SCIENCE CHINA
Information Sciences

• LETTER •

December 2017, Vol. 60 129101:1–129101:3
doi: 10.1007/s11432-016-9241-x

New strategy for searching disturbance vector of SHA-1 collision attack

Hao FENG[1*], Guang ZENG[1,2*], Wenbao HAN[1*] & Yang YANG[3,4*]

[1]*State Key Laboratory of Mathematical Engineering and Advanced Computing, Wuxi* 214125*, China;*
[2]*Centrum Wiskunde & Informatica, Amsterdam* 1098 XG*, The Netherlands;*
[3]*Zhengzhou Information Science and Technology Institute, Zhengzhou* 450001*, China;*
[4]*Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China*

Received May 25, 2017; accepted August 16, 2017; published online November 8, 2017

Citation Feng H, Zeng G, Han W B, et al. New strategy for searching disturbance vector of SHA-1 collision attack. Sci China Inf Sci, 2017, 60(12): 129101, doi: 10.1007/s11432-016-9241-x

Dear editor,

We present a new algorithm to search for effective disturbance vectors with a complexity of 2^{38} based on the following two properties of disturbance vectors. One property is the weight correlation between the first 16-step disturbance vectors and the 60-step disturbance vectors. The other property requires all the differences of the active bit positions of the first 16-step disturbance vectors to be less than 3 if the weight of the 60-step disturbance vectors is less than 35. This algorithm is not only able to retrieve all known disturbance vectors, but can also identify all 60-step disturbance vectors of which the weight is less than 35. Significantly, some of the identified disturbance vectors belong to neither Type-I nor Type-II [1], i.e., the categories of known disturbance vectors.

The attack complexity of the modular differential attack [2, 3] mainly depends on whether the properties of the disturbance vector and the differential path are effective. Thus, searching for and identifying an effective disturbance vector is a precondition to constructing a good differential path. At present, many heuristic cost functions [4–6] have been proposed to evaluate the disturbance vector, where the weight of the last 60-step words of the disturbance vector is viewed as the basic evaluation criterion. However, the identification of all the lower weight 60-step disturbance vectors in the whole search space with a size of 2^{512} is difficult.

A new algorithm, which is capable of identifying all optimal disturbance vectors, is presented. The main difficulty is to determine how to diminish the searching space. This problem was overcome by designing an algorithm consisting of four steps to reduce the searching space gradually. The first and second steps provide two necessary conditions to search the disturbance. The third step serves to exclude any impossible weight distribution of the initial messages. The last step is to determine the active bit distribution. In this way, all 60-step disturbance vectors with a weight less than 35 can be identified. The weight of the last 60-step words is an important factor to evaluate the disturbance vector. Thus, the optimum cannot possibly exist in the residual search space.

Definition 1 (Initial disturbance vector)**.** The first 16 words $w_0, w_1, \ldots, w_{15}$ of the disturbance vectors are termed the initial messages.

Definition 2 (The weight distribution)**.** The weight of each of the initial messages, $h(w_0), h(w_1), \ldots, h(w_{15})$, is termed the weight dis-

* Corresponding author (email: fh070301@163.com, seanzg@yeah.net, wb.han@netease.com, yangyang-wawa@sina.com)

The authors declare that they have no conflict of interest.

tribution of the initial messages.

Definition 3 (The active bit and the active bit position)**.** The non-zero bit of the disturbance vector is termed the active bit, and the position of the active bit is termed the active bit position.

Definition 4 (The active bit distribution)**.** If all active bits of the initial messages are determined, then these determined active bits are termed the active bit distribution of the initial messages.

Definition 5 (Tracking equation)**.** The relation expression between w_k^b and the initial messages are determined by message expansion, and the expression set is termed the tracking equation set. Each item in the set is termed a tracking equation, where w_k^b denotes the b-th bit of w_k, and $16 \leqslant k \leqslant 59$, $0 \leqslant b \leqslant 31$.

First step. This step is intended to prove by way of computation that $h(w_i) \leqslant 3$ is a necessary condition to search for disturbance vectors with a weight less than 35. Hence, the maximum weight of the initial messages satisfies $h(w_i) \leqslant 3$, where $0 \leqslant i \leqslant 15$.

Each 60-step disturbance vector is viewed as a matrix that is denoted by W, consists of 60 rows, and has a length of 32. Let C_l denote the l^{th} column of W and let $C_l[t]_{t=0}^{t=59}$ denote the value of C_l.

Lemma 1. If matrix W is a non-zero matrix $W \neq 0$ and the weight of W is less than 35, then it has at least one zero column.

Based on Lemma 1, Theorem 1 can be attained.

Theorem 1. For each initial message, if $h(w_i) > 3$, then the weight of the 60-step disturbance vector must be more than 34.

Second step. The purpose of this step is to determine the upper bound of the weight of the initial messages in three ways: parity calculation, using a formula to calculate the weight, and by applying Theorem 2.

First, the parity calculation, which is used to demonstrate how to estimate the weight of a 60-step disturbance vector by employing a tracking equation, is discussed. Let n_i^k denote the frequency of w_i in the tracking equation of w_k. Apparently, the prerequisite of $h(w_k) = 0$ is $\sum_{i=0}^{15} h(w_i) n_i^k = 0$. Hence, if $\sum_{i=0}^{15} h(w_i) n_i^k$ is even, then let $h(w_k) = 0$; otherwise, $h(w_k) = 1$. Therefore, the estimated weight is $\sum_{k=16}^{59} \sum_{i=0}^{15} h(w_i) n_i^k$. However, the use of the parity calculation to estimate the weight of the disturbance vectors is approximate. More precisely, if only one word has the active bit in the tracking equation of w_k, we calculate the real minimum weight as this situation, which is easy to analyze.

Second, the formula used to calculate the weight of the disturbance vector is discussed. Since the relations between the active bit positions and the bit positions are neglected, it is difficult to calculate the real minimum weight of the disturbance vector by parity calculation. Thus, a formula for calculating the weight of the disturbance vectors is presented to further estimate the weight of these vectors.

Property 1. In the tracking equation of w_k, if one of the differences of the active bit position is equal to that of one bit, i.e., $x_{i_p,j_q} = d^k_{i_a,j_b}$, then the p^{th} active bit of the a^{th} word w_i can eliminate the q^{th} active bit of the b^{th} word w_j, where $0 \leqslant i \leqslant 14$ and $i < j \leqslant 15$, and x_{i_p,j_q} denotes the difference of the active bit position between the p^{th} active bit position of w_i and the q^{th} active bit position of w_j, $d^k_{i_a,j_b}$ denotes the difference of the bit position between the bit position of the a^{th} word w_i and the bit position of the b^{th} word w_j in the tracking equation of w_k.

According to Property 1, the exponential function $\mathrm{e}^{-c\left(x_{i_p j_q} - d^k_{i_a j_b}\right)^2}$ can denote whether the p^{th} active bit of the a^{th} word w_i eliminates the q^{th} active bit of the b^{th} word w_j in the tracking equation of w_k. Let $y_{i_p,j_q} = \delta(\sum_{d^k_{i_a,j_b}} \mathrm{e}^{-c\left(x_{i_p,j_q} - d^k_{i_a,j_b}\right)})$ denote whether x_{i_p,j_q} can be eliminated in the tracking equation of w_k. Thus, the eliminated number in the tracking equation of w_k is as follows:

$$\begin{aligned} n_k &= \sum_{i_p,j_q} u_{i_p} u_{j_q} y_{i_p,j_q} \qquad (1) \\ &= \sum_{i_p} \sum_{j_q} u_{i_p} u_{j_q} \delta \left(\sum_{d^k_{i_a,j_b}} \mathrm{e}^{-c\left(x_{i_p,j_q} - d^k_{i_a,j_b}\right)} \right), \end{aligned}$$

where $u_{i_p} u_{j_q} = (1 - \delta(\sum_{s=0}^{i} \sum_{t=1}^{p-1} u_{s_t} u_{i_p} y_{s_t,i_p})) \cdot (1 - \delta(\sum_{s=0}^{i} \sum_{t=1}^{p-1} u_{s_t} u_{j_q} y_{s_t,j_q})) \cdot (1 - \delta(\sum_{s=i}^{j} \sum_{t=1}^{q-1} u_{i_p} u_{s_t} y_{i_p,s_t}))$, $\delta(x) = \{ {}^{1,\, x > 0}_{0,\, x = 0}$.

Eq. (1) provides a method to directly calculate the weight of w_k by the differences in the active bit positions, and this is the formula used to calculate the weight of the disturbance vector. Many impossible weight distributions can be excluded by (1). For $16 \leqslant k_1 \leqslant 31$, there are fewer initial disturbance vectors in the tracking equation of w_{k_1}. Hence, by (1), it is easy to verify whether $h(w_{k_1}) = 0$ by the parity calculation. If this result holds, then according to Theorem 3 below, the active bit differences must be less than 3. Thus, it is not difficult to retrieve all possible active bit distributions satisfying $h(w_{k_1}) = 0$ by exhausting 3 bits. Further, those distributions can be used to verify whether $h(w_{k_2}) = 0$ calculated by the parity calculation with $32 \leqslant k_2 \leqslant 59$. Hence, the possi-

ble weight distribution of the initial disturbance vector attained by the parity calculation is a precise estimation of those initial disturbance vectors. This is helpful for excluding additional impossible weight distributions of the initial disturbance vectors.

The parity calculation and the formula calculating the weight of the disturbance vector can be used to prove Theorem 2 as follows.

Theorem 2. The minimum weight of the 44-step disturbance vector is 12.

According to Theorem 2, it is not difficult to observe that

$$h_{60} = \sum_{i=0}^{59} h(w_i) = \sum_{i=0}^{15} h(w_i) + \sum_{i=16}^{59} h(w_i).$$

The upper bound of $\sum_{i=0}^{15} h(w_i)$ can be obtained by calculating the minimum weight of the 44-step disturbance vectors. Actually, Theorem 2 was used to calculate all 44-step disturbance vectors with a weight less than 15; hence, it limits the weight of the initial messages to less than 20 when we search for the 60-step disturbance vector with a weight less than 35. Thus, $\sum_{i=0}^{15} h(w_i) \leqslant 20$ is the second necessary condition to search for disturbance vectors with a weight less than 35.

Third step. This step is intended to exclude almost all of the impossible weight distributions of the disturbance vectors by using the adopted methods of parity calculation and formula computation of the weight of the disturbance vector.

At first, by using the parity calculation, a large number of impossible initial messages can be excluded. In terms of the remaining messages, the formula calculating the weight of the disturbance vector can further exclude initial messages that are impossible. Finally, most of the impossible initial messages are removed by way of selection.

Fourth step. This step is used to determine the active bit distribution of initial messages based on Theorem 3.

Each disturbance vector is also viewed as a matrix M. In addition, let $|x_{\max,\min}|$ denote the maximum active bit position difference, and let the columns in which the active bit $x_{\max}$ and $x_{\min}$ are located be denoted as $C_{\max}$, $C_{\min}$, where $x_{\max}$ and $x_{\min}$ represent the maximum and minimum active bit positions of the initial messages. Before proving Theorem 3, the following three important lemmas are presented.

Lemma 2. If the weight of 60-step disturbance vectors is less than 35, then for $16 \leqslant j \leqslant 59$, $C_{\min}[j] = 0$.

Lemma 3. Let $C_{l_1}, C_{l_2}, C_{l_3}$ be three consecutive columns to the left of $C_{\min}$, then $h(C_{\min})+h(C_{l_1})+h(C_{l_2})+h(C_{l_3}) > 10$.

Lemma 4. The active bits of $C_{\max}$ can impossibly diffuse to the column $C_{\min}$.

Based on the above lemmas, Theorem 3 can be proved.

Theorem 3. If $|x_{\max,\text{low}}| \geqslant 3$, then the weight of the 60-step disturbance vector is more than 34.

Finally, based on Theorem 3, all active bit position differences of the initial messages are limited to $|d_{i_p,j_q}| < 3$. Hence, it is sufficient to only consider three consecutive bits of the initial messages.

Conclusion. Based on Theorems 1 and 3, the initial disturbance vectors could be limited to $h(w_i) \leqslant 3$ and $\sum_{i=0}^{15} h(w_i) \leqslant 20$, respectively. Then parity calculation and a formula calculating the weight of the disturbance vectors were adopted to exclude the impossible weight distribution of the initial messages. Based on Theorem 3, all active bit position differences of the initial messages were limited to $|d_{i_p,j_q}| < 3$. Hence, it suffices to only consider three consecutive bits of the initial messages to identify all the 60-step disturbance vectors of which the weight is less than 35.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61502532) and Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing (Grant No. 2016A02).

References

1 Manuel S. Classification and generation of disturbance vectors for collision attacks against SHA-1. Des Codes Cryptogr, 2011, 59: 247–263
2 Wang X Y, Yin Y L, Yu H B. Finding collisions in the full SHA-1. In: Proceedings of the 25th Annual International Conference on Advances in Cryptology, Santa Barbara, 2005. 17–36
3 Stevens M, Karpman P, Peyrin T. Freestart collision for full SHA-1. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2016
4 Stevens M. New collision attacks on SHA-1 based on optimal joint local-collision analysis. In: Advances in Cryptology–EUROCRYPT. Berlin: Springer, 2013. 245–261
5 Jutla C S, Patthak A C. A matching lower bound on the minimum weight of SHA-1 expansion code. Iacr Cryptol Eprint Arch, 2005
6 Tang Y C, Zeng G, Han W B. Classification of disturbance vectors for collision attack in SHA-1. Sci China Inf Sci, 2015, 58: 112102