MMH* with arbitrary modulus is always almost-universal

Khodakhast Bibak * Bruce M. Kapron * Venkatesh Srinivasan *†

October 13, 2020

Abstract

Universal hash functions, discovered by Carter and Wegman in 1979, are of great importance in computer science with many applications. MMH* is a well-known $\triangle$-universal hash function family, based on the evaluation of a dot product modulo a prime. In this paper, we introduce a generalization of MMH*, that we call GMMH*, using the same construction as MMH* but with an arbitrary integer modulus $n > 1$, and show that GMMH* is $\frac{1}{p}$-almost-$\triangle$-universal, where p is the smallest prime divisor of n. This bound is tight.

1 MMH*

Universal hashing, introduced by Carter and Wegman [3], is of great importance in computer science with many applications. Cryptography, information security, complexity theory, randomized algorithms, and data structures are just a few areas that universal hash functions and their variants have been used as a fundamental tool. In [5], definitions of various kinds of universal hash functions gathered from the literature are presented; we mention some of them here.

Definition 1.1. Let H be a family of functions from a domain D to a range R. Let ε be a constant such that $\frac{1}{|R|} \leq \varepsilon < 1$. The probabilities below are taken over the random choice of hash function h from the set H.
(i) The family H is a *universal family of hash functions* if for any two distinct $x, y \in D$, we have $\Pr_{h \leftarrow H}[h(x) = h(y)] \leq \frac{1}{|R|}$. Also, H is an *ε-almost-universal* (ε-AU) *family of hash functions* if for any two distinct $x, y \in D$, we have $\Pr_{h \leftarrow H}[h(x) = h(y)] \leq \varepsilon$.
(ii) Suppose R is an Abelian group. The family H is a *$\triangle$-universal family of hash functions* if for any two distinct $x, y \in D$, and all $b \in R$, we have $\Pr_{h \leftarrow H}[h(x) - h(y) = b] = \frac{1}{|R|}$, where ‘ $-$ ’ denotes the group subtraction operation. Also, H is an *ε-almost-$\triangle$-universal* (ε-A$\triangle$U) *family of hash functions* if for any two distinct $x, y \in D$, and all $b \in R$, we have $\Pr_{h \leftarrow H}[h(x) - h(y) = b] \leq \varepsilon$.

*Department of Computer Science, University of Victoria, Victoria, BC, Canada V8W 3P6. Email: {kbibak,bmkapron,srinivas}@uvic.ca
†Centre for Quantum Technologies, National University of Singapore, Singapore 117543.

It is worth mentioning that ε-A$\triangle$U families have also important applications in computer science, in particular, in cryptography. For example, these families can be used in message authentication. Informally, it is possible to design a message authentication scheme using ε-A$\triangle$U families such that two parties can exchange signed messages over an unreliable channel and the probability that an adversary can forge a valid signed message to be sent across the channel is at most ε ([5]).

The following family, named MMH* by Halevi and Krawczyk [5] in 1997, is a well-known $\triangle$-universal hash function family.

Definition 1.2. Let p be a prime and k be a positive integer. The family MMH* is defined as follows:

$$\mathrm{MMH}^* := \{g_{\mathbf{x}} \;:\; \mathbb{Z}_p^k \to \mathbb{Z}_p \mid \mathbf{x} \in \mathbb{Z}_p^k\}, \tag{1.1}$$

where

$$g_{\mathbf{x}}(\mathbf{m}) := \mathbf{m} \cdot \mathbf{x} \pmod{p} = \sum_{i=1}^{k} m_i x_i \pmod{p}, \tag{1.2}$$

for any $\mathbf{x} = \langle x_1, \ldots, x_k \rangle \in \mathbb{Z}_p^k$, and any $\mathbf{m} = \langle m_1, \ldots, m_k \rangle \in \mathbb{Z}_p^k$.

It appears that Gilbert, MacWilliams, and Sloane [4] first discovered MMH* (but in the finite geometry setting). However, many resources attribute MMH* to Carter and Wegman [3]. Halevi and Krawczyk [5] proved that MMH* is a $\triangle$-universal family of hash functions. We also remark that, recently, Leiserson et al. [7] rediscovered MMH* (called it "DOTMIX compression function family") and using the same method as Halevi and Krawczyk, proved that DOTMIX is $\triangle$-universal. They then apply this result to the problem of deterministic parallel random-number generation for dynamic multithreading platforms in parallel computing.

Theorem 1.3. *The family* MMH* *is a $\triangle$-universal family of hash functions.*

2 GMMH*

Given that, in the definition of MMH*, the modulus is a prime, it is natural to ask what happens if the modulus is an arbitrary integer $n > 1$. Is the resulting family still $\triangle$-universal? If not, what can we say about ε-almost-universality or ε-almost-$\triangle$-universality of this new family? This is an interesting and natural problem, and while it has a simple solution (see, Theorem 2.3 below), to the best of our knowledge there are no results regarding this problem in the literature.

First, we define a generalization of MMH*, namely, GMMH*, with the same construction as MMH* except that we use an arbitrary integer $n > 1$ instead of prime p.

Definition 2.1. Let n and k be positive integers ($n > 1$). The family GMMH* is defined as follows:

$$\mathrm{GMMH}^* := \{h_{\mathbf{x}} \;:\; \mathbb{Z}_n^k \to \mathbb{Z}_n \mid \mathbf{x} \in \mathbb{Z}_n^k\}, \tag{2.1}$$

where

$$h_{\mathbf{x}}(\mathbf{m}) := \mathbf{m} \cdot \mathbf{x} \pmod{n} = \sum_{i=1}^{k} m_i x_i \pmod{n}, \tag{2.2}$$

for any $\mathbf{x} = \langle x_1, \ldots, x_k \rangle \in \mathbb{Z}_n^k$, and any $\mathbf{m} = \langle m_1, \ldots, m_k \rangle \in \mathbb{Z}_n^k$.

MMH* has found important applications, however, in applications that, for some reasons, we have to work in the ring $\mathbb{Z}_n$, the family GMMH* may be used.

The following result, proved by D. N. Lehmer [6], is the main ingredient in the proof of Theorem 2.3.

Proposition 2.2. *Let $a_1, \ldots, a_k, b, n \in \mathbb{Z}$, $n \geq 1$. The linear congruence $a_1x_1 + \cdots + a_kx_k \equiv b \pmod{n}$ has a solution $\langle x_1, \ldots, x_k \rangle \in \mathbb{Z}_n^k$ if and only if $\ell \mid b$, where $\ell = \gcd(a_1, \ldots, a_k, n)$. Furthermore, if this condition is satisfied, then there are ℓn^{k-1} solutions.*

Now, we are ready to state and prove the following result about ε-almost-$\triangle$-universality of GMMH*.

Theorem 2.3. *Let n and k be positive integers ($n > 1$). The family* GMMH* *is $\frac{1}{p}$-A$\triangle$U, where p is the smallest prime divisor of n. This bound is tight.*

Proof. Suppose that n has the prime factorization $n = p_1^{r_1} \ldots p_s^{r_s}$, where $p_1 < \cdots < p_s$ are primes and $r_1, \ldots, r_s$ are positive integers. Let $\mathbf{m} = \langle m_1, \ldots, m_k \rangle \in \mathbb{Z}_n^k$ and $\mathbf{m}' = \langle m'_1, \ldots, m'_k \rangle \in \mathbb{Z}_n^k$ be any two distinct messages. Put $\mathbf{a} = \langle a_1, \ldots, a_k \rangle = \mathbf{m} - \mathbf{m}'$. For every $b \in \mathbb{Z}_n$ we have

$$h_{\mathbf{x}}(\mathbf{m}) - h_{\mathbf{x}}(\mathbf{m}') = b \Longleftrightarrow \sum_{i=1}^{k} m_i x_i - \sum_{i=1}^{k} m'_i x_i \equiv b \pmod{n} \Longleftrightarrow \sum_{i=1}^{k} a_i x_i \equiv b \pmod{n}.$$

Note that since $\langle x_1, \ldots, x_k \rangle \in \mathbb{Z}_n^k$, we have n^k ordered k-tuples $\langle x_1, \ldots, x_k \rangle$. Also, since $\mathbf{m} \neq \mathbf{m}'$, there exists some i_0 such that $a_{i_0} \neq 0$. Now, we need to find the maximum number of solutions of the above linear congruence over all choices of $\mathbf{a} = \langle a_1, \ldots, a_k \rangle \in \mathbb{Z}_n^k \setminus \{\mathbf{0}\}$ and $b \in \mathbb{Z}_n$. By Proposition 2.2, if $\ell = \gcd(a_1, \ldots, a_k, n) \nmid b$ then the linear congruence $a_1x_1 + \cdots + a_kx_k \equiv b \pmod{n}$ has no solution, and if $\ell = \gcd(a_1, \ldots, a_k, n) \mid b$ then the linear congruence has ℓn^{k-1} solutions. Thus, we need to find the maximum of $\ell = \gcd(a_1, \ldots, a_k, n)$ over all choices of $\mathbf{a} = \langle a_1, \ldots, a_k \rangle \in \mathbb{Z}_n^k \setminus \{\mathbf{0}\}$. Clearly,

$$\max_{\mathbf{a} = \langle a_1, \ldots, a_k \rangle \in \mathbb{Z}_n^k \setminus \{\mathbf{0}\}} \gcd(a_1, \ldots, a_k, n)$$

is achieved when $a_{i_0} = p_1^{r_1 - 1} p_2^{r_2} \ldots p_s^{r_s} = \frac{n}{p_1}$, and $a_i = 0$ $(i \neq i_0)$. So, we get

$$\max_{\mathbf{a} = \langle a_1, \ldots, a_k \rangle \in \mathbb{Z}_n^k \setminus \{\mathbf{0}\}} \gcd(a_1, \ldots, a_k, n) = p_1^{r_1 - 1} p_2^{r_2} \ldots p_s^{r_s} = \frac{n}{p_1}.$$

Therefore, for any two distinct messages $\mathbf{m}, \mathbf{m}' \in \mathbb{Z}_n^k$, and all $b \in \mathbb{Z}_n$, we have

$$\Pr_{h_{\mathbf{x}} \leftarrow \mathrm{GMMH}^*}[h_{\mathbf{x}}(\mathbf{m}) - h_{\mathbf{x}}(\mathbf{m}') = b] \leq \max_{\mathbf{a} = \langle a_1, \ldots, a_k \rangle \in \mathbb{Z}_n^k \setminus \{\mathbf{0}\}} \frac{n^{k-1} \gcd(a_1, \ldots, a_k, n)}{n^k} = \frac{1}{p_1}.$$

This means that GMMH* is $\frac{1}{p_1}$-A$\triangle$U. Clearly, this bound is tight; take, for example, $a_1 = \frac{n}{p_1}$ and $a_2 = \cdots = a_k = 0$. $\square$

Corollary 2.4. *If in* Theorem 2.3 *we let n be a prime then we obtain* Theorem 1.3.

Proof. When n is prime, $\gcd_{\mathbf{a}=\langle a_1,\ldots,a_k\rangle\in\mathbb{Z}_n^k\setminus\{\mathbf{0}\}}(a_1,\ldots,a_k,n)=1$, so we get $\triangle$-universality. $\square$

We remark that if in the family GMMH* we let the keys $\mathbf{x}=\langle x_1,\ldots,x_k\rangle\in\mathbb{Z}_n^k$ satisfy the general conditions $\gcd(x_i,n)=t_i$ $(1\le i\le k)$, where $t_1,\ldots,t_k$ are given positive divisors of n, then the resulting family, which was called GRDH in [2], is no longer 'always' ε-A$\triangle$U. In fact, it was shown in [2] that the family GRDH is ε-A$\triangle$U for some $\varepsilon<1$ if and only if n is odd and $\gcd(x_i,n)=t_i=1$ (that is, $x_i\in\mathbb{Z}_n^*$) for all i. Furthermore, if these conditions are satisfied then GRDH is $\frac{1}{p-1}$-A$\triangle$U, where p is the smallest prime divisor of n (this bound is also tight). This result is then applied in giving a generalization of a recent authentication code with secrecy. A key ingredient in the proofs in [2] is an explicit formula for the number of solutions of restricted linear congruences (a restricted version of Proposition 2.2), recently obtained by Bibak et al. [1], using properties of Ramanujan sums and of the finite Fourier transform of arithmetic functions.

Acknowledgements

The authors would like to thank the editor and anonymous referees for helpful comments. During the preparation of this work the first author was supported by a Fellowship from the University of Victoria (UVic Fellowship).

References

[1] K. Bibak, B. M. Kapron, V. Srinivasan, R. Tauraso, and L. Tóth, Restricted linear congruences, arXiv: 1503.01806.

[2] K. Bibak, B. M. Kapron, V. Srinivasan, and L. Tóth, On an almost-universal hash function family with applications to authentication and secrecy codes, arXiv: 1507.02331.

[3] J. L. Carter and M. N. Wegman, Universal classes of hash functions, *J. Comput. System Sci* **18** (1979), 143–154.

[4] E. N. Gilbert, F. J. MacWilliams, and N. J. A. Sloane, Codes which detect deception, *Bell Syst. Tech. J.* **53** (1974), 405–424.

[5] S. Halevi and H. Krawczyk, MMH: Software message authentication in the Gbit/second rates, *Fast Software Encryption — FSE 1997*, LNCS **1267**, 1997, 172–189.

[6] D. N. Lehmer, Certain theorems in the theory of quadratic residues, *Amer. Math. Monthly* **20** (1913), 151–157.

[7] C. E. Leiserson, T. B. Schardl, and J. Sukha, Deterministic parallel random-number generation for dynamic-multithreading platforms, *Proceedings of the 17th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming — PPoPP 2012*, 193–204.