Preliminary Analysis of the SHA-256 Message Expansion

Norbert Pramstaller, Christian Rechberger, Vincent Rijmen

Institute for Applied Information Processing and Communications (IAIK) - Krypto Group

Faculty of Computer Science
Graz University of Technology

The work described in this paper hasbeen supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT, and by the Austrian Science Fund (FWF) project P18138.

Outline

- Description of the SHA family of hash functions
- Previous work on members of the SHA-2 family
- Closer look at the Message Expansion - Results
- Conclusions

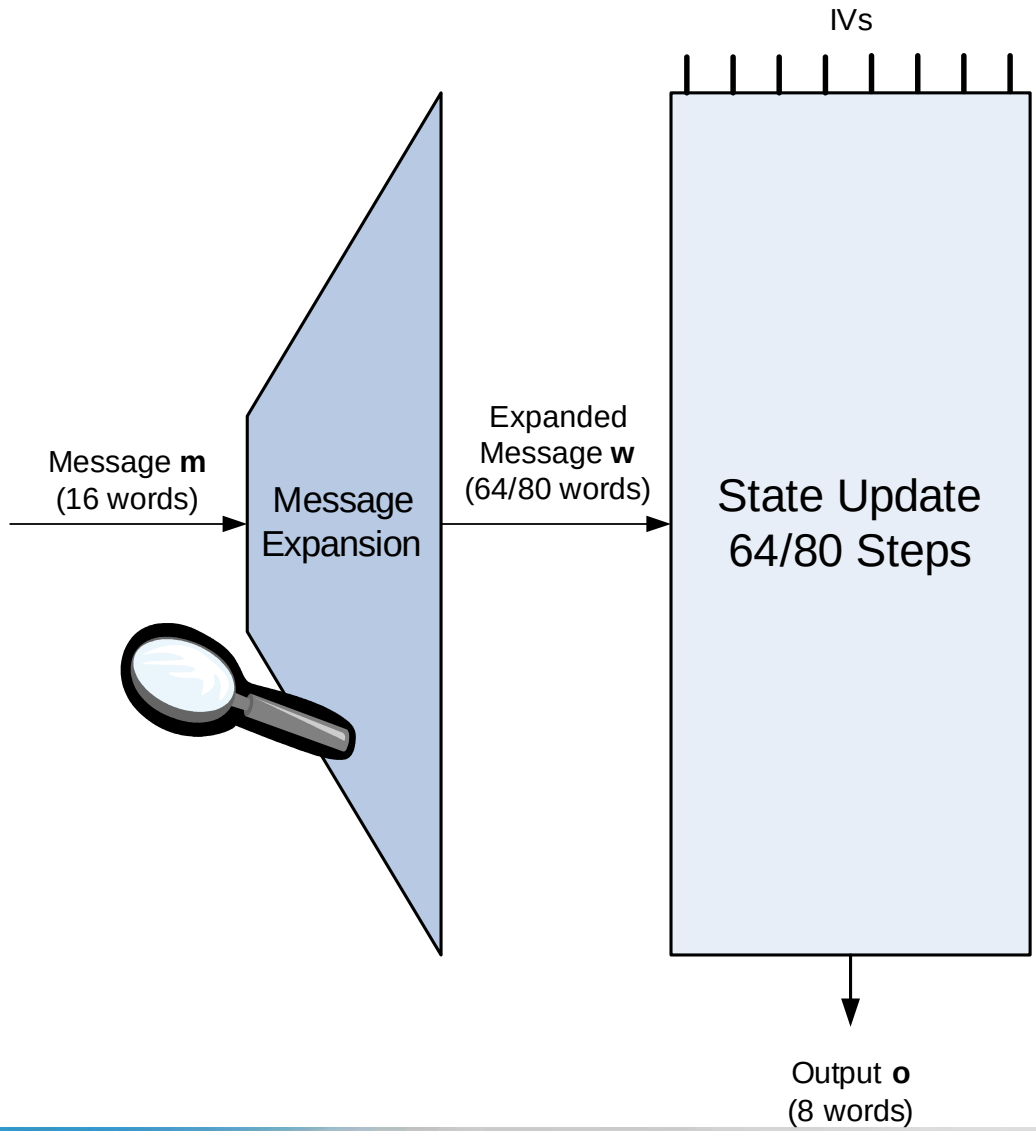
Outline

- **Description of the SHA family of hash functions**
- Previous work on members of the SHA-2 family
- Closer look at the Message Expansion - Results
- Conclusions

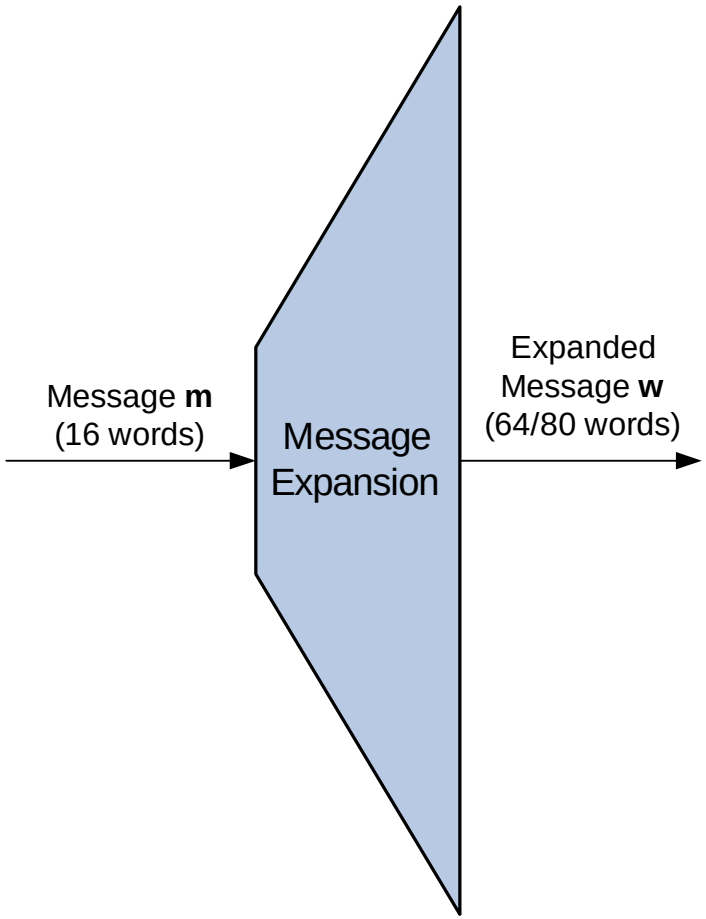
The SHA family of hash functions

		Steps	#Chaining Variables
SHA-0	1992	80	5 (160bit)
SHA-1	1994	80	5 (160bit)
SHA-256	2000	64	8 (256bit)
SHA-512	2000	80	8 (512bit)

Outline of SHA



Outline of SHA – Message Expansion



SHA-1

$$W_t = \begin{cases} M_t & for\,(0 \le t \le 15) \\ ROTL^1(W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}) & for\,(16 \le t \le 79) \end{cases}$$

SHA-256

$$W_t = \begin{cases} M_t & for\,(0 \le t \le 15) \\ \sigma_1(W_{t-2}) + W_{t-7} + \sigma_0(W_{t-15}) + W_{t-16} & for\,(16 \le t \le 63) \end{cases}$$

$$\sigma_0(x) = ROTR^7(x) \oplus ROTR^{18}(x) \oplus SHR^3(x)$$

$$\sigma_1(x) = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x)$$

Outline

- Description of the SHA family of hash functions
- **Previous work on members of the SHA-2 family**
- Closer look at the Message Expansion - Results
- Conclusions

Previous Work

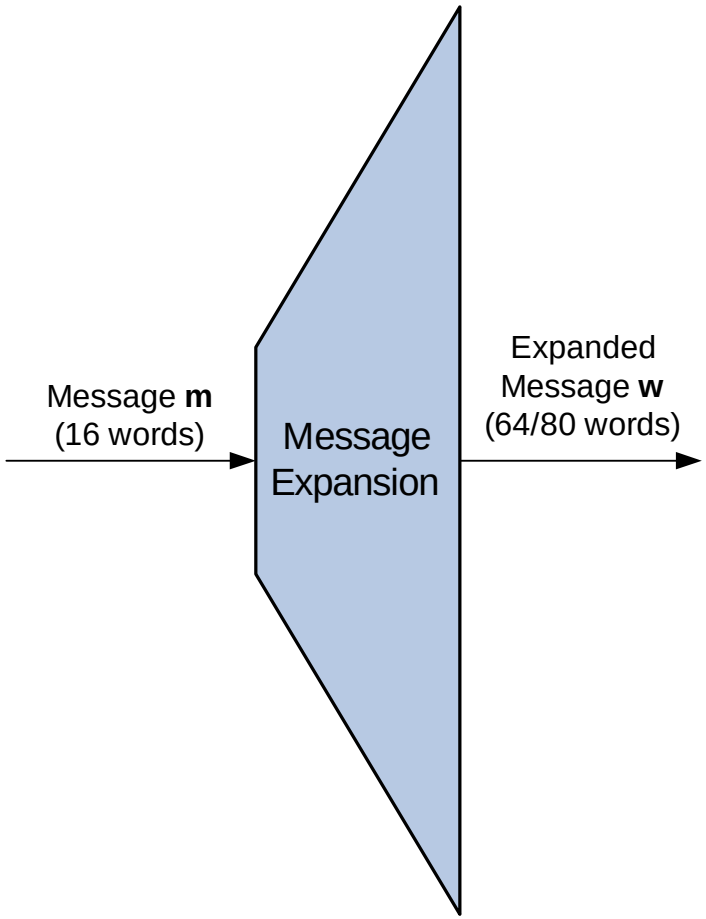
- Gilbert and Handschuh showed at SAC 2003 that SHA-256 and SHA-512 are resistant to an application of the Chabaud/Joux attack on SHA-0
  - Lower bound for probability of a local collision: 2^{-66}
- Hawkes, Paddon and Rose gave an improved probability for the 9-step local collision in the SHA-2 state update
  - Probability for a local collision: between 2^{-39} and 2^{-42}
- Yoshida and Biryukov show at SAC 2005 a pseudo-collision for a simplified variant of SHA-256 (up to 34 steps)

- **What is missing so far:** analysis of the Message Expansion of members of the SHA-2 family
  - Matusiewicz, Pieprzyk, Pramstaller, Rechberger and Rijmen: *"Analysis of simlified variants of SHA-256"*. In Proceedings of WEWoRC 2005, to appear.

Outline

- Description of the SHA family of hash functions
- Previous work on members of the SHA-2 family
- **Closer look at the Message Expansion - Results**
- Conclusions

Outline of SHA – Message Expansion



SHA-1

$$W_t = \begin{cases} M_t & for\,(0 \le t \le 15) \\ ROTL^1(W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}) & for\,(16 \le t \le 79) \end{cases}$$

SHA-256

$$W_t = \begin{cases} M_t & for\,(0 \le t \le 15) \\ \sigma_1(W_{t-2}) + W_{t-7} + \sigma_0(W_{t-15}) + W_{t-16} & for\,(16 \le t \le 63) \end{cases}$$

$$\sigma_0(x) = ROTR^7(x) \oplus ROTR^{18}(x) \oplus SHR^3(x)$$

$$\sigma_1(x) = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x)$$

Effect of a single bit-flip

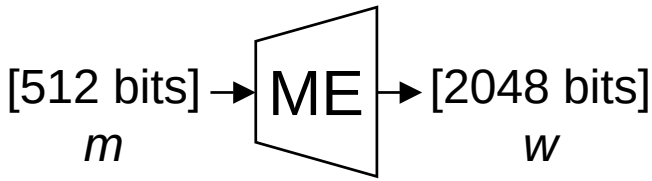
	orig. SHA-1	mod. SHA-256
min (40 steps)	~~18~~ 10	110 ?
max (40 steps)	30	297
min (full)	~~107~~ 44	467 ?
max (full)	174	694

Original SHA-1: $(W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}) <<< 1$ **linear, single rotation**

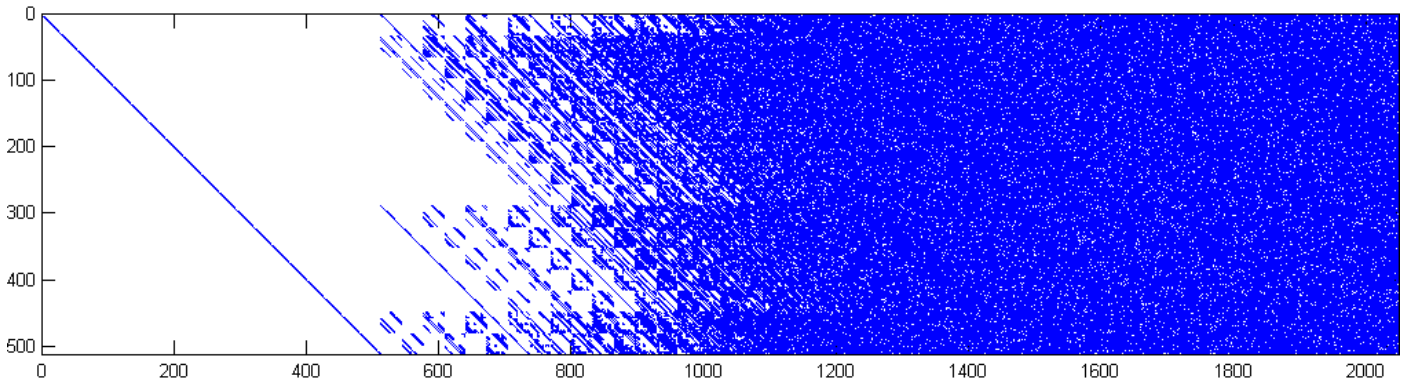
Modified SHA-256: $\sigma_1(W_{t-2}) \oplus W_{t-7} \oplus \sigma_0(W_{t-15}) \oplus W_{t-16}$ **linear, s-boxes**

Linearized SHA-256 Message Expansion

$$W_t = \begin{cases} M_t & for\,(0 \le t \le 15) \\ \sigma_1(W_{t-2}) \oplus W_{t-7} \oplus \sigma_0(W_{t-15}) \oplus W_{t-16} & for\,(16 \le t \le 63) \end{cases}$$

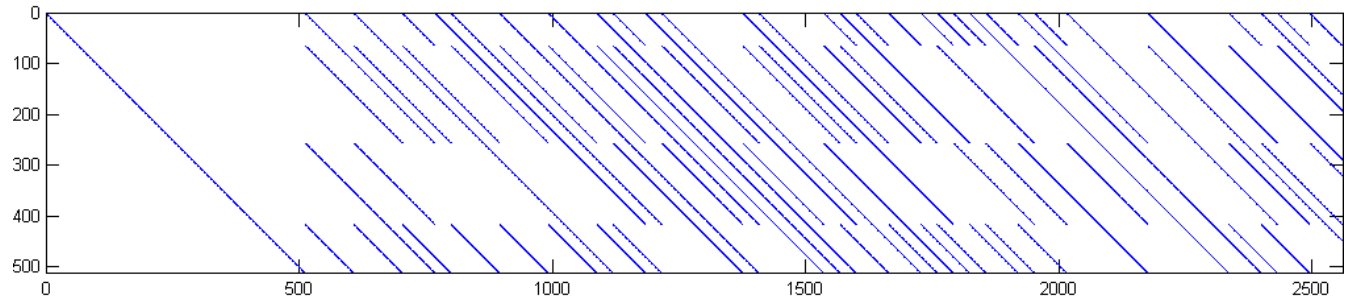


$$E_{512\times2048} = \left[I_{512\times512} F_{512\times1536} \right]$$

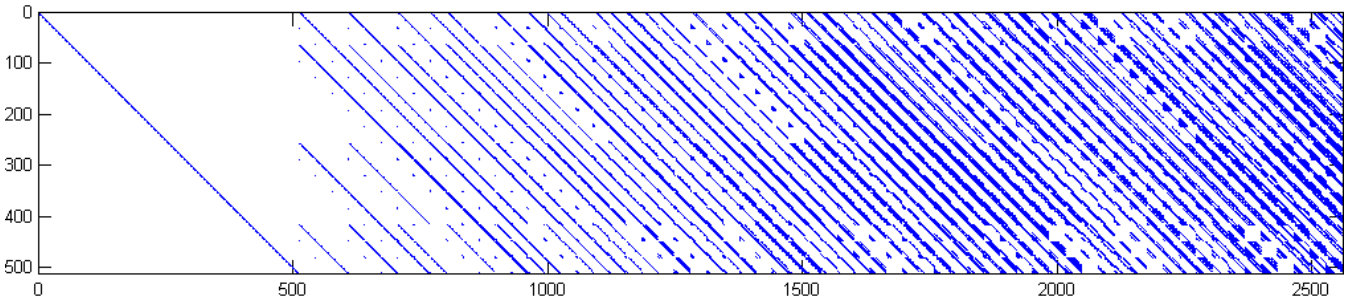


Comparing the message expansions of the SHA family

SHA-0

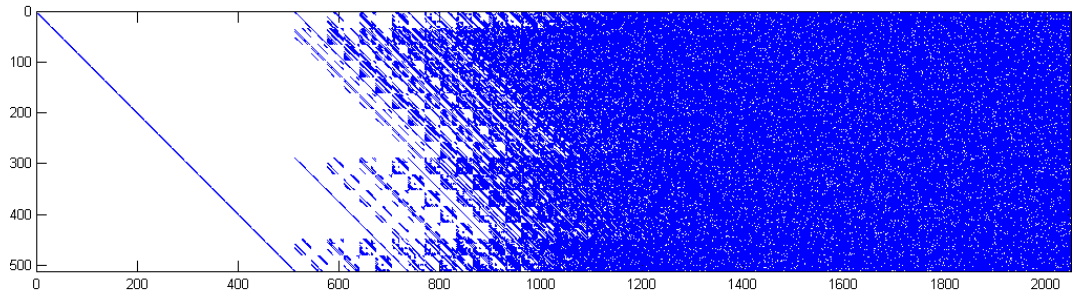


SHA-1

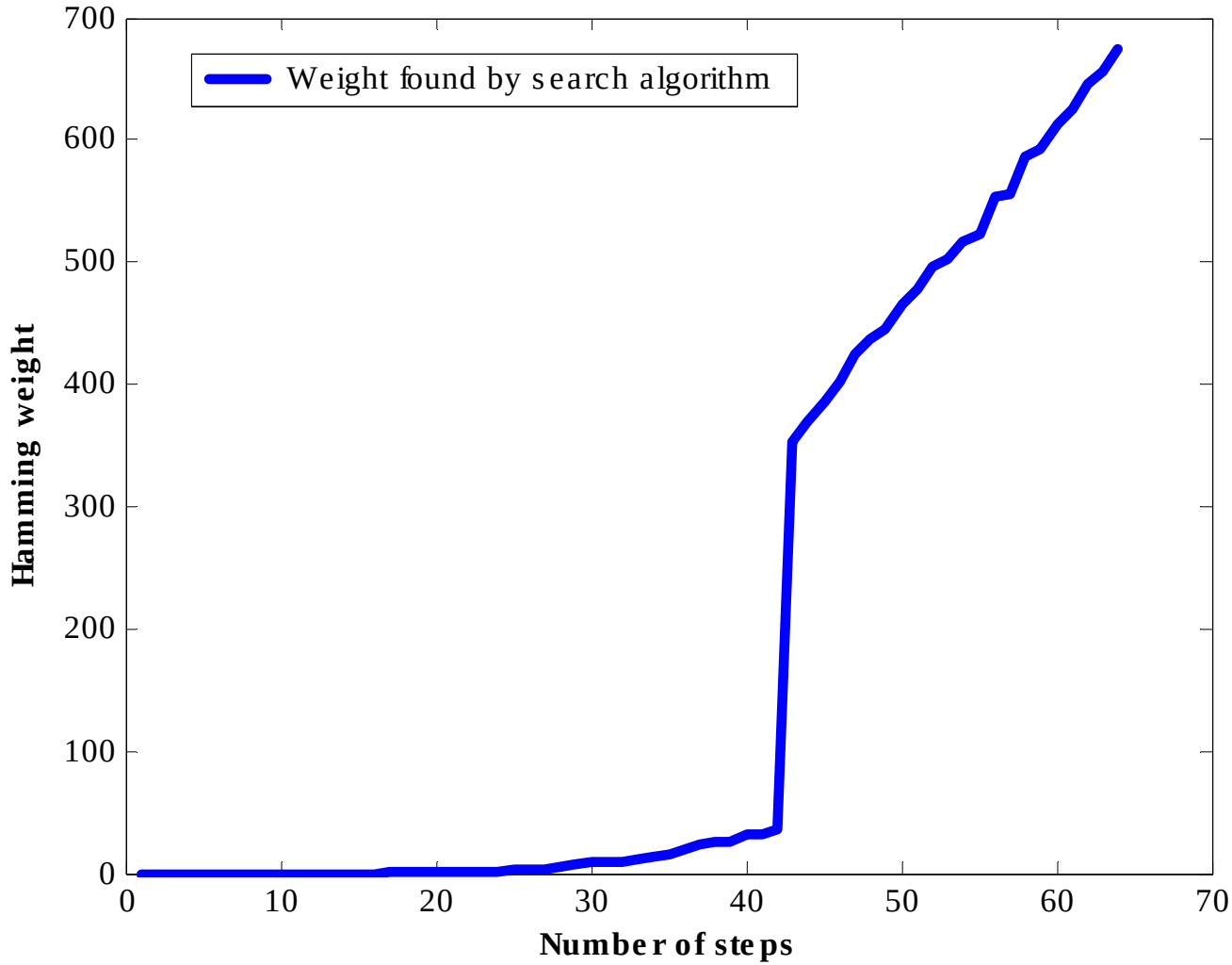


SHA-256

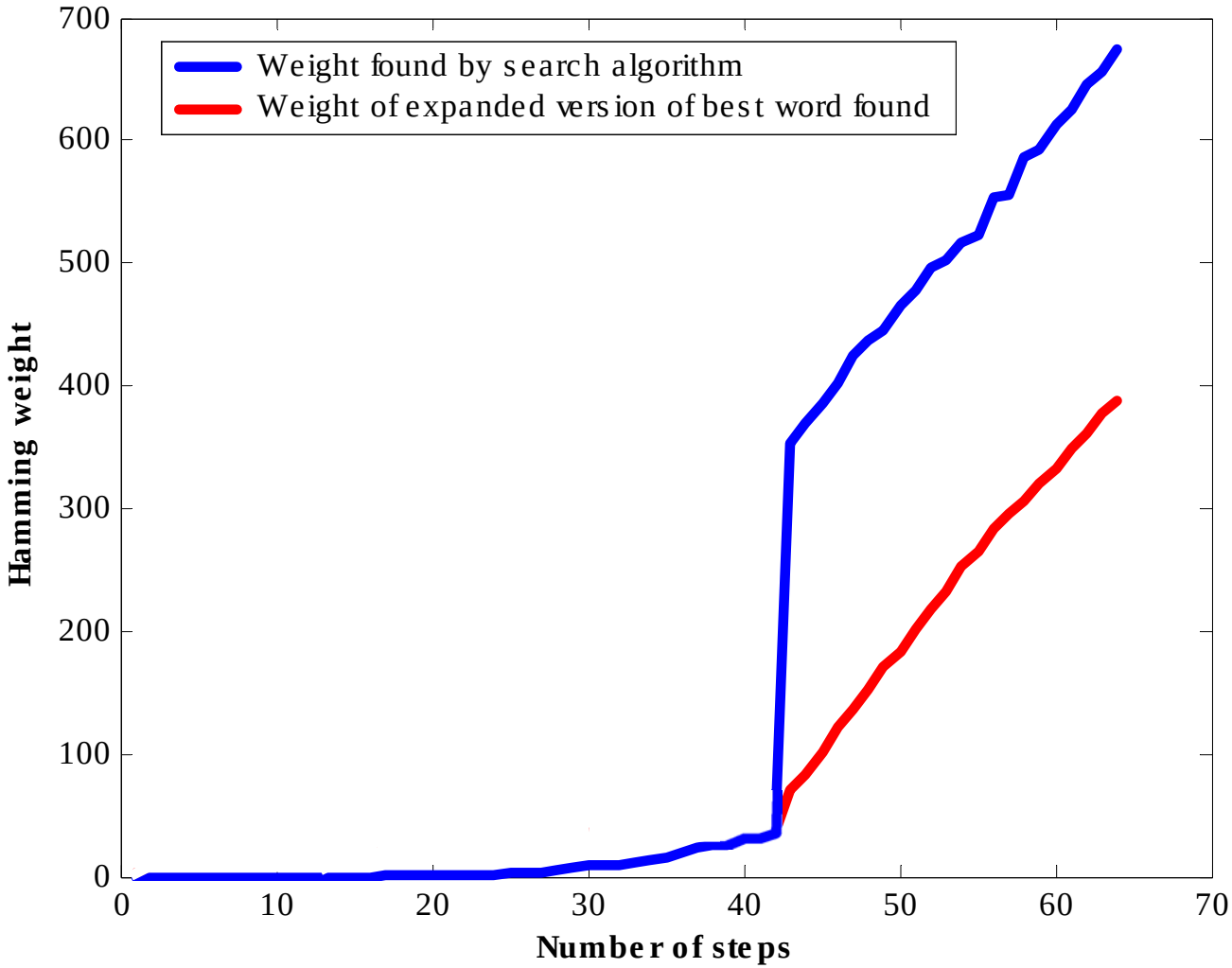
(linearized)



Results of low-weight search



Results of low-weight search



Effect of a single bit-flip

	orig. SHA-1	mod. SHA-256
min (40 steps)	~~18~~ 10	~~110~~ 32
max (40 steps)	30	297
min (full)	~~107~~ 44	~~467~~ ~~388~~ 297
max (full)	174	694

Original SHA-1: $(W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}) <<< 1$ **linear, single rotation**

Modified SHA-256: $\sigma_1(W_{t-2}) \oplus W_{t-7} \oplus \sigma_0(W_{t-15}) \oplus W_{t-16}$ **linear, s-boxes**

Example of a low-weight expanded message

00000001	00040088	00000000	00000000
00000000	00000001	00000000	00000000
00000000	15522028	00000000	00000000
00000000	000A0400	00000000	00000000
00000000	00000000	00000000	00000000
00000000	00000000	00004050	00000000
00000000	00000000	00000000	00000000
00000000	00040088	00000001	00000000
00000000	00000001	00000000	00000000
00000001	00000000	00000000	00000000

Outline

- Description of the SHA family of hash functions
- Previous work on members of the SHA-2 family
- Closer look at the Message Expansion - Results
- **Conclusions**

Conclusions

- Some steps towards understanding the security of SHA-256
- Security of complete SHA-256 is not threatened by our results
- Further analysis of message expansion is an important building block for evaluting the security of members of the SHA-2 family

Preliminary Analysis of the SHA-256 Message Expansion

Contact:

Norbert.Pramstaller@iaik.tugraz.at
Christian.Rechberger@iaik.tugraz.at
Vincent.Rijmen@iaik.tugraz.at

www.iaik.tugraz.at/research/krypto

Institute for Applied Information Processing and Communications (IAIK) - Krypto Group

Faculty of Computer Science
Graz University of Technology

The work described in this paper hasbeen supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT, and by the Austrian Science Fund (FWF) project P18138.

Preliminary Analysis of the SHA-256 Message Expansion

Norbert Pramstaller and Christian Rechberger and Vincent Rijmen
Institute for Applied Information Processing and Communications (IAIK)
Graz University of Technology, Inffeldgasse 16a, A–8010 Graz, Austria

***Abstract*— Recently, cryptanalytic results on popular hash functions such as MD5 and SHA-1 have been announced. In this article we analyze the message expansion of the SHA-2 family of hash functions. Upper bounds on the minimal weight of different variants of the message expansion of SHA-256 are given. Using these results, we expect to find collision-producing differences which in turn allow to find collisions of step-reduced variants of SHA-256 faster than by brute-force search.**

I. Introduction

After recent advances in the analysis of popular and widely used hash functions like MD5 [14], [16] and SHA-1 [1], [12], [17], [15], the analysis of the SHA-2 family of hash functions becomes an increasingly interesting topic. Throughout this article, we consider SHA-256. Having an output size of 256 bits, the expected effort to find a pair of colliding inputs is about 2^{128} hash function executions.

Previous results. Being standardized by NIST in 2000 [10], the first published independent analysis on members of the SHA-2 family was done by Gilbert and Handschuh [4]. They show that there exists a 9-step local collision with probability 2^{-66}. Later on, the result has been improved by Hawkes, Paddon and Rose [5]. By considering modular differences, they increased the probability to 2^{-39}. A variant of SHA-256 is analyzed by Yoshida and Biryukov [18]. The content of this note together with an additional analysis of other variants of SHA-256 can be found in [9].

Our contribution. The local collisions shown so far serve as an important tool to assess the security of members of the SHA-2 family against collision search attacks. However, a crucial thing is missing so far: in order to be able to produce a collision, an attacker has to produce as little local collisions as possible when applying a non-zero difference at the input of the hash function.

To the best of our knowledge there is to data no analysis of the message expansion of any of the members of the SHA-2 family. In this article we analyze the message expansion of SHA-256. We derive some non-trivial upper bounds for the weight of a difference between two expanded messages. Additionally, we consider simplified variants of the message expansion to allow for a comparison at the level of building blocks. We expect to find collision-producing differences without truncated collisions. In turn, these could be used to perform collision-search attacks faster than by brute-force search for step-reduced variants of SHA-256 having a linearized message expansion but the original state update. However, at this point no conclusions on the resistance of SHA-256 against recent attacks on SHA-1 can be given.

TABLE I
Notation

notation	description
$A \oplus B$	addition of A and B modulo 2 (XOR)
$A + B$	addition of A and B modulo 2^{32}
M_t	input message word t (32 bits), $t \geq 0$
W_t	expanded input message word t (32 bits), $t \geq 0$
$A \ggg n$	bit-rotation of A by n positions to the right
$A \gg n$	bit-shift of A by n positions to the right
N	number of steps of the compression function

Outline of the article. In Section II we give a short description of the SHA-256 message expansion. In Section III we derive some upper bounds for the minimum weight of the linearized message expansion. In Section IV we give some basic observations on the used building blocks. We conclude in Section V.

II. Description of the SHA-256 message expansion

In this section we shortly describe the message expansion of SHA-256. For a full description of SHA-256 or other members of the SHA-2 family, refer to [10]. In the remainder of this article we use the notation given in Table I.

The input message is split into 512-bit message blocks (after padding). A single message block is denoted by a row vector m. The message is also represented by 16 32-bit words, denoted by M_t, with $0 \leq t \leq 15$. In the message expansion, this input is expanded into 64 32-bit words W_t, also denoted as the 2048-bit expanded message row-vector w. The words W_t are defined as follows:

$$\begin{aligned} W_t &= M_t, \text{ for } 0 \leq t \leq 15 \\ W_t &= \sigma_1(W_{t-2}) + W_{t-7} + \sigma_0(W_{t-15}) + W_{t-16}, \\ &\quad \text{for } 16 \leq t \leq 63 \end{aligned}$$

where

$$\begin{aligned} \sigma_0(x) &= (x \ggg 7) \oplus (x \ggg 18) \oplus (x \gg 3) \\ \sigma_1(x) &= (x \ggg 17) \oplus (x \ggg 19) \oplus (x \gg 10) \end{aligned}$$

Note that members of the SHA-2 family are the first that use modular additions in their message expansions.

III. Finding low-weight codewords in the code describing the linearized SHA-256 message expansion

In the first attempt to get an idea about the effect of all the changes between the SHA-1 message expansion and the SHA-256 message expansion, we consider single bit differences. Table II illustrates this comparison. We consider variants reduced to 40 steps as well as full variants (80 steps for SHA-1 variants and 64 steps for SHA-256 variants). In variants using modular addition, we used the all-zero vector as a starting point.

By the modified SHA-1 message expansion we refer to a variant where every XOR is replaced by an addition modulo 2^{32}. By the modified SHA-256 message expansion, we refer to a variant where every addition is replaced by an XOR. We observe that both the introduction of modular additions and the replacement of a single bit-shift by a structure using σ_0 and σ_1 significantly increases the number of affected bits in the expanded message.

When talking about the SHA-1 message expansion, it was already observed in [8] and [12] that Hamming weights much smaller than 107 (as given in Table II) can be found. The minimum weight found for the message expansion of SHA-1 is 44. A more recent treatment of low-weight disturbance patterns in SHA-1 can be found in [6] and [11].

Due to the nonlinear behavior of the modular addition, no linear code can describe the SHA-256 message expansion. However, if the modular addition is replaced by a bitwise XOR, a linear code over $\mathbb{Z}_2$ can be constructed. If we consider SHA-256 with N steps, this code can be represented by a $512 \times 32N$ generator matrix G.

Due to the XOR-linearization, every possible difference of two expanded words is also a valid word in this code. Therefore, probabilistic algorithms from coding theory [2], [7], [13] can be used to find low-weight differences for the XOR-linearised SHA-256 message expansion. Some results of this codeword search are depicted in Fig. 1. All minimum weights found for variants of the message expansion up to the full 64 steps are shown in the figure. Until the 42-step variant, our algorithms found reasonable low weights. This is depicted by the solid line. Considering the 40-step variant, the weight of 26 is low compared to a minimal weight of 110 for single-bit differences given in Table II. The 40-step expanded message is given in Table III.

For variants with more than 42 steps, the running time of our algorithms is currently too high to return reasonable low weights. The sudden jump after step 42 is not an intrinsic property of the SHA-256 message expansion, but rather the result of the limited running time of our algorithms.

To show that there indeed are low-weight words for $N > 42$, we proceed as follows. After obtaining a low-weight word for 42 steps we use the expansion process to extend it to the full length word. Weights obtained in this way are depicted by the dashed line. A 42-step word of weight 35 is used there as a starting point. Expanding it to 64 steps gives us a weight of 356. This is considerable lower than 467, which is the minimal weight given for a single bit difference in Table II. However, there is room for improvements.

In contrast to the words found for the SHA-1 message expansion, there are no zero-bands [12] any more. Note that the given expanded message is not necessarily a valid difference in case of the real message expansion since we approximate the modular addition by the bitwise XOR operation. Also note that the given vector cannot directly be used as a collision-producing disturbance pattern as described by Chabaud and Joux in their original attack on SHA-0 [3]. The reason is that there are *truncated local collisions* [3] generated by non-zero words in the backward expansion. These local collisions start before step 0 and would cause additional difficulties for constructing a collision-producing differential characteristic. However, we expect to find input words for reduced variants of the message expansion that can be used to construct a collision-producing difference.

A number of conditions on chaining variables need to be satisfied in order to ensure that the concatenation of local collisions (which hold with a probability between 2^{-39} and 2^{-42}) results in a collision of the output of the compression function. If we do not assume any pre-fulfilled conditions, the maximal weight we allow for a perturbation pattern is 3 (since $2^{-39 \cdot 4} < 2^{-128}$). Considering the weights in Fig. 1, this would mean a maximum of 24 steps.

However, all recent collision search attacks use the fact that conditions on chaining variables in the first steps of the compression are easy to pre-fulfill. Therefore, even vectors

TABLE II

Comparison of the number of affected bits for a single bit difference in various message expansions.

	original SHA-1	modified SHA-1	modified SHA-256	original SHA-256
min (40 steps)	18	18	110	137
max (40 steps)	30	41	297	307
min (full)	107	247	467	507
max (full)	174	354	694	709

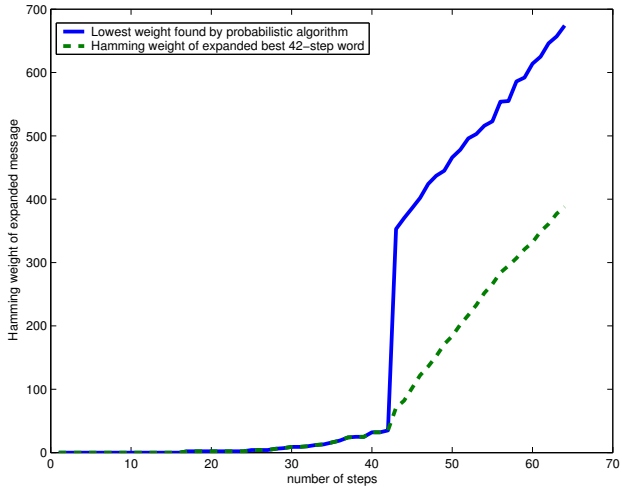


Fig. 1. Hamming weights of low-weight words found for step-reduced variants of the XOR-linearized SHA-256 message expansion.

TABLE III

LOW-WEIGHT EXPANDED MESSAGE FOR THE XOR-LINEARIZED 40-STEP MESSAGE EXPANSION OF SHA-256

00000001	00040088	00000000	00000000
00000000	00000001	00000000	00000000
00000000	15522028	00000000	00000000
00000000	000A0400	00000000	00000000
00000000	00000000	00000000	00000000
00000000	00000000	00004050	00000000
00000000	00000000	00000000	00000000
00000000	00040088	00000001	00000000
00000000	00000001	00000000	00000000
00000001	00000000	00000000	00000000

with considerable higher weights (if they do not have any truncated collisions) can still be used to mount collision-search attacks faster than 2^{128} elementary operations.

IV. OBSERVATIONS ON USED BUILDING BLOCKS

In this section, we list some observations we made on the SHA-256 message expansion.

- σ_0 and σ_1 have both the property to increase the Hamming weight of low-weight inputs. This increase is upper bounded by a factor of 3. The average increase of Hamming weight for low-weight inputs is even higher if three rotations are used instead of two rotations and one bit-shift. However, a reason for this bit-shift is given by the next observation.
- In contrast to all other members of the MD4-family including SHA-1, rotating expanded message words to get new expanded message words is not possible anymore (even in the XOR-linearized case). This is due to the bit-shift being used in σ_0 and σ_1.

V. CONCLUSIONS AND FUTURE WORK

We presented low-weight expanded message words for the step-reduced linearized SHA-256 message expansion. Our results naturally apply to the message expansions of SHA-224, since the message expansion is exactly the same there. In the case of SHA-384 and SHA-512, a slightly changed recurrence relation for the message expansion is used. Additionally, 80 instead of 64 steps are computed. Therefore, the results will be different, the basic observations are however expected to hold. We expect to find collision-producing differences without truncated collisions. For step-reduced variants of SHA-256 having a linearized message expansion but the original state update, they could be used to perform collision-search attacks faster than by brute-force search.

ACKNOWLEDGEMENTS

We would like to thank Florian Mendel for helpful discussions. The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT, and by the Austrian Science Fund (FWF) project P18138.

DISCLAIMER

REFERENCES

[1] Eli Biham, Rafi Chen, Antoine Joux, Patrick Carribault, Christophe Lemuet, and William Jalby. Collisions of SHA-0 and reduced SHA-1. In Ronald Cramer, editor, *Advances in Cryptology – EUROCRYPT'05*, volume 3494, pages 36–57. Springer-Verlag, 2005.

[2] Anne Canteaut and Florent Chabaud. A new algorithm for finding minimum-weight words in a linear code: application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511. *IEEE Transactions on Information Theory*, 44(1):367–378, 1998.

[3] Florent Chabaud and Antoine Joux. Differential collisions in SHA-0. In Hugo Krawczyk, editor, *Advances in Cryptology - CRYPTO'98*, volume 1462 of *LNCS*, pages 56–71. Springer-Verlag, 1998.

[4] Henri Gilbert and Helena Handschuh. Security analysis of SHA-256 and sisters. In Mitsuru Matsui and Robert Zuccherato, editors, *Selected Areas in Cryptography'03*, volume 3006 of *Lecture Notes in Computer Science*, pages 175–193. Springer-Verlag, 2003.

[5] Philip Hawkes, Michael Paddon, and Gregory G. Rose. On corrective patterns for the SHA-2 family. Cryptology ePrint Archive, Report 2004/207, August 2004. http://eprint.iacr.org/.

[6] Charanjit S. Jutla and Anindya C. Patthak. A Matching Lower Bound on the Minimum Weight of SHA-1 Expansion Code. Cryptology ePrint Archive, Report 2005/266, 2005. http://eprint.iacr.org/.

[7] Jeffrey S. Leon. A probabilistic algorithm for computing minimum weights of large error-correcting codes. *IEEE Transactions on Information Theory*, 34(5):1354–1359, 1988.

[8] Krystian Matusiewicz and Josef Pieprzyk. Finding good differential patterns for attacks on SHA-1. In *Proc. International Workshop on Coding and Cryptography, WCC'2005*, LNCS, 2005. to appear.

[9] Krystian Matusiewicz, Josef Pieprzyk, Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Analysis of simplified variants of SHA-256. In *Proceedings of WEWoRC 2005*, LNI, 2005. To appear.

[10] National Institute of Standards and Technology. Secure hash standard (SHS). FIPS 180-2, August 2002.

[11] Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Exploiting Coding Theory for Collision Attacks on SHA-1. In *Cryptography and Coding, 10th IMA International Conference, Cirencester, UK, December 19-21, 2005, Proceedings to appear*, LNCS. Springer, 2005.

[12] Vincent Rijmen and Elisabeth Oswald. Update on SHA-1. In Alfred Menezes, editor, *Topics in Cryptology – CT-RSA 2005*, volume 3376 of *LNCS*, pages 58–71. Springer-Verlag, Feb 2005.

[13] Jacques Stern. A method for finding codewords of small weight. In G. Cohen and J. Wolfmann, editors, *Coding Theory and Applications, 3rd International Colloquium*, volume 388 of *Lecture Notes in Computer Science*, pages 106–113. Springer-Verlag, 1989.

[14] Xiaoyun Wang, Xuejia Lai, Dengguo Feng, Hui Chen, and Xiuyuan Yu. Cryptanalysis of the hash functions MD4 and RIPEMD. In Ronald Cramer, editor, *Advances in Cryptology – EUROCRYPT'05*, volume 3494 of *LNCS*, pages 1–18. Springer-Verlag, 2005.

[15] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding Collisions in the Full SHA-1. In Victor Shoup, editor, *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *LNCS*, pages 17–36. Springer, 2005.

[16] Xiaoyun Wang and Hongbo Yu. How to break MD5 and other hash functions. In Ronald Cramer, editor, *Advances in Cryptology – EUROCRYPT'05*, volume 3494 of *LNCS*, pages 19–35. Springer-Verlag, 2005.

[17] Xiaoyun Wang, Hongbo Yu, and Yiqun Lisa Yin. Efficient Collision Search Attacks on SHA-0. In Victor Shoup, editor, *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *LNCS*, pages 1–16. Springer, 2005.

[18] Hirotaka Yoshida and Alex Biryukov. Analysis of a SHA-256 variant. In Bart Preneel and Stafford Tavares, editors, *Selected Areas in Cryptography (SAC 2005), Kingston, Ontario, Canada, August 11-12, 2005, Proceedings to appear*, LNCS. Springer, 2005.