Research Statement

Xin Li

My past research has been focusing on theoretical computer science, specifically, the use of randomness in computation. Consider the following two kinds of fundamental questions: (1) How to generate high quality random bits from low quality random sources, and (2) How to use randomness to prevent certain adversarial attacks, in the information theoretic setting. The first question is important because an enormous amount of applications in computer science require strictly uniform random bits (especially those in cryptography), while in practice random sources almost always have severe biases, and can leak information to an adversary. Direct use of such random sources may lead to catastrophic situations such as the total loss of security. The second question is important because it provides us the highest level of security, which does not depend on any unproven assumptions. Both questions have been studied extensively for decades, and my work in both areas has led to significant progress and major breakthroughs to long standing open problems.

A specific example, and a central problem of the first question, is the construction of randomness extractors. A randomness extractor is a function that takes as input one or more biased random sources, and outputs an almost uniform probability distribution. Among many interesting models of biased sources, the most important and well studied is independent biased sources, due to its practical use and connections to Ramsey graphs. This was first studied by Chor and Goldreich [2] in 1985, who showed the existence of an extractor for two independent sources of entropy $O(\log n)$, where n is the length of each source. Constructing extractors for a small number of independent sources with small entropy has thus been a major open problem, and one which turns out to be quite difficult. Indeed the progress has been very slow, as until 2013, even with 100 sources we only know how to build extractors for entropy n^{α}, where α is some constant.

A specific example, and a central problem of the second question, is the following privacy amplification problem: two parties want to communicate with each other via a channel to transform a shared private biased random source into shared private uniform random bits, in the presence of an active adversary. In this case, we want to achieve some kind of security guarantee even if the adversary can tamper with the communication protocol. This problem is central because it serves as the basis for several other important cryptographic tasks. First studied by Maurer and Wolf [11] in 1997, the problem has been actively pursued by a lot of researchers, where the important goal is to simultaneously minimize the number of interactions and the entropy loss of the protocol. It is also quite challenging as witnessed by the fact that until 2011, all known protocols either need a large number of interactions or a big entropy loss.

Following the introduction of "non-malleable extractors" by Dodis and Wichs [4], and the first explicit construction of such extractors in one of my joint work with Dodis et. al [3], my work has led to a series of breakthroughs in both problems, with new techniques developed and new connections established between seemingly unrelated areas. Specifically, my work around 2013 [5, 7, 6] established new connections between the problem of constructing randomness extractors and other problems in distributed computing/cryptography. This allows the use of powerful tools from distributed computing/cryptography (including the aforementioned non-malleable extractors) in this problem, and thus pointed out a new way to attack this problem and laid the foundations for an ambitious plan to construct near optimal extractors. To date, through the efforts of myself and other researchers, this plan has been quite successful, leading to numerous new results in the extractor literature, and accumulating in the three-source extractor with exponentially small error for entropy $\log^c n$ by myself [8] (such small error is necessary for cryptographic applications), and the first two-source extractor with polynomially small error for entropy $\log^c n$ by Chattopadhyay and Zuckerman [1], where c is some constant. Their result was subsequently improved by myself in several aspects, and currently my work [10, 9] has the best known construction of two-source extractors, which works

for entropy $\widetilde{O}(\log n)$ and gives the best known explicit construction of Ramsey graphs. These techniques also extend to other models such as affine sources, small space sources and circuit sources.

In turn, the techniques developed in my work have also led to my construction of almost optimal non-malleable extractors and privacy amplification protocols, which simultaneously achieve the optimal number of interactions and the optimal entropy loss [9]. Further, these techniques have been generalized by other researchers to several different settings, such as post-application security and bounded storage model.

In the immediate future, I plan to extend these work and look at several related questions. First of all, now that we have a relatively good understanding of the model of independent sources, it is natural to ask for more general models. We thus ask the following question: what is the most general model of biased sources that allows us to construct randomness extractors? The goal is to both explore new models and to unify existing models in a much larger framework. Second, randomness extractors, as pseudorandom objects, are naturally related to other important objects, such as pseudorandom generators, error correcting codes, hard functions and many more. Can we use our techniques to get improved constructions of these related objects? This direction may lead to new connections and progress in other long-standing open problems. Finally, the area of information-theoretic and tamper-resilient cryptography is an active area that is developing rapidly, and I hope to extend current techniques to other situations where we can offer protection against various tampering adversary. My recent work of constructing a generalization of error correcting codes known as non-malleable codes is one such example.

References

[1] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016.

[2] 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.

[3] Y. Dodis, X. Li, T. D. Wooley, and D. Zuckerman. Privacy amplification and non-malleable extractors via character sums. *SIAM Journal on Computing*, 43(2):800–830, 2014.

[4] Y. Dodis and D. Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 601–610, 2009.

[5] X. Li. Non-malleable extractors, two-source extractors and privacy amplification. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 688–697, 2012.

[6] X. Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 100–109, 2013.

[7] X. Li. New independent source extractors with exponential improvement. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 783–792, 2013.

[8] X. Li. Three source extractors for polylogarithmic min-entropy. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

[9] X. Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. Technical Report TR16-115, ECCC, 2016.

[10] X. Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[11] U. M. Maurer and S. Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology — CRYPTO '97, 17th Annual International Cryptology Conference, Proceedings*, 1997.