Linear Analysis of Reduced-Round CubeHash

Tomer Ashur Orr Dunkelman

Faculty of Mathematics and Computer Science
Weizmann Institute of Science
tomerashur@gmail.com

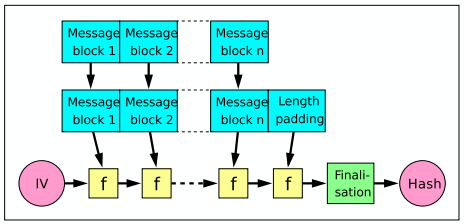
10/6/11

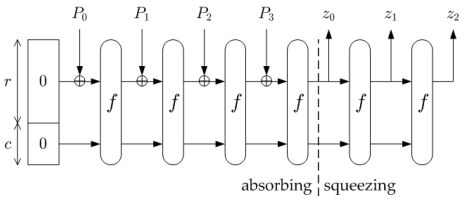
NIST SHA-3 Competition

- In 2007 NIST has announced a public competition for adding a new hash-function to the SHA family.
- 64 submissions, 51 candidates, 14 in the second round.
- Amongst which was CubeHash by Daniel Bernstein.
- CubeHash did not advance to the third round.

Structure of CubeHash

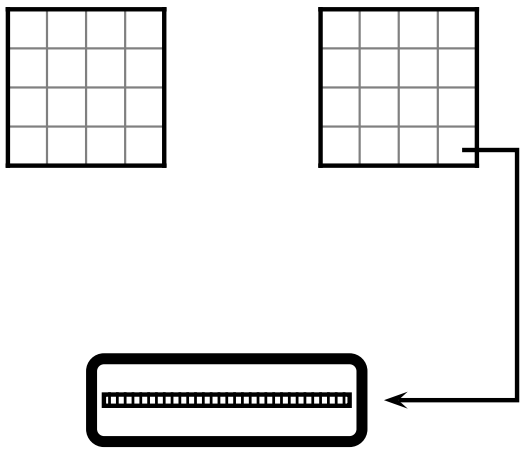
- CubeHash has a unique structure combining both Merkle–Damgård and sponge structures.
- CubeHash has three parameters:
  - h - The digest size.
  - r - The number of times the round function, T, is operated over each message block.
  - b - The size in bytes of each message block.





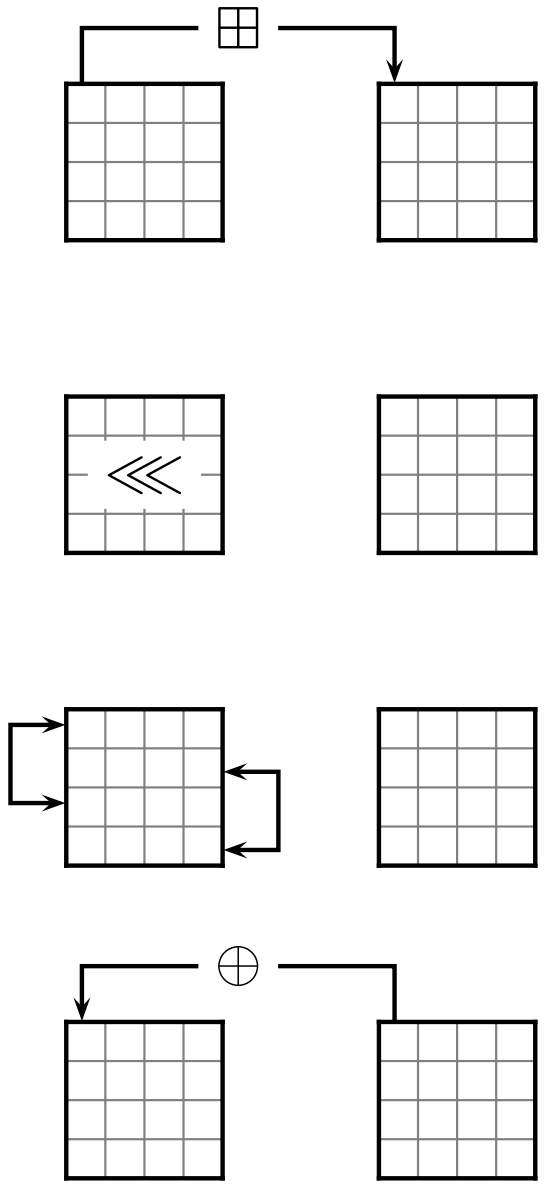
Structure of CubeHash

- The state is of size 1024-bit (treated as 32 32-bit words)
- To initialize CubeHash, h, r and b are loaded into the state.
- In each iteration, the b-byte block is XORed into the state;
- Then, the state is updated by iterating T a total number of $10 \cdot r$ times (which is a lot!).
- Finalization is done by XORing 1 into the state and applying T another $10 \cdot r$ times (which is, again, a lot).



The Round Function

- The round function (denoted as T) is built from 10 steps involving 4 operations (addition mod 2^{32}, XOR, swapping of words and rotation).
- then, the state is updated by applying T^r to the state.
- The full description of the function can be found in the specification and in the paper.



Previous Results on CubeHash

- Preimages can be found in about 2^{512-4b} CubeHash computations. [KNW08, BK09]
- Collisions were found for CubeHash2/120-150 [JP08] and CubeHash4/48 and CubeHash4/64 [Dai08].
- Symmetric properties were found by [A+09, FLM10].

Previous Results on CubeHash

- Preimages can be found in about 2^{512-4b} CubeHash computations. [KNW08, BK09]
  - Preimage attack in a hash function context is the ability to find a matching input for some specific output
- Collisions were found for CubeHash2/120-150 [JP08] and CubeHash4/48 and CubeHash4/64 [Dai08].
  - Collision attack in a hash function context is the ability to find two words that map to the same value.
- Symmetric properties were found by [A+09, FLM10].
  - Symmetric properties are structures that if present in the input, are maintained in the output as well.

Linear Cryptanalysis

- Linear cryptanalysis is a useful cryptanalytic tool to attack block ciphers.
- A linear cryptanalysis attack has two parts: finding a linear approximation and using a linear approximation to attack the cryptosystem.

Finding a Good Linear Approximation

- The adversary tries to approximate the nonlinear operations with other (linear) operations.
- The result is an expression of the form $P_i \oplus \ldots \oplus P_j \oplus C_k \oplus \ldots \oplus C_l = K_m \oplus \ldots \oplus K_n$.
  - The P's are bits from the plaintext, the C's are bits from the ciphertext and the K's are bits from the key.
- Each approximation has a probability p assosicated with it.
- This is usually the less interesting part of the attack.

Finding a Good Linear Approximation

- The adversary tries to approximate the nonlinear operations with other (linear) operations.
- The result is an expression of the form $P_i \oplus \ldots \oplus P_j \oplus C_k \oplus \ldots \oplus C_l = K_m \oplus \ldots \oplus K_n$.
  - The P's are bits from the plaintext, the C's are bits from the ciphertext and the K's are bits from the key.
- Each approximation has a probability p assosicated with it.
- This is usually the less interesting part of the attack.
  - This is the part that we studied.

Linear Attack

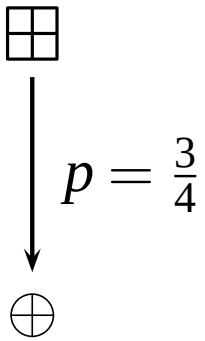
- Once an approximation is found, the adversary can use it to recover bits of the key or to distinguish the function from a random one.
- The adversary asks for pairs of input and output and evaluates the expression using them.
- Each such successful evaluation gives 1-bit worth of information about the key.

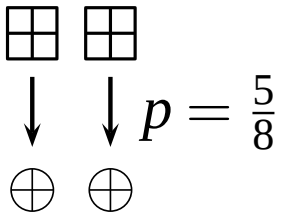
Linear Attack

- Once an approximation is found, the adversary can use it to recover bits of the key or to distinguish the function from a random one.
- The adversary asks for pairs of input and output and evaluates the expression using them.
- Each such successful evaluation gives 1-bit worth of information about the key.
- Unfortonately, since hash functions are unkeyed primitives there is no key to recover.
- Therefore, linear cryptanalysis can only be used to assess the security of the function without actually damaging it.

Linear Approximation of Addition Modulo 2^{32}

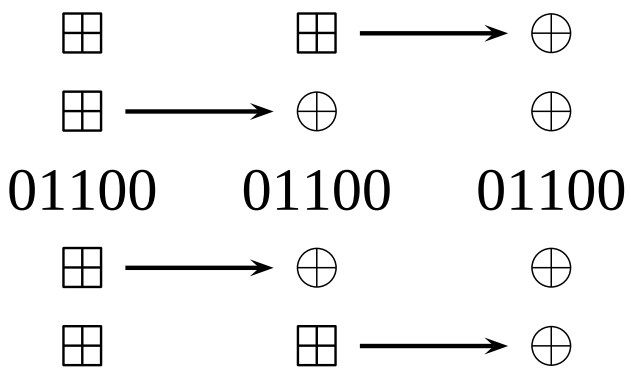
- The only nonlinear operation in CubeHash is the addition modulo 2^{32}.
- However, two consecutive bits entering an addition give rise to a bias of $\frac{1}{4}$ in the output. [Cho and Pieperzyk]
- Moreover, any even number of consecutive bits can be handled as an independant pair (i.e., 4 consecutive bits can be treated as 2 pairs of consecutive bits).





Our Work

- Using a C program we iterated all single and double pairs of consecutive bits when running the round-function both forward and backward.
- The iteration for a certain pair stops when one of these events occured:
  - The rotation operation sent a pair of approximated bits to the MSB and LSB hence not adhering to the Cho and Pieperzyk framework.
  - A XOR operation create a single bit (i.e., $11, 12 \oplus 12, 13 = 11, 13$) hence not adhering to the Cho and Pieperzyk framework.
  - The total bias has become smaller than 2^{-256}.



$$0110 \lll 2 = 1001$$

$$0110 \oplus 0011 = 0101$$

Our Work - cont.

- Once we had a set of partial good approximations we combined the forward and backward approximations to form a long full approximation.
- The best approximation we found was an 11-round approximation offering a bias of 2^{-235}.

Number of Rounds	Bias
9	2^{-157}
10	2^{-199}
11	2^{-235}
12	2^{-289}
13	2^{-347}
14	2^{-407}

Extending the Approximation Using Message Modification

- Once we had our 11-round approximation we have decided to extend our approximation by fixing some input bits to make sure they do not produce any carry during their evaluation.
- Knudsen and Mathiassen has shown that the when making sure the LSB of a pair is 0, an addition operation never produces a carry inside this pair.
- Therefore, this allowed us to extend out 11-round approximation into a 12-round approximation without any effect over the bias in the cost of fixing 116 bits.

Conclusion

- We've shown a linear approximation of 11-round CubeHash with bias 2^{-235}.
- We stress that this analysis does not come merely close to breaking CubeHash, it can be used only to assess its security.
- We continue this work to evaluate other candidates of the final round of SHA-3. We already found an 11-round approximation of Skein.