UNIVERSITY OF CHICAGO

Time-Space Trade-offs in Cryptographic Primitives

by
AKSHIMA

A thesis proposal
submitted in candidacy exam
for the degree of PhD
to the Department of Computer Science
of Physical Sciences Division
at the University of Chicago

Chicago, USA
March 2022

UNIVERSITY OF CHICAGO

Time-Space Trade-offs in Cryptographic Primitives

by

AKSHIMA

Approved by:

Dr. David Cash, Professor — Adviser

Computer Science

Dr. Aloni Cohen, Assistant Professor — Member of Committee

Computer Science

Dr. Hoeteck Wee, Senior Scientist — Member of Committee

NTT Research

Date of thesis defense: April 11, 2022

Abstract of the Thesis Proposal of

Akshima for PhD
in Computer Science

Title: Time-Space Trade-offs in Cryptographic Primitives

The research in complexity theory, for a long time now, has been conscious of memory as a resource in building algorithms with improved asymptotic complexity. There is an understanding to compare time-memory trade-offs as opposed to only running times while choosing between algorithms to solve any problem. While cryptographers have recognized memory to be a resource, there has been little effort to analyze cryptographic primitives in a memory-conscious manner until recently.

This work contributes towards the recent efforts of understanding the role of memory in the security of cryptographic primitives. Our study is two-fold:

1. How much better can any adversary that is capable of performing pre-computation and storing a bounded amount of information about the cryptographic primitive (under attack) do?

2. Are there cryptographic applications which are provably more secure against adversaries with lesser memory?

This work aims to focus on cryptographic hash functions for the first part of the study. The study would analyze properties of collision resistance and resistance against some restricted classes of collisions for these functions.

For the second part of the study, the aim is to analyze some popular constructions of pseudo-random permutations and pseudo-random functions against the memory-bounded adversaries.

Table of Contents

Definitions

Definition 1 ***Adversary*** *is a malicious entity that attempts to prevent any cryptosystem from achieving its goal.*

Definition 2 ***Cryptographic Hash Functions****, or simply referred to as Hash Functions, are functions that map inputs of arbitrary size to fixed size outputs, i.e.,* $\mathcal{H} : \{0,1\}^* \rightarrow \{0,1\}^\ell$ *for some fixed positive integer* ℓ*. They are one-way functions, which means they are hard to invert. Brute force search or rainbow tables are used for inverting these functions. Often times salted hash functions are used in applications to make them harder to invert by brute force search or using rainbow tables. Salted hash functions take an additional fixed size input called salt.*

Definition 3 *A* ***Collision*** *in a hash function* H *is defined as finding two distinct messages in* $\{0,1\}^*$ *that have the same output under* H*.* **m*-way Collision*** *in a hash function* H *is defined as finding* m *distinct messages* $\{0,1\}^*$ *that have the same output under* H *where* m *is a positive integer.*

Definition 4 ***Merkle Damgård*** *is a popular construction scheme for building a collision resistant Hash function for arbitrary input sizes from collision resistant compression functions on fixed input size. The scheme essentially breaks arbitrary sized inputs into blocks of fixed size and applies the compression function in sequence on these blocks.*

Definition 5 ***Random Oracle*** $\mathcal{O}$ *can be thought of as a machine implementing a function* H*. Its internal working are assumed unknown. An input to it is called* ***query****.*

Definition 6 ***Pseudo Random Functions*** *is a family of deterministic functions that are efficiently computable and indistinguishable from a truly random function by any efficient adversary.*

Definition 7 ***Pseudo Random Permutations*** *is a family of deterministic permutations that are efficiently computable and indistinguishable from a truly random permutation by any efficient adversary.*

Definition 8 *To prove security of an application using H in the* ***Random Oracle Model****, analysis assumes H to be a truly random function.*

To prove security of an application using H in the ***Auxiliary-Input Random Oracle Model****, analysis assumes H to be a truly random function and 2-stage adversaries $\mathcal{A} = (\mathcal{A}_0, \mathcal{A}_1)$ such that:*

1. *$\mathcal{A}_0$ gets computationally unbounded access to H and outputs a bounded size information about H, say of S-bits, and called advice.*

2. *$\mathcal{A}_1$ takes the advice output by $\mathcal{A}_0$ and the challenge as input and gets to make a bounded number of queries, say T to H to solve the challenge.*

Definition 9 *To prove security of an application using H in the* ***Bit-Fixing Random Oracle Model****, analysis assumes H to be a truly random function and 2-stage adversaries $\mathcal{A} = (\mathcal{A}_0, \mathcal{A}_1)$ such that:*

1. *$\mathcal{A}_0$ fixes some bounded number of bits of H, say P bits, to obtain say H'*

2. *$\mathcal{A}_1$ takes the challenge as input and gets to make a bounded number of queries, say T queries, to H' to solve the challenge for H'.*

Definition 10 *A* ***streaming adversary*** *is characterized by two parameters m, q such that the adversary gets to make q queries and receives the responses in a stream, i.e., the adversary cannot access the response to a particular query unless it stores that in its memory, which is bounded to m-bits.*

Definition 11 ***Block ciphers*** *are all functions $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ such that for every $K \in \{0,1\}^k$, $E_K := E(K, \cdot)$ is a permutation on $\{0,1\}^n$. For a block cipher E, we write E^{-1} to denote the inverse cipher, i.e. for every $K \in \{0,1\}^k$, $E_K^{-1}(\cdot)$ is inverse permutation of $E_K(\cdot)$.*

Definition 12 *For a block cipher $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$, we define* ***Double Encryption*** *with E as a function in $\{0,1\}^{2k} \times \{0,1\}^n \to \{0,1\}^n$ that on inputs $\hat{K}_1 || \hat{K}_2, x$ returns $E_{\hat{K}_2}(E_{\hat{K}_1}(x))$.*

Definition 13 *For some positive integers $M \leq N$, a permutation Π on $\{0,1\}^{\log N}$, we define* ***Truncated Permutation*** *with Π as a function in $\{0,1\}^{\log N} \to \{0,1\}^{\log M}$ that on input x returns the first $(\log N - \log M)$ bits of $\Pi(x)$.*

Abbreviations

PRP	Pseudo Random Permutation
PRF	Pseudo Random Function
RO	Random Oracle
AI-RO	Auxiliary Input Random Oracle
BF-RO	Bit Fixing Random Oracle
CR	Collision Resistance
MCR	Multi-way Collision Resistance
MD	Merkle Damgård
DE	Double Encryption

Proposal Outline

Chapter 1 gives a summary of the related work for Chapters 2, 3, 5 and 6. Chapters 2-4 give an overview of the works already published or in submission to conferences.

Chapter 2 talks about our results for a restricted class of collisions for MD based hash functions in the AI-RO model. Chapter 3 talks about our follow-up results for a larger class of restricted collisions, again for MD based hash functions in the AI-RO model. Chapter 4 talks about a tangential work on Encrypted databases. The chapter is self- contained in terms of definitions and related work.

Chapters 5 and 6 present the proposed works. Both of these are works in progress. Chapter 5 talks about more restricted classes of collisions and our conjectured results.

Chapter 6 is a proposed work contributing towards the second part of this study. The first part of the chapter focuses on analyzing double encryption scheme as a PRP and the second part focuses on analyzing truncated permutation scheme as a PRF.

Finally, Chapter 7 gives an expected timeline for completing these works and graduation.

Chapter 1

Related Work

Hellman [1] was the first to study pre-processing attacks for inverting functions, which was followed up by the famous work Fiat and Naor[2]. Recently, several works [3], [4] set out to understand the power of such attacks for finding collisions. All of them have studied this question in the auxiliary-input random oracle (AI-RO) model proposed by Unruh [5], for dealing with non-uniform and preprocessing attackers.

Dodis, Guo, and Katz [3] studied the collision resistance of a salted random function (which also corresponds to the $B = 1$ case for Merkle-Damgård). They prove an $\tilde{O}(S/N + T^2/N)$ security upper bound (with respect to a random salt). This bound shows the naive attack which precomputes collisions for S distinct salts as the advice (the $\tilde{O}(T^2/N)$ term is tight due to birthday attack) to be optimal.

Coretti, Dodis, Guo and Steinberger [6] further studied collision finding for salted Merkle-Damgård hash functions (corresponds to arbitrary B). Interestingly, unlike the $B = 1$ case, they show an attack achieving advantage $\tilde{\Omega}(ST^2/N)$, improving the birthday attack by a factor of S. They also prove this attack is optimal.

We mention that time-space lower bounds of attacks (or non-uniform security) against other fundamental cryptographic primitives, such as one-way functions, pseudorandom random generators, discrete log, have been investigated in various idealized models [3], [4], [6]–[11].

Motivated by analyzing post-quantum non-uniform security, several recent works [12], [13] studied the same question in the quantum setting, in which the adversary is given S-(qu)bit of advice and T quantum oracle queries. However, Unlike the classical setting, no matching bounds are known, even for $B = 2$ and $B = T$. The $\Omega(ST^3/N)$ security bound by Guo et al., suggests that the optimal attack may speed up the trivial quantum collision finding by a factor of S. However, the best known attack achieves $O(ST^2/N + T^3/N)$ for every $2 \leq B \leq T$.

Relevant to the second part of this study are works that have studied streaming adversaries. Some of the most noteworthy recent works have revisited the switching lemma, results on popular symmetric encryption designs and authenticated encryption designs and many other results by considering the role of the adversary's memory capability. Memory had traditionally been considered by cryptanalysts, as a large-memory algorithm may be infeasible to implement even if its runtime is within reach. However, from the direction of proving impossibility results, memory was not considered until recently, when it was pointed out [14], [15] that computational reductions in cryptography could be made "memory tight" and sometimes yield a wider range of effective bounds. In particular, when a problem becomes harder for small-memory algorithms (such as LPN), a memory-tight reduction may be important. On the other hand, some problems (like finding a collision in a generic hash function) are solved optimally with small memory and do not benefit from memory-tightness. Subsequent work [16]–[20] has explored possibility and impossibility results on memory tightness. Further work [21], [22] extended the consideration of memory to information-theoretic steps as well.

Jaegar and Tessaro in [22] were the first to study the switching lemma with bounded memory. They realized that giving time-memory lower bounds for adversaries that can repeat queries would be difficult. This is because such an adversary can make use of Pollard-rho type memoryless algorithms to find collisions. Hence, they focused on adversaries that cannot repeat queries. The authors, however, found proving an unconditional bound out of reach. Their proof relied on a combinatorial conjecture on hypergraphs being true to give a conditional loose upper boundon the advantage. Dinur [23] succeeded in proving a bound which has a matching attack up to log factors, by giving a reduction from the Disjointness problem in the communication model to streaming and using the existing results from communication complexity. Another recent work [24] improved the result from [22].

Tessaro and Thiruvengadam [21] showed equivalence between Double Encryption and a special case of Element distinctness, which is list disjointness, and used that to give a conjectured time-space lower bound on Double Encryption.This bound was later proved unconditionally in a restricted model, namely the post filtering model, by Dinur [25]. It must be noted that our aim is to study Double Encryption for a restricted class of adversary, namely the non-adaptive and non-repeating in the streaming model. Follow-up works [16], [17], [21], [22], [26] have studied more primitives, namely authenticated encryption and symmetric encryption.

Chapter 2

Short Collisions

We study short collisions in MD hash functions in the AI-RO model. Formally, by short collisions we mean the colliding messages are restricted in length size. Say the MD hash function uses a one-way compression function h in $[N] \times [M] \rightarrow [N]$, then a B-block collision looks for two distinct messages in $[M]^B$ that have the same output. We consider the following question.

> *How efficiently can an adversary in AI-RO model running in time T and computing an S-bit advice, find 2-block collisions in MD hash functions?*

The starting point of this work is the observation that Hellman's attack based on rainbow tables (or an easy modification of the attack in [6]) can find B-block collisions with success probability roughly STB/N.

While the attack was easy to modify for short collisions, proving that it is optimal constituted the main technical contribution of the work. In order to explain our technique, we recall the approach of [6] used to prove the $O(ST^2/N)$ bound for (unrestricted) collisions in MD hash functions. They used a technical approach (with tighter parameters) first developed by Unruh [5], which connects the AI-RO to the *bit-fixing random oracle (BF-RO)* model. Their work transfers lower bounds in the BF-RO to lower bounds in the AI-RO.

We show that the BF-to-AI template inherently cannot give a lower bound for finding short collisions, *because finding short collisions in the BF-RO is relatively easy.* That is, the lower bound of the form we would need for BF-RO simply does not hold. Thus another approach is required.

2.0.1 *Our lower bound technique.*

Given that the BF-to-AI technique cannot distinguish between finding short and unrestricted length collisions, we used find another approach. We employed an elegant method of Impagliazzo using concentration inequalities.

Compression arguments were previously observed [6] to be difficult (or "intractable") to apply to the setting of MD collision finding despite working in the original non-MD setting [3]. Given that compression was already difficult in this setting, it does not seem promising to extend it to the harder problem of short collisions.

To address these difficulties, we introduce a new technique that first applies a variant of the "constructive" Chernoff bound of Impagliazzo and Kabanets [27] to prove time-space tradeoff lower bounds.

The concentration-based approach to time-space tradeoff lower bounds was, to our knowledge, first introduced by Impagliazzo in an unpublished work, and then later elucidated in an appendix [28] (there an older work of Zimand [29] is also credited). The high-level idea is to first prove that any adversary (with no advice) can succeed on any fixed $U \in [N]$ subset of $\Omega(S)$ of inputs with probability $\varepsilon^{\Omega(S)}$. (In some sense bounding every sufficiently large "moment of the adversary"). The argument continues by applying a concentration bound to the random variable that counts the number of winning inputs for this adversary, showing that it wins on a $O(\varepsilon)$-fraction of inputs except with probability $2^{-\Omega(S)}$. In a final elegant step, one shows that every advice string is likely to be bad via a union bound over all possible 2^S advice strings, to get a final bound of ε.

The technique of Impagliazzo gives a direct and simple proof for the optimal bound on inverting a random permutation. There are two issues in applying it to short MD collisions however. First, as we formally show later, it provable fails ford MD hashing. The issue is that the adversary may simply succeed with probability greater than ε^S on some subsets U (see subsection **??**), so the first step cannot be carried out.

We salvage the technique by showing it is sufficient to bound the adversary's average advantage for *random* subsets U rather than *all* subsets. In the language of probability, we use a concentration bound that only needs *average* of the moments to be bounded by $\varepsilon^{\Omega(S)}$, rather than all of the moments; see Theorem **??**.

So far we have been able to reduce the problem of proving a AI-RO lower bound to the problem of bounding the probability that an adversary with no advice can succeed on every element of a random subset of inputs. For the problems we considered, even this appeared to be complicated. To tame the complexity of these bounds, we apply compression arguments; Note that we are only proving the simpler bound needed for the Impagliazzo technique, but using compression, when previously compression was used for the problem directly. Our variation has the interesting twist that we can not only compress the random function (as other work did), but also the random subset U on which the adversary is being run. This turns out to vastly simplify such arguments.

2.0.2 *Applications of our technique.*

We first apply our technique to reprove the $O(ST^2/N)$ bound for (non-short) collision finding against salted MD hash functions. We then turn to the question of short collisions and prove $O(ST/N)$ bound for finding 2-block collisions which has a matching attack.

We show that there are qualitative gaps between finding 1-block collisions, 2-block collisions, and unrestricted-length collisions. Specifically, while for 1-block collisions we have $\varepsilon = O((S + T^2)/N)$, we show that 2-block collisions are easier when $S > T$, as the optimal bound is $O((ST + T^2)/N)$. For unrestricted-length collisions there is another gap, where the optimal bound is $O(ST^2/N)$. Our bound for length 2 collisions uses our new compression approach used above.

2.0.3 *Bound for a restricted class of attacks.*

In addition we consider ruling out the class of attacks that gives optimal attacks in the known cases. Roughly speaking, these attacks use auxiliary information consisting of S collisions at well-chosen points in the function graph. In the online phase, the attack repeatedly tries to "walk" to these points by taking one "randomizing" step followed by several steps with zero-blocks.

For this class of attacks, we show that the best of choice of collision points will result in $\varepsilon = O(STB/N)$. This result requires carefully analyzing the size of large, low-depth trees in random functional graphs, a result that may be of independent interest.

Chapter 3

Bounded Length Collisions

This is a follow-up of the work talked about in 2. Our main contribution is the following:

For any $2 < B < T$, the advantage of the best adversary with S-bit advice and T queries for finding B-block collisions in Merkle-Damgård hash functions in the AI-RO model, is $\tilde{O}\left(\frac{\kappa TB}{N} + \frac{T^2}{N}\right)$ where $\kappa := \max\{S, \frac{S^2T^2}{N}\}$.

This result implies the following bounds:
For any $2 < B < T$, the advantage of the best adversary with S-bit advice and T queries for finding B-block collisions in Merkle-Damgård hash functions in the AI-ROM, is

1. $\tilde{O}(STB/N + T^2/N)$ when $ST^2 < N$;
2. $\tilde{O}(S^2T^3B/N^2 + T^2/N)$ when $ST^2 > N$.

The first bound supports the STB conjecture from the work in Chapter 2, i.e., for any $2 < B < T$ for the range of S, T such that $ST^2 < N$. This result is interesting despite the restriction on the range of S, T. That is because the only known bound for any B before this work was the ST^2/N upper bound [6] which could not be matched by any known attack. With the first bound, we conclude $\tilde{\Theta}((STB + T^2)/N)$ bound on the advantage when $ST^2 < N$. With the first bound, we can say that no attack can achieve better than $(STB + T^2)/N$ advantage when $ST^2 < N$. In other words, we conclude $\tilde{\Theta}((STB + T^2)/N)$ bound on the advantage when $ST^2 < N$.

For the range of S, T, B such that $ST^2 > N$ and $B < T$, our second bound strictly improves over the known bound of $\tilde{O}(ST^2/N)$ from [6]. For $B = T$, our bounds recover the $\tilde{O}(ST^2/N)$ bound which has a matching attack. Moreover, as $T^2 \leq N$ (from the birthday bound), $S^2T^3B/N^2 = O(S^2TB/N)$ always holds which makes our second bound off by a factor of at most S from the best-known attack.

3.0.1 *Our techniques*

Our initial inspiration is the recent framework of Chung, Guo, Liu, Qian [12] for establishing tight time-space tradeoffs in the quantum random oracle. Generally speaking, they reduce proving the security of a problem with S-bit advice to proving the security of multiple random instances of the problem, presented one at a time, *without* advice. Roughly speaking, they observe that, if any adversary (with no advice) can solve roughly S instances of the problem "sequentially" with advantage at most δ^S, then any adversary with S-bit advice can solve one instance of the problem with advantage at most $O(\delta)$.

This idea of reducing security of a problem with advice to the security of a multi-instance problem without advice was first introduced by Impagliazzo and Kabanets in [27] and also used by our previous work. The difference between [27] and the later works, including this work, is that we reduce to a "sequential" multi-instance game as opposed to "parallel" multi-instance problem. More concretely, in the parallel multi-instance problem, the adversary is presented with all the randomly chosen instances of the challenge problems to solve once at the start. Whereas in the multi-instance game, the adversary gets a new randomly chosen instance of challenge problem one at a time and only after solving all the previous challenges.

Chung et al. [12] recently explicitly demonstrated a separation between "sequential" multi-instance games and "parallel" multi-instance problems in terms of proving tight time-space lower bounds in the context of function inversion. In addition, Guo, Li, Liu and Zhang [13] pointed out connections between "sequential" multi-instance game and the presampling technique —— the main technique used by Coretti et al. [6] for proving tight bounds for $B = T$. Roughly speaking, all results relying on presampling technique can be reproved using "sequential" multi-instance games. That suggested that "sequential" multi-instance games have the potential to prove stronger results. Therefore we are motivated to adapt and take full advantage of "sequential" multi-instance games in the context of collision finding.

To illustrate the power of this method, we show how to recover the $O(ST^2/N)$ bound for the general case via the connection with presampling technique. In a high level, using "sequential" multi-instance games, it is sufficient to bound the advantage of any adversary winning a new game, conditioning on winning all previous games, by $O(ST^2/N)$. The key point is that the adversary made at most ST queries in previous rounds, and it has no advice about the random oracle. Therefore from the view of the adversary, the random oracle is effectively a (convex combination of) bit-fixing random oracles (BF-ROM), where at most ST-positions are known and the rest remains independent and random. Hence, it is sufficient to upper bound its winning probability in the bit-fixing random oracle by $O(ST^2/N)$, which is what exactly offered by the proof of Coretti et al. [6] using the presampling technique.

Work in Chapter 2 pointed out a barrier of using the presampling technique towards proving $B = 2$. In particular, one can achieve $\Omega(ST^2/N)$ in BF-ROM even for

$B = 2$. Our main insight is that, unlike the presampling technique in which bit-fixed values can be arbitrary, the worst case bit-fixed values are not typical and can be tolerated by refining the technique. By identifying the "high knowledge gaining" events and manage to show that they are all unlikely (which is intuitive but non-trivial to prove), we obtain a considerably simpler proof for the $B = 2$ result from 2 using our approach in this work. It is an upside of our technique that it modularises and separates out the bad events, making the overall proof simpler and more intuitive. Following the same structure, we then extend our proof to large B by identifying a few events, and obtain our main result.

Chapter 4

Reconstructing Encrypted Databases

In a separate line of work, we studied Encrypted databases that support range queries. We analyzed an abstraction of two dimensional database encryption schemes that allow range queries but leak access patterns and search patterns or query distribution.

Prior works have shown that any one dimensional database encryption scheme that has similar leakage is catastrophically prone to complete reconstruction up to reflection. Our objective here was to understand whether it is the same for two dimensional databases. Our results indicate that this leakage alone is provably insufficient for complete reconstruction.

We start by expanding on the technical terms used above before moving on to talk about the related work in some more detail. Finally we state our result.

Let $[a] := \{1, \ldots, a\}$ and $[a, b] := \{a, a + 1, \ldots, b\}$ for any positive integer a, b such that $a \leq b$.

Definition 14 *We define* ***one-dimensional database*** *as a collection of finite number of records, say R, where each record is a tuple of (record token, record value) in $[R] \times [N]$ for some positive integer N. For any positive integers N_0, N_1, we define a* ***two-dimensional database*** *as a collection of again R records where each record takes a value in $[N_0] \times [N_1]$.*

Definition 15 *A* ***range query*** *(henceforth we will use query), is defined as a tuple (a_0, a_1, b_0, b_1) where $a_0, b_0 \in [N_0]$, $a_1, b_1 \in [N_1]$, $a_0 \leq b_0$ and $a_1 \leq b_1$. The range query is answered by returning all the records that have values in $[a_0, b_0] \times [a_1, b_1]$.*
Let's denote the set of all range queries by Q, i.e., $Q := \{(a_0, a_1, b_0, b_1) | a_0, b_0 \in [N_0], a_1, b_1 \in [N_1], a_0 \leq b_0, a_1 \leq b_1\}$.
Query distribution *is defined as the distribution over this set Q.*

Note that it is assumed in our abstraction of the searchable encryption schemes that the records and queries are encrypted.

Definition 16 ***Access pattern leakage*** *is the revealing of record tokens for the records returned in response to any queries made to the database.*
Search pattern leakage *is used to determine whether two identical responses correspond to the same query.*

4.0.1 *Related Work*

Kellaris et al. [30] show that for a one-dimensional database over domain $[1, N]$, one can determine the exact record values up to reflection with $O(N^4 \log N)$ uniformly random queries. Also, reconstruction can be done with only $O(N^2 \log N)$ queries if the database is *dense*. Informally, a dense database is one in which each domain value is associated with at least one record. In [31], Lacharitè et al. improve on the dense database attack and present an algorithm that succeeds in reconstructing dense databases with $O(N \log N)$ queries. For large N, these query complexities can quickly become impractical, so they additionally presented an *ϵ-approximate database reconstruction* (ϵ-ADR) attack that recovers all plaintext values up to some additive ϵN error with $O(N \log \epsilon^{-1})$ queries.

The *sacrificial ϵ-ADR* approximation attack by Grubbs et al. [32] is scale free, i.e., its success depends only on the value of ϵ (as opposed to both ϵ and N). The first attack issues $O(\epsilon^{-4} \log \epsilon^{-1})$ queries and the second attack succeeds with $O(\epsilon^{-2} \log \epsilon^{-1})$ queries under the assumption that there exists some record in the database whose value is in the range $[0.2N, 0.3N]$ (or its reflection). Both attacks rely on uniform query distribution. The authors also prove that database reconstruction from known queries can be reduced to PAC learning.

Reconstruction attacks from k-NN queries are presented by Kornaropoulos et al. [33]. For cases when exact reconstruction is impossible, they characterize the space of valid reconstructions and give approximation reconstruction methods. In other work, Kornaropoulos et al. [34] combine access pattern leakage with search-pattern leakage and apply statistical learning methods to reconstruct databases with unknown query distributions from range and k-NN queries.

4.0.2 *Our Work*

This chapter elaborates only the contributions of the proposer to the paper[35]. We showed that there may be an exponential number of databases that have the same access pattern and search pattern leakage. This also characterizes this family of databases with the same leakage profile.

We will repeatedly use a strategy that generalizes the main observation of [30]. There, in trying to determine a point x, they observed that one can compute the proportion of responses in which x appears. Then they could proceed algebraically to limit the number of possible values for x. In particular, in one dimension, this narrowed x down to two values.

We generalized the notion of query density to two dimensions. And we show that any database with records in $[N_0] \times [N_1]$ can be partitioned into components and databases may have a polynomial in N_0, N_1 number of reflectable components(i.e., reflecting component along a diagonal does not changes its leakage profile). Reflecting a subset of these reflectable component results in a different database with the same leakage profile as the original database. This gives a family of possibly an exponential number of databases with the same leakage profile.

Chapter 5

More Types of Collisions

We give a description of more restricted types of collisions that we would like to study in the AI-RO model.

1) m-way Collisions

Finding multi-collisions is harder than finding collisions in RO model but that may not be the case in AI-ROM. We can surely say the following for now: - For $B = 1$, $\theta(S/N + T^2/N)$ is the bound for finding collisions and $\theta(S/N + T^m/N^{m-1})$ is for m-way collisions. So there is clearly a difference between the two problems for $B = 1$. - For unbounded length collisions, we can show $\tilde{\theta}(ST^2/N)$ for m-way collisions which is the same as the known bound for collisions.

It would be interesting to explore this for B length collisions even though it is reasonable to conjecture $\tilde{\theta}(STB/N)$ for $SB >= T$. It might be possible to prove the $\tilde{O}(STB/N)$ bound for multi-collisions. It is meaningful to study whether there is a difference between CR and MCR in AI-ROM.

2) Target Collisions

The notion of Target collision resistance was introduced by Bellare and Rogaway. It is the ability of an adversary to find the 2nd preimage for a fixed message (as opposed to a random message in the notion of 2nd preimage resistance).

Analyzing Target collision finding and how it stands in comparison to collision finding in the precomputation model should be interesting. Our conjecture is the advantage of finding unbounded length collisions for fixed messages like the all 0s message should match the collision finding bounds, that is there is an attack that achieves $\Omega(ST^2/N)$ advantage.

3) Fixed Prefix Target Collisions

This further restricts the desirable set of collisions by giving the adversary a prefix P alongside a random salt a in the online phase such that for a fixed message M, adversary has to output M' such that $P||M' \neq M$ and they collide in MD_h.

Chapter 6

Streaming Adversary

6.0.1 *Design for Pseudo Random Permutation: Double Encryption*

The DES block cipher was famously deployed with key length too short (56 bits) to provide meaningful security against modern adversaries. In order to increase security while reducing the difficulty of redeploying with replacement block ciphers, Triple-DES **X9.52**, which runs DES three times under some keying strategy, has become the standard in some domains. Triple encryption, rather than double (which would already have a key length of 108 bits), is used to defeat generic attacks which work faster than brute force. The famous *meet-in-the-middle attack* (MITM) against double encryption runs in about the time required to brute force a single key, leading to the common-knowledge that double encryption is no more secure than the original block cipher.

The MITM attack has the well-known drawback of requiring a large amount of memory to achieve constant advantage. A bounded-memory version of MITM, where one can store m blocks of memory and run in time q, achieves advantage about $mq/2^{2k}$ (the attack is usually presented with $m \approx q \approx 2^k$, in which case it achieves constant advantage). Thus, the common knowledge that double-encryption is no more secure than single encryption needs to be reassessed: From the lack of a better attack, it appears that there *is* an increase in security once on takes memory complexity into account.

We show that this is indeed the case for a large class of non-adaptive attacks against double encryption (that includes MITM). We prove it in the ideal-cipher model (ICM) that $mq/2^{2k}$ is the optimal advantage and double encryption can provide non-trivial security beyond 2^k queries against a restricted class of non-adaptive adversaries with bounded memory.

6.0.2 *Design for Pseudo Random Function: Truncated Permutations*

Switching lemma suggests that any random permutation on $\log N$-bits is indistinguishable from a $\log N$-bit random function by any adversary making q queries as far as $q^2 = o(N)$. However, the matching attack which looks for collisions in a brute-force manner requires $\tilde{O}(q)$ memory to achieve $\theta(q^2/N)$ advantage. A bounded memory version of the attack only achieves $\theta(mq/N)$ advantage. Dinur's work [23] proved that no streaming adversary with m-bit memory making q queries can do better than $O(mq/N)$, proving the attack to be optimal.

More constructions such as the truncated permutations have been shown to defeat the brute force attacks in the memory unbounded regime. For instance, the truncated permutation construction that truncates last $\log M$ bits of the output from a random permutation for some positive integer $M \leq N$, is indistinguishable from a $\log(N/M)$-bit random function up to $N \cdot M$ queries [36], [37]. We want to analyze these constructions against adversaries with bounded memory.

Chapter 7

Expected Timeline

	Work and/or Current Status	Expected Completion
1	The project summarized in Chapter 3 is currently under submission and needs some rewriting	April 2022
2	The project on analyzing streaming adversaries, proposed in Chapter 6, is currently in progress	May 2022
3	The project on studying more types of collisions, proposed in Chapter 5 has been started	July 2022
4	Thesis writing	July/August 2022
5	Thesis defense	August 2022

Table 7.1: Expected Timeline

Bibliography

[1] M. Hellman, "A cryptanalytic time-memory trade-off," *IEEE Trans. Inf. Theor.*, vol. 26, no. 4, pp. 401–406, Jul. 1980.

[2] A. Fiat and M. Naor, "Rigorous time/space tradeoffs for inverting functions," in *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, 1991, pp. 534–541.

[3] Y. Dodis, S. Guo, and J. Katz, "Fixing cracks in the concrete: Random oracles with auxiliary input, revisited," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2017, pp. 473–495.

[4] S. Coretti, Y. Dodis, and S. Guo, "Non-uniform bounds in the random-permutation, ideal-cipher, and generic-group models," in *Annual International Cryptology Conference*, Springer, 2018, pp. 693–721.

[5] D. Unruh, "Random oracles and auxiliary input," in *Annual International Cryptology Conference*, Springer, 2007, pp. 205–223.

[6] S. Coretti, Y. Dodis, S. Guo, and J. Steinberger, "Random oracles and non-uniformity," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2018, pp. 227–258.

[7] A. De, L. Trevisan, and M. Tulsiani, "Time space tradeoffs for attacks against one-way functions and prgs," in *Annual Cryptology Conference*, Springer, 2010, pp. 649–665.

[8] D. Chawin, I. Haitner, and N. Mazor, "Lower bounds on the time/memory tradeoff of function inversion," in *Theory of Cryptography - 18th International Conference, TCC 2020, Durham, NC, USA, November 16-19, 2020, Proceedings, Part III*, 2020, pp. 305–334. DOI: 10.1007/978-3-030-64381-2_11. [Online]. Available: https://doi.org/10.1007/978-3-030-64381-2%5C_11.

[9] H. Corrigan-Gibbs and D. Kogan, "The discrete-logarithm problem with preprocessing," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2018, pp. 415–447.

[10] ——, "The function-inversion problem: Barriers and opportunities," in *Theory of Cryptography Conference*, Springer, 2019, pp. 393–421.

[11] N. Gravin, S. Guo, T. C. Kwok, and P. Lu, “Concentration bounds for almost k-wise independence with applications to non-uniform security,” in *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, 2021, pp. 2404–2423. DOI: 10.1137/1.9781611976465.143. [Online]. Available: https://doi.org/10.1137/1.9781611976465.143.

[12] K.-M. Chung, S. Guo, Q. Liu, and L. Qian, “Tight quantum time-space tradeoffs for function inversion,” in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, 2020, pp. 673–684.

[13] S. Guo, Q. Li, Q. Liu, and J. Zhang, “Unifying presampling via concentration bounds,” in *Theory of Cryptography Conference*, Springer, 2021, pp. 177–208.

[14] B. Auerbach, D. Cash, M. Fersch, and E. Kiltz, “Memory-tight reductions,” in *Annual International Cryptology Conference*, Springer, 2017, pp. 101–132.

[15] Y. Wang, T. Matsuda, G. Hanaoka, and K. Tanaka, “Memory lower bounds of reductions revisited,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2018, pp. 61–90.

[16] A. Ghoshal, J. Jaeger, and S. Tessaro, “The memory-tightness of authenticated encryption,” in *Annual International Cryptology Conference*, Springer, 2020, pp. 127–156.

[17] A. Ghoshal and S. Tessaro, “On the memory-tightness of hashed elgamal,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2020, pp. 33–62.

[18] D. Diemert, K. Gellert, T. Jager, and L. Lyu, *Digital signatures with memory-tight security in the multi-challenge setting*, Cryptology ePrint Archive, Report 2021/1220, https://eprint.iacr.org/2021/1220, 2021.

[19] R. Bhattacharyya, “Memory-tight reductions for practical key encapsulation mechanisms,” 2020, pp. 249–278. DOI: 10.1007/978-3-030-45374-9_9.

[20] A. Ghoshal, R. Ghosal, J. Jaeger, and S. Tessaro, *Hiding in plain sight: Memory-tight proofs via randomness programming*, Cryptology ePrint Archive, Report 2021/1409, https://eprint.iacr.org/2021/1409, 2021.

[21] S. Tessaro and A. Thiruvengadam, “Provable time-memory trade-offs: Symmetric cryptography against memory-bounded adversaries,” in *Theory of Cryptography Conference*, Springer, 2018, pp. 3–32.

[22] J. Jaeger and S. Tessaro, “Tight time-memory trade-offs for symmetric encryption,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2019, pp. 467–497.

[23] I. Dinur, “On the streaming indistinguishability of a random permutation and a random function,” in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2020, pp. 433–460.

[24] I. Shahaf, O. Ordentlich, and G. Segev, "An information-theoretic proof of the streaming switching lemma for symmetric encryption," in *IEEE International Symposium on Information Theory, ISIT 2020, Los Angeles, CA, USA, June 21-26, 2020*, IEEE, 2020, pp. 858–863.

[25] I. Dinur, "Tight time-space lower bounds for finding multiple collision pairs and their applications," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2020, pp. 405–434.

[26] W. Dai, S. Tessaro, and X. Zhang, "Super-linear time-memory trade-offs for symmetric encryption," in *Theory of Cryptography Conference*, Springer, 2020, pp. 335–365.

[27] R. Impagliazzo and V. Kabanets, "Constructive proofs of concentration bounds," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, 2010, pp. 617–631.

[28] R. Impagliazzo, "Relativized separations of worst-case and average-case complexities for np," in *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity*, ser. CCC '11, IEEE Computer Society, 2011, pp. 104–114, ISBN: 978-0-7695-4411-3. DOI: 10.1109/CCC.2011.34. [Online]. Available: https://doi.org/10.1109/CCC.2011.34.

[29] M. Zimand, "How to privatize random bits," University of Rochester, Tech. Rep., Apr. 1996.

[30] G. Kellaris, G. Kollios, K. Nissim, and A. O'Neill, "Generic attacks on secure outsourced databases," in *Proc. ACM Conf. on Computer and Communications Security 2016*, ser. CCS 2016, 2016.

[31] M.-S. Lacharité, B. Minaud, and K. G. Paterson, "Improved reconstruction attacks on encrypted data using range query leakage," in *Proc. IEEE Symp. on Security and Privacy 2018*, ser. S&P 2018, 2018.

[32] P. Grubbs, M. Lacharité, B. Minaud, and K. G. Paterson, "Learning to reconstruct: Statistical learning theory and encrypted database attacks," in *Proc. IEEE Symp. on Security and Privacy 2019*, ser. S&P 2019, 2019.

[33] E. M. Kornaropoulos, C. Papamanthou, and R. Tamassia, "Data recovery on encrypted databases with k-nearest neighbor query leakage," in *Proc. IEEE Symp. on Security and Privacy 2019*, ser. S&P 2019, 2019.

[34] ——, "The state of the uniform: Attacks on encrypted databases beyond the uniform query distribution," in *Proc. IEEE Symp.on Security and Privacy 2020*, ser. S&P 2020, 2020.

[35] F. Falzon, E. A. Markatou, Akshima, *et al.*, “Full database reconstruction in two dimensions,” in *CCS ’20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*, J. Ligatti, X. Ou, J. Katz, and G. Vigna, Eds., ACM, 2020, pp. 443–460. DOI: 10.1145/3372297.3417275. [Online]. Available: https://doi.org/10.1145/3372297.3417275.

[36] A. J. Stam, “Distance between sampling with and without replacement,” *Statistica Neerlandica*, vol. 32, no. 2, pp. 81–91, 1978.

[37] S. Gilboa and S. Gueron, “The advantage of truncated permutations,” *Discrete Applied Mathematics*, vol. 294, pp. 214–223, 2021.