More Efficient Privacy Amplification with Less Random Seeds via Dual Universal Hash Function

Masahito Hayashi[1,2] Toyohiro Tsurumaru[3]
[1] Graduate School of Mathematics, Nagoya University, Japan,
[2]Centre for Quantum Technologies, National University of Singapore, Singapore
[3] Mitsubishi Electric Corporation, Information Technology R&D Center, Kanagawa 247-8501, Japan

Abstract

We explicitly construct random hash functions for privacy amplification (extractors) that require smaller random seed lengths than the previous literature, and still allow efficient implementations with complexity $O(n \log n)$ for input length n. The key idea is the concept of *dual* universal$_2$ hash function introduced recently. We also use a new method for constructing extractors by concatenating δ-almost dual universal$_2$ hash functions with other extractors.

Besides minimizing seed lengths, we also introduce methods that allow one to use non-uniform random seeds for extractors. These methods can be applied to a wide class of extractors, including dual universal$_2$ hash function, as well as to conventional universal$_2$ hash functions. The technical details are in arXiv:1311.5322 (2013).

Index Terms

privacy amplification, universal hash function, minimum entropy, quantum cryptography

Background: Even when a random source at hand is partially leaked to an eavesdropper, one can amplify its secrecy by applying a random hash function. This process is called the *privacy amplification*. In this process, the amplification of secrecy is realized with the help of another auxiliary random source, which is public and is called a *random seed*. The random hash functions used for this purpose are often called *extractors*. There is also a similar but distinct process called two-sources-extractors [9], where the auxiliary random source is not public. The most typical random hash function for these purposes is the universal$_2$ hash function [5], [45]. There are many security theorems which assumes the use of the universal$_2$ hash function. In particular, the leftover hashing lemma [4], [14] has several extensions and various applications in the classical and quantum setting [30], [39], [17], [18], [24], [16], [19], [20], [27].

The universal$_2$ hash function has now become indispensable for privacy amplification of quantum key distribution (QKD) [3], [30], [40], [23], [22]. The most widely used universal$_2$ hash function for this purpose is the one that uses the (modified) Toeplitz matrix, mainly because it can be implemented efficiently with complexity $O(n \log n)$ for input length n (see, e.g., [32], [43]). Here we note that the usual notion of efficiency (i.e., the algorithm finishes in polynomial time) is not sufficient, but a stricter criterion of the complexity being $O(n \log n)$ is desirable for QKD. This is because, for typical QKD systems, the finite size effect requires the input length n to be $n \geq 10^6$ [40], [23], [22] , and thus algorithms that are efficient in the usual sense, e.g., $O(n^2)$, are useless.

Another important criterion for practical hash functions is how much randomness is required for the random seed. This can be measured in two way, i.e., by the required length of a uniformly random seed, and also by the entropy of the seed. While the importance of minimizing the former is obvious, the latter is also equally important, since it is quite difficult to prepare a perfect random number generator for real cryptographic systems.

The main goal of this paper is to construct explicitly random hash functions for privacy amplification that require smaller random seed lengths than in the previous literature, and still allow efficient implementations with complexity $O(n \log n)$ for input length n. For achieving this goal, we use the concept of δ-almost *dual* universal$_2$ hash function. We also use a new method for constructing extractors by concatenating δ-almost *dual* universal$_2$ hash functions and conventional extractors.

In addition to minimizing the seed lengths, we also present general methods that enable the use of non-uniform random seeds. These methods are general in the sense that they can be applied a wide class of extractors, including dual universal$_2$ hash function, as well as to conventional universal$_2$ hash functions. Here the minimum entropy is used as a measure that describes the randomness of the non-uniform random seed.

The concept of the δ-almost *dual* universal$_2$ hash function, as well as the extended leftover hashing lemma for it were proposed in Refs. [11], [43]. In [43], we also gave the explicit inclusion relation with the (conventional) universal$_2$ hash function; e.g., if an arbitrary linear and surjective hash function is universal$_2$ (with $\delta = 1$), then it is automatically δ-almost dual universal$_2$. In this sense, the δ-almost dual universal$_2$ function can be regarded as an extension of the conventional universal$_2$ function. Several classical and quantum security evaluations have been obtained based on this new class of hash functions [16], [19]. In particular, finite-length security analysis has been done with this class [23], [22].

Our proposed hash function:

Based on properties of conventional and dual universal$_2$ hash functions, the corresponding security criteria, and the corresponding leftover hashing lemmas. we propose a new method to construct random hash functions by concatenating given random hash functions. While a method is already known for concatenating two (conventional) δ-almost universal$_2$ hash functions [37], we are here rather interested in other combinations including δ-almost *dual* universal$_2$ hash functions. Then by exploiting these results, we present secure hash functions that require less random seed length h than previous methods, and can be implemented with complexity $O(n \log n)$. That is, we explicitly construct a set of extractors whose seed lengths are $\min(m, n-m)$ asymptotically, where n is the input length and m the output length. Recall that many of existing random hash functions, such as the one using the (modified) Toeplitz matrix (see the attachment) and the ones proposed recently [41], require seed length n or $2m$ asymptotically (see Table I). Here, we improve them by giving four types of hash functions explicitly. Namely, we first present $f_{\mathrm{F1},R}$ suitable for $m/n \geq 1/2$, and $f_{\mathrm{F2},R}$ suitable for $m/n \leq 1/2$, both requiring seed length $n-m$. Then by concatenating $f_{\mathrm{F2},R}$ and its dual $f^{\perp}_{\mathrm{F2},R}$, we construct $f_{\mathrm{F3},R}$ and $f_{\mathrm{F4},R}$ which require seed length m asymptotically.

In order to demonstrate that hash functions $f_{\mathrm{F1},R}, \ldots, f_{\mathrm{F4},R}$ can indeed be implemented efficiently with complexity $O(n \log n)$, we also give a set of explicit algorithms in the attachment. This algorithm set uses multiplication algorithm for finite field $\mathbb{F}_{2^k}$ developed, e.g., in Refs. [35], [26], and works for parameter k satisfying certain conditions related to Artin's conjecture [36, Chap. 21]. We numerically check the existence of so many such integers up to $k \simeq 10^{50}$, and thus the algorithm can be applied to most practical cases.

TABLE I
COMPARISON OF RANDOM HASH FUNCTIONS

	computational complexity	length of random seeds h & min entropy t when the seeds are uniformly random	
		ϵ const.	$\epsilon = e^{-\beta n^\gamma}$
$f_{\mathrm{F1},R}$, $f_{\mathrm{F2},R}$	$O(n \log n)$	$t = \alpha n + O(1)$ $h = (1-\alpha)n$	$t = \alpha n + 2\beta n^\gamma + O(1)$ $h = (1-\alpha)n$
$f_{\mathrm{F3},R}$	$O(n \log n)$	$t = \alpha n + O(1)$ $h = \alpha n + O(1)$	$t = \alpha n + 2\beta n^\gamma + O(1)$ $h = \alpha n + 4\beta n^\gamma + O(1)$
$f_{\mathrm{F4},R}$	$O(n \log n)$	$t = \alpha n + O(1)$ $h = \alpha n + O(1)$	$t = \alpha n + 4\beta n^\gamma + O(1)$ $h = \alpha n + 4\beta n^\gamma + O(1)$
Modified Toeplitz matrix	$O(n \log n)$	$t = \alpha n + O(1)$ $h = n$	$t = \alpha n + 2\beta n^\gamma + O(1)$ $h = n$
Trevisan's extractor [42], [6]	poly(n)	$t = \alpha n + O(1)$ $h = O(\log^3 n)$	$t = \alpha n + 4\beta n^\gamma + O(1)$ $h = O(n^{2\gamma} \log^3 n)$
TSSR paper [41]	$O(n \log n)^*$	$t = \alpha n + O(1)$ $h = 2\alpha n + O(1)$	$t = \alpha n + 4\beta n^\gamma + O(1)$ $h = 2\alpha n + 4\beta n^\gamma + O(1)$
ϵ-almost pairwise independent [28]	poly(n)	$t = \alpha n + O(1)$ $h = 4\alpha n + o(n)$	$t = \alpha n + 4\beta n^\gamma + O(1)$ $h = 4\alpha n + 4\beta n^\gamma + o(n)$
Strong blender (classical) [8]	poly(n)	$t = \alpha n + O(1)$ $h = n$	$t = \alpha n + 2\beta n^\gamma + O(1)$ $h = n$

$f_{\mathrm{F1},R}$, $f_{\mathrm{F2},R}$, $f_{\mathrm{F3},R}$, and $f_{\mathrm{F4},R}$ are hash functions proposed in this paper. Parameter n is the length of the input to the hash function, and ϵ is the security level (L_1 distinguishability) of the final key. Parameters h, t, α, γ are defined in order to compare the six schemes for a case where the random seeds are uniformly random: t is the required minimum entropy for the input to a hash function, αn the output length, h the required length of random seeds, and γ a constant in $(0, 1]$. We mainly choose $\gamma > 1/2$. $f_{\mathrm{F3},R}$ is a hash function for the classical case. $f_{\mathrm{F4},R}$ is its quantum modification. *The paper [41] did not evaluate the computational complexity. However, when we employ our construction of finite filed given in the attachment, we find that the computational complexity of the random hash function is $O(n \log n)$.

Comparison with existing hash functions: As to comparisons with the existing methods: Trevisan [42] proposed another efficient random hash function, whose performance was studied in the quantum case by [6]. The paper [41], [28] also proposed other random hash functions. As is also summarized in Table I, the relations with our hash function are as follows.

1) Our random hash functions, $f_{\mathrm{F1},R}$, $\ldots$, $f_{\mathrm{F4},R}$ and $g_{n,l,m}$, and those of Ref. [41] have an efficient algorithm with complexity $O(n \log n)$ for input length n. On the other hand, Ref. [8] only considers algorithms typically with complexity $O(n^3)$ (c.f. the attachment), and Ref. [28] with $\mathrm{poly}(n)$. For Trevisan's random extractor, a pre-computation is required and the complexity of the actual calculation is only shown to be polynomial in n. Although our random hash functions require a search for an integer k mentioned above, it should be noted that k of a desired size up to $k \simeq 10^{50}$ can be found in less than a second, and thus our random hash functions practically have no pre-computation.
2) For the case where the uniform random seeds are uniformly random, we also compare the required length h of random seeds, and the required minimum entropy t of the input to the hash function, as is summarized in Table I. Here we denote the input and output lengths by n and m, their ratio by $\alpha := m/n$, and the security level (L_1 distinguishability) of the final key by ϵ.
   - When both α and ϵ are constant, all random hash functions have almost the same required minimum input entropy t. While Trevisan's random extractor [42], [6] has the minimum value for the required length h of random seeds, the computational complexity is $O(\mathrm{poly}(n))$ and also requires a pre-computation. Our hash function $f_{\mathrm{F1},R}, f_{\mathrm{F2},R}$ or $f_{\mathrm{F3},R}, f_{\mathrm{F4},R}$ realizes the next minimum value dependently of α, and can be implemented efficiently with $O(n \log n)$ and with virtually no pre-computation.
   - Next, we consider the case where α is constant and ϵ is exponentially small with respect to n; that is, we assume that ϵ behaves as $e^{-\beta n^\gamma}$ with $\gamma > \frac{1}{2}$.[1] In this case our random hash function $f_{\mathrm{F1},R}, f_{\mathrm{F2},R}$ or $f_{\mathrm{F3},R}, f_{\mathrm{F4},R}$ achieves the minimum values of the required length h of random seeds and the required minimum input entropy t at least in the first order n, dependently of α.

Conclusion: We have proposed new random hash functions $f_{\mathrm{F1},R}$, $\ldots$, $f_{\mathrm{F4},R}$ using a finite field with a large size, which are designed based on the concepts of the δ-almost dual universal$_2$ hash function. The proposed method realizes the two advantages simultaneously. First, it requires the smallest length of random seeds. Second, there exist efficient algorithms for them achieving the calculation complexity of the smallest order, namely $O(n \log n)$. Note that no previously known methods, such as the one using the modified Toeplitz matrix, as well as those given in Refs. [6], [41], [28], can realize these two at the same time.

Although there are now several security analyses done with the δ-almost dual universality$_2$ [16], [19], a larger part of existing security analyses are still based on the conventional version of universality$_2$. The results obtained here clarify advantages of the δ-almost dual universal$_2$ hash function over the conventional one, and also demonstrate that they can be easily constructed in practice. We believe that these facts suggest the importance of further security analyses based on the δ-almost dual universality$_2$, from theoretical and practical viewpoints.

Acknowledgment

MH thanks Prof. Toru Uzawa, Prof. Ryutaroh Matsumoto, and Dr. Marco Tomamichel for valuable comments. MH also thanks Prof. Yaoyun Shi for explaining the concept "ϵ-almost pairwise independent hash function", and references [28], [2], [29], [13]. The authors are grateful to the referee of the previous version for explaining an idea for non-uniform random seeds. The authors are partially supported by the National Institute of Information and Communication Technology (NICT), Japan. MH is also partially supported by a MEXT Grant-in-Aid for Scientific Research (A) No. 23246071. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

References

[1] T. Asai, and T. Tsurumaru, "Efficient Privacy Amplification Algorithms for Quantum Key Distribution" (in Japanese), *IEICE technical report*, ISEC2010-121 (2011).

[1] Recall that, as is numerically shown in [44], when ϵ is too small in comparison with n, it is better to describe ϵ as an exponential function of n.

[2] N. Alon, O. Goldreich, J. HÃĕstad, and R. Peralta. "Simple constructions of almost k-wise independent random variables," *Random Structures & Algorithms*, 3(3):289-304, 1992.

[3] C. H. Bennett and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing", Proceedings of IEEE International Conference on Computers Systems and Signal Processing, Bangalore India, pp.175-179, December 1984.

[4] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized Privacy Amplification," *IEEE Trans. Inform. Theory*, **41**, 1915 (1995).

[5] L. Carter and M. Wegman, "Universal classes of hash functions," *J. Comput. System Sci.*, vol. **18**, No. 2, 143–154, 1979.

[6] Anindya De, Christopher Portmann, Thomas Vidick, Renato Renner, "Trevisan's extractor in the presence of quantum side information," *SIAM Journal on Computing*, 41(4):915-940, (2012).

[7] N. Dedić, D. Harnik, and L. Reyzin, "Saving Private Randomness in One-Way Functions and Pseudorandom Generators," Theory of Cryptography, *Lecture Notes in Computer Science*, Vol. 4948, 2008, pp 607-625

[8] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz, "Improved Randomness Extraction from Two Independent Sources" Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, *Lecture Notes in Computer Science*, Vol. 3122, 2004, pp 334-344.

[9] Y. Dodis, and R. Oliveira, "On Extracting Private Randomness over a Public Channel," Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques, *Lecture Notes in Computer Science*, Vol. 2764, 2003, pp 252-263.

[10] Y. Dodis and A. Smith. "Correcting Errors Without Leaking Partial Information," In *Proceedings of the 37th symposium on Theory of computing, STOC05*, pp. 654-663. ACM, 2005.

[11] S. Fehr and C. Schaffner. "Randomness Extraction via Delta-Biased Masking in the Presence of a Quantum Attacker," Theory of Cryptography Fifth Theory of Cryptography Conference, TCC 2008 New York, USA, March 19-21, *Lecture Notes in Computer Science*, Vol 4948, pp 465-481 (2008).

[12] G. H. Golub, and C. F. Van Loan, Matrix Computation, Third Edition, The John Hopkins University Press, 1996.

[13] V. Guruswami. List decoding with side information. In *Proceedings of IEEE Conference on Computational Complexity*, p. 300. IEEE Computer Society, 2003.

[14] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby, "A Pseudorandom Generator from any One-way Function," *SIAM J. Comput.* **28**, 1364 (1999).

[15] M. Hayashi, "Upper bounds of eavesdropper's performances in finite-length code with the decoy method," *Physical Review A*, Vol. **76**, 012329 (2007).

[16] M. Hayashi, "Large deviation analysis for quantum security via smoothing of Rényi entropy of order 2," arXiv:1202.0322 (2012); Accepted for *IEEE Trans. Inform. Theory*.

[17] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inform. Theory*, Vol. **57**, No. 6, 3989-4001, (2011).

[18] M. Hayashi, "Tight exponential analysis of universally composable privacy amplification and its applications," *IEEE Trans. Inform. Theory*, vol. **59**, No. 11, 7728 – 7746, 2013.

[19] M. Hayashi, "Security analysis of ε-almost dual universal$_2$ hash functions," arXiv:1309.1596.

[20] M. Hayashi, "Precise evaluation of leaked information with universal$_2$ privacy amplification in the presence of quantum attacker," *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2012)*, Cambridge, MA, USA, July, 1-6, 2012, pp. 890 - 894.

[21] M. Hayashi and R. Matsumoto, "Secure Multiplex Coding with Dependent and Non-Uniform Multiple Messages," arXiv:1202.1332 (2012).

[22] M. Hayashi and R. Nakayama, "Security analysis of the decoy method with the Bennett-Brassard 1984 protocol for finite key lengths," *New J. Phys.* **16**, 063009 (2014).

[23] M. Hayashi and T. Tsurumaru, "Concise and Tight Security Analysis of the Bennett-Brassard 1984 Protocol with Finite Key Lengths," *New J. Phys.* **14**, 093014 (2012).

[24] M. Hayashi and S. Watanabe, "Non-Asymptotic and Asymptotic Analyses on Markov Chains in Several Problems," arXiv:1309.7528 (2013).

[25] J. Justesen and T. Hoholdt, *Course In Error Correcting Codes*, European Mathematical Society (2004).

[26] A. Mahalanobis "The discrete logarithm problem in the group of non-singular circulant matrices," *Groups Complexity Cryptology*, vol.2, pp.83-89, (2010).

[27] R. Matsumoto and M. Hayashi, "Universal Strongly Secure Network Coding with Dependent and Non-Uniform Messages," arXiv:1111.4174 (2011).

[28] C. A. Miller, and Y. Shi, "Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices," arXiv:1402.0489.

[29] J. Naor and M. Naor, "Small-bias probability spaces: Efficient constructions and applications," *SIAM Journal on Computing*, 22(4):838-856, 1993.

[30] R. Renner, "Security of Quantum Key Distribution," PhD thesis, Dipl. Phys. ETH, Switzerland, 2005; arXiv:quantph/0512258.

[31] R. Renner, and R. König, "Universally composable privacy amplification against quantum adversaries," Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, J.Kilian (ed.) *Lecture Notes in Computer Science*, vol. 3378, pp. 407-425, (2005).

[32] M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, A. Tanaka, K. Yoshino, Y. Nambu, S. Takahashi, A. Tajima, A. Tomita, T. Domeki, T. Hasegawa, Y. Sakai, H. Kobayashi, T. Asai, K. Shimizu, T. Tokura, T. Tsurumaru, M. Matsui, T. Honjo, K. Tamaki, H. Takesue, Y. Tokura, J. F. Dynes, A. R. Dixon, A. W. Sharpe, Z. L. Yuan, A. J. Shields, S. Uchikoga, M. Legre, S. Robyr, P. Trinkler, L. Monat, J.-B. Page, G. Ribordy, A. Poppe, A. Allacher, O. Maurhart, T. Langer, M. Peev, and A. Zeilinger, "Field test of quantum key distribution in the Tokyo QKD Network," *Optics Express*, Vol. 19, Issue. 11, pp. 10387-10409 (2011).

[33] Y. Shi, private communication 2014.

[34] V. Shoup, *A Computational Introduction to Number Theory and Algebra, 2nd Ed.*, (Cambridge University Press, 2009).

[35] J. H. Silverman, "Rings of Low Multiplicative Complexity," *Finite Fields and Their Applications* **6**, 175-191 (2000).

[36] J. H. Silverman, *A Friendly Introduction to Number Theory, Third Edition*, (Pearson Education Inc., 2006).
[37] D. R. Stinson. "Universal hash families and the leftover hash lemma, and applications to cryptography and computing," *J. Combin. Math. Combin. Comput.* **42**, pp.3-31 (2002).
[38] K. Tamaki, M. Curty, G. Kato, H.-K. Lo, and K. Azuma, "Loss-tolerant quantum cryptography with imperfect sources," arXiv:1312.3514 [quant-ph].
[39] M. Tomamichel and M. Hayashi, "Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks," *IEEE Trans. Inform. Theory*, vol. **59**, No. 11, 7693-7710 (2013).
[40] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, "Tight Finite-Key Analysis for Quantum Cryptography" *Nat. Commun.* **3**, 634 (2012)
[41] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, "Leftover Hashing Against Quantum Side Information," *IEEE Trans. Inform. Theory*, vol. 57, No. 8, 5524-5535 (2011).
[42] L. Trevisan, "Extractors and pseudorandom generators," *J. ACM*, **48**, pp. 860-879 (2001).
[43] T. Tsurumaru and M. Hayashi, "Dual Universality of Hash Functions and Its Applications to Quantum Cryptography," *IEEE Trans. Inform. Theory*, vol. **59**, No. 7, 4700–4717 (2013).
[44] S. Watanabe and M. Hayashi, "Non-asymptotic analysis of privacy amplification via Rényi entropy and inf-spectral entropy," *Proceedings of the 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, 2013, pp. 2715-2719.
[45] M. N. Wegman and J. L. Carter, "New Hash Functions and Their Use in Authentication and Set Inequality," *J. Comput. System Sci.* **22**, 265–279 (1981).
[46] Ran Raz, "Extractors with weak random seeds," In *Proceedings of the 37th symposium on Theory of computing, STOC05*, pages 11-20. ACM, 2005.
[47] M. Tomamichel, private communication (2014).
[48] M. Tomamichel, "A Framework for Non-Asymptotic Quantum Information Theory," PhD thesis, Dipl. Phys. ETH, Switzerland, 2005;
[49] R. König, R. Renner, and C. Schaffner, "The Operational Meaning of Min-and Max-Entropy," *IEEE Trans. Inform. Theory*, vol. **55**, no. 9, 4337-4347, (2009).