COMS W6261: Advanced Cryptography
Spring 2020: Information Theoretic Cryptography
Instructor: Tal Malkin

Information Theoretic Cryptography: Lectures, Readings, Proposed Project Topics

Last updated: 2/9/20

This is an evolving document, summarizing lecture topics and readings. In addition to papers presented in class, we include an annotated list of other optional readings and point out some open problems. The goal is to give an idea of the state of the art, and for people who may want to study a certain area in more depth (eg towards their project). The content is biased towards lectures that already happened (hence the document is evolving), as well as towards the interests of the instructor and students in the class (if you're a student and would like more refined resources on a specific topic, let us know!).

This is meant as an informal document to help students in the class, and not as an exhaustive survey. Omissions and mistakes are possible (please let us know if you find mistakes! no monetary compensation, but we will be grateful.)

1 Some General Resources

- BIU Winter School On Information Theoretic Cryptography. This is forthcoming (February 17-20, 2020), and should have the materials (videos, slides, lecture notes) available soon after completion.
- Ronald Cramer, Ivan Damgård, Jesper Buus Nielsen, Secure Multiparty Computation and Secret Sharing. This is a textbook focusing on the information theoretic aspects.
- Vinod Vaikuntanathan, Some Open Problems in Information Theoretic Cryptography, FSTTCS 2017.

2 Lecture 1: Secret Sharing

Lecture Summary

After an introduction to the class, we defined threshold secret sharing, showed Shamir's secret sharing scheme, and proved its security. We also provided an informal overview of how this is used for secure computation (BGW protocol), which we will see next week.

Additional Directions and Resources for Secret Sharing

One could spend a whole semester on the area secret sharing (this is true also for the other areas we will talk about this semester). In class we focused only on (perfect) secrecy, and on an upper bound for a threshold access structure.

Secret sharing can also be defined for *general (monotone) access structures:* a minimal set of allowed subsets, so that any set of players that contains an allowed subset can reconstruct the secret, while any other set of players gains no information about the secret. Some common access structures that have been considered, in addition to threshold, include monotone formulae and graph-based access structures (as well as arbitrary monotone structures).

One could also consider secret sharing in the case of *malicious parties*. For example, if some of the parties send wrong shares for reconstruction, can the secret still be recovered? (this is called "robustness"). For Shamir's $t+1$-out-of-n secret sharing scheme, in the case that $t < n/3$, the answer is yes. Specifically, given n shares (which are supposed to be points on a polynomial of degree t), the polynomial is unique and can be reconstructed, even if up to t of the points do not lie on the polynomial. This is due to Berlekamp-Welch, and boils down to the fact that Shamir's secret sharing is actually an error correcting code (it is the Reed-Solomon error correcting code). In fact, there's a strong general connection between secret sharing and error correction (although in the ECC literature secrecy is not required).

Another type of malicious behavior in the context of secret sharing, is when the *dealer* is malicious, providing shares that were not generated correctly, and cannot be used to reconstruct any secret (in Shamir's scheme, the shares may not be on a degree t polynomial). This comes up in applications of secret sharing, e.g. for secure computation, if the parties are malicious. This is addressed by the stronger notion of *verifiable secret sharing (VSS)*.

Much literature considered upper and lower bounds on the size of the shares for secret sharing. Note that Shamir's (threshold) secret sharing has share length $\log n$ (for secrets of size at most $\log n$), as we are working in a field $\mathbb{F}$ of size $|\mathbb{F}| > n$, and the share (as well as the secret) is a field element.[1] Moreover, as discussed in class, the reconstruction of the shares (polynomial interpolation) is a *linear* operation (which can be computed by multiplying the vector of shares by an appropriate matrix). Such schemes are called *linear secret sharing* schemes.

The best upper bounds on the share size in secret sharing with general access structures (an arbitrary monotone function) is $O(2^n)$. This is known to be optimal for linear secret sharing. However, for general (non-linear) secret sharing, the only lower bound is $\Omega(n/\log n)$.

This exponential gap between best known upper and lower bounds is a recurring theme in information-theoretic cryptography, and reducing this gap is an important open problem in various domains (see Vaikuntanathan's short overview of these open problems from 2017).

Secret sharing scheme are tightly connected to several other topics we mention in this class, including PSM, CDS, PIR, and secure computation more generally.

Supplementary Readings:

Here are some classic results and a few papers demonstrating connections of secret sharing to other topics we are studying,[2] as well as examples of very recent research.

- Amos Beimel, Secret Sharing Schemes: A Survey, IWCC 2011
- Mitsuru Ito, Akira Saito, Takao Nishizeki, Secret Sharing Scheme Realizing General Access Structure, 1987. A $O(2^n)$ upper bound on the size of the share for general access structures.

[1] For every n there's a prime p such that $n < p \leq 2n$, so we can work in a finite field where each element can be specified with $\log n + 1$ bits.

[2] Some papers about these connections appear here, and some under the other topics – the division is fairly arbitrary (and this is only a sample of relevant papers).

- Laszlo Csirmaz, The Size of A Share Must Be Large, 1989. A $\Omega(n/\log n)$ lower bound on the size of the share for general access structures.
- Amos Beimel, Yuval Ishai, On the Power of Nonlinear Secret Sharing, CCC 2001. Section 3 in this paper shows a perfect PSM/DARE for quadratic residuosity, and how it can be used to get a nonlinear secret sharing scheme (where an efficient linear secret sharing scheme is not known to exist, as this function is conjectured not to be in NC).
- Tianren Liu, Vinod Vaikuntanathan, Breaking the Circuit-Size Barrier in Secret Sharing, STOC 2018. This result goes through the connection to CDS.
- Kasper Green Larsen, Mark Simkin Towards an Exponential Lower Bound for Secret Sharing, 2019.
- Applebaum, Beimel, Nir, Peter, Better Secret-Sharing via Robust Conditional Disclosure of Secrets, 2020.

Additional Project Directions

There are various open problems in all the above, both for upper and lower bounds (let me know if you'd like more concrete suggestions, I may have some!).

Additional project directions related to secret sharing include studying *visual cryptography.* Here is the original Naor and Shamir Eurocrypt 1994 paper, and here is a recent Bogdanov, Mande, Thaler, Williamson RANDOM-APPROX 2019 paper on the complexity of reconstructing this AND function (and its connection to approximate degree).

In another direction, a recent line of work studies *non-malleable secret sharing* (this is related to the non-malleable codes topic, as well as non-malleable extractors).

More elaboration on these (and additional resources) available upon request.

3 Lecture 2-3: BGW and Information-Theoretic Secure Computation

Lecture Summary

We defined secure multiparty computation (MPC) in the semi-honest (aka honest-but-curious or passive), perfect correctness and perfect privacy setting. For this setting, we showed the Ben-Or, Goldwasser, Wigderson (BGW) protocol, which is secure when there is an honest majority ($t < n/2$).

We mentioned that the "honest majority" requirement is necessary in general, as there are functions that cannot be computed with perfect correctness and privacy when $t = n/2$. For example, the AND function for two parties[3] (each party holds a bit, and one of the parties should output the logical AND of the bits) cannot be computed with perfect correctness and privacy (even in the semi-honest case). The proof was left as an exercise.

The rest of the discussion (summarized below) was at a higher level, without definitions/proofs/details.

[3] Note that in the two party case (2PC), it doesn't make sense to consider "honest majority", as even one party who just looks at its own transcript already gives $n = 2, t = 1$, namely $t = n/2$.

We mentioned that the BGW protocol has a version that is secure against *malicious* adversaries, when $t < n/3$. Again this threshold is tight (for the information-theoretic setting), and this stems from impossibility of byzantine agreement/broadcast (which is required, and can be viewed as a special case of secure computation). However, if the parties are provided a broadcast channel, general secure computation (with statistical security) can be achieved with $t < n/2$ even for malicious adversaries.

We defined the two party (1 out of 2) oblivious transfer (OT) functionality, and mentioned that this cannot be computed securely information-theoretically, but can be computed securely in the computational setting, under some computational assumptions (eTDP, which follows from standard assumptions like hardness of factoring and others). It can also be computed securely (information-theoretically) in a "correlated randomness" model, where the parties start off with some joint randomness (which is independent of the inputs).

In turn, OT can be used to achieve secure computation against higher corruption thresholds (both in the semi-honest and malicious cases). In particular, if we consider the "OT-hybrid model", where parties have access to an oracle computing the OT functionality, we can achieve secure MPC for $t < n$ (ie when all but one party collude against the last party). As a corollary, we get secure MPC for $t < n$ in the computational model (under cryptographic assumptions), or in the correlated-randomness model.

One way to achieve this is the Goldreich, Micali, Wigderson (GMW) protocol, which also uses secret sharing and gate-by-gate computation of the circuit. They use a Boolean circuit (and additive, n-out-of-n secret sharing), dealing with multiplication (AND) gates by invoking OT (we didn't show how).

These results demonstrate that secure computation is feasible for general functions(!) in many settings. There is a lot of other research on secure computation, in these and other settings, for general tasks or subclasses, from very theoretical to practical implementations. Even defining secure computation is quite challenging in many cases. We did not (and will not) cover these in class.

Instead, going back to the setting we addressed – semi-honest, honest majority, perfect security – what remains to be improved over the BGW protocol?

Round complexity. The BGW protocol has round complexity that is proportional to the multiplicative depth of the circuit (since there's interaction required for every multiplication gate). The GMW protocol for the computational setting is similar in this respect.

Can we achieve constant round complexity? If the function can be represented as a constant degree polynomial, the BGW already gives constant round complexity (the degree of the polynomial). But what about general functions (e.g., AND of n bits, which has degree n?) The answer is yes, as we will discuss in the next topic. Specifically, randomized encodings (RE) give a unified way to describe constant-round protocols (and many other applications!) in various settings. Some of these results were known before RE were defined (e.g., constant round 2PC in the computational/OT-hybrid setting, originally due to Yao), and some were proved using the RE machinery. A recent paper of Benny Applebaum, Zvika Brakerski, and Rotem Tsabary (TCC 2018) achieves (optimal) round complexity 2.

Communication complexity. The communication complexity in the BGW protocol is proportional to the size of the circuit (as well as the number of parties n and the share size), which means it can be exponential in the input size. Can we achieve communication that is polynomial in the input/output size (independent of the circuit size)? This is another example of a problem with an exponential gap between the best upper bounds (which are exponential) and the best lower bounds (which are polynomial). A recent paper that was mentioned in class in this context is the following Ivan Damgård, Kasper Green Larsen, Jesper Buus Nielsen CRYPTO 2019 communication lower bounds paper.

The large gap motivates looking at more restricted, minimalist, models of secure computation such as PSM and its special case CDS, as we will discuss. The results for communication complexity are also related to techniques from PIR literature.

Readings and Open Problems/ Project Directions

To read up on what was presented in class, and MPC more generally, you can use general textbooks, as well as the following (the workshop/school links include also videos and slides):

- Gilad Asharov, Yehuda Lindell, A Full Proof of the BGW Protocol for Perfectly-Secure Multiparty Computation (Journal of Cryptology, 2017). This paper includes also the malicious setting (construction and proof).
- BIU Winter School on Secure Computation and Efficiency (2011) (you can see also the one on Practical Multiparty Computation (2015)).
- Simons Workshop on Securing Computation (summer 2015).

We include readings (and open problems) related to 2PC and round complexity of secure computation with the next section.

A historical note: The BGW protocol came out at the same time (STOC 1988) as another paper by Chaum, Crépeau, and Damgård (CCD), who achieved a similar result (secure MPC, even in malicious setting, when less than a third of the parties are corrupted). We presented BGW because it provides perfect secrecy (while CCD is statistical), and because a full proof of its security is available (see above). The construction for the honest majority case given a broadcast channel is due to Tal Rabin and Michael Ben-Or (STOC 1989)

For open problems and project directions, since the area is so extensive, we will only mention open problems that came up in class, or motivated by student questions/requests.

One question was already mentioned above: reduce the gap between the upper and lower bounds on communication complexity of general MPC (there is more progress and many papers in the PSM / CDS models, see below).

Another question mentioned in class, is classifying which functions can be computed without an honest majority. In the two-party case, this is understood (we mentioned above that AND is not possible; can you think of an example where it is possible?). For more parties (e.g., 3) a characterization is not known, and remains an open problem. [4]

There are several works (with many remaining open problems) addressing secure computation in *incomplete networks*. One such area is *topology hiding computation (THC)*, where the privacy of the graph topology itself should be protected. Here's a recent paper by Ball, Boyle, Cohen, Malkin, Moran (TCC 2019) on information theoretic THC.

[4] The case of deterministic functionalities for two parties was resolved by Eyal Kushilevitz in FOCS'89. Surprisingly, the case of randomized functionalities is still not understood (See Data & Prabhakaran at PKC18). For more than 2-parties the situation is perhaps even more poorly understood. See Agarwal, Anand, and Prabhakaran at Eurocrypt 2019 and references within (particularly Halevi et al. from TCC'18) for recent progress on (a special case of) this question.

4 Lectures 3-4: Randomized Encodings, RE for Circuits (Information-Theoretic Yao), and RE for Branching Programs

Lecture Summary:

The teachers for this lecture (actually, close to two lectures) are **Lalita Devadas** and **Tim Randolph**, with support from **Ruth Wang**.

A summary of the class and various directions and open problems is forthcoming. In the meantime, here is a partial list.

Readings for Class

Required reading:

- Randomization Techniques for Secure Computation, Yuval Ishai, 2013, sections 1 and 2. This is a survey of Randomized Encodings – we will cover some of it (as well as later work) in class.

Papers presented in class:

1. The same Ishai survey above.
2. Perfect Constant-Round Secure Computation via Perfect Randomizing Polynomials, Yuval Ishai, Eyal Kushilevitz, ICALP 2002.

Note: garbled circuits correspond to source 1 sections 4.1.2, 4.2.3, and source 2 section 3, while RE from branching programs correspond to source 1 sections 4.1.3, 4.2.4, and source 2 section 4. These are the technical parts you're responsible for if you want to select one of these papers as the "in-depth" reading requirement,[5]. For this purpose, "one paper" could be one of the above papers, including intro + prelims + all the sections mentioned above. Alternatively, you can read the garbled circuit parts from both papers, or the branching program parts from both papers (the presentations are a bit different between the papers).

Supplementary Reading and Additional Directions

(To be expanded)

Garbled Circuits. Note: the above presentation of information-theoretic Yao is not the typical/original way that Yao's garbled circuit construction is presented. The following papers have a mix of the "traditional" presentation and the RE-based abstraction.

- Benny Applebaum's survey on "Garbled circuits as randomized encodings of functions: a primer" (you can also see his PhD thesis).
- Bellare, Hoang, Rogaway, CCS 2012 paper, "Foundations of Garbled Circuits".
- Yehuda Lindell and Benny Pinkas, A Proof of Yao's Protocol for Secure Two-Party Computation

[5] Recall that this requires emailing the teaching staff before class

- Vladimir Kolesnikov, Ranjit Kumaresan, Improved Secure Two-Party Computation via Information-Theoretic Garbled Circuits, SCN 2012.

Open problems include *Adaptive Garbling*. This is the setting where the part that doesn't depend on the input (the garbled circuit itself, or the part of the RE that depends only on the randomness) is selected first, and then the inputs are chosen *adaptively* after seeing it (this is particularly well motivated in the secure 2PC application). The proof of security of the original Yao protocol does not extend (it is not known if the protocol is adaptively secure or not), and there are open problems regarding the required communication complexity for adaptive security.

Decomposable RE, or PSM. Private simultaneous message model (PSM) is basically the same as *decomposable* RE: full decomposability corresponds to n-party (or multiparty) PSM, while 2-party PSM (which is the original model presented and often the default interpretation of "PSM") corresponds to 2-decomposability. There are many open problems related to the PSM model, we will provide some of them in the next topic below.

Locality of RE. We are not spending much time in class on the locality (or "parallel complexity") measure, but it is in fact a very important application of RE. It leads to results such as showing that if there are PRGs in NC^1 (which is true under standard cryptographic assumptions), there are also PRGs in NC^0_4. This is because for any n-ary function, there is a RE in NC^0_4 (namely where each output bit depends only on 4 input bits).

It is open whether the same can be shown for NC^3_0. There are several other interesting and important open problems here (both from a cryptographic perspective, and a complexity theoretic one).

5 Lecture 5+: a taste of PSM and CDS (Lower Bounds)

As mentioned above, Private simultaneous message model (PSM) can be viewed as decomposable randomized encoding, but can also be viewed as a very simple model of secure computation, where parties with correlated randomness send a single message to a "referee", who should learn the output of the function on their inputs, without learning anything else.

Besides being a natural and simple model (adding a privacy requirement to a standard communication complexity problem), it turns out to be very related to other topics such as PIR and secret sharing. Since we've seen some upper bounds (all the decomposable RE schemes that we've seen last class), in the next lecture we will show a lower bound for PSM (here too, the gap between the best upper and lower bounds is exponential, and making it smaller is an important open problem).

We will then give an overview of some *Conditional Disclosure of Secrets (CDS)* results, which can be viewed as an even simpler version (and a special case) of PSM. Here the parties (say there are two: Alice and Bob) hold a secret bit, in addition to their inputs and correlated randomness, and they each send one message to the referee, Carol. Carol knows the inputs, but doesn't know the secret bit. The goal is for Carol to output the secret if a predicate on the inputs evaluates to 1, and learn no information about the secret otherwise. Note that this is easy to achieve with a single bit of communication if privacy is not required (sending the secret in the clear).

Readings

Papers presented in class:

1. On the Complexity of Decomposable Randomized Encodings, Ball, Holmgren, Ishai, Liu, Malkin, ITCS 2020. We will focus on the information theoretic lower bound.

2. Gay, Kerenidis, Wee, Communication Complexity of Conditional Disclosure of Secrets and Attribute-Based Encryption, CRYPTPO 2015.

Supplementary Reading and Additional Directions

For now, this is a list of papers with some notes. The list will be curated (and cut down), to highlight open problems and directions.

- Feige, Kilian, Naor, A Minimal Model for Secure Computation, STOC 1994. This is the original paper that introduced the PSM model.

- Ishai, Wee, Partial Garbling Schemes and Their Applications, ICALP 2014.

- Applebaum, Arkis, Raykov, Vasudevan, Conditional Disclosure of Secrets: Amplification, Closure, Amortization, Lower-bounds, and Separations

  This paper contains a variety of results on CDS, including lower bounds.

  One cool result in this paper shows how to adapt a technique for generically amortizing space complexity in a certain sense (Potechin 2016) to conditionally disclose *long* secrets. (This was later improved Applebaum and Arkis ”On the power of amortization in secret sharing: d-uniform secret sharing and CDS with constant information rate.” TCC 2018.)

- Liu, Vaikuntanathan, Wee, Conditional Disclosure of Secrets via Non-Linear Reconstruction, CRYPTO 2017

  This paper presents two CDS constructions for general functions $f : [N] \times [N] \to \{0, 1\}$ with complexity that is $o(\sqrt{N})$.

  This was later extended to the multiparty setting (and this extension is the basis of recent breakthroughs in “normal” Secret Sharing) in:

  Liu, Vaikuntanathan, Wee. ”Towards breaking the exponential barrier for general secret sharing.” Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Cham, 2018.

- Applebaum, Holenstein, Mishra, Shayevitz, The Communication Complexity of Private Simultaneous Messages, Revisited, Eurocrypt 2018.

  Fixes a 2-party PSM lower bound from [FKN94] and show that it yields something for CDS as well. The lower bound is linear (!), essentially just $3N$ for functions $f : [N] \times [N] \to \{0, 1\}$, but still very non-trivial.

- Applebaum, Vasudevan “Placing Conditional Disclosure of Secrets in the Communication Complexity Universe”, ITCS 2019

In general, relates CDS complexity of a predicate to the communication complexity of the predicate in other models. One result of this is several improved lower bounds for simple, concrete functions.

- Assouline, Liu Multi-Party PSM, Revisited

  Gives an $O(N^{\frac{k-1}{2}})$ upper bound for arbitrary k-party functions $f : [N]^k \to \{0,1\}$. Improves on the paper below:

  Beimel, Kushilevitz, Nissim. "The complexity of multiparty PSM protocols and related models." EUROCRYPT 2018.

  First $O(\sqrt{N})$ bound for the 2-party case is due to Beimel, Ishai, Kumaresan, Kushilevitz "On the cryptographic complexity of the worst functions." TCC 2014.

6 Lecture 6,7: Upper Bounds and Private Information Retrieval (PIR)

The information theoretic setting for PIR involves 2 or more non-communicating servers holding a database, and a client who wants to retrieve one bit of the data, without revealing its query. PIR has many applications (both directly and indirectly), and is also closely related to the other topics studied in this class (either in terms of formal equivalence relationships, or in terms of techniques used in known constructions). In particular, state of the art upper bounds for PSM, CDS, and Secret Sharing all rely on PIR techniques.

Specific papers to be covered, and supplementary reading and directions, to be added.

7 Other Topics

We will have some selection of the following topics. If you're particularly interested in one of these topics to be presented in class (possibly by you?) let me know. All these topics are good areas for potential research projects.

7.1 Distributed Computing: Byzantine Agreement, Consensus, Broadcast

This is a special case of malicious MPC, which has been studied by the distributed computing community for a long time (with both upper and lower bounds), and is very relevant to current research as well (from MPC to cryptocurrencies).

Specific papers to be covered, and supplementary reading and directions, to be added.

7.2 Interactive Proof systems and (Statistical) Zero Knowledge

Details to be added. Here's a webpage for a related whole semester class: Chiesa and Shinkar's 2017 Berkeley class on Probabilistically Checkable and Interactive Proof Systems.

7.3 Randomness Extractors and Cryptographic Applications

7.4 Cryptographic Coding Theory

This can include topics such as non-malleable codes, locally decodable codes (equivalent to PIR), noisy channels, and more.