring, but it clearly shows that narrow-pipe designs have a property when we count the calls to the hash function as a whole, the birthday paradox bound of $2^{n/2}$ calls to the hash function is clearly broken. This is yet another property in a series of similar "non-ideal random" properties (like HMAC or PRF constructions) that narrow-pipe hash function manifest and that are described in [1] and [2].

1 Introduction

Merkle-Damgård construction, introduced in 1989 ([3], [4]), is the most used method how to design hash functions. Even before the formal proposal of the Merkle-Damgård construction, there were known results in the thesis of Merkle from 1979 [5] that say that when an adversary is given 2^k distinct target hashes, (second) preimages can be found after hashing about 2^{n-k} messages, instead of expected 2^n different messages.

For the first generic attack against Merkle-Damgård construction we can treat the length-extension attack.

Then in 2004 we saw another generic attack described in the paper of Joux [6]. The Joux analysis showed that the attacker can find multi-collisions much more faster than expected: r messages with the same hash value can be found in $ln_2 r \times 2^{n/2}$ instead of $2^{n(r-1)/r}$ calls of hash function.

Soon after that, in 2005, Kelsey and Schneier extended ideas of Joux in [7] to find second preimages of messages long 2^k-message-blocks with complexity $k \times 2^{n/2+1} + 2^{n-k+1}$ which is below the generic bound of 2^n.

In this note we show another generic attack against Merkle-Damgård and narrow-pipe constructions when hashing long messages of 2^k blocks. Our attack reduces the collision search, from the generic bound of $2^{n/2}$ to $2^{n/2-k/2}$ number of hash calls, where hashing is done over messages of length 2^k blocks.

2 Notations

For the definition of narrow and wide-pipe hash function, let us denote:

- $C(h, m)$ - a compression function C with chaining variable h and message block variable m.
- *hlen* - the length of the chaining variable, i. e. the length of compression function output.
- *mlen* - the length of the message block.
- *hashlen* - the length of the hash function output.

If the compression function has the property, that for every value m the function $C(h, m) \equiv C_m(h)$ is an **i**deal **r**andom **f**unction of the variable h, we denote it as $IRF(h)$.

If the compression function has the property, that for every value h the function $C(h, m) \equiv C_h(m)$ is an **i**deal **r**andom **f**unction of the variable m, we denote it as $IRF(m)$.

The hash function is defined by a narrow-pipe compression function (NPCF), iff $hashlen = hlen = \frac{mlen}{2}$ and the compression function is $IRF(h)$ and $IRF(m)$.

The hash function is defined by a wide-pipe compression function (WPCF), iff $hashlen = \frac{hlen}{2} = \frac{mlen}{2}$ and the compression function is $IRF(h)$ and $IRF(m)$.

3 Main contribution of this note

Theorem 1. *Suppose that the hash function $H : \{0,1\}^* \to \{0,1\}^n$ is defined by a narrow-pipe compression function $C : \{0,1\}^n \times \{0,1\}^{mlen} \to \{0,1\}^n$. Then we can find a collision (M, M') for the hash function H using much less than $2^{n/2}$ calls to the hash function H (the lower bound of the birthday paradox).*

Proof. For the sake of simplicity, let us suppose $n = hashlen = 256$. The general case is analogous. In this case, the hashed message is padded and divided into 512-bit blocks. Let us suppose that a message M (for instance the content of a hard disk or a RAM memory) is divided into two parts A and B, i.e. $M = A||B$, where the part A consist of just one message block of 512 bits, and the number of 512-bit blocks in the part B is $N = 2^{35}$ (in case of current 2TByte HDD). Let us denote by hA the intermediate chaining value, obtained after hashing the part A of the message M and let us suppose that the content of the part B is never changing - so it consists of constant message blocks $const_1$, $const_2$, ..., $const_N$ (note that if padding is a part of the definition, it is also a constant block). We compute the final hash with the following iterative procedure:

$$\begin{gathered} h_1 = C(hA, const_1) \\ h_2 = C(h_1, const_2) \\ h_3 = C(h_2, const_3) \\ \dots \\ h_N = C(h_{N-1}, const_N) \\ H(M) = h_N \end{gathered}$$

If the compression function C is $IRF(h)$, then the chaining values are loosing the entropy in every of the N steps above. From Corollary 3[2] we obtain that the entropy of the final hash h_N is equal to

$$E(hash) = hashlen + 1 - log_2(N),$$

and for $N = 2^{35}$ it gives

$$E(hash) = 222.$$

If we compute hash values for 2^{111} different parts A (whereas the part B remains unchanged), we will obtain 2^{111} hash values h_N. According to the birthday paradox it is sufficient for finding a collision in the set of these values with probability around $\frac{1}{2}$. Cryptographically strong hash function H should require approximately 2^{128} hash computations. □

Corollary 1. *For hash functions $H()$ constructed as in Theorem 1, finding a pair of colliding messages (M, M') that are long $N = 2^k$ blocks, can be done with $O(2^{n/2-k/2})$ calls to the hash function $H()$.* □

Note 1: If we count the number of calls to the compression function $C(H_i, M_i)$, then with our collision strategy we are calling actually more

times the compression function. Namely, $2^{111} \times 2^{35} = 2^{145}$. So, our finding does not reduces the $\frac{n}{2}$ bits of collision security that narrow-pipe functions are declaring, but we clearly show that narrow-pipe designs have a property when we count the calls to the hash function as a whole, the birthday paradox bound of $2^{n/2}$ calls to the hash function is clearly lowered.

Note 2: This technique is not applicable to wide-pipe hash functions because the entropy reduction after applying the compression function $C(H_i, M_i)$ to different message blocks starts from the value $hlen$ which is two times bigger than $hashlen$ i.e. $hlen = 2hashlen$. So the final reduction from $hlen$ to $hashlen$ bits will make the technique described in this note ineffective against wide-pipe designs.

References

1. D. Gligoroski: "Narrow-pipe SHA-3 candidates differ significantly from ideal random functions defined over big domains", NIST hash-forum mailing list, 7 May 2010.
2. D. Gligoroski, V. Klima: "Practical consequences of the abberation of narrow-pipe hash designs form ideal random functions", IACR eprint archive Report 384/2010, `http://eprint.iacr.org/2010/384.pdf` (2010/08/08).
3. R. C. Merkle: "One Way Hash Functions and DES", Proceedings of CRYPTO '89, Santa Barbara, California, USA, Lecture Notes in Computer Science, Vol. 435, Springer, 1990, pp. 428 - 446.
4. I. Damgård: "A Design Principle for Hash Functions", Proceedings of CRYPTO '89, Santa Barbara, California, USA, Lecture Notes in Computer Science, Vol. 435, Springer, 1990, pp. 416 - 427.
5. R. C. Merkle - Secrecy, authentication, and public key systems, Ph.D. thesis, Stanford University, 1979, pp. 12 - 13, `http://www.merkle.com/papers/Thesis1979.pdf` (2010/08/08).
6. A. Joux: "Multicollisions in Iterated Hash Functions. Application to Cascaded Constructions", Proceedings of CRYPTO'04, Lecture Notes in Computer Science, Vol. 3152, Springer, 2004, pp. 306 - 316.
7. J. Kelsey, B. Schneier: "Second Preimages on n-Bit Hash Functions for Much Less than 2^n Work, Proceedings of EUROCRYPT'05, Lecture Notes in Computer Science, Vol. 3494, Springer, 2005, pp. 474 - 490.