Budapest University of Technology and Economics
Department of Networked Systems and Services

Cryptographic Problems in the Context of Data Markets

Collection of Ph.D. Theses
of
Máté Horváth

Supervisor: Levente Buttyán, Ph.D.

Budapest, Hungary
2020

Acknowledgement

I would like to acknowledge the financial support of the National Research, Development and Innovation Office (NKFIH) of Hungary,[1,2] the MELLODDY project,[3] the CELSA Research Fund, and the Sándor Csibi Grant.

List of Abbreviations

2PC	secure two-party computation
ABE	attribute-based encryption
CKA2	adaptive chosen keyword attack
CP-ABE	ciphertext-policy attribute-based encryption
CPFE	controlled private function evaluation
DB	data broker
DDH	decisional Diffie–Hellman assumption
DO	data owner
EU-CMA	existential unforgeability under chosen message attack
FE	functional encryption
GID	global identifier
IND-CKA2	indistinguishability under adaptive chosen keyword attack
IND-CPA	indistinguishability under chosen plaintext attack
IoT	internet of things
MAC	message authentication code
OT	oblivious transfer
PFE	private function evaluation
rCPFE	relaxed controlled private function evaluation
RL	revocation list
SSE	searchable symmetric-key encryption
SXDH	symmetric external Diffie–Hellman assumption
VAS	value added service
VASP	value added service provider

[1]The research presented in this work was supported by the National Research, Development and Innovation Office (NKFIH) of Hungary under Grant Contract No. 116675 (K).

[2]The presented work was carried out partly within the SETIT Project (2018-1.2.1-NKP-2018-00004). Project no. 2018-1.2.1-NKP-2018-00004 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2018-1.2.1-NKP funding scheme.

[3]This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N° 831472. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

1 Introduction

The past decades show that our everyday life is going through a continuous paradigm shift: from physical to virtual. The past years demonstrates that the pace of this change is speeding up. One of the biggest challenges of this phenomenon is that in the virtual world, we have to take care of things that we have taken for granted earlier in the physical world. A prominent example of this is trust. When Alice walked into a fashion store, spent there an hour finding her favourite dress, paid with cash, and finally went home cheerfully, for sure she did not have to worry that someone is stealing her last month salary, spying on her to see which dresses she liked and which she did not, which size is fitting on her, or where she was coming from and going to. To do these misdeeds in the physical world requires determination and significant effort, so people usually trust each other not doing them. However, the same activity in the virtual setting (e.g. in an online store) is much more "dangerous" as an entire industry is built around spying on Alice and her fellows to collect data about them [EN16], which turned out to be extremely valuable. The result is that trust cannot be evident anymore.

According to the neat definition of Boaz Barak, "cryptography is about replacing trust with mathematics" [Bar16]. This observation highlights the increasing role of cryptography in our everyday life. Existing cryptographic methods find newer and newer applications whenever a new area develops in the virtual world, or a new cryptographic challenge appears when there are no ready to use solutions or their adoption is non-trivial.

My dissertation studies the emerging concept of data markets and its connections to cryptography. The growing importance of data is beyond question today. According to an EU report [CMM+20] "the value of the data economy, which measures the overall impacts of the data market on the economy as a whole, exceeded the threshold of €400 billion in 2019 for the EU27 plus the United Kingdom, with a growth of 7.6% over the previous year." In spite of the large numbers, data economy today mainly involves companies and the end users remain one of the main sources of data collection instead of having real role in the market. Current technological trends, such as the proliferation of smart devices and the internet of things (IoT), can change this situation as the rapidly increasing amount of data is waiting for utilization. The main barrier of this is that in most cases the collected data is only available for the user and manufacturer of a sensor or smart device. One possible way of exploiting the full potential of this information is to build an ecosystem around it. This is exactly the idea of data markets [OLJ+19], the basic concept of which is depicted in Fig. 1. Here a data broker (DB) buys data from the data owners (DOs) and resell the collected data (possibly together with computing resources) to third parties that provides some value added service (VAS) to their users. These services can typically help predictions or support optimization via analysing a wide range of data. While there are several approaches to realize data markets in practice [Oce, Dat17, IOT, Dat], I focused on the research questions that the security of these markets pose.

As a first step, I provided a general system model for data markets and relying on

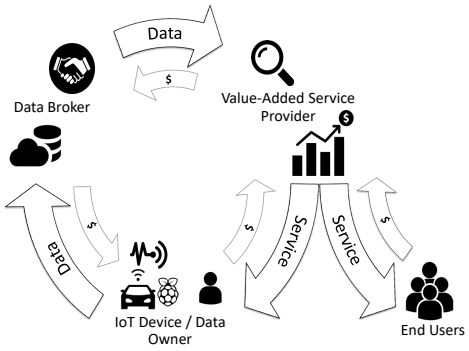


Figure 1: Centralized data market model.

this, analysed data markets based on the possible security requirements of the different participants. I provide a problem-domain structuring, in which I systematically identify the possible scenarios that are determined by the possible goals, trust relations, and requirements of the participants. Then these scenarios are connected to different areas of cryptography that also helps to identify open problems in the area. I improve upon the state of the art, in the following three different areas.

I investigated problems where one has to face inherent barriers stemming from interests that essentially contradict with each other but should be satisfied concurrently. As one can not have a cake and eat it too, I was always looking for some trade-off that takes us closer to a satisfactory solution.

The first problem is the conflict of interests of participants in private function evaluation (PFE) protocols. One of them aims to hide an input to a function that is the secret of the other one. Assuming that the function provider has access to the output, the maximum that we can hope for is that the protocol reveals no more information about the input that is already leaked by the output. Even if the entire input is not possible to leak in this way (output length is typically shorter than input length), when launching the protocol, the input provider cannot know what will be revealed about the input. I initiate the study of partial input information leakage in this context and propose the notion of controlled PFE and a relaxation of it. I show generic realizations of these new notions and demonstrate the applicability of the protocol fulfilling the relaxed requirements by implementing it for the inner product functionality (see more details §2.2 and §4 of the dissertation).

The next topic, I have dealt with is fine-grained access control to encrypted data, more precisely access right revocation in the context of attribute-based encryption (ABE). The challenge of user revocation in ABE schemes comes from their extreme flexibility in user identification. Concretely, it is very easy to identify various user groups when describing individuals with their non-unique attributes. However, when using an attribute-based description of users, it becomes very hard a to efficiently identify a single user, who may share his or her attributes with multiple other users. It follows that

achieving both flexible access control and efficient user revocation in the same system requires compromise. To this end, I propose a revocable multi-authority ciphertext-policy attribute-based encryption (CP-ABE) scheme and prove its security (for the details of this contribution, I refer to §2.3 and §5 of the dissertation).

Finally, the last part of my dissertation is dedicated to the investigation of the inherent tension that appears during the design of searchable encryption. The problem is that in order to facilitate efficient search (i.e. breaking the $O(n)$ search complexity barrier) one has to introduce some structure in the encrypted database. However, in case of an update, it requires special care to not leak information when the new entry is inserted in the structure. This care, typically incurs significant efficiency loss. In this context, I revisit the so-called "forward index" approach and show its relevance when updates are frequent and searching in parts of the database is enough. For the details of the related contributions, see §2.4 (and §6 of the dissertation).

The possible applications of my results are summarized in §3.

2 New Results

This section is dedicated to the brief introduction of my results that are discussed in details in my dissertation.

2.1 Problem Domain Analysis for Data Markets

When considering the security and privacy aspects of a concept, it is natural to first assess the problem domain. As to the best of my knowledge data markets were never investigated from this aspect, in [C3] I examined the trust relations between the different participants of the market and explored the effect of these relations on how data has to be handled in order to satisfy the requirements of each party.

THESIS 0: *I proposed a general system model, describing the possible interests, roles, and activities of participants in a data market, in order to be able to formalize their possible security and privacy goals. I investigated the different constellations of these goals and identified the possible scenarios that emerge from the meaningful combinations of the goals. Finally, I connected the resulting scenarios to the relevant areas of cryptography.*

My problem domain analysis considered the following possible trust relationships of the participants. DOs either allow access to their plaintext data to the DB or it is only allowed to see ciphertexts. Similarly, value added service providers (VASPs) can either buy plaintext data or only information that is derived from raw data, meaning that a VASP can obtain raw data only in an encrypted form. The relationship between DB and VASPs is more diverse. Assuming that VASPs can buy computational resources together with data, they might intend to hide the function to be computed, the input value(es) used in the computation, the metadata of the used input, or the output. From the perspective of the DB the computation is either not restricted or possibly the DB intends

to verify the function before computing it. The combinations of the listed requirements determine 2^7 different scenarios (see Table 1), some of which are contradictory, some are trivial to realize, and some can be connected to different areas of cryptography. The analysis of these scenarios in §3 of the dissertation also highlights several open problems motivating further research on certain cryptographic primitives.

2.2 Controlling Partial Input Information in Private Function Evaluation

Private function evaluation (PFE) protocols enable two parties, Alice and Bob, to jointly compute the function of Bob, on the input of Alice, such that at the end, either one or both of them obtain the output (we are going to consider those cases when Bob can see the result). The privacy requirement of PFE is that by executing the protocol, none of participants should obtain any more information about the contribution of the other, than what can already be extracted from the output of the computation. Clearly, function privacy prevents Alice to control or even to track what information about her data was revealed to Bob, which problem leads to the following questions that I investigated in [C2]:

Is it possible to enable the input provider to rule out the leakage of specific sensitive information in PFE without exposing what partial information she wants to hide? What kind of trade-offs between input and function privacy can lead to efficient protocols with meaningful security?

THESIS 1.1: *I proposed the novel notion of controlled private function evaluation (CPFE) by introducing a definitional framework to prevent partial input information leakage for arbitrary information while capturing different flavours of function privacy. The presented definitions consider simulation based security against semi-honest participants.*

The formal security definitions of CPFE and relaxed CPFE require that the protocol messages, obtained by the participants, has to be indistinguishable from so-called simulated messages. Intuitively, security is guaranteed by the fact that the simulated messages has to be prepared while only having access to the ideal functionality of the protocol (i.e. the inputs and outputs of a party). The proposed ideal functionalities of CPFE and relaxed controlled private function evaluation (rCPFE) are depicted on Fig. 2. For more details, see §4.3.1 of the dissertation.

THESIS 1.2: *I proposed a generic realisation of CPFE, based on universal circuits and secure two-party computation (2PC) protocols. The security of the resulting protocol relies on the security of the underlying 2PC.*

The protocol, proposed in §4.3.2 of the dissertation, is conceptually simple and demonstrates the general realizability of CPFE. The main idea is to handle the constraints of the input provider on the evaluated function as an additional input to a secure function

	function	input value(s)	input metadata	output	Trusted Data Broker (stores plaintext data): No restriction on the computed function (0)	Trusted Data Broker (stores plaintext data): Limited function queries (1)	Untrusted Data Broker (without access to plaintext data): No restriction on the computed function (0)	Untrusted Data Broker (without access to plaintext data): Limited function queries (1)
Trusted VASP (can access to plaintext data)	0	0	0	0	DB = VASP	As VASP is trusted by the DO, it can access to plaintext data that can be used for any computation without the consent of the DB	Revealing the input to the DB contradicts with the data owner's will to hide data from DB	As VASP is trusted by the DO, it can access to plaintext data that can be used for any computation without the consent of the DB
	0	0	1	0	public input⇔secret metadata			
	0	0	0	1	Function and input determine the output			
	0	0	1	1				
	0	1	0	0	Metadata reveals input		As VASP can access to plaintext data the evaluated function is not revealed to DB	
	0	1	0	1				
	0	1	1	0	As VASP can access to plaintext data the evaluated function is not revealed to DB			
	0	1	1	1				
	1	0	0	0	VASP computes locally and publishes the output		Revealing the input to the DB contradicts with the data owner's will to hide data from DB	
	1	0	0	1	VASP computes locally			
	1	0	1	0	Public input⇔secret metadata			
	1	0	1	1				
	1	1	0	0	Metadata reveals input		Fig. 3.2a with published output	
	1	1	0	1			❷ – Fig. 3.2a (see §2.3)	
	1	1	1	0	VASP downloads all data, computes locally, publishes output		Fig. 3.2b with published output	
	1	1	1	1	VASP downloads all data and computes locally		❸ – Fig. 3.2b (see §2.4)	
Untrusted VASP (without access to plaintext data)	0	0	0	0	Without restriction on the computable functions VASP could access the output of the identity function contradicting with the requirement that it should not get plaintext data	Function verification with outsourced computation	Without restriction on the computable functions VASP could access the output of the identity function contradicting with the requirement that it should not get plaintext data	Revealing the input to the DB contradicts with the data owner's will to hide data from DB
	0	0	1	0		Public input⇔ secret metadata		
	0	0	0	1		Function and input determine the output		
	0	0	1	1				
	0	1	0	0		Metadata reveals input		Fig. 3.4a with published output
	0	1	0	1				Fig. 3.4a
	0	1	1	0		As Fig. 3.3a with published output		Fig. 3.4b with published output
	0	1	1	1		Fig. 3.3a		Fig. 3.4b
	1	0	0	0		Fig. 3.3b with published output		Revealing the input to the DB contradicts with the data owner's will to hide data from DB
	1	0	0	1		❶ – Fig. 3.3b (see §2.2)		
	1	0	1	0		Public input ⇔ secret metadata		
	1	0	1	1				
	1	1	0	0		Metadata reveals input		Fig. 3.4c with published output
	1	1	0	1				Fig. 3.4c
	1	1	1	0		Fig. 3.3c with published output		Fig. 3.4d with published output
	1	1	1	1		Fig. 3.3c		Fig. 3.4d

Table 1: Summary of the identified scenarios and their ideal solutions (figure numbers refer to figures of the dissertation), including the contradictory cases (denoted with grey), the trivial ones (green), the ones that are interesting from the viewpoint of cryptography (light and dark orange). Numbered scenarios are the relevant ones for the described new results.

PARAMETERS: participants P_1, P_2, a class $\mathcal{F} = \{f : \mathcal{X} \to \mathcal{Y}\}$ of deterministic functions *[and an integer $\kappa > k$]*

FUNCTIONALITY:
On inputs $x_1, \ldots, x_d \in \mathcal{X}$ and $\mathcal{F}_A \subset \mathcal{F}$ from P_1; and $\mathcal{F}_B = \{f_1, \ldots, f_k\} \subset \mathcal{F}$ from P_2

- P_1 receives no output, *[or P_1 receives $\mathcal{F}_R$ s.t. $\mathcal{F}_B \subset \mathcal{F}_R \subset \mathcal{F}$ and $|\mathcal{F}_R| = \kappa$]*
- P_2 obtains $\{y'_{i,j} = f'_j(x_i)\}_{i\in[d],j\in[k]} \subset \mathcal{Y} \cup \{\perp\}$ for

$$f'_j(x_i) = \begin{cases} f_j(x_i) & \text{if } f_j \notin \mathcal{F}_A \\ \perp & \text{otherwise.} \end{cases}$$

Figure 2: Ideal functionalities for $\mathcal{F}_{\mathbf{CPFE}}$ formulated generally for multiple inputs and multiple functions. The extensions in brackets lead to the ideal functionality $\mathcal{F}_{\mathbf{rCPFE}}$ of rCPFE that guaranties a κ-anonimity type function privacy [SS98].

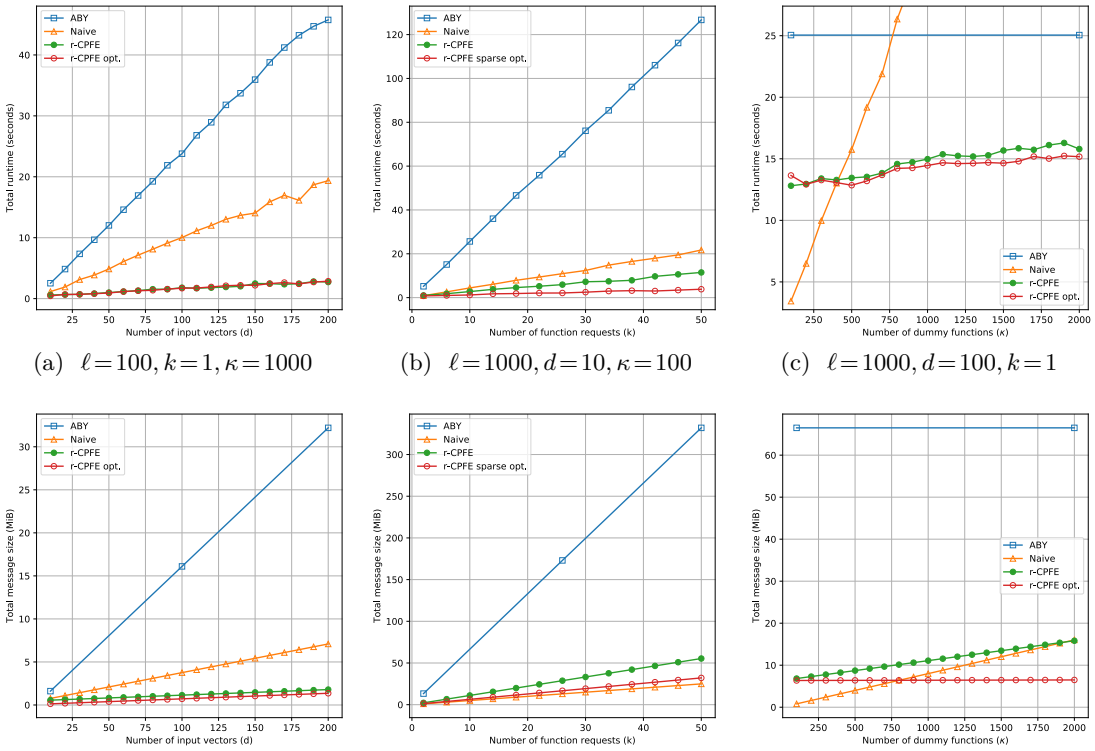
evaluation protocol. At the same time, the resulting protocol highlights the importance of meaningful relaxations to the security requirements to make the concept practical.

THESIS 1.3: *I proposed a protocol for generic functions, that enables the reusability of the protocol messages in case of multiple function evaluations. As a result, when evaluating the same function(s) on multiple, say d inputs, the communication and online computation overhead is additive and proportional to d compared to evaluation(s) on a single input. This is in contrast to the multiplicative overhead in case of traditional PFE. I proved that the proposed protocol fulfils the rCPFE security requirements as long as the underlying primitives satisfy standard security requirements.*

The idea of the generic rCPFE realisation is that Alice can send the inputs to be used to Bob in an encrypted form if she can carefully restrict the decryptability of these ciphertexts. Functional encryption (FE) enables exactly this, as a functional secret key, issued for a specific function, enables special decryption that is integrated with function evaluation. Decrypting a ciphertext corresponding to x, with a functional key for f reveals $f(x)$ but nothing more. The challenge of applying FE is to resolve the availability of proper functional keys to Bob, which is solved with the help of oblivious transfer (OT). The precise statement about the security and efficiency of the protocol can be found in Theorem 4.3.2 and Corollaries 4.3.2.1 and 4.3.2.2 of the dissertation together with the security proof.

THESIS 1.4: *The proposed rCPFE protocol is practical when instantiated for the inner product functionality. I demonstrated this via a proof of concept implementation and optimizations together with a comparison with the state of the art secure arithmetic inner product computation method.*

Instantiating the proposed rCPFE protocol with the k-NA-SIM secure inner product FE scheme of [ALS16] and the semi-honest 1 out of κ OT protocol of [Tze04] led to an inner product rCPFE protocol with security under the decisional Diffie–Hellman assumption (DDH). This instantiation was analysed together with two possible optimizations: one utilizing an opportunity for precomputation, and another that assumes that one of the vectors in the inner product is sparse. The performance of these protocols were compared with a naive OT-based portocol, and with the state of the art ABY framework [DSZ15]. The results are depicted in Fig. 3 and further elaborated in §4.4 of the dissertation.



(a) $\ell = 100, k = 1, \kappa = 1000$ (b) $\ell = 1000, d = 10, \kappa = 100$ (c) $\ell = 1000, d = 100, k = 1$

(d) $\ell = 100, k = 1, \kappa = 1000$ (e) $\ell = 1000, d = 10, \kappa = 100$ (f) $\ell = 1000, d = 100, k = 1$

Figure 3: Comparisons of the overall running times (3a–3c) and communication costs (3d–3f) of our rCPFE protocols with the ABY framework [DSZ15] and a naive OT-based approach for inner product computation (ℓ denotes vector dimension, d and k are the number of input and "function" vectors, while κ is the number of dummy vectors). For our experiments we used a commodity laptop with a 2.60GHz Intel® Core™ i7-6700HQ CPU and 4GB of RAM.

2.3 User Revocation in Multi-Authority Attribute-Based Encryption

Attribute-based encryption (ABE) [SW05] is intended for one-to-many encryption in which ciphertexts are encrypted for those who are able to fulfil certain requirements. In so-called ciphertext-policy attribute-based encryption (CP-ABE), ciphertexts are associated with access policies, determined by the encryptor, and attributes describe the user, accordingly attributes are embedded in the users' secret keys that are issued by attribute authorities. A ciphertext can be decrypted by someone if and only if, his or her attributes satisfy the access policy given in the ciphertext. Note that using ABE, data sharing is possible without prior knowledge of who will be the receiver that can preserve the flexibility of the cloud even after encryption.

Flexible identification of user groups has its own price that we have to pay when an individual user has to be identified. The typical example, when we have to do this is user revocation. In everyday use, a tool for changing a user's rights is essential as unexpected events may occur and affect these. An occasion when someone has to be revoked can be dismissal or the revealing of malicious activity. Revocation is especially hard problem in ABE, since different users may hold functionally the same secret keys related with the same attribute set (aside from randomization).

THESIS 2.1: *I proposed a game based security model to assess* **IND-CPA** *security of multi-authority CP-ABE cryptosystems that support direct, i.e. identity based, user revocation.*

In the context of multi-authority CP-ABE systems identity-based revocation was first considered in [C5]. In order to argue security, one needs a model capturing the capabilities of an attacker. I followed a game based approach where security is defined through a *security game* between an attacker algorithm $\mathcal{A}$ and a challenger. We assume that adversaries can corrupt authorities only statically but key queries are made adaptively. The definition reflects the scenario where all users in the revocation list (RL) get together and collude (this is because the adversary can get all of the private keys for the revoked set). Informally, $\mathcal{A}$ can determine a set of corrupted attribute authorities, ask for any identity and attribute keys and specify messages, on which it will be challenged using the revocation list and access matrix of its choice. The only (natural) restriction in the above choices is that $\mathcal{A}$ cannot ask for a set of keys that allow decryption, in combination with any keys that can be obtained from corrupt authorities in case of a non-revoked global identifier (GID). In case of revoked identities, we can be less restrictive: corrupted attributes alone cannot satisfy the access policy, but it might be satisfied together with attributes from honest authorities. $\mathcal{A}$ wins the game if it respects the rules and can decide which of its challenge messages were encrypted by the challenger. The formal security game consists of the following rounds:

Setup. The challenger $\mathcal{C}$ runs the setup algorithm of the scheme to obtain the global public parameters GP. $\mathcal{A}$ specifies a set $AA' \subseteq AA$ of corrupt attribute authorities and uses the corresponding authority setup algorithm to obtain public and

private keys. For honest authorities in $AA \setminus AA'$ and for the central identity issuer authority, $\mathcal{C}$ obtains the corresponding keys by running the corresponding setup algorithms, and gives the resulting public keys to the attacker.

Key Query Phase. $\mathcal{A}$ adaptively issues private key queries for identities GID_k (which denotes the kth *GID* query). The challenger provides to $\mathcal{A}$ with the corresponding identity keys. Let UL denote the set of all queried GID_k. $\mathcal{A}$ also makes attribute key queries by submitting pairs of (i, GID_k) to the challenger, where i is an attribute belonging to a honest authority. The challenger responds by giving the attacker the corresponding attribute secret key.

Challenge. $\mathcal{A}$ gives $\mathcal{C}$ two messages $\mathfrak{m}_0, \mathfrak{m}_1$, a set $RL \subseteq UL$ of revoked identities and an access policy $\mathcal{P}$.

RL and $\mathcal{P}$ must satisfy the following constraints. Let V denote the subset of attributes controlled by corrupt authorities. For each identity $GID_k \in UL$, let V_{GID_k} denote the subset attributes i for which $\mathcal{A}$ has queried (i, GID_k). For any $GID_k \in UL \setminus RL$, attributes in $V \cup V_{GID_k}$ should not satisfy $\mathcal{P}$, while for $GID_k \in RL$, only attributes in V should not satisfy $\mathcal{P}$.

The attacker must also give the challenger the public keys for any corrupt authorities whose attributes appear in $\mathcal{P}$

The challenger flips a random coin $\beta \in (0, 1)$ and sends the attacker an encryption of M_β under access policy $\mathcal{P}$ with the revoked set *RL*.

Key Query Phase 2. The attacker may submit additional attribute key queries (i, GID_k), as long as they do not violate the constraint on the challenge revocation list *RL* and policy $\mathcal{P}$.

Guess. $\mathcal{A}$ must submit a guess β' for β. The attacker wins if $\beta' = \beta$. The attacker's advantage in this game is defined to be $\mathbb{P}(\beta' = \beta) - \frac{1}{2}$.

Definition 1. *We say that a multi-authority CP-ABE system with identity-based revocation is chosen-plaintext secure against static corruption of attribute authorities if, for all revocations sets RL of size polynomial in the security parameter, all polynomial time adversary $\mathcal{A}$ has at most a negligible advantage in the above defined security game.*

THESIS 2.2: *I proposed a revocable multi-authority ciphertext-policy attribute-based encryption scheme and I proved that it is secure in the generic bilinear group and random oracle models.*

In [C5, J2], I built upon the results of [LW11] to construct a CP-ABE scheme where multiple, independent attribute authorities can issue attribute keys and in which direct revocation of specific users with all of their attributes is possible without updating attribute public and secret keys (neither periodically, nor after revocation event). With this method, in the proposed cryptosystem it is possible to avoid the expensive re-encryption of all ciphertexts the access policy of which contain a subset of attributes

of a revoked user. The revocation right can be given directly to the encryptor, just like the right to define the access structure which fits to the cloud computing and data market scenarios. To realize the targeted features, I used ideas from public key broadcast encryption systems [LSW10]. The ciphertexts are constructed in a way that they contain a special secret sharing. Reconstructing the secret is necessary for decryption, but any user with a revoked identity GID_k^* will not be able to incorporate the kth share and thus cannot decrypt the message. This approach presents the following challenges. First, it must be ensured that every decryptor needs to compare his or her GID with the ones in the revocation list RL even if his or her attributes satisfy the access policy of the ciphertext. Second, we need to make sure that no revoked user with GID_k^* can obtain any information about share s_k. Third, we need to worry about collusion attacks between multiple revoked users. The solution for these challenges and the proposed construction are described in §5.3, while the security proof of the scheme is given in §5.4 of the dissertation.

2.4 Searchable Symmetric-Key Encryption for Restricted Search

The concept of searchable symmetric-key encryption (SSE) allows the secure storage of sensitive data on untrusted servers in the cloud without losing all the flexibility that plaintext data would allow. More precisely it supports keyword search over the ciphertexts in the following way: encrypted queries called trapdoors can be sent to the server which can test whether any of the stored ciphertexts matches the keyword underlying the trapdoor.

The two natural approaches towards realizing SSE are called "forward" and "inverted index". The first one is to attach (or even include) one-way mappings of searchable keywords to the encrypted data. This leads to linear search complexity in the number of documents as the server has to go through all of them with a sequential scan to find all the matches for a trapdoor. A more sophisticated arrangement of the ciphertexts is to build an "inverted index". In this case, the documents (or their IDs) are sorted based on the one-way mappings of keywords which are related to them. The latter solution allows logarithmic search complexity in the number of keywords. This clear benefit caused that the inverted index approach became prevalent in SSE design [PCY+17]. At the same time, these solutions are rather complex and while operating smoothly on huge *static* databases, handling the *rapid expansion* of the database turns out to be more troublesome as the underlying data structure has to be updated without information leakage (see Table 2 for details). [C4, J1] and §6 of the dissertation are dedicated to the following question:

What is the best suitable method for realizing encrypted search when the encrypted database is rapidly growing but it is enough to search parts of the database?

The inverted index approach, optimized for efficient search, is less beneficial in this case, than the less explored forward index method.

THESIS 3.1: *I proposed a game based security model to assess the* **IND-CKA2** *security of forward index searchable symmetric-key encryption cryptosystems.*

The commonly used security model for SSE was defined by [CGKO06] to capture the intuition that, in the course of using the scheme, the remotely stored files and search queries together do not leak more information about the underlying data than the search pattern and the search outcome. The proposed security definition follows this intuition, but it is formulated in the context of a forward index.

In the security game, the adversary has to recognize which one of two challenge datasets (consisting of messages and their keywords chosen by herself) was encrypted by the challenger. Note that in a forward index even the knowledge of the order of ciphertexts can help the attacker, that is why our challenger provides her with a random permutation of ciphertexts prepared from the randomly chosen challenge message set. The adversary has access not only to the encryptions themselves but also to a trapdoor generation oracle that can be queried adaptively with pairs of keywords corresponding to the two challenge sets. The oracle answers consistently with a trapdoor for that keyword which belongs to the encrypted challenge data set. The only restriction is that the queried keywords cannot separate the two challenge sets, as we are interested in information leakage beyond the search result.

For the ease of exposition, we can assume that there is a single keyword for each message but this can be easily generalized. The formal definition of indistinguishability under adaptive chosen keyword attack (**IND-CKA2**) for forward index SSE is the following:

Definition 2 (**IND-CKA2** security)**.** *Let SSE = (**Setup**, **Enc**, **TrpdGen**, **Dec**, **Test**) be a secret-key searchable encryption scheme, $\lambda \in \mathbb{N}$ a security parameter, and $\mathcal{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_{q+1})$ a non-uniform adversary. Consider* $\mathbf{IND\text{-}CKA2}_{\mathbf{SSE},\mathcal{A}}(\lambda)$*, the probabilistic experiment depicted on Fig. 4 with the restriction that the number of keyword matches between the challenge message sets and the corresponding trapdoor queries are equal, i.e.*

$$\#\{i|\hat{w}_j^0 = w_i^0 \text{ for } j \in [k]\} = \#\{i|\hat{w}_j^1 = w_i^1 \text{ for } j \in [k]\}$$

for all $k = 1, \ldots, q$, where q is some polynomial of the security parameter λ. We say that an SSE scheme is secure in the sense of adaptive indistinguishability if for all polynomial-time adversaries $\mathcal{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_{q+1})$,

$$\Pr(\mathbf{IND\text{-}CKA2}_{\mathbf{SSE},\mathcal{A}}(\lambda) = 1) \leq \frac{1}{2} + \mathsf{negl}(\lambda).$$

IND-CKA2$_{\mathbf{SSE},\mathcal{A}}(\lambda)$ Security Game

$\mathsf{sk} \leftarrow\!\!\$\ \mathbf{SSE.Setup}(1^\lambda)$
$b \leftarrow\!\!\$\ \{0,1\}$
$(\mathsf{state}_{\mathcal{A}_0}, D^0, D^1) \leftarrow \mathcal{A}_0(1^\lambda)$
parse D^b as $\{(m_1^b, w_1^b), \ldots, (m_n^b, w_n^b)\}$
for $1 \leq j \leq n$,
  $C_i^b \leftarrow\!\!\$\ \mathbf{SSE.Enc}(\mathsf{sk}, m_i^b, w_i^b)$,
$C^b := (C_{\pi_1}^b, \ldots, C_{\pi_n}^b)$ for a random permutation π,
for $1 \leq j \leq q$,
  $(\mathsf{state}_{\mathcal{A}_j}, \hat{w}_j^0, \hat{w}_j^1) \leftarrow \mathcal{A}_j(\mathsf{state}_{\mathcal{A}_{j-1}}, C^b, \{T_i^b\}_{i\in[j]})$
  $T_j^b \leftarrow\!\!\$\ \mathbf{SSE.TrpdGen}(\mathsf{sk}, \hat{w}_j^b)$
$b' \leftarrow \mathcal{A}_{q+1}(\mathsf{state}_{\mathcal{A}_q}, C^b, \{T_j\}_{i=1,\ldots,q})$
return $b = b'$

Figure 4: **IND-CKA2** security game for forward index SSE schemes.

THESIS 3.2: *Based on* **IND-CPA** *secure symmetric-key encryption and* **EU-CMA** *secure message authentication, I proposed an* **IND-CKA2** *secure searchable symmetric-key encryption in the standard model, under the so-called SXDH assumption.*

The intuition behind the proposed construction is fairly simple. The trapdoors for a specific keyword and the keyword related ciphertext components are constructed in a symmetric manner: both are randomised message authentication codes (MACs) of the underlying keyword, however, represented in distinct groups $\mathbb{G}_1$ or $\mathbb{G}_2$. This enables equality testing by "mixing" the ciphertext and the trapdoor in two different ways (using the pairing operation) that are equal only if the underlying keywords are the same. Using distinct groups prevents the testability both among ciphertexts and among trapdoors. In more detail, the algorithms are described in §6.3.3 of the dissertation, while Theorem 6.4.1 and its proof show **IND-CKA2** security of the scheme as long as the symmetric external Diffie–Hellman assumption (SXDH) holds. For a comparison with other SSE schemes, see Table 2.

3 Application of Results

The brief summary of results is closed by a short description of the applications of the new results in data markets. For other application areas, see §4.1.2, §5.1.2, and §6.1.2 of the dissertation. Next, circled numbers refer to the corresponding scenarios of Table 1.

Scenario ❶. Let us assume that a DB periodically collects location-based information from DOs in vector form, where vector elements correspond to information related

Scheme	Model	Security	Fw/Bw Privacy	Update Complexity	Update Privacy	Search Complexity
[SWP00]	Standard	**IND-CPA**	× / ×	$O(b)$	×	$O(n \cdot b)$
[vLSD+10]	Standard	**IND-CKA2**	× / ×	$O(w_D)^*$	×	$O(\log W)$
[KPR12]	ROM	**CKA2**	× / ×	$O(w_D)$	×	$O(n_w)$
[KP13]	ROM	**CKA2**	× / ×	$O(\log n)^*$	✓	$O(n_w \log n)$
[CJJ+14]	ROM	**CKA2**	× / ×	$O(w_D + W \log n)$	×	$O(n_w + a + d)$
[SPS14]	ROM	**CKA2**	✓ / ×	$O(w_D \log(nW))^*$	✓	$O(n_w + d)$
[HK14]	ROM	**CKA2**	× / ×	$O(nw_D/W)^{**}$	×	$O(nw_D/W)^{**}$
[YG15]	ROM	**CKA2**	✓ / ✓	$O(W)$	✓	$O(W)$
[Gaj16]	Standard	**IND-CKA2**	× / ×	$O(w_D \cdot W)$	×	$O(\log n)$
[KKL+17]	ROM	**CKA2**	✓ / ×	$O(w_D)^*$	✓	$O(n_w)^*$
Our scheme	Standard	**IND-CKA2**	✓ / ✓	$O(w_D)$	✓	$O(n \cdot w_D)$

Table 2: Comparison of our results and dynamic SSE Schemes (n denotes the number of documents (data entries), w_D is the number of keywords per a specific document, W is the total number of distinct keywords in the database, n_w is the number of documents matching the searched keyword w, a is the total number of additions to the database and d is the total number of deletions, b is the bit length of encrypted documents. * indicates that update requires some rounds of interaction between the server and the client and ** denotes amortized complexity).

to specific positions. Such data can be important for VASPs, offering location-based services, without the proper infrastructure to collect the necessary data. During their interaction that can be an inner product computation[4], the VASP should hide the location of its users, while the DB may want to protect the exact information in specific locations or to adjust higher price if specific measurements are used. These can be achieved by having control over the possible queries of a VASP using the proposed inner product rCPFE protocol.

Scenario ❷. DO can control the access rights to his or her data through the use of CP-ABE. Having a central role in the data market, the DB is capable issuing identity keys for the VASPs upon sign up to the data market. Maintenance and publication of a revocation list can also be part of the DB's responsibilities. Attribute authorities can be run by any entities, that are independent of the market (i.e. has no interest of obtaining data, sold in the market) and can be trusted by the DOs to provide authentic information about the VASPs. Examples for such entities may include regulatory bodies, certification authorities, but also non-governmental organizations, consumer protection offices, etc. The proposed revocable multi-authority CP-ABE enables DOs to encrypt their data without prior knowledge of who exactly will decrypt it. At the same time, constraints on the possible decryptor can be determined in encryption time (through the access control policy) based on attributes of VASPs, which were recognised by independent

[4]For example, multiplying the data vector with a position vector (that is non-zero in all positions representing locations close to the user – possibly containing weights depending on the distance – and zero otherwise) can give useful information.

authorities.

Scenario ❸. Assume that metadata consists of multiple records that describes the data in different granularity. In this case, the most general information is possibly not sensitive, e.g. a time frame when the data was recorded, and the VASP would only want to hide more specific informations of the bought data. In that case, only a portion of the database has to be searched that is possible after the following extension of the mechanism, sketched above regarding scenario ❷. DOs encrypt fine-grained metadata of their data as well, using SSE. The SSE keys of a DO are encrypted with ABE just like the symmetric key (used for data encryption) with the only difference that the SSE keys are not new for every data entry but used for a longer time. Before accessing some data, the VASPs have to obtain[5] the SSE key of a certain DO to be able to create a trapdoor, that allows finding the wished entry in the encrypted database without revealing to the DB, what exactly was accessed.

[5]Of course, this is only possible if the access policy of the DO, allows the VASP to search his or her data.

List of Publications

Conference and Workshop Publications

[C1] Gergely Biczók, Máté Horváth, Szilveszter Szebeni, István Lám and Levente Buttyán. The cost of having been pwned: A security service provider's perspective. *The 3rd International Workshop on Emerging Technologies for Authorization and Authentication (Co-Located with ESORICS 2020) – ETAA 2020.*

[C2] Máté Horváth, Levente Buttyán, Gábor Székely and Dóra Neubrandt. There Is Always an Exception: Controlling Partial Information Leakage in Secure Computation. *J. H. Seo (Ed.): Information Security and Cryptology - ICISC 2019, LNCS 11975, pp. 133–149, 2020.*

[C3] Máté Horváth, Levente Buttyán. Problem Domain Analysis of IoT-Driven Secure Data Markets. In: E. Gelenbe et al. (eds) *Security in Computer and Information Sciences. Euro-CYBERSEC 2018. Communications in Computer and Information Science*, vol 821. pp. 57–67, Springer, 2018.

[C4] Máté Horváth, István Vajda. Searchable symmetric encryption: Sequential scan can be practical. *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp 1–5, 2017.

[C5] Máté Horváth. Attribute-Based Encryption Optimized for Cloud Computing. G.F. Italiano et al. (eds), *SOFSEM 2015: Theory and Practice of Computer Science – 41st International Conference on Current Trends in Theory and Practice of Computer Science, Pec pod Sněžkou, Czech Republic, January 24–29, 2015. Proceedings*, *Lecture Notes in Computer Science* vol. 8939, pp. 566–577, Springer, 2015.

[C6] Máté Horváth. Private Key Delegation in Attribute-Based Encryption. *Mesterpróba – Conference for last year MSc and first year PhD students*, pp. 21–24, 2015.

Journal Publications

[J1] Máté Horváth, István Vajda. Searchable Symmetric Encryption for Restricted Search. *Journal of Communications Software and Systems*, 14(1):104–111, 2018.

[J2] Máté Horváth. Attribute-Based Encryption Optimized for Cloud Computing. *Infocommunications Journal*, 7(2):1–9, 2015.

Book

[B1] Máté Horváth, Levente Buttyán. Cryptographic Obfuscation: A Survey. *Springer Briefs in Computer Science, ISBN 978-3-319-98040-9.* Springer, 2020. Manuscript is available here: `http://eprint.iacr.org/2015/412`.

References

[ALS16] Shweta Agrawal, Benoît Libert, and Damien Stehlé. Fully secure functional encryption for inner products, from standard assumptions. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016, Proceedings, Part III*, volume 9816 of *Lecture Notes in Computer Science*, pages 333–362. Springer, 2016.

[Bar16] Boaz Barak. Lecture notes from the Cryptography course at Harvard University, Spring 2016. https://intensecrypto.org/public/lec_17_SFE.html, Accessed 4. October 2020.

[CGKO06] Reza Curtmola, Juan A. Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors, *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006*, pages 79–88. ACM, 2006.

[CJJ+14] David Cash, Joseph Jaeger, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Dynamic searchable encryption in very-large databases: Data structures and implementation. In *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*, 2014.

[CMM+20] Gabriella Cattaneo, Giorgio Micheletti, Chrysoula Mitta, Mike Glennon, and Carla La Croce. *The European data market monitoring tool, Key facts & figures, first policy conclusions, data landscape and quantified stories: d2.9 final study report.* Publications Office of the EU, July 2020.

[Dat] Data Market Austria Project. https://datamarket.at/en/, Accessed 4. October 2020.

[Dat17] Datum Network GmbH. Datum Network - The decentralized data marketplace (White Paper V15). Technical report, Datum Network GmbH., 2017. https://datum.org/, Accessed 4. October 2020.

[DSZ15] Daniel Demmler, Thomas Schneider, and Michael Zohner. ABY - A framework for efficient mixed-protocol secure two-party computation. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015*. The Internet Society, 2015.

[EN16] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communi-*

cations Security, Vienna, Austria, October 24-28, 2016, pages 1388–1401. ACM, 2016.

[Gaj16] Sebastian Gajek. Dynamic symmetric searchable encryption from constrained functional encryption. In *Topics in Cryptology - CT-RSA 2016 - The Cryptographers' Track at the RSA Conference 2016, San Francisco, CA, USA, February 29 - March 4, 2016, Proceedings*, pages 75–89, 2016.

[HK14] Florian Hahn and Florian Kerschbaum. Searchable encryption with secure and efficient updates. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pages 310–320, 2014.

[IOT] IOTA Data Marketplace. https://data.iota.org/, Accessed: 4. October 2020.

[KKL+17] Kee Sung Kim, Minkyu Kim, Dongsoo Lee, Je Hong Park, and Woo-Hwan Kim. Forward secure dynamic searchable symmetric encryption with efficient updates. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017*, pages 1449–1463. ACM, 2017.

[KP13] Seny Kamara and Charalampos Papamanthou. Parallel and dynamic searchable symmetric encryption. In *Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers*, pages 258–274, 2013.

[KPR12] Seny Kamara, Charalampos Papamanthou, and Tom Roeder. Dynamic searchable symmetric encryption. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012*, pages 965–976, 2012.

[LSW10] Allison Lewko, Amit Sahai, and Brent Waters. Revocation systems with very small private keys. In *IEEE Symposium on Security and Privacy*, pages 273–285, 2010.

[LW11] Allison Lewko and Brent Waters. Decentralizing attribute-based encryption. In *Advances in Cryptology–EUROCRYPT 2011*, pages 568–588. Springer, 2011.

[Oce] Ocean Protocol Foundation Ltd. A decentralized data exchange protocol to unlock data for artificial intelligence (technical primer). https://oceanprotocol.com/, Accessed 4. October 2020.

[OLJ+19] Boris Otto, Dominik Lis, Jan Jürjens, Jan Cirullies, Falk Howar, Sven Meister, Markus Spiekermann, Heinrich Pettenpohl, Frederik Möller, Jakob Rehof, and Sebastian Opriel. Data ecosystems. conceptual foundations, constituents and recommendations for action. Technical report, Fraunhofer Institute for Software and Systems Engineering ISST, 10 2019.

[PCY+17] Geong Sen Poh, Ji-Jian Chin, Wei-Chuen Yau, Kim-Kwang Raymond Choo, and Moesfa Soeheila Mohamad. Searchable symmetric encryption: Designs and challenges. *ACM Comput. Surv.*, 50(3):40:1–40:37, 2017.

[SPS14] Emil Stefanov, Charalampos Papamanthou, and Elaine Shi. Practical dynamic searchable encryption with small leakage. In *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*, 2014.

[SS98] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. techreport, SRI-CSL-98-04, SRI Computer Science Laboratory, 1998. Technical report.

[SW05] Amit Sahai and Brent Waters. Fuzzy identity-based encryption. In *Advances in Cryptology–EUROCRYPT 2005*, pages 457–473. Springer, 2005.

[SWP00] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *2000 IEEE Symposium on Security and Privacy, Berkeley, California, USA, May 14-17, 2000*, pages 44–55, 2000.

[Tze04] Wen-Guey Tzeng. Efficient 1-out-of-n oblivious transfer schemes with universally usable parameters. *IEEE Trans. Computers*, 53(2):232–240, 2004.

[vLSD+10] Peter van Liesdonk, Saeed Sedghi, Jeroen Doumen, Pieter H. Hartel, and Willem Jonker. Computationally efficient searchable symmetric encryption. In *Secure Data Management, 7th VLDB Workshop, SDM 2010, Singapore, September 17, 2010. Proceedings*, pages 87–100, 2010.

[YG15] Attila Altay Yavuz and Jorge Guajardo. Dynamic searchable symmetric encryption with minimal leakage and efficient updates on commodity hardware. In *Selected Areas in Cryptography - SAC 2015 - 22nd International Conference, Sackville, NB, Canada, August 12-14, 2015, Revised Selected Papers*, pages 241–259, 2015.