(一) 研究計畫之背景

I. 萃取器 (Extractors)

隨機性(Randomness)已被證實在資訊科學(computer science)的許多領域都是非常有用的。舉個例子來說，在密碼學上，我們常使用隨機性來隱藏秘密，以避免在傳輸的過程中被第三者竊聽而洩露資訊。此外，在許多情況下，使用隨機演算法(probabilistic algorithm)比我們已知的傳統決定性演算法(deterministic algorithm)在時間以及空間的複雜度(time and space complexity)上來的更有效率[MR95, Gol98]。而這些隨機演算法都需要使用所謂的"真正的隨機元(truly random bits)"。不幸地，在真實的世界中並不存在真正的隨機元，因此，我們只能退而求其次地使用一個決定性的函數(deterministic function)從一些可取樣的弱隨機源(weak random sources)中萃取出真正的隨機元，其中，弱隨機源所能保證的事只有任一個字串出現的機率都不會非常高。如果我們說一個弱隨機源的(statistical) min-entropy 為 k 則表示每個字串出現的機率都不會超過 2^{-k}，直觀地我們認為這樣的弱來源"包含" k 個隨機元。

1988 年時，Chor 等人證明不存在一個決定性的函數可以從單一個(statistical) min-entropy <n 的弱隨機源中萃取出一個隨機元[CG88]。於是，研究者們試圖在單一個弱隨機源外，再加上一個很短(相較於弱隨機源的長度)的真正隨機元，以萃取出隨機元。我們稱這個很短的真正隨機元為種子(seed)，而這種萃取器為種子萃取器(seeded extractor)。在近幾十年來，國內外的理論學家們都致力於建造出使用最少種子，而能從各種弱隨機源中萃取出非常靠近真正隨機元的種子萃取器，如[ILL89、Zuc97、RSW00、RVW00、TSZS01、TSUZ01]等。最後終於造出了幾近完美的種子萃取器[LRVW03]。

然而在使用種子萃取器時，我們仍然需要種子，其為一些真正的隨機元。在一些應用中，我們可以解決此問題(比如在解隨機 BPP 問題中列舉所有可能的種子)，而在其他應用中，則又回歸到最初的問題: 如何獲得一些真正的隨機元呢? 這個爭議讓學者們轉而建造不需種子輔助的萃取器，即決定性萃取器(deterministic extractors or seedless extractors)。

當弱隨機源擁有某些特殊的性質時，我們確實可以從一個隨機源中萃取隨機元。一個(n, k)-固定某些位元的來源 (A (n,k)-bit fixing source) 為一個在 $\{0,1\}^n$ 的分布 $X=(X_1,\cdots,X_n)$，其中 n-k 個位元是固定的，而其餘的 k 個位元則是均等的(uniform)且彼此獨立的(independent)。Kamp 等人[KZ03]提出從一個固定某些位元的來源中萃取出一些隨機元的方法，而 Gabizon 等人則在 2004 年提出改進的方法使之萃取出幾乎所有隨機源所含的隨機性[GRS04]。

此外，我們亦可從兩個甚至多個弱隨機源中不使用種子而直接萃取出很靠近均等分布(uniform distribution)的隨機元。針對兩個弱隨機源，研究者們致力於放寬對此兩個隨機源的(statistical) min-entropy 的限制[CG88、DO03、DEOR04、LLTT05、Raz05]，最後，Bourgain 造出只要兩個(statistical) min-entropy 均略小於 n/2 的弱隨機源即可萃取出 $\Omega(n)$ 個隨機元的萃取器[Bou05]。而另一方面，理論學家們也設法從越少個且含有越少(statistical) min-entropy 的隨機源中萃取出隨機元[BIW04、BKS+05、Raz05]。

事實上，固定某些位元的來源以及多個隨機源的來源可以下列觀點看成是兩種極端。這兩種來源均包含多個部分而且這些部份彼此都是獨立的。固定某些位元的來源可看成是包含許多個部份，且每個部分只有單一個隨機的或者為固定的位元。而多個隨機源的來源則可看成包含相對少數個部份，可是每個部份為多個位元且含足夠數量的隨機元。我們考慮一個介於其中的來源，稱為獨立符

號的來源(independent-symbol source)，寫成 (n,D,k)-來源，其為一個在[D]^n([D]={1,2,...,N})上的分布 $X=(X_1,\cdots,X_n)$，其中 $X_1,\cdots,X_n$ 是彼此獨立的，且 X 的 min-entropy 是 k。 我們不難看出一個(n, k)-固定某些位元的來源其實就是一個(n,2,k)-來源。換句話說，固定某些位元的來源只是我們所考慮的獨立符號來源的一個特例。而對較小的 n 及較大的 D，此種來源即可涵蓋多個隨機源的來源。Kamp 等人[KRVZ06]及 Lee 等人[LLT06]均提出從一個獨立符號來源中萃取出幾乎所有隨機性的萃取器。

II. Computational min-entropy

在過去的文獻中，大多是考慮那些在統計上仍有隨機性(亦即 statistical min-entropy 不為 0)的隨機源。在本計劃中，我們將換個角度，考慮那些在統計上並不具有任何的隨機性，但在那些計算複雜度有所限制的觀察者而言仍有些許隨機性的隨機源。換句話說，我們將從傳統的統計觀點，轉換為計算觀點。由於這些隨機源在統計上本就不具任何的隨機元，因此我們也必須將萃取器的輸出限制稍作修改。相對於在原本的定義中，我們要求萃取器的輸出要在統計上(亦即不限制計算複雜度的觀察者眼中)極靠近真正的隨機元，現在我們只要求這些輸出在那些計算複雜度有限的觀察者眼中看似隨機即可。值得注意的是，這些輸出有可能在統計上是距離真正隨機元極遠的，甚或是根本不具任何隨機性的。事實上，在密碼學中，尤其是有關利用 one-way functions 建造 pseudo-random generators 的研究中，已經有一些隱含的結果存在[Yao82, GL89, HILL99]。

何謂"看似隨機"呢? 事實上，已經有一個非常好的"看似幾乎隨機"的定義[Yao82]，但關於一個隨機源"看來有些隨機"的定義以及衡量此隨機源中隨機性之測量值的定義仍不清楚。目前有一些看似合理的定義，但目前已有些證據可以證明這些定義之間有不一致之處[BSW03, HLR07]。

在本計劃中，我們考慮採用[HLR07]中的定義來計算一個隨機源在計算上的隨機量。我們在這裡所考慮的來源是一個條件的型式(f(X)|X)，其中 X 是一個在 $\{0,1\}^{n_1}$ 的分布，而 $f:\{0,1\}^{n_1}\to\{0,1\}^n$ 為某個函數。假設當輸入是依照分布 X 所產生時，任何大小為 2^k 的電路最多只有 2^{-k} 的機率可以猜對 f 的函數值時，則我們說這種條件分布(f(X)|X)擁有 computational min-entropy k。當 f 是一個決定性的函數時，亦即 f(x)的值完全由 x 所決定，則當給定 x 之後，f(x)並不具有任何在統計上的隨機性。然而，依照我們的定義，只要函數 f 是很難計算的，則(f(X)|X)還是可以擁有很大的 computational min-entropy，使得我們可以從中萃取隨機元。更準確地說，從一個分佈 f(X)中，我們希望能萃取出一些隨機性，使得即使在給定 X 時，這些隨機性在某些大小的電路看來幾乎是非常隨機的。

就如同於統計的架構下的結果，我們依然無法從單一個長度為 n，computational min-entropy 為 n-2 的隨機源中萃取出隨機元，即使我們僅希望萃取出一個隨機元[LLT09]。然而，[LLTT05]的萃取器可從一個擁有 computational min-entropy 為 $k_1=n-k+O(k/\log k)$ 的隨機源與另一個擁有 statistical min-entropy 為 k 的隨機源中萃取出隨機元[LLT09]。事實上， [LLTT05]的萃取器亦可從一個擁有 computational min-entropy 為 $k_1=n-k+O(k/\log k)$ 的隨機源與另一個擁有 computational min-entropy 為 k 的隨機源中萃取出隨機元[LLT09]。

（二）研究目的

一、考慮在計算角度下的獨立符號來源

在本計劃中，我們考慮計算的獨立符號來源(computational independent-symbol source)。如同獨立符號來源，每個計算的(n, D, k, s)-來源 (computational (n, D, k, s)-source)$(V|X)=(V_1|X_1)\circ\cdots\circ(V_n|X_n)$，包含 n 個彼此獨立的部分 $(V_1|X_1),\cdots,(V_n|X_n)$，其中每個$V_i$都是分布在[D]={1, 2, …, D}中，而$X_i$則是分布在$\{0,1\}^{\ell_i}$上，且對每一個$i\in\{1,2,\cdots,n\}$，和每一個大小為 s 的電路 C，$\Pr[C(X_i)=V_i]\le 2^{-k_i}$對某個值 $k_i\le\log D$，並滿足$\sum_{i=1}^{n}k_i=k$。我們將試著證明在統計觀點中針對獨立符號來源的萃取器亦可從此種計算的獨立符號來源中萃取出隨機元。

（三）研究方法與結果

A. 延伸有名的核心集引理(hardcore set lemma)

Impagliazzo 的核心集引理［Imp95］告訴我們，如果一個布林函數$f:\{0,1\}^{\ell}\to\{0,1\}$，其滿足對每個大小為 s 的電路 h，$\Pr_{x\in\{0,1\}^{\ell}}[h(x)\neq f(x)]>\delta$，則對每個ε>0，存在一個大小至少為 $\delta 2^{\ell}$ 的 hardcore set $H\subseteq\{0,1\}^{\ell}$ 使得對任何大小為 $\Omega(\varepsilon^2\delta^2 s)$ 的電路 C，$\Pr_{x\in H}[C(x)\neq f(x)]>\frac{1}{2}-\varepsilon$。我們延伸此結果考慮一個函數$f:\{0,1\}^{\ell}\to[D]$，其滿足對每個大小為 s 的電路 h，$\Pr_{x\in\{0,1\}^{\ell}}[h(x)\neq f(x)]>\delta$，並證明此時依然存在一群總大小至少為$(\delta/2)2^{\ell}$的二元核心集(binary hardcore sets) $T_1,\cdots,T_r$，亦即對每一個$i\in[r]$，存在一個大小為 2 的子集合$I_i\subseteq[D]$使得$T_i\subseteq f^{-1}(I_i)$，且對每個大小為$\Omega(\varepsilon^2\delta^2 s/D^6)$的電路 C，$\Pr_{x\in T_i}[C(x)\neq f(x)]>\frac{1}{2}-\varepsilon$。

B. 建造出針對計算獨立符號來源的萃取器

我們首先利用[STV01]的方法以及延伸的核心集引理證明對每一個來源$(V_i|X_i)$都存在一

個來源Y_i，使得任何小電路都無法分辨(X_i,V_i)以及(X_i,Y_i)，且來源$(Y_i|X_i)$的 statistical min-entropy 與$(V_i|X_i)$的 computational min-entropy 有關。所以給定一個計算的(n, D, k, s)-來源$(V|X)=(V_1|X_1)\circ\cdots\circ(V_n|X_n)$，其中$k\geq\Omega(2^{2m}\log^2 D)$，對任何大小為$\Omega\left(s(\log n/nkD)^2\right)$的電路 C，

$$\left|\Pr\left[C(X_1,\cdots,X_n,V_1,\cdots,V_n)=1\right]-\Pr\left[C(X_1,\cdots,X_n,Y_1,\cdots,Y_n)=1\right]\right|\leq\frac{2^{2m}\log n}{k}\ 。$$

進一步地，我們可以證明對任何大小為$\Omega\left(s(\log n/nkD)^2\right)$的電路 C，

$$\left|\Pr\left[C\left(X_1,\cdots,X_n,\sum_{i=1}^{n}V_i\right)=1\right]-\Pr\left[C\left(X_1,\cdots,X_n,\sum_{i=1}^{n}Y_i\right)=1\right]\right|\leq\frac{2^{2m}\log n}{k}\ 。$$

另一方面，有$1-e^{-\Omega(k/\log D)}$的機率，即使在知道$X_1,\cdots,X_n$後，來源$Y=Y_1\circ\cdots\circ Y_n$的 min-entropy 至少也有$\Omega(k/\log D)$。再加上$(V_1|X_1),\cdots,(V_n|X_n)$是彼此獨立的，我們可推得$Y_1,\cdots,Y_n$也是彼此獨立的。因此，我們可以利用[LLT06]中，

$$Ext(V_1,\cdots,V_n)=\sum_{i=1}^{n}V_i$$

是針對獨立符號來源的萃取器的結果，證明

$$\Delta\left(\left(X_1,\cdots,X_n,\sum_{i=1}^{n}Y_i\right),(X_1,\cdots,X_n,U_m)\right)\leq e^{-\Omega\left(k/2^{2m}\log D\right)}\ 。$$

結合以上的結果，我們推得對任何大小為$\Omega\left(s(\log n/nkD)^2\right)$的電路 C，

$$\left|\Pr\left[C\left(X_1,\cdots,X_n,\sum_{i=1}^{n}V_i\right)=1\right]-\Pr\left[C(X_1,\cdots,X_n,U_m)=1\right]\right|\leq O\left(\frac{2^{2m}\log n}{k}\right)\ ,$$

亦即$Ext(V_1,\cdots,V_n)=\sum_{i=1}^{n}V_i$為針對計算獨立符號來源的萃取器。

C. 延伸著名的 XOR 引理

我們發現上述針對計算獨立符號來源的萃取器的結果其實可以推得延伸的 XOR 引理。Yao 的 XOR 引理［Yao82］告訴我們，如果一個布林函數 $f:\{0,1\}^{\ell}\to\{0,1\}$，其滿足對每個大小為 s 的電路 h，$\Pr_{x\in\{0,1\}^{\ell}}\left[h(x)\neq f(x)\right]>\delta$，則對 $\varepsilon>2(1-\delta)^{t}$，任何大小為 $\Omega\left(\varepsilon^{2}s/\ell\right)$ 的電路 C，

$$\Pr_{x_1,\cdots,x_t\in\{0,1\}^{\ell}}\left[C(x_1,\cdots,x_t)=\sum_i f(x_i)\right]<\frac{1}{2}+\varepsilon \text{ 。}$$

我們延伸此結果考慮 n 個函數 $f_1,\cdots,f_n$，其中每個函數 $f_i:\{0,1\}^{\ell_i}\to[D]$，其滿足對每個大小為 s 的電路 h，$\Pr_{x_i\in\{0,1\}^{\ell_i}}\left[h(x_i)\neq f_i(x_i)\right]>\delta_i$，並證明如果 $\delta=\sum_{i=1}^{n}\delta_i\geq\Omega\left(M^{2}\log D\right)$，則對每個大小為 $\Omega\left(s\left(\log n/nD\delta\right)^{2}\right)$ 的電路 C，

$$\Pr_{x_1\in\{0,1\}^{\ell_1},\cdots,x_n\in\{0,1\}^{\ell_n}}\left[C(x_1,\cdots,x_n)=\sum_{i=1}^{n}f_i(x_i)\right]\geq\frac{1}{M}+\frac{M^{2}\log n}{\delta} \text{ 。}$$

D. 黑箱子建造法中二元核心集大小的上限

在延伸的核心集引理中，我們證明存在若干個二元核心集，其大小總和至少為 $(\delta/2)2^{\ell}$。大家可能會猜想是否存在單一個夠大的核心集，比如說大小為 $(\delta/D)2^{\ell}$。我們最後證明一個對任何黑箱子建造法的二元核心集大小的上限。假如一個演算法 $\mathrm{Dec}^{(\cdot)}$ 滿足對任意的函數 $f:\{0,1\}^{\ell}\to[D]$，以及任何的函數集合 $G=\left\{g_I\,\middle|\,I\subseteq[D],|I|=2\right\}$，其中任何在 G 裡面的函數 g_I，以及任何大小為 s 的子集 $H\subseteq f^{-1}(I)$，$\Pr_{x\in H}\left[g_I(x)\neq f(x)\right]\leq(1-\varepsilon)/2$，可推得

$$\Pr_{x\in\{0,1\}^{\ell}}\left[Dec^{G}(x)\neq f(x)\right]\leq\delta \text{ ，}$$

則我們說演算法 $\mathrm{Dec}^{(\cdot)}$ 是一個核心集的黑箱子 (δ, ε, D)-建造法 (black-box (δ, ε, D)-construction of a hardcore set)，而其中 s 為黑箱子建造法的大小複雜度(size complexity)。

我們最後一個結果即利用機率方法證明當 $\delta\geq\Omega(1/D^{c})$ 對某個常數 c ，$D\geq4$，以及 $\varepsilon<1/5$ 時，任何核心集的黑箱子(δ, ε, D)-建造法的大小複雜度為 $O\left(\delta 2^{\ell}/D^{2}\right)$。

計畫成果自評

一、我們延伸核心集引理從原本只考慮函數 f 為布林函數的情況到考慮 $f:\{0,1\}^{\ell}\to[D]$。

二、我們利用延伸的核心集引理證明[LLT06]中針對獨立符號來源的萃取器亦可從計算的獨立符號來源中萃取出隨機元，其在小電路的眼中看起來是非常隨機的。

三、我們亦利用證明針對計算獨立符號來源之萃取器的方法證明延伸的 XOR 引理，其將原本只考慮函數 f 為布林函數的情況延伸到考慮 $f:\{0,1\}^{\ell}\to[D]$。

四、最後我們證明一個核心集的黑箱子建造法中二元核心集大小的上限。

參考文獻

[BIW04] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting Ranodmness from Few Independent Sources. FOCS 2004.

[BKS+05] B.Barak, G. Kindler, R. Shaltiel, B. Sudakov, A. Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers and Extractors. In *Proc. 37th STOC*. ACM, 2005.

[BSW03] B. Barak, R. Shaltiel, and A. Wigderson. Computational Analogues of Entropy. In RANDOM-APPROX 2003, 200-215.

[Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. International Journal of Number Theory, 1:1-32, 2005.

[CG88] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. SIAM Journal on Computing. 17(2):230-261, 1988.

[DO03] Y. Dodis, R. Oliveira. On Extracting Private Randomness over a Public Channel. RANDOM-APPROX 2003, 252-263.

[DEOR04] Y. Dodis, A. Elbaz, R. Oliveira, R. Raz. Improved Randomness Extraction from Two Independent Sources. RANDOM-APPROX 2004.

[GL89] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In Proc. 21st Annual ACM Symposium on Theory of Computing (STOC'89),1989.

[Gol98] O. Goldreich. Modern Cryptography, Probabilistic Proofs and Pseudorandomness. Springer-Verlag, Algorithms and Combinatorics, 1998.

[GRS04] A. Gabizon, R. Raz and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. In Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, 2004.

[HILL99] J. Hastad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. SIAM J. Comput., 28(4):1364--1396, 1999.

[HLR07] C. Y. Hsiao, C. J. Lu, and L. Reyzin. Conditional computational entropy, or toward separating pseudoentropy form compressibility. In Proc. Advances in Cryptology-EUROCRYPT07, 2007.

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudorandom generation from one-way functions. In Proceedings of the 21st ACM Symposium on Theory of Computing, 1989.

[Imp95] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In Proc. 36th IEEE Symposium on Foundations of Computer Science(FOCS'95), 1995.

[KRVZ06] J. Kamp, A. Rao, S. Vadhan, and D. Zuckerman. Deterministic Extractors for Small-Space Sources. In Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06), pages 691-700, May 2006.

[KZ03] J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In Proceedings of the 44th Annual IEEE

Symposium on Foundations of Computer Science, 2003.

[LLTT05] C. J. Lee, C. J. Lu, S. C. Tsai, and W. G. Tzeng. Extracting randomness from n independent weak random sources. IEEE Transactions on Information Theory (SCI), 51(6) 2224-2227, 2005.

[LLT06] C. J. Lee, C. J. Lu, and S. C. Tsai, Deterministic extractors for independent-symbol sources. In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), pages 84-95, 2006.

[LLT09] C.J. Lee, C.J. Lu, S.C. Tsai. Extracting Computational Entropy and Learning Noisy Linear Functions. COCOON 2009.

[LRVW03] C. J. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: Optimal up to Constant Factors. In Proceedings of the 35th ACM Symposium on Theory of Computing, pages 601-611, 2003.

[MR95] R. Motwani and P. Raghavan. Randomized algorithms. Cambridge University press, 1995.

[Raz05] R. Raz. Extractors with Weak Random Seeds. Proceeding of the 37th STOC, 2005, pp. 11-20.

[RSW00] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. In Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, 2000.

[RVW00] O. Reigold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, 2000.

[STV01] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the XOR lemma. Journal of Computer and System Sciences, 62(2): 236-266, 2001.

[TSUZ01] A. Ta-Shma, C. Umans, and D. Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, 2001.

[TSZS01] A. Ta-Shma, D. Zucherman and S. Safra. Extractors from Reed-Muller codes. In Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science, 2001.

[Yao82] A. Yao. Theory and applications of trapdoor functions. In Proc. 23rd Annual Symposium on Foundations of Computer Science (FOCS'82), 1982.

[Zuc97] D. Zuckerman. Randomness-optimal oblivious sampling. Random Structures and Algorithms, 11:345-367, 1997.