Cryptographic Tools for Privacy Preservation and Verifiable Randomness

CARLO BRUNETTA

Division of Network and System
Department of Computer Science & Engineering
Chalmers University of Technology and Gothenburg University
Gothenburg, Sweden, 2018
Cryptographic Tools for Privacy Preservation and Verifiable Randomness

CARLO BRUNETTA

Copyright ©2018 Carlo Brunetta
except where otherwise stated.
All rights reserved.

Technical Report No 188L
ISSN 1652-876X
Department of Computer Science & Engineering
Division of Network and System
Chalmers University of Technology and Gothenburg University
Gothenburg, Sweden

This thesis has been prepared using \LaTeX.

Printed by Chalmers Reproservice,
Gothenburg, Sweden 2018.
“Whoever controls the media, controls the mind.”
(Unsourced) - James Douglas Morrison
Abstract

Our society revolves around communication. The Internet is the biggest, cheapest and fastest digital communication channel used nowadays. Due to the continuous increase of daily communication among people worldwide, more and more data might be stolen, misused or tampered. We require to protect our communications and data by achieving privacy and confidentiality.

Despite the two terms, “privacy” and “confidentiality”, are often used as synonymous, in cryptography they are modelled in very different ways. Intuitively, cryptography can be seen as a tool-box in which every scheme, protocol or primitive is a tool that can be used to solve specific problems and provide specific communication security guarantees such as confidentiality. Privacy is instead not easy to describe and capture since it often depends on “which” information is available, “how” are these data used and/or “who” has access to our data.

This licentiate thesis raises research questions and proposes solutions related to: the possibility of defining encryption schemes that provide both strong security and privacy guarantees; the importance of designing cryptographic protocols that are compliant with real-life privacy-laws or regulations; and the necessity of defining a post-quantum mechanism to achieve the verifiability of randomness.

In more details, the thesis achievements are:

(a) defining a new class of encryption schemes, by weakening the correctness property, that achieves Differential Privacy (DP), i.e., a mathematically sound definition of privacy;

(b) formalizing a security model for a subset of articles in the European General Data Protection Regulation (GDPR), designing and implementing a cryptographic protocol based on the proposed GDPR-oriented security model, and;

(c) proposing a methodology to compile a post-quantum interactive protocol for proving the correct computation of a pseudorandom function into a non-interactive one, yielding a post-quantum mechanism for verifiable randomness.

Keywords

Cryptography, Confidentiality, Privacy, Differential Privacy, GDPR, Verifiable Randomness
Acknowledgment

I would like to thank **Katerina**. We may have had some issues in the beginning but we manage to solve them. Without you I would not be here. Thank you for giving me this opportunity to grow as a human and as a researcher.

Thank you to all the people in **my division**, Network and System, for sharing the good and bad moments of the daily work. I would really love to thank all the **administration**

**“moms and dads”** for all the **support** that they give me either if it’s “**work related**” or “it’s a blue day”. **Tack <3**

A big thank you to the uncountable number of **friends** that crossed my life here in Chalmers. Thank you for all the good fika, the beers and the infinite discussions.

I would love to thanks all my Sahlgrenska’s real-science friends such as **Tugce, Lydia, Jasmine, Eleni, Axel, Giacomo and many more** for the good moments outside the Chalmers walls!

Thank you, **Bei and Georgia**, for the research discussions and shared funny moments. I am open to continue our future collaborations and our friendship!

A special thank you goes to **Pablo, Lara and Oliver**. I wish you all the best, always!

We shared a lot and we changed the band-name a lot of times. Thank you **Marco, Enzo, Pier, Evgenii and Grischa** for all the good bohemian-moments and the jams. You are like a family to me and even if we will have to split, our band will always be a beautiful memory in my music-career. I think that “**Band with Many Names**” is the best name we had!

A heart-graxie goes to **Elena**. You mean a lot to me and you were here, next to me in my **crazy journey**. I’m really lucky to know you and call you a Friend. I will always be grateful to you and I hope the best for you, anywhere on Earth. (==)

I know I shouldn’t cite you if there’s not a good impact-factor but thank you, **Eridan**! “**Rolling From Teh Kitchen**” is my family. Glad to have you in it :)

Looking forward to all the pizzas, films and tv-series!

My biggest thank you goes to my love, **Aura**. You are my precious “**stella alpina**”.

**You are the reason of my smiles and, to me, that is all that matters in life.**

A big thank you to all the rest of the people that might, or not, be present in this too-small page. **No-margin will be able to contain you all!** I think this Pink Floyd quote from **Breathe** can explain what I think about all of you in my daily R-life:

All you touch and all you see,  
is all your life will ever be.

---

**Breathe - Dark Side of the Moon**  
**Pink Floyd**
List of Publications

Appended publications

This thesis is based on the following publications:

**Paper A: C. Brunetta, C. Dimitrakakis, B. Liang, and A. Mitrokotsa**
“A Differentially Private Encryption Scheme”

**Paper B: E. Pagnin, C. Brunetta, P. Picazo-Sanchez**
“HIKE: Walking the Privacy Trail”

**Paper C: C. Brunetta, B. Liang, and A. Mitrokotsa**
“Lattice-Based Simulatable VRFs: Challenges and Future Directions”
Other publications

The following publications were published during my PhD studies, or are currently under submission. However, they are not appended to this thesis.

(a) **C. Brunetta**, M. Calderini, and M. Sala
    “On hidden sums compatible with a given block cipher diffusion layer”
    *Discrete Mathematics (Journal), Vol. 342 Issue 2, 2018* [10]

(b) **C. Brunetta**, B. Liang, and A. Mitrokotsa
    “Strong Functional Signature”
    *Under submission*
Research Contribution

In Paper A, I was involved in the initial brainstorming with Katerina and Christos who proposed me the idea of including differential privacy in the cryptographic domain. I had the idea of relaxing the correctness property of an encryption scheme, the key-idea that allows to define differentially private encryption schemes. I further formalized, defined and proved all the contents of the paper. In the final stage, I wrote the implementation and the statistical tests.

In Paper B, after many fruitful morning-fika and brainstorming with Elena and Pablo (and Oliver!), we all together traced the main structure and motivation for the HIKE protocol. During the development of the paper, I was the relay figure for the translation between theory and implementation. More specifically, I wrote the draft of some proofs and I was responsible for the theoretical aspects necessary for the implementation. Finally, I am the corresponding author of this work and I finalised the camera ready version.

In Paper C, I participated to the initial brainstorming discussion with Bei and Katerina after Bei’s suggestion on the specific topic of constructing a post-quantum verifiable Pseudo Random Function. I completely wrote the first draft of the paper. After receiving some useful external feedback on the paper, I participated in finding different possible solutions while Bei and Katerina revised the draft. In this final and much shorter version, I conceived the summary of the entire research-exploration and I was responsible for the introduction-background sections of the final paper.
# Thesis Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>vii</td>
</tr>
<tr>
<td>List of Publications</td>
<td>ix</td>
</tr>
<tr>
<td>Personal Contribution</td>
<td>xi</td>
</tr>
<tr>
<td>From Caves to the Internet: Privacy and Cryptography</td>
<td>1</td>
</tr>
<tr>
<td>1 Sketchy 3-Sets-Data Privacy Model</td>
<td>4</td>
</tr>
<tr>
<td>2 The Thesis’ Contributions</td>
<td>10</td>
</tr>
<tr>
<td>Paper A - A Differentially Private Encryption Scheme</td>
<td>15</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>18</td>
</tr>
<tr>
<td>2 Preliminaries</td>
<td>22</td>
</tr>
<tr>
<td>3 Our Definition of $\alpha_{m_1,m_2}$-correct Encryption Scheme</td>
<td>23</td>
</tr>
<tr>
<td>4 Equality Between DP-then-Encrypt and Encrypt+DP</td>
<td>28</td>
</tr>
<tr>
<td>5 Example of an $\alpha_{m_1,m_2}$-Correct Homomorphic Encryption Scheme</td>
<td>31</td>
</tr>
<tr>
<td>6 Conclusions &amp; Future Work</td>
<td>34</td>
</tr>
<tr>
<td>Paper B - HIKE: Walking the Privacy Trail</td>
<td>37</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>40</td>
</tr>
<tr>
<td>2 Preliminaries</td>
<td>42</td>
</tr>
<tr>
<td>3 Labelled Elliptic-curve ElGamal (LEEG)</td>
<td>45</td>
</tr>
<tr>
<td>4 FEET: Feature Extensions to LEEG</td>
<td>48</td>
</tr>
<tr>
<td>5 The HIKE protocol</td>
<td>51</td>
</tr>
<tr>
<td>6 Security model and proofs for HIKE</td>
<td>54</td>
</tr>
<tr>
<td>7 Implementation details and results</td>
<td>59</td>
</tr>
<tr>
<td>8 Conclusions and directions for future work</td>
<td>61</td>
</tr>
<tr>
<td>Paper C - Lattice-Based Simulatable VRFs: Challenges and Future Directions</td>
<td>65</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>68</td>
</tr>
<tr>
<td>2 Applying Lindell’s Transformation</td>
<td>70</td>
</tr>
<tr>
<td>3 Translation of Boneh’s PRF</td>
<td>76</td>
</tr>
<tr>
<td>4 Challenges and Future Directions</td>
<td>79</td>
</tr>
<tr>
<td>Bibliography</td>
<td>83</td>
</tr>
</tbody>
</table>
From Caves to the Internet: Privacy and Cryptography

Who are you? Why do you hide in the darkness and listen to my private thoughts?

Romeo and Juliet
William Shakespeare

We are more than animals. We are social animals [43] who need to communicate, socialize and share our feelings, thoughts and ideas to others. Our communication methods evolved during our history and from primitive languages and abstract-paintings, we now use “advanced” languages and complex technologies that allow us to feel “closer” to other people that are not physically close to us.

Nowadays, our life is becoming more and more digital and the Internet is the biggest channel we use to communicate by using e-mails, social networks, blogs, instant messaging, video-calls and many others. Digital communication is cheap, fast and practical for the standard user: sending an email to someone on the other side of the world is just a matter of typing the words on a keyboard, clicking on the “send” button and almost instantaneously the email is sent and received.

What does “digital” mean?

Whenever we write an email, our thoughts are translated into words, and the words are typed into a keyboard which encode them into something easier to transmit. For this reason, the digital-world is dominated by discrete and finite sets of symbols, for example the alphabet in any latin-derived language. There is a finite amount of symbols that are composed in order to create words of which we, as humans, give a special meaning.

One of these finite alphabets is the binary-set with only the symbols 0 and 1. Claude Shannon gave a name to the symbols in this set: the bits. By combining bits, we obtain bit-strings and with these, we can just encode our previous latin-alphabet into bit-strings and share these 0s and 1s. It is possible to encode everything into a bit-string and facilitate its employment into technologies since only two different symbols are transmitted and not, for example, 26 latin-letters! For this and other reasons, the bit is the building block of Information.
Theory [73], the mathematical field that, briefly, studies how information is exchanged during communication. A more tangible physical result of information theory is the computer as the “automatic bits machine” which is able to manipulate, communicate and use pieces of information encoded into bit-strings.

We love computers. They make our life easier and allow us to digitally communicate with anybody at any time. Research on computers is evolving into new directions and one the most promising and breakthrough ideas is building a quantum computer, based on quantum bits, or qubits. As the name suggest, a quantum computer is still a computer but with something more. Quantum computers are expected to offer the biggest performance boost in the history of technology and break the computational limits of current computers. Even if there is not yet any practical quantum computer, different quantum algorithms are ready to be deployed having a strong impact in mathematics, cryptography and many other fields. In particular, Shor’s [74] and Grover’s [40] algorithms will make possible to break down the security of some well-known cryptographic primitives used in our daily life in order to protect our bank accounts or emails. The primitives that will remain secure even when a quantum computer will be used, belong to what is called post-quantum cryptography.

Why should we care about security?

Everyone of us has at least one secret. Maybe it is our PIN code for the credit-card, the password for the email account, our health condition or anything that we do not want to share with too many people. Secrets are important in personal relationships in the sense that if I want to share with a friend a small piece of information that I want to keep secret, then, trust is needed. In other words, I need confidentiality guarantees.

It might sometimes be implicit, but we are used to require privacy or confidentiality when interacting with other people or computers. Is it practical to require “secure communications” or to ask “keep it confidential!”?

But, what does it mean? What is privacy and confidentiality?

What does “secure communication” mean?

Historically, cryptography is a tool-set that allows us to protect our sensitive information and communications [67]. Cryptography evolved around the concept of confidentiality of a message and security of a communication channel. At the basic level, confidentiality and security guarantee that only authorized people can obtain the message and access its authentic contents. With a picturesque metaphor, we can think of cryptography as locks and keys that can be used to lock the chests that contain our gold coins. We know that a specific key will open the chest and we will allow a friend to open the chest by providing him/her the key. Therefore, confidentiality is the guarantee that the lock is so well-made and designed that no pirate can open the locked chest or deduce
the amount of coins in the chest without having access to the key. Nowadays, cryptography is not made with chests, locks and keys. Cryptography is based on mathematics, functions, bits and it is widely used on a daily basis for all the sensitive digital communication we have such as managing our bank accounts, browsing the Internet, storing our health test results, messaging with a friend or in other scenarios not so obvious, e.g., in cars, public transportation, etc.

Cryptography has a pretty neat distinction between secret and public information, so neat that a piece of information cannot be both secret and public at the same time. This concept is quite old, tracing back to the ancient Greece. Aristotle, a 4th-century B.C. philosopher and scientist, wrote a collection of political-philosophical books titled “Πολιτικά” or “Politics”. In this collection, Aristotle explained the need for every politician to a neat separation between the public sphere called πόλις, or “polis”, related to the personal political-life, and a personal sphere called οἶκος, or “oikos”, that contains the family-life.

Aristotle’s idea was the starting concept that developed into privacy. Since his initial spheres’ distinction, the concept of privacy evolved in our society and technology and it can be considered as a modern human invention. In 1890, the journal Harvard Law Review contains a law review titled “The Right To Privacy” that expresses the need of laws that can protect the “right to be alone”. At the time, letters were read by post-employees and/or phones were easily eavesdropped or wiretapped. The World Wars moved even further the necessity to confidential military communication but governments started profiling, identifying and threatening individual people. The mass surveillance phenomenon was rising.

Society, as a whole, has managed to take a strong turn and today, we consider privacy as a form of a personal-right of an individual to selectively express, share information or seclude themselves from the others. In other words, the right of privacy is the human unconditional action of protecting and hiding chosen information to other entities. This means that no one, without the personal permission of an individual, can use, read, store or sell that information. One of the biggest examples of legislative documents, on human’s privacy-rights, is the European General Data Protection Regulation (GDPR) [18], focusing on the right of any European citizen to maintain his/her privacy over the data he/she generates.

Our digitalized society requires privacy, confidentiality and security.

Every day, our data are generated by our devices, such as our own digital fingerprints, stored in big companies servers and used to generate better services, advertising and help us in our daily life. Data are the fuel of our markets but can be used against us. Information is the Power and the Weapon of our digital-era. Our contemporary history shows us the personal damage that can be done when bigger IT-companies, like Facebook [38], or governmental security agencies, like the NSA [39], abuse their information-power. For this reason, societies start
raising questions and the will of achieving the right to be safe intellectually and to protect their own identity by asking:

Can we protect our data? How can we use cryptography to achieve privacy? How is privacy related to confidentiality?

In this licentiate thesis, I address these questions and provide some concrete cryptographic tools that can be deployed in real-life to protect our data.

1 Sketchy 3-Sets-Data Privacy Model

Confidentiality and privacy are similar concepts when considered into the Aristotle’s sphere distinction between private and public information and they are commonly synonymous in our daily discussions with friends. They are indeed interconnected however there are substantial differences. In order to explain how privacy and cryptography differ, let us describe and explain a simplistic model that represent our data and their properties.

Firstly, let us define an order $<$ to compare different data and their “publicity”. For example, let us consider the information $x = \text{“Carlo is a PhD student”}$ and $y = \text{“Carlo’s credit-card PIN is $\star\star\star$”}$. It is pretty clear that $x$ is more public with respect to $y$, therefore we can denote it with $x < y$. Even if sometimes it is easy to decide how to order two pieces of information, it is required to state the axiom:

Informal Axiom 1. The order $<$ is not-objective, in the sense that every person has his/her own order.

The second step is to define three sets to distinguish three categories of data, or levels of publicity:

- **Confidential Data** $\mathcal{C}$ is the set of information that we are not willing to share with anybody because these are sensitive data that can easily hurt us, in some sense. Some examples might be the PIN code of our credit-card or the passwords of our e-mail account.

- **Shared Data** $\mathcal{S}$ is the set of information that is sensitive but we are willing or we need to share for some reason. In this set we can find our personal health-measurements shared with a family doctor, our home address, our annual-income or similar data that will be shared with some specific people/entity but we do not want to disclose that information to the whole world.

- **Public Data** $\mathcal{P}$ is the set of information that is not sensitive and we do not mind to share it with the whole world.
Informal Corollary 1. Since the Informal Axiom 1 holds, the sets are different for each person and the membership problem is ill-defined, i.e., given an information $x$, deciding which set contains $x$ is a non-objective problem.

In different words, the order $<$ and the distinct sets $\mathcal{C}, \mathcal{S}, \mathcal{P}$ are different among different people. We describe this non-objectiveness with the term personal privacy-perception, i.e., the individual perception of publicity. For example, Carlo might find that $x = \text{“my personal email address”}$ is public information, i.e., $x \in \mathcal{P}$, while it might exist someone that thinks that it is just a piece of information that can be shared, i.e., $x \in \mathcal{S}$.

The last point of our construction is describing how data relate with respect to other data. We define the inference/deduction of information as the process that takes as input some set of data $\{x_i\}_{i \in I}$ and outputs a new information $z$, denoted as $\{x_i\}_{i \in I} \rightarrow z$. In a nutshell, imagine that somebody knows that $x = \text{“Carlo loves cooking”}$ and $y = \text{“Carlo is Italian”}$, then she might infer that $z = \text{“Carlo loves Italian restaurants”}$. In real-life, an advertising company will “bet” on $z$ and it will start advertising Italian restaurants to Carlo. This inferred-concept is at the base of all the big advertising companies online.

Informal Axiom 2. Data are always dependent with respect to other data: for every information $z$, there always exists a set of information $\{x_i\}_{i \in I}$ that infers about $z$, i.e., $\{x_i\}_{i \in I} \rightarrow z$.

The “always dependency” axiom is strong and scary but it is exactly what research in advertising, machine learning and other fields, is willing to achieve in a close future. For example, let us imagine a future where an internet search engine will be able to predict Carlo’s research query before he finishes typing the query [41]. This can help Carlo in getting useful advertisements while he is navigating the internet. On the other hand, suppose Carlo is addicted to thai-food, if Carlo’s next predicted advertising describes “the cheapest thai-restaurant” with extremely high probability, Carlo will never be able to beat his addiction.

Additionally, inferring new information might have bigger consequences: suppose the input data $\{x_i\}_{i \in I}$ are public data, we infer $\{x_i\}_{i \in I} \rightarrow z$ and the output $z$ is a shared or a confidential information. This is what we define either as a privacy breach or a security breach. These two breaches differ only on when the breach happens. If the breach happens during the communication phase of the data, then it is a security breach. Otherwise, if it happens after the communication is finished, then it is a privacy breach. We distinguish and define four different types of breaches named direct privacy breach, indirect privacy breach, security breach and fake news phenomenon. The entire 3-Sets-Data privacy model is depicted in Figure 1.
Figure 1: The Sketchy 3-Sets-Data Privacy Model. The black arrow indicates the communication between Carlo and his friend. The red arrows indicate all the possible inferences that are breaches: direct privacy breach from public to shared/private, indirect privacy breach from publishing Carlo’s shared information with his friend, security breach from the communication between Carlo and his friend, fake information from Carlo’s public data infer and publish a wrong inference.

1.1 Direct Privacy Breach

The direct privacy is the property that someone finds hard to infer secret information that he/she didn’t receive. Equivalently, whenever someone is able to infer a less public datum from the ones we gave him/her, this is a direct privacy breach.

When considering a direct privacy breach, we have to better understand how it is possible to have this breach. Let us consider a database in which Carlo has some sensitive information $x$. The database is not publicly available but it is possible to perform queries in order to obtain statistical analysis on the data-points contained, i.e., it is not possible to directly query $x$ but it is allowed to query some function $f$ over some aggregation of sensitive data and obtain a public evaluation $f(x, \cdots)$. For example, Carlo is pretty cautious not to share his birth-date with anybody. On the other hand, Carlo gave his birth-date in the national census form. The government now offers a free-of-charge web-service in which anyone can ask and get statistics over a population. It is therefore possible to get the answer to the query “how many people are born every month” but it is not possible to ask for “the month when Carlo was born”.

In 2006, Dwork et al. [22] presented the concept of Differential Privacy (DP) in which even by cleverly querying the database, it is impossible to infer information about $x$. For example, let us consider two queries with a specific difference that leak information: the first query is “how many people are born every month” while the second is “how many people are born every month except Carlo”. It is easy to understand that Carlo’s birth-month will be the difference of the two
queries. This is a consequence of the Informal Axiom 2, the “always dependence” axiom, because Carlo’s information is always contained in some group, either as the fact that Carlo is participating in a dataset by just sharing his information or simply because it is easy to create groups in statistics, for example “counting with respect to categories” as in “count the number of people born each month”.

For this reason, a direct privacy breach can be an individual breach, if the data used for the inference are just Carlo’s public data. Otherwise, if the data are collected from statistics over groups in which Carlo is (or not) a member, then we have a group breach. The direct privacy breach concept is depicted in Figure 2.

![Figure 2: From Paper A: Individual and group privacy breaches.](image_url)

To avoid this information leakage, Dwork et al. [22] proposed the DP framework that starts by measuring the accuracy of the query \( f \). In other terms, this means that it is necessary to compute “how precisely the query \( f \) can allow some differential static privacy breach” or, more empirically, compute all the possible differences between the statistics over the database. The result is used to define a random variable distribution and, whenever the database has to reply to the query \( f \), a noise value is sampled and summed to the query result. The final result is a noisy output that makes impossible to infer information on a single user’s sensitive data, even when multiple and similar queries are performed to the database.

### 1.2 Indirect Privacy Breach

The indirect privacy is the property that someone will not be able to publish our shared information that he/she has. Equivalently, if someone is malicious, in the sense that he/she wants to hurt us, then a indirect privacy breach happens whenever he/she publishes some shared secret that was not allowed to share with third parties.

The indirect privacy notion is hard to formally define because it highly depends on the trust given to the receiver of the data. Let us consider a specific scenario common in our daily-life. This setting is depicted in Figure 3. Carlo, as a client,
uploads his data to a database in a cloud-service and allows a service-provider to compute some statistical functions on his data.

![Diagram](image)

**Figure 3:** From Paper B: Setting: users send data to a database and enjoy some service.

Despite the pretty simple model, developing a protocol between the parties while achieving indirect privacy, this is not trivial to solve especially with the European General Data Protection Regulation (GDPR) [18] in mind. The GDPR is the new European collections of rules, procedures, rights and obligations that anyone has to follow when handling data from any European citizen.

To give an example, Carlo uploads all of his workout run-sessions and allows a fitness application called Strava [78]. Strava monitors Carlo’s data and provide him some statistical analysis on his workout and suggestions on how to improve. Apart from that, Strava also provides an interactive map in which Carlo can take a look for his running paths from his GPS-data and the path that other runners do but only if they have approved to publicly share their GPS-data with the rest of the users of the application. Unluckily, the option in the Strava application for “sharing the GPS-position of run-sessions” was set to true by default. Therefore, people were sharing their private (and sensitive) information without noticing it. The threats of online-sport social networks and the sensitivity of publishing GPS-data were theoretically studied in 2014 [77] and, in the beginning of 2018, it was possible to infer the life-style of a specific Strava user, i.e., an American soldier running around his secret military base in Afghanistan [42].

Strava’s example demonstrates that handling data in a privacy-preserving way is hard to describe via cryptographic protocols because the receiver do not understand the level of confidentiality that other people give to data. This is the reason why we need specific privacy-laws, such as the GDPR, to protect how other people use the data we generate and, by changing focus from law to cryptography, try to solve the question:

*Is it possible to design privacy-preserving protocols that comply with some privacy-policies, such as the European GDPR?*

### 1.3 Security Breach

A security breach is indeed the leakage of confidentiality and it is therefore connected to the cryptographic property of the crypto-primitives used in communications. Let us consider, Carlo is willing to share his $x_0 = \text{“secret lasagna}$
receipt” with Elena. Simplistically, in order to share it, Carlo and Elena encrypt their messages while communicating, e.g., the encryption of the secret receipt is $\text{Enc}(x_0)$. Since communication is made over a public channel, every eavesdropper can collect all the ciphertexts $\{\text{Enc}(x_i)\}_{i \in I}$ and, if it is possible to infer $\{\text{Enc}(x_i)\}_{i \in I} \rightarrow x_0$, then we have a security breach.

Whenever a cryptographic primitive or protocol is not secure in a general sense, it means that it is somehow possible to have a security breach. To avoid this problem, one of the characteristics that an encryption scheme must have is that it produces ciphertexts that look like chaotic messages without any connection to the original plaintext. This “chaos” is indeed connected to the most important concept in cryptography: randomness. Despite the concept of randomness is easy to understand, when stated as “sample a random value in a set” or “flip a fair-coin”, designing a function that should have a random-like output is not an easy task. In fact, in cryptography we always refer to pseudo-randomness because “it looks random” but in fact “it is not random”. To be precise, such functions with random-looking outputs are called PseudoRandom Functions (PRF). It is therefore of extreme importance to always use proved-secure cryptographic primitives, such as proved-pseudorandom PRFs, in order to avoid security breaches.

Additionally, it might be of vital importance to prove the correct evaluation of a function. Imagine that Carlo needs to prepare a vegan-pizza for one of his friend, Elena, which means that Carlo has to use vegan-cheese. After cooking, Elena will require some guarantees that Carlo’s pizza is indeed vegan. Therefore Carlo, while preparing the pizza, will take some videos as proofs and show them to Elena before eating. Elena will verify, by checking the pizza and the videos, that the correct cheese was used.

This proof-idea can be extended to PRFs into the definition of Verifiable Random Functions (VRFs) which are PRFs that evaluate on pseudo-random output and provide an additional proof used to verify the output is a correct computation of a PRF.

### 1.4 Fake News Phenomenon

A possible problem that can arise is that the deducted results might be incorrect and it can be used in a malicious way by just publicly sharing these fake information which can be connected to a bigger phenomenon identifiable as fake news phenomenon. Publishing fake information may lead to defamation, the act of damaging another person’s reputation by publicly sharing wrong and/or malicious data.

With the help of social-networks and better communications tools, the spread of fake news increased year by year, making it hard to judge the sources of the news and the correctness of information. Our own society is affected by fake news since people changed their political ideas [3] or changed their memories about
the past [72].

It is hard to state how cryptography can help solving the fake-news problem. The only exception is whenever we think of re-designing the whole news-publishing process. In a nutshell, if it is possible to use cryptography to, at least, guarantee the correct handling of photographies by creating tamper-proof and certified photo-cameras, then we might help journalists to provide a “proof of correctness” that allow readers to verify the correct handling of pictures and prove that the photos are indeed real and untampered.

The under submission paper “Strong Functional Signature” by myself, B. Liang and A. Mitrokotsa, provides a cryptographic primitive that can be used to help journalists to provide these proofs. Since it has not been published yet, it is not included in this dissertation.

2 The Thesis’ Contributions

In this thesis, we focus on providing mechanisms that can be employed to combat the different privacy and security breaches. Our goal is to provide new cryptographic tools that help protecting our data and/or prove that cryptography can be used to design more privacy-oriented primitives while maintaining their confidentiality.

2.1 Direct Privacy Breach - Paper A

An encryption scheme has to be correct, in the sense that the decryption of a ciphertext needs to be the original message. When compared to an encryption scheme, a DP mechanism always replies with an almost-correct answer. For this reason, Paper A replies to the following question:


Is there a way to define/construct a differentially private encryption scheme?

The differential privacy mechanism is different from an encryption algorithm while both can be seen as frameworks. Paper A studies the relation between the formal definitions of an encryption scheme and a differential private mechanism and merges them into a single cryptographic primitive.

To achieve this, we relax the encryption scheme’s correctness property. This means that the encryption scheme has to “wrongly decrypt” with some bounded probability, i.e., the decryption of a ciphertext can return a wrong message $m'$ with some probability $\alpha_{m, m'}$ that depends on the original message $m$ and the
2. THE THESIS’ CONTRIBUTIONS

final wrong message $m'$. By knowing these probabilities for all the messages, it is possible to prove that the “faulty” encryption scheme achieves differential privacy.

Additionally, we abstract and prove that using these “faulty” encryption schemes is equivalent of using a correct encryption scheme and a DP mechanism as two separate frameworks.

As a final contribution, an implementation is provided as a proof-of-concept.

2.2 Indirect Privacy Breach - Paper B

The main goal of Paper B is to provide a model/scheme with an implementation designed to provide privacy-guarantees with respect to privacy-policies/regulations, such as the GDPR, that are not always fully described in mathematical formalism. By considering the scenario of Figure 3, the paper answers the following question:

**Question B: HIKE: Walking the Privacy Trail [62]**

Is it possible to design privacy-preserving protocols that comply with some privacy-policies, such as the European GDPR?

We start by selecting some specific articles contained in the GDPR and describe as formal cryptographic properties:

(a) data has to be encrypted when stored;

(b) the user decides to selectively allow third parties to access his/her data; and

(c) the user can always delete his/her data from the database (*right to be forgotten*).

In order to describe the “client, cloud and service provider” model, we use the concept of labelled encryption scheme [4] in which every message, or ciphertext, has a *label* that can be seen as a unique public identifier for that message. With the labels and an algebraic “magic trick”, better described as associativity and commutativity in a group, we are able to define some *decryption token* generated by the client. This allows a service provider to decrypt some specific label-ciphertext. Finally, we exploit the additive homomorphic property in order to allow homomorphic evaluations on the client’s ciphertexts. Additionally, the client is able to generate a single token for a *labelled-program*, which is the homomorphic function to evaluate with a list of labels that are the inputs to the function.
Since the function to evaluate is seen by the clients, the clients can refuse to provide the decryption token and therefore not-disclose their data.

More concretely, we start from the ElGamal encryption scheme [25], we describe the scheme as a labelled encryption scheme called LEEG, expand it with some specific features regarding the decryption token into FEET and finally obtaining the HIKE protocol, depicted in Figure 4, that is then proven secure in the GDPR-oriented security model we defined.

As a final contribution, all our ideas are implemented and our code for the HIKE protocol is publicly available at https://github.com/Pica4x6/HIKE.

![HIKE protocol](image)

**Figure 4:** From Paper B: The HIKE protocol.

### 2.3 Security Breach - Paper C

We start by considering the general problem of achieving a post-quantum version of every cryptographic primitive. We focus on verifiable random functions and in particular on simulatable VRFs (sVRFs). In a nutshell, sVRFs are a family of VRFs in a public parameter security model, such as the common reference string. Paper C provides some directions in order to address the following question:

**Question C: Lattice sVRF: Challenges and Future Directions [12]**

Is it possible to define a post-quantum sVRF?

We proposed the possibility of defining a lattice-based membership-hard with efficient sampling language which can be used to define a lattice-based dual-mode commitment scheme. We partially conjecture the possibility to combine the dual-mode commitment scheme with Libert et al.’s protocol [52] and Lindell’s transformation [54] and obtain an sVRF under post-quantum assumptions.

Given the non-triviality of the task, we raise and identify different open challenges in lattice-based cryptography and possible future directions for achieving a post-quantum sVRF.


