Efficient Collision-Resistant Hashing

Rachit Garg
CS14B050

November 19, 2017

Based on paper
Efficient Collision-Resistant Hashing from Worst-Case Assumptions on Cyclic Lattices
Published by Chris Peikert and Alon Rosen
Work appeared in 3 rd Theory of Cryptography Conference (TCC 2006) [8].

Contents

1 Problem Motivation

Collision-resistant hash functions are one of the most widely-employed cryptographic primitives. Their applications include integrity checking, user and message authentication, commitment pro-

tocols, and more. Many of the applications of collision-resistant hashing tend to invoke the hash function only a small number of times. Thus, the efficiency of the function has a direct effect on the efficiency of the application that uses it. Collision-resistance can be obtained from many well-studied complexity assumptions, but the resulting hash functions are not efficient enough for practical use. Instead, faster heuristic constructions such as MD5 and SHA-1 are often employed. Unfortunately, recent cryptanalytic analysis of many popular hash functions casts doubt on the heuristic approach [9],[10]. This paper proposed an efficient collision-resistant hash function with rigorous security guarantees.

2 Context and problem statement

Collision resistant functions: A function family $\{f_a\}, a \in A$ is said to be collision-resistant if given a uniformly chosen $a \in A$, it is infeasible to find elements $x_1 \neq x_2$ so that $f_a(x_1) = f_a(x_2)$.
Generalized Knapsacks: For a ring R, key $a = (a_1, ..., a_m) \in R^m$, and input $x = (x_1, ..., x_m)$.

$$f_a(x) = \sum_{i=1}^{m} a_i x_i$$

where each x_i is restricted to some large subset $S \subseteq R$. This generalization was proposed by Micciancio, who suggested a specific choice of the ring R and subset S for which inverting the function (for random a, x) is at least as hard as solving certain worst-case problems on cyclic lattices [5].
The authors mention that even though many knapsack systems have been broken heuristically, there is still no asymptotically-efficient attack on the general function. The paper studied continued Micciancios line of study, and showed that for a different choice of $S \subset R$, the generalized knapsack function can enjoy even stronger cryptographic properties.

2.1 Cyclic Lattices

Lattices admit worst-case to average-case reductions. Ajtai first constructed a one-way function [1], which was later observed to also be collision-resistant [2]. These constructions tended to be asymptotically more efficient than those based on, e.g., modular exponentiation. An interesting special case is presented by cyclic lattices.
A lattice L is said to be cyclic if for any vector $x \in L$, its cyclic rotation also belongs to L. The cyclic rotation of $x = (x_0, ..., x_{n1})^T \in R^n$ is defined as $(x_{n-1}, x_0, ..., x_{n-2})^T$.
Currently no hardness results are known for problems on cyclic lattices (even in their exact versions), and the additional structure may indeed reduce the underlying hardness. However, state-of-the-art lattice algorithms appear not to benefit from cyclicity, and it seems reasonable to conjecture that standard problems on cyclic lattices are intractable, at least for small approximation factors.

3 Comparison with known solutions

	Security	**Efficiency**	**Lattice Class**	**Assumption**	**Approx. Factor**
Ajtai	CRHF	$O(n^2)$	General	SVP etc.	poly(n)
Cai, Nerurkar	CRHF	$O(n^2)$	General	SVP etc.	$n^{4+\epsilon}$
Micciancio	OWF	$\tilde{O}(n)$	Cyclic	GDD	$n^{1+\epsilon}$
Micciancio, Regev	CRHF	$O(n^2)$	General	SVP etc.	$\tilde{O}(n)$
This work	CRHF	$\tilde{O}(n)$	Cyclic	SVP etc.	$\tilde{O}(n)$

Table 1: Comparison of results in lattice-based cryptographic functions with worst-case to average case security reductions, to date. Efficiency means the key size and computation time, as a function of the lattice dimension n. Security denotes the functions main cryptographic property.

The work studied was similar to Micciancio's work on cylic lattices. However the reduction used to establish collision-resistance differs in many significant ways. Micciancios function is proven to be one-way, while the authors is collision-resistant. On the other hand, Micciancio relies on a presumably weaker worst-case assumption than theirs. The stronger assumption, combined with the algebraic view of cyclic lattices, makes the security reduction tighter and conceptually simpler.

Other works considered are slower in efficiency. The structure of cyclic lattices and the choice of ring admits very efficient implementations of the knapsack function: using a Fast Fourier Transform algorithm. The resulting time complexity of the function is $O(m \cdot n \cdot poly(\log n))$, with key size $O(m \cdot n \log n)$. Where m denotes the number of elements in the key.

4 Paper results

They formulated a reduction showing that for cyclic lattices of prime dimension n, the short independent vectors problem $SIVP$ reduces to (a slight variant of) the shortest vector problem SVP with only a factor of 2 loss in approximation factor. They note that factor of loss in approximation is not trivial and the prime dimension constraint is not restricting. For general lattices, the best known reduction loses a $\sqrt{n}$ factor [6]; furthermore, that reduction performs manipulations on its input lattice that can destroy the cyclicity property. Hence their reduction can be seen as the first connection between SIVP and SVP on cyclic lattices.

Also in using the Gaussian techniques of [7], they also establish a new bound on the discrete Gaussian distribution over general lattices, which may be of independent interest.

Their main result is that certain instantiations of the generalized knapsack function are collision resistant, assuming it is infeasible to approximate the shortest vector in cyclic lattices up to factors $\tilde{O}(n)$ almost linear in the dimension n. The construction is also efficient as noted in the previous section. To motivate their choice of knapsack function, they also show that Micciancios original one-way function is not collision-resistant, nor even universal one-way.

5 Techniques and ideas

The techniques used in the paper use the fact that cyclic lattices are closed under cyclic convolution with integer vectors. Furthermore, the lattice points naturally correspond to polynomials in $\mathbb{Z}[\alpha]/(\alpha^n - 1)$.
Convolution is defined as follows. For any $x = (x_0, ..., x_{n-1})^T \in \mathbb{R}^n$, define the rotation of x, denoted as $rot(x)$, to be the vector $(x_{n-1}, x_0, ..., x_{n2})^T$; similarly $rot_i(x) = rot(rot(x))$ is defined to be the rotation of x, taken i times. A lattice $\mathcal{L}$ is cyclic if for all $x \in \mathcal{L}$, $rot(x) \in \mathcal{L}$. For any integer $d \geq 1$, define the rotation matrix $Rot^d(x)$ to be the matrix $[x|rot(x)||rot^{d-1}(x)]$.

For any ring R, the (cyclic) convolution product of $x, y \in R^n$ is the vector $x \otimes y = Rot^n(x) \cdot y$, with entries

$$(x \otimes y)_k = \sum_{i+j=k \bmod n} x_i \cdot y_j$$

Hence we observe that in a cyclic lattice $\mathcal{L}$, the convolution of any $x \in \mathcal{L}$ with any integer vector $y \in \mathbb{Z}_n$ is also in the lattice: $x \otimes y \in \mathcal{L}$. This is because all the columns of $Rot^n(x)$ are in $\mathcal{L}$, and any integer combination of points in $\mathcal{L}$ is also in $\mathcal{L}$.

The divisors of $(\alpha^n - 1)$ in $\mathbb{Z}[\alpha]$ correspond to special cyclotomic linear subspaces of $\mathbb{R}^n$. These subspaces admit a natural partitioning into complementary pairs of orthogonal subspaces. Even more importantly, the subspaces are closed under cyclic rotation of vector coordinates, and under certain other conditions, these rotations are linearly independent. These facts imply a new connection between the SIVP and SVP problems in cyclic lattices.

The security of the knapsack function comes from using all this structure to impose an algebraic restriction on the function domain. Looking ahead to the security reduction, this restriction ensures that collisions in the function are very likely to yield "useful" and short lattice points in a desired subspace.

6 Definitions

We go through some definitions that will help to show the main worst case to average case reduction.

Definition 6.1. *Cyclotomic Subspace*

$$H_\Phi = \{x \in \mathbb{R}^n : \Phi(\alpha) \text{ divides } x(\alpha) \in \mathbb{R}[\alpha]\}.$$

Definition 6.2. *SubSIVP*

The cyclotomic (generalized) short independent vectors problem, $SubSIVP_\gamma^\zeta$, given an n-dimensional full-rank cyclic lattice basis B and an integer polynomial $\Phi(\alpha) \neq 0 \bmod (\alpha^n - 1)$ that divides $\alpha^n - 1$, asks for a set of $dim(H_\Phi)$ linearly independent (sub)lattice vectors $S \subset \mathcal{L}(B) \cap H_\Phi$ such that $||S|| \leq \gamma(n)\zeta(\mathcal{L}(B) \cap H_\Phi)$.

Definition 6.3. *SubSVP*

The cyclotomic (generalized) short vectors problem, $SubSVP_\gamma^\zeta$, given an n-dimensional full-rank cyclic lattice basis B and an integer polynomial $\Phi(\alpha) \neq 0 \bmod (\alpha^n - 1)$ that divides $\alpha^n - 1$, asks for a (sub)lattice vector $c \in \mathcal{L}(B) \cap H_\Phi$ such that $||c|| \leq \gamma(n)\zeta(\mathcal{L}(B) \cap H_\Phi)$.

Definition 6.4. *SubIncSVP*

The cyclotomic (generalized) short vectors problem, $SubIncSVP_\gamma^\zeta$, given an n-dimensional full-rank cyclic lattice basis B and an integer polynomial $\Phi(\alpha) \neq 0 \bmod (\alpha^n - 1)$ that divides $\alpha^n - 1$, and a nonzero (sub)lattice vector $c \in \mathcal{L}(B) \cap H_\Phi$ such that $||c|| > \gamma(n)\zeta(\mathcal{L}(B) \cap H_\Phi)$, asks for a non-zero (sub)lattice vector $c' \in \mathcal{L}(B) \cap H_\Phi$ such that $||c'|| \leq ||c||/2$.

7 Lemmas

Lemma 7.1. *H_Φ is closed under rotation, thats if $c \in H_\Phi$, then $rot(c) \in H_\Phi$*

Proof. $$rot(c) = \alpha \cdot c(\alpha) \bmod (\alpha^n - 1)$$
$$(\alpha^n - 1) | rot(c) - \alpha \cdot c(\alpha)$$
$$\Phi(\alpha) | rot(c) - \alpha \cdot c(\alpha)$$
$$\Phi(\alpha) | rot(c)$$

□

We omit other proofs and just mention the lemmas and ideas in the paper.

Lemma 7.2. *Let $c \in \mathbb{Z}^n$, and suppose $\Phi(\alpha) \in \mathbb{Z}[\alpha]$ divides $(\alpha^n - 1)$ and is coprime to $c(\alpha)$. Then $c, rot(c)...., rot^{deg(\Phi)-1}(c)$ are linearly independent.*

Lemma 7.3. *Let $a, b \in \mathbb{R}^n$ with $a(\alpha) \cdot b(\alpha) = 0$ mod $(\alpha^n - 1)$. Then $\langle a, b \rangle = 0$.*

Lemma 7.4. *H_Φ is a linear subspace of $\mathbb{R}^n$ of dimension $n - deg(\Phi)$.*

8 Reductions

Proposition 8.1. *For any $\zeta, \gamma(n)$, there is a deterministic, polynomial-time sublattice-preserving reduction from $SubSVP_\gamma^\zeta$ to $SubIncSVP_\gamma^\zeta$.*

Informal Proof: Given an instance of $(B, \Phi(\alpha))$, iteratively reduce the length of c by invoking oracle for $SubIncSVP_\gamma^\zeta$ on $(B, \Phi(\alpha), c)$. If oracle fails we have solved $SubSVP_\gamma^\zeta$. (Note it is easy to show that the iterative process lasts poly number of times).

Proposition 8.2. *For any $\zeta, \gamma(n)$, there is a deterministic, polynomial-time sublattice-preserving reduction from $SubSIVP_\gamma^\zeta$ to $SubSVP_\gamma^\zeta$ which makes one oracle call where $\Phi(\alpha) = (\alpha^n - 1)/\Phi_k(\alpha)$ for some $k|n$.*

Informal Proof: We use the $SubSVP$ oracle to find a short vector in $\mathcal{L}(B) \cap H_{\Phi_1}$, then rotate it to yield $n - 1$ linearly independent vectors, and output that answer for $SSubSIVP$.

Proposition 8.3. *For any $\zeta, \gamma(n)$, there is a deterministic, polynomial-time lattice-preserving reduction from $SIVP_{\max(n,2\gamma)}$ on a cyclic lattice of prime dimension to $SubSVP_\gamma^{\lambda_1}$ which makes one oracle call where $\Phi(\alpha) = \Phi_1(\alpha) = \alpha - 1$.*

Informal Proof: The main idea behind the proof is as follows: first, we use the $SubSVP$ oracle to find a short vector in $\mathcal{L}(B) \cap H_{\Phi_1}$, then rotate it to yield $n - 1$ linearly independent vectors. For the nth vector, we take the shortest vector in $\mathcal{L}(B) \cap H_{\Phi_n}$, which can be found efficiently;

furthermore, it is an n-approximation to the shortest vector in $L(B) \setminus H_{\Phi_1}$.

$s_i = \sum_{j=1}^{n} (\mathbf{b}_i)_j = \mathbf{b}_i(1)$ for $i = 1, ...n$. Because $\alpha - 1$ cannot divide every $b_i(\alpha)$ (otherwise $\mathcal{L}(B) \subset H_{\Phi_1}$, so $\mathcal{L}(B)$ would not be full-rank), some s_i must be non-zero.
$\mathbf{s}_i = \mathbf{b}_i \otimes (1, 1, ...) = (s_i, s_i....s_i) \in \mathcal{L}(B)$. Let $g = gcd(s_1, s_2....s_n)$. Output $\mathbf{g} = (g, g....g)$ as the shortest vector. By the extended Euclidean algorithm, g is an integer combination of the s_i vectors, hence $g \in \mathcal{L}(B)$.

Proposition 8.4. *For any $\gamma(n)$, there is a deterministic, polynomial-time lattice-preserving reduction from $SVP_{\max(n,\gamma)}$ on a cyclic lattice of prime dimension to $SubSVP_{\gamma}^{\lambda_1}$ which makes one oracle call where $\Phi(\alpha) = \Phi_1(\alpha) = \alpha - 1$.*

Informal Proof The idea is same as in the previous reduction where we used the oracle to solve $SIVP$. To use the oracle to solve SVP, output minimum of norm of c, and the norm of the n^{th} vector as chosen before.

Hence we have claimed reductions from $SVP_{\max(n,\gamma)}$ to $SubSVP$, and $SubSVP$ to $SubIncSVP$. Thus we now try to show a reduction from $SubIncSVP$ to finding collisions in generalized knapsack with a specific choice of ring. This will show a reduction from $SVP_{\max(n,\gamma)}$ to the main collision resistant hash funcion.

9 Finding Collisions

We try to find collisions in the general knapsack function and try to motivate the papers choice of ring R and subset S.

Generalized Compact Knapsacks: For any ring R, subset $S \subset R$ and integer $m \geq 1$, the generalized function family $H(R, S, m) = \{f_a : S^m \to R\}_{a \in R^m}$ is defined by:

$$f_a(x) = \sum_{i=1}^{m} x_i \cdot a_i$$

We observe that f_A is linear:

$$f_A(X) + f_A(X') = f_A(X + X')$$

For random key A, to find a collision with X' it suffices to find a non-zero $X \in S^m$ such that $f_A(X) = 0$, and $||X||_\infty$ is small.(So that the bound on $||X||_\infty$ is still satisfied).

$$f_a(x) = \sum_{i=1}^{m} x_i(\alpha) \cdot a_i(\alpha) mod(\alpha^n - 1)$$

We define $X = (x_1, x_2...x_m)$ as follows, for any q that is a divisor of n:

$$x_1(\alpha) = \frac{\alpha^n - 1}{\alpha^q - 1}$$
$$x_j(\alpha) = 0 | j \neq 1$$

Now $f_A(X) = 0$ if $a_1(\alpha)$ is divisible by $a^q - 1$, note that this happens with probability $1/p^q$. Over a uniform choice of A, so we have found a specific $X \neq 0$, such that $f_A(X) \neq 0$ with non-negligible probability.

The fact enabling this attack is that $(\alpha^n - 1)$ is not irreducible in $\mathbb{Z}_p[\alpha]$. So it easy to find $x(\alpha)$ with small coefficients such that $a(\alpha)x(\alpha) = 0 \bmod (\alpha^n - 1)$ and for each divisor of $(\alpha^n - 1)$ either $x(\alpha) = 0$ or $a(\alpha) = 0$.
To prevent this we enforce an algebraic constraint on X, informally we require every $x_i(\alpha)$ to be divisible over $\mathbb{Z}[\alpha]$ by $\frac{\alpha^n-1}{\Phi_k(\alpha)}$ for some fixed k dividing n. Now essentially the evaluation is performed mod $\Phi_k(\alpha)$. Their choice of subset $S_{D,\Phi}$ is as follows.

$$S_{D,\Phi} = \{x \in \mathbb{Z}_p^n : ||x||_\infty \leq D \text{ and } \Phi(\alpha) \text{ divides } x_\mathbb{Z}(\alpha) \text{ in } \mathbb{Z}[\alpha]\}$$

10 The main reduction

We reduce from $SubIncSVP_\gamma^{\eta_\epsilon}$ to collision function $H(\mathbb{Z}_{p(n)}, S_{D(n),\Phi}, m(n))$. Note that this is a worst case to average case reduction on cyclic lattices.
Here we try to show three main properties:

1. Sampling an average instance.
2. Outputting a new vector and showing that it belongs in $\mathcal{L}(B) \cap H$.
3. It has a norm smaller than half the norm of input vector in $SubIncSVP$, with non negligible probability.

10.1 The algorithm

1. For $i = 1$ to m,
   - Generate uniform $\mathbf{v}_i \in \mathcal{L}(B) \cap H \cap P(Rot^d(\mathbf{c}))$. [6]
   - Generate noise $\mathbf{y}_i \in H$, according to $D_{H,s}$ for $s = 2||c||/\gamma(n)$. Let $y_i' = y_i \bmod P(B)$.
   - Choose b_i so that $Rot^n(\mathbf{c}) \cdot \mathbf{b} = \mathbf{v_i} + \mathbf{y_i'}$, and let $\mathbf{a_i} = \lfloor \mathbf{b_i} \cdot p \rceil$
2. Pass A to collision finding oracle, and get collission pairs X, X', let $Z = X - X'$, such that $||Z||_\infty \leq 2D$ and $\Phi(\alpha)$ divides every $z_i(\alpha)$.
3. Output

$$c' = \sum_{i=1}^{m} (\mathbf{v_i} + \mathbf{y_i'} - \mathbf{y_i}) \otimes \mathbf{z_i} - \mathbf{c} \otimes \frac{\sum_{i=1}^{m} \mathbf{a_i} \otimes \mathbf{z_i}}{p}$$

10.2 Correctness

- *Property - Instance sampled is average instance*: It is uniform because for choosing b we sample the latter half uniformly from I^{n-d}. And the former half is chosen such that $(I_{d\times d})^{-1}(v_i + y_i' - w)$, where $w = Rot^n(c) \cdot (0, 0, ...(b_i)_d, .., (b_i)_n)^T$. Since y' is statistically uniform from our choice of c such that the spread in the guassian is more than the smoothing parameter, and v is uniform. We have the former half is uniform, and since the latter half is already sampled uniformly we have that the $b_i's$ are uniform.

- *Property - Outputting a new vector and showing that it belongs in* $\mathcal{L}(B) \cap H$: Convolution of a lattice vector with an integer vector also lies in the lattice, hence the second property holds true. Note that the second term in the output vector is c convoluted with an integer vector due to a convolution z being congruent to zero mod p as we have constructed z from a collision.

- *Property - Outputted vector is smaller than half of the input vector*: Idea is to use Markov's Inequality and bound on the expected value of the new vector that is outputted.

11 Future Work

In the same year Lyubashevsky and Micciancio [3] obtained exceedingly similar results but expressed them in different mathematical language. In particular, by making many of the same algebraic insights, they constructed collision-resistant hash functions with nearly identical parameters, based on a worst-case hardness assumption that was similar to the paper studied. They also presented a more general algebraic framework for constructing hash functions, which can be related to problems in algebraic number theory. Due to its generality, their framework may have the potential to admit better constructions, though its current best application essentially matches the collision resistant function considered here.

A more practical instantiation of the function considered here was presented in [4]. They propose a collection of compression functions that are highly parallelizable and admit very efficient implementations on modern microprocessors. Their constructions were supported with a detailed security analysis of concrete instantiations, and a high-performance software implementation that exploits the inherent parallelism of the FFT algorithm. They claim that the throughput of their implementation is competitive with that of $SHA - 256$, with additional parallelism yet to be exploited.

References

[1] M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, 1996.

[2] Mihir Bellare and Daniele Micciancio. A new paradigm for collision-free hashing: Incrementality at reduced cost. In *Proceedings of the 16th Annual International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT'97, 1997.

[3] Vadim Lyubashevsky and Daniele Micciancio. Generalized compact knapsacks are collision resistant. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, 2006.

[4] Vadim Lyubashevsky, Daniele Micciancio, Chris Peikert, and Alon Rosen. Fast software encryption. chapter SWIFFT: A Modest Proposal for FFT Hashing. 2008.

[5] Daniele Micciancio. Generalized compact knapsacks, cyclic lattices, and efficient one-way functions from worst-case complexity assumptions. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, 2002.

[6] Daniele Micciancio and Shafi Goldwasser. *Complexity of Lattice Problems: a cryptographic perspective*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 2002.

[7] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Comput.*

[8] Chris Peikert and Alon Rosen. Efficient collision-resistant hashing from worst-case assumptions on cyclic lattices. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, 2006.

[9] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full sha-1. In *Proceedings of the 25th Annual International Conference on Advances in Cryptology*, CRYPTO'05, 2005.

[10] Xiaoyun Wang and Hongbo Yu. How to break md5 and other hash functions. In *Proceedings of the 24th Annual International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'05.