A Combinatorial Approach to Leakage Abuse Attacks and their Mitigation

Francesca Falzon

March 2022

Abstract

With the rise of remote cloud services and the consequent rise in data breaches, there is an increased need for the secure outsourcing of data. The problem of enabling query processing over encrypted data without decryption is a challenging one, and approaches ranging from software to hardware solutions have been proposed. In this work, we take a closer look at a class of solutions that are efficient and deployable in the near-term future and that employ the use of light weight symmetric key primitives. In exchange for this added efficiency, these schemes leak certain information about the underlying data and queries. We explore the limitations of what a passive server-side adversary can learn from this information leakage, and present practical constructions that minimize leakage while supporting complex queries.

In the first two parts of this work, we present the first attacks on schemes that support range queries over multi-attribute (or multi-dimensional) data. We describe the information theoretic limitation of reconstruction attacks in two settings and show that, in both cases, there can be an exponential number of distinct databases that produce equivalent leakage. We present a full database reconstruction attack that reconstructs the database when all queries are observed. We then relax these assumptions, and present an order reconstruction attack and an approximate database reconstruction attack that require only a strict subset of the possible range queries to succeed.

In the third part of this work, we shift our focus to schemes that support shortest path queries on graph-structured data. We initiate our study by describing a query recovery attack on a graph encryption scheme, which we call the GKT scheme (Ghosh et al. AsiaCCS 2021). We then present a modified version of the GKT scheme with reduced leakage at the expensive of interactive queries and increased storage overhead. Our proposed scheme uses data-structure techniques to decompose the graph into non-intersecting sub-paths which are then encrypted. We support our scheme with a detailed cryptanalysis and explain why this new approach provably mitigates our previous attack.

1 Introduction

> **Thesis statement:** Systems that enable the efficient processing of complex queries over encrypted data often leak information about the underlying data and their queries. We take a combinatorial approach to cryptanlyzing this leakage and mitigating these attacks.

Data breaches have been occurring with alarming frequency in the last few years, with billions of accounts being compromised in the Yahoo, Alibaba, and LinkedIn data breaches alone [1]. When such information is compromised, the effects - such as identity theft and financial fraud - can be devastating. Encryption can mitigate the risk of a data breach, however, the straightforward solution for encrypting data prevents the server from being able to search over the encrypted data without possession of the decryption key.

Encrypted databases (EDBs) provide one of the only plausible solutions for strongly mitigating such attacks. EDBs have been extensively studied, and many solutions have been suggested to enable server query processing over encrypted data on behalf of clients. One potential solution is to use heavy theoretical cryptographic solutions such has ***fully-homomorphic encryption*** [2] or ***oblivious RAM (ORAM)*** [3]. While these provide strong security guarantees, they are still not as efficient as we would like them to be despite many recent advances. Another solution is to use ***trusted hardware*** solutions, such as Intel SGX [4,5,6]. However, while readily deployable with today's technologies, trusted hardware has been shown be vulnerable to powerful timing attacks [7], attacks leveraging rogue data cache loads that can extract secret keys [8], and code-reuse attacks that do not even require kernel privilege [9]. Hardening applications against such attacks often requires re-writing the application code [10] or using ORAM [11]. More over, trusted hardware still requires trusting a third party i.e. the company that manufactures the hardware.

The third class of solutions, which we will refer to as ***structured encryption (STE)*** [12], is what this thesis is concerned with. Such solutions use light weight symmetric key primitives to encrypt structured data. This structured data may take on various forms ranging from document-oriented data to tree-structured data to matrix-structured data. The efficiency of STE schemes comes at a cost: these schemes are inherently "leaky" and reveal certain information about the underlying data and queries being issued. In fact, a number of attacks that leverage this leakage have been described (e.g. [13, 14, 15, 16, 17]). As such, the state of the art is still far from being as secure as we would hope. Understanding the security of practical STE schemes is a key-question in applied cryptography and computer security.

This thesis aims at advancing our understanding of the security of STE schemes and making strides towards more practical schemes that support complex queries. In particular, we propose to investigate schemes that support range queries over multi-attribute (multi-dimensional) datasets and shortest path queries over graphs. We plan to develop novel constructions and cryptanalyze them. Our work will include both theoretical and experimental contributions.

2 Proposal Outline

We now describe the organization of this thesis proposal.

Chapters 3-4 overview work we have already performed. In Chapter 3, we present our paper "Full Database Reconstruction Attack in Two Dimensions," joint work with Evangelia Anna Markatou, Akshima, David Cash, Adam Rivkin, Jesse Stern, and Roberto Tamassia which appeared in CCS 2020 [17]. Chapter 4 covers our follow-up paper "Reconstructing with Less: Leakage Abuse Attacks in Two Dimensions," joint work with Evangelia Anna Markatou, Roberto Tamassia, and William Schor which appeared in CCS 2021 [18].

In Chapter 5, we present our proposed work. The first part of this chapter describes ongoing research on an efficient query recovery attack against a recently published graph encryption scheme [19], for which we have already obtained promising preliminary results. In the second part of the chapter, we outline our plans for developing a novel graph encryption scheme that employs data-structure techniques to reduce leakage and mitigate our described query recovery attack.

References

[1] accessed on February 25, 2022.

[2] Craig Gentry and Dan Boneh. *A fully homomorphic encryption scheme*, volume 20:09. Stanford university Stanford, 2009.

[3] Oded Goldreich. Towards a theory of software protection and simulation by oblivious rams. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, page 182–194, New York, NY, USA, 1987. Association for Computing Machinery.

[4] Ittai Anati, Shay Gueron, Simon Johnson, and Vincent Scarlata. Innovative technology for cpu based attestation and sealing. In *Proceedings of the 2nd international workshop on hardware and architectural support for security and privacy*, volume 13. ACM New York, NY, USA, 2013.

[5] Matthew Hoekstra, Reshma Lal, Pradeep Pappachan, Vinay Phegade, and Juan Del Cuvillo. Using innovative instructions to create trustworthy software solutions. In *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*, HASP '13, New York, NY, USA, 2013. Association for Computing Machinery.

[6] Frank McKeen, Ilya Alexandrovich, Alex Berenzon, Carlos V. Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R. Savagaonkar. Innovative instructions and software model for isolated execution. In *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*, HASP '13, New York, NY, USA, 2013. Association for Computing Machinery.

[7] Jo Van Bulck, Frank Piessens, and Raoul Strackx. Nemesis: Studying microarchitectural timing leaks in rudimentary cpu interrupt logic. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, page 178–195, New York, NY, USA, 2018. Association for Computing Machinery.

[8] Jo Van Bulck, Marina Minkin, Ofir Weisse, Daniel Genkin, Baris Kasikci, Frank Piessens, Mark Silberstein, Thomas F. Wenisch, Yuval Yarom, and Raoul Strackx. Foreshadow: Extracting the keys to the intel SGX kingdom with transient Out-of-Order execution. In *27th USENIX Security Symposium (USENIX Security 18)*, page 991–1008, Baltimore, MD, aug 2018. USENIX Association.

[9] Andrea Biondo, Mauro Conti, Lucas Davi, Tommaso Frassetto, and Ahmad-Reza Sadeghi. The guard's dilemma: Efficient Code-Reuse attacks against intel SGX. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 1213–1227, Baltimore, MD, aug 2018. USENIX Association.

[10] Ernie Brickell, Gary Graunke, Michael Neve, and Jean-Pierre Seifert. Software mitigations to hedge aes against cache-based software side channel vulnerabilities, 2006.

[11] Emil Stefanov, Marten Van Dijk, Elaine Shi, T.-H. Hubert Chan, Christopher Fletcher, Ling Ren, Xiangyao Yu, and Srinivas Devadas. Path oram: An extremely simple oblivious ram protocol. *J. ACM*, 65(4), apr 2018.

[12] Melissa Chase and Seny Kamara. Structured encryption and controlled disclosure. In *Advances in Cryptology – ASIACRYPT 2010*, 2010.

[13] Mohammad Saiful Islam, Mehmet Kuzu, and Murat Kantarcioglu. Access pattern disclosure on searchable encryption: Ramification, attack and mitigation. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012*. The Internet Society, 2012.

[14] Georgios Kellaris, George Kollios, Kobbi Nissim, and Adam O'Neill. Generic attacks on secure outsourced databases. In *Proc. ACM Conf. on Computer and Communications Security 2016*, CCS 2016, 2016.

[15] Marie-Sarah Lacharité, Brice Minaud, and Kenneth G Paterson. Improved reconstruction attacks on encrypted data using range query leakage. In *Proc. IEEE Symp. on Security and Privacy 2018*, S&P 2018, 2018.

[16] F. Betül Durak, Thomas M. DuBuisson, and David Cash. What else is revealed by order-revealing encryption? In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1155–1166, New York, NY, USA, 2016. Association for Computing Machinery.

[17] Francesca Falzon, Evangelia Anna Markatou, Akshima, David Cash, Adam Rivkin, Jesse Stern, and Roberto Tamassia. Full Database Reconstruction in Two Dimensions. In *Proc. ACM Conf. on Computer and Communications Security*, CCS, 2020.

[18] Evangelia Anna Markatou, Francesca Falzon, Roberto Tamassia, and William Schor. Reconstructing with less: Leakage abuse attacks in two-dimensions. In *Proc. ACM Conf. on Computer and Communications Security*, CCS, 2021.

[19] Esha Ghosh, Seny Kamara, and Roberto Tamassia. Efficient graph encryption scheme for shortest path queries. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*, ASIA CCS '21, page 516–525, New York, NY, USA, 2021. Association for Computing Machinery.